**Flight Delay Prediction Using Linear Regression**

**1. Introduction**  
This project focuses on predicting flight arrival delays using departure delay data from real-world airline records. The objective was to build a simple yet effective regression model that can estimate how much a late departure affects the arrival time. This kind of predictive model is essential for operational decision-making in aviation logistics.

**2. Dataset Overview**  
The dataset, sourced from publicly available flight records, includes over 300,000 flight entries with columns such as:

* Departure and arrival times
* Scheduled times
* Airline name
* Distance, air time, origin, and destination airports

Missing values in the dep\_delay and arr\_delay columns were cleaned to prepare the dataset for modeling.

**3. Exploratory Data Analysis (EDA)**  
Key steps included:

* Checking for missing values
* Visualizing average departure delays by hour, month, and airline
* Identifying strong correlation (0.91) between departure and arrival delays

**4. Regression Model**  
A simple Linear Regression model was applied using dep\_delay as the predictor and arr\_delay as the target. The dataset was split into training and testing sets, and the model was trained on the training subset.

**Results:**

* **MAE (Mean Absolute Error):** 13.13 minutes
* **RMSE (Root Mean Squared Error):** 18.07 minutes
* **R² Score:** 0.83

These metrics indicate that the model explains 83% of the variability in arrival delay based on departure delay, with relatively low error margins.

**5. Interpretation and Use Cases**  
The high correlation between dep\_delay and arr\_delay makes it feasible to use this model in:

* Scheduling adjustments
* Resource allocation for ground staff
* Real-time alert systems for passengers

**6. Future Improvements**  
The project can be enhanced by:

* Including multiple predictors such as air time, carrier, and destination
* Trying regularized regression models (Ridge, Lasso)
* Building airline-specific models for comparison

**7. Conclusion**  
This project demonstrates that even a basic regression approach can yield valuable insights for operational forecasting. The methodology is transparent and easily extendable for more advanced machine learning techniques.

**Project Summary**

**Flight Delay Prediction Using Linear Regression**  
Developed a regression model to forecast flight arrival delays based on departure delays. Achieved an R² score of 0.83, with average prediction error under 15 minutes. Used real-world airline data (flights.csv), performed data cleaning, visualization, and model evaluation. Ideal for real-time alert systems and delay risk assessment in aviation operations.