Uber Fare Data Analysis Project

A comprehensive data analysis project exploring Uber ride fare data using **Python (Jupyter Notebooks)** for cleaning and **Power BI** for visualization.
The goal is to identify ride patterns, fare distributions, and geographic trends to support business decisions.

Project Objectives

- Understand fare amount distribution and ride behavior.
- Identify time-based and location-based ride trends.
- Use data to make useful business suggestions.

Tools & Technologies

- **Data Source**: Kaggle Uber Dataset
- **Data Cleaning**: Python (Pandas, NumPy in Jupyter Notebook)
- **Data Visualization**: Power BI

---

1. Data Preparation

Steps Taken:

- Removed missing or incorrect values.
- Converted date/time columns into proper datetime format.
- Extracted hour, weekday, and month from timestamp.
- Filtered out rides with invalid location (latitude/longitude outliers).
- Removed fare values less than 0 and extremely high outliers.

---

2. Exploratory Data Analysis (EDA)

🔹 Fare Amount Distribution

- Most fares are under a certain amount (long-tail distribution).
- Boxplots helped identify outliers in fare values.

🔹 Ride Time Analysis

- Grouped data by hour, weekday, and month.
- Found peak hours (early mornings and late evenings).
- More rides occurred during weekends and late afternoons.

🔹 Temporal Trends

- Created time series visuals to observe trends across hours, days, and months.
- Identified growth in ride frequency on weekends and during holiday seasons.

🔹 Geographic Insights

- Plotted pickup locations using latitude and longitude.
- City centers and busy areas showed high ride concentration.

---

3. Power BI Dashboard

The Power BI dashboard includes the following:

- **Fare Distribution**: Histograms and box plots
- **Ride Duration Trends**: Time-based charts (hour/day/month)
- **Time Series Analysis**: Visuals showing ride patterns over time
- **Geographic Mapping**: Map visual of pickup locations
- **Geographic Mapping**: Map visual of pickup locations
- **Filters**: Day of the week, hour of the day, fare range
- **Drill-down Features**: Click to explore specific time ranges or locations

---

4. Key Findings

- Most Uber fares are low-priced with a few high-value outliers.
- Ride volume increases during rush hours and weekends.
- Certain locations have very high ride activity.
- Seasonal and daily ride trends are visible and predictable.

---

5. Business Recommendations

- Apply dynamic pricing during busy hours.
- Add more drivers near high-demand areas.
- Offer promotions during low-demand hours.
- Use real-time ride and fare patterns for better planning.

Folder Structure

```
├── cleaned_data/           # Cleaned dataset (CSV)
├── notebooks/              # Jupyter Notebook files for data cleaning
├── powerbi_dashboard.pbix  # Power BI dashboard file
├── visuals/                # Screenshots of visualizations (optional)
└── README.md               # Project report (this file)
```