**Project Title: Flight Delays – Data Cleaning, EDA & Dashboard Visualization**

**1. Introduction**

This project focuses on analyzing flight delays in the United States. The primary goal is to clean and preprocess the dataset, perform exploratory data analysis (EDA), and create an interactive Power BI dashboard to identify patterns, root causes, and trends of flight delays. The project is part of a data analytics learning path, integrating Python-based EDA and Business Intelligence tools.

**2. Dataset Description**

* **Source:** U.S. Bureau of Transportation Statistics (BTS)
* **Structure:** The dataset includes flight arrival statistics with the following key columns:
  + year, month, carrier, airport, arr\_flights, arr\_del15, carrier\_ct, weather\_ct, nas\_ct, security\_ct, late\_aircraft\_ct, and their respective delay times.
* **Time Period:** Multiple years (usually 2015–2022)
* **Observations:** Includes thousands of rows representing flights arriving at various U.S. airports.

**3. Tools and Technologies**

* **Python (Jupyter Notebook):** Data cleaning and exploratory analysis.
* **Pandas, NumPy, Matplotlib, Seaborn:** For data manipulation and visualization.
* **PostgreSQL:** Optional storage and management of the cleaned dataset.
* **Power BI:** Dashboard creation with interactive graphs and filters.

**4. Procedure Description**

**Step 1: Data Cleaning**

* Checked and handled missing values.
* Removed rows with more than 50% nulls in critical columns.
* Filled nulls in arr\_del15 with 0, assuming no delay.
* Exported cleaned data to CSV for further analysis.

**Step 2: Exploratory Data Analysis (EDA)**

* Performed distribution plots for delay causes.
* Used bar charts and heatmaps to examine correlations.
* Grouped data by airline, month, and delay type to uncover trends.

**Step 3: Dashboard Creation in Power BI**

* Loaded the cleaned dataset into Power BI.
* Created:
  + A delay cause breakdown per airline and airport.
  + Monthly delay trends.
  + KPIs (% cancelled flights, % diverted, total delay time).
  + Filters by year, airline, and airport.
* Optimized axis ranges, legends, and layout for clarity and usability.

**5. Results and Conclusions**

* **Insights:**
  + Late aircraft and NAS delays are among the top contributors.
  + Weather-related delays increase in winter months.
  + Certain carriers consistently report higher delays.
* **Visualization Outcomes:**
  + The dashboard allows quick filtering and understanding of delay trends.
  + Stakeholders can identify performance issues by airport and carrier.

**Conclusion:** This project demonstrates an end-to-end process of cleaning, analyzing, and visualizing aviation delay data using Python and Power BI, helping to transform raw data into actionable insights.

README

# Flight Delays Analysis – EDA & Power BI Dashboard

## Introduction

This repository showcases a data analysis project focused on flight delays in the United States. It covers dataset cleaning using Python, exploratory data analysis (EDA), and a final Power BI dashboard for visualization.

---

## Dataset Description

- \*\*Source:\*\* U.S. Bureau of Transportation Statistics

- \*\*Key Fields:\*\* Year, Month, Carrier, Airport, Delay Types, Arrival Flights, etc.

- \*\*Goal:\*\* Identify major delay causes and patterns across airlines and airports.

---

## Tools Used

- Python (Jupyter Notebook)

- Pandas, NumPy, Matplotlib, Seaborn

- PostgreSQL (optional)

- Power BI

---

## Procedure Overview

1. \*\*Data Cleaning:\*\* Handle missing values and remove rows with >50% nulls.

2. \*\*EDA:\*\* Analyze delay causes and trends by carrier, airport, and time.

3. \*\*Dashboard:\*\* Build an interactive Power BI dashboard with KPIs and visual filters.

---

## Results and Insights

- Late aircraft and NAS delays dominate most of the cases.

- Delay patterns vary significantly by airline and season.

- Power BI enables dynamic filtering and business-focused visuals.

---

## Files Included

- `EDA\_Flights\_Delay.ipynb` – Python script for data cleaning and EDA

- `Dashboard Flights Delay.pbix` – Power BI dashboard file

---

## Screenshot

![Dashboard Preview](link\_to\_image\_if\_available)

---

## Contact

Developed by \*\*Mónica Prieto\*\* – for academic and portfolio purposes.