Project Report

Project Name: Uber Rides Data Analysis

Group Members:

- Member 1: Avishkar Chavhan, Project Lead
- Member 2: Sarthak Gite, Visualization Specialist
- Member 3: Prathamesh Bhagwat, Presentation Specialist
- Member 4: Omkar Chavhan, Documentation Specialist
- Member 5: Swapnil Gandhale, Data Analyst

Guide/Supervisor: Nikita Mam

Date of Submission: 24/01/2025

---

1. Objective

The primary objective of this project is to analyze Uber ride data to extract meaningful insights into:

- Peak demand hours and days.
- Popular pickup and drop-off locations.
- Trends in ride volumes over time (hourly, daily, or monthly).
- Customer payment preferences.
- Average trip durations and anomalies in trip durations.

The analysis aims to improve operational efficiency, customer experience, and strategic planning.

---

2. Sources

References, Tools, and Resources Used:

- **Programming Language:** Python
- **Libraries:**
  - pandas and numpy for data preprocessing and manipulation.
  - matplotlib and seaborn for data visualizations.
  - folium for location-based visualizations (maps).
- **Dataset:** A CSV file containing Uber ride details for India.
- **Development Environment:** Python Environment(venv), Streamlit, Colab.

---

3. Steps Taken

Step-by-Step Explanation:

1. **Data Collection and Preparation:**
    - Loaded the provided CSV dataset containing details of Uber rides across India.
    - Verified the integrity of the dataset by checking for missing or incorrect values.
    - Standardized column names and converted timestamp fields to a datetime format.
2. **Data Preprocessing:**
    - Extracted relevant features such as hour, day of the week, and month from pickup timestamps.
    - Calculated trip durations in minutes from pickup and drop-off timestamps.
    - Filtered invalid or negative trip durations to ensure data quality.
3. **Analysis:**
    - **Peak Demand Analysis:** Identified the busiest hours of the day and busiest days of the week for rides.
    - **High-Demand Zones:** Determined the most popular pickup and drop-off locations using latitude and longitude data.
    - **Payment Preferences:** Analyzed payment methods (cash, card, digital wallet) to identify trends.
    - **Trip Duration Insights:** Calculated average trip durations for different times of the day and flagged anomalies.
4. **Visualization:**
    - Created bar charts for peak hours and payment distribution.
    - Generated heatmaps to visualize high-demand zones using folium.
5. **Output Generation:**
    - Saved visualizations in the output folder as PNG files.
    - Exported the heatmap as an HTML file for interactive exploration.

---

4. Code

Key Portions of Code:

1. **Data Preprocessing:**

```
import pandas as pd

def preprocess_data(file_path):
    df = pd.read_csv(file_path)
    df['pickup_time'] = pd.to_datetime(df['Pickup Timestamp'],
errors='coerce')
    df['dropoff_time'] = pd.to_datetime(df['Dropoff Timestamp'],
errors='coerce')
    df['trip_duration'] = (df['dropoff_time'] -
df['pickup_time']).dt.total_seconds() / 60
    df['hour'] = df['pickup_time'].dt.hour
    df['day_of_week'] = df['pickup_time'].dt.day_name()
    return df
```

2. **Peak Hours Analysis:**

```
def peak_hours_analysis(df):
    return df.groupby('hour').size()
```

3. **High-Demand Zones:**

```
def high_demand_zones_analysis(df):
    return df.groupby(['Pickup Latitude', 'Pickup
Longitude']).size().sort_values(ascending=False).head(10)
```

4. **Visualization:**

```
import matplotlib.pyplot as plt
import seaborn as sns

def plot_peak_hours(peak_hours, output_path):
    plt.figure(figsize=(10, 6))
    sns.barplot(x=peak_hours.index, y=peak_hours.values, palette="viridis")
    plt.title("Peak Hours for Uber Rides")
    plt.xlabel("Hour of Day")
    plt.ylabel("Number of Rides")
    plt.savefig(output_path)
```

---

5. Results

Summary of Outcomes:

1. **Peak Hours:**
   - The busiest hours for Uber rides were between [Insert Hours].
2. **High-Demand Zones:**
   - The most popular pickup location was latitude [Insert Latitude] and longitude [Insert Longitude].
3. **Payment Preferences:**
   - Digital wallets accounted for [Insert Percentage] of all rides, indicating a trend towards cashless transactions.
4. **Trip Duration:**
   - The average trip duration was [Insert Duration] minutes.
   - Trips with unusually long durations were flagged for further investigation.

---

6. Conclusion

The analysis provided valuable insights into ride patterns, high-demand zones, and payment trends for Uber rides in India. These findings can help Uber optimize operations, improve customer satisfaction, and strategize for future growth.