|  |  |
| --- | --- |
|  | **DEPARTMENT OF COMPUTER ENGINEERING** |

**Project Report**

|  |  |
| --- | --- |
| Semester | B.E. Semester VII – Computer Engineering |
| Subject | BDA |
| Subject Professor In-charge | Prof. Pankaj Vanwari |
| Assisting Teachers | Prof. Pankaj Vanwari |

|  |  |
| --- | --- |
| Roll Number | Name of Students |
| 21102A0052 | Ayush Makade |
| 21102A0009 | Nikhil Dhumal |
| 21102A0011 | Swaraj Andhale |
| 21102A0001 | Digvijay Mawale |

**Name of the Project: Flight Delay Analysis and Airline Performance Evaluation Using R**

**Project Details:**

**R**: Used for data manipulation, analysis, and visualization.

**Packages**:

* **ggplot2**: For creating visualizations and plots.
* **dplyr**: For data wrangling and manipulation.
* **data.table**: For efficient data loading and handling large datasets.
* **lubridate**: For working with date and time formats.

**Why This Dataset?**

This flight delay dataset offers significant insights into the performance and behavior of airlines. It allows us to:

* Analyze flight delays across different airlines, time periods, and days of the week.
* Investigate patterns of cancellations and diversions to identify performance issues.
* Evaluate how different factors like airline and day of the week affect delay times.
* Measure the impact of delays on airline schedules and understand how they vary by airline and flight path.

**Methodology**

**Data Preprocessing**

* Missing Values: We checked for missing values and handled them by removing rows with missing values in key columns like FlightDate, Airline, and DepTime.
* Date and Time Conversion: Columns like FlightDate and DepTime were converted to appropriate date and numeric formats for accurate analysis.
* Feature Engineering: A new feature, total\_delay, was created to measure the total delay by combining DepDelayMinutes and ArrDelayMinutes. Another feature, day\_of\_week, was created from FlightDate to analyze delays by the day of the week.

**Analysis and Visualizations**

**1. Line Plot: Flight Counts per Day of the Week**

* **Purpose:** Understand flight traffic trends over the days of the week.
* **Insight:** Flight counts vary by day, with specific days experiencing heavier traffic, which may help predict delays or congestion.

**2. Bar Plot: Average Departure Delay by Airline**

* **Purpose:** This analysis aimed to measure the average departure delay across different airlines.
* **Insight:** Some airlines consistently have higher average delays than others, offering insights into performance variability across airlines.

**3. Bar Plot: Average delay by Time of the Day**

* Purpose: To analyze how average flight delays vary across different times of the day.
* Insight: Helps understand trends in opening preferences. The analysis helps us to understand that evening flights are most prone to delays, likely due to accumulated operational issues throughout the day.

**4. Line Plot: Average Delay per Day of the Week**

* **Purpose:** Measure how delays change based on the day of the week.
* **Insight:** Some days exhibit higher average delays, offering clues on which days travellers might face more issues.

**5. Bar Plot: Flight Volume by Time Block**

* **Purpose:** To analyse the distribution of flight departures across different time blocks throughout the day.
* **Insight:** The highest flight volume occurs in the morning (0600-1159), indicating peak operational hours and higher demand for morning flights.

**Conclusion and Insights**

This project provides valuable insights into flight delays and airline performance:

* **Average Delays by Airline**: Some airlines show higher average delays, which could impact their reliability.
* **Day-of-Week Analysis:** Flight counts and delays vary across the week, with certain days exhibiting more flight activity and longer delays.
* **Delay Distribution:** Departure delays are concentrated around shorter durations, but significant outliers exist that cause operational issues.
* **Departure vs Arrival Time**: There is a visible relationship between departure and arrival times across airlines, highlighting punctuality and recovery from delays.

**GitHub Repository Link (Public):** [GITHUB-LINK](https://github.com/nikhil-dhumal/BDA_MiniProject)

**DATA-SET LINK(PUBLIC):** [**KAGGLE**](https://www.kaggle.com/datasets/robikscube/flight-delay-dataset-20182022?select=Combined_Flights_2022.csv)

**Output Screenshots:**

A graph with a line

Description automatically generated

A graph with different colored squares

Description automatically generated

A close-up of a list of airline flights

Description automatically generated

A graph with a line

Description automatically generated

A red squares with black numbers

Description automatically generated

A graph of blue bars

Description automatically generated

A graph of different colored bars

Description automatically generated