Aviation Analysis Using SQL – Project Report

1. Introduction

The aviation industry produces massive amounts of structured and unstructured data related to flights, operations, and passenger movement. Efficient analysis of such data plays a crucial role in understanding delay patterns, improving operational workflows, and enhancing customer experience. This project involves performing in-depth SQL analysis on a dataset containing over **50 lakh (5M+) flight records pulled into a SQL Server** local instance. The objective was to uncover trends in flight delays, focusing particularly on weekday vs. weekend performance and delay contributors.

2. Project Objective

• Handle and process large-scale datasets using SQL Server.
• Perform data cleaning, transformation, and optimization for analysis.
• Identify and analyze flight delay patterns across different time periods.
• Derive actionable insights to showcase SQL query mastery and big data handling capabilities.

3. Tools & Technologies Used

• SQL Server (Local Instance) – Database storage & execution engine.
• SSMS (SQL Server Management Studio) – Query execution & management.
• CSV Files (5M+ rows) – Source dataset imported using bulk operations.

4. Data Preparation & ETL

The raw dataset consisting of large CSV files was loaded into SQL Server using optimized **My SQL LOAD DATA INFILE** command operations.

Data cleaning steps included:
• Removing duplicates and invalid rows.
• Standardizing date-time formats.
• Fixing inconsistent airline or airport codes.
• Handling missing delay values with domain appropriate techniques.

5. Exploratory Data Analysis (EDA)

The analysis performed using SQL included:
a) **Weekday vs Weekend Delay Comparison** – Understanding delay frequency on different days.
b) **Flight Delay Trend Analysis** – Examining scheduled vs. actual times.
c) **Airline & Route Performance Ranking** – Identifying high delay carriers and congested routes.
d) **Peak Delay Time Windows** – Detecting hours with maximum operational delays.

6. Key Insights Derived

• Weekend flights showed statistically higher delays compared to weekdays.
• A few specific airlines maintained strong OTP (OnTime Performance).
• Certain routes experienced significantly higher congestion induced delays.
• Late evening and night flights were more prone to extended delays.
• Proper indexing improved query performance by more than 40% on large tables.

7. Conclusion

This aviation analysis project demonstrates expertise in handling big datasets, writing optimized SQL queries, and performing analytical investigations. Insights derived from this project emphasize the importance of data-driven decisions in aviation operations. The work highlights proficiency suitable for data analyst roles involving SQL, big data, and EDA.

8. Challenges and solution

Challenge 1: Handling 50 lakh+ rows from multiple files
Solution: Imported 50 lakh+ records efficiently using the MySQL LOAD DATA INFILE command to handle large CSV files with high performance.

Challenge 2: Duplicate & inconsistent records
Solution: Applied DISTINCT, ROW_NUMBER(), and data validation logic.

Challenge 3: Maintaining unique records across files
Solution: Created relational keys and standardized data before joining