DBMS Project Deliverable -1

Student 1 Name and Odin ID: Nida Mariam Sheikh Aslam – (nidama)

Student 2 Name and Odin ID: Snehil Shrivastava (snehils)

Subject Area: Flight Operations Analysis at JFK Airport

The subject area focuses on the daily operations of flights taking off from John F. Kennedy International Airport (JFK), one of the busiest airports in the United States and a major international air travel hub. JFK Airport serves as a critical point in the global air transportation network, handling flights to and from hundreds of destinations worldwide. The operational efficiency, weather conditions, air traffic control, and scheduling at JFK have a significant impact on flight punctuality, safety, and passenger satisfaction. Analysing the data related to flights departing from JFK can provide insights into patterns of delays, the influence of weather conditions on flight operations, and the performance of different airlines in managing their schedules. This analysis can help stakeholders, including airlines and airport authorities, to make informed decisions to improving operational efficiency and enhancing the overall travel experience.

Potential Questions for Analysis

1. How many flights departed from JFK to each destination between Nov 2019-Dec-2020?
2. What are the average departure delays across different airlines at JFK?
3. What flights were delayed by more than 60 minutes and by how much?
4. Are there any noticeable trends in departure delay based on time of day?
5. Which destinations have the highest incidence of departure delays?
6. Which destination received the most flights from JFK in November 2019?
7. Which day of the week experiences the highest average departure delay?
8. Which aircraft models are most frequently delayed?
9. Are longer flights (in terms of distance) more susceptible to departure delays?
10. How do specific holidays or events impact flight operations?
11. Is there a difference in departure delays between weekdays and weekends?
12. On average how many flights are departing and arriving to JFK?
13. Do the number of flights departing and arriving to JFK vary based on holidays?
14. How do scheduled departure and arrival times compare to actual times across different carriers?
15. What is the overall on-time departure rate for flights departing from JFK?
16. For each type of wind condition, list the average departure delay and total number of affected flights.
17. Between November 2019 and December 2020, which month has the highest number of delays?
18. Are there any significant changes in departure delay patterns over the course of the dataset period (Nov 2019-Dec 2020)?
19. How do wind speed and direction affect flight operations?
20. What percentage of flights depart early in the morning (before 6 AM) and their punctuality?

Data Source and Ingestion Plan

The primary source of data for this project will be the JFK Airport Flight Take-off Data available on Kaggle:
[https://www.kaggle.com/datasets/deepankurk/flight-take-off-data-jfk-airport?resource=download].
This dataset includes information on flight departures from JKF airport between Nov 2019-Dec-2020.
The parameters included in the table are:

MONTH - Month

DAY_OF_MONTH - Date of flight

DAY_OF_WEEK - Day of the week

OP_UNIQUE_CARRIER - Carrier Code (Should generally be carrier company)

TAIL_NUM - Airflight Number

DEST - Destination

DEP_DELAY - Departure delay of the flight

CRS_ELAPSED_TIME - Scheduled journey time of the flight

DISTANCE - Distance of the flight.

CRS_DEP_M - Scheduled Departure Time.

DEP_TIME_M - Actual Departure Time (Gate checkout of the flight not the take-off time)

CRS_ARR_M - Scheduled Arrival Time

Temperature - Temp.

Dew Point - Dew

Humidity - Humidity

Wind - Wind

Wind Speed - Wind speed

Wind Gust - Wind Gust

Pressure - Pressure

Condition - Condition of the climate

sch_dep - No. of flights scheduled for arrival.

sch_arr - NO. of flights scheduled for departure.

TAXI_OUT - Taxi-out time (Run away time) (Target variable)

The dataset has 28821 rows. In order to ingest the data into a database, the CSV file obtained from Kaggle will be imported into a PostgreSQL database using DataGrip. The process involves creating a database schema reflective of the dataset's structure, followed by importing the CSV file directly into the database through DataGrip's import functionality.