**AirSprint Aviation Project**

Data have been pivotal since the beginning of time, I am mostly motivated on how data- driven insight informed strategic decision and changing business trajectory positively. That been said, concerning the project at hand, below are the steps taken.

Understanding Dataset and how AirSprint Private Aviation company operates: - This is the foremost step before delving into any form of analysis or programming. It entails, on one hand understanding the concept of AirSprint Private Aviation – what they represent in the business space and the society, who are the target customers and how they operate their business to satisfy customers and maximize profit.

Foregoing, it preluded me to study and understand the dataset I was provided with and below are the steps administered.

In my terms, dataset can be seen as a person with an information to identity a problem or solve one.

Know the industry your dataset belongs to. just as knowing where your friends are from. It is obvious the AirSprint dataset belongs to the Aviation industry.

Understanding the relationships between your dataset to know the primary and foreign key concept, a concept that enables us to structure or organize dataset for effective and efficient use by connecting table records to another. Just as you would want to know the relationships of your friends to know who their parents and siblings are. This awareness permeated How AirSprint Aviation objects where connected to creating a unique record across the business model.

Lastly, understanding the kind of terms or words familiar to your dataset that is companies that word keep. Just as you would want to know the companies that your friend keeps. This awareness gave me the edge to follow appropriate naming conventions according AirSprint business policies.

After due diligence on the above steps explained, I was able to draw the crucial relationships between the models in AirSprint Aviation dataset according to their business model which enable me to create the database objects leveraging SQL Developer, Oracle database and Oracle HR User to support loading step of the ETL process. Please find attached the file that have the SQL Scripts.

**Prior to creating database objects:**

Prior to database object creation, after understanding AirSprint aviation business concept and identify the crucial relationships between the models in the dataset given: I leverage Python to develop the ETL process and Apache Airflow to automate its execution viz:

**Data Extraction/Collection:**

Import the AirSprint file or dataset to python workspace or editor, this is the step where each models in the dataset is imported to Python and in the case of this project there was also need to extract flight operations data from Fl3xx.com through API (Application Programming Interface) to initiate further process.

**Transformation/Data Wrangling**

Data Discovery/Exploration: Major part of this step have been told in the initial step of this project viz: Understand the structure, content and quality of the dataset, using python quick hacks like df.info () to identify missing value, data types and column names and indexes, df.describe() quickly identify the statistical measures of the dataset usually numerical.

**Data Cleaning:**

This step process was minimal due to the nature of the dataset. Actions taken to validate Oracle date format and data inserting includes

* formatting the date to validate Oracle database date format to avoid error.
* replacing empty values of a column with null.
* Removed duplicates.
* Converting each column of the dataset to suit Oracle data types.

Other possible actions include handling missing value either by imputation or deletion and also outliers with any statistical replacement or interpolation etc.

**Data Structuring:**

It entails normalization i.e. rearranging my dataset to be consistent with the primary and foreign key relationship where one or two models share a common value for easier data retrieval. This step was the major part of the project because I have to tie all the models of my dataset together so that a single unique id can retrieve a whole record of data. There is also denormalization which is reducing the models to the barest number possible with the same aim of faster and easier data retrieval with a single unique identifier.

**Data Enrichment:**

I had to add this because the flight Ids I was given whose record I need to retrieve from Fl3xx.com through API have one Id missing a record and I decided to add the information in compliance with the business rule and flight record logic.

**Data Validation:**

this is the holistic steps of making sure data is consistent, accurate and complete after cleaning to make sure it complies with the AirSprint Aviation business concepts and rules.

**Loading:**

Having completed the above two steps, the models were loaded into Oracle 19C database using execute many query with Python OracleDB Library.

**Automations**-Python Apache Airflow

Please find attached the SQL scripts that created all the database objects and Python Jupiter file.

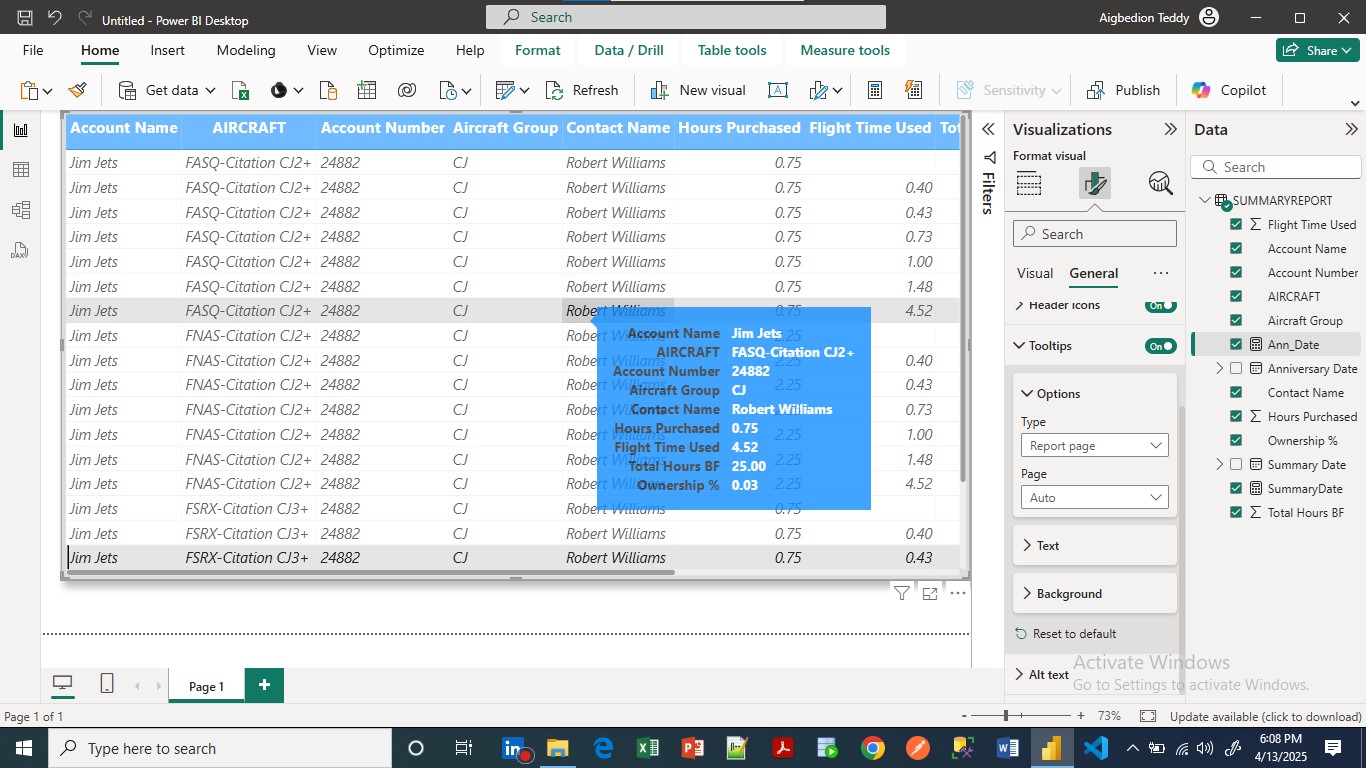
**Visualization Task**- To achieve this task, I leveraged Microsoft Power Business Intelligence(PBI).

**How?** By integrating PBI directly to the database Oracle 19C .

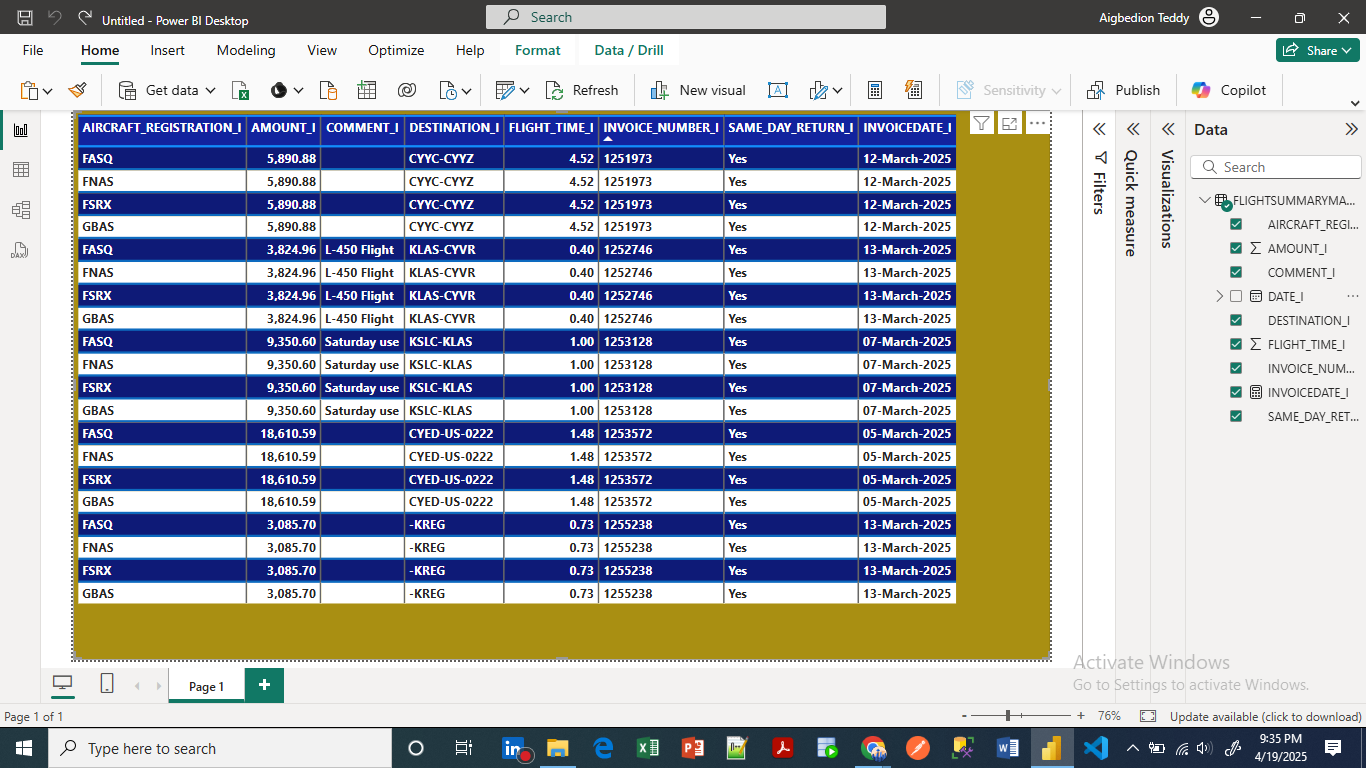
**Reason:** for scalable data source:

* This will make the virtual auto update to current data structure, thereby enhancing real-time insights
* The model used for my visualization was developed as a view in the database using SQL: This prompt having to integrate PBI with the database to report from the view created seated in the database

Please find attached the PBI visualization file and a screenshot below for easy sighting

Visualization 1:

Visualization 2:

Visualization 2: