

# RETO 2: Histórico Oferta Justjoin.it - Colisiones

---

## Objetivo

En este reto se pretende poner en práctica los conceptos aprendidos en clase acerca de los mapas no ordenados del módulo No. 2 y sus tablas de hash (Separate Chaining y Linear Probing) e integrarlo a los conceptos aprendidos en el módulo No. 1. Como lo son: (i) las listas, pilas y colas, (ii) los algoritmos de búsqueda y (iii) los algoritmos de ordenamiento.

Específicamente se pretende:

1. **Practicar** los conceptos sobre estructuras de tablas no ordenadas (tablas de Hash).
2. **Integrar** el uso de las estructuras lineales (Lista, pilas y colas) y las tablas no ordenadas.
3. **Practicar** los algoritmos de ordenamientos y búsquedas lineales.
4. **Aprender** a cargar y procesar en memoria datos en formato CSV.
5. **Utilizar** adecuadamente el administrador de versiones GIT y GitHub.
6. **Aprender** a trabajar en equipo.

## Fecha Límite de Entrega

El reto podrá entregarse hasta el martes **9 de abril**, 11:59 p.m. Ver las condiciones de entrega en la sección Entrega.

## Contexto

Just Join IT es una plataforma líder en la publicación de ofertas de trabajo especializada en el ámbito de la tecnología de la información (TI) y la programación en Polonia. Fundada para conectar profesionales de TI con oportunidades laborales relevantes, la plataforma se ha consolidado como recurso integral para el mercado laboral tecnológico.

La misión principal de Just Join It es facilitar el proceso de búsqueda y oferta de empleo para aquellos especializados en el campo de la tecnología, ofreciendo una amplia variedad de ofertas de trabajo permitiéndoles explorar oportunidades que se ajusten a sus habilidades y aspiraciones profesionales. Al mismo

tiempo, brindan a las empresas del sector de TI un medio efectivo para atraer y reclutar talento altamente calificado.

Desde su fundación en 2017, Just Join IT ha experimentado un notable crecimiento a lo largo de los años. Su creciente base de datos de empleos y su sólida reputación en el ámbito tecnológico han contribuido al procesamiento de más de 290.000 desarrolladores al mes en el mercado laboral polaco y europeo. La plataforma se ha convertido en un punto de referencia tanto para empresas como un “portal de contenidos Geek” que proporciona información valiosa a los entusiastas del sector.

Just Join IT los ha seleccionado para llevar a cabo un análisis del histórico de ofertas de trabajo publicadas en su plataforma a lo largo de los años. El primer grupo de datos abarcará del 2022-04 al 2023-09 para un total de 3.358.641 de ofertas de trabajo, las cuales incluyen un total de 15 millones de habilidades solicitadas y 7 millones de tipos de empleo en las ofertas. Los datos resultantes del análisis darán una herramienta fundamental para identificar el comportamiento pasado del mercado ofertas laborales en TI y predecir futuros comportamientos que permitirán que las empresas publiquen ofertas laborales más competitivas.

## Carga de Datos

Los datos del reto fueron tomados del proyecto **Kaggle** publicado por *Jakub Szafran* titulado “**JustJoinIT job offers data (2021.10 – 2023-09)**” que recopila la información de las ofertas de más de 2 años.<sup>1</sup> Los datos útiles para el proyecto están preparados y disponibles para los estudiantes en el aula unificada en Bloque Neón (BrightSpace).

En particular, la información está conformada por cuatro archivos CSV en formato UTF-8 que se describen a continuación:

- El archivo **jobs.csv**, contiene la información relevante a las ofertas de trabajo publicadas. Como se ve en **Tabla 1**, este archivo incluye identificación de la oferta, título de la oferta, dirección de la oferta, tipo de lugar de trabajo, nombre la organización, dirección web de la empresa, fecha de publicación de la oferta y si está disponible a contratar ucranianos.
- El archivo **skills.csv**, contiene la información relevante sobre las habilidades solicitadas en las ofertas de trabajo. Como se ve en **Tabla 2**, este archivo incluye el id de la oferta de trabajo, nombre de la habilidad y nivel mínimo requerido para la oferta. Una oferta de trabajo puede tener varias habilidades solicitadas.
- El archivo **employment-types.csv**, contiene la información sobre los tipos de contratación disponibles en la oferta de trabajo. Como se ve en la **Tabla 3**, este archivo incluye el id de la oferta de trabajo, el tipo de contrato, la divisa de la oferta, salario mínimo y salario máximo de la oferta. Una oferta de trabajo puede tener una o varios tipos de contratación
- El archivo **multilocation.csv**, contiene la información sobre las otras ubicaciones de la empresa que publica la oferta de trabajo. Como se ve en la **Tabla 4**, este archivo incluye la ciudad de las sedes, dirección de las sedes y el id de la oferta de trabajo.

↓ un trabajo puede tener muchos

---

<sup>1</sup> JustJoinIT job offers data (2021.10 – 2023-09), Kaggle, URL: [./www.kaggle.com/datasets/jszafranqb/justjoinit-job-offers-data-2021-10-2023-09](https://www.kaggle.com/datasets/jszafranqb/justjoinit-job-offers-data-2021-10-2023-09)

NOMBRE DE LA COLUMNA	DESCRIPCIÓN
<i>title</i>	Título de publicación de la oferta de trabajo.
<i>street</i>	Calle de la ubicación de la oferta de trabajo. Algunas publicaciones no definen una calle, por lo cual puede haber campos con el valor “NOT DEFINED”
<i>city</i>	Nombre de la ciudad de la oferta de trabajo.
<i>country_code</i>	Código del país de la oferta de trabajo.
<i>address_text</i>	Dirección completa de la sede de la oferta de trabajo. En los casos que el valor “Street” no esté definido, la dirección completa también tendrá un “NOT DEFINED” en la posición de la calle.
<i>marker_icon</i>	Tipo de icono usado para identificar la oferta en un mapa
<i>workplace_type</i>	Tipo del lugar de trabajo de la oferta de trabajo. Puede tomar los valores “office”, “remote” o “partly_remote”
<i>company_name</i>	Nombre de la empresa que publica la oferta de trabajo.
<i>company_url</i>	Dirección url de la empresa que publica la oferta de trabajo.
<i>company_size</i>	Número de empleados que tiene la empresa que publicó la oferta de trabajo. Este valor es numérico, pero puede haber valores “Undefined” en los casos que no se definió el tamaño de la empresa en la publicación.
<i>experience_level</i>	Nivel de experticia buscado en la oferta de trabajo. Puede tomar los valores “junior”, “mid” o “senior”
<i>published_at</i>	Fecha de la publicación de la oferta de trabajo (con formato “yyyy-mm-ddThh:mm:ss”)
<i>remote_interview</i>	Indicador si la entrevista es remota. Esta columna es True o False.
<i>open_to_hire_ukrainians</i>	Indicador que la oferta de trabajo está abierta a contratar ucranianos. Esta columna es True o False.
<i>id</i>	Identificador de la oferta de trabajo. Este id es usado para identificar las habilidades, tipos de contratación y multi-ubicación de los otros archivos.
<i>display_offer</i>	Indicador que la oferta se encuentra visible en la plataforma web. Esta columna es True o False.

Tabla 1. Nombres y descripción de las propiedades de datos de las ofertas de trabajo.

NOMBRE DE LA COLUMNA	DESCRIPCIÓN
<i>name</i>	Nombre de la habilidad solicitada en la oferta laboral
<i>level</i>	Nivel mínimo solicitado para la habilidad
<i>id</i>	Identificador de la oferta de trabajo

Tabla 2. Nombres y descripción de las habilidades solicitadas en las ofertas laborales.

NOMBRE DE LA COLUMNA	DESCRIPCIÓN
<i>type</i>	Tipo de contratación ofrecida. Puede tomar los valores “permanent” o “b2b”.
<i>id</i>	Identificador de la oferta de trabajo
<i>currency_salary</i>	Divisa del salario ofertado.
<i>salary_from</i>	Valor mínimo ofertado (en la divisa de la oferta). Este valor puede ser vacío.
<i>Salary_to</i>	Valor máximo ofertado (en la divisa de la oferta). Este valor puede ser vacío.

Tabla 3. Nombres y descripción de las propiedades de datos tipos de contratación.

NOMBRE DE LA COLUMNA	DESCRIPCIÓN
<i>city</i>	Nombre de la ciudad donde se encuentra la sede.
<i>street</i>	Calle en la que se encuentra la sede de la empresa.
<i>id</i>	Identificador de la oferta

Tabla 4. Nombres y descripción de las propiedades de datos tipos de las sedes de las ofertas.

Para evitar problemas de buffer en la lectura de los archivos se recomienda aumentar el tamaño de los campos de lectura de la librería **Python CSV** al máximo posible para el sistema con el siguiente comando en la librería CSV en el **controller.py** del Reto.

```
import csv
...
csv.field_size_limit(2147483647)
```

En algunos casos experimentales puede que Python y el IDE declaren que se alcanzó el límite de recursión con un mensaje “**RecursionError: maximum recursion depth exceeded in comparison**”, en este caso se recomienda actualizar en el **view.py** este límite con las siguientes líneas de código:

```
import sys
...
default_limit = 1000
sys.setrecursionlimit(default_limit*10)
```

## Trabajo Propuesto

### Parte 1: Configuración Repositorio

Complete los siguientes pasos para configurar su repositorio de trabajo:

- p.1.** Cree en GitHub un repositorio basado en la plantilla propuesta para el reto, el cual se encuentra en el URL: <https://github.com/ISIS1225DEVS/Reto2-Template>
- p.2.** Renombre el repositorio de su reto con el formato **Reto2-G<<Número del grupo>>** ej.: **Reto2-G01** para el grupo 1 de la sección.
- p.3.** Edite el **README** del repositorio e incluya los nombres completos, correo Uniandes y códigos de los miembros del equipo de trabajo.
- p.4.** Realice el procedimiento según lo aprendido en clase para clonar el repositorio en su máquina local y sincronizarlo con su repositorio en GitHub.
- p.5.** Descargue los datos desde la sección unificada del curso y cópielos en la carpeta **data** del repositorio local.

### Parte 2: Carga de Datos

En la sección de **Reto 2**, de la sección unificada en Bloque Neón, encontrarán los datos oficiales del proyecto. El ZIP contiene varios archivos con los prefijos **small-**, **medium-** y **large-**. Estos son archivos con diferente número de registros. (ej.: el archivo **small-jobs.csv** contiene una tercera parte de los datos con 67000 registros y el archivo **large-Jobs.csv** contiene la totalidad de los datos con 203000 registros). Esto facilita la implementación y pruebas en computadores con memoria RAM y procesadores reducidos.

Para responder los requerimientos deberán cargar la información de los archivos entregados; recuerde que solo se permite leer una vez la información de cada archivo y que las pruebas finales sobre sus algoritmos serán sobre los archivos **large.csv**.

Al completar la carga de datos de las ofertas de trabajo (**jobs-utf8.csv**) se debe reportar los siguiente:

- El total de ofertas de trabajo publicadas cargadas.
- La información de las tres primeras y tres últimas ofertas de trabajo publicadas ordenadas por la fecha de publicación. Cada elemento debe contener los siguientes detalles:
  - Fecha de publicación.
  - Título de la oferta
  - Nombre de la empresa que publica
  - Nivel de experticia de la oferta
  - País de la oferta
  - Ciudad de la oferta

#### **Recomendaciones:**

- Todos los órdenes cronológicos de las ofertas de trabajo se definen desde el más reciente al más antiguo registrado.
- Se recomienda utilizar las librerías nativas de Python de datetime<sup>2</sup> para manejar el procesamiento y comparaciones de los datos temporales dentro de los archivos.
- Se recomienda utilizar librerías por extensión de Python como tabulate<sup>3</sup> para imprimir adecuadamente los resultados como se muestran en los ejemplos.
- En caso de que no exista algún valor dentro de los campos solicitados, completarlos con el valor **“Desconocido”** o **“Unknown”**.

---

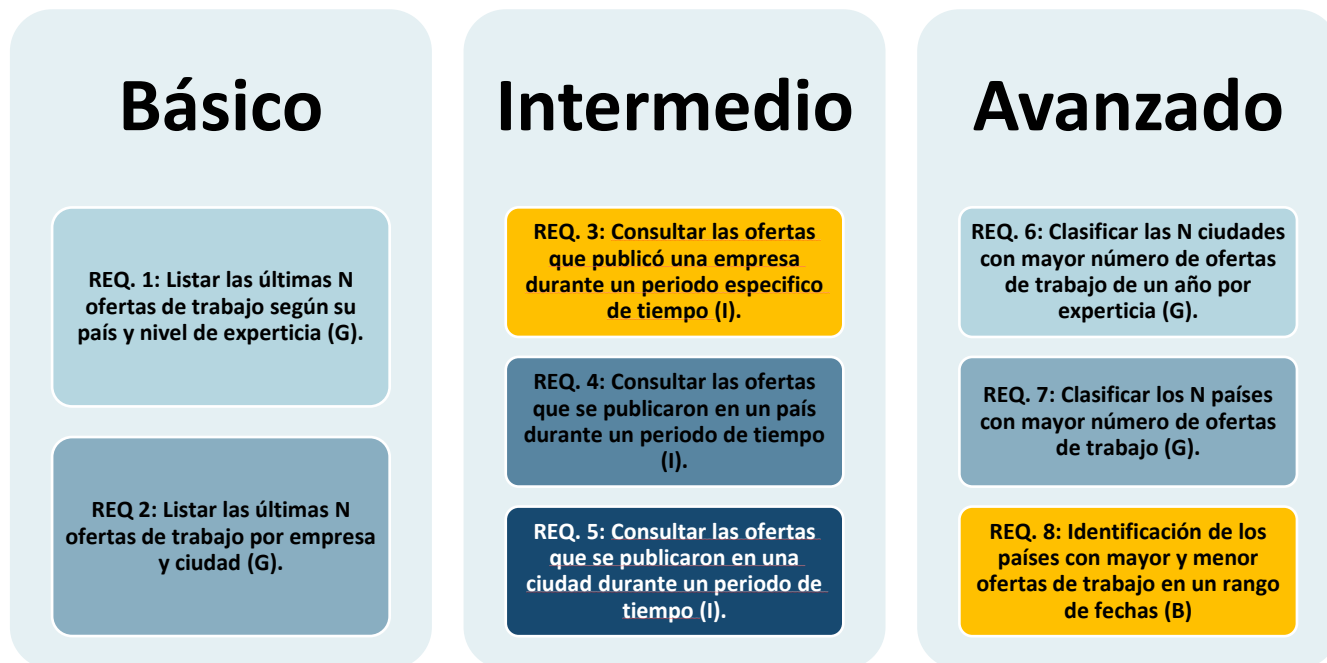
<sup>2</sup> Librería Python datetime, Python URL: <https://docs.python.org/3/library/datetime.html>

<sup>3</sup> librería Python tabulate, PyPi URL: <https://pypi.org/project/tabulate/>

### Parte 3: Desarrollo de los Requerimientos

Para este reto se ha identificado **ocho (8) requerimientos: siete (7) obligatorios y un (1) bono**. Divididos según dificultad en básico, intermedio y avanzado. Adicionalmente, es importante resaltar que **tres (3)** de estos requerimientos se deben desarrollar de forma **individual** y los **cinco (5)** restantes son **grupales**.

El resumen de los requerimientos se muestra en la siguiente tabla y se explican detalladamente en la siguiente sección.



**NOTA:** Si el equipo este conformado solamente por **dos (2)** integrantes deberán resolver solamente **cinco (5) requerimientos obligatorios: uno (1)** de los requerimientos **básicos**, **dos (2)** de los requerimientos **intermedios**, y los **dos (2)** primeros requerimientos **avanzados**. Opcionalmente, el último requerimiento **avanzado (Bono)**.

## Requerimiento No. 1 (Grupal): Listar las últimas N ofertas de trabajo según su país y nivel de experticia

Como analista de datos **quiero** listar las últimas **N** ofertas de trabajo (v.gr. las más recientes) ofrecidas en un país filtrando por el nivel de experticia del puesto (junior, mid, senior).

Los **parámetros de entrada** de este requerimiento son:

- El número (N) de ofertas a listar (ej.: 3, 5, 10 o 20).
- Código del país (ej.: PL, CO, ES, etc).
- Nivel de experticia de las ofertas a consultar (**junior, mid o senior**).

La **respuesta esperada** debe contener:

- El total de ofertas de trabajo ofrecidas según el país.
- El total de ofertas de trabajo ofrecidas según la condición (junior, mid o senior).
- Para cada una de las ofertas de la consulta debe presentar la siguiente información:
  - Fecha de publicación de la oferta
  - Título de la oferta
  - Nombre de la empresa de la oferta
  - Nivel de experticia de la oferta (es el mismo del filtro)
  - País de la empresa de la oferta
  - Ciudad de la empresa de la oferta
  - Tamaño de la empresa de la oferta
  - Tipo de ubicación de trabajo (remote, partialy, remote, office)
  - Disponible a contratar ucranianos (Verdadero o Falso)

### Recomendaciones:

- Integre el uso de mapas a las listas según lo considere pertinente.
- En caso de obtener una lista de ofertas demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.

## Requerimiento No. 2 (Grupal): Listar las últimas N ofertas de trabajo por empresa y ciudad

Como analista de datos **quiero** conocer las últimas **N** ofertas (v.gr. las más recientes) de una empresa dado su nombre y la ciudad.

Los **parámetros de entrada** de este requerimiento son:

- El número (N) de ofertas a listar (ej.: 3, 5, 10 o 20).
- Nombre completo de la empresa a consultar.
- Ciudad de la oferta (nombre de la ciudad)

La **respuesta esperada** debe contener:

- El total de ofertas ofrecida por la empresa y ciudad.
- Para cada una de las ofertas en la consulta debe presentar la siguiente información:
  - Fecha de publicación de la oferta.
  - País de la oferta
  - Ciudad de la oferta (es la misma del filtro)
  - Nombre de la empresa de la oferta (es la misma del filtro)
  - Título de la oferta.
  - Nivel de experticia de la oferta
  - Formato de aplicación de la oferta
  - Tipo de trabajo, si es remoto o no

### Recomendaciones:

- Integre el uso de mapas a las listas según lo considere pertinente.
- En caso de obtener una lista de ofertas demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.



## Requerimiento No. 3 (Individual): Consultar las ofertas que publicó una empresa durante un periodo específico de tiempo

**Como** analista de datos **quiero** consultar las ofertas de trabajo publicadas por una empresa en un rango de fechas dado.

Los **parámetros de entrada** de este requerimiento son:

- Nombre de la empresa.
- La fecha inicial del periodo a consultar (con formato "%Y-%m-%d").
- La fecha final del periodo a consultar (con formato "%Y-%m-%d").

La **respuesta esperada** debe contener:

- Número total de ofertas.
- Número total de ofertas con experticia junior.
- Número total de ofertas con experticia mid.
- Número total de ofertas con experticia senior.
- El listado de ofertas de la empresa ordenados cronológicamente por fecha y país (v.gr. Para dos ofertas con la misma fecha, el orden lo decide el país de forma alfabética). Donde para cada uno de los elementos resultantes contendrá la siguiente información:
  - Fecha de la oferta.
  - Título de la oferta.
  - Nivel de experticia requerido
  - Ciudad de la empresa de la oferta
  - País de la empresa de la oferta
  - Tamaño de la empresa de la oferta
  - Tipo de lugar de trabajo de la oferta.
  - Disponible a contratar ucranianos (Verdadero o Falso).

### Recomendaciones:

- Integre el uso de mapas a las listas según lo considere pertinente.
- En caso de obtener una lista de ofertas demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.

## Requerimiento No. 4 (Individual): Consultar las ofertas que se publicaron en un país durante un periodo de tiempo

Como analista de datos **deseo** consultar las ofertas de trabajo publicadas en un país dado su código y un rango de fechas.

Los **parámetros de entrada** de este requerimiento son:

- Código del país para la consulta (ej.: PL, CO, ES, etc).
- La fecha inicial del periodo a consultar (con formato "%Y-%m-%d").
- La fecha final del periodo a consultar (con formato "%Y-%m-%d").

La **respuesta esperada** debe contener:

- El total de ofertas en el país en el periodo de consulta. ✓
- El total de empresas que publicaron al menos una oferta en el país de consulta. ✓
- Número total de ciudades del país de consulta en las que se publicaron ofertas. ✓
- Ciudad del país de consulta con mayor número de ofertas y su conteo ✓
- Ciudad del país de consulta con menor número de ofertas (al menos una) y su conteo ✓
- El listado de ofertas publicadas ordenados cronológicamente por fecha y nombre de la empresa (v.gr. Para dos ofertas con la misma fecha, el orden lo decide la empresa de forma alfabética). Donde cada uno de los elementos resultantes contendrá la siguiente información: ✓
  - Fecha de publicación de la oferta.
  - Título de la oferta.
  - Nivel de experticia requerido
  - Nombre de la empresa de la oferta
  - Ciudad de la empresa de la oferta
  - Tipo de lugar de trabajo de la oferta.
  - Tipo de trabajo, si es remoto o no
  - Disponible a contratar ucranianos.

### Recomendaciones:

- Integre el uso de mapas a las listas según lo considere pertinente.
- En caso de obtener una lista de ofertas demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.
- Si una sola empresa cumple con la condición, presente esta empresa como la empresa con mayor y menor número de ofertas.

## Requerimiento No. 5 (Individual): Consultar las ofertas que se publicaron en una ciudad durante un periodo de tiempo

Como analista de datos **quiero** consultar las ofertas de trabajo publicadas en una ciudad dado un rango de fechas

Los **parámetros de entrada** de este requerimiento son:

- Nombre de la ciudad.
- La fecha inicial del periodo a consultar (con formato "%Y-%m-%d").
- La fecha final del periodo a consultar (con formato "%Y-%m-%d").

La **respuesta esperada** debe contener:

- El total de ofertas publicadas en la ciudad en el periodo de consulta.
- El total de empresas que publicaron por lo menos una oferta en la ciudad de consulta.
- Empresa con mayor número de ofertas y su conteo
- Empresa con menor número de ofertas (al menos una) y su conteo
- El listado de ofertas publicadas ordenadas cronológicamente por fecha y nombre de la empresa (v.gr. Para dos ofertas con la misma fecha, el orden lo decide la empresa de forma alfabética). Cada uno de los elementos debe presentar la siguiente información:
  - Fecha de publicación de la oferta
  - Título de la oferta
  - Nombre de la empresa de la oferta
  - Tipo de lugar de trabajo de la oferta
  - Tamaño de la empresa de la oferta
  - Tipo de lugar de trabajo de la oferta

### Recomendaciones:

- Integre el uso de mapas a las listas según lo considere pertinente.
- En caso de obtener una lista de ofertas demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.
- Si una sola empresa cumple con la condición, presente esta empresa como la empresa con mayor y menor número de ofertas.

## Requerimiento No. 6 (Grupal): Clasificar las N ciudades con mayor número de ofertas de trabajo de un año por experticia

Como analista de datos **deseo** clasificar las **N** ciudades con mayor cantidad de ofertas de trabajo dado el año de publicación y nivel de experticia de la oferta.

Los **parámetros de entrada** de este requerimiento son:

- El número (N) de ciudades para consulta (ej.: 3, 5, 10 o 20).
- Nivel de experticia de las ofertas de interés (**junior, mid, senior, indiferente**)
- El año de la consulta (con formato "%Y").

La **respuesta esperada** debe contener:

- El total de ciudades que cumplen con las condiciones de la consulta (valor menor o igual a N)
- El total de empresas que cumplen con las condiciones de la consulta
- El total de ofertas publicadas que cumplen con las condiciones de la consulta
- Nombre de la ciudad con mayor cantidad de ofertas de empleos y su conteo
- Nombre de la ciudad con menor cantidad de ofertas de empleos y su conteo
- El listado de las ciudades ordenadas por el número de ofertas publicadas y nombre de la ciudad (siguiendo ese orden de prioridad). El orden de las ciudades es de mayor a menor cantidad de ofertas de trabajo. Donde cada una de las ciudades resultantes contendrá la siguiente información:
  - Nombre de la ciudad.
  - País de la ciudad.
  - El total de ofertas hechas en la ciudad.
  - Promedio de salario ofertado en la ciudad
  - Número de empresas que publicaron por lo menos una oferta en la ciudad
  - Nombre de la empresa con mayor número de ofertas y su conteo
  - La información de la mejor oferta por salario en la ciudad (considerando el tope más alto dado en una oferta). Se consideran solo ofertas que tengan información de salario.
  - La información de la peor oferta por salario en la ciudad (considerando el tope más bajo dado en una oferta). Se consideran solo ofertas que tengan información de salario.

### Recomendaciones:

- Integre el uso de mapas a las listas según lo considere pertinente.
- Al ingresar el parámetro **indiferente** como nivel de experticia se tiene en cuenta todos los niveles de las ofertas de trabajo.
- En caso de obtener una lista de ofertas demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.
- Si una sola empresa cumple con la condición, presente esta empresa como la empresa con mayor y menor número de ofertas.

- Si al ordenar por número de ofertas dos o más ciudades presentan el mismo número, ordene las ciudades por el promedio de salario de las ofertas hechas.

## Requerimiento No. 7 (Grupal): Clasificar los N países con mayor número de ofertas de trabajo.

**Como** analista de datos, quiero identificar las habilidades solicitadas en los niveles de experticia publicados en los primeros N países con mayor número de ofertas de trabajo publicas dado el año y el mes.

Los **parámetros de entrada** de este requerimiento son:

- El número (N) de países para consulta (ej.: 3, 5, 10 o 20).
- El año de la consulta (con formato "%Y").
- El mes de la consulta (con formato "%m").

La **respuesta esperada** debe contener:

- El total de ofertas de empleo.
- Número de ciudades donde se ofertó en los países resultantes de la consulta.
- Nombre del país con mayor cantidad de ofertas y su conteo
- Nombre de la ciudad con mayor cantidad de ofertas y su conteo
- Para el conjunto de las ofertas de trabajo en los países resultantes de la consulta, por cada uno de los tres niveles de experticia (junior, mid y senior) calcule y presente la siguiente información:
  - Conteo de habilidades diferentes solicitadas en ofertas de trabajo
  - Nombre de la habilidad más solicitada y su conteo en ofertas de trabajo
  - Nombre de la habilidad menos solicitada y su conteo en ofertas de trabajo
  - Nivel mínimo promedio de las habilidades
  - Conteo de empresas que publicaron una oferta con este nivel de experticia
  - Nombre de la empresa con mayor número de ofertas y su conteo
  - Nombre de la empresa con menor número de ofertas (al menos una) y su conteo
  - Número de empresas que publicaron una oferta en este nivel de experticia que tienen una o más sedes

### Recomendaciones:

- Integre el uso de mapas a las listas según lo considere pertinente.
- En caso de obtener una lista de ofertas demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.
- Si solo una empresa cumple con la condición, presente esta empresa como la empresa con mayor y menor número de ofertas.
- En caso de que al ordenar por número de ofertas dos o más empresas presentan el mismo número de ofertas, sobre estas empresas realice un orden alfabético ascendente.

## Requerimiento No. 8 (Bono Grupal): Identificación de los países con mayor y menor ofertas de trabajo en un rango de fechas

**Como** analista de datos quiero identificar los países con mayor y menor ofertas salariales para una experticia en un periodo de tiempo dada la divisa de publicación. Para este análisis, se tendrá en cuenta solo ofertas que tengan información de rango salarial publicado. Para cada oferta laboral se tomará como referencia de salario el promedio de su rango salarial. Este salario de referencia servirá para el cálculo del promedio salarial de las ofertas de trabajo para un nivel de experticia.

Los **parámetros de entrada** de este requerimiento son:

- Nivel de experticia a consultar (**junior, mid, senior o indiferente**).
- Divisa de la consulta (ej: usd, eur, etc).
- La fecha inicial del periodo a consultar (con formato "%Y-%m-%d").
- La fecha final del periodo a consultar (con formato "%Y-%m-%d").

La **respuesta esperada** se divide en dos partes. La primera, enfocada a las estadísticas generales de las ofertas de trabajo publicadas resultado de la consulta. La segunda, enfocada en los países con mayor y menor oferta salarial.

La **primera parte** debe contener la siguiente información:

- Número de empresas que publicaron por lo menos una oferta
- Número total de ofertas de empleo que cumplen la consulta
- Número de países que cumplan la consulta
- Número de ciudades que cumplan la consulta
- Número de ofertas publicadas con rango salarial
- Número de ofertas publicadas con valor fijo de salario
- Número de ofertas publicadas sin salario
- Listado de los países ordenados del mayor al menor promedio de oferta salarial. Donde cada uno de los países resultantes contendrá la siguiente información:
  - Nombre del país
  - Promedio de oferta salarial (para las ofertas publicadas con rango salarial).
  - Número de empresas que publicaron
  - Número de ofertas publicadas
  - Número de ofertas publicadas con rango salarial
  - Número de habilidades promedio solicitadas por oferta en el país

La **segunda parte** se concentra en el país con mayor y menor oferta salarial en el periodo y nivel de experticia de consulta. Del país con mayor (primero) y menor (último) promedio de salario presente la siguiente información:

- Código del país.
- Total de ofertas de empleo.
- Promedio de salario ofertado.
- Número de ciudades donde se ofertaron.

- Número de empresas que publicaron ofertas en el país.
- Valor del mayor salario ofertado.
- Valor del menor salario ofertado.
- Número de habilidades promedio solicitadas por oferta en el país.

#### **Recomendaciones:**

- Integre el uso de mapas a las listas según lo considere pertinente.
- Al ingresar el parámetro **indiferente** como nivel de experticia se tiene en cuenta todos los niveles de las ofertas de trabajo.
- En caso de obtener una lista de países demasiado extensa (más de 10 elementos) presente solo los 5 primeros y últimos elementos de la lista con los detalles solicitados.
- Antes de empezar el desarrollo del requerimiento analice los archivos e identifique posibles valores para datos como código de país, nivel de experticia, nombre de la empresa y fechas de consulta.
- Para la conversión de divisas cree alguna función que se encargue de recibir como parámetro el valor a convertir, la divisa del valor y la divisa común a la que quiere hacer la conversión. Esta función **no** tiene como obligación consultar en tiempo real las tasas de cambio entre divisas, por lo que las tasas pueden ser valores predefinidos durante la creación de la función. Se propone como divisa común de conversión el US dólar por ser la divisa en la cual normalmente se hacen las transacciones internacionales.
- Si una sola empresa cumple con la condición, presente esta empresa como la empresa con mayor y menor número de ofertas.
- En caso de que al ordenar por número de ofertas dos o más empresas presentan el mismo número de ofertas, sobre estas empresas realice un orden alfabético ascendente por su nombre.



## Parte 4: Análisis de resultados

Dentro del proyecto debe incluir, en la carpeta **Docs**, un documento en formato **PDF** donde se evidencie el análisis de complejidad y las pruebas de tiempos de ejecución para cada requerimiento. Se sugiere que el documento tenga la siguiente distribución del contenido:

- Nombres, código y correo Uniandes de los integrantes del grupo.
- Para los requerimientos individuales se debe indicar que estudiante del equipo lo realizó.
- Análisis de complejidad de los requerimientos en Notación O.
- Pruebas de tiempos de ejecución para cada uno de los requerimientos. En estas pruebas se deben incluir:
  - Las tablas de tiempos de ejecución registrados.
  - Las gráficas comparativas de los experimentos.
  - Un análisis de resultados comparándolo los resultados obtenidos con el análisis de complejidad realizado.

### Recomendaciones:

- Tomen como guía las herramientas, metodología y análisis realizados en los laboratorios; en especial el laboratorio 4, 5 y 7
- Ejecute las pruebas de los requerimientos siempre con los mismos parámetros de entrada y con los archivos más grandes que pueda procesar su computador.
- Ejecute las pruebas de los requerimientos con la configuración optima del catálogo (LINKED\_LIST, ARRAY\_LIST, PROBING, CHAINING) y los algoritmos de ordenamiento (Shell, Insertion, Merge, etc.).

## Entrega

Para realizar la entrega del reto deben:

- 1) Agregar los usuarios de los monitores y profesores del curso a su organización de GitHub para hacer la entrega adecuada de la actividad,
- 2) Dar permisos adecuados repositorio a los monitores y al profesor, de lo contrario el reto **NO** podrá ser calificado,
- 3) Asegurarse que la visibilidad del repositorio entregado sea privada y que solo pueda accederse con los permisos configurados para los integrantes del grupo,
- 4) Enviar el enlace de GitHub en la actividad correspondiente dentro de Bloque Neón Uniandes (BrightSpace) de su sección.
- 5) Incluir en el repositorio GIT todo el material, código y documentos solicitados durante la actividad.

**IMPORTANTE:** Recuerde que solo se calificará el material hasta el último **COMMIT** realizado previo a la Fecha/Hora Límite de Entrega indicada al inicio de este enunciado.