❖ **Project** : **Airline referral prediction Analysis**

- ➢ **Tools Used**: Python
- ➢ **Industry Context**:-
  - i. The travel and hospitality industry, especially aviation, is growing rapidly and becoming more complex.
  - ii. Airlines and airports face challenges in managing flight schedules, minimizing delays, and improving customer satisfaction.
  - iii. Data analytics helps address these challenges by providing insights that can enhance operational efficiency and improve passenger experiences.
- ➢ **Recent Examples in Industry Context:**-
  - i. **Delta Air Lines' Operational Overhaul**: In 2023, Delta implemented advanced analytics to revamp its scheduling and delay management processes, resulting in a 20% reduction in average delays.
  - ii. **Heathrow Airport's Digital Transformation**: Heathrow introduced AI-driven analytics to manage passenger flow and resource allocation, reducing wait times by 15% during peak hours in 2022.
  - iii. **Singapore Airlines' Fleet Optimization**: By leveraging data analytics, Singapore Airlines optimized their fleet usage, cutting down maintenance costs by 10% and improving on-time performance in 2023.
- ➢ **Project Objective: -**
  - i. As a data analyst my objective for this project was to:
  - ii. Analyse extensive aviation datasets to understand flight operations, delay patterns, and airport traffic dynamics.
  - iii. Provide actionable recommendations to improve flight schedules, reduce delays, and enhance overall operational efficiency.
  - iv. Develop interactive EDA to visualize key aviation metrics, making it easier for stakeholders to explore data and make informed decisions and create ML models to predict the key findings.
- ➢ **About Data:-**
- ➢ The project involved three primary datasets, all from the year 2015**:**
  - i. **Flights Dataset**: Contains detailed flight information, including schedules, delays, and cancellation details.
  - ii. **Airlines Dataset**: Provides information about various airlines.

iii. **Airports Dataset**: Contains information about different airports, including location and traffic details.

➢ **Data Overview and Dataset Description**:

➢

- **airline**: Name of the airline.
- **overall**: Overall rating of the trip (1 to 10).
- **author**: Author of the review.
- **reviewdate**: Date of the review.
- **aircraft**: Type of aircraft used.
- **travellertype**: Type of traveler (e.g., business, leisure).
- **cabin**: Cabin class.
- **seatcomfort**: Comfort of the seat (1-5 rating).
- **cabinservice**: Cabin service rating (1-5).
- **foodbev**: Food and beverage rating (1-5).
- **groundservice**: Ground service rating (1-5).
- **valueformoney**: Value for money rating (1-5).

➢ **Data Preprocessing: -**

➢ To prepare the data for analysis, I undertook the following steps: -

i. **Data Cleaning**: Ensured consistency in data types and handled missing values using methods such as mean, median and mode filling.

ii. **Feature Engineering**: Added custom columns to enrich the dataset, such as calculating total delay times and categorizing delay reasons

➢ **Key Results: -**

➢ In traveller type, it is evident that people have given both 0 and 1 which we can take it as positive and negative recommendation to interpret to this to family type. positive recommendation given may be due to low price and negative recommendation may be due to lack of proper services and infrastructure. But this is just and assumption from the data we have received.

➢ From month vs no. of recommendation. We can see that people tents to travel most in the month of July considering the total of positive and negative recommendation combined.

➢ In cabin service rating people has given highest recommendation rating to cabin service rating 5 as compared to its other counterparts, It is evident that the cabin service is performing good.

- In Food and Bev department, It is evident that highest negative recommendation has been given to rating 1 which is a matter of concern and the department has to improve the quality of food and services.
- In Entertainment, It is evident that highest negative recommendation has been given to rating 1 which is a matter of concern and the department has to improve the their entertainment system.
- In value for money also we can see most people has given highest negative recommendation to value for money rating 1 which shows that airline has to make their flight service more cost effective.
- Here the accuracy of Random Forest model is higher compared to all other models with a score of 0.957166 and logistic regression and support vector machines have second highest accuracy of 0.956831. The combined score of Precision, Recall and F1-score and ROC_AUC_SCORE is highest in Random Forest model hence we will be selecting random forest for further prediction.

- **Business Impact:**
  - **i. Operational Efficiency Enhancement:**
    - Improved on-time performance and reduced delays enhance customer satisfaction.
  - **ii. Resource Allocation Optimization:**
    - Efficient resource allocation reduces wait times and enhances passenger experiences.
  - **iii. Strategic Planning for Airlines and Airports:**
    - Anticipating seasonal trends optimizes schedules and capacity, improving operations.
  - **iv. Continuous Improvement Initiatives:**
    - Implementation of best practices leads to higher customer satisfaction and operational efficiency.
- **Expected Interview Questions**
  - i. Can you describe the main objectives of your project in the travel and hospitality analytics domain?
  - ii. What specific challenges do airlines and airports face that data analytics can help address?
  - iii. Could you walk us through the steps you took in preprocessing the aviation datasets for analysis?
  - iv. How did you handle missing values in the dataset? Why did you choose those specific methods (mean, median, mode)?
  - v. What types of feature engineering did you perform on the dataset, and how did it enrich the analysis?

vi. Can you explain some of the actionable insights you derived from the EDA you developed?
vii. How did you identify peak times for delays and common delay causes using your analysis?
viii. What metrics did you use to compare airline efficiency, and how did you present this information in your Project?
ix. Could you elaborate on how you analysed airport traffic dynamics and its relevance to resource allocation optimization?
x. What were some of the key results you obtained from your analysis, particularly in terms of flight operations efficiency and airport traffic dynamics?
xi. How do you think the business impacts of your project, such as operational efficiency enhancement and resource allocation optimization, contribute to the overall goals of airlines and airports?
xii. Can you discuss any continuous improvement initiatives you recommended based on your findings?
xiii. What were some of the challenges you encountered during this project, and how did you overcome them?
xiv. How do you ensure the accuracy and reliability of your analytical findings when working with large and complex datasets?
xv. In what ways do you see the insights and recommendations from this project shaping the future of the aviation industry?