

“Análisis de la oferta y la demanda laboral durante el año 2023”

por Agustín Rizzo

9/3/2024

CODERHOUSE

Contenido

1. Introducción.....	3
2. Objetivo específico.....	4
3. Hipótesis.....	4
4. Fuente de datos	4
5. Enlace del data set.....	4
6. Alcance del análisis.....	5
7. Herramientas tecnológicas utilizadas.....	5
8. Nivel de aplicación del análisis.....	5
9. Glosario.....	5
10. Descripción del dataset.....	6
11. Diagrama de entidad relación.....	12
12. Tabla de versiones	13
13. Procesamiento de los datos.....	13
13.1 Transformación para la normalización de salarios.....	13
12.2 Transformación para la cuantificación de beneficios totales por empleo.....	20
12.3 Transformación para la cuantificación de beneficios específicos / rubro.....	21
12.4 Transformación para la cuantificación de la experiencia.....	21
13.5 Transformación para la cuantificación de habilidades requeridas por empleo.....	21
13.6 Transformación para la cuantificación de la popularidad de los empleos.....	22
13.7 Transformación para obtener las variables categóricas “Salario SI o Salario NO”.....	24
13.8 Transformación para obtener las variables categóricas “Beneficio SI o Beneficio NO”.....	25
13.9 Transformación para obtener la cantidad de palabras de la columna “Description”.....	25
14. Medidas calculadas	25
15. Modelo de datos en PowerBI.....	28
16. Descripción del reporte.....	29
17. Futuras líneas.....	35

1. Introducción:

El trabajo no solo forma una parte crucial de nuestras vidas, sino que es una pieza clave de la sociedad entera, delineando cada uno de sus aspectos y determinando así su presente, su pasado y su futuro. Entender la dinámica del mercado laboral y conocer las variables que la afectan hoy en día se considera una necesidad principal porque contar con esta información nos brindaría la oportunidad para la toma de decisiones pertinentes, tanto a nivel micro o individual como a escala macro o social. En este escenario, el análisis de datos moderno basado en el uso de herramientas que permiten la recolección, transformación y visualización de conjuntos de datos masivos emerge como una disciplina fundamental, no solo por el volumen de información que nos permite manejar sino también por la capacidad para responder en tiempo real y de forma dinámica a los rápidos cambios que atraviesa el mercado laboral en la actualidad. "Análisis de la oferta y la demanda laboral durante el año 2023" es un proyecto en el que se propone el uso de herramientas del universo del "Data analytics" como SQL y PowerBI para analizar y describir las características del mercado laboral tanto desde el punto de vista de la oferta como de la demanda de empleo. El año 2023 se lo considera un año bisagra, signado por los efectos de la pandemia que lo precedió y que generó grandes consecuencias en muchos aspectos de nuestra vida, incluido el aspecto laboral. Este nuevo escenario posibilitó el surgimiento de un nuevo paradigma del trabajo en el que las tecnologías de la comunicación y la información pasaron al frente, con la creación de nuevos roles tales como los "nómades virtuales" y los "creadores de contenido". Con el fin de adentrarse en esta nueva revolución del empleo, este proyecto se centrará en el análisis de una base de datos de empleos publicados en una plataforma digital que cuenta con una gran cantidad de información relacionada tanto a la oferta como a la demanda de trabajo. Por un lado, se plantea establecer las características principales que ofrecen los empleadores en términos de habilidades



requeridas, nivel de experiencia, beneficios y salarios, mientras que por otro lado se buscará cuantificar el interés de las personas por los distintos empleos a través de indicadores tales como el número de veces que se visitó la oferta laboral y el número de personas que se postuló al empleo.

2. Objetivo específico:

- Establecer cuáles son las características de los empleos más solicitados en la actualidad.
- Establecer cuáles son las características del empleado más buscadas por parte de los empleadores en la actualidad.

3. Hipótesis:

- Los principales factores que determinan el interés de una persona por un determinado empleo son el salario y la modalidad de contratación (remoto vs presencial).
- Las principales habilidades que buscan los empleadores son aquellas relacionadas al uso de la tecnología de la información (IT).

4. Fuente de datos:

La base de datos se obtuvo de la plataforma Kaggle y fue creada por un usuario de LinkedIn mediante el uso de dos scripts de “scraping” o rastrillaje. Un primer script recolectó los avisos de forma inmediata luego de ser publicados, mientras que el segundo script actualizó dicha publicación un tiempo después a una determinada hora aleatoria. Estos scripts se corrieron en paralelo durante dos días consecutivos dos veces obteniendo así dos data sets: el primero se obtuvo con datos de empleos publicados durante los días 23 y 24 de agosto del 2023, y el segundo con datos de empleos publicados durante los días 3 y 4 de noviembre del 2023. En el caso de este segundo data set, se añadió un nuevo campo en el que se registró la fecha y hora en la que el script de actualización registró los datos. Esto es fundamental para poder analizar la popularidad de los empleos publicados en el corto plazo, dado que permite normalizar la cantidad de visitas o aplicaciones en función del tiempo que transcurrió entre que se publicó el empleo y se registró el dato.

5. Enlace del data set:

<https://www.kaggle.com/datasets/arshkon/linkedin-job-postings/data>

6. Alcance del análisis:

Si bien se trata de un data set que registró datos de empleos publicados durante 2 días consecutivos, y por lo tanto no representa con fidelidad lo transcurrido durante todo el año 2023, su estructura nos permite obtener información sobre el impacto de los empleos publicados en el corto plazo. Dicha información podría ser de relevancia para aumentar la efectividad de las publicaciones de empleo generadas por el equipo de recursos humanos de una empresa. Además, el dataset nos permite obtener información sobre las características más comunes de los empleos ofertados y las cualidades que se demandan en ellos, resultando de utilidad para aquellas personas que se encuentren en la búsqueda activa de empleo. Al tener datos recolectados en los meses de agosto y de noviembre, también se puede realizar un estudio comparativo para determinar si hubo cambios en las características de los empleos ofertados entre estos dos puntos temporales.

7. Herramientas tecnológicas utilizadas

Se usaron Microsoft Excel y SQL server para realizar la exploración preliminar de los datos; Draw.io para el diseño del diagrama entidad-relación; y Microsoft PowerBI para el modelado, transformación y visualización de los datos.

8. Nivel de aplicación del análisis:

Se trata de un análisis de nivel operativo dirigido tanto para aquellas personas que se encuentren en la búsqueda de un nuevo empleo como para equipos de recursos humanos de empresas que se encuentren en un proceso activo de reclutamiento, puesto que les permitiría diseñar publicaciones de empleo con mayor impacto.

9. Glosario

- ☛ Salario/hora : es la cantidad de dinero en dólares que se paga por hora trabajada.
- ☛ Beneficios : son atributos que ofrece el empleo por fuera del salario. Los beneficios pueden ser del tipo seguro médico, seguro dental, seguro de visión, seguro de discapacidad, licencia por maternidad, licencia por paternidad, jardín maternal, compensación por viáticos, compensación por educación, préstamo estudiantil o pensión.
- ☛ Experiencia requerida: es el nivel de experiencia que se solicita en cada empleo cuantificada en una escala que va del 1 al 10, siendo 10 el mayor nivel de experiencia requerido (nivel Directivo).
- ☛ Relación salario/experiencia: es la cantidad de salario ofrecida en dólares por unidad de experiencia solicitada.

📌 Visitas / hora: Representa la cantidad de veces que se accedió a la publicación del empleo en 1 hora.

📌 Aplicaciones/hora: Representa la cantidad de personas que aplicaron al empleo en 1 hora.

Aplicaciones / Visitas: Representa la cantidad de aplicaciones por cada unidad de visita. Es una medida de la capacidad de la publicación para transformar visitas en aplicaciones.

📌 Tamaño de la empresa: Se cuantifica usando una escala del 1 al 10, siendo 10 las empresas de mayor tamaño. La relación entre el tamaño y la cantidad de empleados se describe a continuación:

1: 11 a 50 empleados

2: 51 a 200 empleados

3: 201 a 500 empleados

4: 501 a 1,000 empleados

5: 1,001 a 5,000 empleados

6: 5,001 a 10,000 empleados

10.Descripción del dataset:

Tabla 1 - job_postings: es la tabla principal, la tabla de hechos. Contiene todos los registros de empleos publicados en LinkedIn. Posee las siguientes columnas:

📌 job_id: Es un dato de tipo int. Se trata de un número que identifica cada registro nuevo. Es una PK. Permite la vinculación con las tablas 2, 4, 5 y 9.

📌 company_id: Es un dato de tipo int. Se trata de un número que identifica la compañía que publicó el empleo. Es una FK. Permite la vinculación con las tablas 6, 7, 8 y 11.

📌 title: Es un dato de tipo Nvarchar. Es el nombre de la posición del empleo publicado.

📌 description: Es un dato de tipo Nvarchar. Ofrece una descripción general del empleo, en la cual se detallan características de la compañía que lo ofrece, así como las tareas a realizar en el cargo.

📌 formatted_work_type: Es un dato de tipo Nvarchar. Indica el tipo de contratación, es decir si se trata de un empleo tiempo completo, medio tiempo, temporal, voluntariado, etc.

- 📍 **location:** Es un dato de tipo Nvarchar. Indica la ciudad en la que se va a desarrollar el empleo.
- 📍 **State:** Es un dato de tipo Nvarchar. Indica el estado al que pertenece la ciudad en la que se va a desarrollar el empleo.
- 📍 **Country:** Es un dato de tipo Nvarchar. Indica el país en el que se desarrollará el empleo. La gran mayoría son empleos en EE. UU.
- 📍 **applies:** Es un dato de tipo int. Indica la cantidad de personas que aplicaron a ese empleo, desde que se publicó hasta que se registró el dato.
- 📍 **original_listed_time:** Es un dato de tipo int. Se trata de la fecha y hora en la que se publicó el empleo, indicada en formato Unix time. es un sistema para la descripción de instantes de tiempo: se define como la cantidad de segundos transcurridos desde la medianoche UTC del 1 de enero de 1970, sin contar segundos intercalares.
- 📍 **fecha_publicacion:** Es un dato de tipo Nvarchar. Se refiere a la fecha de publicación en formato "día/mes/año" calculada a partir del Unix time de la columna anterior.
- 📍 **hora_publicacion:** Es un dato de tipo Nvarchar. Se refiere a la hora de publicación en formato hora/minutos/segundos calculada a partir del Unix time de la columna anterior.
- 📍 **remote_allowed:** Es un dato de tipo Int. Cobra el valor de 1 si se trata de un trabajo con la posibilidad de ser realizado de forma remota.
- 📍 **views:** Es un dato de tipo int. Indica la cantidad de personas que entraron al enlace del empleo para ver la oferta desde que se publicó hasta que se registró el dato.
- 📍 **job_posting_url:** Es un dato de tipo Nvarchar. Indica el enlace para acceder a la publicación de empleo de LinkedIn.
- 📍 **application_url:** Es un dato de tipo Nvarchar. Indica el enlace para acceder a la aplicación del empleo de LinkedIn.
- 📍 **application_type:** Es un dato de tipo Nvarchar. Describe el tipo de proceso para aplicar al empleo.
- 📍 **expiry:** Es un dato de tipo Nvarchar. Indica la fecha de vencimiento de la publicación del empleo en formato Unix time.
- 📍 **closed_time:** Es un dato de tipo Nvarchar. Indica la hora en la que se cerrará la publicación del empleo en formato Unix Time.

- ☛ **formatted_experience_level:** Es un dato de tipo Nvarchar. Indica el nivel de experiencia requerido para la posición.
- ☛ **skills_desc:** Es un dato de tipo Nvarchar. Es una descripción de las habilidades que se requieren para ocupar la posición.
- ☛ **listed_time:** Es un dato de tipo Nvarchar. Indica la fecha y hora en la que se realizó un repost del empleo en formato Unix Time. Para aquellos empleos que no tuvieron repost, el listed_time tiene el mismo valor que el original_listed_time.
- ☛ **posting domain:** Es un dato de tipo Nvarchar. Ofrece el dominio de la página web que ofrece el empleo.
- ☛ **sponsored:** Es un dato de tipo int. Si su valor es igual a 1, Indica si el empleador pagó para que la oferta de empleo alcance mayor difusión en LinkedIn. Si su valor es 0, significa que la publicación no fue patrocinada.
- ☛ **work_type:** Es un dato de tipo Nvarchar. Indica el tipo de trabajo en términos de contratación (medio tiempo, tiempo completo, temporal, etc.)
- ☛ **scraped:** Es un dato de tipo int. Indica la fecha y hora en la que se relevaron y actualizaron los datos del empleo tales como las vistas y las aplicaciones en formato Unix Time.
- ☛ **fecha_muestreo:** Es un dato de tipo Nvarchar. Indica la fecha en la que se relevaron y actualizaron los datos del empleo tales como las vistas y las aplicaciones en formato día/mes/año.
- ☛ **hora_muestreo:** Es un dato de tipo Nvarchar. Indica la fecha en la que se relevaron y actualizaron los datos del empleo tales como las vistas y las aplicaciones en formato día/mes/año.
- ☛ **diferencia_publicado_muestreo:** Es un dato de tipo int. Es la diferencia entre las columnas scraped y original_listed_time. Indica el tiempo que pasó entre que se publicó el empleo y se actualizó en formato Unix time.
- ☛ **diferencia_hora:** Es un dato de tipo Nvarchar. Es el cálculo de la diferencia entre el tiempo que pasó entre que se publicó el empleo y se relevó la actualización en formato hora:minutos:segundos

Tabla 2 - job_skills: indica las habilidades que se demandan para cada empleo a través de un código abreviado de letras.

- ☛ **job_id:** Es un dato de tipo int. Es una FK. Identifica la publicación de empleo y permite la vinculación con la tabla 1, 4, 5 y 9 a través de la clave.

☛ skill_abr: Es un dato de tipo Nvarchar. Es una FK. Identifica a través de un código de letras las habilidades solicitadas para la posición del empleo y permite la vinculación con la tabla 3.

Tabla 3 – skills: describe las habilidades que fueron abreviadas en la tabla anterior.

☛ skill_abr: Es un dato de tipo Nvarchar. Es una PK. Identifica a través de un código de letras las habilidades solicitadas para la posición del empleo y permite la vinculación con la tabla 2.

☛ skill_name: Es un dato de tipo Nvarchar. Describe la habilidad solicitada para la posición de empleo codificada.

Tabla 4 – salaries: ofrece información en relación con los salarios que son ofrecidos por las distintas publicaciones.

☛ salary_id: Es un dato de tipo int. Es una PK. Identifica todos los registros de salarios.

☛ job_id: Es un dato de tipo int. Es una FK. Identifica la publicación de empleo y permite la vinculación con la tabla 1, 2, 5 y 9 a través de la clave.

☛ max_salary: Es un dato de tipo int. Es el valor máximo del salario ofrecido por el empleador.

☛ med_salary: Es un dato de tipo int. Es el valor promedio del salario ofrecido por el empleador.

☛ min_salary: Es un dato de tipo int. Es el valor mínimo del salario ofrecido por el empleador.

☛ pay_period: Es un dato de tipo Nvarchar. Describe el tipo de pago en término de tiempo, es decir si se trata de un salario por hora, mensual o anual.

☛ currency: Es un dato de tipo Nvarchar. Identifica la moneda en que se pagará el sueldo.

☛ compensation type: Es un dato de tipo Nvarchar. Indica el tipo de compensación.

☛ salario_normalizado_hora: Es un dato de tipo float. Es el cálculo del salario pagado por hora, incluso para aquellos salarios que se publicaron por mes o año.

Tabla 5 – benefits: describe los beneficios que son ofertados en cada publicación.

☛ job_id: Es un dato de tipo int. Es una FK. Identifica la publicación de empleo y permite la vinculación con la tabla 1, 2, 4 y 9 a través de la clave.

- ☛ **inferred:** Es un dato de tipo int. Indica si el beneficio fue inferido a través del texto por LinkedIn o fue especificado explícitamente a través de un tag.
- ☛ **type:** Es un dato de tipo Nvarchar. Describe el tipo de beneficio ofrecido por el empleador.

Tabla 6 – **companies:** ofrece información relativa a las distintas empresas que ofrecen la posición de empleo.

- ☛ **company_id:** Es un dato de tipo int. Es un PK. Identifica la compañía que ofertó el empleo y permite la vinculación con la tabla 1, la tabla 7, la tabla 8 y la tabla 11.
- ☛ **name:** Es un dato de tipo Nvarchar. Describe el nombre de la compañía que ofertó el empleo.
- ☛ **description:** Es un dato de tipo Nvarchar. Ofrece una descripción de la compañía incluyendo su visión y valores.
- ☛ **company_size:** Es un dato de tipo int. Indica el tamaño de la compañía a través de un código numérico: 0 se le asigna a las compañías más pequeñas y 7 se le asigna a las más grandes.
- ☛ **state:** Es un dato de tipo Nvarchar. Identifica el estado en el que se encuentra la compañía.
- ☛ **country:** Es un dato de tipo Nvarchar. Identifica el país en el que se encuentra la compañía.
- ☛ **city:** Es un dato de tipo Nvarchar. Identifica la ciudad en el que se encuentra la compañía.
- ☛ **zip_code:** Es un dato de tipo Nvarchar. Identifica el código postal de la zona en la que se encuentra la compañía.
- ☛ **address:** Es un dato de tipo Nvarchar. Identifica la dirección en la que se encuentra la compañía.
- ☛ **url:** Es un dato de tipo Nvarchar. Ofrece el enlace de la página web de la compañía.

Tabla 7 – **company_specialities:** indica la especialidad en la que se desempeña la empresa.

- ☛ **company_id:** Es un dato de tipo int. Es una FK. Identifica la compañía que ofertó el empleo y permite la vinculación con la tabla 6, la tabla 8, la tabla 11 y la tabla 1.
- ☛ **speciality:** Es un dato de tipo Nvarchar. Describe la especialidad en la que se desenvuelve la compañía.

Tabla 8 – company_industries: indica el rubro industrial en el que se desempeña la empresa.

- ☛ company_id: Es un dato de tipo int. Es una FK. Identifica la compañía que ofertó el empleo y permite la vinculación con la tabla 6, la tabla 7, la tabla 11 y la tabla 1.
- ☛ industry: Es un dato de tipo Nvarchar. Describe la industria a la que pertenece la compañía.

Tabla 9 – job_industries: indica el rubro industrial al que se relaciona la posición ofertada a través de un identificador.

- ☛ job_id: Es un dato de tipo int. Es una FK. Identifica la publicación de empleo y permite la vinculación con la tabla 1, 2, 4 y 5 a través de la clave.
- ☛ industry_id: Es un dato de tipo int. Es una FK. Identifica la industria a la que pertenece el empleo y permite la vinculación con la tabla 10.

Tabla 10 – industries: describe el rubro industrial al que se relaciona cada empleo asociado a cada identificador.

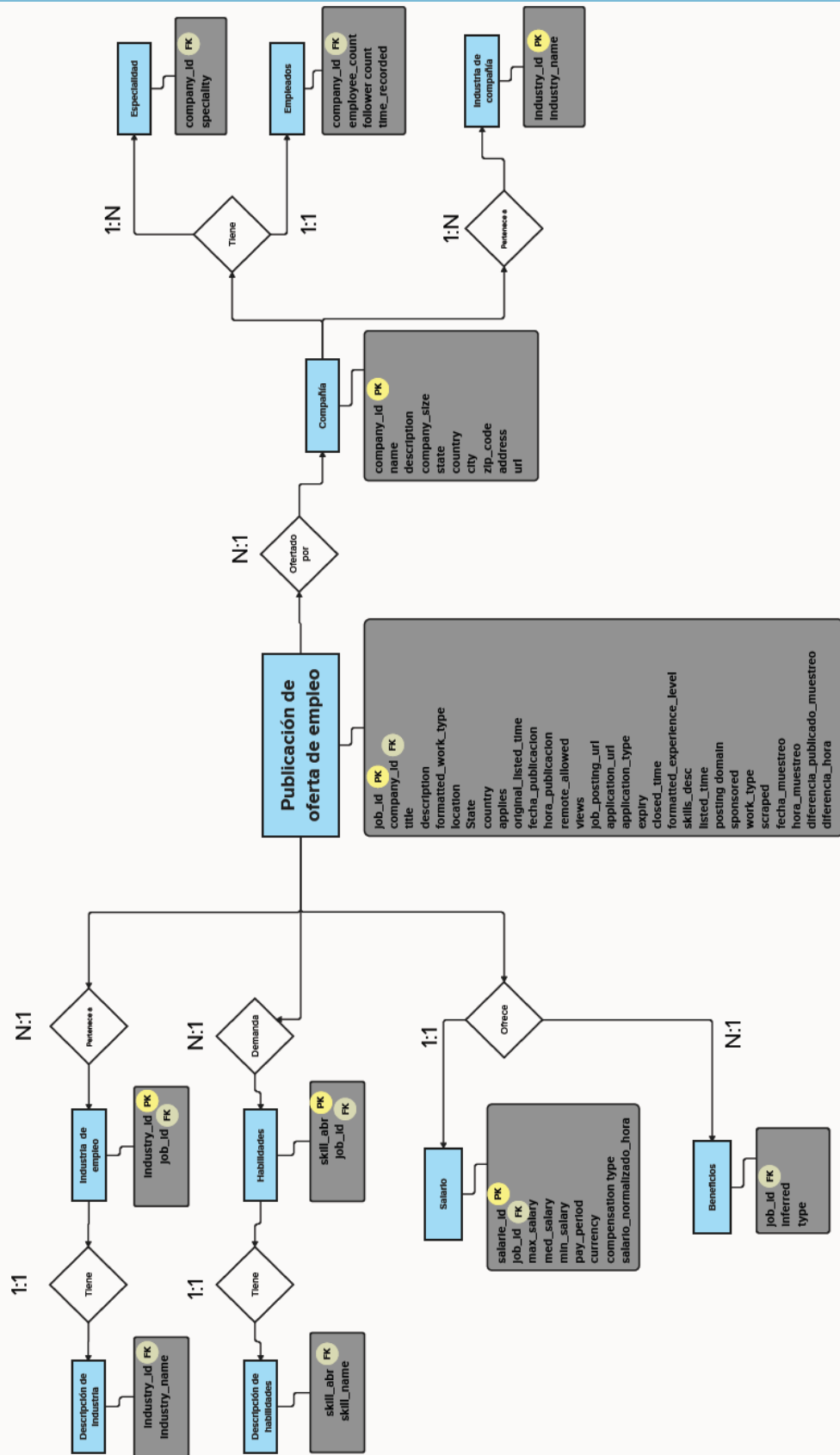
- ☛ industry_id: Es un dato de tipo int. Es una PK. Identifica la industria a la que pertenece el empleo y permite la vinculación con la tabla 9.
- ☛ industry_name: Es un dato de tipo Nvarchar. Describe la industria a la que pertenece el empleo.

Tabla 11 – employee_counts: ofrece información adicional relacionada al número de empleados y seguidores que tiene la empresa en LinkedIn.

- ☛ company_id: Es un dato de tipo int. Es una FK. Identifica la compañía que ofertó el empleo y permite la vinculación con la tabla 6, la tabla 7, la tabla 8 y la tabla 1.
- ☛ employee_count: Es un dato de tipo int. Indica el número de empleados en la compañía.
- ☛ follower count: Es un dato de tipo int. Indica el número de seguidores que tiene la cuenta de la compañía en LinkedIn.
- ☛ time_recorded: Es un dato de tipo Nvarchar. Indica la fecha y hora en la que se recolectaron estos datos en formato Unix Time.

11. Diagrama de entidad relación

Diagrama entidad - relación



12. Tabla de versiones

Versión	Acciones
V00	Definición del proyecto y elección del dataset Carga del data set en excel Análisis exploratorio en excel
V01	Definición de los lineamientos estéticos del proyecto Elección de imágenes y paleta de colores a incluir Diseño de las páginas "Portada" y Diseño del primer Mockup
V02	Transformación de datos para la normalización de salarios , beneficios, escala de experiencia y habilidades requeridas. Cálculos de medidas para las métricas referidas a salarios, beneficios y experiencia requerida Creación de las páginas "General empleados", "Insights Salarios", "Insights Beneficios", "Insights experiencia"
V03	Transformación de datos para a cuantificación de la popularidad de empleos; para la obtención de variables categóricas Salario SI/NO y Beneficios SI/NO; y para la obtención de la cantidad de palabras de la columna "description" Cálculo de medidas para la cuantificación de la popularidad de empleos Creación de las páginas "Empleadores 1", "Empleadores 2" y "Empleadores 3" Creación de tabla calendario
Versión final	Ensamblado y paginación del reporte Creación de la solapa "glosario"

13. Procesamiento de los datos:

13.1 Transformación para la normalización de salarios:

Los salarios registrados en la base de datos tenían un formato heterogéneo, de modo tal que fue necesario realizar una serie de transformaciones para normalizarlos y de esa forma hacerlos comparables entre sí. Las diferencias entre salario se encontraban principalmente en tres aspectos:


1. En primer lugar, tal como se observa en el ejemplo siguiente, en algunos registros se indica el salario máximo y el mínimo, mientras que en otros solo se registra el salario medio. Para normalizar esta diferencia, en aquellos salarios en donde se registró un máximo y un mínimo, se calculó el promedio de estos en el Excel, mediante la función AVERAGE() antes de cargar la base de datos a SQL server.

A	B	C	D	E	F	G	H
salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
57	3696990096	3060000		2080000	YEARLY	USD	BASE_SALARY
58	3696990991		40000		MONTHLY	USD	BASE_SALARY



2. Por otro lado, mientras que en algunos registros los últimos dos dígitos del salario representaban decimales, en otros solo el último dígito correspondía a un número decimal. Por ejemplo, en el caso del registro con salary_id = 19 se registró un salario máximo de 2555250 y al evaluar la publicación del empleo en LinkedIn se ofrece un salario de 255525\$, indicando que el último 0 agregado al registro corresponde a un número decimal. Por otro lado, en el caso del registro con salary_id=20, observamos que se registró un salario máximo de 6414 mientras que en la publicación de dicho empleo se ofrece un salario máximo de 64,14\$, indicando que los últimos dos dígitos del número registrado corresponden a decimales. Para proceder a la normalización de decimales se encontraron ciertos patrones que permitieron decidir si el salario tenía uno o dos dígitos decimales:

salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
19	369638859	2555250	null	1857000	YEARLY	USD	BASE_SALARY
20	3696394452	6414	null	6414	HOURLY	USD	BASE_SALARY




Physician - CenterWell - Las Vegas, NV
CenterWell Senior Primary Care - Las Vegas, NV · Publicado de nuevo hace 5 meses

185,700 \$/año - 255,525 \$/año · Presencial · Jornada completa

De 1.001 a 5.000 empleados · Cuidados y atención sanitaria

Mira las tendencias de contratación recientes de CenterWell Senior Primary Care.

Ya no se aceptan solicitudes



Device Hardware Asset Manager
Talent Table · Oakland, CA · hace 6 meses

64,14 \$/hr · En remoto · Contrato por obra · Intermedio


De 11 a 50 empleados · Dotación y selección de personal

Ve una comparación con los otros 13 solicitantes. [Probar Premium p...](#)

Ya no se aceptan solicitudes

Se encontraron salarios por hora con registros de salarios mínimos de 4 dígitos y salarios máximos de 3 dígitos. Como el salario máximo no puede ser menor que el mínimo, lo que ocurrió en estos casos es que el salario máximo tenía un solo dígito decimal, mientras que el mínimo tenía dos. Para normalizarlos se calculó el promedio de ambos y se aplicó como corrección una división en un factor de 10 en el salario mínimo (líneas 1 a 3 del código desarrollado a continuación).

A	B	C	D	E	F	G	H
salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
115	3700549721	219		1455	HOURLY	USD	BASE_SALARY



Retail Team Lead (PT) ✓
New Balance · Thornton, CO · Publicado de nuevo hace 5 meses

14,55 \$/h - 21,90 \$/h · Presencial · Media jornada · Intermedio

De 5.001 a 10.000 empleados · Fabricación de artículos de deporte

Ve una comparación con los otros 5 solicitantes. [Probar Premium por 0 AR\\$](#)

Para aquellos salarios por hora en los cuales el salario mínimo y el máximo tenían solo 3 dígitos, se observó que sólo el último dígito representaba el decimal, de modo tal que solo se calculó el promedio de estos (líneas 4-5 del código)

salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
117	3700550106		390		310	HOURLY	USD BASE_SALARY

Permanent Staff Acute Care Float RN job in Greenville, SC - Make \$31 to \$39/hour ✓

Greenville, SC · Publicado de nuevo hace 5 meses

31 \$/h - 39 \$/h · Presencial · Jornada completa · Sin experiencia

Noche · 12 horas

De 5.001 a 10.000 empleados · Dotación y selección de personal

Para aquellos salarios con registros de salarios máximos y mínimos de 4 dígitos, se observó que los últimos dos dígitos correspondían a números decimales, de modo que se calculó el promedio de ellos corrigiendo por un factor de 10 (líneas 6-7 del código)

salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
137	3700552599		3462		1538	HOURLY	USD BASE_SALARY

Intern - Information Security (Fall 2023) ✓

Itron Inc. · Liberty Lake, WA · Publicado de nuevo hace 5 meses

15,38 \$/h - 34,62 \$/h · Híbrido · Prácticas · Prácticas

De 5.001 a 10.000 empleados · Servicios públicos

Ve una comparación con otros 100 solicitantes. [Probar Premium por 0 AR\\$](#)

Para aquellos salarios en los que el registro de salario máximo presentaba 4 dígitos y el de salario mínimo 3, la transformación fue más compleja, porque en algunos casos los últimos dos dígitos del salario máximo representaban números decimales, mientras que en otros sólo el último dígito era decimal. Lo que se observó a través del filtrado y ordenado de salarios máximos, fue que la primera condición se cumplía para registros de salarios mínimos mayores a 750, mientras que la segunda se cumplía para salarios máximos menores a dicho valor. Para corregir este fenómeno se aplicó el Código observado entre las líneas 8-11.

salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
161	3700559014	2322		180	HOURLY	USD	BASE_SALARY



Residential Behavior Tech/Care Coordinator

Behavioral Health ✓

Newport Healthcare · Shingle Springs, CA · Publicado de nuevo hace 5 meses

18 \$/h - 23,22 \$/h · Presencial · Jornada completa · Sin experiencia

De 1.001 a 5.000 empleados · Atención a la salud mental

Mira las tendencias de contratación recientes de Newport Healthcare. [Probar Premium por 0 AR\\$](#)

salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
374	3693044890	1150		780	HOURLY	USD	BASE_SALARY



Compliance Manager with 10+ years experience

Wissman Group · Mechanicsburg, PA · hace 6 meses


78 \$/h - 115 \$/h · Presencial · Contrato por obra · Algo de responsabilidad

De 1.001 a 5.000 empleados · Dotación y selección de personal

Para aquellos salarios por hora sin registros de salario máximo y con registros de salario medio se aplicó la siguiente regla: si el registro tenía 3 dígitos se lo dejó sin modificaciones, mientras que si tenía 4 dígitos se le aplicó una corrección


dividiéndolo por un factor de 10, asumiendo que los dos últimos dígitos correspondían a números decimales (líneas 12-15 del código)


salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
418	3693045318			2425	HOURLY	USD	BASE_SALARY




CDL A Truck Driver


Hustle Partners · Sparks, NV · hace 6 meses

 24,25 \$/hr · Presencial · Jornada completa

 Entre 1 y 10 empleados


 Ve una comparación con los otros 4 solicitantes. [Probar Premium por 0 AR\\$](#)


salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
451	3693045715			380	HOURLY	USD	BASE_SALARY




Diagnostic Sonographers

MD Imaging · Redding, CA · hace 6 meses


 38 \$/hr · Presencial · Jornada completa

 De 51 a 200 empleados

 Mira las tendencias de contratación recientes de MD Imaging. [Probar Premium por 0 AR\\$](#)


Para el caso de los salarios mensuales y anuales, se observó que aquellos que terminaban en 0 o 5, solo el último dígito correspondía a un número decimal, mientras que en aquellos que terminaban en otro número los últimos dos dígitos correspondían a números decimales. Para detectar estos últimos casos se aplicó una regla que calculaba el resto de la división por 10 o por 5, de modo que si el resto era distinto de 0 se asumía que los últimos dos dígitos del salario correspondían a números decimales y se les aplicaba una corrección por 10 (líneas 16-23 del código).


salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
9133	3757449515		524467		MONTHLY	USD	BASE_SALARY




Operations Focal Point, Flight Management

IOM-USRAP DC · Nueva York, NY · hace 3 meses

 5.244,67 \$/month · Presencial · Jornada completa

 Más de 10.001 empleados

 Ve una comparación con los otros 46 solicitantes. [Probar Premium por 0 AR\\$](#)

salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
9285	3757485638		20000		MONTHLY	USD	BASE_SALARY



Registered Nurse - RN

CareerStaff Unlimited · Laredo, TX · hace 3 meses

 2.000 \$/month · Presencial · Jornada completa

 Cuidados a largo plazo · Día · 8 horas

 De 5.001 a 10.000 empleados · Dotación y selección de personal

 Mira las tendencias de contratación recientes de CareerStaff Unlimited. [Probar Premium por 0 AR\\$](#)

Finalmente, para los salarios semanales se observó que el último dígito corresponde a un número digital de modo que se dejaron sin modificaciones (línea 24 del código)

salary_id	job_id	max_salary	med_salary	min_salary	pay_period	currency	compensation_type
8165	3757478680	23680		22150	WEEKLY	USD	BASE_SALARY



Speech Therapist - Home Health

Host Healthcare, Inc. · Mason, OH · Publicado de nuevo hace 3 meses

2.215 \$/semana - 2.368 \$/semana · Presencial · Jornada completa · Sin experiencia

De 1.001 a 5.000 empleados · Hospitales y atención sanitaria

Mira las tendencias de contratación recientes de Host Healthcare, Inc., [Probar Premium por 0 AR\\$](#)

#NORMALIZACION DE DECIMALES

```

= Table.AddColumn("#Changed Type1", "Custom", each if
[pay_period]="HOURLY" and [min_salary] > 999 and [max_salary] < 1000 then
((([min_salary]/10)+[max_salary])/2)

else if [pay_period]="HOURLY" and [min_salary] < 1000 and [max_salary] < 1000
then ([min_salary]+[max_salary])/2

else if [pay_period]="HOURLY" and [min_salary] > 999 and [max_salary] > 999
then ((([min_salary]/10)+([max_salary]/10))/2)

else if [pay_period]="HOURLY" and [min_salary] < 750 and [max_salary] > 999
then ([min_salary]+([max_salary]/10))/2

else if [pay_period]="HOURLY" and [min_salary] > 750 and [max_salary] > 999
then ([min_salary]+[max_salary])/2

else if [pay_period]="HOURLY" and [min_salary] = null and [med_salary] < 1000
then [med_salary]

else if [pay_period]="HOURLY" and [min_salary] = null and [med_salary] > 999
then [med_salary]/10

else if [pay_period]="MONTHLY" and Number.Mod([med_salary], 10.0)= 0.0 or
Number.Mod([med_salary], 5.0)= 0.0 then [med_salary]

else if [pay_period]="MONTHLY" and Number.Mod([med_salary], 10.0) > 0.0
then [med_salary]/10

```

```

else if [pay_period]="YEARLY" and Number.Mod([med_salary], 10.0)= 0.0 or
Number.Mod([med_salary], 5.0)= 0.0 then [med_salary]

else if [pay_period]="YEARLY" and Number.Mod([med_salary], 10.0) > 0.0 then
[med_salary]/10

else [med_salary])

```

Con esta sentencia de código M se logró hacer que el último dígito de todos los salarios promedio corresponda a un número decimal.

3. En tercer lugar, tal como se observa en los ejemplos anteriores, los salarios se ofrecen con distintos tipos de período de pago: algunos son anuales, otros mensuales, otros semanales y otros por hora. Para normalizarlos y hacerlos comparables entre sí se aplicó el siguiente código para transformar todos a salario por hora, asumiendo que un día tiene 8 horas de jornada laboral, una semana tiene 5 días laborales, un mes tiene en promedio 21 días laborales y un año tiene en promedio 249 días laborales. En la normalización, se dividió el valor del salario promedio calculado anteriormente por el siguiente factor según si se trataba de salarios anuales, mensuales, semanales o por hora:

1 año = 1992 horas

1 mes = 168 horas

1 semana = 40 horas

Además, se agregó a la división un factor de 10 para corregir el último número decimal. Para ello se usó la siguiente sentencia de código M:

```

#NORMALIZACION POR TIEMPO DE PAGO
= Table.AddColumn("#Removed Columns", "Salario / hora", each if
[pay_period]="YEARLY" then [med_salary]/19920
else if [pay_period]="MONTHLY" then [med_salary]/1680
else if [pay_period]="WEEKLY" then [med_salary]/400
else [med_salary]/10)

```

12.2 Transformación para la cuantificación de beneficios totales por empleo:

Se realizó el recuento de beneficios totales ofrecido por cada empleo mediante la función de recuento de filas agrupadas por job_id en la tabla benefits\$. Se guardó esta consulta en una nueva tabla llamada benefits_transformado\$

12.3 Transformación para la cuantificación de beneficios específicos / rubro:

Para poder realizar el análisis desglosado de la frecuencia con la que los distintos rubros ofrecen diferentes beneficios se generó para cada beneficio una tabla nueva uniendo las tablas `job_industries$` e `industries$` a través de la columna `industry_id` usando la función `"Merge query"`. Luego, con esta misma función esta tabla se fusionó con la tabla `benefits$` a través de la columna `job_id`. Finalmente, en esta tabla se filtró uno a uno por cada tipo de beneficio (columna `"type"`) y luego de cada filtrado se empleó la función de conteo de filas agrupadas por `industry_name`. Esta información se almacenó en distintas tablas (una por cada tipo de beneficio) en la que se registró en una columna el rubro (`industry_name`) y en otra el conteo. Finalmente, usando la función `"Merge query"`, todas las tablas se fusionaron en una sola a través de la columna `"industry_name"`. Además, en esta tabla se registró en otra columna el número total de empleos ofertados por cada tipo de industria. Esta tabla matriz se empleó para alimentar el gráfico de tipo map tree en la solapa `"Insights beneficios"` del reporte.

12.4 Transformación para la cuantificación de la experiencia:

En el dataset original la experiencia se registró de forma cualitativa a través de 5 categorías fijas que establece la plataforma de LinkedIn en la columna `"formatted_experience_level"` de la tabla `"job_postings"`. Para poder transformar esta información a una escala cuantitativa y de esa forma comparar las distintas exigencias de cada empleo, se le asignó a cada categoría un número directamente proporcional al nivel de experiencia de cada categoría. Para ello se creó una nueva columna llamada `"Escala_experiencia"` en la tabla `"job_postings"` y se asignó el valor de la celda en función de la categoría de la columna `"formatted_experience_level"` de la siguiente manera:

Internship = 0

Entry level = 1

Associate = 3

Mid-Senior level = 5

Executive = 8

Director = 10

13.5 Transformación para la cuantificación de habilidades requeridas por empleo:

Se realizó el recuento de beneficios totales ofrecido por cada empleo mediante la función de recuento de filas agrupadas por job_id en la tabla skills\$. Se guardó esta consulta en una nueva tabla llamada skills_transformados\$.

13.6 Transformación para la cuantificación de la popularidad de los empleos:

Se generó una copia de la tabla job_postings\$ a la cual se le eliminaron las siguientes columnas: "company_id", "title", "description", "location", "State", "Country", "original_listed_time", "job_posting_url", "application_url", "expiry", "closed_time", "formatted_experience_level", "skills_desc", "listed_time", "posting_domain", "sponsored", "work_type", "scraped", "diferencia_publicado_muestreo_listed" y "Escala_experiencia". Al ordenar las filas en orden descendente a través de la columna diferencia_publicado_muestro_original, que representa la cantidad de segundos que pasaron entre que se publicó el empleo y se midió la cantidad de aplicaciones o visitas se encontró una inconsistencia: algunos empleos tenían una diferencia negativa. Para eliminar dicha inconsistencia se procedió a crear una nueva columna llamada "diferencia_hora_corregida" usando el siguiente código de lenguaje M:

```
if [diferencia_publicado_muestro_original] < 0 then "null"
else [hora_diferencia_original])
```

De esa forma, a aquellos empleos con diferencia de tiempo negativa se les asignó un valor null.

Luego, a la columna "remote_allowed" se le reemplazó el valor de 1 por el valor "SI" y a las celdas vacías por el valor "NO".

Luego, tanto a la columna "applies" como a la columna "views" se le reemplazaron las celdas vacías por el valor 0.

Para normalizar la cantidad de visitas o aplicaciones por unidad de tiempo, se procedió a crear 2 columnas usando los siguientes códigos de lenguaje M:

```
#normalización de visitas/unidad de tiempo
```

```
= Table.AddColumn("#Replaced Value3", "Visitas_segundo", each if [views]>0 and
[diferencia_publicado_muestro_original] > 0 then
[views]/([diferencia_publicado_muestro_original]/3600)
else null)
```

```
#normalización de aplicaciones/unidad de tiempo
```

```
= Table.AddColumn("#Renamed Columns", "aplicaciones_hora", each if [applies]>0 and  
[diferencia_publicado_muestro_original]>0 then  
[applies]/([diferencia_publicado_muestro_original]/3600)  
else null)
```

De esta manera cada columna registró la cantidad de visitas o aplicaciones por hora.

Finalmente, para poder establecer la cantidad de horas que transcurrieron entre que el empleo se publicó y que se registraron las aplicaciones o visitas se creó una nueva columna llamada "horas_transcurridas" mediante la siguiente sentencia de código M:

```
= Table.AddColumn("#Replaced Errors", "Custom", each if  
Time.Hour([hora_diferencia_original]) > 0 and  
Time.Hour([hora_diferencia_original]) <= 1 then 1  
  
else if Time.Hour([hora_diferencia_original]) > 1 and  
Time.Hour([hora_diferencia_original]) <= 2 then 2  
  
else if Time.Hour([hora_diferencia_original]) > 2 and  
Time.Hour([hora_diferencia_original]) <= 3 then 3  
  
else if Time.Hour([hora_diferencia_original]) > 3 and  
Time.Hour([hora_diferencia_original]) <= 4 then 4  
  
else if Time.Hour([hora_diferencia_original]) > 4 and  
Time.Hour([hora_diferencia_original]) <= 5 then 5  
  
else if Time.Hour([hora_diferencia_original]) > 5 and  
Time.Hour([hora_diferencia_original]) <= 6 then 6  
  
else if Time.Hour([hora_diferencia_original]) > 6 and  
Time.Hour([hora_diferencia_original]) <= 7 then 7  
  
else if Time.Hour([hora_diferencia_original]) > 7 and  
Time.Hour([hora_diferencia_original]) <= 8 then 8  
  
else if Time.Hour([hora_diferencia_original]) > 8 and  
Time.Hour([hora_diferencia_original]) <= 9 then 9  
  
else if Time.Hour([hora_diferencia_original]) > 9 and  
Time.Hour([hora_diferencia_original]) <= 10 then 10  
  
else if Time.Hour([hora_diferencia_original]) > 10 and  
Time.Hour([hora_diferencia_original]) <= 11 then 11  
  
else if Time.Hour([hora_diferencia_original]) > 11 and  
Time.Hour([hora_diferencia_original]) <= 12 then 12
```

```

else if Time.Hour([hora_diferencia_orginal]) > 12 and
Time.Hour([hora_diferencia_orginal]) <= 13 then 13

else if Time.Hour([hora_diferencia_orginal]) > 13 and
Time.Hour([hora_diferencia_orginal]) <= 14 then 14

else if Time.Hour([hora_diferencia_orginal]) > 14 and
Time.Hour([hora_diferencia_orginal]) <= 15 then 15

else if Time.Hour([hora_diferencia_orginal]) > 15 and
Time.Hour([hora_diferencia_orginal]) <= 16 then 16

else if Time.Hour([hora_diferencia_orginal]) > 16 and
Time.Hour([hora_diferencia_orginal]) <= 17 then 17

else if Time.Hour([hora_diferencia_orginal]) > 17 and
Time.Hour([hora_diferencia_orginal]) <= 18 then 18

else if Time.Hour([hora_diferencia_orginal]) > 18 and
Time.Hour([hora_diferencia_orginal]) <= 19 then 19

else if Time.Hour([hora_diferencia_orginal]) > 19 and
Time.Hour([hora_diferencia_orginal]) <= 20 then 20

else if Time.Hour([hora_diferencia_orginal]) > 20 and
Time.Hour([hora_diferencia_orginal]) <= 21 then 21

else if Time.Hour([hora_diferencia_orginal]) > 21 and
Time.Hour([hora_diferencia_orginal]) <= 22 then 22

else if Time.Hour([hora_diferencia_orginal]) > 22 and
Time.Hour([hora_diferencia_orginal]) <= 23 then 23

else if Time.Hour([hora_diferencia_orginal]) > 23 and
Time.Hour([hora_diferencia_orginal]) <= 24 then 24

else null)

```

Toda esta información quedó registrada en la tabla "Job_popularity"

13.7 Transformación para obtener las variables categóricas "Salario SI o Salario NO"

Se procedió a crear una nueva tabla llamada "Empleos_con_sin_salario" a través de la fusión de las tablas "salarios_transformados" y la tabla "job_postings" usando la columna "job_id". Se eliminaron todas las columnas a excepción de "salarios_transformados.Salario/hora" y "job_id". Luego se agregó una nueva columna llamada "Salario_si_no" usando la siguiente sentencia de código M:


```
= Table.AddColumn("#Reordered Columns", "Salario_si_no", each if
[#salarios_transformados.Salario/hora]=null then "NO"
else "SI")
```

13.8 Transformación para obtener las variables categóricas “Beneficio SI o Beneficio NO”

Se procedió a crear una nueva tabla llamada “Empleos_con_sin_salario” a través de la fusión de las tablas “benefits_transformado” y la tabla “job_postings” usando la columna “job_id”. Se eliminaron todas las columnas a excepción de “benefits_transformado.Salario/hora” y “job_id”. Luego se agregó una nueva columna usando la siguiente sentencia de código M:

```
= Table.AddColumn("#Reordered Columns", "Salario_si_no", each if
[#salarios_transformados.Salario/hora]=null then "NO"
else "SI")
```

Se reemplazaron los valores null por el valor 0 y se creó una nueva columna llamada “Beneficio_si_no” usando la siguiente sentencia de código M:

```
= Table.AddColumn("#Replaced Value", "Beneficio_si_no", each if
[#benefits_transformado$.Count] > 0 then "SI"
else "NO")
```

13.9 Transformación para obtener la cantidad de palabras de la columna “Description”

Se creó una nueva tabla llamada “Descripciones” duplicando la tabla “job_postings” y eliminando todas las columnas a excepción de “job_id” y “description”. Luego se creó una nueva columna llamada “cantidad de palabras” usando la siguiente sentencia de código M:

```
Table.AddColumn("#Removed Other Columns", "Custom", each
List.Count(Text.Split([description], " ")))
```

14. Medidas calculadas:

👤 **Promedio global salarios:** es el promedio de todos los salarios normalizados presentes en el dataset. Se calculó usando la función AVERAGE.

Promedio global salarios = AVERAGE(salarios_transformados[Salario/hora])

📌 **Mediana global salarios** : es la mediana de todos los salarios normalizados presentes en el dataset. Se calculó usando la función MEDIAN.

📌 **Salario/hora promedio por industria**: es el promedio de los salarios normalizados para una determinada industria. Se calculó mediante la función AVERAGEX, agrupando por industria.

Salario/hora promedio por industria =
`AVERAGEX(KEEPFILTERS(VALUE('industries$'[industry_name])),CALCULATE(AVERAGE('salarios_transformados'[Salario/hora])))`

📌 **Cantidad de beneficios promedio por industria**: es el promedio de la cantidad de beneficios ofrecido por los empleos de una industria determinada. Se calculo aplicando la función AVERAGEX sobre la tabla `benefits_transformado$` mencionada anteriormente.

Cantidad de beneficios promedio por industria =
`AVERAGEX(KEEPFILTERS(VALUE('industries$'[industry_name])),CALCULATE(AVERAGE('benefits_transformado$'[Count])))`

📌 **Experiencia promedio por industria**: es el promedio de nivel de experiencia requerido expresado en la escala cuantitativa que se exige para una industria determinada. Se calculó empleando la función AVERAGEX sobre la columna "Escala_experiencia" en la tabla "job_postings\$" mencionada anteriormente:

Experiencia promedio por industria = `AVERAGEX(KEEPFILTERS(VALUE('industries$'[industry_name])),CALCULATE(AVERAGE('job_postings$'[Escala_experiencia])))`

📌 **Promedio de cantidad de trabajos por industria**: es el recuento de empleos publicados por cada rubro. Se calculo usando la siguiente función:

Promedio de cantidad de trabajos por industria =
`AVERAGEX(KEEPFILTERS(VALUE('industries$'[industry_name])),CALCULATE(COUNTA("'"job_industries $"'[job_id]")))`

📌 **% de empleos que ofrecen (beneficio)** : Es el % del total de empleos que ofrecen un determinado beneficio. Se calculo dividiendo la cantidad de empleos que ofrecen un beneficio sobre el total de empleos expresado en forma de porcentaje usando la tabla `Beneficios_industria` mencionada anteriormente.

Ejemplo:

% empleos que ofrecen jardín maternal =

```
DIVIDE(SUM('Beneficios_industria'[Child_care.Child Care Support]),  
SUM('Beneficios_industria'[Total empleos]))
```

☛ **Aplicaciones por visitas:** es la cantidad de aplicaciones por unidad de visitas que obtuvo cada empleo. Se calcula de la siguiente manera:

Aplicaciones/visitas =

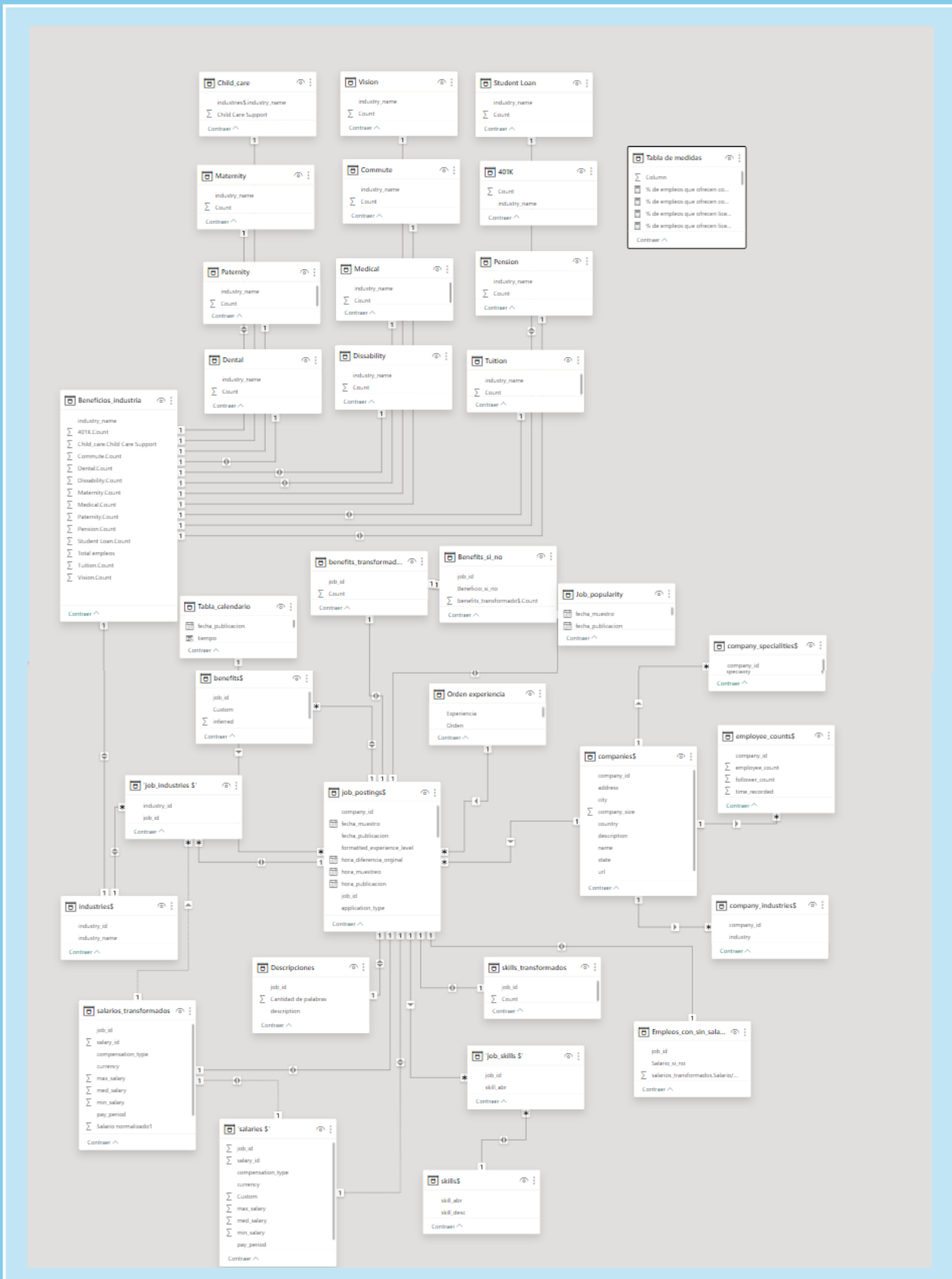
```
DIVIDE(  
    AVERAGE('Job_popularity'[aplicaciones_hora]),  
    AVERAGE('Job_popularity'[Visitas_hora]))
```

☛ **Salario/Experiencia:** es la cantidad de salario por unidad de experiencia requerida agrupado por industria. Se calcula con la siguiente función:

Salario/Experiencia =

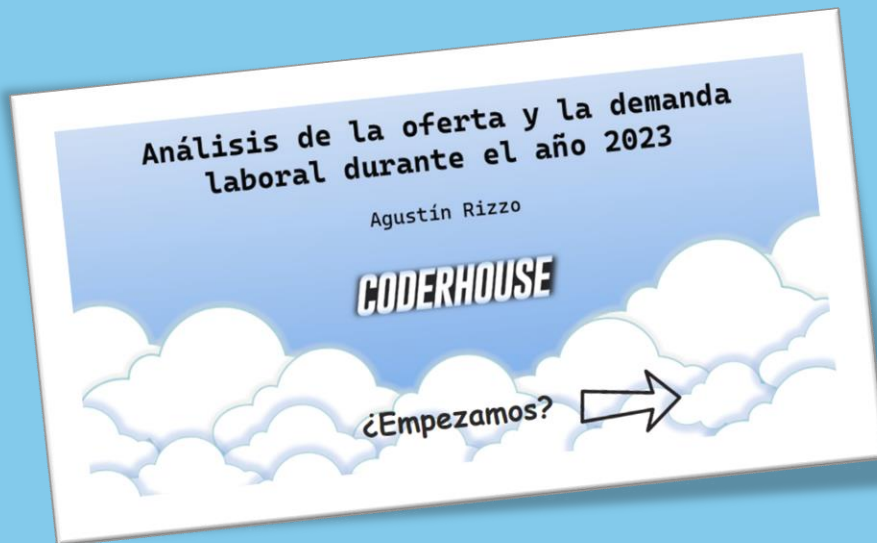
```
DIVIDE(  
    [Salario/hora promedio por industria],  
    [Experiencia promedio por industria])
```

15. Modelo de datos en PowerBI



16. Descripción del reporte:

- **Portada** En esta solapa inicial se describe el título y autor del proyecto. A través del botón de tipo flecha se puede acceder a la siguiente página del reporte (solapa "Elección")

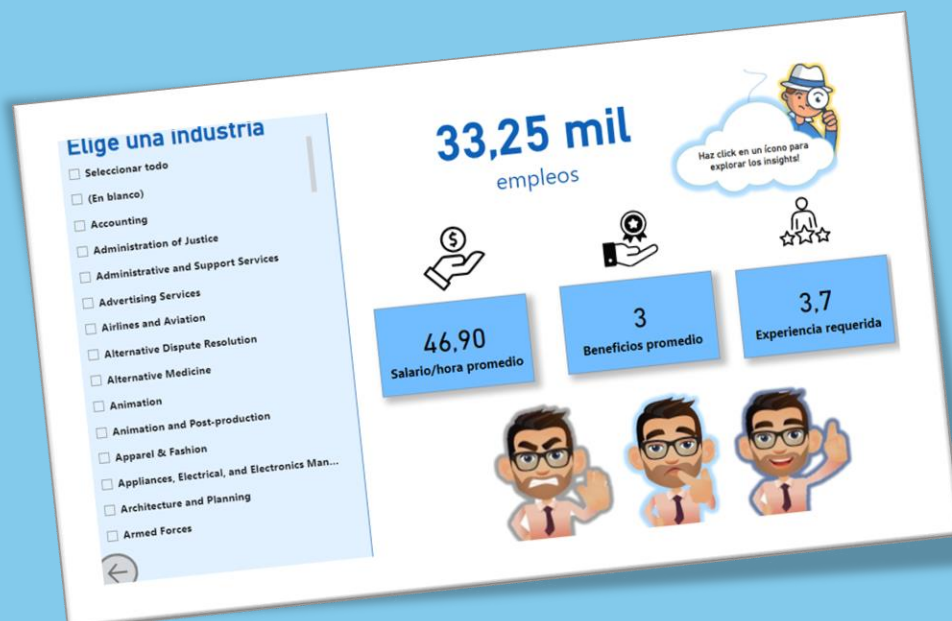


- **Elección** Esta solapa refleja los dos grandes tipos de análisis de los datos que se pueden realizar con este dataset: por un lado el análisis de la oferta laboral, representado por el cartel "busco empleo" que le permite al usuario introducirse en una serie de estadísticas y métricas relacionadas con las características principales de los distintos empleos publicados en la plataforma de LinkedIn; Por otro lado, el análisis de la demanda de empleos, representado por el cartel "busco empleado", a



través del cual el usuario puede explorar métricas relacionadas con aquellas cualidades de los empleos con mayor impacto o popularidad.

- **General empleados:** A esta solapa se accede mediante la selección del botón "Busco empleo" ubicado en la portada. Le permite al usuario seleccionar un rubro de empleo de interés y explorar rápidamente 3 métricas que podría considerar de relevancia: el promedio del "salario/hora" normalizado, el promedio de la cantidad de beneficios y el nivel de experiencia promedio de aquellos empleos ofrecidos por el rubro seleccionado. Además, las métricas dentro de las tarjetas responden a un código de color: Para el caso del salario/hora y el promedio de beneficios, aquellas métricas ubicadas por debajo del percentil 25 se muestran de color gris, aquellas ubicadas entre el percentil 25 y 75 se muestran de color celeste, y aquellas que se ubican por arriba del percentil 75 adoptan el color azul. Para el caso del promedio del nivel de experiencia el código de color se muestra a la inversa: aquellos empleos que requieren menor experiencia se muestran de color azul, los intermedios en celeste y los que exigen mayor experiencia en color gris.

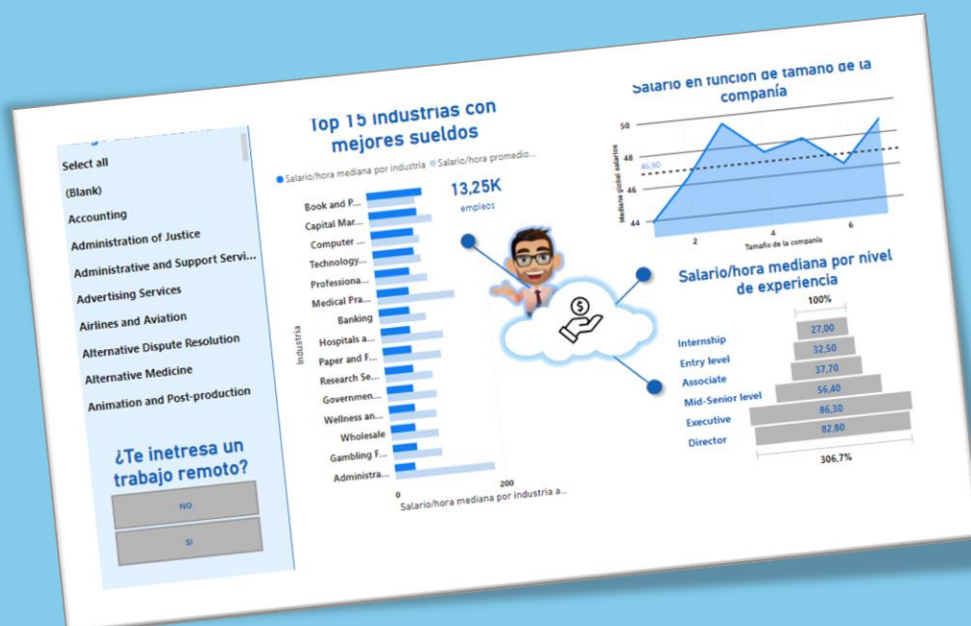


- **Insights salarios:** A esta solapa se accede a través del icono ubicado dentro de la tarjeta de salario/hora promedio en la solapa "General empleados". Le permite al usuario profundizar en algunos aspectos relacionados con el sueldo ofrecido por los distintos empleos. En particular se muestran las 15 industrias que ofrecen en promedio

los mejores salarios. Además, se muestra un gráfico en el que se puede observar la proporcionalidad directa que existe entre el tamaño de la empresa y el monto promedio del salario ofrecido. El tamaño de la empresa se encuentra codificado en una escala predeterminada por el dataset original que responde a la siguiente relación:

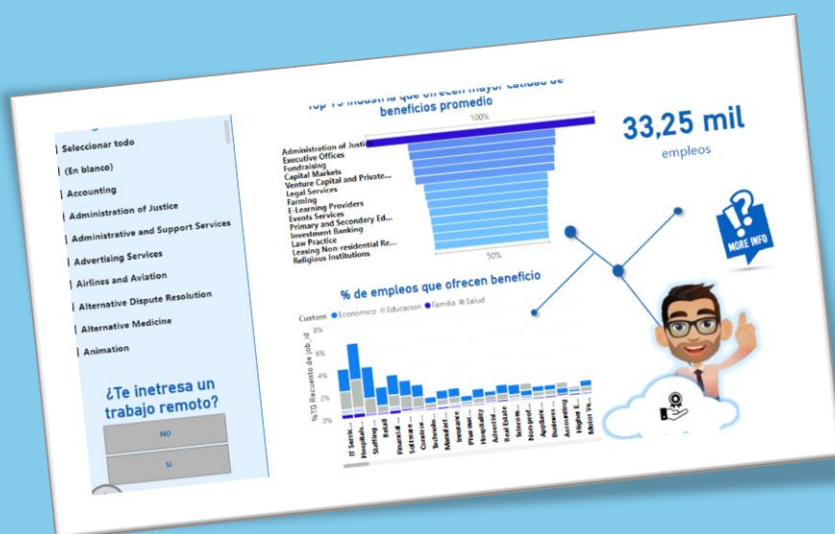
- 0: menos de 10 empleados
- 1: 11 a 50 empleados
- 2: 51 a 200 empleados
- 3: 201 a 500 empleados
- 4: 501 a 1,000 empleados
- 5: 1,001 a 5,000 empleados
- 6: 5,001 a 10,000 empleados
- 7: más de 10,000 empleados

Por último, en la parte inferior izquierda se observa un gráfico que permite establecer la mediana de salario/hora en función de la escala de experiencia. La solapa cuenta con 2 filtros, uno para seleccionar la industria de interés y otro para seleccionar trabajos remotos.

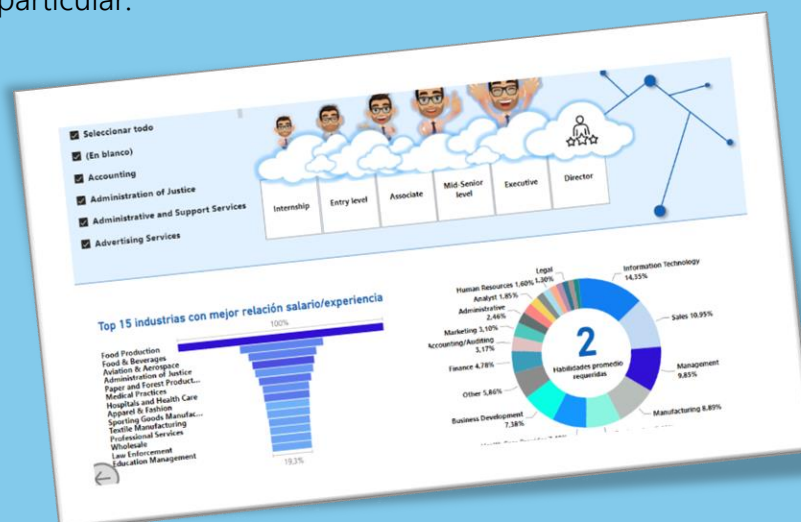


- **Insights Beneficios:** A esta solapa se accede a través del icono ubicado dentro de la tarjeta de beneficios promedio en la solapa "General empleados" y le permite al usuario explorar con mayor profundidad cuáles son los principales beneficios que

ofrece cada rubro. Para ello el usuario puede elegir un rubro de interés en el filtro ubicado a la izquierda y automáticamente se actualizará el gráfico de tipo tree map en el que se reflejará el porcentaje de empleos de dicho rubro que ofrece cada una de las 12 categorías de beneficios. El gráfico además tiene un código de colores que agrupa los distintos beneficios en 3: los beneficios relacionados con la familia tales como licencia por maternidad, licencia por paternidad o servicio de jardín maternal; beneficios relacionados con la salud del empleado tales como seguro médico, seguro por discapacidad, seguro de vista y plan dental; y otros beneficios de tipo económico tales como compensación de viáticos, financiamiento de estudios, préstamos estudiantiles o pensión.



- **Insights Experiencia:** A esta solapa se accede a través del icono ubicado dentro de la tarjeta de experiencia promedio en la solapa "General empleados" y le permite al usuario explorar con mayor profundidad cuáles son las principales habilidades que se requieren en distintos empleos con distintas exigencias de experiencia. El usuario puede hacer click en el filtro superior y seleccionar un nivel de experiencia de interés. Automáticamente en el gráfico de barras inferior se mostrará en orden decreciente el % de empleos que solicitan los distintos tipos de habilidades para esa categoría en particular.



- **Empleadores 1:** A esta solapa se accede a través de botón “Busco empleado” en la solapa “Elección”. La idea de esta parte del reporte es proveer al reclutador de empleo información relacionada al contenido de la publicación y su relación con la popularidad del empleo cuantificada a través de la relación de visitas o aplicaciones por hora. Se observa por ejemplo como aquellos salarios en los que se especifica el salario o la posibilidad de realizar el trabajo de forma remota tienen mayores visitas y aplicaciones por hora. A su vez esta solapa cuenta con dos filtros temporales que fueron creados usando la tabla calendario. El primero permite elegir el rango de meses en que se publicaron los empleos y el segundo permite cambiar el rango de horas que transcurrieron desde que se publicó el empleo hasta que se registró el número de visitas o aplicaciones. Estos filtros nos permiten comparar la popularidad de los empleos en distintos contextos temporales.



- **Empleadores 2:** A esta página del reporte se accede a través del botón en forma de flecha ubicada en la parte inferior derecha de la página “Empleadores 1”. La página responde a una lógica similar a la solapa anterior: cuenta con filtros temporales y gráficos en los que se relaciona la cantidad de visitas o aplicaciones con el tipo de formato con el que el usuario se postula al empleo. Además, cuenta con un filtro adicional que permite abrir por la categoría “remoto si” o “remoto no”. La página muestra una variable nueva en este caso que es la relación entre el número de visitas y el número de aplicaciones. Esta métrica nos da una idea de cuantas visitas se transforman en aplicaciones y por lo tanto se relaciona con el impacto que genera la publicación en aquella persona que la visita. Finalmente, la página cuenta con un gráfico de tipo esquema jerárquico que le da la posibilidad al reclutador de combinar las distintas variables categóricas exploradas anteriormente y calcular la mediana de aplicaciones por hora.



- Empleadores 3: A esta página del reporte se accede a través del botón en forma de flecha ubicada en la parte inferior derecha de la página "Empleadores 2" y se muestra un análisis un poco más profundo y específico en el que se relaciona la popularidad del empleo con la cantidad de palabras que se usaron en la columna "descripción". Tal como se puede observar, existe una relación inversa entre la cantidad de palabras y la popularidad del empleo. Además, se agregó un gráfico de análisis inteligente que permitió establecer que aquellas publicaciones de empleo más resumidas con un número de palabras menor a 846 generan más impacto en términos de aplicaciones por unidad de visitas. Esta página cuenta con los mismos filtros que la página anterior con la diferencia que se agregó un filtro que permite al usuario filtrar empleos en función de la cantidad de palabras de la descripción.



17. Futuras líneas

Este dataset que se trabajó en el proyecto tiene el potencial para obtener una gran cantidad de información y de esa forma respaldar decisiones tanto para aquella persona que se encuentra en la búsqueda de un nuevo empleo, como para aquellas empresas que se encuentren en un proceso activo de reclutamiento. Por ejemplo, se podría aumentar el número de registros tomando muestras todos los meses del año para hacer comparaciones temporales y describir posibles fluctuaciones de empleos, salarios, etc. Además, se podrían incluir estrategias de análisis más avanzadas con el uso de herramientas de inteligencia artificial para obtener insights más profundos. Como iniciativa, proponemos el uso de “Natural language processing” (NLP) para tratar de encontrar que características del lenguaje de las descripciones de empleos se asocian con una mayor popularidad.