



Organización  
Internacional  
del Trabajo



Organización  
Internacional  
del Trabajo

CINTERFOR

# INFORME FINAL

## *República Dominicana: análisis de oferta y demanda de empleo y trazabilidad de egresados Infotep, a partir de bases de datos*

*Elaborado por: Juan Carlos Guataquí  
con el apoyo de Diana Isabel Londoño*

CONTRATO DE COLABORACIÓN EXTERNA 40324779 / 0

## CONTENIDO

<b>Resumen Ejecutivo .....</b>	<b>3</b>
<b>INTRODUCCIÓN .....</b>	<b>6</b>
<b>1. ANÁLISIS DE VACANTES ON-LINE DISPONIBLES EN EL SISTEMA DE INTERMEDIACIÓN LABORAL DE REPÚBLICA DOMINICANA.....</b>	<b>8</b>
a. Descargue de datos en Web-scraping .....	8
b. Reporte estadístico del total de descargas .....	11
c. Depuración y homogenización de las bases de vacantes descargadas .....	12
d. Análisis de las vacantes en el mercado laboral de República Dominicana .....	13
<b>2. ANÁLISIS DE PATRONES DE INSERCIÓN LABORAL DE EGRESADOS DE INFOTEP 2016-2019 ...</b>	<b>18</b>
a. Deserción .....	20
b. Egresados de Cursos Cortos 2016-2019 .....	22
c. Egresados de Cursos Largos 2016-2019 .....	26
d. Características de empleabilidad de egresados INFOTEP 2016-2019: Encuesta a Egresados.....	29
e. Análisis econométrico de determinantes de la relación entre tenencia de trabajo y formación recibida en INFOTEP .....	38
f. Trazabilidad de los egresados en Encuesta a Egresados INFOTEP en el Sistema Integrado de Registro Laboral (SIRLA) .....	42
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>48</b>
<b>ANEXO 1. Caracterización del mercado laboral de República Dominicana .....</b>	<b>49</b>
<b>ANEXO 2. Caracterización de la oferta laboral .....</b>	<b>58</b>
<b>ANEXO 3. Análisis exploratorio de la prima salarial para la población egresada de INFOTEP que se encuentra empleada .....</b>	<b>65</b>
<b>ANEXO 4. Algoritmos utilizados para la obtención de bases Web-scraping software utilizado Python .....</b>	<b>69</b>

## RESUMEN EJECUTIVO

Este estudio aborda el análisis de dos elementos específicos<sup>1</sup> del mercado laboral de República Dominicana, con el propósito de desarrollar iniciativas analíticas del mercado laboral a partir de metodologías adscritas a un enfoque de Big Data<sup>2</sup>:

1. Compilación, procesamiento y análisis de la información de vacantes on-line disponibles en servicios (websites) de intermediación laboral, obtenidas mediante web-scraping.
2. Análisis de los determinantes de la inserción laboral de receptores de formación técnica de INFOTEP, mediante la integración de bases de datos primarias con registros administrativos, en el periodo 2016-2020.

El estudio aborda también la caracterización estructural y coyuntural del mercado de trabajo de República Dominicana entre 2008 y 2020, procesando y analizando información de la Encuesta Nacional Continua de Fuerza de Trabajo (ENCFT).

### Resultados y Reflexiones

Los resultados de la compilación, procesamiento y análisis de vacantes llaman la atención sobre varios elementos de importancia en la política pública. El estudio logró compilar 11.000 vacantes no estructuradas, procesar la información en ellas contenida y generar elementos de análisis ocupacionales (profesiones, áreas de formación, remuneraciones y habilidades blandas). Los sectores económicos con mayor demanda de vacantes que requieren competencias duras y bilingüismo (servicios asociados a la industria turística, servicios financieros, entre otros), vacantes de alta rotación y recurrencia, asociadas a actividades comerciales y la notoria ausencia de vacantes altamente calificadas, de competencias complejas, alta productividad y probablemente con remuneraciones diferenciales, lo cual sugiere que dichas vacantes no se reclutan mediante estos servicios de intermediación.

Como un todo, el ejercicio recalca la necesidad de dotar al Ministerio de Trabajo con herramientas legales que lleven a consolidar la información estructurada de las vacantes que, sobre puestos disponibles de trabajo, se ponen a disponibilidad en los servicios de intermediación laboral on line. El web-scraping se ha constituido en una alternativa de compilación y análisis de vacantes en entornos donde los ministerios de trabajo y los servicios públicos de empleo no cuentan con la capacidad de enforcement legal (reactiva) o el liderazgo efectivo (proactivo) para compilar, estructurar y analizar las vacantes existentes entre los diferentes actores que realizan actividades de intermediación laboral (por ejemplo, ver Guataquí et Al, 2014, 2015 para ilustración del caso colombiano). Sin embargo, una vez la

---

<sup>1</sup> Las bases de datos involucradas en el estudio son las siguientes: INFOTEP (bases de datos de egresados de cursos cortos o cursos largos 2016-2019, base de datos de deserción 2016-2019, base de datos Encuesta de Impacto a Egresados 2019, SIRLA), base de datos de vacantes por web-scraping (EmpleateYA, Tu Empleo RD y Computrabajo), base de datos de individuos inscritos a EmpleateYA.

<sup>2</sup> El informe se concentra en describir detalladamente el proceso de aplicación de dos iniciativas de Big Data al análisis laboral. Como anexos se incluyen: un análisis del mercado laboral de República Dominicana 2008-2020, la caracterización de la oferta laboral mediante el análisis de los individuos inscritos en EmpleateYa, un ejercicio estadístico de análisis de la prima salarial de los egresados de INFOTEP, y, finalmente, el código del proceso de Web-Scraping de las vacantes.

normatividad requerida dota al ministerio de trabajo con las herramientas para solicitar a los operadores de intermediación ciertos campos estructurados de las vacantes que anuncian, es posible realizar análisis coyunturales o estructurales sobre las vacantes así almacenadas.

En relación con la inserción laboral de egresados de INFOTEP, este componente del estudio es un piloto de utilización de técnicas de ciencia de datos para el análisis de la interacción entre la oferta y la demanda en el mercado de trabajo, analizando la empleabilidad de los egresados a partir de la integración y análisis de cuatro bases de datos (ver cita pie de página 2). El estudio llama la atención sobre la importancia de implementar sistemas de seguimiento basados en la Cédula de Identidad y Electoral como mecanismo de vinculación entre las bases de datos y los registros administrativos. De acuerdo con el análisis de inserción laboral, es importante enfatizar una perspectiva de género en el análisis de la inserción laboral de egresados de formación técnica y /o tecnológica. En el periodo analizado (2016-2019) INFOTEP convocó una mayor matrícula femenina que masculina. Las mujeres son notoriamente más demandantes de cursos cortos (317.095 egresadas, 72.7% del total de egresados de estos cursos) que de cursos largos (1.826 egresadas, 52.59% del total de egresados de cursos largos). Sin embargo, dicho diferencial no se ve aprovechado en el proceso de formación: la deserción es comparativamente más femenina (69%) que masculina (31%) y el 87.77% de los egresados que expresan que requieren formación adicional son mujeres. Finalmente, para todos los niveles de formación, las mujeres muestran niveles de empleo menores que los de sus contrapartes masculinas. El número de mujeres con trabajo, como porcentaje del total de mujeres con el respectivo nivel de formación siempre es menor que el porcentaje de hombres con trabajo en dicho nivel de formación.

La Encuesta de Impacto a Egresados encuentra que el 44.74% de los egresados estaba laborando al momento de serles aplicada la Encuesta. Entre los elementos más determinantes de la obtención de puestos de trabajo están la experiencia laboral, la pertinencia laboral de la formación recibida, la modalidad de formación dual (para mujeres en particular) y haber recibido apoyo de INFOTEP en el proceso de búsqueda de empleo.

Adicional a la importancia de ofrecer una perspectiva analítica de género al proceso de formación e inserción laboral, el ejercicio permite llegar a las siguientes reflexiones / recomendaciones.

- La formación dual o contar con oportunidades de empleo simultáneas a la formación es importante, toda vez que el 64.7% de los egresados INFOTEP que estaban trabajando al momento de la Encuesta, también trabajaban al momento de recibir la formación.
- La pertinencia de la formación recibida es importante, pero no definitivamente influyente en la posibilidad de obtener un empleo formal: el 54% de los egresados que al momento de aplicación de la encuesta tenían un trabajo formal, consideraban que dicho trabajo estaba relacionado con la formación recibida.
- El documento de identidad es un instrumento eficiente y directo imprescindible para la integración de bases de datos, dada la idiosincrasia particular de los nombres latinoamericanos, y en particular caribeños, y la posibilidad de encontrar múltiples acepciones escritas a nombres que fonéticamente suenan semejantes. Esto es extremadamente importante ante la creciente y prevalente práctica de realizar las encuestas vía telefónica.

Finalmente, al analizar el componente de oferta del mercado de trabajo a través de los perfiles de individuos inscritos en el servicio EmpléateYA entre 2008 y 2018, ha sido posible identificar varios

aspectos de interés. El primero es la relación positiva existente entre el comportamiento histórico de los desempleados cesantes en el país y la inscripción de individuos en búsqueda de empleo en EmpléateYA. Dicha relación se mantuvo entre 2009 y 2017, pero a partir de dicho año, a pesar de que los cesantes siguieron incrementándose (de 189.739 en 2016 a 215.700 en 2017, 229.092 en 2018), los inscritos en EmpléateYA se redujeron: de 62.623 (número más alto en todo el periodo analizado) en 2016, a 52.075 en 2017 y finalmente a 17.867 en 2018. Esto no puede atribuirse a la efectividad del servicio, toda vez que la base de datos sólo analiza los inscritos, no su efectiva inserción laboral. Uno de cada 5 inscritos en EmpléateYA desea laborar en áreas administrativas y 15% en servicios, seguidos de lejos por hotelería (4.8%) e informática (3.8%). Hay evidencia de segregación ocupacional por género entre los buscadores de empleo: las áreas de preferencia para laborar son totalmente diferentes entre mujeres y hombres. Por ejemplo, las áreas administrativas, farmacia y salud tienen un porcentaje de mujeres inscritas de 72.2%, 80.4% y 82.3% respectivamente. Esto, a contrastar con las áreas de energía, sector automotriz y construcción, donde el 95.1%, el 94.1% y el 77.8% de los inscritos son hombres. Resulta interesante ver que el 85% de los inscritos no revela su aspiración salarial, lo cual genera inquietudes sobre la existencia y nivel de un eventual salario de reserva.

Finalmente, es importante enfatizar el costo de oportunidad en que se incurre en no contar con un manejo más detallado de la información disponible en SIRLA. En términos prácticos, la información de pagos de seguridad social en formato de registro administrativo constituye un panel de datos. Y dado este aspecto cabe recordar que está técnicamente comprobado que el panel de datos es la herramienta más rigurosa y precisa para analizar los determinantes y efectos de los estados laborales y de las transiciones entre los mismos. Sin embargo, como operación estadística, el panel de datos es demasiado costoso para los presupuestos de los países en desarrollo, y la mejor muestra de ello es que muy pocos países cuentan con operaciones estadísticas con esta concepción. Sin embargo, en la práctica, los registros de seguridad social obtenidos mediante planillas de cotización constituyen paneles de datos, obtenidos a mínimo costo y susceptibles de ofrecer excelente información si su diseño operativo incluye información detallada sobre las remuneraciones y las posiciones ocupacionales. Es comprensible que las firmas y aportantes no puedan ser exhaustivamente indagados sobre competencias laborales o motivaciones del cambio de estado laboral, pero con la información básica de perfil socio-económico individual y remuneraciones, compilada y analizada a lo largo del tiempo se pueden construir valiosos indicadores adicionales, como experiencia general o específica, variaciones en el ingreso y transiciones ocupacionales. En el caso de SIRLA, estas innovaciones son viables y tienen enorme potencial. Y en el caso de otros países con registros consolidados de seguridad social, dichos registros constituyen un repositorio de información laboral extremadamente valiosa: un panel de datos obtenido como by-product de una política alterna cuyo propósito es administrativo, no estadístico ni analítico.

## INTRODUCCIÓN

El principal objetivo de este estudio ha sido el de implementar y desarrollar herramientas de caracterización y análisis del mercado de trabajo de República Dominicana, a partir de un enfoque de Big Data<sup>3</sup>. En este estudio hemos buscado alcanzar dicho objetivo analizando dos temas principales. El primero de ellos es la compilación, caracterización y análisis de vacantes On Line disponibles para el mercado laboral de República Dominicana en las bases de datos de cuatro<sup>4</sup> websites de intermediación laboral (incluido el portal público EmpléateYa), proceso que ha sido posible mediante el uso de técnicas de Web-Scrapping. El segundo tema es la aproximación a un seguimiento de egresados mediante la combinación de bases de datos de INFOTEP, compiladas en diferentes momentos y con distintos propósitos: (i) bases de datos de inscritos y deserciones, (ii) bases de datos de seguimiento a egresados y, por último, complementando el análisis con los registros de afiliación a seguridad social (SIRLA).

El uso de herramientas Big Data en la medición de brechas de oferta demanda laboral ofrece varias ventajas: oportunidad (el rezago entre la disponibilidad del instrumento y el momento de análisis es menor que en encuestas de hogares o de establecimientos); costo (la aplicación de instrumentos y su procesamiento se realizan con costos menores que instrumentos laborales estándar) y capacidad de procesamiento (se pueden manejar, procesar y analizar altos volúmenes de información no estructurada y estructurarla para propósitos de análisis).

Este informe se dirige a la presentación detallada de la aplicación de técnicas de Big Data a (i) la compilación, procesamiento y análisis de información de vacantes y (ii) seguimiento a egresados de formación técnica. Está dividido en dos grandes componentes.

Dentro del primer componente, su sección inicial presenta la primera aplicación de Big Data al análisis de la oferta y demanda laboral de República Dominicana. Esto es, la compilación, estructuración y análisis de más de 11.000 vacantes obtenidas mediante web-scrapping de tres websites de intermediación laboral (Tu Empleo RD, Computrabajo y EmpléateYa). Dichas vacantes tienen diferente comportamiento histórico. En el caso de “Tu Empleo” se analizó todo el stock histórico de vacantes publicadas en dicho portal, es decir que analizamos vacantes publicadas desde 2018. En el caso de “Computrabajo” se analizaron vacantes publicadas durante los tres primeros meses del 2021 y para EmpléateYa, el Ministerio de Trabajo puso a nuestra disposición la base de datos con el consolidado de finales del año 2020 y el primer trimestre de 2021.

La segunda sección del informe analiza los patrones de inserción laboral de los egresados de diferentes programas del Instituto Nacional de Formación Técnica Profesional (INFOTEP) que recibieron su formación entre 2016 y 2019, analizados en tres momentos de su proceso de inserción: al momento de recibir la formación (438.151 egresados de cursos cortos, 3.472 egresados de cursos largos, 83.524

---

<sup>3</sup> Como Big Data entendemos conjuntos extremadamente grandes de datos, que pueden analizarse computacionalmente para revelar patrones, tendencias y asociaciones, especialmente en relación con el comportamiento y las interacciones humanas.

<sup>4</sup> EmpléateYA, Tu Empleo RD, Computrabajo y Quala.

personas que desertaron a la formación), en la Encuesta de Impacto a Egresados, realizada en 2019 (1.065 egresados) y en la base de datos de SIRLA (donde fueron ubicados 352 egresados).

En los anexos a este informe el lector podrá encontrar los siguientes documentos complementarios o de contexto: Un análisis del comportamiento del mercado laboral de República Dominicana entre 2008 y 2020 (Anexo 1), la caracterización de la oferta laboral mediante el análisis de los individuos inscritos en la búsqueda de empleo en el portal EmpléateYA (Anexo 2), un análisis econométrico de los determinantes del ingreso laboral de los egresados de INFOTEP identificados en la base SIRLA (Anexo 3) y en el Anexo 4, los algoritmos de programación utilizados para el proceso de Web-Scraping de vacantes. Almente, la cuarta sección aborda el análisis del componente de oferta: los individuos que se han registrado en el servicio de “EmpléateYa” para buscar empleo. Se analiza la información disponible para 515.174 inscritos en dicho servicio entre 2008 y 2018. En promedio, el 60% de los inscritos están desempleados, y de dichos 60%, el 35% son mujeres. No hay expresiones claras de aspiración salarial: el 85% de los inscritos no reporta un salario de expectativa. La posibilidad de hacer emparejamiento entre la información de las vacantes y la información de los inscritos en EmpléateYa es baja, pues las vacantes son recurrentes y de alta rotación<sup>5</sup>, y los elementos de vínculo son genéricos (nivel educativo y habilidades blandas<sup>6</sup>). El Ministerio de Trabajo no cuenta actualmente con las herramientas legales para definir y analizar de forma más exigente y eficiente las vacantes, pero si puede mejorar los formatos de captura de información de los desempleados y aspirantes, en relación a descripción más detallada de la experiencia laboral y las competencias laborales.

Finalmente, es importante enfatizar el costo de oportunidad en que se incurre en no contar con un manejo más detallado de la información disponible en SIRLA. En términos prácticos, la información de aportes a seguridad social constituye un panel de datos: tiene las dimensiones de recurrencia en el tiempo, temáticas permanentes de reporte y los individuos analizados. Una vez reconocido este aspecto, es importante recordar que está técnicamente comprobado que el panel de datos es la herramienta más rigurosa y precisa para analizar los determinantes y efectos de los estados laborales y de las transiciones entre los mismos, pero como operación estadística, el panel de datos es demasiado costoso para los presupuestos de los países en desarrollo, y la mejor muestra de ello es que muy pocos países cuentan con operaciones estadísticas con esta concepción. Sin embargo, en la práctica, los registros de seguridad social obtenidos mediante planillas de cotización constituyen paneles de datos, obtenidos a mínimo costo y susceptibles de ofrecer excelente información si su diseño operativo incluye información detallada sobre las remuneraciones y las posiciones ocupacionales. Es comprensible que las firmas y aportantes no puedan ser exhaustivamente indagados sobre competencias laborales o motivaciones del cambio de estado laboral, pero con la información básica de perfil socio-económico individual y remuneraciones, compilada y analizada a lo largo del tiempo se pueden construir valiosos indicadores adicionales, como experiencia general o específica, variaciones en el ingreso y transiciones ocupacionales. En el caso de SIRLA, estas innovaciones son viables y tienen enorme potencial. Y en el caso de otros países con registros consolidados de seguridad

---

<sup>5</sup> Los vendedores representan el 15% de las vacantes compiladas, el 32% de las vacantes requieren baja calificación. El 38% de las vacantes no especifica el área de formación requerida.

<sup>6</sup> Las habilidades blandas más recurrentemente requeridas son el trabajo en equipo, la gestión de personas, organización y comunicación eficiente.



social, dichos registros constituyen un repositorio de información laboral extremadamente valiosa: un panel de datos obtenido como by-product de una política alterna cuyo propósito es administrativo, no estadístico ni analítico.

## **ANÁLISIS DE VACANTES ON-LINE DISPONIBLES EN EL SISTEMA DE INTERMEDIACIÓN LABORAL DE REPÚBLICA DOMINICANA**

En términos metodológicos, el análisis de las vacantes para República Dominicana se realizó a partir del uso de metodologías cuantitativas de extracción y procesamiento de información, obtenida de fuentes de datos no estructurados. En concreto, la propuesta metodológica que se desarrolló tiene como propósito principal capturar un gran número de observaciones (a partir de las bases de datos de sitios web de empleo), permitiendo identificar diferentes niveles de formación requeridos para cada cargo, competencias, habilidades y requerimientos específicos. En concreto, el componente cuantitativo se desarrolló en dos grandes pasos. El primero fue la descarga de las bases de datos de vacantes de los principales portales web de empleo a través de Web Scrapping durante los meses de febrero y marzo de 2021. El segundo paso fue la limpieza de estas fuentes de información, su depuración, exploración y categorización de las vacantes con la clasificación de ocupaciones CIUO 08 Internacional. Finalmente, se realizó un análisis a profundidad de la información consolidada y se generaron indicadores para ahondar y entender las dinámicas del mercado laboral en República Dominicana. El proceso está ilustrado en la figura 1.

### **a. Descargue de datos en WEB-SCRAPING**

Esta etapa se orientó a lograr la unificación de un proceso de desarrollo e implementación de diversas fuentes de información estructurada (bases de datos) o no estructuradas (Documentos, páginas web). Para tal fin se realizó un esquema de trabajo en dos fases:

#### **Primera fase: Identificación y recolección de vacantes disponibles en páginas web.**

Para la obtención de la información se implementaron técnicas de Web Scrapping, las cuales consisten en la extracción de grandes cantidades de datos disponibles en la web. A la hora de aplicar el web scraping se presentan varios desafíos, dentro de los principales se debe enfatizar la infraestructura en la cual se encuentra desarrollada la página web o portal de empleo, dentro de los cuales se destacan HTML, JAVA, XML y ANGULAR. Cada sitio en la web tiene una infraestructura única y debido a esto cada proceso de web scraping requiere un script único. La alteración de las páginas por parte de los desarrolladores web suele causar una reestructuración del código de nuestro web scraping.

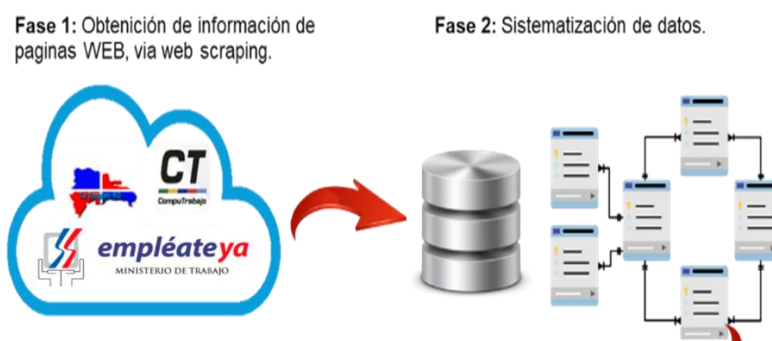
Para el análisis de los principales portales de ofertas web de República Dominicana, se construyeron algoritmos por medio del software estadístico python, y sus librerías de Selenium y BeautifulSoup, la cual tiene como principal objetivo facilitar la conexión automatizada entre páginas web, para descargar de manera sistemática y organizada la información que se le solicite.



**Segunda fase: fase de sistematización, en la cual se realiza un procesamiento de datos.**

La información de los portales de empleo, por su naturaleza, contiene información completa sobre la vacante, sin embargo, esta información no se encuentra estructurada en una base de datos organizada. Por tal motivo, se procede por medio de las funciones, a reestructurar y organizar la información obtenida en una base de datos estructurada, que tenga como mínimo los siguientes campos: i) Nombre de la vacante, ii) Descripción, iii) Fecha de publicación, iv) Salario, v) Empresa, iv) Entre otros requisitos mínimos.

**Figura 1. Fases del Proceso de download de vacantes por Web-Scrapping**



Fuente: Elaboración de los autores.

En la figura 2 se presenta un ejemplo de la información extraída para una vacante en uno de los portales de empleo consultados. Como un todo, los algoritmos de programación contruidos para la obtención y sistematización de las vacantes están incluidos en el anexo 5.

Figura 2. Ejemplos de Estructura de Avisos de Vacante: Campos de información extraída

<p> titulo : Chofer de Cabezote para Trasladar mercancías  url : <a href="https://www.tuempleord.do/2021/01/20/chofer-de-cabezote-para-trasladar-mercancias/">https://www.tuempleord.do/2021/01/20/chofer-de-cabezote-para-trasladar-mercancias/</a>  Requisitos: : Mayor de edad.  Funciones: : Trasladar mercancías en los cabezotes.  Correo Electrónico: : arodriguez@ccn.net.do  titulo : Mensajero área Administración para Santo Domingo Oeste  url : <a href="https://www.tuempleord.do/2021/01/22/mensajero-area-administracion-para-santo-domingo-oeste/">https://www.tuempleord.do/2021/01/22/mensajero-area-administracion-para-santo-domingo-oeste/</a>  Mensajero con Bachillerato : Experiencias previa en posiciones concluido similares  Beneficios: : Salario competitivo  url_de_la_empresa : <a href="https://www.computrabajo.com.do//maxx-extintores">https://www.computrabajo.com.do//maxx-extintores</a>  empresa : Maxx Extintores  url_de_la_vacante : <a href="https://www.computrabajo.com.do//ofertas-de-trabajo/oferta-de-trabajo-de-ayudante-de-chofer-repartidor-preferible-chofer-en-santo-domingo-oeste-61A9956AE76B2B8661373E686DCF3405">https://www.computrabajo.com.do//ofertas-de-trabajo/oferta-de-trabajo-de-ayudante-de-chofer-repartidor-preferible-chofer-en-santo-domingo-oeste-61A9956AE76B2B8661373E686DCF3405</a>  titulo : Ayudante de chófer repartidor - Preferible chofer  descripcion [13]  fecha agregado : 2021-01-22 19:45:00.139634  salario : 12.000,00 \$ (Neto mensual) .  region : Santo Domingo Oeste,  ciudad : Santo Domingo .  hora : 21 enero </p>
<p> url_de_la_empresa : <a href="https://www.computrabajo.com.do//empresas/ofertas-de-trabajo-de-bairesdev-sa-C8A646DBC3A7DE6E">https://www.computrabajo.com.do//empresas/ofertas-de-trabajo-de-bairesdev-sa-C8A646DBC3A7DE6E</a>  empresa : BairesDev SA  url_de_la_vacante : <a href="https://www.computrabajo.com.do//ofertas-de-trabajo/oferta-de-trabajo-de-business-development-operations-analyst-en-santo-domingo-este-6F18ACB3C19A4A2861373E686DCF3405">https://www.computrabajo.com.do//ofertas-de-trabajo/oferta-de-trabajo-de-business-development-operations-analyst-en-santo-domingo-este-6F18ACB3C19A4A2861373E686DCF3405</a>  titulo : Business Development Operations Analyst  descripcion [14]  fecha agregado : 2021-01-22 19:45:11.812805  salario : no especificado  region : Santo Domingo Este,  ciudad : Santo Domingo .  hora : 20 enero (actualizada) </p>

Fuente: Elaboración de los autores.

## b. Reporte Estadístico del Total de Descargas

Luego de la consolidación del script, la descarga de la información se inició la última semana de enero de 2021 y se finalizó la primera semana de marzo del mismo año. En total se consultaron 4 páginas web, desde las cuales sólo se pudo descargar información de 3 de ellas. En la Tabla 1 se relacionan las fuentes de información consultadas y el número de descargas.

**Tabla 1. Resultados de descargas**

Fuente	Fecha Inicial	Número de vacantes
Portal Tu Empleo RD	25 enero 2021	11.065*
Portal Computrabajo	25 enero 2021	674
Portal EmpleateYA	25 enero 2021	1055*
Quala	No se puede descargar la información, dado que la pagina se encuentra indexada a otro website que cuenta con restricción en las descargas.	

\*El portal de Tu Empleo RD permite capturar todo el histórico de las vacantes que se han publicado a través de este portal web. Para las vacantes del portal de EmpleateYA el Ministerio de Trabajo compartió la base de datos histórica desde finales de 2020 hasta marzo de 2021.

### **Limitaciones de la descarga de datos**

A continuación, se describen las principales limitantes de obtención de datos por medio de la web.

- **Derechos de propiedad intelectual:** Puede existir restricciones por el uso de la información obtenida, tales como resultados propios de metodologías exclusivas de las páginas web o aprovechamiento indebido de la reputación o esfuerzo ajeno, por ejemplo, puntajes de servicios o calidad de las empresas que publican vacantes.
- **Violación de los términos legales y condiciones de uso:** Aplica en los casos en los cuales las páginas indican de manera explícita el no uso de métodos de web-scraping o robots para descarga automatizada de datos que violen la normativa en protección de datos y la vulneración de los derechos de los titulares de los datos personales. En este caso, ninguna de las tres páginas presenta estas restricciones.
- **Disponibilidad de información en línea:** Se han identificado limitantes en la obtención de datos, dado que estas páginas web, limitan la publicación de vacantes solo para las últimas semanas, retirando de sus bases de datos públicas vacantes no disponibles.

### **c. Depuración y homogenización de las Bases de Vacantes descargadas**

El proceso de depuración, homologación e imputación de los códigos ocupacionales se realizó a través de los softwares estadísticos de R y Stata (ver anexo 4). Una vez se consolida las diferentes fuentes, se construye una única base de datos. Esta base de datos cuenta inicialmente con información sobre título, descripción, fecha de publicación y algunas características adicionales que son homogéneas a través de los diferentes portales. Al consolidar la información en una sola base de datos, existe la posibilidad de que una misma vacante haya sido publicada a través de los diferentes portales por lo que se debe hacer un ejercicio de eliminación de duplicidad basados en las siguientes razones:

1. La primera razón para encontrar vacantes duplicadas es que las vacantes son reportadas a diario, y por tanto en la mayoría de los casos una misma vacante puede ser reportada durante varios días, mientras se mantenga activa (duplicados dentro de la misma plataforma)
2. La segunda causa de duplicidad se presenta porque algunos empleadores pueden recurrir a diferentes prestadores para buscar su recurso humano y en ese caso una misma vacante puede ser reportada varias veces, aunque en diferentes prestadores (duplicados entre prestadores).

Para realizar la eliminación de los duplicados, se debió tomar cada caso por aparte. Para el primer caso en donde se presentan duplicados dentro de la misma plataforma que reporta, se toman las variables de título, descripción, y demás características de la información contenida en las bases de cada uno de los portales y se hace una primera eliminación de duplicidad. Una vez eliminado el primer tipo de duplicados, se procede a la eliminación de las vacantes duplicadas del segundo tipo (duplicados entre plataformas), si bien no se puede tener certeza de que se está eliminado la totalidad de vacantes repetidas que se hayan reportado a diferentes portales, el método utilizado es una buena aproximación a la realidad, para esto se supone que un empleador que reporta una misma vacante en diferentes prestadores utilizará la misma información en cada uno y por tanto se deberá realizar una eliminación a partir de las características título, descripción, fecha publicación, y demás variables.

Una vez se depuró la base de vacantes, se procedió a asignar los nombres a las variables y a generar las variables día, mes y año de publicación de las vacantes. Adicionalmente se corrigieron los caracteres especiales como tildes, puntos, comas, etc y los posibles errores ortográficos que presenten las variables título y descripción. Para generar las variables adicionales necesarias para complementar los análisis de vacantes y teniendo en cuenta la naturaleza de algunas de estas, se recurre a técnicas de análisis de texto a través de dos comandos en Stata (Strpos y Regexm<sup>7</sup>) que permiten identificar si en una observación se encuentra una palabra específica. Para cada variable se utilizan las palabras o raíces buscadas en el título y la descripción de la vacante, estas son dos de las variables más importantes de la base de datos, dado que contienen gran cantidad de información sobre las características de la vacante que es solicitada por los empleadores. La desagregación de estas variables a través del análisis de texto permitirá ampliar y profundizar la información de las vacantes reportadas. Para iniciar con el análisis de la base de datos, se generaron algunas variables de nivel educativo, duración de la vacante, salario, habilidades, entre otras.

Luego del proceso de limpieza de los datos y eliminación de duplicidad, se realizó la imputación de los códigos de ocupaciones. Para la imputación de estos códigos, se utilizaron las variables de título y

---

<sup>7</sup> Estos comandos identifican si una palabra aparece en cada observación de una variable determinada.

descripción de la vacante. Posteriormente a la limpieza de los “Corpus” Lingüísticos de estas variables se hizo uso de la librería *LabourR* dentro del paquete estadístico *R*. Esta librería permite hacer la imputación de los códigos ocupacionales, ya que mapea los cuerpos de texto generados a ontologías jerárquicas existentes como la ESCO y la ISCO. En nuestro caso, por criterio de estandarización de República Dominicana, la clasificación se hace hacia la ISCO-08. Tras la limpieza de los textos de entrada, el paquete *tokeniza* el texto, este proceso se entiende como generar vectores numéricos que indican la existencia o no de los diversos términos posibles dentro de los textos analizados mediante unos y ceros. Estos vectores son comparados con los vectores de la clasificación estandarizada a usar. Para este proceso se calcula la frecuencia de término – frecuencia inversa de documento, Tf-idf por sus siglas en inglés. Este puntaje es precisamente el que permite recuperar la ontología sugerida, mediante la implementación de metodologías de aprendizaje de máquina. Específicamente esta librería, y en el proceso aquí descrito, se hace uso del algoritmo K-NN, o K vecinos más cercanos en español, aunque cabe resaltar que también permite la utilización de algoritmos como el Clasificador Bayesiano Ingenuo. Este ejercicio se realiza dos veces; en primer lugar, se imputa el código solo con la información que contiene el título de la vacante y, en segundo lugar, se imputa el código solo con la información contenida en la descripción de la vacante. Este proceso dual se debe a que algunos títulos de las vacantes no cuentan con suficiente información para ser categorizados por lo cual, se ajusta la categorización de las vacantes que no resultaron clasificadas a través del título, utilizando la descripción. Para este ejercicio particular con los datos de las vacantes de los diferentes portales, a partir de este proceso se logró categorizar el 86.3% de las vacantes.

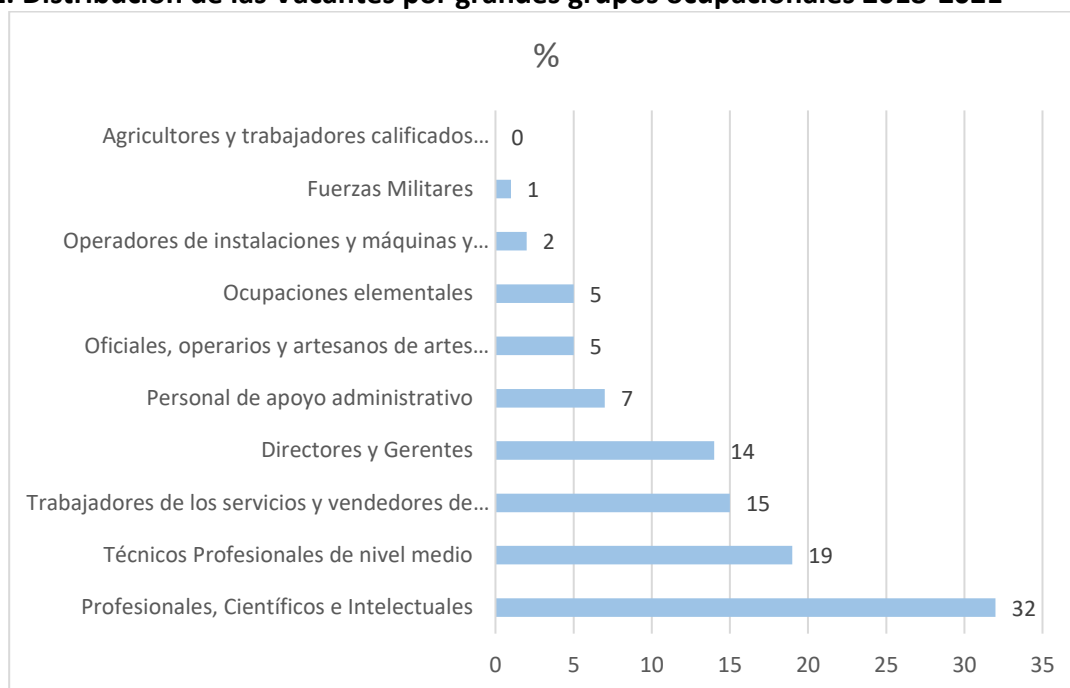
#### **d. Análisis de las Vacantes en el mercado laboral de República Dominicana**

En términos generales, el análisis de las vacantes para República Dominicana se realizó a partir del uso de metodologías cuantitativas de extracción y procesamiento de información a partir de fuentes de datos no estructuradas. En concreto, la propuesta metodológica que se desarrolló tiene como propósito principal capturar un gran número de observaciones (a partir de las bases de datos de portales web de empleo), permitiendo identificar diferentes niveles de formación requeridos para cada cargo, competencias, habilidades y requerimientos específicos. En concreto, el componente cuantitativo se desarrolla en dos grandes pasos. El primero es la descarga de las bases de datos de vacantes de los principales portales web de empleo a través de Web Scrapping durante los meses de enero y principios marzo de 2021. El segundo paso es la limpieza de estas fuentes de información, su depuración, exploración y categorización de las vacantes con la clasificación de ocupaciones CIUO 08 Internacional. Finalmente, se realiza un análisis a profundidad de la información consolidada y se generan indicadores que permiten ahondar y entender las dinámicas del mercado laboral en República Dominicana.

A partir de la información descargada en los portales de empleo, se capturó información entre 2018 y marzo 2021. La información de los años anteriores a 2021 corresponde a la reportada en el portal de TuEmpleoRd, se acumulan un total de 12,091 vacantes. En su mayoría estas vacantes fueron del 2018 (43.7%), año previo a la aparición de la pandemia por COVID-19. Para el año 2019 se tuvieron 2.668 (22.06%) y en 2020 el número de vacantes capturadas fue de 4,139 vacantes. Entre enero y principios de marzo de 2021 a través de los portales se registraron 2,136 vacantes.

Para caracterizar las vacantes de los años recientes el gráfico 1 muestra las vacantes capturadas por grandes grupos ocupacionales, esto se realiza a partir de la asignación de código ocupación CIUO 08 a un dígito, a las vacantes descargadas. El grupo ocupacional de clasificación demandada más recurrente es el de Profesionales, científicos e intelectuales que representó el 31.25% del total de vacantes. Entre este grupo se resaltan las vacantes para profesionales de ventas técnicas y médicas que representaron el 4.3%, las de analistas financieros con 3.7% y las especialistas en políticas y servicios de personal con 3.4%.

**Gráfico 1. Distribución de las Vacantes por grandes grupos ocupacionales 2018-2021**



Fuente: Elaboración de los autores a partir de vacantes obtenidas.

El siguiente gran grupo ocupacional de mayor importancia fue el de técnicos profesionales de nivel medio que representó el 19% de las vacantes y dentro del cual resaltan por ejemplo las vacantes para secretarios administrativos y ejecutivos que representaron 3.3% del total de vacantes, las de otros profesionales de nivel medio en actividades culturales y artísticas con 2.9% y secretarios jurídicos con 2.8%.

Los dos grandes grupos ocupacionales mencionados anteriormente, son seguidos en importancia por el de trabajadores de los servicios y vendedores de comercios y mercados con 15.2% del total de vacantes. Este grupo cuenta además con la ocupación con el mayor número de vacantes en el mercado laboral de República Dominicana, los asistentes de venta de tiendas y almacenes que con 694 vacantes equivalen al 5.7% del total. En este grupo también se tienen a los guardias de protección que representaron 2.3%. Por último, los directores y gerentes representan el cuarto grupo ocupacional con mayor participación sobre el total de vacantes con 13.5%, dentro de este grupo resaltan las vacantes para gerentes de centros deportivos, de esparcimiento y culturales con 2.2% del total de vacantes y los

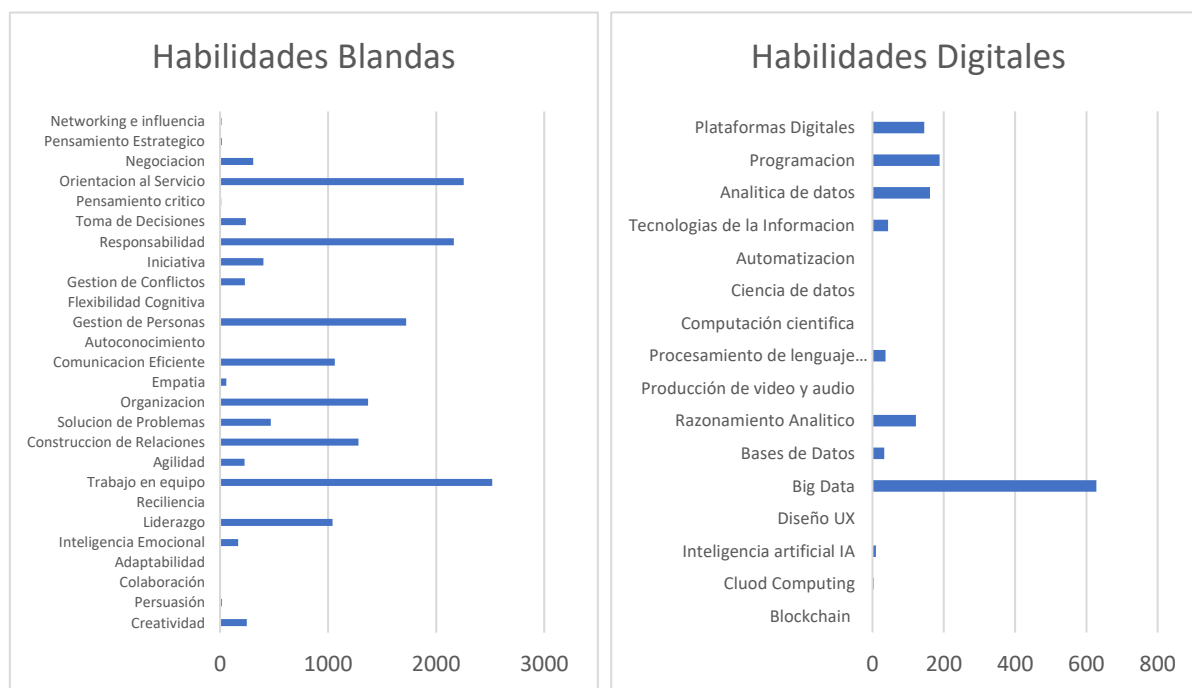
gerentes de comercios al por mayor y al por menor y los directores financieros, ambos con una participación del 1.9% sobre el total de vacantes.

Si aproximamos los niveles educativos requeridos en las vacantes a expresiones del nivel de cualificación requerido, en su mayoría las vacantes requieren un nivel de cualificación alto (45.5%), el siguiente nivel de cualificación más común es baja cualificación (34.2%) y el de menor número de vacantes es el nivel de cualificación medio (18.8%). En cuanto a la experiencia requerida son pocas las vacantes de entrada, que no requieren experiencia, solo 2.3% del total de vacantes requiere no tener experiencia. La gran mayoría de las vacantes requieren 1 año de experiencia (29.9%) o 2 años (29.4%). La experiencia requerida también muestra alta concentración en 3 años (14.9%) y 5 años (10.2%). En general, la concentración de vacantes en el nivel alto de cualificación concuerda con los altos requerimientos de experiencia en el mercado. Sin embargo, cabe resaltar que la importante participación de vacantes de baja cualificación no se representa en un gran número de vacantes que no requiera o requiera poca experiencia, el 86.9% de las vacantes requiere 1 año o más.

El gráfico 2 resume las habilidades más demandadas por las vacantes en República Dominicana. Se resaltan entre las habilidades blandas el trabajo en equipo, la orientación al servicio, la responsabilidad y la gestión de personas. Entre las habilidades digitales la más importante son los conocimientos de *Big Data*, y con una mucho menor participación los conocimientos de programación, analítica de datos y plataformas digitales. Coherente con esta distribución de habilidades demandas, las palabras o frases más comunes dentro de las vacantes son “trabajo en equipo”, “construcción de relaciones”, “liderazgo”, “gestión de personas”, “organización” y “comunicación eficiente”.



**Gráfico 2. Habilidades más demandadas**



Fuente: Elaboración de los autores. Datos en el eje horizontal en número de vacantes que expresan el requerimiento.

Por último, para complementar el análisis de las vacantes laborales en los años recientes se exploran las áreas de formación más demandadas, aunque cabe resaltar que la mayoría de las vacantes no especifica claramente una profesión, como título profesional (34.5%). Las áreas de formación más comúnmente demandadas son Publicidad y mercadeo con 18.3% del total de vacantes, Ingeniería de sistemas con 6.9%, administración con 6.6% y gestión empresarial con 5.7%.

La tabla 2 presenta los principales criterios exhibidos por las vacantes compiladas, en la coyuntura laboral 2020-2021. Las vacantes se concentran en cuatro grandes grupos ocupacionales, en ocupaciones relacionadas con actividades comerciales, 44% de ellas requiere cualificación alta. Cerca del 59% de las vacantes requieren entre 1 y 2 años de experiencia y el área de formación más demandada es Publicidad y Mercadeo (16%).

**Tabla 2. Principales Características de las Vacantes 2020-2021. Criterios más demandados por participación porcentual.**

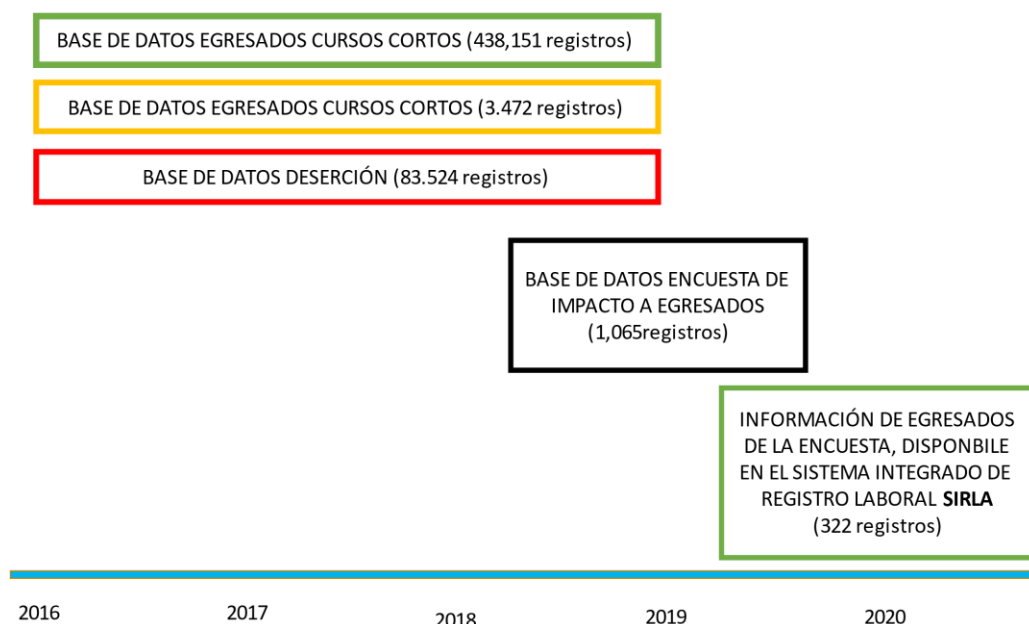
GRUPO OCUPACIONAL	% en 2020	% en 2021
Profesionales, Científicos e Intelectuales	30,5	29,4
Técnicos Profesionales de nivel medio	18,9	18,5
Trabajadores de los servicios y vendedores de comercios y mercados	15,5	15,1
Directores y Gerentes	13,5	13,5
OCUPACION		
Asistente de venta de tienda / almacén	4,9	5,8
Profesional de ventas técnicas y médicas	4,3	3,8
Otros profesionales de nivel medio en actividades culturales y turísticas	3,4	3,5
Ocupaciones elementales no clasificadas	3,3	3,4
Especialistas en políticas y servicios de personal	3,2	3,4
NIVEL DE CUALIFICACION		
Alto	44,6	44
Medio	35,3	35,8
Bajo	18,9	18,5
EXPERIENCIA REQUERIDA		
1 año	28,8	29,3
2 años	29,8	30,2
3 años	13,8	15,6
5 años	11,4	10,1
No requiere experiencia	2,2	2,3
AREA DE FORMACIÓN		
Publicidad y Mercadeo	16,8	17,6
Ingeniería de Sistemas	6,5	6,2
Administración	6,1	6
Gestión Empresarial	5,6	4,9
Sociología	4,1	3,8
No Especifica	38,4	37,4

Fuente: Elaboración de los autores. Para consideraciones de habilidades digitales requeridas en las vacantes ver Gráfico 2.

## 1. ANÁLISIS DE PATRONES DE INSERCIÓN LABORAL DE EGRESADOS DE INFOTEP 2016-2019<sup>8</sup>

Con el fin de analizar la posibilidad de establecer un mecanismo de seguimiento a la inserción laboral de los egresados de INFOTEP, en este capítulo se presenta el proceso de consolidación y análisis de diferentes bases de datos. La figura 3 ilustra el proceso cronológico y conceptual desarrollado.

**Figura 3. Análisis Egresados INFOTEP. Proceso Conceptual de Integración de Bases de Datos.**



Fuente: Elaboración propia.

A partir de las bases de datos de egresados (cursos cortos / cursos largos) y de deserción, se estableció el consolidado de egresados 2016-2019, dicho consolidado se analiza en la sección (a) de este capítulo. INFOTEP aplicó una Encuesta de Seguimiento a Egresados entre mayo y junio del año 2019. La sección (b) de este capítulo analiza en detalle las características laborales de los egresados, utilizando el microdato de la Encuesta, y mediante la integración de los datos de identificación, compara sus circunstancias en dos momentos: al recibir la capacitación y luego, al serles aplicada la Encuesta. A partir de la información de la Encuesta, combinada con las bases de datos originales, se realiza una estimación de los determinantes de empleo. Finalmente, los datos de los egresados de la Encuesta fueron indagados por el Ministerio de Trabajo, en el Sistema Integrado de Registro Laboral (SIRLA), repositorio y administrador de la información de cotización a seguridad social de la República Dominicana. La tercera sección de este capítulo analiza en detalle las características de estos individuos y, dada la integración de bases de datos, evalúa sus circunstancias laborales en tres momentos en el tiempo, utilizando tres diferentes bases de datos, dos primarias y una de registro administrativo, obtenida por procesamiento de un data warehouse como SIRLA.

<sup>8</sup> Agradecemos de la forma más cordial a Ondina Marte, Arelis Tolentino y Nancy Salcedo por su apoyo permanente al desarrollo de este componente analítico de la consultoría.

Como un todo, este componente del estudio mostró interesantes oportunidades analíticas y decepcionantes resultados en el componente final de la integración, es decir, el análisis de los egresados en la base de datos SIRLA. Desde la perspectiva estadística, los 1060 egresados que fueron sujeto de la Encuesta de Impacto constituían una muestra aceptable del total de egresados de la institución en un periodo reciente, que es cuando resulta fructífero analizar el impacto de la formación sobre la inserción laboral. Sin embargo, los problemas metodológicos fueron debilitando el proceso. El formulario de la Encuesta no indagaba sobre el número de documento de identidad, o al menos el microdato de la Encuesta no lo había compatibilizado. Esto es particularmente problemático dada la sintaxis y diferentes acepciones lingüísticas de los nombres caribeños, los cuales dan lugar a equívocos, que podrían ser subsanados con el simple número de cédula de identidad y electoral. Un periodo sustantivo de tiempo tuvo que ser dedicado a reconstruir los números de cédula de los 1060 egresados encuestados.

Una vez se contó con el listado de números de cedula de identidad y electoral de estos egresados, la búsqueda en SIRLA, es decir, en los registros de seguridad social, ofreció varios resultados decepcionantes. Uno, quizá en gran parte por la coyuntura económica de la pandemia (2020) o por el hecho de que SIRLA sólo se concentra en empleo formal dependiente, sólo fue posible ubicar a 352 egresados en dicho sistema. Dos, la información obtenida era extremadamente precaria en términos de las condiciones laborales y los montos y criterios de remuneración. Y tres, la información era estática: sólo se pudo obtener la información del estado laboral de estas 352 personas en el momento de la consulta, es decir, no pudimos acceder a la información de los diferentes estados laborales en que los 352 egresados pudieron haber estado en SIRLA, en particular, los momentos en que no aparecían como aportantes, con lo cual se podrían analizar sus periodos de desempleo / inactividad. Sin embargo, consideramos que existía información valiosa a ser analizada, en particular dados los diferentes momentos analizados y las posibles transiciones que los egresados han experimentado. Por ejemplo, los niveles de empleo por nivel de formación son semejantes cuando se analizan en la Encuesta de Egresados y en SIRLA. En esta sección del documento se presenta el análisis del proceso de inserción de egresados de INFOTEP mediante la integración de bases de datos.

#### **a. Características de los Egresados 2016-2019.**

Gracias a la amable gestión y apoyo de las directivas de INFOTEP, hemos podido tener acceso y capacidad de procesamiento y análisis de las siguientes bases de datos:

**Tabla 3. Bases de Datos Análisis INFOTEP.**

Base de Datos	Tamaño del Archivo	Número de Registros
Egresados de Cursos Cortos (2016-2019).	27.093 kb	438.151
Egresados de Cursos Largos (2016-2019).	254 kb	3.472
Personas que Desertaron (2016-2019).	5.851 kb	83.524
Encuesta de Impacto a Egresados.	342 kb	1.065

#### a. Deserción

Para los 4 años de análisis (2016-2019, ver Tabla 4) no parece existir una incidencia cronológica particular, aunque si se aprecia un incremento en los niveles de deserción. Del total de personas que desertan de los cursos (83.541 personas), el número se ha ido incrementando (22,38% en 2016 a 28,5% en 2019). Se aprecian mayores niveles de deserción en mujeres que en hombres, en una relación de aproximadamente 7 mujeres por cada 3 hombres, un patrón que, si bien puede reflejar una circunstancia específica de salida, también puede estar influido por elementos de entrada, pues las mujeres también estarían más representadas en el proceso de ingreso, con lo cual estadísticamente no sería particularmente llamativo, desde el punto de vista conceptual, que también lo estén en el proceso de deserción.

**Tabla 4. INFOTEP: Deserción 2016-2019.**

<b>AÑO</b>	<b>PERSONAS QUE ABANDONARON LOS CURSOS</b>	<b>% DEL TOTAL DE DESERTORES 2016-2019</b>	<b>MUJERES</b>	<b>HOMBRES</b>
2016	18.696	22,38%	12.535	6.161
2017	18.834	22,54%	12.870	5.964
2018	22.036	26,38%	15.399	6.637
2019	23.975	28,70%	17.136	6.839
Total	83.541	100	69,18%	30,81%

Fuente: INFOTEP. Procesamiento de los Autores.

Al combinar los criterios de sexo y edad para analizar el comportamiento de la deserción en el periodo analizado (Tabla 5) se encuentra una amplia representación del espectro de edades, con ciertas notoriedades: una baja representatividad de jóvenes (20 a 24 años), que sólo representan el 0,09% del total de personas que desertan a la formación y el hecho de que la mayor parte de la deserción se encuentra concentrada en el componente de 25 a 34 años (49% aproximadamente). De nuevo, al interior de estos grupos, la representación de las mujeres es mayor a la de hombres en una relación de 7 a 3.

**Tabla 5. Deserción por Sexo y Edad. 2016-2019**

Edad	Femenino	Masculino	Total	%
20-24	78	0	78	0,09%
	100%	0%	100%	
25-29	12.881	7.028	19.909	23,83%
	64,70%	35,30%	100%	
30-34	13.802	7.078	20.880	24,99%
	66,10%	33,90%	100%	
35-39	9.770	4.287	14.057	16,83%
	69,50%	30,50%	100%	
40-44	7.121	2.718	9.839	11,78%
	72,38%	27,62%	100%	
45-49	5.113	1.598	6.711	8,03%
	76,19%	23,81%	100%	
50-54	3.809	1.077	4.886	5,85%
	77,96%	22,04%	100%	
55-59	2.518	792	3.310	3,96%
	76,07%	23,93%	100%	
60-64	1.424	492	1.916	2,29%
	74,32%	25,68%	100%	
65-69	785	280	1.065	1,27%
	73,71%	26,29%	100%	
70-74	410	138	548	0,66%
	74,82%	25,18%	100%	
75-79	140	65	205	0,25%
	68,29%	31,71%	100%	
>80	72	42	114	0,14%
	63,16%	36,84%	100%	
Total	57.923	25.595	83.518	100%
	69,35%	30,65%	100%	

Fuente: INFOTEP. Procesamiento de los Autores.

Los niveles educativos de las personas que desertaron de la formación de INFOTEP en el periodo de análisis se concentran en la educación media (45.913 personas) con una marcada participación de mujeres (67% de este nivel educativo y 37% del total). Dado el papel que tiene la formación técnica y tecnológica como parte de un potencial ciclo propedéutico, la participación de personas con educación media no es sorprendente, pero si resulta serlo la de personas con educación universitaria (15,87%, 13.266 personas), un aspecto que podría sugerir problemas de pertinencia en la educación superior universitaria.

**Tabla 6. Deserción por Nivel Educativo del Inscrito.**

Nivel Educativo	Femenino	Masculino	Total
Alfabetizado	8.628	2.532	11.160
	77,31	22,69	100
Básica	6.468	3.072	9.540
	67,8	32,2	100
Inicial	1.547	602	2.149
	71,99	28,01	100
Media	31.112	14.801	45.913
	67,76	32,24	100
Técnico	804	709	1.513
	53,14	46,86	100
Universitario	9.381	3.885	13.266
	70,71	29,29	100
Total	57.940	25.601	83.541
	69,36	30,64	100

Fuente: INFOTEP. Procesamiento de los Autores.

#### **b. Egresados de Cursos Cortos 2016-2019**

A modo de estadísticas descriptivas y previo a la depuración por duplicidades (una persona tomando varios cursos, un aspecto de rigor metodológico que no afecta la robustez de los resultados agregados, en particular los patrones porcentuales), el perfil de los egresados de cursos cortos muestra los siguientes elementos: predominancia femenina (72%), particularmente de mujeres entre los 30 y 34 años (constituyen el 16.18% del total de egresados de cursos cortos).



**Tabla 7. Egresados de Cursos Cortos por Edad y Sexo.2016-2019**

Edad	Femenino	Masculino	Total
20-24	291	0	291
	100	0	100
25-29	59.438	29.478	88.916
	66,85	33,15	100
30-34	70.890	31.199	102.089
	69,44	30,56	100
35-39	53.910	20.357	74.267
	72,59	27,41	100
40-44	42.248	13.652	55.900
	75,58	24,42	100
45-49	31.845	8.963	40.808
	78,04	21,96	100
50-54	24.295	6.307	30.602
	79,39	20,61	100
55-59	16.295	4.662	20.957
	77,75	22,25	100
60-64	9.335	2.859	12.194
	76,55	23,45	100
65-69	5.165	1.758	6.923
	74,61	25,39	100
70-74	2.266	1.018	3.284
	69	31	100
75-79	698	494	1.192
	58,56	41,44	100
>80	318	265	583
	54,55	45,45	100
Total	316.994	121.012	438.006
	72,37	27,63	100

Fuente: INFOTEP. Procesamiento de los Autores.

El nivel educativo de los egresados de cursos cortos es claramente intensivo en personas con educación media (64.45% del total), en el cual predominan las mujeres (71.9%), al punto de que constituyen el grupo etareo y de sexo más representativo del total, seguido de sus contrapartes masculinas (18%). De

nuevo, un elemento llamativo es la participación de personas con educación universitaria (14.3%), de los cuales 3 de cada cuatro son mujeres.

**Tabla 8. Egresados de Cursos Cortos INFOTEP por sexo y nivel educativo. 2016-2019**

Nivel Educativo	Femenino	Masculino	Total
Alfabetizado	23.954	5.796	29.750
	5,47%	1,32%	6,79%
Básica	34.524	14.766	49.290
	7,88%	3,37%	11,25%
Inicial	5.191	1.757	6.948
	1,18%	0,40%	1,59%
Media	203.121	79.255	282.376
	46,36%	18,09%	64,45%
Técnico	4.171	2.901	7.072
	0,95%	0,66%	1,61%
Universitario	46.134	16.581	62.715
	10,53%	3,78%	14,31%
Total	317.095	121.056	438.151
	72,37%	27,63%	100

Fuente: INFOTEP. Procesamiento de los Autores.

En relación con el perfil de formación recibido, el portafolio de formación de cursos cortos hace referencia a 74 cursos, de los cuales los 10 más demandados se presentan en la Tabla 9.

**Tabla 9. INFOTEP: 10 Cursos Cortos con mayor número de egresados. 2016-2019.**

FAMILIA	# Personas	% (por Familia o al interior de la familia)
CC INFORMATICA	62.532	<b>14,27</b>
Hombres	19.113	30,57
Mujeres	43.319	69,43
CC MANUALIDADES	40.438	<b>9,23</b>
Hombres	2.905	7,18
Mujeres	37.533	92.82
CC CONFECCION DOMESTICA DE PRENDAS DE VESTIR	34.325	<b>7,83</b>
Hombres	1.223	3,56
Mujeres	33..102	96,44
CC BELLEZA Y PELUQUERIA	33.093	<b>7,55</b>
Hombres	5.460	16,5
Mujeres	27.633	83,5
CC FARMACOLOGIA	25.316	<b>5,78</b>
Hombres	2.645	10,45
Mujeres	22.671	89,55
CC VENTAS	24.206	<b>5,52</b>
Hombres	7.916	32,7
Mujeres	16.290	67,3
CC CONTABILIDAD	22.591	<b>5,16</b>
Hombres	5.562	24,62
Mujeres	17.029	75,38
CC PANADERIA Y REPOSTERIA	20.881	<b>4,77</b>
Hombres	2.306	11,04
Mujeres	18.575	88,96
CC INSTALACION Y MANTENIMIENTO ELECTRICO	14.891	<b>3,4</b>
Hombres	14.177	95,21
Mujeres	714	4,79
CC ARTE CULINARIO	14.715	<b>3,36</b>
Hombres	2.245	15,26
Mujeres	12.470	84,74

Entre los perfiles de formación de los 10 cursos cortos con mayor número de egresados existen algunos que, a pesar de generar condiciones de empleabilidad, están extremadamente segregados: Instalación y Mantenimiento Eléctrico (95% egresados hombres), Panadería y Repostería (88.9% egresadas mujeres) y Farmacología (89.5% mujeres).

### **c. Egresados de Cursos Largos 2016-2019**

En promedio, 868 egresaron de cursos largos impartidos por INFOTEP en el periodo 2016-2019. A diferencia de los cursos cortos, existe una distribución más homogénea de los egresados, al analizar su composición por sexo (52.5% mujeres y 47.5% hombres).

**Tabla 10. Egresados Cursos Largos INFOTEP. 2016- 2019**

AÑO	Femenino	Masculino	Total
2016	509	446	955
	53,3	46,7	100
2017	453	330	783
	57,85	42,15	100
2018	374	428	802
	46,63	53,37	100
2019	490	442	932
	52,58	47,42	100
Total	1.826	1.646	3.472
	52,59	47,41	100

Fuente: INFOTEP. Procesamiento de los Autores.

Los niveles educativos de los egresados de cursos largos están casi en su totalidad concentrados en educación media. Existen muy pocas mujeres egresadas que contaban sólo con educación básica (4, 0,12% del total de egresados), pero las mujeres con educación media representan el 52,47% del total de egresados, mucho más que su contraparte masculina (41,12%).

**Tabla 11. Egresados Cursos Largos INFOTEP. 2016-2019.**

Nivel Educativo	Femenino	Masculino	Total
Básica	4	218	222
	1,8%	98,2%	100%
Media	1.822	1.428	3.250
	56,06%	43,94%	100%
Total	1.826	1.646	3.472
	52,59%	47,41%	100%

Fuente: INFOTEP. Procesamiento de los Autores.

El portafolio de cursos largos abarca 37 perfiles de formación. De ellos los 10 cursos más demandados se presentan en la tabla 12. Se aprecia una marcada segregación, sólo un curso (Ventas) tiene una distribución relativamente homogénea de egresados hombres y egresadas mujeres (48,16% y 51,69%). De los 10 cursos más demandados, tres (Enfermería, Diseño y Decoración de Interiores, Cosmiatría) tienen niveles de egresadas mujeres superiores al 93%.

**Tabla 12. Los Diez Cursos Largos con más egresados INFOTEP. 2016-2019**

FAMILIA	# Personas	% (por Familia o al interior de la familia)
ENFERMERIA	1.266	36,46
Hombres	85	6,71%
Mujeres	1.181	93,29%
INSTALACION Y MANTENIMIENTO ELECTRICO	315	9,07
Hombres	312	99,05%
Mujeres	3	0,95%
MECANICA GENERAL	298	8,58
Hombres	292	97,99%
Mujeres	6	2,01%

MECÁNICA AUTOMOTRIZ	187	5,39
Hombres	186	99,47%
Mujeres	1	0,53%
VENTAS	179	5,16
Hombres	86	48,04%
Mujeres	93	51,96%
CONTABILIDAD	154	4,44
Hombres	29	18,83%
Mujeres	125	81,17%
MANTENIMIENTO DE REFRIGERACION Y AIRE	132	3,8
Hombres	132	100%
Mujeres	0	0%
DISEÑO Y DECORACION DE INTERIORES	100	2,88
Hombres	5	5%
Mujeres	95	95%
SOLDADURA	92	2,65
Hombres	92	100%
Mujeres	0	0%
COSMIATRIA	65	1,87
Hombres	1	1.54%
Mujeres	64	98,46%

Fuente: INFOTEP. Procesamiento de los Autores.

**d. Características de empleabilidad de egresados INFOTEP 2016-2019: Encuesta a Egresados.**

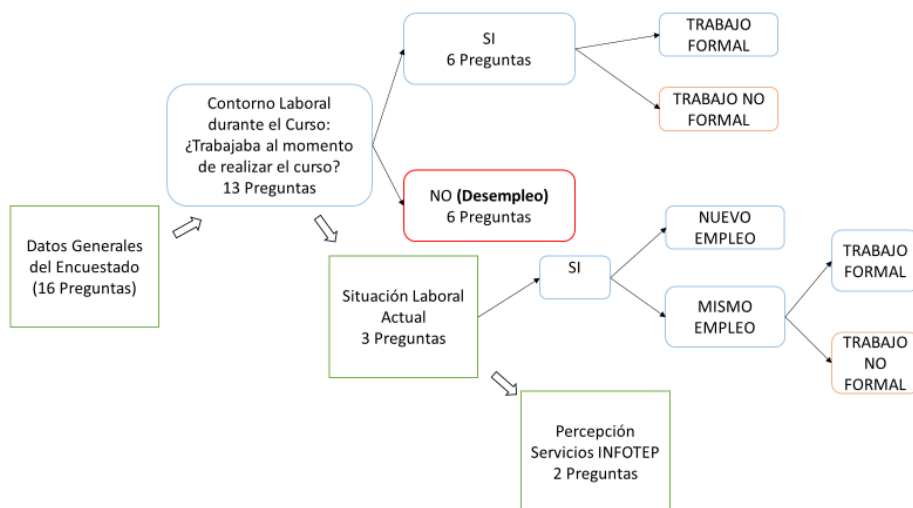
Entre el 28 de mayo y el 14 de junio de 2019, INFOTEP (Gerencia de Planificación Estratégica e Inteligencia Institucional / Departamento de Investigación y Estadísticas de Mercados Laborales) aplicó su Encuesta de Seguimiento a Egresados de la Formación Técnica 2019. El cuestionario estuvo compuesto por 4 módulos (Datos Generales, Situación Laboral durante el curso, Situación Laboral Actual (2019) y Servicios del INFOTEP).

El microdato generado por el procesamiento de esta encuesta fue amablemente puesto a nuestra disposición por INFOTEP. A continuación, presentamos las principales características de los egresados que diligenciaron la Encuesta (1060 egresados). El 60.8% de ellos (645) son mujeres y el 39.2% (415) hombres, distribuidos en los siguientes rangos de edad:

- 14 a 22 años (18.3% y 25.1%, total 20.9%),
- 23 a 28 años (26.8% y 39.8%, total 31.9%),
- 29 a 34 años (20.5% y 15.9%, total 18.7%)
- 35 a 40 años (14.9% y 10.4%, total 13.1%)
- 41 y más (19.5% y 8.9%, total 15.4%)

Como puede apreciarse los rangos de edad son adecuadamente representativos de las estructuras de edad de los egresados de cursos cortos / largos. Existe una relativa sobre-representación de los hombres de 23 a 28 años, pero las estructuras por edad y para cada rango de edad por sexo, se encuentran dentro de intervalos representativos de la población de egresados.

**Figura 4. Mapa Conceptual del a Encuesta de Seguimiento a Egresados 2019**



FUENTE: Elaboración de los autores a partir del Cuestionario de la Encuesta.

La Tabla 13 muestra la distribución de la población encuestada de acuerdo con su sexo y nivel educativo. En relación con el nivel educativo, los dos niveles más representativos son secundaria completa (11.3%) y universitaria incompleta (36.7%).



**Tabla 13. Población Encuestada por Nivel Educativo.**

Último Nivel aprobado	Mujeres	Hombres	Total
Primaria completo	10	6	16
%	1,6%	1,4%	1,5%
Primaria incompleto	13	1	14
%	2,0%	0,2%	2,2%
Secundaria completo	238	197	435
%	36,9%	30,5%	67,4%
Secundaria incompleto	44	29	73
%	6,8%	4,5%	11,3%
Técnico	2	0	2
%	0,3%	0,0%	0,3%
Universitario completo	82	38	120
%	12,7%	5,9%	18,6%
Universitario incompleto	250	139	389
%	38,8%	21,6%	60,3%
Maestría y/o Posgrado	6	5	11
%	0,9%	0,8%	1,7%
Total	645	415	1060
%	60,8%	39,2%	100%

Fuente: INFOTEP. Procesamiento de los autores. NOTA: 5 mujeres no declaran su nivel educativo.

#### **i.Estatus Laboral al momento de recibir la capacitación INFOTEP.**

Como lo ilustra la Figura 3, la Encuesta indagó inicialmente sobre las circunstancias laborales al momento en que simultáneamente estaban recibiendo su formación. Las personas que afirmaron estar trabajando al momento de realizar / recibir el curso de formación técnica se les indagó sobre el carácter formal / no formal de su trabajo, y a quienes lo consideraban formal, la relación de su ocupación con la formación que estaban recibiendo, el sector económico de su trabajo, cargo e ingresos mensuales. Para los ocupados no formales, se les preguntó sobre la dependencia laboral del trabajo (cuenta propia o emprendedor).

La Tabla 14 relaciona el estatus laboral al momento de realizar el curso INFOTEP (y las eventuales características de dicho estatus laboral, de implicar un trabajo), con el sexo del(a) egresado(a) encuestado(a). Si bien habría sido conceptualmente plausible desagregar dicho estatus laboral de acuerdo con nivel educativo de los encuestados, buscando evidencia de la influencia / pertinencia del

capital humano en el estatus laboral, estaríamos con ello incurriendo en una falacia intertemporal, pues no podemos establecer con certeza que los logros educativos ocurrieron previa o simultáneamente al momento en que los individuos tomaron el curso.

Aproximadamente uno de cada tres egresados (34.74%, 370 personas) estaba laborando al momento de tomar el curso. Si bien en niveles la representación es relativamente semejante, cuando se divide dicho número (174 mujeres y 196 hombres entre sus respectivos totales, 645 y 415) se ve que la ocupación efectiva de los hombres (47.72%) era casi el doble de las mujeres (26.69%).

**Tabla 14. Estatus laboral al momento de tomar la capacitación INFOTEP.**

CRITERIO		¿Trabajaba al Momento de tomar la formación?				TOTAL
		NO	SI			
			¿El trabajo era formal o informal?			
			FORMAL		INFORMAL	
			¿El trabajo estaba relacionado con la formación que estaba recibiendo?			
			NO	SI		
Sexo del(a)	Femenino	476	60	74	40	650
Encuestado(a)	Masculino	219	74	77	45	415
Egresado(a)						
TOTAL		695	134	151	85	1065

Fuente: INFOTEP. Procesamiento de los autores.

Si asumimos el estatus laboral al momento de realizar el curso como una variable proxy a la acumulación de experiencia laboral, que eventualmente podría explicar la inserción laboral de los egresados, este aspecto debe ser tenido en cuenta. Del total de ocupados (370), 285 (77.02%) manifestaron tener un trabajo formal. En este caso, los aparentes niveles semejantes al interior del empleo formal (134 mujeres, 151 hombres) reflejan una situación más equitativa en términos de formalidad laboral. Las mujeres representan el 47.01% del empleo formal y el porcentaje de su empleo formal dentro del empleo total femenino es de 77.01%, y los hombres representan el 52.99% del empleo formal y el porcentaje de su empleo formal dentro del empleo total masculino es de 77.04%.

Para los ocupados formales, la pertinencia de la formación en función de la actividad laboral que desarrollaban en dicho momento no es particularmente definitiva: el 55.22% de las mujeres y el 49% de los hombres ocupados en trabajos formales al momento de recibir la formación en INFOTEP la consideraban relacionada con su formación. Posteriormente, al analizar el estado laboral de los egresados al momento de desarrollar la encuesta, retomaremos el eventual efecto de esta relación.

## **ii. Desempleo al momento de recibir la capacitación / Desempleo al momento de desarrollar la Encuesta a Egresados**

Como ya anotamos, la redacción de la pregunta sobre el eventual desempleo del egresado al momento de recibir la capacitación, implica un cambio de foco, pues no indaga sobre las circunstancias en el momento de tomar la capacitación, sino sobre los seis meses inmediatamente anteriores al momento de aplicación de la Encuesta a Egresados. Esto es particularmente nocivo si tenemos en cuenta que las preguntas detalladas sobre desempleo (preguntas 14.3 a 14.7 del cuestionario: búsqueda activa de empleo, razón del rechazo, necesidad de formación adicional, etc) se le aplicaron a quienes manifestaban estar desempleados al momento de recibir la formación (pregunta 14, 422 personas) y no a quienes declararon estar desempleados al momento en que se les aplica la Encuesta a Egresados (pregunta 15, 474 personas), con lo cual se cuenta con información detallada sobre el desempleo de 422 personas, en un momento previo al que efectivamente enfrenta un egresado (474 personas desempleadas al momento de contestar la Encuesta a Egresados). La Tabla 15 nos permite ilustrar comparativamente los dos patrones:

**Tabla 15. Encuesta a Egresados INFOTEP. Composición del Desempleo por Sexo y Nivel Educativo.**

Momento del Desempleo	Mujeres	Hombres	TOTAL
No tenían un trabajo al momento de tomar la formación INFOTEP			
Maestría y/o Posgrado	3	1	4
%	0,6%	0,2%	0,8%
Primaria completo	9	3	12
%	1,9%	0,6%	2,5%
Primaria incompleto	9	0	9
%	1,9%	0,0%	1,9%
Secundaria Completa	178	109	287
%	37,4%	22,9%	60,3%
Secundaria incompleto	35	16	51
%	7,4%	3,4%	10,7%
Técnico	0	0	0
%	0,0%	0,0%	0,0%
Universitario Completo	51	14	65
%	10,7%	2,9%	13,7%
Universitario Incompleto	191	76	267
%	40,1%	16,0%	56,1%
TOTAL	476	219	695
No tenían un trabajo al momento de aplicación de la Encuesta			

Maestría y/o Posgrado	1	0	1
%	0,3%	0,0%	0,3%
Primaria completo	8	4	12
%	2,2%	3,4%	1205,9%
Primaria incompleto	8	1	9
%	2,2%	0,8%	2109,0%
Secundaria Completa	131	49	180
%	36,8%	41,5%	20187,3%
Secundaria incompleto	31	12	43
%	8,7%	10,2%	24506,2%
Técnico	1	0	1
%	0,3%	0,0%	24606,5%
Universitario Completo	30	7	37
%	8,4%	5,9%	28320,9%
Universitario Incompleto	146	45	191
%	41,0%	38,1%	47500,0%
TOTAL	356	118	474

Fuente: INFOTEP. Procesamiento de los autores.

Tenemos entonces dos fuentes de información sobre desempleo de los egresados. Una, la provista por egresados que estaban desempleados al momento de recibir su formación, y dos, la provista por egresados desempleados al momento de serles aplicada la encuesta. La Tabla 16 ilustra la transición.

Un número mayor de personas expresaron estar desempleadas al momento de tomar la capacitación INFOTEP (695) que al momento de serles aplicada la Encuesta a Egresados (474), el cual es un elemento a resaltar, de cara a analizar la inserción laboral de los egresados. Para el segundo momento, 273 personas que habían estado desempleadas, habían conseguido un trabajo, pero, por otra parte, 45 personas que tenían un trabajo al momento de recibir su formación, no tenían uno al momento de serles aplicada la Encuesta a Egresados.

**Tabla 16. Encuesta a Egresados INFOTEP. Evolución de la situación laboral.**

CRITERIO		¿Trabaja actualmente?		TOTAL
		NO	SI	
¿Trabajaba al momento de tomar la formación?	NO	422	273	695
	SI	52	318	370
TOTAL		474	591	1065

Fuente: INFOTEP. Procesamiento de los autores.

Al analizar el estado laboral (trabajo / desempleo) se puede encontrar el nivel de formación que, según los desempleados, requieren dadas sus circunstancias. De acuerdo con la Tabla 17, el 73.23% (509 personas, 359 mujeres y 150 hombres) de los desempleados que respondieron la pregunta sobre requerimientos adicionales de formación, consideran que SI es necesaria dicha formación adicional.

**Tabla 17. Requerimiento de Formación Adicional por parte de los desempleados.**

Modalidad de la Formación Recibida	¿Requiere de más formación?	Mujeres	Hombres	TOTAL
Complementación	SI	55	20	75
	NO	4	5	9
	NS/NR	6	1	7
Formación Continua en Centros	SI	59	19	78
	NO	21	5	26
	NS/NR	3	2	5
Formación de maestros técnicos	SI	5	2	7
	NO	2	3	5
	NS/NR	0	1	1
Formación dual	SI	4	35	39
	NO	3	15	18
	NS/NR	0	2	2
Habilitación	SI	236	74	310
	NO	67	32	99
	NS/NR	11	3	14
TOTALES	SI	359	150	409
	NO	97	60	157
	NS/NR	20	9	29

Fuente: INFOTEP. Procesamiento de los autores.

En términos específicos, el 73.28% de los recipientes de formación a nivel de habilitación, que se encuentran desempleados, consideran que requieren de formación adicional. Para los restantes niveles de formación, estas son las expresiones de requerimiento: formación dual (66%), maestros técnicos (53.84%), formación continua (71.55%), y complementación (82.41%). Tal vez es conveniente que INFOTEP aborde un enfoque específico de género: el requerimiento de mayor formación es expresado en un 87.77% (359 de 409 egresados) por mujeres.

### iii. Egresados que contaban con trabajo al momento de aplicación de la Encuesta.

Al momento de aplicación de la Encuesta a Egresados, el 55.49% de los mismos manifestó estar trabajando. De dicho total, el 49.74% eran mujeres y el 50.26% restante, hombres. Al analizar la empleabilidad por nivel de formación, una preocupante apreciación directa es que para todos los niveles de formación, las mujeres muestran niveles de empleo menores que los de sus respectivas contrapartes masculinas: el número de mujeres con trabajo, como porcentaje del total de mujeres con el respectivo nivel de formación siempre es menor que el porcentaje de hombres con trabajo en dicho nivel de formación. La menor brecha de ocupación se da para el nivel de formación dual (3.6 puntos porcentuales), nivel para el cual el porcentaje de mujeres ocupadas es el mayor (66.67%). Claramente, este elemento (alta ocupación de egresadas de cursos de formación dual) es la explicación de una menor brecha de ocupación, toda vez que en los restantes niveles de formación, los hombres exhiben niveles de ocupación superiores (entre el 66.67% y el 84.21%) al nivel más alto de ocupación de las mujeres, que como ya mencionamos, es el de egresadas de formación dual (66.67%).

**Tabla 18. Encuesta a Egresados INFOTEP. Porcentaje de Ocupados Total, por sexo y nivel de formación.**

¿Trabaja Actualmente?	Sexo	Modalidad de la Formación					
		Complementación	Formación continua en centros	Formación de maestros técnicos	Formación dual	Habilitación	Total
No	Femenino	51	49	5	4	247	356
	Masculino	21	10	3	22	62	118
	Total	72	59	8	26	309	474
Si	Femenino	41	63	4	8	178	294
	Masculino	42	34	16	52	153	297
	Total	83	97	20	60	331	591
Total	Femenino	92	112	9	12	425	650
	Masculino	63	44	19	74	215	415
	Total	155	156	28	86	640	1065
Brecha de Ocupación Hombres - Mujeres		22,10%	21,02%	39,77%	3,60%	29,28%	26,34%

Fuente: INFOTEP. Procesamiento de los autores.

Analizando la pertinencia de la formación sobre la inserción laboral, analizamos la relación entre la modalidad de la formación y el empleo actual de lo(a)s egresado(a)s que tienen trabajo, en la Tabla 20. De ellos, el 44% (204) consideran que su trabajo actual no está relacionado con la formación recibida en INFOTEP. El 56% restante si considera que existe dicha relación. El 46% de los egresados que recibieron formación tipo habilitación laboran en actividades relacionadas con dicha formación, porcentaje que es del 75% para egresados de formación dual, 84% para formación de maestros técnicos, 65.33% para formación continua y 54.38% para egresados de formación en complementación.

**Tabla 19. Encuesta a Egresados INFOTEP. Relación entre el trabajo actual (al momento de la encuesta) y la modalidad de formación.**

Modalidad de Formación	¿Su trabajo actual se relaciona con su formación?				TOTAL
	NO		SI		
	Mujeres	Hombres	Mujeres	Hombres	
Complementación	17	9	11	20	57
Formación Continua	15	11	37	12	75
Formación de maestros técnicos	1	2	3	13	19
Formación Dual	2	10	5	31	48
Habilitación	64	73	61	57	255
TOTAL	99	105	117	133	454

Fuente: INFOTEP. Procesamiento de los autores.

Para tener una relativa consideración de pertinencia laboral de la formación de los egresados contemplados en la Encuesta (desde la perspectiva de la oferta, es decir, el egresado, no su empleador) analizamos en la Tabla 20 los patrones de respuesta a la combinación de preguntas sobre el estado laboral del egresado (si tenía un trabajo al momento de la Encuesta) y si las labores se relacionaban con su formación, pero en este caso presentamos la información como porcentaje relativo del total de egresados formados por modalidad, de la Encuesta.



**Tabla 20. Encuesta a Egresados INFOTEP. En caso de tener un trabajo (al momento de la encuesta), ¿dicho trabajo se relaciona con su formación?**

Modalidad de Formación	Respuesta	Sexo		Total
		Femenino	Masculino	
Complementación	NS/NR	31,71%	30,95%	31,33%
	No	41,46%	21,43%	31,33%
	Si	26,83%	47,62%	37,35%
Formación continua en centros	NS/NR	17,46%	32,35%	22,68%
	No	23,81%	32,35%	26,80%
	Si	58,73%	35,29%	50,52%
Formación de maestros técnicos	NS/NR	0,00%	6,25%	5,00%
	No	25,00%	12,50%	15,00%
	Si	75,00%	81,25%	80,00%
Formación dual	NS/NR	12,50%	21,15%	20,00%
	No	25,00%	19,23%	20,00%
	Si	62,50%	59,62%	60,00%
Habilitación	NS/NR	28,09%	15,03%	22,05%
	No	37,64%	47,71%	42,30%
	Si	34,27%	37,25%	35,65%
Total	NS/NR	25,51%	19,87%	22,67%
	No	34,69%	35,35%	35,03%
	Si	39,80%	44,78%	42,30%

Fuente: INFOTEP. Procesamiento de los autores.

La formación dual (60%) y la formación de maestros técnicos (80%) parecen ser las modalidades con mayor pertinencia laboral. Esto debe tomarse con beneficio de inventario por dos aspectos: uno, el elevado nivel de NS/NR en todas las modalidades y en el total como un todo. Y dos, el bajo número de egresados de formación como maestros técnicos.

Para finalizar esta sección, la Tabla 21 presenta el resumen final de las características más notorias de los egresados ocupados. No existe una diferenciación de posición ocupacional por sexo (los valores y los porcentajes son extremadamente semejantes). La informalidad laboral es mayor entre egresadas (una de cada cuatro egresadas ocupadas tiene un trabajo informal) que entre egresados (uno de cada cinco egresados tiene un trabajo informal). Finalmente, el 54% de los trabajos formales se relacionan con la formación recibida, y dicha relación es mayor para los hombres (55%) que para las mujeres (53.42%).

**Tabla 21. Encuesta a Egresados INFOTEP. Características de los Egresados Ocupados (al momento de la encuesta).**

Criterio	Mujeres	Hombres	TOTAL
Posición Ocupacional			
Cuenta Propia	90	87	177
Emprendedor	6	7	13
Asalariado	190	194	384
Otras Ocupaciones	8	9	17
TOTAL	294	297	591
Si tiene trabajo, ¿Es el mismo empleo?			
No	143	123	266
Si	151	174	325
TOTAL	294	297	591
¿Su trabajo actual es formal?			
No	75	59	134
Si	219	238	457
TOTAL	294	297	591
Si su trabajo actual es formal, ¿Se relaciona con su formación?			
No	102	105	207
Si	117	133	250
TOTAL	219	238	457

Fuente: INFOTEP. Procesamiento de los autores.

#### **e. Análisis econométrico de determinantes de la relación entre tenencia de trabajo y formación recibida en INFOTEP**

En esta sección buscamos identificar los determinantes de la probabilidad de que una persona formada en INFOTEP tenga trabajo, a partir de características como recibir apoyo de parte de INFOTEP, estar estudiando, haber solicitado cursos de formación, haber estado buscando Empleo, el género, la edad, el ultimo nivel educativo aprobado, la región y la modalidad del curso tomado a través de INFOTEP. Dicha estimación se realiza para el conjunto de egresados encuestados por INFOTEP.

Dado dicho conjunto muestral, el modelo probabilístico de variable dicotómica a estimar en este caso particular tiene las siguientes especificaciones:

$$Pr(y_i = \text{estar empleado}) = F(X_i\beta + \varepsilon_i) \quad (1)$$

En donde:

→  $y_i = 1$  si el egresado encuestado  $i$  está trabajando actualmente

→  $\beta$  es el vector de parámetros de interés, que relacionan las variables con la probabilidad de que el hogar reciba remesas

→  $X_i$  es un vector de características de los emigrantes a nivel de egresado encuestado. Entre estas están:

- a) INFOTEP: si la persona encuestada considera que INFOTEP le ayudo.
- b) Sexo: hombre y mujer (categoría excluyente de comparación).
- c) Estudia: si el encuestado actualmente se encuentra estudiando
- d) Educación: último nivel educativo que posee el encuestado
- e) Edad: rangos de edad que van desde los 14 a los 41 o más
- f) Modalidad: se refiere al tipo de formación educativa que tomó a través de INFOTEP.
- g) Región: control geográfico del lugar donde la persona tomó los cursos de formación.

### Resultados de la Estimación

La tabla 25 presenta las estadísticas descriptivas de las variables utilizadas en la estimación del modelo. En promedio, el 55,5% de los egresados encuestados se encuentran trabajando actualmente. Esta tabla, también se puede observar que el 34,9% de los egresados encuestados reportaron que el proceso que surtieron con INFOTEP les ayudó a conseguir Empleo. De igual forma, el 60% de los egresados encuestados son de sexo femenino. Asimismo, el 38,8% de los encuestados reportaron estar estudiando y el 60% reportaron haber tomado curso de habilitación con INFOTEP en la zona Central. Finalmente, el 48,8% asegura haber buscado trabajo en los últimos 6 meses.

**Tabla 22. Estadísticas descriptivas de los egresados encuestados por INFOTEP, 2019**

Variable	Mean	Std. Dev.
Trabaja actualmente	0.555	0.497
Le ayudo Infotep	0.349	0.477
Mujer	0.608	0.488
Hombre	0.392	0.488
Educación primaria	0.028	0.166
Educación secundaria	0.479	0.500
Educación superior	0.482	0.500
Educación posgrado	0.010	0.101
14 a 22	0.209	0.407
23 a 28	0.319	0.466
29 a 34	0.187	0.390
35 a 40	0.131	0.338
41 ó más	0.154	0.361
Complementación	0.146	0.353

Variable	Mean	Std. Dev.
Formación continua en centros	0.146	0.353
Formación de maestros técnicos	0.026	0.160
Formación dual	0.081	0.273
Habilitación	0.600	0.490
Estudia actualmente	0.388	0.488
Ha buscado trabajo	0.482	0.500
Solicito cursos de formación	0.213	0.410
Regional Central	0.428	0.495
Regional Este	0.133	0.340
Regional Norte	0.345	0.476
Regional Sur	0.093	0.291

**Fuente:** elaboración de los autores a partir de Encuesta egresados INFOTEP, 2019

La tabla 23 presenta los resultados obtenidos de la estimación del modelo probabilístico presentado en la ecuación (1). En esta se puede observar como el hecho de ser hombre, aumenta la probabilidad de estar trabajando actualmente en un 18,7% comparado con ser mujer. De igual forma, tener un nivel de educación de posgrado aumenta la probabilidad de estar empleado en un 50% respecto a los egresados con educación básica primaria. El hecho de haber sido apoyado por INFOTEP para conseguir trabajo aumenta la probabilidad de estar empleado en un 37,4% en comparación con no haber sido apoyado. Finalmente estar estudiando o haber solicitado cursos de formación reduce la probabilidad de estar empleado en un 10,2% y 9,5% respectivamente.

**Tabla 23. Modelo probabilístico de estimación de los determinantes de empleabilidad**

VARIABLES	Trabaja actualmente Efectos marginales
Le ayudo Infotep	0.374*** (0.052)
Hombre	0.187*** (0.056)
Estudia actualmente	-0.102* (0.052)
Ha buscado trabajo	-0.052 (0.049)
Solicito cursos de formación	-0.095* (0.057)
Nivel referencia: Educación primaria	
Educación secundaria	0.090 (0.195)
Educación superior	0.164 (0.194)
Educación posgrado	0.499*** (0.170)
Edad referente: 14 a 22	
Edad 23 a 28	0.017 (0.066)
Edad 29 a 34	-0.011 (0.079)
Edad 35 a 40	-0.028 (0.091)
Edad 41 o más	-0.023 (0.087)
Modalidad referente cursos: Complementación	
Modalidad Formación continua en centros	0.139 (0.087)
Modalidad Formación de maestros técnicos	-0.128 (0.171)
Modalidad Formación dual	0.183

VARIABLES	Trabaja actualmente
	Efectos marginales
	(0.115)
	0.021
Modalidad Habilitación	(0.069)
Región referente: Central	
	0.134*
Regional Este	(0.081)
	-0.046
Regional Norte	(0.055)
	-0.116
Regional Sur	(0.081)
Observaciones	509
Pseudo R2	0.17

Errores estándar robustos entre paréntesis.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Fuente:** elaboración de los autores a partir de Encuesta egresados INFOTEP, 2019

#### f. Trazabilidad de los egresados en Encuesta a Egresados INFOTEP en el Sistema Integrado de Registro Laboral (SIRLA).<sup>9</sup>

Gracias al apoyo de INFOTEP, fue posible ubicar el número de CIE (Cédula de Identidad y Electoral) de 900 egresados incluidos en la Encuesta a Egresados. La eficiente gestión del Ministerio de Trabajo fue requerida para realizar la búsqueda de la información personal de estos individuos en el SIRLA (Sistema Integrado de Registro Laboral), con el fin de determinar sus eventuales circunstancias laborales para el año 2020 (como cabe recordar, la Encuesta a Egresados fue aplicada entre mayo y junio del 2019). La búsqueda de los egresados INFOTEP en SIRLA se realizó con dos propósitos. Uno, metodológico, la integración de bases de datos como aplicación de Big Data. Y el segundo, conceptual, analizar las circunstancias laborales del 84.5% de los egresados a quienes se les aplicó la Encuesta de Egresados de 2019.

El Ministerio de Trabajo, luego de procesar el listado de 900 egresados que habían participado en la Encuesta de Egresados de 2019, y de quienes se pudo ubicar su número de CIE, pudo identificar

<sup>9</sup> En el Anexo 3 se incluye un ejercicio de determinantes salariales (ecuación minceriana) de los egresados INFOTEP identificados en SIRLA. Dado que la información de SIRLA no incluye remuneraciones, se imputaron las remuneraciones reportadas en la Encuesta de Egresados. Se presenta como un ejercicio del potencial de trabajo a partir de integración de bases de datos, pero dada la incipiente información de ingresos reportada en SIRLA, no se pretende que sea considerado como un análisis de las remuneraciones a 2020.

información relacionada con 352 egresados, que en el año 2020 trabajaban para alguna de las empresas registradas en SIRLA. Algunos de ellos habían sido beneficiarios del Programa FASE, desarrollado por el gobierno dominicano para ofrecer apoyo ante la crisis de la Pandemia.

Resulta inevitable abordar las circunstancias que pueden explicar un margen de formalidad<sup>10</sup> / retención / ubicación del 39.11% (352/900):

- La primera de ellas tiene que ver con los efectos laborales de la Pandemia en República Dominicana: reducción del empleo y del desempleo, con un notorio incremento de la inactividad laboral. En la primera sección de este informe detallamos que el empleo se redujo en 227.746 personas. Los desempleados, en lugar de incrementarse, se redujeron en casi 80.000 personas (79.980, el 27% de los desempleados del 2019). El paso a la inactividad implicó la salida de 307.727 personas de la fuerza de trabajo (Población Económicamente Activa).
- La segunda es el efecto de la informalidad laboral, dadas las características del sistema de registro de protección social (Sistema Integrado de Registro Laboral, SIRLA). Desde la perspectiva de las diferentes categorías ocupacionales (formas de trabajo y empleo) el SIRLA se concentra en la relación laboral de dependencia (empleo)<sup>11</sup> y como tal, “es para registrar las empresas, establecimientos y reportar sus trabajadores fijos”<sup>12</sup>. Dada esta premisa, aún si no existiesen las circunstancias económicas de la Pandemia, no aparecerían en SIRLA 668 egresados de INFOTEP (474 desempleados, 177 ocupados cuenta propia y 17 con otras ocupaciones). Un punto adicional sería considerar a los 13 emprendedores como empleadores formales y, por ende, sujetos a registrar su empresa y empleados en SIRLA.
- El tercer punto a tener en cuenta es la expresa informalidad de las relaciones laborales que manifiestan enfrentar los egresados de INFOTEP. Esto se aprecia en dos circunstancias de evidencia:
  - La primera, en el empleo al momento de recibir la capacitación. En dicho momento 370 egresados manifestaban haber estado laborando. De ellos, 85 (22%) consideraban que su trabajo no era formal.
  - Al momento de aplicación de la Encuesta a Egresados (primer semestre del 2019), 591 (55.49%) estaban trabajando. De ellos, 134 (de nuevo, el 22%) no consideraban su trabajo formal.

El bajo nivel de ubicación de egresados INFOTEP en SIRLA impidió el análisis a profundidad de sus circunstancias laborales, toda vez que no se pueden hacer inferencias estadísticas a partir de 352 egresados que a su vez han sido extraídos de una encuesta con 1065 participantes, la cual, si bien puede revestir criterios conceptuales de significancia como para ser considerada representativa, no puede considerar estos 352 egresados como representativos de los patrones de inserción laboral de los egresados INFOTEP (representan el 0.0007 de los 441.609 egresados documentados en la base de egresados).

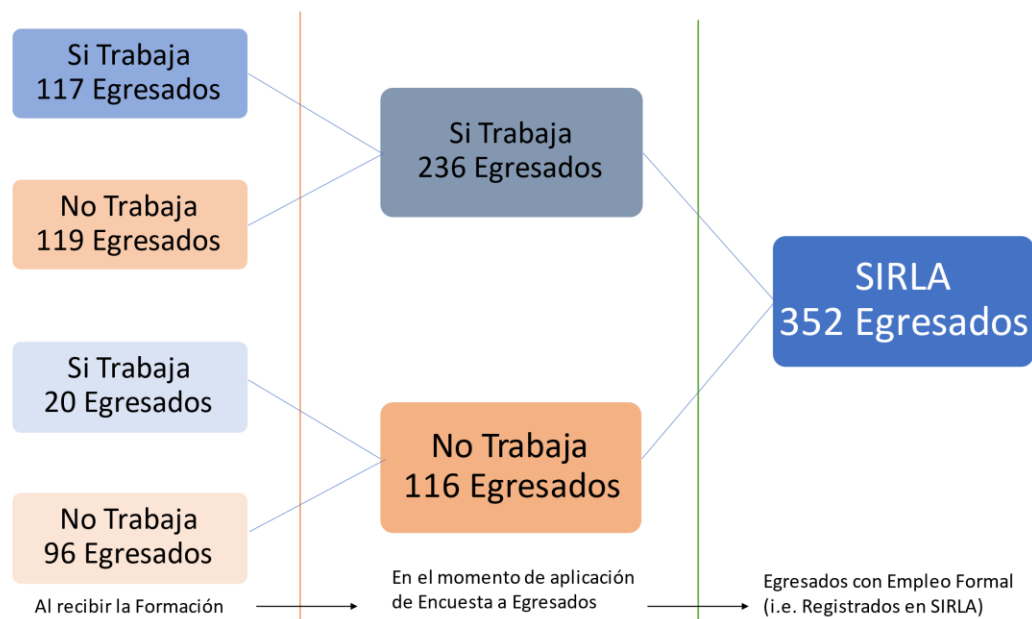
---

<sup>10</sup> El hecho de que un egresado no se ubique en SIRLA no necesariamente implica que esté desempleado, y por ello nuestra redacción es cuidadosa de no definir a los 548 egresados de INFOTEP no ubicados en SIRLA como “desempleados”. Definimos a los 352 detectados como ocupados formales, pero no podemos hacer inferencias sobre el estado laboral de los 548 no detectados (ya sea si son ocupados informales, desempleados o laboralmente inactivos).

<sup>11</sup> Cabe anotar que SIRLA permite el registro de “personas físicas no normalizadas” (Ministerio de Trabajo, “Instructivo para el Registro de los Formularios DGT-3 y DGT-4”, pág. 3).

<sup>12</sup> “Qué es el SIRLA y cual (SIC) son los requisitos”. <https://yotrabajo.info/que-es-sirla/>

**Figura 5. Transiciones Laborales entre empleo y desempleo. Egresados INFOTEP en SIRLA 2020.**



Fuente: Combinación de bases de datos. Procesamiento de los autores.

Resulta coherente argumentar que estos 352 egresados representan la formalidad laboral perfecta de la cual se puede beneficiar un egresado de INFOTEP. Tienen un empleo asalariado, están registrados como empleados de una firma inscrita en SIRLA y gran parte de ellos han sido beneficiarios de una o dos etapas del Programa FASE. En razón a lo anterior, a continuación haremos un breve análisis de sus características. El proceso de empleabilidad de los 352 egresados de INFOTEP que registran tener un empleo formal en SIRLA se ilustra en la Figura 5. El 33.23% de estos egresados han mantenido un trabajo remunerado en los tres momentos analizados (al recibir la capacitación, en la Encuesta de Egresados y en el SIRLA 2020). En este sentido, la formación parece redundar no sólo en la posibilidad de obtener un trabajo, sino que dicho trabajo sea asalariado formal, potenciando la relación entre formación y experiencia laboral: el 27.27% de los egresados INFOTEP con empleo formal en SIRLA, no tenían trabajo al momento de recibir la capacitación y tampoco al momento de la encuesta, pero si lo tienen en el año 2020.

#### **i. Formación Recibida por egresados INFOTEP ubicados en SIRLA**

La Tabla 22 presenta la clasificación de los egresados INFOTEP ubicados en SIRLA en el año 2020. Del total de 352 egresados, 183 (52%) son hombres y 169 (48%) mujeres. Como punto de referencia, incluimos también la misma estructura, pero tomada de la Encuesta a Egresados, aquellos que manifestaron estar ocupados al momento de dicha encuesta. Como mencionamos al analizar la Tabla 22, la relación entre modalidad de formación y empleabilidad no resulta causal ni directa y debe tomarse con cautela y rigor: existe un alto porcentaje de no respuesta y la relación de pertinencia sólo supera el 50% de respuesta para una modalidad de formación (maestro técnico).

Desde la perspectiva de la modalidad de formación, una medida relativa de inserción laboral formal es el porcentaje de egresados INFOTEP por modalidad de formación, que efectivamente aparecen



registrados en el SIRLA. Dichos porcentajes son respectivamente 71.67% de los egresados de la Formación Dual, 60% de los egresados de la Formación de Maestros Técnicos, 58.31% de Habilitación, 57.83% de Complementación y 57.73% de Formación Continua en Centros, y se obtienen analizando, para cada modalidad de formación, qué porcentaje de los egresados que al momento de la Encuesta tenían un trabajo, aparecen efectivamente en SIRLA (por ejemplo, 48 egresados con Complementación son el 57.83% de 83 egresados con Complementación que, para el momento de la Encuesta estaban ocupados).

**Tabla 24. Egresados ocupados? de acuerdo con SIRLA 2020 y con Encuesta a Egresados 2019 por Modalidad de Formación y Sexo.<sup>13</sup>**

Modalidad de Formación	Sexo del Egresado		TOTAL
	Femenino	Masculino	
ENCUESTA A EGRESADOS 2019 (Ocupados)			
Complementación	41	42	83 (14.04%)
Formación Continua en Centros	63	34	97 (16.41%)
Formación de Maestros Técnicos	4	16	20 (3.38%)
Formación Dual	8	52	60 (10.15%)
Habilitación	178	153	331 (56.01%)
TOTAL EN LA ENCUESTA	294	297	591
SIRLA 2020			
Complementación	28	20	48 (13.64%)
Formación Continua en Centros	38	18	56 (15.91%)
Formación de Maestros Técnicos	4	8	12 (3.41%)
Formación Dual	2	41	43 (12.22%)
Habilitación	97	96	193 (54.83%)
TOTAL EN SIRLA 2020	169	183	352

Fuente: Ministerio de Trabajo (SIRLA) e INFOTEP. Procesamiento de los autores. Como tasa de ocupación se define el % de egresados ocupados, por modalidad de formación, entre el total de egresados formados en dicha modalidad de formación.

<sup>13</sup> Complementación: Curso técnico de perfeccionamiento en una ocupación. Formación Continua en Centro: Curso relativamente largo de formación en habilidades y conocimientos tecnológicos sobre una ocupación. Formación de Maestro Técnico: Formación avanzada dirigida a trabajador calificado, sobre actividades de supervisión y asesoría de la producción, desarrollo de planes de producción y desarrollo de nuevos trabajadores. Formación Dual: Carrera técnica de modalidad combinada docente – empresarial dirigida a formación de técnico en una ocupación específica. Habilitación: curso técnico en que se imparten los conocimientos básicos para iniciarse en una ocupación. Tomado de Formación Presencial Infotep. ([www.infotep.gob.do](http://www.infotep.gob.do))

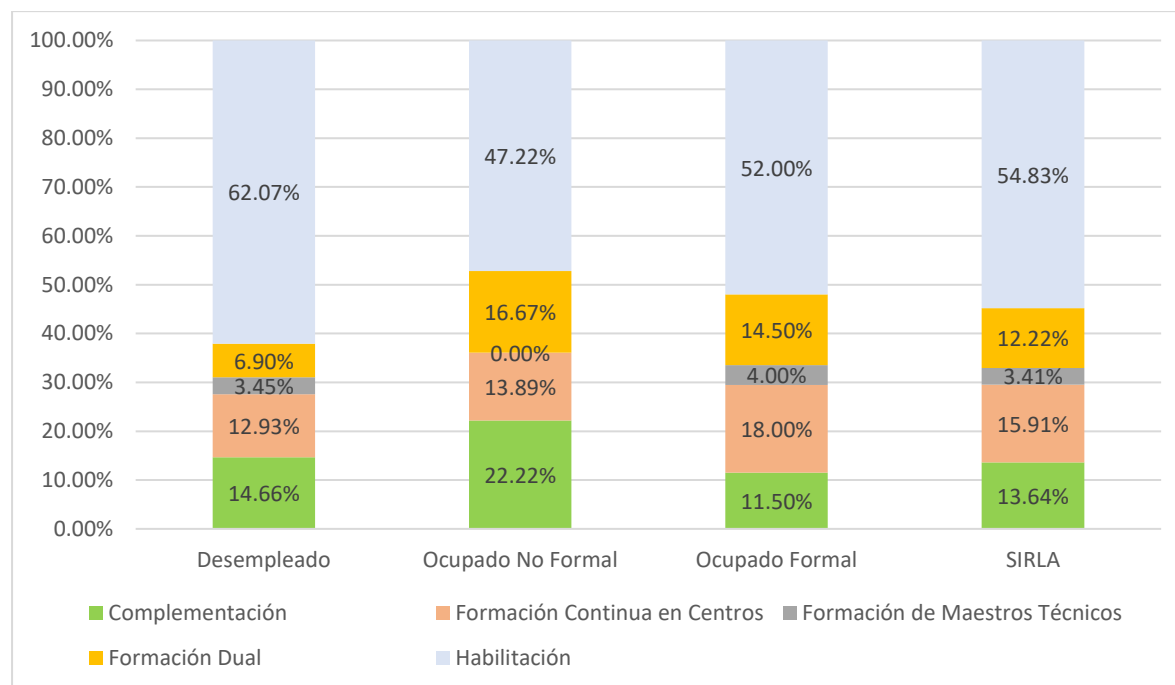
## **ii. Trayectorias en Trabajo Formal**

Para profundizar en la relación entre modalidad de formación y trabajo podemos retomar los patrones de respuesta a las preguntas sobre formalidad laboral en la Encuesta a Egresados provistos por los egresados identificados en SIRLA, y extenderlos a su condición laboral formal actual, puesto que por definición estos últimos son ocupados formales.

El análisis de formalidad laboral permite reforzar las particulares características laborales de los 352 egresados identificados en SIRLA, puesto que parece existir un círculo virtuoso que redundaba en la mejora de sus condiciones laborales. Al momento de responder la Encuesta, el 67% (236) estaban ocupados, y de ellos el 84.74% (200) manifestaban tener trabajos formales. Para el año 2020, de acuerdo con las condiciones para estar afiliado en SIRLA, no sólo el 100% estaban ocupados, sino que todos tenían trabajos formales.

Desde la perspectiva de cómo las modalidades de formación se relacionan con el empleo formal, podemos analizar su incidencia en diferentes circunstancias laborales. El gráfico 3 ilustra la descomposición porcentual de las modalidades entre los desempleados y ocupados (formales y no formales) al momento de la Encuesta a Egresados, y cómo se distribuyen dichas modalidades en los egresados ubicados en SIRLA. La formación continua en centros tiene una clara influencia sobre el empleo formal, así como la formación de maestros técnicos. La alta participación de la formación en habilitación dentro del total de egresados capacitados hace que el efecto de esta modalidad no sea fácilmente discernible.

**Gráfico 3. Composición porcentual de las categorías laborales por modalidad de formación. Encuesta a Egresados vs. Seguimiento en SIRLA.**



Fuente: Combinación de bases de datos. Procesamiento de los autores.

Las columnas de desempleado, ocupado no formal y ocupado formal hacen referencia a los posibles estados laborales declarados por los egresados al momento de la Encuesta (2019) de acuerdo con su modalidad de formación. Dado su nivel de profundización y requerimiento previo, la modalidad de maestro técnico se ve reflejada en empleo formal. De forma opuesta, la habilitación, como formación inicial no diferenciada, se ve representada de forma notoria en el desempleo identificado en la Encuesta.

Podemos resaltar algunos resultados del gráfico 3 a la luz de ciertos patrones recurrentes de la inserción laboral de egresados de formación técnica:

- Los egresados de habilitación y cursos cortos, que suelen contar con competencias más básicas, tienden a ser mayoría en los desempleados y por oposición los egresados de formación dual – usualmente atada a una práctica en la empresa- tienen muy bajo peso en la categoría de desempleados.
- De acuerdo a la Encuesta de Impacto, Los egresados de complementación trabajan en su mayoría en el sector informal.
- La formación en centros presenta más ocupados formales tanto en la encuesta como en el SIRLA.
- En los casos formación en centro y dual, los desempleados son menores que los ocupados

## **REFERENCIAS BIBLIOGRÁFICAS**

- Guataquí, J., Cárdenas, J. y Montaña, J. (2015), “Metodología para el análisis de demanda laboral mediante datos de Internet: el caso colombiano”, *Revista de Economía del Rosario*, 18(1).
- Guataquí, J., Cárdenas, J. y Montaña, J. (2014), “La problemática del análisis laboral de demanda en Colombia”, *Perfil de Coyuntura Económica*, 24.
- Johnson, E. M. (2016). *Can Big Data Save Labor Market Information Systems?* RTI Press.
- Mezzanzanica, M., y Mercurio, F. (2019). *Big data for labour market intelligence: An introductory guide*. Torino: European Training Foundation.
- Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining*. John Wiley & Sons.
- Jurka, T. P., Collingwood, L., Boydston, A. E., Grossman, E., & van Atteveldt, W. (2013). RTextTools: A Supervised Learning Package for Text Classification. *R Journal*, 5(1).
- Wowczko, I. A. (2015, December). Skills and vacancy analysis with data mining techniques. In *Informatics* (Vol. 2, No. 4, pp. 31-49). Multidisciplinary Digital Publishing Institute

## ANEXO 1. CARACTERIZACION DEL MERCADO LABORAL DE REPÚBLICA DOMINICANA

Para tener una mejor comprensión del mercado de trabajo se ha elaborado este análisis que se ofrece como anexo.

### **a. Tendencias de Largo Plazo 2008 - 2019**

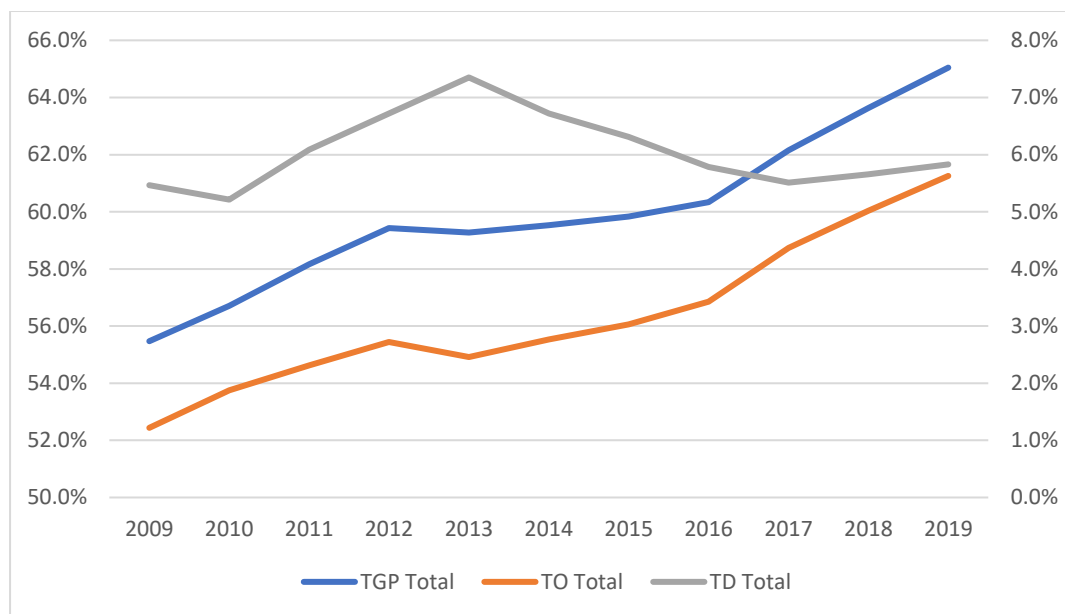
A modo de marco general, los principales aspectos laborales analizados (participación, ocupación y desocupación) configuran el siguiente panorama:

- Una baja participación con tendencia creciente, dinamizada en particular por el relativo incremento en la participación femenina en el último lustro de los 2010s.
- La tasa de ocupación mantiene una tendencia promedio creciente en el periodo analizado, con un notorio aporte de la participación femenina, que pasó a niveles del 40% en el último quinquenio.
- La desocupación exhibe dos periodos claramente definidos (2008-2013 y 2014 en adelante), con un primer patrón de crecimiento (pasa de 4.8% a 7.4%) y uno posterior de reducción (baja de 7.4% a 5.7%). Existe cierta tendencia a converger a una tasa de desocupación cercana al 5.5%.

Como puede apreciarse en el gráfico 1, existe una tendencia creciente a largo plazo tanto de la participación laboral como de la ocupación. Si bien la brecha entre la tasa de participación y la tasa de ocupación es de 3.5 puntos porcentuales promedio en el periodo 2008 – 2019, el comportamiento en el segundo lustro de los 2010s refleja verdaderos logros en la generación de empleo, pues si bien la brecha entre la ocupación y la participación es de 3.6-3.8 pp, dicha brecha se da cuando los niveles de participación y desempleo son superiores al 60%. En particular, los logros en generación de empleo más notorios se pueden identificar en el lapso 2014-2017, en que la tasa de crecimiento de la ocupación duplica a la de la participación laboral.

Si tomamos las tasas de crecimiento de TGP y TD como indicadores de la pendiente del comportamiento de la variable, podemos apreciar una sorprendente estabilidad de las dos variables a lo largo del periodo de análisis (tasa de crecimiento de la TGP, 1%; tasa de crecimiento de la TO, 1.2%).

**Gráfico 1. República Dominicana. Participación Laboral – Ocupación y Desocupación.**

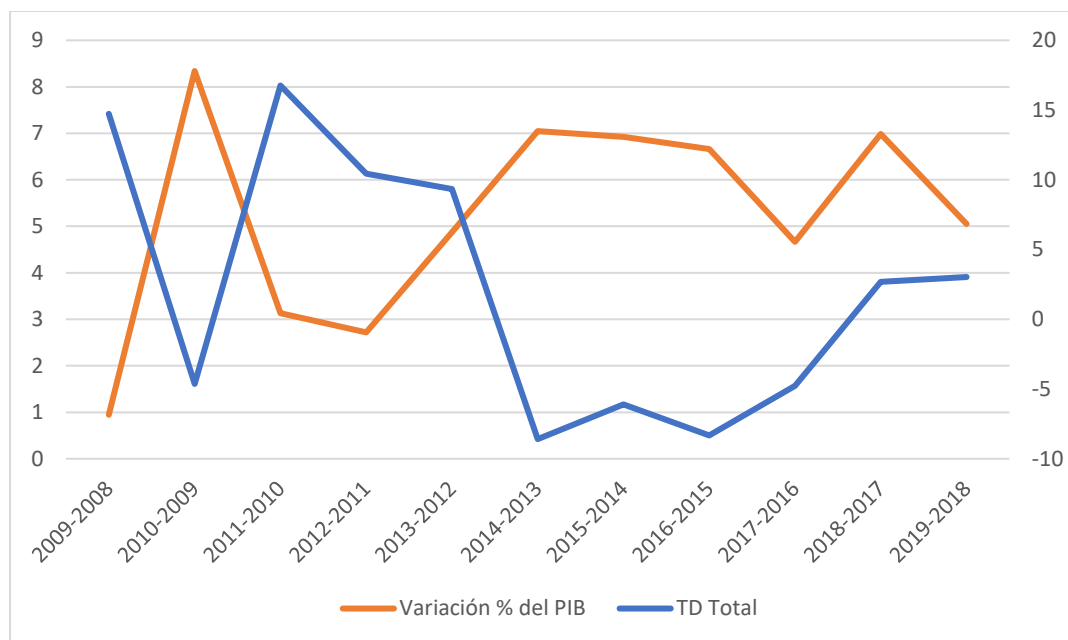


Fuente: ENFT – ECNFT. BCRD.

Existe otro aspecto importante a tener en cuenta en el análisis, y es el efecto del nivel de las variables analizadas sobre las tasas de crecimiento. El orden de magnitud de los niveles de los indicadores laborales se debe tener en consideración cuando se relacionan con la actividad económica buscando aproximarse de alguna forma a la tipología de la desocupación. De acuerdo con cifras del Banco Mundial, en el contexto económico del siglo XXI, el crecimiento porcentual anual del PIB de República Dominicana (PIB real a precios constantes) alcanzó su máximo nivel en el año 2005 (9.428%), a partir de dicho año se produjo una caída pronunciada en el ritmo de la actividad económica hasta el año 2009, cuando el PIB sólo creció 0.946%.

El gráfico 2 ilustra el comportamiento de la variación porcentual del PIB y el crecimiento de la tasa de desocupación. La relación de “espejo” muestra la forma en que el desempleo está negativamente asociado al ritmo de la actividad económica, tanto en su nivel como en la persistencia de su tendencia al alza o a la baja. Dicho comportamiento inverso parece mantener su patrón durante todo el periodo analizado, pero la reacción de respuesta es particularmente preocupante a partir del año 2017, por dos elementos llamativos: i) a pesar de que el PIB tiene un crecimiento anual entre 2018 y 2017 de 6.98%, la desocupación crece en 2.69%, y ii) si bien el PIB crece a una tasa menor entre 2019 y 2018 (5%), pero aún positiva, esto no implica una reducción de la desocupación: pasa de 5.7 a 5.8%, un crecimiento del 3%.

**Gráfico 2. Tasa de Crecimiento de la Desocupación y el PIB 2008-2019**



Fuente: ENFT – ECNFT. BCRD. Cálculos de los autores.

Los comportamientos aquí anotados en torno a la relación entre el nivel y ritmo de la actividad económica y la forma en que la misma afecta la desocupación llevan a dos aspectos extremadamente llamativos. Uno, el efecto del nivel de las variables analizadas sobre las tasas de crecimiento. Una tasa de desocupación abierta que oscila entre el 4.8% (nivel mínimo, 2008) y 7.4% (nivel máximo, 2013) implica que variaciones cuantitativamente pequeñas se vean reflejadas en tasas de crecimiento inusualmente elevadas. Esto, y la reciente falta de respuesta de la tasa de desocupación al crecimiento relativamente sostenido del PIB entre 2014 y 2019 llevan a una interesante conclusión, la cual en cierto modo da un enorme sustento al objetivo de este estudio: la tasa de desocupación no parece ser un indicador oportuno, dinámico y responsivo del mercado laboral de República Dominicana, desde la perspectiva de políticas públicas, y por ende es importante avanzar en otras caracterizaciones instrumentales de diferentes facetas de dicho mercado laboral.

## b. Coyuntura Laboral 2018 – 2020

Esta sección se ha desarrollado con tres propósitos: uno, ofrecer un análisis específico de la coyuntura laboral reciente, dejando a la sección anterior a cargo de las tendencias estructurales y de largo plazo. Dos, para ofrecer una caracterización regional específica y tres, para profundizar en ciertas variables o aspectos de particular interés para los propósitos de este estudio. En esta sección realizaremos un análisis de aspectos específicos de interés para la coyuntura laboral que comprende los años 2018 y 2019 y el primer semestre de 2020.

### Indicadores Laborales por Edad y Sexo

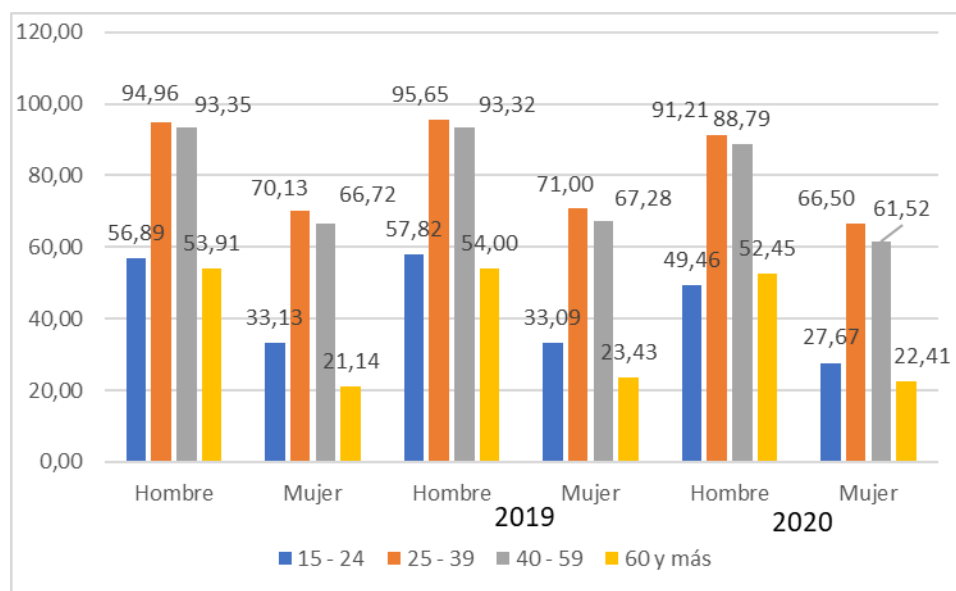
Teniendo como principal criterio de análisis los periodos de escolaridad y dos periodos diferenciados de vida laboral, los indicadores laborales fueron divididos en cuatro grandes grupos: 15-24 años, 25-39

años, 40-59 y 60 y más años de edad. Analizaremos la estructura de las tasas de participación, ocupación y desocupación, por sexo, en tres diferentes puntos del tiempo: 2018, 2019 y 2020.

### • Participación Laboral

Resultan sorprendentes los comportamientos específicos que se pueden encontrar al analizar la participación laboral de los diferentes grupos etareos, si tenemos en cuenta que la tasa global de participación, como un todo, asumió valores de 64.47, 65 y 60.13% para 2018, 2019 y el primer semestre de 2020 respectivamente. Los patrones de participación, analizados sólo desde el punto de vista del criterio de edad, reflejan el comportamiento esperado del ciclo de vida laboral, teniendo las edades de 25-39 y 40-59 como los puntos más altos de participación. La brecha de participación por sexo, ya anotada en el punto 1 de este reporte, se puede apreciar para cada uno de los rangos de edad analizados. El rango de edad donde la participación masculina es mayor que la femenina es el de 40-59 años de edad (26 puntos porcentuales de diferencia). En términos comparativos, para cada uno de los rangos de edad, hay un diferencial de participación de 25 puntos porcentuales.

**Gráfico 3. Tasas de Participación Laboral por Nivel de Edad y Sexo 2018-2020**

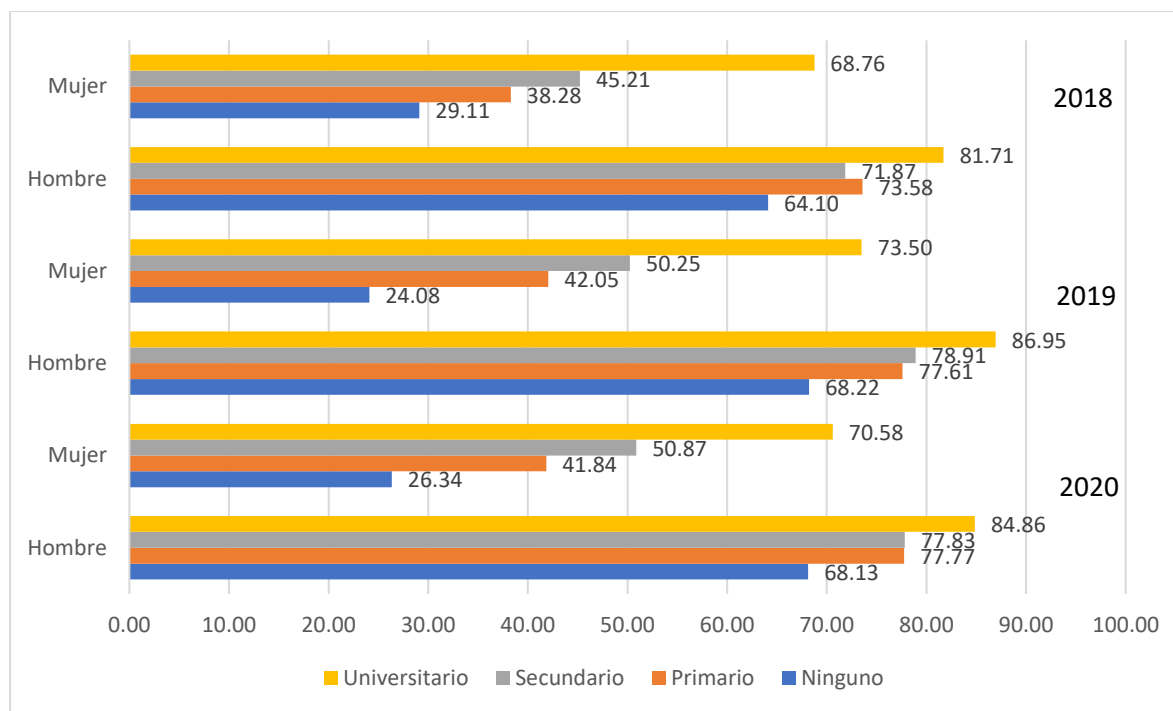


Fuente: ENFT – ECNFT. BCRD. Cálculos de los autores.

Cuando las tasas de participación se analizan con el criterio de nivel educativo y sexo se encuentran dos aspectos radicalmente diferenciados: (i) una enorme brecha de participación entre hombres y mujeres sin ningún nivel educativo, atribuible a la dedicación de las mujeres a la economía del cuidado, y (ii) un hecho evidente de la inversión en capital humano: a un mayor nivel educativo alcanzado, se puede apreciar una mayor participación laboral. Los mayores logros en participación se aprecian para las mujeres con título universitario, las cuales superan los niveles de participación de las mujeres con educación secundaria en 20 puntos porcentuales. En cuanto al efecto “inactividad” detectado como consecuencia de los primeros meses de confinamiento en el año 2020, afectó tanto a hombres como a mujeres en todos los niveles educativos, con excepción de las mujeres con ninguna educación, las cuales incrementaron su participación en 4 puntos porcentuales.



**Gráfico 4. Tasas de Participación por Nivel Educativo y Sexo.**

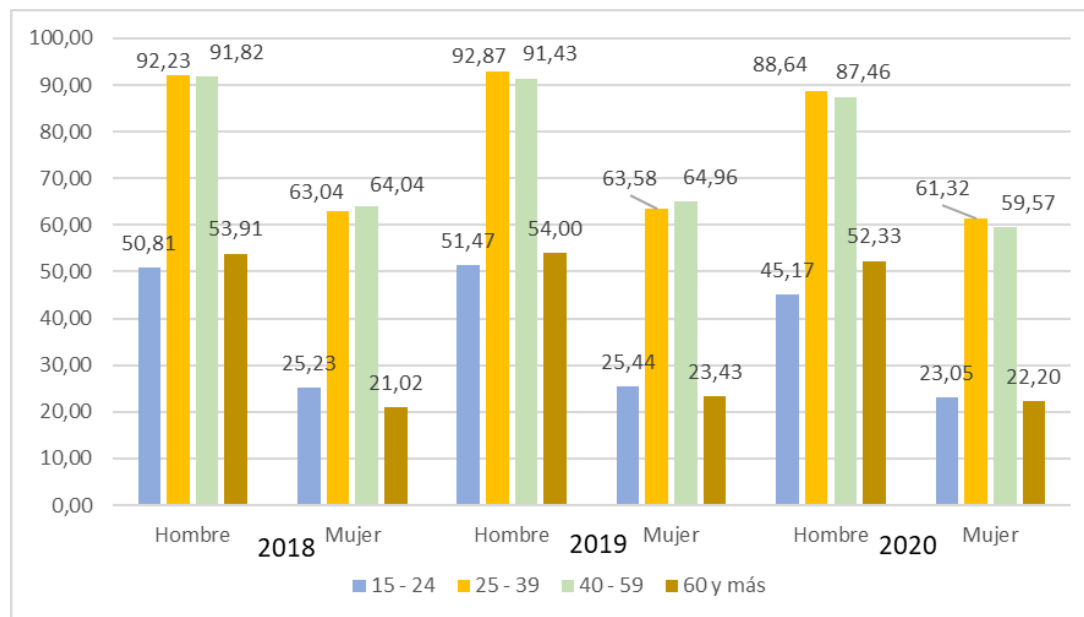


Fuente: ENFT – ECNFT. BCRD. Cálculos de los autores.

#### • Ocupación

El patrón de la ocupación entre niveles de edad es muy semejante al de la participación: El grupo etareo con mayor nivel de ocupación es el de 25-39 años para el periodo analizado. Sin embargo, empieza a apreciarse un patrón diferencial: la brecha detectada en el caso de la participación se incrementa para el caso del desempleo. Esto es particularmente más notorio para la edad premium de ocupación, en la cual la brecha entre hombres y mujeres se incrementa en 4 puntos porcentuales. Cabe anotar que las brechas entre edades, sexo y tasas de participación y ocupación son totalmente comparables, puesto que tienen como nivel común de comparación la Población en Edad de Trabajar, que es el denominador utilizado para calcular ambas tasas.

**Gráfico 5. Tasa de Ocupación 2018-2020**



Fuente: ENFT – ECNFT. BCRD. Cálculos de los autores.

La ocupación revela patrones más específicos de brechas de género al ser analizada por nivel educativo. La tabla 1 muestra las tasas de ocupación por nivel educativo y sexo para los tres periodos analizados. El hecho de que existan niveles crecientes de educación por nivel educativo es reflejo del concepto teórico de inversión en capital humano, y muestra que no parecen existir problemas de desempleo estructural (*mismatch* entre perfiles de formación universitaria y sus requerimientos por la demanda de trabajo), dados los altos niveles de tasa de ocupación en personas con educación universitaria terminada.

**Tabla 1. Tasa de Ocupación por Nivel Educativo y Sexo. 2018-2020**

Nivel Educativo	2018		2019		2020	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Ninguno	67,88	25,34	67,91	23,37	63,74	27,50
Primario	76,08	39,30	75,78	40,24	72,33	36,47
Secundario	74,41	45,04	75,02	44,01	68,74	41,43
Universitario	80,46	63,90	82,89	66,91	78,94	64,48
TO Total	60,71		61,255		57,43	

Fuente: ENFT – ECNFT. BCRD. Cálculos de los autores.

Es interesante apreciar aspectos estructurales y aspectos coyunturales en la ocupación. Por una parte, notorios diferenciales de ocupación entre hombres y mujeres, a los mismos niveles educativos, y dicha brecha se reduce a medida que se incrementa el nivel educativo. Por otra parte, la tasa de ocupación refleja el efecto de la pandemia a través de la caída de los niveles de ocupación para ambos sexos en todos los niveles educativos.

- **Desocupación**

La tasa de desocupación tuvo valores de 5.84% (2018), 5.82% (2019) y 4.49% (2020). ¿Cómo puede explicarse en el primer semestre de 2020, cuando ya tuvieron lugar varios de los múltiples efectos económicos del confinamiento y la pandemia, el desempleo se haya reducido en lugar de incrementarse? Las estadísticas reflejan el primer efecto del confinamiento: la inactividad. La tabla 2 ilustra los principales conceptos poblacionales laborales para el periodo analizado.

**Tabla 2. Conceptos Poblacionales Laborales. 2018 – 2020.**

Periodo	Población total	PET	PEA	PEI	Ocupados	Desocupados
2018	10'300.522	7'563.659	4'876.624	3'612.938	4'591.969	284.656
2019	10'323.754	7'611.455	4'950.904	3'564.108	4'662.421	288.483
2020	10'425.631	7'722.208	4'643.178	4'004.797	4'434.675	208.503
Variación 2019-2020	101.877	110.753	- 307.727	40.689	-227.746	- 79.980

Fuente: ECNFT.

Como puede apreciarse, el panorama identificado para el primer semestre de 2020 (una tasa de desempleo de 4.49%, mientras que la de 2019 fue de 5.82%) refleja una reducción del número de desempleados (79.980) y también del número de ocupados (227.746). Pasaron directamente a la inactividad, con una contribución de 110.753 personas que para el 2020 llegaron a la edad de trabajar y una vez entraron a ser parte de la PET, no lo hicieron participando laboralmente (en la PEA) sino directamente siendo parte de la Población Económicamente Inactiva (se incrementó en 440.689 personas).

## **i.Comportamientos Regionales**

Utilizando como criterio analítico el de macro-regiones y provincias de la división política de República Dominicana, se realizaron los cálculos de indicadores laborales de participación, ocupación y desocupación para los tres años de análisis coyuntural.

**Tabla 3. Indicadores Laborales por Macro-Región. 2018-2020**

TASAS POR REGION	2018		2019		2020	
TGP	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Este	81,64	53,47	82,41	55,44	78,10	48,51
Gran Santo Domingo	76,09	52,84	77,43	54,87	69,62	48,44
Norte o Cibao	80,13	48,92	81,17	49,21	76,85	49,64
Sur	77,72	52,22	76,17	48,54	72,59	41,37
TGP TOTAL	64,47		65,04		60,13	
TO	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Este	77,47	47,72	77,87	49,67	75,57	44,75
Gran Santo Domingo	73,37	48,04	74,40	50,37	67,63	45,95
Norte o Cibao	78,27	44,90	79,22	45,11	74,53	45,59
Sur	74,22	47,04	72,83	43,47	70,42	38,91
TO TOTAL	60,71		61,25		57,43	
TD	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Este	5,11	10,75	5,50	10,41	3,25	7,76
Gran Santo Domingo	3,57	9,09	3,92	8,21	2,87	5,15
Norte o Cibao	2,33	8,22	2,40	8,33	3,02	8,17
Sur	4,50	9,92	4,39	10,44	2,99	5,93
TD TOTAL	5,84		5,82		4,49	

Fuente: ENFT – ECNFT. BCRD. Cálculos de los autores.

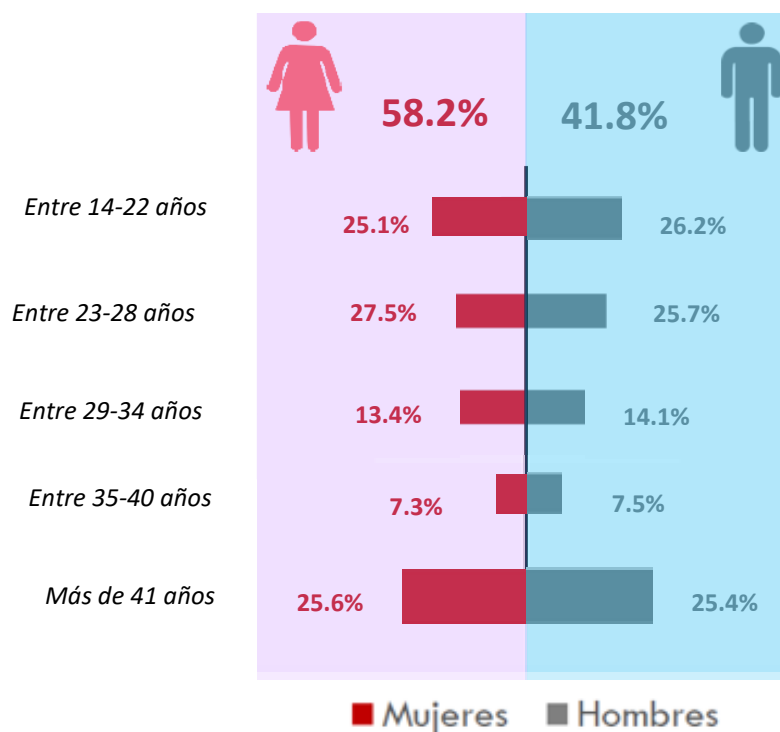
En términos de participación la región con niveles más altos de TGP (tanto para hombres como para mujeres) es la del Este, seguida por Norte / Cibao, Sur y Gran Santo Domingo, en su orden respectivo.

El efecto de caída en participación (inactividad), al que hemos hecho referencia como principal consecuencia laboral de la pandemia, se aprecia de forma más notoria en Gran Santo Domingo (7.8 y 6.4 puntos porcentuales de reducción en la TGP de hombres y mujeres respectivamente). Es particularmente llamativo el efecto diferencial en la región Este, donde la participación femenina, que había sido la más alta entre las regiones analizadas, se redujo en 7 puntos porcentuales, y es el único caso en que la participación femenina se redujo más que la masculina.

## ANEXO 2. CARACTERIZACIÓN DE LA OFERTA LABORAL

En términos del análisis del componente de oferta (las personas que se registran en búsqueda de empleo), se analizó la oferta laboral capturada a partir de los inscritos a través de la plataforma EmpléateYA del Ministerio de Trabajo. Dicho servicio concentra cerca de 515,174<sup>14</sup> inscritos; de los cuales el 58.2% son mujeres y el 41.8% hombres (ver figura 1). Al clasificar la población por rangos de edad, se observa que las personas que cuentan con mayor porcentaje de inscripción son los jóvenes con un 25.6% para los que se encuentran entre los 14 y 22 años, 26.8% para los que se encuentran entre los 23 y 28 y 13.7% para los que se encuentran entre los 29 y 34 años. La figura 1, presenta la distribución de inscritos por rango de edad y sexo, observándose que el 66% de los inscritos tanto para mujeres como para hombres se encuentran entre los 14 y 34 años.

**Figura 5. Distribución de los oferentes registrados por sexo y rangos de edad**



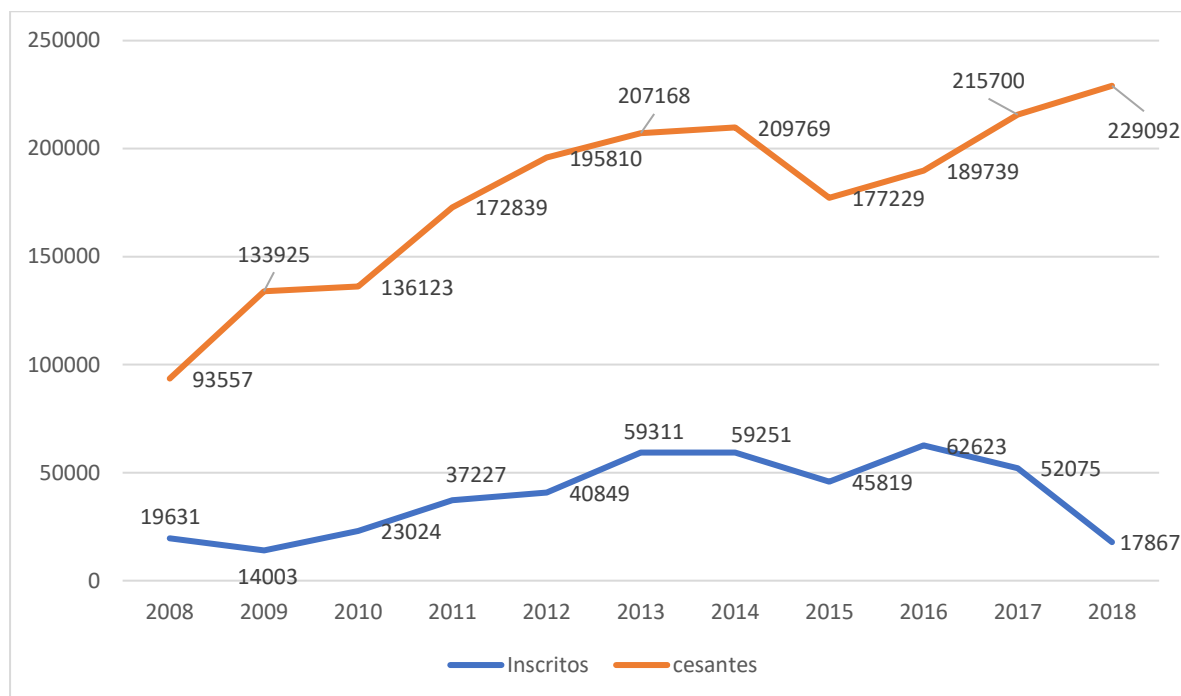
**Nota:** para el 1% de hombres y mujeres inscritos no se cuenta con información de edad registrada. **Fuente:** elaboración de los autores a partir de datos de EmpléateYA.

En términos del periodo temporal de inscripción de los oferentes al sistema, se ha encontrado que en el periodo entre 2008 a 2018 hubo unos picos de registro entre 2013 y 2014 y 2016. Para estos años, se inscribieron alrededor de 60 mil personas únicas. Los meses en donde se presenta mayor registro son los meses de febrero y octubre con un total de inscritos que oscila entre 8% y 10% respectivamente sobre el total de inscritos durante cada año. Dado que esta información podría estar mostrando indicios sobre los periodos en donde mayores tasas de crecimiento de la tasa de desempleo está

<sup>14</sup> Esta cifra es el resultado de un proceso de depuración y homogenización de la información.

enfrentando el país, la contrastamos con el dato de desempleados cesantes obtenido de la Encuesta Continua Nacional de Fuerza de Trabajo para el mismo periodo, y efectivamente, al menos hasta el año 2016 la tendencia de inscritos en EmpléateYa sigue el comportamiento del desempleo cesante.

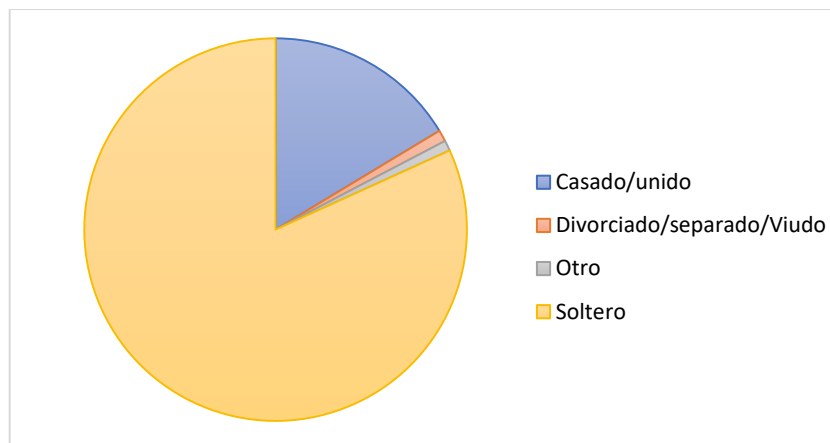
**Gráfico 8. Distribución de los oferentes registrados para los últimos 10 años**



**Fuente:** elaboración de los autores a partir de datos de EmpléateYA.

Como aspectos descriptivos de la población inscrita, en relación con su estado marital, la mayoría de la población inscrita son personas que en su estado civil han reportado estar solteros (68.3%), seguido por casados con un 13.7%. Por sexo, esta proporción se mantiene, para los hombres inscritos el 67.5% reportan estar solteros y el 14.6% están casados o viven en unión libre; y para las mujeres el 69% reporto estar solteras y el 13% casada. Para el caso de los separados o divorciados solo el 1% tanto para hombres como para mujeres se han inscrito al sistema.

**Gráfico 9. Distribución de los oferentes registrados según estado civil**



**Fuente:** elaboración de los autores a partir de datos de EmpléateYA.

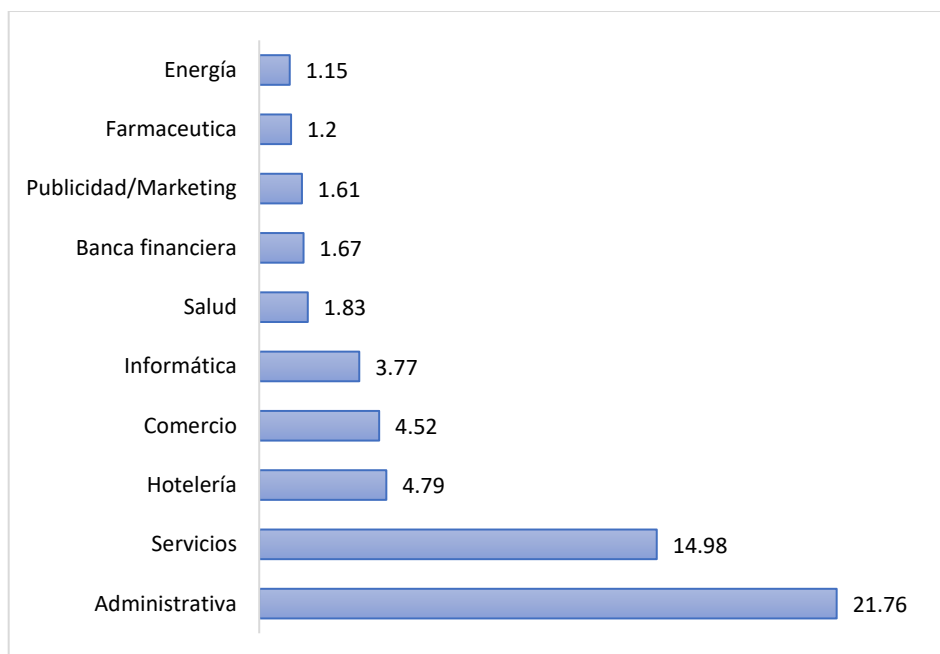
Al examinar si la población inscrita se trata de jefes de hogar, se encuentra que solo el 5.6% lo son. La gran mayoría de los inscritos (99%) se inscribieron a través del portal en los puntos de atención y el 1% (4,673 inscritos) restante se inscribió en alguna de las ferias de empleo. La mayoría de las personas registradas durante las ferias de empleo son jóvenes entre los 14-28 con un 61%, seguido por adultos entre 29 y 34 años con un 18.7% y los adultos de más de 41 años con un 10.2%.

Pasando a aspectos más relacionados con la índole laboral, en relación con las áreas ocupacionales y de conocimiento<sup>15</sup> en donde los inscritos desearían trabajar, el gráfico 10 compila el top 10 de las más frecuentes. Entre estas se encuentra en primer lugar el área administrativa con un 21.8% de los oferentes, seguido por servicios con 15%, hotelería con 4.8%, informática con 3.8%. En general, se observa que las áreas ocupacionales y de conocimiento son bastante transversales a todos los sectores de la economía del país. Al explorar este top de áreas de acuerdo al sexo del inscrito se encuentra que para las áreas administrativa, comercio, farmacia, y salud, la gran mayoría de inscritos son mujeres con participación del 72.2%, 54.4%, 80.4% y 82.3% respectivamente. Para el caso de los hombres, las áreas de conocimiento con mayor participación son energía, agropecuaria, automotriz y construcción con 95.1%, 68.1%, 94.1% y 77.8%.

<sup>15</sup> El formato de inscripción no solicita ni orienta al inscrito a ofrecer información en términos de un sector económico o área ocupacional. Es de suponer que se trata de información orientadora, solicitada con el propósito de facilitar la vinculación laboral del inscrito, y por ende tiene un amplio rango de interpretación.



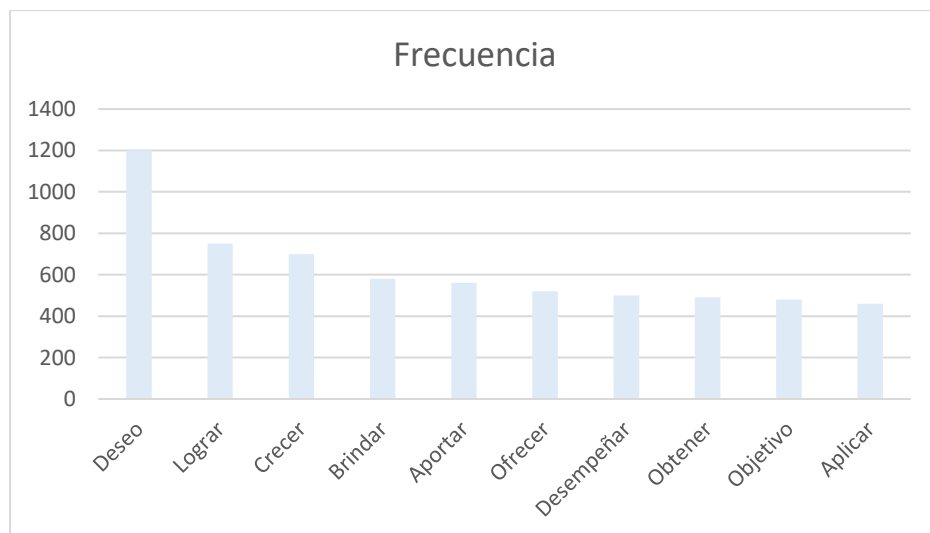
**Gráfico 10. Top 10 áreas de conocimiento en donde los oferentes desean trabajar**



**Fuente:** elaboración de los autores a partir de datos de EmpléateYA.

La compilación de la información del inscrito y las opciones de texto abierto sobre sus aspiraciones e intereses, permiten realizar análisis de conceptos y motivaciones laborales, mediante un ejercicio exploratorio de análisis de texto sobre el objetivo que esperan los oferentes una vez se coloquen.

**Gráfico 11. Opciones de Visualización de Análisis de Texto: Motivación de los oferentes**



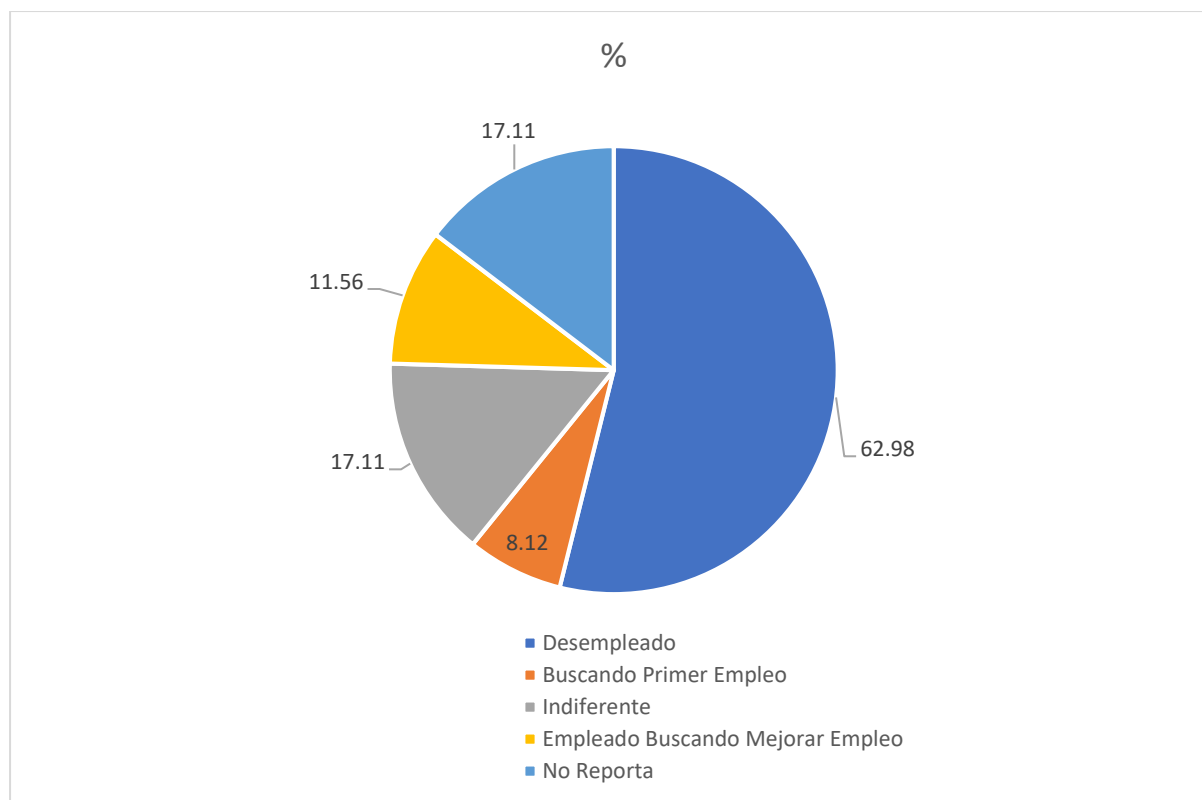


**Fuente:** elaboración de los autores a partir de datos de EmpléateYA.

El gráfico 11 ilustra en dos modalidades los patrones más frecuentes. Dentro de estos, se pueden mencionar ejemplos como: *ofrecer servicios a la Institución con eficiencia en el ambiente profesional para compartir las responsabilidades y éxitos en términos operacionales y funcionales. Alcanzar metas y nuevos retos mediante el trabajo en equipo y la toma de decisiones dentro del límite de tiempo requerido y contribuir con el crecimiento de la Institución de forma profesional. Aportar mis conocimientos para el buen desempeño laboral. Ingresar a una empresa en la que pueda poner en práctica todos mis conocimientos, colaborando con el desarrollo de esta, además que me ofrezca la oportunidad de crecer en el área laboral, personal e intelectual. Algunos otros ejemplos son querer ayudar por ejemplo en el crecimiento de la empresa, brindar conocimientos y experiencia, lograr metas propuestas, aprender nuevas cosas, emprender, buscar la mejora constante, desempeñarse de la mejor manera posible, aportar, servir, etc.*

Al analizar la situación laboral de los oferentes, se encuentra que el 62.9% (324,450 inscritos) se encuentra desempleado, el 11.6% (59,573 inscritos) se encuentra empleado, pero buscando mejorar su situación en otro empleo y el 8.1% (41,819 inscritos) se encuentra buscando su primer empleo (ver figura 5). Del total de inscritos que reportaron estar desempleados, 189,939 son mujeres y 134,511 son hombres. Esta población en su gran mayoría se encuentra entre los 23 y 28 años (97,262 inscritos), seguido por los que se encuentran entre 14 y 22 años (89,059 inscritos). Para el caso de los que se encuentran buscando un mejor empleo, 31,533 corresponden a mujeres y 28,040 a hombres; esta población se concentra en población entre los 23 y 28 años (21,307 inscritos). Finalmente, la población que está buscando su primer empleo se concentra en los jóvenes entre los 14 y 22 años, con cerca de 24,254 inscritos, que en su gran mayoría son mujeres (70%). Finalmente, a nivel regional la provincia en donde los oferentes reportaron estar buscando o desean trabajar, se concentra en Santo Domingo con un 9.8%, seguido por Santiago con 6.3%, La Romana con 2%, Altagracia con 1.3%, San Cristóbal con 1.3% y San Pedro de Macorís con 1%.

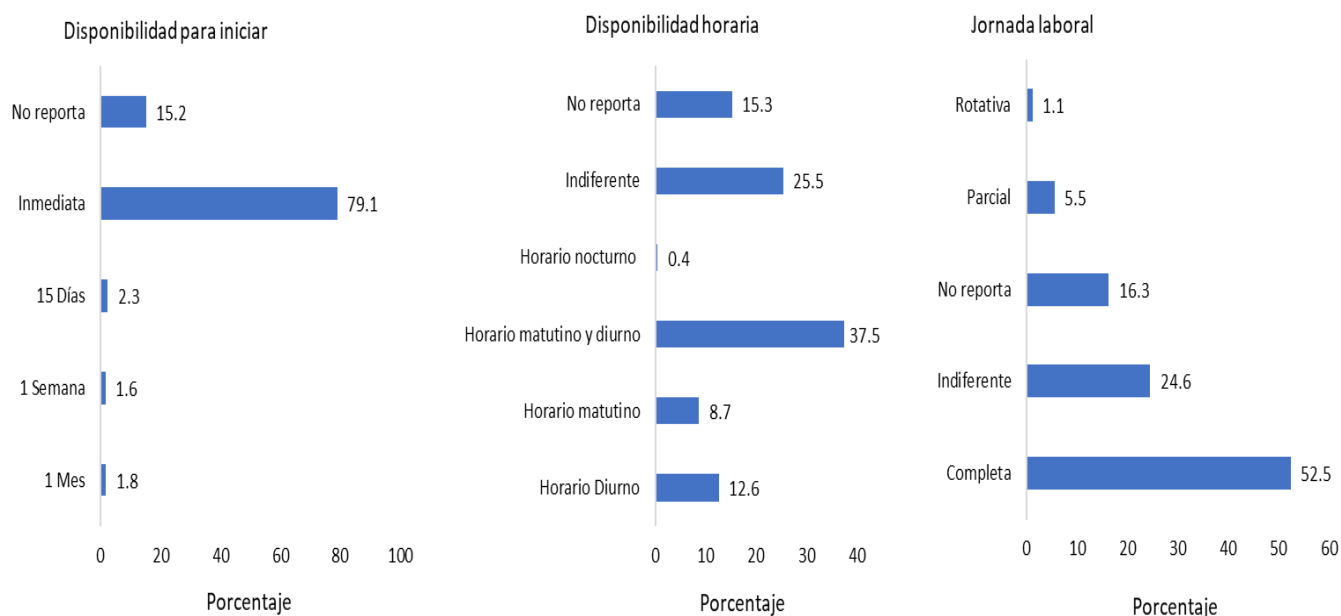
**Gráfico 12. Situación laboral de los oferentes**



**Fuente:** elaboración de los autores a partir de datos de EmpléateYA.

En cuanto a las características de empleabilidad de los oferentes, la figura 6 presenta algunas de las más relevantes. En esta se encuentra la disponibilidad para iniciar labores, disponibilidad horaria y tipo de jornada laboral. En general se observa que el 79.1% de los oferentes se encuentra disponible para iniciar labores en caso de ubicarse en algún empleo. En cuanto a la disponibilidad horaria, el 25.5% es indiferente a el horario propuesto, pero el 37.5% prefiere jornadas laborales en horarios matutinos y/o diurnos. Además, el 52.5% prefiere jornadas laborales completas. Finalmente, y a pesar de que el 85.7% de los oferentes no revela su salario de reserva (Tabla 28), entre aquellos que reportan el salario al que aspiran, la gran mayoría está dispuesta a tener salarios que oscilan entre los RD\$ 10,001.00 a RD\$ 15,000.00 (4.3% inscritos), seguido por salarios entre RD\$ 5,000.00 a RD\$ 10,000.00 (3.5% inscritos) y sólo el 1.3% de los inscritos estaría dispuesto a recibir un salario mínimo.

**Figura 6. Características de empleabilidad de los oferentes**



**Fuente:** elaboración de los autores a partir de datos de EmpléateYA.

**Tabla 28. Perspectiva de salario esperado por los oferentes**

Salario	Oferentes inscritos	Porcentaje
De RD\$ 10,001.00 a RD\$ 15,000.00	22,155	4.30
De RD\$ 15,001.00 a RD\$ 20,000.00	11,841	2.30
De RD\$ 20,001.00 a RD\$ 25,000.00	5,963	1.16
De RD\$ 25,001.00 a RD\$ 35,000.00	4,400	0.85
De RD\$ 35,001.00 a RD\$ 45,000.00	2,321	0.45
De RD\$ 5,000.00 a RD\$ 10,000.00	18,001	3.49
Mas de RD\$ 45,000.00	2,262	0.44
Salario mínimo	6,692	1.30
No reporta	441,539	85.71
Total	515,174	100.00

**Fuente:** Elaboración de los autores a partir de datos de EmpléateYA.

### ANEXO 3. ANÁLISIS EXPLORATORIO DE LA PRIMA SALARIAL PARA LA POBLACIÓN EGRESADA DE INFOTEP QUE SE ENCUENTRA EMPLEADA

Como lo hemos mencionado, el ejercicio de Big Data de vinculación de bases de datos de Egresados, Encuesta a Egresados y SIRLA permitió identificar 352 egresados ocupados en empleo formal. En la información sobre estos 352 egresados recabada de SIRLA y puesta a nuestra disposición por el Ministerio de Trabajo, se incluían los datos de remuneración de 218 egresados (61.9%). Las tres bases de datos ofrecían suficiente información para desarrollar un ejercicio minceriano de determinantes de ingresos. De acuerdo a este proceso metodológico y la información que permite compilar, en esta sección se estudia la importancia de características como recibir apoyo de parte de INFOTEP, el género, la edad, el último nivel educativo aprobado, la región y la modalidad del curso tomado a través de INFOTEP sobre la determinación de los salarios de los egresados que reportan estar empleados o vinculados al mercado laboral.

Para esto, se estima la ecuación de precios, bajo una especificación lineal de tipo Mincer para el egresado que se encuentra empleado  $i$  en la ciudad  $j$  en el periodo  $t$ .

$$\text{Log}W_{ijt} = \beta_{jt}X_{ijt} + \alpha_{jt}Z_{ijt} + \varepsilon_{ijt}$$

Donde  $\log W_{ijt}$  es el salario<sup>16</sup>,  $X_{ijt}$  es un conjunto de características socioeconómicas como la edad, el sexo, escolaridad;  $Z_{ijt}$  son características asociadas a haber recibido apoyo de infotep y la regional en donde tomo los cursos de infotep y  $\varepsilon_{ijt}$  son los factores no observables,  $\beta$ ,  $\alpha$  son parámetros para estimar. Los resultados se pueden observar para el logaritmo del salario o para el valor del salario.

#### Análisis de Resultados

La tabla 28 presenta los resultados obtenidos de la estimación del modelo de prima salarial presentado en la ecuación (1). En esta se puede observar en general que características de capital humano como escolaridad, edad y sexo afectan las primas salariales. En el caso de la edad, el hecho de tener 41 años o más en promedio aumenta un 23.5% (3,983) el salario al compararse con la población que se tiene entre 14 y 22 años. En este aspecto, resultaría plausible que la edad sea una variable proxy a la experiencia laboral, la cual cuenta teóricamente como factor que incrementa el salario, en la teoría de capital humano. En el caso de la escolaridad, el aumento del salario depende de si la persona cuenta con educación secundaria o de posgrado. El salario aumenta en un 26% (5,370) si la persona cuenta con educación secundaria y aumenta un 46.6% (9,548) si cuenta con educación de posgrado en comparación con una persona con solo educación primaria. Por su parte, hay posible evidencia de discriminación salarial de género, dado que en igualdad de circunstancias (edad, nivel educativo), las mujeres reciben un salario 21.4% (4,014) menor que hombres en circunstancias comparables. Por su parte, el hecho de haber recibido apoyo de parte de Infotep reduce el salario en un 10% (1,940). Este aspecto implica elementos cualitativos y cuantitativos de análisis. La ayuda de INFOTEP en la obtención de empleo, dados los análisis realizados, se concentra en individuos con episodios de desempleo, con

<sup>16</sup> El cálculo de los salarios es una imputación del valor medio del rango de salario que reporta el egresado a través de la encuesta.

lo cual el no haberla requerido haría referencia a individuos con un efectivo premium salarial, que no han estado desempleados y por ende no han requerido el apoyo de intermediación de INFOTEP. Asimismo, si el egresado tomó los cursos en el regional sur, se redujo su salario en un 34.8% en comparación de si el egresado tomó los cursos en la regional central. Esto no necesariamente hace referencia a calidad diferencial de la formación entre regionales de INFOTEP, sino al lugar efectivo donde los individuos desarrollan sus actividades laborales luego de recibir la formación. Es decir, individuos que recibieron su formación en la región central laboran en dicha región, donde los salarios son más altos.

**Tabla 26**

**Modelo prima salarial población empleada encuesta egresados**

Variables	Salario	log Salario	
Femenino	-4,041.476***	-0.214***	
	(1,150.426)	(0.058)	
Le ayudo Infotep	-1,940.086*	-0.109*	
	(1,145.269)	(0.058)	
Nivel referencia: Educación primaria			
Educación secundaria	5,370.856***	0.260***	
	(1,871.797)	(0.087)	
Educación superior	1,563.780	0.079	
	(1,186.992)	(0.062)	
Educación posgrado	9,548.407**	0.466***	
	(3,881.353)	(0.173)	
Edad referente: 14 a 22			
Edad 23 a 28	1,560.424	0.083	
	(1,307.801)	(0.073)	
Edad 29 a 34	1,777.318	0.101	
	(1,663.116)	(0.088)	
Edad 35 a 40	2,022.412	0.078	
	(1,821.510)	(0.097)	
Edad 41 o más	3,983.896*	0.235**	
	(2,255.734)	(0.116)	
Región referente: Central			
Regional Este	1,001.732	0.050	
	(1,802.877)	(0.093)	
Regional Norte	974.707	0.035	
	(1,211.105)	(0.061)	
Regional Sur	-5,276.086***	-0.348**	
	(1,533.383)	(0.170)	
Constante	16,977.298***	9.671***	
	(1,557.392)	(0.085)	

Observaciones	218	218	
Pseudo R2	0.148	0.156	
Errores estándar robustos entre paréntesis.			
*** p<0.01, ** p<0.05, * p<0.1			

**Fuente:** elaboración de los autores a partir de Encuesta egresados INFOTEP, 2019



## ANEXO 4. ALGORITMOS UTILIZADOS PARA LA OBTENCIÓN DE BASES WEB SCRAPPING SOFTWARE UTILIZADO PYTHON

### 1. Portal Tu Empleo RD

```
from selenium import webdriver

import threading

import json

import datetime

import time

from time import sleep

from bs4 import BeautifulSoup

import os

import pandas as pd


def get_num_pages(driver):

    driver.get('https://www.tuempleord.do/page/{}/'.format(2))

    sleep(1)

    num_pages = driver.find_element_by_css_selector('body > div.contenedor > div > div.col-md-8.lista-de-trabajos > div.paginacion > a:nth-child(7)').get_attribute('innerText')

    num_pages = int(num_pages.replace('.', ''))

    print(f"There are {num_pages} pages to read")

    return(num_pages)


def get_urls_empleos(driver , num):

    driver.get('https://www.tuempleord.do/page/{}/'.format(num))

    sleep(1)

    todo_el_html = driver.find_element_by_css_selector('body').get_attribute('innerHTML')

    soup = BeautifulSoup(todo_el_html , 'lxml')

    tags_empleos = soup.findAll('h2' , {'class':'resumido'})


    lista_url_empleos = []

    for i in tags_empleos:

        for j in i:

            lista_url_empleos.append(j['href'])

    return(lista_url_empleos)


def procesar_url_empleo(driver , url_empleo):
```

```
driver.get(url_empleo)
todo_el_html = driver.find_element_by_css_selector('body').get_attribute('innerHTML')
soup = BeautifulSoup(todo_el_html , 'lxml')
titulo = soup.find('h2' , {'class':'titulo-individual'}).getText()
fecha = soup.find('span' , {'class':'date'}).getText()
provincia = soup.find('a' , {'class':'pequeno provincia'}).getText()
categoria = soup.find('a' , {'class':'pequeno categoria'}).getText()
contenido = soup.findAll('div' , {'class':'contenido'})

full_content = []
for i in contenido:
    for j in i:
        full_content.append(j)
contenido_interesante = (full_content[:3])

datos = {
    'fecha':fecha,
    'provincia':provincia,
    'categoria':categoria,
    'titulo':titulo,
    'full_content':str(contenido_interesante),
    'hora_subido':str(datetime.datetime.now())
}
return(datos)
```

```
def procesar_bloque(df_resultados, pag_inicial , pag_final):
    options = webdriver.ChromeOptions()
    options.add_argument('--disable-gpu')
    options.add_argument('--headless')
    options.add_argument('--no-sandbox')
    options.add_argument('--disable-dev-shm-usage')
    driver = webdriver.Chrome(options = options)
    for i in range(pag_inicial, pag_final):
        lista = get_urls_empleos(driver , i)
        for k,j in enumerate(lista):
```

```
print("processing ... {}".format(j))
datos = procesar_url_empleo(driver , j)
fecha = datetime.datetime.now()
anio = str(fecha.year)
mes = str(fecha.month)
dia = str(fecha.day)
serial = (((i-1) * 25) + k)
serial = "{0:05d}".format(serial)
nombre = f"tuempleord/vacantes/tuempleord_{anio}-{mes}-{dia}_pag{i}-emp{k+1}.json"
```

```
if not os.path.exists('tuempleord/vacantes'):
    os.makedirs('tuempleord/vacantes')
```

```
with open(nombre , 'w') as json_file:
    json.dump(datos , json_file)
    sleep(1)
```

```
global df_ofertas
df_ofertas = df_resultados.append(datos, ignore_index=True)
```

```
driver.quit()
```

```
def cuts(num_pages, num_chunks):
    amplitud = num_pages//num_chunks + 1
    print(f'amplitude: {amplitud}')
    chunks = [ x for x in range(0,num_pages+1) if x == 0 or x % amplitud == 0 or x == num_pages]
    return chunks
```

```
options = webdriver.ChromeOptions()
options.add_argument('--disable-gpu')
options.add_argument('--headless')
options.add_argument('--no-sandbox')
options.add_argument('--disable-dev-shm-usage')
driver = webdriver.Chrome(options=options)
```

```
num_pages = get_num_pages(driver)
driver.quit()

num_chunks = 2 # TODO: A calcular de la máquina
data_cuts = cuts(num_pages, num_chunks)
df_ofertas = pd.DataFrame(columns=['fecha', 'provincia', 'categoria', 'titulo', 'full_content', 'hora_subido'])

thread_list = list()

for x, cut in enumerate(data_cuts):
    if (x < len(data_cuts) - 1 ):
        print('Job # ', x + 1 , [data_cuts[x] + 1 , data_cuts[x+1]])
        t = threading.Thread(name='Test {}'.format(x), target=procesar_bloque, args=(df_ofertas, data_cuts[x] + 1 ,
data_cuts[x + 1], ) )
        thread_list.append(t)
        t.start()
        print(t.name + ' started!')

for x, cut in enumerate(data_cuts):
    if (x < len(data_cuts) - 1 ):
        print('Job # ', x + 1 , [data_cuts[x] + 1 , data_cuts[x+1]])

# Wait for all threads to complete
for thread in thread_list:
    thread.join()

print('Data retrieval completed!')

# Json to CSV

from os import listdir
from os.path import isfile, join
import pandas as pd
import json
from pandas.io.json import json_normalize
```

```
mypath= 'vacantes'
onlyfiles = [f for f in listdir(mypath) if isfile(join(mypath, f))]

#lista para almacenar los temp_df
frames=[]

#Abre cada archivo json y lo guarda como un temp df
for camino in onlyfiles:
    with open('vacantes/'+camino) as f:
        data = json.load(f)
        temp_df = json_normalize(data)
        frames.append(temp_df)

#concatena todos los temp_df en un solo df
df=pd.concat(frames)

#Guarda el df como un archivo csv con codificación UTF-8
df.to_csv('rd_tuempleord.csv', index = False,encoding='UTF-8')
```

## **2. Portal Computrabajo**

```
import base64

from selenium.webdriver.chrome.options import Options
import datetime
from selenium import webdriver
import json
import requests as r
from time import sleep
from bs4 import BeautifulSoup

def get_ofertas(d

    river, paginador):
    driver.get('https://www.computrabajo.com.do/ofertas-de-trabajo/?p={}'.format(paginador))
    body_html = driver.find_element_by_xpath('/html/body/div[1]/section').get_attribute('innerHTML')
```

```
soup = BeautifulSoup(body_html , 'lxml')
elementos = soup.findAll('a' , {'class':'js-o-link'})
links = []
for i in elementos:
    link = 'https://www.computrabajo.com.do/' + i['href']
    links.append(link)
return(links)
```

```
def get_info_url(url):
    req = r.get(url).content
    soup = BeautifulSoup(req , 'lxml')
    caja = soup.findAll('ul' , {'class':'p0 m0'})
    tit = soup.find('h1' , {'class':'m0'})
    desc = soup.find('ul' , {'class':'p0 m0'})
    header = soup.find('p' , {"class":'fc80 mt5'})
    texto = header.getText()
    texto = texto.split('\n')
    emp = soup.find('a' , {'id':'urlverofertas'})
    empresa = emp.getText()
    url_empresa = emp['href']
    url_empresa = 'https://www.computrabajo.com.do/' + url_empresa
    algo =1
    if(len(texto) == 7):
        salario = texto[1]
        region = texto[3]
        ciudad = texto[4]
        hora = texto[5]
    else:
        salario = "no especificado"
        region = texto[2]
        ciudad = texto[3]
        hora = texto[4]

    #salario = salario.strip()
    #region = region.replace(" ", "")
```

```
#ciudad = ciudad.replace(" ", "")
#hora = hora.replace(" ", "")
salario = ("salario" , str(salario).strip())
region = ("region " , str(region).strip())
ciudad = ("ciudad" , str(ciudad).strip())
hora = ("hora" , str(hora).strip())
salario = salario[1]
region = region[1]
ciudad = ciudad[1]
hora = hora[1]

titulo = (tit.getText())
array = []
for i in desc:
    array.append(str(i))
#codifique esta vuelta en base64 para poderlo subir como json |vie ene 22 16:12:01 -05 2021
info = {
    'url_de_la_empresa':url_empresa,
    'empresa':empresa,
    'url_de_la_vacante':url,
    'titulo':titulo,
    'descripcion':array,
    'fecha_agregado':str(datetime.datetime.now()),
    'salario':salario,
    'region':region,
    "ciudad":ciudad,
    'hora':hora
}
return(info)

#def get_info(driver , link):
#driver.get(link)
#sleep(1)
#nombre_box = driver.find_element_by_xpath('/html/body/div[2]/div[2]/article/div[1]/div[1]/h1')
#nombre = nombre_box.get_attribute('innerText')
```

```
#description_box = driver.find_element_by_xpath('/html/body/div[2]/div[3]/article[1]/section[1]/ul/li[2]')
#description = description_box.get_attribute('innerText')

#req1 = driver.find_element_by_xpath('/html/body/div[2]/div[3]/article[1]/section[1]/ul/li[4]')
#idiomas =
driver.find_element_by_xpath('/html/body/div[2]/div[3]/article[1]/section[1]/ul/li[5]').get_attribute('innerText')
#edad =
driver.find_element_by_xpath('/html/body/div[2]/div[3]/article[1]/section[1]/ul/li[6]').get_attribute('innerText')
#disponibilidad_viaje =
driver.find_element_by_xpath('/html/body/div[2]/div[3]/article[1]/section[1]/ul/li[7]').get_attribute('innerText')
#cambio_resistencia =
driver.find_element_by_xpath('/html/body/div[2]/div[3]/article[1]/section[1]/ul/li[8]').get_attribute('innerText')
#educacion = (req1.get_attribute('innerText'))

#info = {
# 'nombre':nombre,
# 'descripcion':description,
# 'educacion':educacion,
# 'idiomas':idiomas,
# 'disponibilidad para viajar':disponibilidad_viaje,
# 'disponibilidad de cambio de residencia':cambio_resistencia
#}
#return(info)

#MAIN |vie ene 22 14:57:42 -05 2021|

options = Options()
options.headless = True
driver = webdriver.Chrome(options = options)
err = 0
i = 1
cont = 0
while(err < 10):
    try:
        ofertas = get_ofertas(driver , i)
        i+=1
```



```
for j in ofertas:
    nombre = 'vacantes/vacante_{0:05d}.json'.format(cont)
    info = get_info_url(j)
    with open(nombre, 'w') as json_file:
        json.dump(info, json_file)
        print("saved")
    cont += 1
    print(cont)
err = 0
except:
    err += 1
    i += 1
    cont += 1
    print("error", err)
```

```
driver.close()
```

### **3. Portal EmpleateYA**

```
from selenium import webdriver
```

```
from selenium.webdriver.chrome.options import Options
```

```
import datetime
```

```
import json
```

```
from time import sleep
```

```
options = Options()
```

```
options.headless = True
```

```
def grab_ultimo(driver):
```

```
    driver.get('http://empleateya.mt.gob.do/#/empleo/buscar-empleo')
```

```
    sleep(1)
```

```
    ultimo_button = driver.find_element_by_xpath('/html/body/app-root/app-buscar-empleo/main/article/section[2]/div[14]/button').click()
```

```
    sleep(0.5)
```

```
    ultima_pagina = driver.find_element_by_xpath('/html/body/app-root/app-buscar-empleo/main/article/section[2]/div[12]/button').get_attribute('innerText')
```

```
    return(int(ultima_pagina))
```

```
def guardar(vacante , num):
    nombre = 'vacantes/vacante_{0:03d}.json'.format(num)
    with open(nombre, 'w') as archivo_json:
        json.dump(vacante , archivo_json)

driver = webdriver.Chrome(options = options)

cont = 1
err = 0
total = 1

ultimo = grab_ultimo(driver)
driver.get('http://empleateya.mt.gob.do/#/empleo/buscar-empleo')

print("hay {} paginas".format(ultimo))
total = 1
for j in range(0,ultimo):
    sleep(1)
    try:
        for i in range(1,7):
            elem = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[1]/article[2]/div/div[2]/div[2]'.format(i))
            nombre_box = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[1]/article[2]/div/div[2]/div[2]/h3'.format(i))
            nombre = nombre_box.get_attribute('innerText')
            try:
                provincia_box = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[1]/article[2]/div/div[2]/div[2]/ul/li[1]'.format(i))
                provincia = provincia_box.get_attribute('innerText')
                provincia = provincia.replace("Provincia:" , "")
            except:
                provincia = "no especificado"
            try:
                jornada_box = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[1]/article[2]/div/div[2]/div[2]/ul/li[2]'.format(i))
```

```
jornada = jornada_box.get_attribute('innerText')
jornada = jornada.replace('Jornada:', '')
if(jornada == ""):
    jornada = "no especificada"
except:
    jornada = "no especificada"
try:
    salario_box = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[1]/article[2]/div/div[2]/ul/li[3]'.format(i))
    salario = salario_box.get_attribute('innerText')
    salario = salario.replace("Salario:", "")
except:
    salario = 'no especificado'
print(nombre , provincia , jornada , salario)
elem.click()

caja = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[1]/article[2]/div/div[8]')
caja_html = caja.get_attribute('innerText')
empleo = {
    'nombre_empleo':nombre,
    'provincia':provincia,
    'jornada':jornada,
    'salario':salario,
    'descripcion_completa':caja_html,
    'hora_scrapeada':str(datetime.datetime.now())
}
guardar(empleo , total)
afuera = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[1]/article[2]/div/div[8]/section[1]/span')
afuera.click()
print(total , i)
total += 1
sleep(0.3)
```

```
next_button = driver.find_element_by_xpath('/html/body/app-root/app-buscar-
empleo/main/article/section[2]/div[13]/button')
next_button.click()
sleep(0.5)
except:
    print("no hay mas empleo")
    break

print('done')
driver.close()
```

---

#### ALGORITMO USADO PARA IMPUTAR LA CLASIFICACIÓN DE OCUPACIONES

```
install.packages("NPL")
install.packages("tm")
install.packages("wordcloud")
install.packages("SnowballC")
install.packages("grep")
install.packages("grepl")
install.packages("regexr")
install.packages("ggplot2")
install.packages("RColorBrewer")
install.packages("wordcloud2")
install.packages("NLP")
install.packages("gtools")

install.packages("labourR")
install.packages("magrittr")
install.packages("textcat")
install.packages(c("readxl", "doParallel", "stringdist", "phonics", "dplyr", "plyr", "tidyr", "data.table"))
install.packages("lubridate")

library(tm)
library(RColorBrewer)
library(wordcloud)
```

```
library(ggplot2)
library(NLP)
library(wordcloud2)
library(readr)
library(gtools)
library(dplyr)
library(labourR)
library(data.table)
library(magrittr)
library(textcat)
library("readxl", lib.loc=~R/win-library/3.4")
library("doParallel")
library("stringdist")
library("phonics")
library("dplyr")
library("plyr")
library("tidyr")
library("data.table")
library("stringr")
library(lubridate)

#seteamos el directorio de trabajo
setwd("C:/Users/Administrador/Dropbox/ScrapperOIT")

#Funcion de limpieza

corrector <- function(base0){
  limpiador <- function(x){ iconv(x, "UTF-8", "latin1") %>% iconv(., to="ASCII//TRANSLIT") %>% tolower() %>%
gsub("(?:#|@)[a-zA-Z0-9]+ ?", "", .) %>% gsub("[[:punct:]]", "", .) %>% gsub("^ (.+)", "\\1", .) %>% gsub("(.) $", "\\1",
.) %>% gsub("(.) $", "\\1", .) %>% gsub("(.) $", "\\1", .)
}
  return(base0[, c(lapply(.SD, limpiador))])
}

saca_departamentos <- function(lista_ent){
```

```
lista_entaux <- data.table(lista_ent[, "id"], lista_ent[, tstrsplit(computrabajo$ciudad, ",")]) %>% corrector(.) %>%  
return(lista_entaux)  
}
```

```
abrestop <- function(vec){  
  vec <- str_split(vec, " ")[[1]]  
  vec <- vec[!(vec %in% stopwords("spanish"))]  
  return(vec)  
}
```

#### Empleo

```
#Cargue y depuracion de las bases de vacantes para Republica Dominicana  
tuempleord <- read_csv("csv/rd_tuempleord.csv")  
View(tuempleord)
```

```
tuempleord$cat
```

```
#Eliminacion duplicidad y limpieza de caracteres
```

```
tuempleord <- data.table(tuempleord)  
tuempleordv2 <- unique(tuempleord, by=c("titulo", "full_content", "categoria", "provincia")) %>%  
corrector(.)
```

```
names(tuempleordv2)[4] <- "Titulo_oferta"  
names(tuempleordv2)[5] <- "Descripcion"
```

```
write.csv(tuempleordv2, "tuempleordvf.csv")
```

###

```
#Cargue y depuración de las bases de vacantes EmpleateYA
```

```
library(readxl)
```

```
VacantesEmpeateYA <- read_excel("VacantesEmpeateYA.xlsx")
```

```
View(VacantesEmpeateYA)
```

```
co_empleo <- VacantesEmpeateYA
```

```
co_empleo$id <- paste("empleo", 1:nrow(co_empleo), sep="," )
```

```
#Eliminación de duplicidad y limpieza de caracteres
```

```
names(co_empleo)[2] <- "Titulo_oferta"
```

```
names(co_empleo)[5] <- "Descripcion"
```

```
co_empleov1 <- select(co_empleo, c(id, Titulo_oferta, Descripcion, ACTIVIDAD_ECONOMICA, CANTIDAD, BENEFICIOS_GENERALES))
```

```
co_empleov1 <- data.table(co_empleov1)
```

```
co_empleov2 <- unique(co_empleov1, by=c("Titulo_oferta", "Descripcion", "BENEFICIOS_GENERALES", "ACTIVIDAD_ECONOMICA")) %>%
```

```
corrector(.)
```

```
###
```

```
write.csv(co_empleov2, "empleovf.csv")
```

```
#####Asignación de la Clasificación de ocupaciones
```

```
computrabajov4 <- tuempleordv2
```

```
co_empleov4 <- co_empleov2
```

```
title_empleo <- select(co_empleov4, c(id, Titulo_oferta, Descripcion))
```

```
title_comput <- select(computrabajov4, c(id, Titulo_oferta, Descripcion))
```

```
base_ocu <- rbind(title_empleo, title_comput )
```

```
base_ocu <- unique(base_ocu, by=c("Titulo_oferta", "Descripcion")) %>%  
  corrector(.)
```

```
names(base_ocu)[2] <- "titulos"
```

```
#Asignacion por titulo  
corpus <- data.table(  
  id = base_ocu$id,  
  text = base_ocu$titulos  
)
```

```
# Define lengauge -- Cambio respecto original  
langv <- textcat(corpus$text)  
langv[langv=="english" & !is.na(langv)] <- "en"  
langv[langv!="en" | is.na(langv)] <- "es"
```

```
corpus[, language := langv]  
languages <- unique(corpus$language)
```

```
class_ocup <- corpus[, 1:2]  
leaves <- 10  
isco <- 4
```

```
ocupos <- lapply(languages, function(lang) {  
  classify_occupation(  
    corpus = corpus[language == lang],  
    lang = lang,  
    isco_level = isco,  
    num_leaves = leaves  
  )  
}) %>% rbindlist  
prev.names <- names(class_ocup)  
class_ocup <- left_join(class_ocup, ocupos)
```



```
names(class_ocup) <- c(prev.names, paste("ocu_", isco, "_", leaves, sep=""),
                        paste("label_", isco, "_", leaves, sep=""))
write.csv(class_ocup, "ocup.csv")

class_ocupTITULO <- class_ocup

# Descripcion
corpus <- data.table(
  id = base_ocu$id,
  text = base_ocu$Descripcion
)

# Define language -- Cambio respecto original
langv <- textcat(corpus$text)
langv[langv=="english" & !is.na(langv)] <- "en"
langv[langv!="en" | is.na(langv)] <- "es"

corpus[, language := langv]
languages <- unique(corpus$language)

class_ocup <- corpus[, 1:2]
leaves <- 10
isco <- 4

ocup <- lapply(languages, function(lang) {
  classify_occupation(
    corpus = corpus[language == lang],
    lang = lang,
    isco_level = isco,
    num_leaves = leaves
  )
}) %>% rbindlist

desc_ocup <- left_join(class_ocup, ocup)
names(desc_ocup) <- c("id", "text", "isco_desc", "ocup_desc")
```

```
desc_ocup$isco <- as.numeric(desc_ocup$isco)

names(class_ocupTITULO)[3] <- "isco_tit"
names(class_ocupTITULO)[4] <- "ocu_title"

base_clasivf <- merge(class_ocupTITULO, desc_ocup, by="id")
base_clasivf$isco <- NULL
base_clasivf$text.x <- NULL
base_clasivf$text.y <- NULL
write.csv(base_clasivf, "ocupos_titulo&Desc.csv")
```

---

#### ANEXO. ALGORITMO UTILIZADO EN EL ANÁLISIS DE MERCADO LABORAL CON LA ENCUESTA DE HOGARES: SOFTWARE UTILIZADO STATA

\*Indicadores Mercado Laboral

```
use "C:\Users\Dropbox\RepublicaDominicana\Datoos\2020M.dta", clear
*use "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019M.dta", clear
*drop if ANO==2018
*población total
```

```
rename FACTOR_EXPANSION FACTOR_EXPANSION1
gen FACTOR_EXPANSION=FACTOR_EXPANSION/2
```

```
egen pt=sum(FACTOR_EXPANSION)
egen ptarea=sum(FACTOR_EXPANSION), by (SEXO DES_PROVINCIA)
egen ptgenero=sum(FACTOR_EXPANSION), by (SEXO GRUPO_EDAD)
egen pteducacion=sum(FACTOR_EXPANSION), by (SEXO GRUPO_EDUCACION)
egen ptregion=sum(FACTOR_EXPANSION), by (SEXO GRUPO_REGION)
```

```
*GRUPO_OCUPACION GRUPO_OCUPACION GRUPO_OCUPACION GRUPO_OCUPACION GRUPO_OCUPACION
GRUPO_OCUPACION_PET10 GRUPO_OCUPACION GRUPO_SECTOR GRUPO_OCUPACION
```

\*poblacion en la pet para colombia esta definida en area rural 10 años y en area urbana 12, para efectos de analisis de la encuesta se utiliza mayor a 12

\*p6040: cuantos años cumplidos tiene a la fecha

```
egen pet=sum(FACTOR_EXPANSION) if PET==1
egen petarea=sum(FACTOR_EXPANSION) if PET==1, by (SEXO DES_PROVINCIA)
egen petgenero=sum(FACTOR_EXPANSION) if PET==1, by (SEXO GRUPO_EDAD)
egen peteducacion=sum(FACTOR_EXPANSION) if PET==1, by (SEXO GRUPO_EDUCACION)
egen petregion=sum(FACTOR_EXPANSION) if PET==1, by (SEXO GRUPO_REGION)
```

\*poblacion economicamente activa

```
egen pea=sum(FACTOR_EXPANSION) if PEA==1
egen peaarea=sum(FACTOR_EXPANSION) if PEA==1, by(SEXO DES_PROVINCIA)
egen peagenero=sum(FACTOR_EXPANSION) if PEA==1, by(SEXO GRUPO_EDAD)
egen peaduracion=sum(FACTOR_EXPANSION) if PEA==1, by(SEXO GRUPO_EDUCACION)
egen pearegion=sum(FACTOR_EXPANSION) if PEA==1, by (SEXO GRUPO_REGION)
```

\*poblacion inactiva

```
egen pei=sum(FACTOR_EXPANSION) if INACTIVO==1
egen peiarea=sum(FACTOR_EXPANSION) if INACTIVO==1, by(SEXO DES_PROVINCIA)
egen peigenero=sum(FACTOR_EXPANSION) if INACTIVO==1, by(SEXO GRUPO_EDAD)
egen peieducacion=sum(FACTOR_EXPANSION) if INACTIVO==1, by(SEXO GRUPO_EDUCACION)
egen peiregion=sum(FACTOR_EXPANSION) if INACTIVO==1, by (SEXO GRUPO_REGION)
```

\*poblacion empleada

```
egen empleados=sum(FACTOR_EXPANSION) if OCUPADO==1
egen empleadosarea=sum(FACTOR_EXPANSION) if OCUPADO==1, by (SEXO DES_PROVINCIA)
egen empleadosgenero=sum(FACTOR_EXPANSION) if OCUPADO==1, by (SEXO GRUPO_EDAD)
egen empleadoseducacion=sum(FACTOR_EXPANSION) if OCUPADO==1, by (SEXO GRUPO_EDUCACION)
egen empleadosregion=sum(FACTOR_EXPANSION) if OCUPADO==1, by (SEXO GRUPO_REGION)
```

\*poblacion desempleada

```
egen desempleados=sum(FACTOR_EXPANSION) if DESOCUPADO==1
egen desempleadosarea=sum(FACTOR_EXPANSION) if DESOCUPADO==1, by (SEXO DES_PROVINCIA)
egen desempleadosgenero=sum(FACTOR_EXPANSION) if DESOCUPADO==1, by (SEXO GRUPO_EDAD)
egen desempleadoseducacion=sum(FACTOR_EXPANSION) if DESOCUPADO==1, by (SEXO GRUPO_EDUCACION)
egen desempleadosregion=sum(FACTOR_EXPANSION) if DESOCUPADO==1, by (SEXO GRUPO_REGION)
```

**\*Cálculo de las tasas\***

%pET: muestra la relacion porcentual entre el numero de personas que componen la poblacion en edad de trabajar, frente a la poblacion total.

gen Tpet=(pet/pt)\*100

gen Tpetarea=(petarea/ptarea)\*100

gen Tpetgenero=(petgenero/ptgenero)\*100

gen Tpeteducacion=(peteducacion/pteducacion)\*100

gen Tpetregion=(petregion/ptregion)\*100

\*Tasa Global de participacion -TGp es uno si esta en la pEA Es la relacion porcentual entre la poblacion economicamente activa y la poblacion en edad de trabajar

gen TGp1=1 if PEA==1

gen TGp= (pea/pet)\*100

gen TGparea= (peaarea/petarea)\*100

gen TGpgenero= (peagenero/pet)\*100

gen TGpeducacion= (peaducacion/peteducacion)\*100

gen TGpreregion= (pearegion/petregion)\*100

**\* Tasa de Ocupacion total**

gen TO1=1 if OCUPADO==1

gen TO=(empleados/pet)\*100

gen TOarea=(empleadosarea/petarea)\*100

gen TOgenero=(empleadosgenero/petgenero)\*100

gen TOeducacion=(empleadoseducacion/peteducacion)\*100

gen TOregion=(empleadosregion/petregion)\*100

**\* Tasa de Desempleo total**

gen TD=(desempleados/pea)\*100

gen TDarea=(desempleadosarea/peaarea)\*100

gen TDgenero=(desempleadosgenero/peagenero)\*100

gen TDeducacion=(desempleadoseducacion/peaducacion)\*100

```
gen TDregion=(desempleadosregion/pearegion)*100
```

```
*keep pt ptarea ptgenero pteducacion ptregion pet petarea petgenero peteducacion petregion pea peaarea peagenero  
peaducion pearegion pei peiarea peigenero peieducacion peiregion empleados empleadosarea empleadosgenero  
empleadoseducacion empleadosregion desempleados desempleadosarea desempleadosgenero desempleadoseducacion  
desempleadosregion Tpet Tpetarea Tpetgenero Tpeteducacion Tpetregion TGp1 TGp TGparea TGpgenero TGpeducacion  
TGpreregion TO1 TO TOarea TOgenero TOeducacion TOregion TD TDarea TDgenero TDeducacion TDregion
```

```
collapse (max) pt pet pea pei empleados desempleados Tpet TGp TO TD
```

```
export excel using "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019nal.xls", firstrow(varlabels)
```

```
collapse (max) ptarea petarea peaarea peiarea empleadosarea desempleadosarea Tpetarea TGparea TOarea TDarea,  
by(SEXO DES_PROVINCIA)
```

```
tostring SEXO, replace
```

```
replace SEXO="HOMBRE" if SEXO=="1"
```

```
replace SEXO="MUJER" if SEXO=="2"
```

```
export excel using "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019prosex.xls", firstrow(varlabels)
```

```
collapse (max) ptgenero petgenero peagenero peigenero empleadosgenero desempleadosgenero Tpetgenero TGpgenero  
TOgenero TDgenero, by(SEXO GRUPO_EDAD)
```

```
tostring SEXO, replace
```

```
replace SEXO="HOMBRE" if SEXO=="1"
```

```
replace SEXO="MUJER" if SEXO=="2"
```

```
export excel using "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019sexedad.xls", firstrow(varlabels)
```

```
collapse (max) pteducacion peteducacion peaducion peieducacion empleadoseducacion desempleadoseducacion  
Tpeteducacion TGpeducacion TOeducacion TDeducacion, by(SEXO GRUPO_EDUCACION)
```

```
tostring SEXO, replace
```

```
replace SEXO="HOMBRE" if SEXO=="1"
```

```
replace SEXO="MUJER" if SEXO=="2"
```

```
export excel using "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019sexedu.xls", firstrow(varlabels)
```

```
collapse (max) ptreion petregion pearegion peiregion empleadosregion desempleadosregion Tpetregion TGregion  
TOregion TDregion, by(SEXO GRUPO_REGION)
```

```
tostring SEXO, replace
```

```
replace SEXO="HOMBRE" if SEXO=="1"
```

```
replace SEXO="MUJER" if SEXO=="2"
```

```
export excel using "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019sexreg.xls", firstrow(varlabels)
```

```
*use "C:\Users\Dropbox\RepublicaDominicana\Datoos\2020M.dta", clear
```

```
use "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019M.dta", clear
```

```
*drop if ANO==2019
```

```
*poblacion total
```

```
rename FACTOR_EXPANSION FACTOR_EXPANSION1
```

```
gen FACTOR_EXPANSION=FACTOR_EXPANSION1/2
```

```
replace INTERNET=0 if INTERNET==2
```

```
collapse (sum) FACTOR_EXPANSION, by(SEXO DES_PROVINCIA INTERNET)
```

```
tostring SEXO, replace
```

```
replace SEXO="HOMBRE" if SEXO=="1"
```

```
replace SEXO="MUJER" if SEXO=="2"
```

```
tostring INTERNET, replace
```

```
replace INTERNET="No tiene" if INTERNET=="0"
```

```
replace INTERNET="Tiene" if INTERNET=="1"
```

```
export excel using "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019interpsex.xls", firstrow(varlabels) replace
```

```
collapse (sum) FACTOR_EXPANSION, by(SEXO GRUPO_EDUCACION INTERNET)
```

```
tostring SEXO, replace
```

```
replace SEXO="HOMBRE" if SEXO=="1"
```

```
replace SEXO="MUJER" if SEXO=="2"
```

```
tostring INTERNET, replace
```

```
replace INTERNET="No tiene" if INTERNET=="0"
```

```
replace INTERNET="Tiene" if INTERNET=="1"
```

```
export excel using "C:\Users\Dropbox\RepublicaDominicana\Datoos\2019interedusex.xls", firstrow(varlabels) replace
```

---

#### ANEXO. ALGORITMO UTILIZADO EN EL ANÁLISIS DE LA ENCUESTA DE EGRESADOS DE INFOTEP: SOFTWARE UTILIZADO STATA

\*Base de datos encuesta a egresados

```
import excel "C:\Users\Dropbox\RepublicaDominicana\Datoos\Base Egresados.xlsx", sheet("Hoja1") firstrow
```

```
duplicates drop Nombreencuestado Ultimogradoaprobado Sexo Edad, force
```

```
cd "C:\Users\Dropbox\RepublicaDominicana\Datos"
```

```
use "C:\Users\Dropbox\RepublicaDominicana\Datos\EncuestaEgresadovregresion.dta", clear
```

```
gen trabaja= 0 if Trabajaustedactualmente=="No"
```

```
replace trabaja= 1 if Trabajaustedactualmente=="Si"
```

```
gen gender=0 if Sexo=="Femenino"
```

```
replace gender=1 if Sexo=="Masculino"
```

```
encode Ultimogradoaprobado, gen (educacion)
```

```
gen niveledu=1 if educacion==2 | educacion==3
```

```
replace niveledu=2 if educacion==4 | educacion==5
```

```
replace niveledu=3 if educacion==6 | educacion==7 | educacion==8
```

```
replace niveledu=4 if educacion==1
```

```
encode Edad, gen (edad1)
```

```
encode Regionaldonderealizólaformac, gen (region)
```

```
encode Modalidaddelaformación, gen (modalidad)
```

```
gen trabajamientras= 0 if Trabajabaustedalmomentodere=="No"
```

```
replace trabajamientras= 1 if Trabajabaustedalmomentodere=="Si"
```

```
gen habuscado= 0 if Habuscadotrabajoenlosúltimo=="No"  
replace habuscado= 1 if Habuscadotrabajoenlosúltimo=="Si"
```

```
gen requiereF= 0 if Requiereusteddeotrasformacio=="No"  
replace requiereF= 1 if Requiereusteddeotrasformacio=="Si"
```

```
gen solicitudF= 0 if Hasolicitadolasformaciones=="No"  
replace solicitudF= 1 if Hasolicitadolasformaciones=="Si"
```

```
gen Estudia= 0 if Estudiaactualmente=="No"  
replace Estudia= 1 if Estudiaactualmente=="Si"
```

```
gen Continua= 0 if Continuaenelmismotrabajo=="No"  
replace Continua= 1 if Continuaenelmismotrabajo=="Si"
```

```
gen formal= 0 if Trabajoformal=="No"  
replace formal= 1 if Trabajoformal=="Si"
```

```
gen infotep= 0 if LeayudóelINFOTEPaconseguir=="No"  
replace infotep= 1 if LeayudóelINFOTEPaconseguir=="Si"
```

```
encode Sectoreconómicotrabajoactual, gen (Sectoreconómicotrabajoactual1)  
gen sector="Comercio" if Sectoreconómicotrabajoactual1==3  
replace sector="Servicios" if Sectoreconómicotrabajoactual1==6  
replace sector="Industrial" if Sectoreconómicotrabajoactual1==5  
replace sector="OtrosSectores" if Sectoreconómicotrabajoactual1==2 | Sectoreconómicotrabajoactual1==4 |  
Sectoreconómicotrabajoactual1==7  
encode sector, gen (sector1)
```

```
tab region, gen(region_)  
tab niveledu, gen(niveledu_)
```



```
tab edad1, gen(edad1_)
```

```
tab modalidad, gen(rmodalidad_)
```

```
gen log_ing=log(ingreso)
```

\*Regresion determinantes de empleabilidad

```
*reg trabaja infotep gender i.niveledu i.edad1 i.modalidad Estudia habuscado solicitudF i.region
```

```
probit trabaja infotep gender Estudia habuscado solicitudF region_2 region_3 region_4 niveledu_2 niveledu_3 niveledu_4  
edad1_2 edad1_3 edad1_4 edad1_5 rmodalidad_2 rmodalidad_3 rmodalidad_4 rmodalidad_5, vce(robust)
```

```
mfex
```

```
outreg2 using probit, mfx ctitle(mfx) se append excel auto(2) bdec(3) tdec(3) lab addstat(Pseudo R2, e(r2_p))
```

```
summarize trabaja infotep i.gender i.niveledu i.edad1 i.modalidad Estudia habuscado solicitudF i.region
```

```
probit trabaja infotep gender i.niveledu i.edad1 i.modalidad Estudia habuscado solicitudF i.region, vce(robust)
```

```
mfex
```

```
outreg2 using probit, mfx ctitle(mfx) se append excel auto(2) bdec(3) tdec(3) lab addstat(Pseudo R2, e(r2_p))
```

\*Regresión prima salarial tipo mincer

```
keep if trabaja==1
```

```
reg log_ing b1.gender i.infotep i.niveledu i.edad1 i.modalidad , vce(robust)
```

```
outreg2 using POOL, se excel auto(2) bdec(3) tdec(3) lab
```

```
reg ingreso b1.gender i.infotep i.niveledu i.edad1 i.modalidad , vce(robust)
```

```
outreg2 using POOL1, se excel auto(2) bdec(3) tdec(3) lab
```

---

## ANEXO. ALGORITMO UTILIZADO PARA LA CONSTRUCCIÓN DE NUBES DE PALABRAS SOFTWARE UTILIZADO R Y RSTUDIO

```
install.packages("tm")
install.packages("wordcloud")
install.packages("SnowballC")
install.packages("grep")
install.packages("grepl")
install.packages("regexr")
```

---

### INSTALL.PACKAGES("GGPLOT2")

```
install.packages("RColorBrewer")
```

```
library(tm)
library(RColorBrewer)
library(wordcloud)
library(ggplot2)
library(NLP)
```

```
setwd("C:\\Users\\Administrador\\Dropbox\\ScrapperOIT\\NubesdePalabras\\Wordcloud")
```

```
txt<-readLines("Descripcion2020.csv",encoding="UTF-8")
```

```
#configurar texto
```

```
txt=objeto
```

```
txt=iconv(txt, to="ASCII//TRANSLIT")
```

```
txt
```

```
#Construye Corpus (crear una copia para manipular el texto)
```

```
corpus <- Corpus(VectorSource(txt))
```

```
#Pasar a minuscula
```

```
d <-tm_map(corpus, tolower)

#quitar espacios en blanco
d <-tm_map(d, stripWhitespace)
d

#quitar puntuacion
d <-tm_map(d, removePunctuation)

# quita los números
d <- tm_map(d, removeNumbers)

#Cargar el archivo de palabras que voy a eliminar
sw<- readLines("vacias.txt",encoding="UTF-8")
sw=iconv(sw, to="ASCII//TRANSLIT")

#Remover al archivo d las palabras características del español: preposiciones...
d <-tm_map(d, removeWords, stopwords("spanish"))
d <-tm_map(d, removeWords, stopwords("english"))

#Remover palabras vacías personalizadas
d <- tm_map(d, removeWords, sw)

#Volver el archivo plano (con todas las correcciones)
corpus_clean<-tm_map(d, PlainTextDocument)

corpus2 <- Corpus(VectorSource(corpus_clean))

#Crear la matriz de terminos
tdm <- TermDocumentMatrix(corpus2)

#encontrar las palabras que se repitan más de 5 veces en el archivo depurado
findFreqTerms(tdm, lowfre=30)
```

```
#
tdm = as.matrix(tdm)

tdmm = sort(rowSums(tdm), decreasing = TRUE)

m <- as.matrix(tdm) #lo vuelve una matriz
v <- sort(rowSums(m),decreasing=TRUE) #lo ordena y suma
df <- data.frame(word = names(v),freq=v) # lo nombra y le da formato de data.frame
write.table(df, file = "2020.csv", row.names=FALSE, na="",col.names=T, sep=";")
df<-X2020

#Trazar la frecuencia de palabras
barplot(df[1:10,]$freq, las = 2, names.arg = df[1:10,]$word,
        col = "lightblue", main = "PALABRAS MÁS FRECUENTES", ylab = "Frecuencia de palabras")

wordcloud(names(tdmm),tdmm, min.freq = 8,
          colors=brewer.pal(6,"Dark2"),random.order=FALSE)

install.packages("wordcloud2")
library(wordcloud2)

wordcloud2(df, size=1.2)
wordcloud2(df, size = 0.5, ellipticity = 0.1)
```

#### **ALGORITMO UTILIZADO EN EL PROCESAMIENTO DE LA BASE DE VACANTES: SOFTWARE UTILIZADO STATA**

```
use "C:\Users\Dropbox\ScrapperOIT\baseFinalRD.dta", clear
***Corrección de errores de ortografía de la base, específicamente para arreglar los nombres de las vacantes
*poner en minusculas

foreach var of varlist Titulo_oferta Descripcion actividad_economica beneficios_generales provincia categoria nameInter
{

replace `var' = lower(`var')
```

}

\*reemplazando caracteres raros y tildes

```
foreach var of varlist Titulo_oferta Descripcion actividad_economica beneficios_generales provincia categoria nameInter
{
replace `var' = subinstr(`var', "Ã", "a", .)
replace `var' = subinstr(`var', "Ã³", "o", .)
replace `var' = subinstr(`var', "Ã²", "o", .)
replace `var' = subinstr(`var', "Ã´", "o", .)
replace `var' = subinstr(`var', "Ã³", "o", .)
replace `var' = subinstr(`var', "Ã©", "e", .)
replace `var' = subinstr(`var', "Ã%", "e", .)
replace `var' = subinstr(`var', "Ã", "u", .)
replace `var' = subinstr(`var', "Ã¼", "u", .)
replace `var' = subinstr(`var', "Ãš", "u", .)
replace `var' = subinstr(`var', "Ã±", "ñ", .)
replace `var' = subinstr(`var', "á", "a", .)
replace `var' = subinstr(`var', "é", "e", .)
replace `var' = subinstr(`var', "í", "i", .)
replace `var' = subinstr(`var', "Ã", "i", .)
replace `var' = subinstr(`var', "i", "i", .)
replace `var' = subinstr(`var', "ç", "o", .)
replace `var' = subinstr(`var', "à", "o", .)
replace `var' = subinstr(`var', "Ö", "i", .)
replace `var' = subinstr(`var', "á", "a", .)
replace `var' = subinstr(`var', "é", "e", .)
replace `var' = subinstr(`var', "í", "i", .)
replace `var' = subinstr(`var', "ó", "o", .)
replace `var' = subinstr(`var', "ú", "u", .)
replace `var' = subinstr(`var', "i", "ñ", .)
replace `var' = subinstr(`var', "i-", "i", .)
replace `var' = subinstr(`var', "Á", "a", .)
replace `var' = subinstr(`var', "É", "e", .)
replace `var' = subinstr(`var', "Í", "i", .)
replace `var' = subinstr(`var', "Ó", "o", .)
}
```

```
replace `var' = substr(`var', "Ú", "u", .)
replace `var' = substr(`var', "Ñ", "ñ", .)
replace `var' = substr(`var', "?", "e", .)
replace `var' = substr(`var', "i", "i", .)
replace `var' = substr(`var', "'", "", .)
replace `var' = substr(`var', "¨", "", .)
}

foreach var of varlist Titulo_oferta Descripcion {

replace `var' = substr(`var', " ", " ", .)
}

*****

**para documentar las vacantes nuevas inscritas***
gen año=substr(fecha,-4,.)
gen mes2=substr(fecha,4,3)
gen mes=substr(fecha,4,2)

foreach var of varlist mes mes2{

replace `var' = substr(`var', "/", " ", .)
}

gen mes3=mes
destring mes3, force replace

gen mes4=1 if mes2=="ene"
replace mes4=2 if mes2=="feb"
replace mes4=3 if mes2=="mar"
```

```
replace mes4=4 if mes2=="abr"  
replace mes4=5 if mes2=="may"  
replace mes4=6 if mes2=="jun"  
replace mes4=7 if mes2=="jul"  
replace mes4=8 if mes2=="ago"  
replace mes4=9 if mes2=="sep"  
replace mes4=10 if mes2=="oct"  
replace mes4=11 if mes2=="nov"  
replace mes4=12 if mes2=="dic"
```

```
replace mes4=mes3 if mes4==.
```

```
gen dig1=substr(isco,1,1)  
gen nivel="Alto" if dig1=="0" | dig1=="1" | dig1=="2"  
replace nivel="Medio" if dig1=="3"  
replace nivel="Bajo" if nivel==""
```

```
gen Educacion=Descripcion
```

```
***Educación
```

```
gen primaria=strpos(Educacion, "primar") | strpos(Educacion, "primaria") | strpos(Educacion, "preesco") >0  
gen secundaria=strpos(Educacion, "secun") | strpos(Educacion, "secunda") | strpos(Educacion, "secundaria") >0  
gen bachillerato=strpos(Educacion, "bachil") | strpos(Educacion, "bachillerato") | strpos(Educacion, "bachiller") |  
strpos(Educacion, "media") >0  
gen tecnico=strpos(Educacion, "tecni") | strpos(Educacion, "tecnico") | strpos(Educacion, "auxili") | strpos(Educacion,  
"tecnica") | strpos(Educacion, "tecnca") >0  
gen tecnologo=strpos(Educacion, "tecnologo") | strpos(Educacion, "tecnol") | strpos(Educacion, "tecnologia") |  
strpos(Educacion, "tecnologo") | strpos(Educacion, "tecnologico") | strpos(Educacion, "enfermeria") >0  
gen universitario=strpos(Educacion, "profes") | strpos(Educacion, "universi") | strpos(Educacion, "administra") |  
strpos(Educacion, "agronom") | strpos(Educacion, "bacteriologia") | strpos(Educacion, "archivi") | strpos(Educacion,  
"antropol") | strpos(Educacion, "arquite") | strpos(Educacion, "biblio") | strpos(Educacion, "biolog") | strpos(Educacion,  
"ciencia") | strpos(Educacion, "comerc") | strpos(Educacion, "comunica") | strpos(Educacion, "comunica") |  
strpos(Educacion, "contadur") | strpos(Educacion, "derech") | strpos(Educacion, "dibu") | strpos(Educacion, "econom") |  
strpos(Educacion, "ecolog") | strpos(Educacion, "diseñ") | strpos(Educacion, "fisica") | strpos(Educacion, "fisio") |  
strpos(Educacion, "fonia") | strpos(Educacion, "gerenc") | strpos(Educacion, "geolog") | strpos(Educacion,  
"geogra") | strpos(Educacion, "histor") | strpos(Educacion, "licencia") | strpos(Educacion, "ingenier") | strpos(Educacion,  
"instrumenta") | strpos(Educacion, "lengua") | strpos(Educacion, "metemati") | strpos(Educacion, "medicina") |  
strpos(Educacion, "mercade") | strpos(Educacion, "microbiol") | strpos(Educacion, "negocio") | strpos(Educacion,  
"nutricio") | strpos(Educacion, "optomet") | strpos(Educacion, "odonto") | strpos(Educacion, "sociolog") |  
strpos(Educacion, "psicolog") | strpos(Educacion, "quimic") | strpos(Educacion, "salud") | strpos(Educacion,
```

```
"terapia")|strpos(Educacion, "social")| strpos(Educacion, "universitaria")| strpos(Educacion, "traduccion")|
strpos(Educacion, "traducto")| strpos(Educacion, "terapia")>0

gen especializacion=strpos(Educacion, "especia") | strpos(Educacion, "especializacion") | strpos(Educacion, "diplomad") |
strpos(Educacion, "especializacion") >0

gen maestria=strpos(Educacion, "master") | strpos(Educacion, "maestria") >0

gen doctorado=strpos(Educacion, "doctorad") | strpos(Educacion, "doctorado") >0


gen primaria1="pri" if primaria==1
gen secundaria1="sec" if secundaria==1
gen bachillerato1="bac" if bachillerato==1
gen tecnico1="tec" if tecnico==1
gen tecnologo1="ten" if tecnologo==1
gen universitario1="uni" if universitario==1
gen especializacion1="esp" if especializacion==1
gen maestria1="mae" if maestria==1
gen doctorado1="doc" if doctorado==1


egen educa=concat(primaria1 secundaria1 bachillerato1 tecnico1 tecnologo1 universitario1 especializacion1 maestria1
doctorado1)
gen edu_fin=substr(educa,1,3)


gen educacion_vaca=1 if edu_fin=="pri"
replace educacion_vaca=2 if edu_fin=="sec"
replace educacion_vaca=3 if edu_fin=="bac"
replace educacion_vaca=4 if edu_fin=="tec"
replace educacion_vaca=5 if edu_fin=="ten"
replace educacion_vaca=6 if edu_fin=="uni"
replace educacion_vaca=7 if edu_fin=="esp"
replace educacion_vaca=8 if edu_fin=="mae"
replace educacion_vaca=9 if edu_fin=="doc"
replace educacion_vaca=0 if edu_fin=="


drop primaria secundaria bachillerato tecnico tecnologo universitario especializacion maestria doctorado primaria1
secundaria1 bachillerato1 tecnico1 tecnologo1 universitario1 especializacion1 maestria1 doctorado1 educa edu_fin
rename educacion_vaca educacion_mod

gen edu=0 if educacion_mod==0
```



\*\*\*Profesiones\*\*\*\*\*

gen Titulo\_vacante=Titulo\_oferta

gen	fisic	=	strpos(Titulo_vacante, "fisic")		strpos(Descripcion, "fisic")	>0
gen	quimic	=	strpos(Titulo_vacante, "quimic")		strpos(Descripcion, "quimic")	>0
gen	astrono	=	strpos(Titulo_vacante, "astrono")		strpos(Descripcion, "astrono")	>0
gen	meteor	=	strpos(Titulo_vacante, "meteor")		strpos(Descripcion, "meteor")	>0
gen	geolo	=	strpos(Titulo_vacante, "geolo")		strpos(Descripcion, "geolo")	>0
gen	geofi	=	strpos(Titulo_vacante, "geofi")		strpos(Descripcion, "geofi")	>0
gen	matema	=	strpos(Titulo_vacante, "matema")		strpos(Descripcion, "matema")	>0
gen	actua	=	strpos(Titulo_vacante, "actua")		strpos(Descripcion, "actua")	>0
gen	estadi	=	strpos(Titulo_vacante, "estadi")		strpos(Descripcion, "estadi")	>0
gen	biolo	=	strpos(Titulo_vacante, "biolo")		strpos(Descripcion, "biolo")	>0
gen	botani	=	strpos(Titulo_vacante, "botani")		strpos(Descripcion, "botani")	>0
gen	zool	=	strpos(Titulo_vacante, "zool")		strpos(Descripcion, "zool")	>0
gen	agrono	=	strpos(Titulo_vacante, "agrono")		strpos(Descripcion, "agrono")	>0
gen	silvic	=	strpos(Titulo_vacante, "silvic")		strpos(Descripcion, "silvic")	>0
gen	zootec	=	strpos(Titulo_vacante, "zootec")		strpos(Descripcion, "zootec")	>0
gen	ingen	=	strpos(Titulo_vacante, "ingen")		strpos(Descripcion, "ingen")	>0
gen	arquit	=	strpos(Titulo_vacante, "arquit")		strpos(Descripcion, "arquit")	>0
gen	planif	=	strpos(Titulo_vacante, "planif")		strpos(Descripcion, "planif")	>0
gen	topog	=	strpos(Titulo_vacante, "topog")		strpos(Descripcion, "topog")	>0
gen	diseña	=	strpos(Titulo_vacante, "diseña")		strpos(Descripcion, "diseña")	>0
gen	paisaj	=	strpos(Titulo_vacante, "paisaj")		strpos(Descripcion, "paisaj")	>0
gen	cartog	=	strpos(Titulo_vacante, "cartog")		strpos(Descripcion, "cartog")	>0
gen	arqueol	=	strpos(Titulo_vacante, "arqueol")		strpos(Descripcion, "arqueol")	>0
gen	bacteriol	=	strpos(Titulo_vacante, "bacteriol")		strpos(Descripcion, "bacteriol")	>0
gen	medic	=	strpos(Titulo_vacante, "medic")		strpos(Descripcion, "medic")	>0
gen	enferm	=	strpos(Titulo_vacante, "enferm")		strpos(Descripcion, "enferm")	>0
gen	partera	=	strpos(Titulo_vacante, "partera")		strpos(Descripcion, "partera")	>0
gen	paramed	=	strpos(Titulo_vacante, "paramed")		strpos(Descripcion, "paramed")	>0
gen	instrumenta	=	strpos(Titulo_vacante, "instrumenta")		strpos(Descripcion, "instrumenta")	>0

```

gen veteri =      strpos(Titulo_vacante, "veteri") |      strpos(Descripcion, "veteri")      >0
gen odonto =      strpos(Titulo_vacante, "odonto") |      strpos(Descripcion, "odonto")      >0
gen farmace =      strpos(Titulo_vacante, "farmace") |      strpos(Descripcion, "farmace")      >0
gen fisiotera =      strpos(Titulo_vacante, "fisiotera") |      strpos(Descripcion, "fisiotera")      >0
gen dietis =      strpos(Titulo_vacante, "dietis") |      strpos(Descripcion, "dietis")      >0
gen nutrici =      strpos(Titulo_vacante, "nutrici") |      strpos(Descripcion, "nutrici")      >0
gen fonoaudi =      strpos(Titulo_vacante, "fonoaudi") |      strpos(Descripcion, "fonoaudi")      >0
gen optom =      strpos(Titulo_vacante, "optom") |      strpos(Descripcion, "optom")      >0
gen salud_ocupa =      strpos(Titulo_vacante, "salud")& strpos(Titulo_vacante, "ocupaci") |
strpos(Descripcion, "salud") & strpos(Descripcion, "ocupaci")      >0
gen terapeu =      strpos(Titulo_vacante, "terapeu") |      strpos(Descripcion, "terapeu")      >0
gen ortodon =      strpos(Titulo_vacante, "ortodon") |      strpos(Descripcion, "ortodon")      >0
gen audiolo =      strpos(Titulo_vacante, "audiolo") |      strpos(Descripcion, "audiolo")      >0
gen profeso =      strpos(Titulo_vacante, "profeso") |      strpos(Descripcion, "profeso")      >0
gen profesu =      strpos(Titulo_vacante, "profesi") |      strpos(Descripcion, "profesi")      >0
gen maestra =      strpos(Titulo_vacante, "maestra") |      strpos(Descripcion, "maestra")      >0
gen maestro =      strpos(Titulo_vacante, "maestro") |      strpos(Descripcion, "maestro")      >0
gen administra =      strpos(Titulo_vacante, "administra") |      strpos(Descripcion, "administra")      >0
gen contad =      strpos(Titulo_vacante, "contad") |      strpos(Descripcion, "contad")      >0
gen finan =      strpos(Titulo_vacante, "finan") |      strpos(Descripcion, "finan")      >0
gen gestion =      strpos(Titulo_vacante, "gestion") |      strpos(Descripcion, "gestion")      >0
gen vent =      strpos(Titulo_vacante, "vent") |      strpos(Descripcion, "vent")      >0
gen comerc =      strpos(Titulo_vacante, "comerc") |      strpos(Descripcion, "comerc")      >0
gen publici =      strpos(Titulo_vacante, "publici") |      strpos(Descripcion, "publici")      >0
gen audito =      strpos(Titulo_vacante, "audito") |      strpos(Descripcion, "audito")      >0
gen softw =      strpos(Titulo_vacante, "softw") |      strpos(Descripcion, "softw")      >0
gen multime =      strpos(Titulo_vacante, "multime") |      strpos(Descripcion, "multime")      >0
gen web =      strpos(Titulo_vacante, "web") |      strpos(Descripcion, "web")      >0
gen program =      strpos(Titulo_vacante, "program") |      strpos(Descripcion, "program")      >0
gen licenciad =      strpos(Titulo_vacante, "licenciad") |      strpos(Descripcion, "licenciad")      >0
gen redes =      strpos(Titulo_vacante, "redes") |      strpos(Descripcion, "redes")      >0
gen derecho =      strpos(Titulo_vacante, "derecho") |      strpos(Descripcion, "derecho")      >0
gen aboga =      strpos(Titulo_vacante, "aboga") |      strpos(Descripcion, "aboga")      >0
gen juez =      strpos(Titulo_vacante, "juez") |      strpos(Descripcion, "juez")      >0
gen archivi =      strpos(Titulo_vacante, "archivi") |      strpos(Descripcion, "archivi")      >0

```

```

gen curad = strpos(Titulo_vacante, "curad") | strpos(Descripcion, "curad") >0
gen bibliotec= strpos(Titulo_vacante, "bibliotec") | strpos(Descripcion, "bibliotec") >0
gen documental = strpos(Titulo_vacante, "documental") | strpos(Descripcion, "documental") >0
gen economi= strpos(Titulo_vacante, "economi") | strpos(Descripcion, "economi") >0
gen sociolo = strpos(Titulo_vacante, "sociolo") | strpos(Descripcion, "sociolo") >0
gen antropo = strpos(Titulo_vacante, "antropo") | strpos(Descripcion, "antropo") >0
gen filoso = strpos(Titulo_vacante, "filoso") | strpos(Descripcion, "filoso") >0
gen histor = strpos(Titulo_vacante, "histor") | strpos(Descripcion, "histor") >0
gen psicol = strpos(Titulo_vacante, "psicol") | strpos(Descripcion, "psicol") >0
gen social = strpos(Titulo_vacante, "social") | strpos(Descripcion, "social") >0
gen escritor = strpos(Titulo_vacante, "escritor") | strpos(Descripcion, "escritor") >0
gen periodi = strpos(Titulo_vacante, "periodi") | strpos(Descripcion, "periodi") >0
gen traduct = strpos(Titulo_vacante, "traduct") | strpos(Descripcion, "traduct") >0
gen interpret = strpos(Titulo_vacante, "interpret") | strpos(Descripcion, "interpret") >0
gen linguis = strpos(Titulo_vacante, "linguis") | strpos(Descripcion, "linguis") >0
gen esculto = strpos(Titulo_vacante, "esculto") | strpos(Descripcion, "esculto") >0
gen pintor = strpos(Titulo_vacante, "pintor") | strpos(Descripcion, "pintor") >0
gen composit = strpos(Titulo_vacante, "composit") | strpos(Descripcion, "composit") >0
gen politol = strpos(Titulo_vacante, "politol") | strpos(Descripcion, "politol") >0
gen comunicad = strpos(Titulo_vacante, "comunicad") | strpos(Descripcion, "comunicad") >0
gen docent = strpos(Titulo_vacante, "docent") | strpos(Descripcion, "docent") >0

```

\*\*\*\*\*tablita de profesiones

```

gen profesiones=1 if fisic==1
replace profesiones=2 if quimic==1
replace profesiones=3 if astrono==1
replace profesiones=4 if meteor==1
replace profesiones=5 if geolo==1
replace profesiones=6 if geofi==1
replace profesiones=7 if matema==1
replace profesiones=8 if actua==1
replace profesiones=9 if estadi==1
replace profesiones=10 if biolo==1

```

```
replace profesiones=11 if botani==1
replace profesiones=12 if zool==1
replace profesiones=13 if agrono==1
replace profesiones=14 if silvic==1
replace profesiones=15 if zotec==1
replace profesiones=16 if ingen==1
replace profesiones=17 if arquit==1
replace profesiones=18 if planif ==1
replace profesiones=19 if topog==1
replace profesiones=20 if diseñ==1
replace profesiones=21 if paisaj==1
replace profesiones=22 if cartog==1
replace profesiones=24 if arqueol==1
replace profesiones=25 if bacteriol==1
replace profesiones=26 if medic==1
replace profesiones=27 if enferm==1
replace profesiones=28 if partera==1
replace profesiones=29 if paramed==1
replace profesiones=30 if instrumenta==1
replace profesiones=31 if veteri==1
replace profesiones=32 if odonto==1
replace profesiones=33 if farmace==1
replace profesiones=34 if fisioterapia==1
replace profesiones=35 if dietis==1
replace profesiones=36 if nutrice==1
replace profesiones=37 if fonoaudi==1
replace profesiones=38 if optom==1
replace profesiones=39 if salud_ocupa==1
replace profesiones=40 if terapeu==1
replace profesiones=41 if ortodon==1
replace profesiones=42 if audiolo==1
replace profesiones=43 if profesor==1
replace profesiones=44 if profesista==1
replace profesiones=45 if maestra==1
replace profesiones=46 if maestro==1
```

replace profesiones=47 if administra==1  
replace profesiones=48 if contad==1  
replace profesiones=49 if finan==1  
replace profesiones=50 if gestion==1  
replace profesiones=51 if vent==1  
replace profesiones=52 if comerc==1  
replace profesiones=53 if publici==1  
replace profesiones=54 if audito==1  
replace profesiones=55 if softw==1  
replace profesiones=57 if multime==1  
replace profesiones=58 if web==1  
replace profesiones=59 if program==1  
replace profesiones=60 if licenciad==1  
replace profesiones=61 if redes==1  
replace profesiones=62 if derecho==1  
replace profesiones=63 if aboga==1  
replace profesiones=64 if juez==1  
replace profesiones=65 if archivi==1  
replace profesiones=66 if curad==1  
replace profesiones=67 if bibliotec==1  
replace profesiones=68 if documental==1  
replace profesiones=69 if economi==1  
replace profesiones=70 if sociolo==1  
replace profesiones=71 if antropo==1  
replace profesiones=72 if filoso==1  
replace profesiones=73 if histor==1  
replace profesiones=74 if psicol==1  
replace profesiones=75 if social==1  
replace profesiones=76 if escritor==1  
replace profesiones=77 if periodi==1  
replace profesiones=78 if traduct==1  
replace profesiones=79 if interpret==1  
replace profesiones=80 if linguis==1  
replace profesiones=81 if esculto==1  
replace profesiones=82 if pintor==1

```
replace profesiones=83 if composit==1
replace profesiones=84 if politol==1
replace profesiones=85 if comunicad==1
replace profesiones=86 if docent==1
```

\*\*\*Nota: para sacar las estadísticas tener en cuenta que se debe sumar ciertas profesiones como profesor, maestro, docente son una misma

```
gen Profe="NA"
replace Profe="Publicidad y Mercadeo" if profesiones==52
replace Profe="Publicidad y Mercadeo" if profesiones==51
replace Profe="Sociología" if profesiones==75
replace Profe="Ingeniería de Sistemas" if profesiones==59
replace Profe="Administración" if profesiones==47
replace Profe="NA" if profesiones==44
replace Profe="Gestión empresarial" if profesiones==50
replace Profe="Economía" if profesiones==69
replace Profe="Finanzas" if profesiones==49
replace Profe="Ingeniería" if profesiones==16
replace Profe="Licenciatura" if profesiones==86
replace Profe="Contaduría" if profesiones==54
replace Profe="Ingeniería de Sistemas" if profesiones==61
replace Profe="Sicología" if profesiones==74
replace Profe="Ingeniería de Sistemas" if profesiones==55
replace Profe="Finanzas" if profesiones==8
replace Profe="Contaduría" if profesiones==48
replace Profe="Lenguas\Idiomas" if profesiones==79
replace Profe="Enfermería" if profesiones==27
replace Profe="Medicina" if profesiones==26
replace Profe="Publicidad y Mercadeo" if profesiones==53
replace Profe="Salud ocupacional" if profesiones==39
replace Profe="Periodismo" if profesiones==68
replace Profe="Periodismo" if profesiones==77
replace Profe="Derecho" if profesiones==62
replace Profe="Ingeniería de Sistemas" if profesiones==58
replace Profe="Física" if profesiones==1
```

replace Profe="Derecho" if profesiones==63  
replace Profe="Química" if profesiones==2  
replace Profe="Historia" if profesiones==73  
replace Profe="Arquitectura" if profesiones==17  
replace Profe="Química" if profesiones==33  
replace Profe="Artes" if profesiones==82  
replace Profe="Nutrición" if profesiones==36  
replace Profe="NA" if profesiones==30  
replace Profe="Fisioterapia" if profesiones==40  
replace Profe="Comunicación Social" if profesiones==85  
replace Profe="Diseño" if profesiones==20  
replace Profe="Licenciatura" if profesiones==60  
replace Profe="Licenciatura" if profesiones==46  
replace Profe="Ingeniería civil" if profesiones==19  
replace Profe="Fonoaudiología" if profesiones==42  
replace Profe="Licenciatura" if profesiones==43  
replace Profe="NA" if profesiones==65  
replace Profe="Optometría" if profesiones==38  
replace Profe="Astronomía" if profesiones==3  
replace Profe="NA" if profesiones==76  
replace Profe="Matemáticas" if profesiones==7  
replace Profe="Bacteriología" if profesiones==25  
replace Profe="Odontología" if profesiones==32  
replace Profe="NA" if profesiones==67  
replace Profe="Artes" if profesiones==66  
replace Profe="Veterinaria" if profesiones==31  
replace Profe="Enfermería" if profesiones==29  
replace Profe="Ingeniería de Sistemas" if profesiones==57  
replace Profe="Estadística" if profesiones==9  
replace Profe="NA" if profesiones==22  
replace Profe="Biología" if profesiones==10  
replace Profe="Lingüística" if profesiones==80  
replace Profe="Agronomía" if profesiones==13  
replace Profe="Antropología" if profesiones==71  
replace Profe="Filosofía" if profesiones==72

```
replace Profe="Odontología" if profesiones==41
replace Profe="Licenciatura" if profesiones==45
replace Profe="NA" if profesiones==4
replace Profe="NA" if profesiones==14
replace Profe="NA" if profesiones==21
replace Profe="Fisioterapia" if profesiones==34
replace Profe="Geología" if profesiones==5
replace Profe="Geología" if profesiones==6
replace Profe="Veterinaria" if profesiones==12
replace Profe="Derecho" if profesiones==64
replace Profe="Sociología" if profesiones==70
replace Profe="Lenguas\Idiomas" if profesiones==78
replace Profe="Ciencias Políticas" if profesiones==84
```

\*\*\*\*\*

\*Experiencia

\*primero hacer un arreglo con largo de 50 caracteres

```
gen bus_exp=strpos(Descripcion, "experiencia")
```

```
gen ext_exp1=substr(Descripcion, bus_exp-15,50) if bus_exp>0
```

```
gen Experiencia_meses_cc=""
```

```
replace Experiencia_meses_cc="0" if regexm(ext_exp1, "sin")>0
```

```
replace Experiencia_meses_cc="0" if regexm(ext_exp1, "no requiere")>0
```

```
replace Experiencia_meses_cc="1" if (regexm(ext_exp1, "1 mes") | regexm(ext_exp1, "un mes"))>0
```

```
replace Experiencia_meses_cc="2" if (regexm(ext_exp1, "2 mes") | regexm(ext_exp1, "dos mes"))>0
```

```
replace Experiencia_meses_cc="3" if (regexm(ext_exp1, "3 mes") | regexm(ext_exp1, "tres mes"))>0
```

```
replace Experiencia_meses_cc="4" if (regexm(ext_exp1, "4 mes") | regexm(ext_exp1, "cuatro mes"))>0
```

```
replace Experiencia_meses_cc="5" if (regexm(ext_exp1, "5 mes") | regexm(ext_exp1, "cinco mes"))>0
```

```
replace Experiencia_meses_cc="6" if (regexm(ext_exp1, "6 mes") | regexm(ext_exp1, "seis mes"))>0
```

```
replace Experiencia_meses_cc="7" if (regexm(ext_exp1, "7 mes") | regexm(ext_exp1, "siete mes"))>0
```

```
replace Experiencia_meses_cc="8" if (regexm(ext_exp1, "8 mes") | regexm(ext_exp1, "ocho mes"))>0
```

```
replace Experiencia_meses_cc="9" if (regexm(ext_exp1, "9 mes") | regexm(ext_exp1, "nueve mes"))>0
```

```
replace Experiencia_meses_cc="10" if (regexm(ext_exp1, "10 mes") | regexm(ext_exp1, "diez mes"))>0
```



```
replace Experiencia_meses_cc="11" if (regexm(ext_exp1, "11 mes") | regexm(ext_exp1, "once mes"))>0
replace Experiencia_meses_cc="12" if (regexm(ext_exp1, "12 mes") | regexm(ext_exp1, "doce mes"))>0
replace Experiencia_meses_cc="12" if (regexm(ext_exp1, "1 año") | regexm(ext_exp1, "un año"))>0
replace Experiencia_meses_cc="24" if (regexm(ext_exp1, "2 año") | regexm(ext_exp1, "dos año"))>0
replace Experiencia_meses_cc="36" if (regexm(ext_exp1, "3 año") | regexm(ext_exp1, "tres año"))>0
replace Experiencia_meses_cc="48" if (regexm(ext_exp1, "4 año") | regexm(ext_exp1, "cuatro año"))>0
replace Experiencia_meses_cc="60" if (regexm(ext_exp1, "5 año") | regexm(ext_exp1, "cinco año"))>0
replace Experiencia_meses_cc="72" if (regexm(ext_exp1, "6 año") | regexm(ext_exp1, "seis año"))>0
replace Experiencia_meses_cc="12" if (regexm(ext_exp1, "1 ano") | regexm(ext_exp1, "un ano"))>0
replace Experiencia_meses_cc="24" if (regexm(ext_exp1, "2 ano") | regexm(ext_exp1, "dos ano"))>0
replace Experiencia_meses_cc="36" if (regexm(ext_exp1, "3 ano") | regexm(ext_exp1, "tres ano"))>0
replace Experiencia_meses_cc="48" if (regexm(ext_exp1, "4 ano") | regexm(ext_exp1, "cuatro ano"))>0
replace Experiencia_meses_cc="60" if (regexm(ext_exp1, "5 ano") | regexm(ext_exp1, "cinco ano"))>0
replace Experiencia_meses_cc="72" if (regexm(ext_exp1, "6 ano") | regexm(ext_exp1, "seis ano"))>0
```

\*con largo de 100 caracteres

```
gen ext_exp2=substr(Descripcion, bus_exp-15,100) if bus_exp>0
replace Experiencia_meses_cc="0" if regexm(ext_exp2, "sin")>0 & Experiencia_meses_cc=="
replace Experiencia_meses_cc="0" if regexm(ext_exp2, "no requiere")>0 & Experiencia_meses_cc=="
replace Experiencia_meses_cc="1" if (regexm(ext_exp2, "1 mes") | regexm(ext_exp2, "un mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="2" if (regexm(ext_exp2, "2 mes") | regexm(ext_exp2, "dos mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="3" if (regexm(ext_exp2, "3 mes") | regexm(ext_exp2, "tres mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="4" if (regexm(ext_exp2, "4 mes") | regexm(ext_exp2, "cuatro mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="5" if (regexm(ext_exp2, "5 mes") | regexm(ext_exp2, "cinco mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="6" if (regexm(ext_exp2, "6 mes") | regexm(ext_exp2, "seis mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="7" if (regexm(ext_exp2, "7 mes") | regexm(ext_exp2, "siete mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="8" if (regexm(ext_exp2, "8 mes") | regexm(ext_exp2, "ocho mes"))>0 &
Experiencia_meses_cc=="
replace Experiencia_meses_cc="9" if (regexm(ext_exp2, "9 mes") | regexm(ext_exp2, "nueve mes"))>0 &
Experiencia_meses_cc=="
```

```
replace Experiencia_meses_cc="10" if (regexm(ext_exp2, "10 mes") | regexm(ext_exp2, "diez mes"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="11" if (regexm(ext_exp2, "11 mes") | regexm(ext_exp2, "once mes"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="12" if (regexm(ext_exp2, "12 mes") | regexm(ext_exp2, "doce mes"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="12" if (regexm(ext_exp2, "1 año") | regexm(ext_exp2, "un año"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="24" if (regexm(ext_exp2, "2 año") | regexm(ext_exp2, "dos año"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="36" if (regexm(ext_exp2, "3 año") | regexm(ext_exp2, "tres año"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="48" if (regexm(ext_exp2, "4 año") | regexm(ext_exp2, "cuatro año"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="60" if (regexm(ext_exp2, "5 año") | regexm(ext_exp2, "cinco año"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="72" if (regexm(ext_exp2, "6 año") | regexm(ext_exp2, "seis año"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="12" if (regexm(ext_exp2, "1 ano") | regexm(ext_exp2, "un ano"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="24" if (regexm(ext_exp2, "2 ano") | regexm(ext_exp2, "dos ano"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="36" if (regexm(ext_exp2, "3 ano") | regexm(ext_exp2, "tres ano"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="48" if (regexm(ext_exp2, "4 ano") | regexm(ext_exp2, "cuatro ano"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="60" if (regexm(ext_exp2, "5 ano") | regexm(ext_exp2, "cinco ano"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="72" if (regexm(ext_exp2, "6 ano") | regexm(ext_exp2, "seis ano"))>0 &  
Experiencia_meses_cc==""
```

\*3 paso para los que colocan ejemplo: 3 (tres) año

```
replace Experiencia_meses_cc="1" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "un"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="2" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "dos"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="3" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "tres"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="4" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "cuatro"))>0 &  
Experiencia_meses_cc==""
```

```
replace Experiencia_meses_cc="5" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "cinco"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="6" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "seis"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="7" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "siete"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="8" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "ocho"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="9" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "nueve"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="10" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "diez"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="11" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "once"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="12" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "doce"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="1" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "un"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="2" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "dos"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="3" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "tres"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="4" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "cuatro"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="5" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "cinco"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="6" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "seis"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="1" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "un"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="2" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "dos"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="3" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "tres"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="4" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "cuatro"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="5" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "cinco"))>0 &
Experiencia_meses_cc=="

replace Experiencia_meses_cc="6" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "seis"))>0 &
Experiencia_meses_cc=="
```

\*en numeros

```
replace Experiencia_meses_cc="1" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "1"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="2" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "2"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="3" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "3"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="4" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "4"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="5" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "5"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="6" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "6"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="7" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "7"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="8" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "8"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="9" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "9"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="10" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "10"))>0 &
Experiencia_meses_cc=""

replace Experiencia_meses_cc="11" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "11"))>0 &
Experiencia_meses_cc=""

replace Experiencia_meses_cc="12" if (regexm(ext_exp1, "mes") & regexm(ext_exp2, "12"))>0 &
Experiencia_meses_cc=""

replace Experiencia_meses_cc="1" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "1"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="2" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "2"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="3" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "3"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="4" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "4"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="5" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "5"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="6" if (regexm(ext_exp1, "año") & regexm(ext_exp2, "6"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="1" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "1"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="2" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "2"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="3" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "3"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="4" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "4"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="5" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "5"))>0 & Experiencia_meses_cc=""
replace Experiencia_meses_cc="6" if (regexm(ext_exp1, "ano") & regexm(ext_exp2, "6"))>0 & Experiencia_meses_cc=""
```

\*4 paso para los que colocan ejemplo: 3 (tres) año

```
replace Experiencia_meses_cc="1" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "un"))>0 &
Experiencia_meses_cc=""

replace Experiencia_meses_cc="2" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "dos"))>0 &
Experiencia_meses_cc=""

replace Experiencia_meses_cc="3" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "tres"))>0 &
Experiencia_meses_cc=""

replace Experiencia_meses_cc="4" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "cuatro"))>0 &
Experiencia_meses_cc=""
```

replace Experiencia\_meses\_cc="5" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "cinco"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="6" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "seis"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="7" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "siete"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="8" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "ocho"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="9" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "nueve"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="10" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "diez"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="11" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "once"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="12" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "doce"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="1" if (regexm(ext\_exp2, "año") & regexm(ext\_exp2, "un"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="2" if (regexm(ext\_exp2, "año") & regexm(ext\_exp2, "dos"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="3" if (regexm(ext\_exp2, "año") & regexm(ext\_exp2, "tres"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="4" if (regexm(ext\_exp2, "año") & regexm(ext\_exp2, "cuatro"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="5" if (regexm(ext\_exp2, "año") & regexm(ext\_exp2, "cinco"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="6" if (regexm(ext\_exp2, "año") & regexm(ext\_exp2, "seis"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="1" if (regexm(ext\_exp2, "ano") & regexm(ext\_exp2, "un"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="2" if (regexm(ext\_exp2, "ano") & regexm(ext\_exp2, "dos"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="3" if (regexm(ext\_exp2, "ano") & regexm(ext\_exp2, "tres"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="4" if (regexm(ext\_exp2, "ano") & regexm(ext\_exp2, "cuatro"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="5" if (regexm(ext\_exp2, "ano") & regexm(ext\_exp2, "cinco"))>0 &  
Experiencia\_meses\_cc==""

replace Experiencia\_meses\_cc="6" if (regexm(ext\_exp2, "ano") & regexm(ext\_exp2, "seis"))>0 &  
Experiencia\_meses\_cc==""

\*en numeros

replace Experiencia\_meses\_cc="1" if (regexm(ext\_exp2, "mes") & regexm(ext\_exp2, "1"))>0 & Experiencia\_meses\_cc==""

```
replace Experiencia_meses_cc="2" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "2"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="3" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "3"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="4" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "4"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="5" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "5"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="6" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "6"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="7" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "7"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="8" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "8"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="9" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "9"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="10" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "10"))>0 &
Experiencia_meses_cc==""
replace Experiencia_meses_cc="11" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "11"))>0 &
Experiencia_meses_cc==""
replace Experiencia_meses_cc="12" if (regexm(ext_exp2, "mes") & regexm(ext_exp2, "12"))>0 &
Experiencia_meses_cc==""
replace Experiencia_meses_cc="1" if (regexm(ext_exp2, "año") & regexm(ext_exp2, "1"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="2" if (regexm(ext_exp2, "año") & regexm(ext_exp2, "2"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="3" if (regexm(ext_exp2, "año") & regexm(ext_exp2, "3"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="4" if (regexm(ext_exp2, "año") & regexm(ext_exp2, "4"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="5" if (regexm(ext_exp2, "año") & regexm(ext_exp2, "5"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="6" if (regexm(ext_exp2, "año") & regexm(ext_exp2, "6"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="1" if (regexm(ext_exp2, "ano") & regexm(ext_exp2, "1"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="2" if (regexm(ext_exp2, "ano") & regexm(ext_exp2, "2"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="3" if (regexm(ext_exp2, "ano") & regexm(ext_exp2, "3"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="4" if (regexm(ext_exp2, "ano") & regexm(ext_exp2, "4"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="5" if (regexm(ext_exp2, "ano") & regexm(ext_exp2, "5"))>0 & Experiencia_meses_cc==""
replace Experiencia_meses_cc="6" if (regexm(ext_exp2, "ano") & regexm(ext_exp2, "6"))>0 & Experiencia_meses_cc==""

drop bus_exp ext_exp1 ext_exp2

rename Experiencia_meses_cc Experiencia_meses
replace exp_orig=. if exp_orig>144
replace exp_orig=. if exp_orig<0
```

\*\*\*\*\*

**\*\*Grupos CIUO08\*\***

gen ciuo08= isco

tostring isco, replace

gen ciuo08\_2= substr( ciuo08 ,1,2)

gen ciuo08\_1= substr( ciuo08 ,1,1)

**\*\*\*GENERACION DE VAR DE HABILIDADES MÁS DEMANDADAS\*\*\***

**\*\*Habilidades Blandas\*\***

gen A = regexm(Descripcion,"Creatividad") + regexm(Descripcion,"Creativo") + regexm(Descripcion,"creatividad")+  
regexm(Descripcion,"creatividad")

gen B = regexm(Descripcion,"Persuasión") + regexm(Descripcion,"persuasivo")

gen C = regexm(Descripcion,"Colaboración") + regexm(Descripcion,"colaborativo") + regexm(Descripcion,"colaborativa")

gen D = regexm(Descripcion,"Adaptabilidad") + regexm(Descripcion,"adaptación")

gen E = regexm(Descripcion,"Inteligencia Emocional") + regexm(Descripcion,"inteligencia emocional") +  
regexm(Descripcion,"Inteligencia emocional") + regexm(Descripcion,"inteligencia Emocional") +  
regexm(Descripcion,"Inteligencia") + regexm(Descripcion,"inteligencia") + regexm(Descripcion,"Emociones") +  
regexm(Descripcion,"emocional") + regexm(Descripcion,"manejo de emociones")

gen F = regexm(Descripcion,"Liderazgo") + regexm(Descripcion,"liderazgo") + regexm(Descripcion,"Lider") +  
regexm(Descripcion,"lider")

gen G = regexm(Descripcion,"Reciliencia") + regexm(Descripcion,"reciliencia")

gen H = regexm(Descripcion,"Trabajo en equipo") + regexm(Descripcion,"trabajo en equipo") +  
regexm(Descripcion,"equipo") + regexm(Descripcion,"Equipo")

gen I = regexm(Descripcion,"Agilidad") + regexm(Descripcion,"agilidad") + regexm(Descripcion,"Agil") +  
regexm(Descripcion,"agil")

gen J = regexm(Descripcion,"Construccion de Relaciones") + regexm(Descripcion,"construccion de relaciones") +  
regexm(Descripcion,"relaciones") + regexm(Descripcion,"Relaciones") + regexm(Descripcion,"relaciones interpersonales")  
+ regexm(Descripcion,"Relaciones Interpersonales")

gen K = regexm(Descripcion,"Solucion de Problemas") + regexm(Descripcion,"solucion de problemas") +  
regexm(Descripcion,"problemas") + regexm(Descripcion,"Problemas") + regexm(Descripcion,"resolucion de conflictos") +  
regexm(Descripcion,"conflictos") + regexm(Descripcion,"Conflictos")

gen L = regexm(Descripcion,"Organizacion") + regexm(Descripcion,"organizacion") + regexm(Descripcion,"Organizado") +  
regexm(Descripcion,"organizado") + regexm(Descripcion,"Organizada") + regexm(Descripcion,"organizada")

gen M = regexm(Descripcion,"Empatia") + regexm(Descripcion,"empatia") + regexm(Descripcion,"Empatico") +  
regexm(Descripcion,"Empatica")

gen N = regexm(Descripcion,"Comunicacion Eficiente") + regexm(Descripcion,"comunicacion eficiente") +  
regexm(Descripcion,"Comunicacion eficiente") + regexm(Descripcion,"comunicacion") +  
regexm(Descripcion,"Comunicacion") + regexm(Descripcion,"Facilidad de expresion") + regexm(Descripcion,"facilidad de  
expresion") + regexm(Descripcion,"expresion") + regexm(Descripcion,"Expresion")

gen O = regexm(Descripcion,"Autoconocimiento") + regexm(Descripcion,"autoconocimiento")

gen P = regexm(Descripcion,"Gestion de Personas") + regexm(Descripcion,"gestion de personas") +  
regexm(Descripcion,"Gestion de personas") + regexm(Descripcion,"Gestion") + regexm(Descripcion,"gestion") +

regexm(Descripcion,"Manejo del recurso humano") + regexm(Descripcion,"manejo del recurso humano") +  
regexm(Descripcion,"recursos humanos") + regexm(Descripcion,"Recurso humano") + regexm(Descripcion,"Recurso  
Humano") + regexm(Descripcion,"Recursos humanos")

gen Q = regexm(Descripcion,"Flexibilidad Cognitiva") + regexm(Descripcion,"Flexibilidad cognitiva") +  
regexm(Descripcion,"flexibilidad cognitiva")+ regexm(Descripcion,"facilidad de aprender")+  
regexm(Descripcion,"Aprender")

gen R = regexm(Descripcion,"Gestion de Conflictos") + regexm(Descripcion,"Gestion de conflictos") +  
regexm(Descripcion,"gestion de conflictos") + regexm(Descripcion,"conflictos") + regexm(Descripcion,"Resolución de  
conflictos") + regexm(Descripcion,"resolución de conflictos") + regexm(Descripcion,"conflictos") +  
regexm(Descripcion,"Conflictos")

gen S = regexm(Descripcion,"Iniciativa") + regexm(Descripcion,"iniciativa") + regexm(Descripcion,"Capacidad de  
emprender") + regexm(Descripcion,"emprender") + regexm(Descripcion,"Emprender") +  
regexm(Descripcion,"Innovacion") + regexm(Descripcion,"innovacion") + regexm(Descripcion,"innovador") +  
regexm(Descripcion,"Innovador")

gen T = regexm(Descripcion,"Responsabilidad") + regexm(Descripcion,"responsabilidad") +  
regexm(Descripcion,"responsable") + regexm(Descripcion,"Responsable")

gen U = regexm(Descripcion,"Toma de Decisiones") + regexm(Descripcion,"toma de decisiones") +  
regexm(Descripcion,"Toma de decisiones") + regexm(Descripcion,"decisiones") + regexm(Descripcion,"Decisiones")

gen V = regexm(Descripcion,"Pensamiento critico") + regexm(Descripcion,"pensamiento critico") +  
regexm(Descripcion,"Pensamiento Critico")

gen W = regexm(Descripcion,"Orientacion al Servicio") + regexm(Descripcion,"Orientacion al servicio") +  
regexm(Descripcion,"orientacion al servicio") + regexm(Descripcion,"servicio") + regexm(Descripcion,"Servicio")

gen X = regexm(Descripcion,"Negociacion") + regexm(Descripcion,"negociacion") + regexm(Descripcion,"negociaciador")  
+ regexm(Descripcion,"Negociaciador")

gen Y = regexm(Descripcion,"Pensamiento Estrategico") + regexm(Descripcion,"Pensamiento estrategico") +  
regexm(Descripcion,"pensamiento estrategico")

gen Z = regexm(Descripcion,"Networking e influencia") + regexm(Descripcion,"networking e influencia") +  
regexm(Descripcion,"Networking") + regexm(Descripcion,"influencia") + regexm(Descripcion,"Influencia")

#### **\*\* Habilidades Digitales\*\***

gen AA = regexm(Descripcion,"Blockchain") + regexm(Descripcion,"blockchain")

gen BB = regexm(Descripcion,"Cluod Computing") + regexm(Descripcion,"Cluod computing") + regexm(Descripcion,"cluod  
computing") + regexm(Descripcion,"Cluod") + regexm(Descripcion,"cluod") + regexm(Descripcion,"Computing") +  
regexm(Descripcion,"computing")

gen CC = regexm(Descripcion,"Inteligencia artificial IA") + regexm(Descripcion,"inteligencia artificial") +  
regexm(Descripcion,"artificial") + regexm(Descripcion,"Artificial")

gen DD = regexm(Descripcion,"Diseño UX") + regexm(Descripcion,"diseño UX") + regexm(Descripcion,"Diseño") +  
regexm(Descripcion,"diseño")

gen EE = regexm(Descripcion,"Big Data") + regexm(Descripcion,"Bigdata") + regexm(Descripcion,"big data") +  
regexm(Descripcion,"datos") + regexm(Descripcion,"Datos")

gen FF = regexm(Descripcion,"Bases de Datos") + regexm(Descripcion,"Bases de datos") + regexm(Descripcion,"bases de  
datos")

gen GG = regexm(Descripcion,"Razonamiento Analitico") + regexm(Descripcion,"Razonamiento analitico") +  
regexm(Descripcion,"razonamiento analitico") + regexm(Descripcion,"Analitico") + regexm(Descripcion,"analitico")



gen HH = regexm(Descripcion,"Producción de video y audio") + regexm(Descripcion,"produccion de video y audio") + regexm(Descripcion,"Producción de video") + regexm(Descripcion,"Producción de audio") + regexm(Descripcion,"video y audio")

gen II = regexm(Descripcion,"Procesamiento de lenguaje natural") + regexm(Descripcion,"procesamiento de lenguaje natural") + regexm(Descripcion,"Procesamiento") + regexm(Descripcion,"procesamiento") + regexm(Descripcion,"lenguaje natural")

gen JJ = regexm(Descripcion,"Computación científica") + regexm(Descripcion,"computación científica") + regexm(Descripcion,"computación") + regexm(Descripcion,"Computación")

gen KK = regexm(Descripcion,"Ciencia de datos") + regexm(Descripcion,"Ciencia de Datos") + regexm(Descripcion,"ciencia de datos")

gen LL = regexm(Descripcion,"Automatizcion") + regexm(Descripcion,"automatizcion")

gen MM = regexm(Descripcion,"Tecnologias de la Informacion") + regexm(Descripcion,"Tecnologias de la informacion") + regexm(Descripcion,"tecnologias de la Informacion") + regexm(Descripcion,"Tecnologias") + regexm(Descripcion,"tecnologias")

gen NN = regexm(Descripcion,"Analitica de datos") + regexm(Descripcion,"Analitica de Datos") + regexm(Descripcion,"Analitica") + regexm(Descripcion,"analitica")

gen OO = regexm(Descripcion,"Programacion") + regexm(Descripcion,"programacion") + regexm(Descripcion,"programador") + regexm(Descripcion,"desarrollador") + regexm(Descripcion,"Programador") + regexm(Descripcion,"Desarrollador")

gen PP = regexm(Descripcion,"Plataformas Digitales") + regexm(Descripcion,"plataformas digitales") + regexm(Descripcion,"Plataformas digitales") + regexm(Descripcion,"Plataformas") + regexm(Descripcion,"digitales") + regexm(Descripcion,"plataformas")