Trabajo Práctico Final

El siguiente trabajo práctico se realizará de manera individual.

El trabajo evaluará los conocimientos adquiridos en el Módulo Foundations.

Las tecnologías a utilizar son **Docker, Bases de Datos Relacionales y Python 3.6+.**

Se debe enviar un email con la entrega a jpamplie@itba.edu.ar antes de la fecha estipulada.
El email con la entrega debe tener en el subject **[TP CDE-FOUNDATIONS]**

La fecha de entrega límite es el lunes 23 de octubre de 2020 a las 00hs.

1. Elegir un dataset fácilmente accesible desde internet que sea de interés para el alumno.
   Crear una breve descripción de las respuestas de negocio que se podrían responder teniendo esos datos en una base consultable.

2. Crear un container con una base de datos a elección (recomendamos usar las imágenes de PostgreSQL o MySQL disponibles en Docker Hub) y exponer el puerto estándar de esa base de datos para recibir conexiones desde el exterior.

3. Crear un script de bash que cree la estructura de tablas de la base.

4. Popular la base de datos con un dataset a elección disponible en Internet. Para eso es necesario crear un script de Python que lea el archivo con los datos crudos, los procese y luego cargue en la base. Esto será el paso de ETL (extract-transform-load) que dejará la data lista para ser consultada.
   Este script debe correr dentro de una imagen de Docker mediante el comando `docker run`.
   La imagen de Docker generada no debe contener los datos crudos que se utilizarían para cargar la base.
   Para pasar los archivos se puede montar un volumen (argumento `-v` de `docker run`) o bien bajarlos directamente desde internet usando alguna librería de Python (como `requests`).

5. Escribir un script de Python que realice al menos 5 consultas SQL que puedan agregar valor al negocio y muestre por pantalla un reporte con los resultados.
   Este script de reporting debe correrse mediante una imagen de Docker con `docker run` del mismo modo que el script anterior de ETL.

Entregables:

- Dockerfiles para generar las imágenes de ETL y generación de reportes.
- Scripts de Bash y Python relevantes.
- Archivo README.md con:
    - Instrucciones para
        - Crear la base.
        - Crear las imágenes usando los Dockerfiles.
        - Utilizar los scripts creados (mencionar argumentos pertinentes).
    - Descripción del negocio al que responde esa base de datos.

Se recomienda usar un repositorio Git para mantener versionado el código y disponibilizar el resultado del trabajo para que sea evaluado fácilmente.

Fuentes de datos abiertas sugeridas:

- https://catalog.data.gov/dataset
- https://datasetsearch.research.google.com/
- https://www.kaggle.com/datasets