

## Prueba data Mailtrack

Tenemos una base de datos que contiene:

- Una tabla de referencia de usuarios registrados (users)
  - id (bigint): clave primaria del usuario
  - email (text): dirección de correo del usuario
  - country\_id (bigint): referencia a countries.id, representa el país del usuario
- Una tabla de conexiones de usuarios (user\_connections) que graba una fila cada vez que el usuario se conecta en la plataforma
  - timestamp (timestamp): timestamp UTC de la conexión del usuario
  - user\_id (bigint): referencia a users.id
- Una tabla de referencia de países (countries)
  - id (bigint): clave primaria del país
  - name (text): nombre del país

**Consideramos que un “usuario activo” es un usuario que se conecta al menos 2 veces en la plataforma dentro del mismo mes calendario.**

### Preguntas SQL

Escribir consultas SQL para obtener la información siguiente (puedes cargar los datos artificiales de los 3 ficheros CSV adjuntos en una base de datos para probar las consultas, recomendamos usar la imagen docker de PostgreSQL). Los timestamp están en el formato YYYY-MM-dd HH:MI:SS y van de 2021-01-01 00:00:00 a 2022-01-01 00:00:00.

Para cada pregunta generar un fichero SQL con la consulta.

**1. Número de usuarios activos por mes calendario en el año 2021, ordenado por mes descendente. Devolver 0 si un mes no tiene ningún usuario activo.**

Ejemplo de formato de salida esperado

Mes	Número usuarios activos
2021-12	984
2021-11	1020
2021-10	0
...	...
2021-01	894

## 2. El dominio email con más usuarios activos en abril 2021

Puedes considerar que el dominio es todo lo que viene después del carácter '@' en el email del usuario.

Ejemplo de formato de salida esperado

Dominio	Número usuarios activos
gmail.com	500

## 3. Para cada mes calendario del año 2021, generar una cadena agregada de nombres de países que no han tenido ningún usuario activo, ordenado por mes descendente

Mes	Lista países
2021-12	Argentina,Brazil,Colombia
2021-11	Alemania,España,Reino Unido
2021-10	China,Japón
...	...

## 4. Variación mes a mes del número de usuarios activos por país, ordenado por país ascendente, mes descendente. Incluir los meses/países que no han tenido ningún usuario activo.

Ej. si hemos tenido 1000 usuarios activos en diciembre de 2021 y 500 en noviembre de 2021 la variación sería de +100% (representado por 100). Si pasamos de 1000 a 500 en el mes siguiente la variación sería de -50% (representado por -50).

Mes	País	Número de usuarios activos	Variación
...	...	...	...
2021-03	España	800	-20.0
2021-02	España	1000	500.0
2021-01	España	200	NULL
...	...	...	...
2021-05	Reino Unido	200	100.0
2021-04	Reino Unido	100	NULL
2021-03	Reino Unido	0	NULL
...	...	...	...

## 5. Dirección de correo de los 10 últimos usuarios conectados por país, ordenados por país ascendente, timestamp de la conexión descendente

Correo	País	Timestamp de conexión
vbarnes@gmail.com	Estados Unidos	2022-05-15 22:01:01.000
gmccall@gmail.com	Estados Unidos	2022-05-15 21:00:32.000
epeterson@gmail.com	Estados Unidos	2022-05-15 18:00:32.000
...	...	...
juliehahn@hotmail.com	Reino Unido	2022-05-15 22:03:32.000
nhall@hotmail.com	Reino Unido	2022-05-15 21:03:32.000
...	...	...

## Pregunta Python

Escribir un programa Python 3 que lee los 3 ficheros de datos proporcionados y generar un fichero CSV conteniendo el resultado para cada una de las preguntas previas.

Puedes asumir que el programa será ejecutado a partir de la raíz del proyecto con:

```
python3 -m prueba_data.py
```

La salida esperada es un conjunto de ficheros CSV:

```
./resultados/pregunta1.csv
./resultados/pregunta2.csv
...
```

El módulo de Python tendrá que contener un método por pregunta para poder identificar fácilmente el código de cada pregunta, por ejemplo:

```
def pregunta1():
    # TODO implement me
    pass
```

En cada de estos métodos, leer y transformar los CSV necesarios y generar el fichero CSV de salida. Por supuesto se pueden añadir más métodos.

El proyecto tendrá que tener al menos la organización de carpetas/ficheros siguiente:

```
./prueba_data.py
./datos/users.csv
./datos/user_connections.csv
./datos/countries.csv
```

Puedes usar cualquier librería de terceros.

Se evaluará:

- la exactitud del código ejecutado

- la calidad del código Python (DRY, limpieza, testing de métodos o clases creadas etc.)
- la estructura del proyecto Python (organización de ficheros, facilidad de instalación de dependencias usadas, documentación)