

Proyecto Final – Ciencia de Datos en Python

Proyecto: Ingeniería de Datos con Python

Tema: Python, Pandas, SQL, ETL, AWS

Fecha y Hora de Entrega: 12/04/2024 23:55

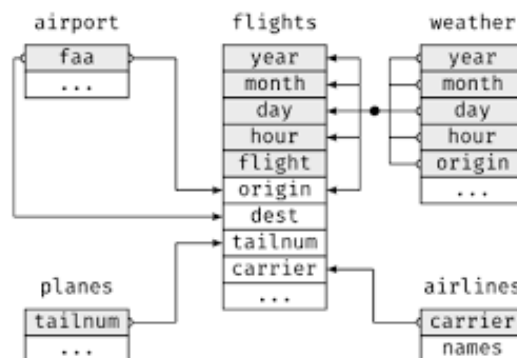
Formato de Entrega: Archivos de Construcción y Video.

Grupo: Grupos de 2 o 3 personas

Calificación: Presentación por medio de Video

DESCROPCIÓN: Para este proyecto usted deberá desarrollar un pipeline de ingeniería de datos utilizando Python, SQL y AWS como herramientas de desarrollo, su proyecto tener los siguientes componentes:

Sistema transaccional: El cual será consturido utilizando una instancia de RDS en AWS y SQL.



Para este sistema transaccional deberá crear el código SQL necesario considerando los tipos de datos adecuados, llaves primarias y llaves foraneas.

Ingestión de datos transaccionales: Deberá poblar la base de datos utilizando a partir de los archivos de .CSV para esto deberá usar Python y librerías de conexión de Python a SQL.

Preguntas de Negocio: Deberá plantear 10 preguntas de negocio que se puedan resolver utilizando la estructura de datos propuesta por el proyecto.

ETL y Analytics: Utilizando Markdown, numpy, pandas, matplotlib y seaborn, deberá elaborarlo un notebook con las respuestas a las preguntas planteadas de forma consistente y lo más profundo que sea posible. Para este caso deberá mostrar el código necesario para resolver cada pregunta planteada.

Adicionalmente deberá responder las siguientes preguntas de negocio utilizando tablas y gráficas:

1. ¿En que país y que avión se encuentra entre el 85% y el 70% de la cantidad de aterrisajes? Proporcione el nombre del país y el nombre del avión.
2. ¿Indique cual es el aeropuerto con la temperatura más alta registrada en los datos?
3. ¿Cuál es la áreaolínea con la menor cantidad de vuelos registrados, indique cuantos vuelos, el código de la áreaolínea, el nombre completo?
4. Indique la media, median, minimo, maximo y desviación estándar de las millas recorridas por cada avión, debe mostrar el nombre del avión y la información estadística en columnas adicionales
5. Muestre un cubo de información incluyendo la información de todas las tablas proporcionadas.
 - a. Indique la cantidad de filas y columnas
 - b. Indique cuantas y cuales son las variables categoricas, continuas, discretas y de fecha y hora.
 - c. Muestre una gráfica
 - i. de barras para la cantidad de las variables categoricas y discretas.
 - ii. De densidad para las variables continuas.
 - iii. Serie de tiempo con el conteo de apariciones para las de fecha y hora.

DETALLES TECNICOS: A continuación se describen los detalles técnicos mínimos que su proyecto debe cumplir:

- Para desarrollar el sistema transaccional podrá utilizar cualquier gestor de base de datos SQL que esté disponible en RDS.
- El procesamiento puede realizarlo en un maquina local o en una instancia de EC2 corriendo Python.

- El Notebook debe contener detalles sobre los procesos necesarios para responder las preguntas de negocio planteadas.
- Notar que no puede usar SQL para hacer la construcción de ninguna estructura salvo para leer de tablas almacenadas en las bases de datos es decir `SELECT * FROM tabla`. Cualquier otra necesidad de procesamiento deberá hacerla en Python con pandas y librerías de procesamiento de información

ENTREGA: Como entrega deberá publicar todos los archivos utilizados por medio de un link de Git, incluyendo la presentación utilizada en el video, los notebooks utilizados, los y los archivos adicionales que requiera. Adicionalmente deberá hacer un video de 5 a 7 minutos máximo donde explique todos los pasos que realizó para desarrollar su proyecto, es decir describir todos los elementos de su proyecto.