# Taxi Trip Data Analysis

Author: Shlok Apte

Data Analyst Intern,

BITINFOTECH,Vadodara

## Abstract:-

Taxi trip datasets provide valuable insights into urban transportation patterns. The dataset contains information about various aspects of taxi trips, including pickup and drop-off times, trip duration, fare amount, passenger count, and payment types. The objective of this project is to perform Exploratory Data Analysis (EDA) to identify key trends and patterns that impact taxi services.

## 1.Problem Statement:-

Taxi trip demand is influenced by multiple factors such as time of the day, location, trip distance, weather conditions, and fare pricing. This analysis aims to uncover the important elements that govern taxi trips and how they impact urban mobility.

## 2. Introduction:-

The taxi trip dataset includes numerical, categorical, and temporal data. It contains columns such as pickup location, drop-off location, trip distance, fare amount, passenger count, and payment type. Our objective is to gain insights into ride patterns, peak demand hours, fare distributions, and the impact of different factors on taxi usage.

## 3. Steps Involved:-

### Data Wrangling

After loading the dataset, we performed various preprocessing steps such as handling missing values, dealing with outliers, and converting data types for better analysis. This step ensures data accuracy and reliability.

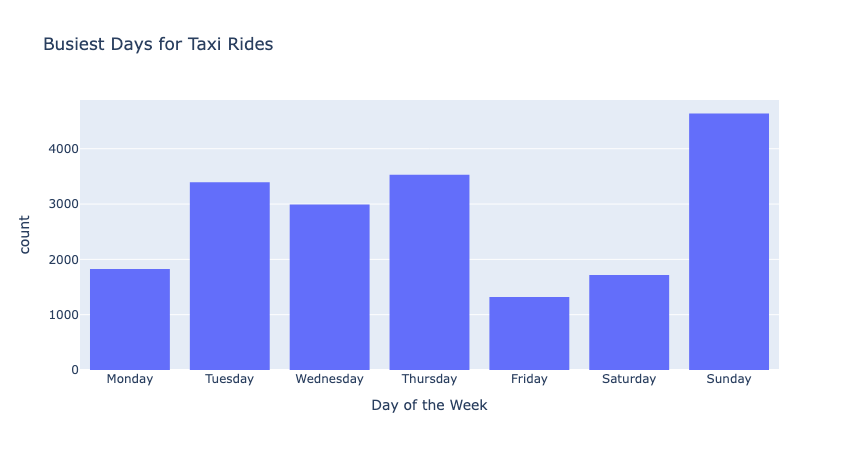
- Handling Missing Values: Missing values were replaced with appropriate estimates based on data trends.

- Outlier Treatment: Extreme fare values and unrealistic trip durations were handled using statistical techniques.

### Exploratory Data Analysis (EDA):-

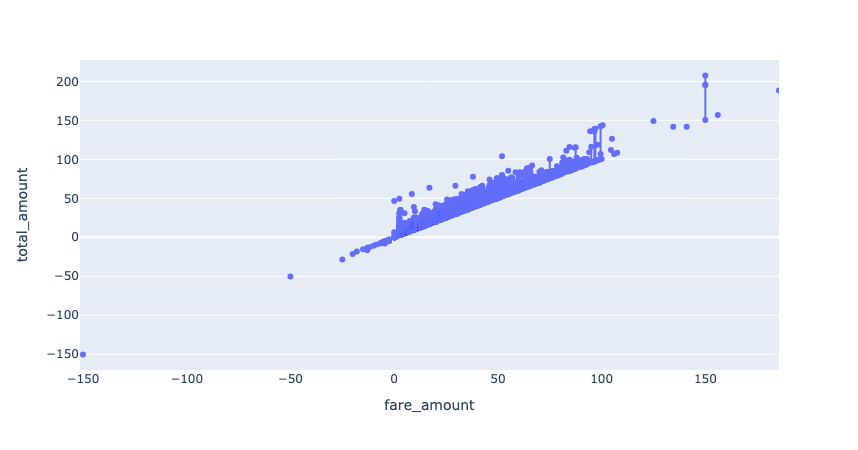
We performed various analytical techniques to extract insights from the dataset:

1. Trip Distribution Analysis: Analyzing trip frequency based on time of day, day of the week, and month.

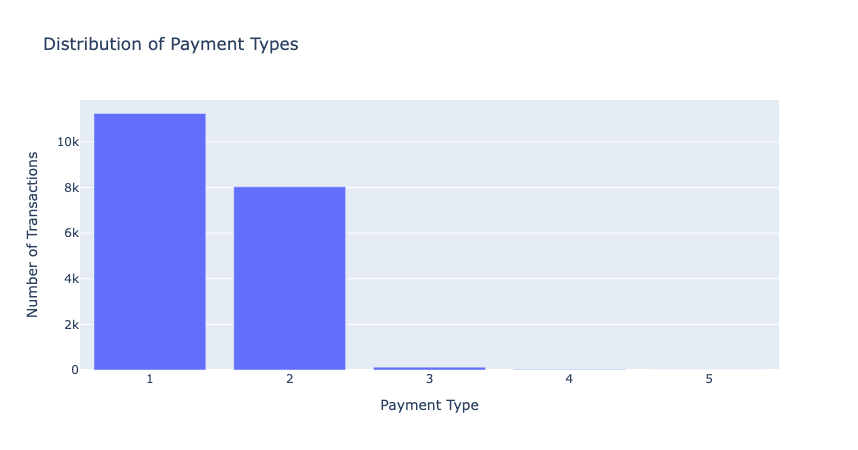




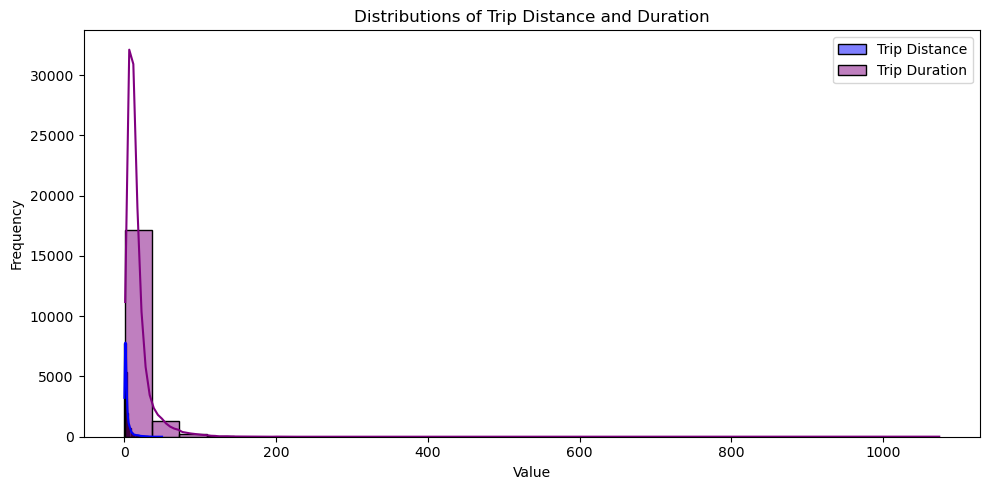
2. Fare Amount Analysis: Understanding the distribution of fares and factors affecting fare variations.



3. Payment Method Trends: Analyzing the preferred payment methods (1-cash vs.2-card transactions 3-Net Banking 4&5-Others).



4. Trip Duration and Distance: Studying how trip durations vary with distance traveled.



## 4. Key Insights

- Peak Demand Hours: Taxi demand is highest during morning and evening rush hours.  
- Fare Trends: Shorter trips have a higher fare per mile, while longer trips benefit from lower fare rates per mile.  
- Passenger Count Influence: Most trips have 1-2 passengers, while group rides are less frequent.  
- Payment Preferences: Credit card payments are more common than cash transactions.  
- Trip Duration Variation: Longer trips are often taken during non-peak hours, while short trips dominate during rush hours.

## 5. Conclusion:-

- Taxi trips are highly dependent on time and location, with rush hours experiencing maximum demand.  
- Digital payments have overtaken cash transactions as the preferred payment method.  
- Fare pricing models should consider optimizing rates based on demand fluctuations.  
- Understanding ride patterns can help taxi services optimize fleet management and improve customer satisfaction.

## 6. References:-

- Kaggle Datasets  
- Pandas & Matplotlib Documentation  
- Data analytics Tutorials from W3Schools, GeeksforGeeks,