# Uber Trip Data Analysis Report

## 1. Introduction: Project Overview and Objectives

This analytical report explores ride fare patterns, pickup/dropoff behavior, and passenger trends based on a dataset of 200,000 Uber rides.  
  
Objectives:  
- Understand fare distribution and detect anomalies (outliers)  
- Identify peak seasons and passenger count behavior  
- Offer recommendations for business improvement based on findings

## 2. Methodology: Data Collection and Analysis Approach

Dataset: 200,000 records of Uber ride transactions  
Tools Used: Python (pandas, seaborn, matplotlib), Jupyter Notebook  
Steps Followed:  
- Data cleaning: removed null values, invalid coordinates, and negative fares  
- Statistical analysis: summary statistics, mean/median/std dev, and outlier detection  
- Visual analysis: plotted seasonal and geographical data patterns

## 3. Analysis: Detailed Findings and Statistical Insights

Summary Statistics:

Fare Amount:  
Mean: 11.36  
Median: 8.50  
Standard Deviation: 9.89  
Max: 499.00  
Min: -52.00  
Outliers Detected: 17,155  
  
Passenger Count:  
Mean: 1.68  
Standard Deviation: 1.38  
Max: 208  
Min: 0

Note: The dataset had some extreme outliers (e.g. fares over $400 and distances with unrealistic coordinates).

Geographical Patterns:  
- Most pickup and dropoff locations are within a small consistent range, likely urban areas like New York City.  
- Some coordinates fall far outside the expected range and were marked as outliers.

Seasonal Trend:  
- A clear spike in Uber usage during Spring was observed, likely due to better weather, tourism, and outdoor activities.

## 4. Results: Key Discoveries and Patterns

- Fare distribution is right-skewed with most rides under $20.  
- Majority of rides had 1 or 2 passengers.  
- Several outliers exist in fare amounts and coordinates.  
- Spring season sees a surge in Uber demand.

## 5. Conclusion: Summary of Main Findings

- The average Uber fare is approximately $11.36.  
- Most rides involve 1 passenger and are priced under $20.  
- Spring shows increased demand.  
- Data inconsistencies (e.g., negative fares, unrealistic coordinates) were flagged during preprocessing.

## 6. Recommendations: Data-Driven Business Suggestions

- Implement fare validation logic to flag extreme fare values (e.g., above $200).  
- Use stricter coordinate filters to correct invalid GPS points.  
- Run marketing campaigns in Spring to capitalize on high demand.  
- Tailor car allocation for 1–2 passenger rides.  
- Use seasonal data for demand forecasting and fleet optimization.