

INSY:5336-001 Python Programming
Uber Fare dataset Prediction

1. Background of study

Uber is a popular ride service in the United States, used for daily commutes, scheduled trips, and transportation when personal vehicles are unavailable. However, finding an affordable rate for an Uber ride can be a challenge, particularly during festivals or events when fares may increase. This study aims to conduct a trend analysis on Uber fare data over time to determine the busiest days of the month.

2. Motivation

Performing a trend analysis on Uber fare data helps to identify patterns and avoid booking a ride during busy times. The main objective is to conduct data analysis using visualization tools in Python and draw conclusions that aid decision-making.

3. Research Question

1. As this is Taxi fare data and we know there are many factors which affect the price of taxi like
 - Travelled distance.
 - Time of Travel
 - Demand and Availability of Taxi
 - Some special places are costlier like Airport or other places where there might be toll.
2. On different day and time there would be different price like during evening price would be more compared to afternoon, during Christmas price would be different and similarly on weekends price would be different compared to week days.

4. Dataset

The dataset contains the following fields:

- key - a unique identifier for each trip
- fare_amount - the cost of each trip in used.
- pickup_datetime - date and time when the meter was started.
- pickup_longitude - the longitude where the meter was started.
- pickup_latitude - the latitude where the meter was started.
- dropoff_longitude - the longitude where the meter was not engaged.
- dropoff_latitude - the latitude where the meter was not engaged.
- passenger_count - the number of passengers in the vehicle (Value entered by Driver)

5. Analysis:

Various data cleaning techniques will be implemented to achieve a consistent dataset for this study. Moreover, the Matplotlib library will be utilized to plot various visualization graphs and identify factors that impact the fare of an Uber ride.

6. Findings:

Through this analysis, I aim to address the research questions mentioned earlier and uncover patterns within the Uber fare dataset Prediction.

7. Discussion:

Predicting Uber fare dataset can be challenging due to the dynamic nature of the ride-hailing industry and external factors that can impact ride demand and pricing. Accurate fare predictions can benefit both riders and drivers and can also help ride-hailing companies adjust pricing models for increased revenue and customer satisfaction. To create accurate prediction models, advanced machine learning techniques can be utilized, but the accuracy of these models is dependent on the quality and completeness of the data.

8. Limitation:

There are several limitations to consider for an Uber fare dataset prediction project. Some of these limitations include:

- Incomplete or missing data, which can impact the accuracy of the prediction models.
- The dynamic nature of the ride-hailing industry, which can result in sudden changes to demand and pricing models.
- Limitations in the dataset, such as only including certain cities or time periods, which may not accurately represent the overall trends in the industry.
- The influence of external factors, such as events, weather conditions, and traffic, on ride demand and pricing.