Globant

Code Challenge

Autor: Luis Velásquez

Fecha de Entrega: 12/10/2023

Alcance

El siguiente documento tiene por objetivo cumplir con la entrega de documentación para estar al tanto de mi avance en el code Challenge de Globant.

La idea de esto es generar cuatro o tres entregables en donde se detallará de manera sencilla los pasos que he ido generando hasta encontrar la solución

También será un desafío ya que nunca e trabajado entregando documentación utilizando metodología ágil

Contexto

Se me ha solicitado como Analista de Datos Junior el Realizar un pequeño proceso ETL el cual debe pasar por distintas fases de Desarrollo.

Las instrucciones que se deben seguir para tener una referencia son las siguientes

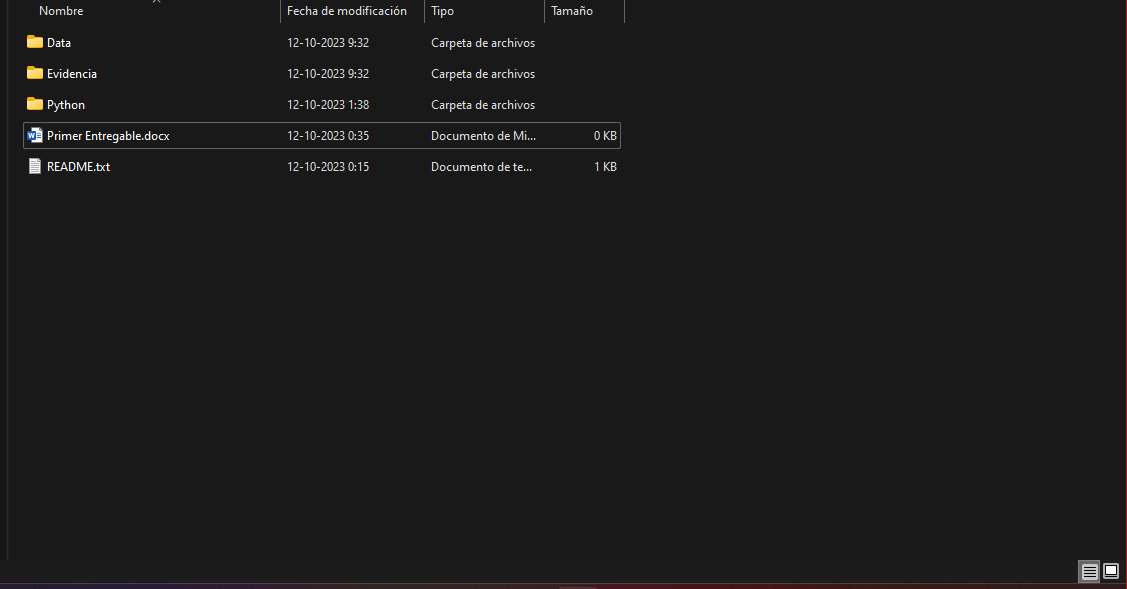
* Elegir un lenguaje de programacion entre Python o java para extraer datos desde un csv o una api ficticia
* Luego de extraer los datos, utilizar SQL para realizar la limpieza de estos
* Utilizar SDK de GCP para cargar los datos a BigQuery, se debe garantizar que la estructura de los datos sea la idónea
* Se debe generar una ingesta de datos de manera periódica (cada 24 horas) se puede utilizar Apache Airflow o Mediante cron Jobs
* Se debe utilizar GitHub para el control de versiones
* Incluir un Readme
* Organizar el desarrollo en Sprint
* Utilizar Trello O jira

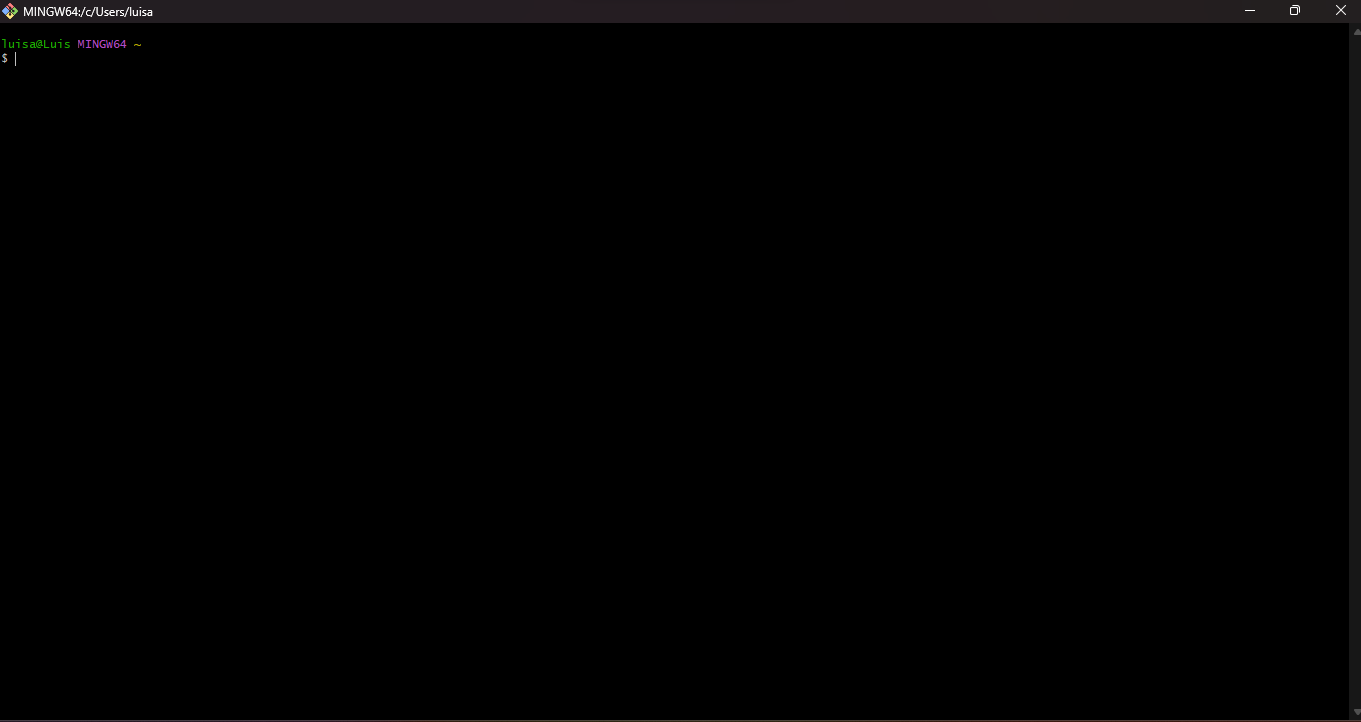
Siguiendo estas instrucciones se debe lograr el objetivo de generar un ETL sencillo

Avance

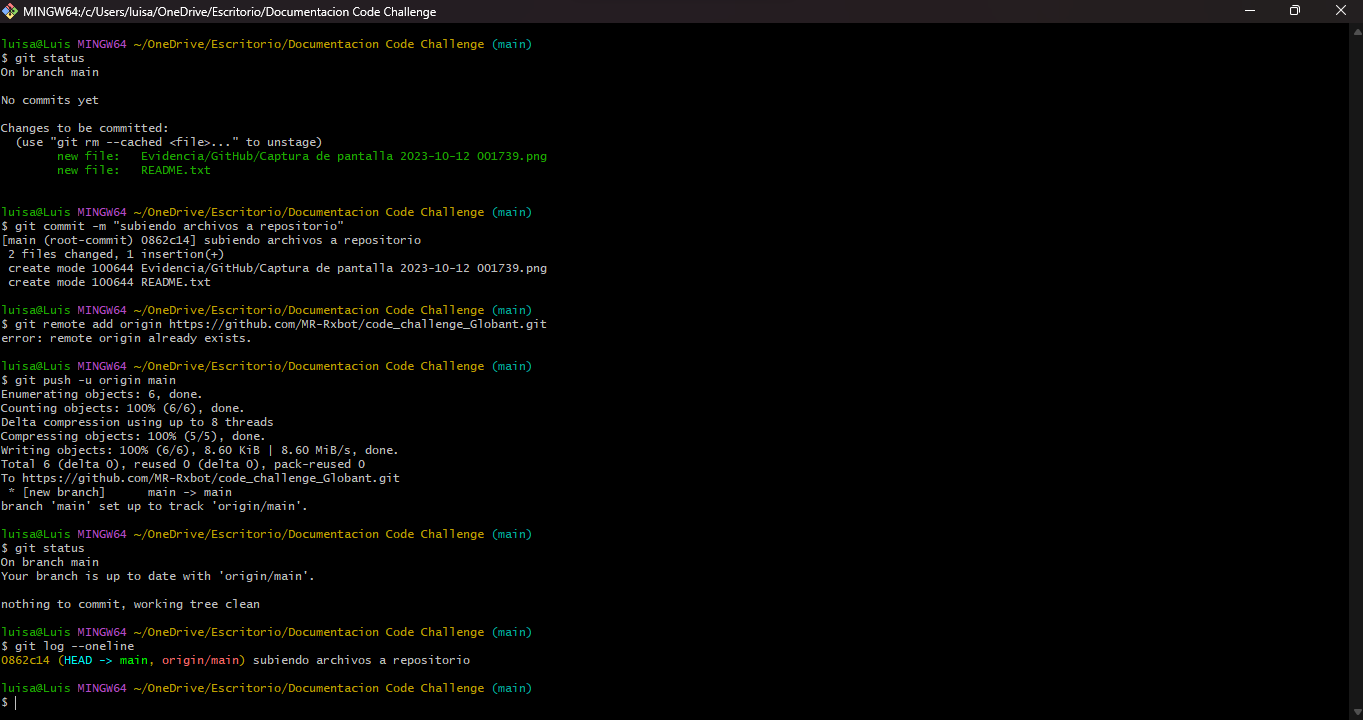
Como primer entregable Mostrare evidencia de la creación de las carpetas, creación de un repositorio para que todo quede guardado de forma segura, también subir el csv que usare para este caso al repositorio, además debo acceder a los datos utilizando un lenguaje de programacion.

Creación de una carpeta donde guardare todos los archivos



Creando Repositorio local

Subiendo archivos a repositorio Remoto



Utilizando Python accederemos a nuestro csv el cual este alojado en el repositorio de GitHub

Texto

Descripción generada automáticamente

Generando un nuevo archivo csv el cual se utilizará para ser cargado en BigQuery utilizando el SDK de GCP

Texto

Descripción generada automáticamente

Nuestro nuevo archivo

Texto

Descripción generada automáticamente

primer entregable

Para finalizar la entrega de la primera documentación doy por finalizado el primer documento en donde dejo evidencia y detallo el cumplimiento de lo requerido para nuestro primer sprint.