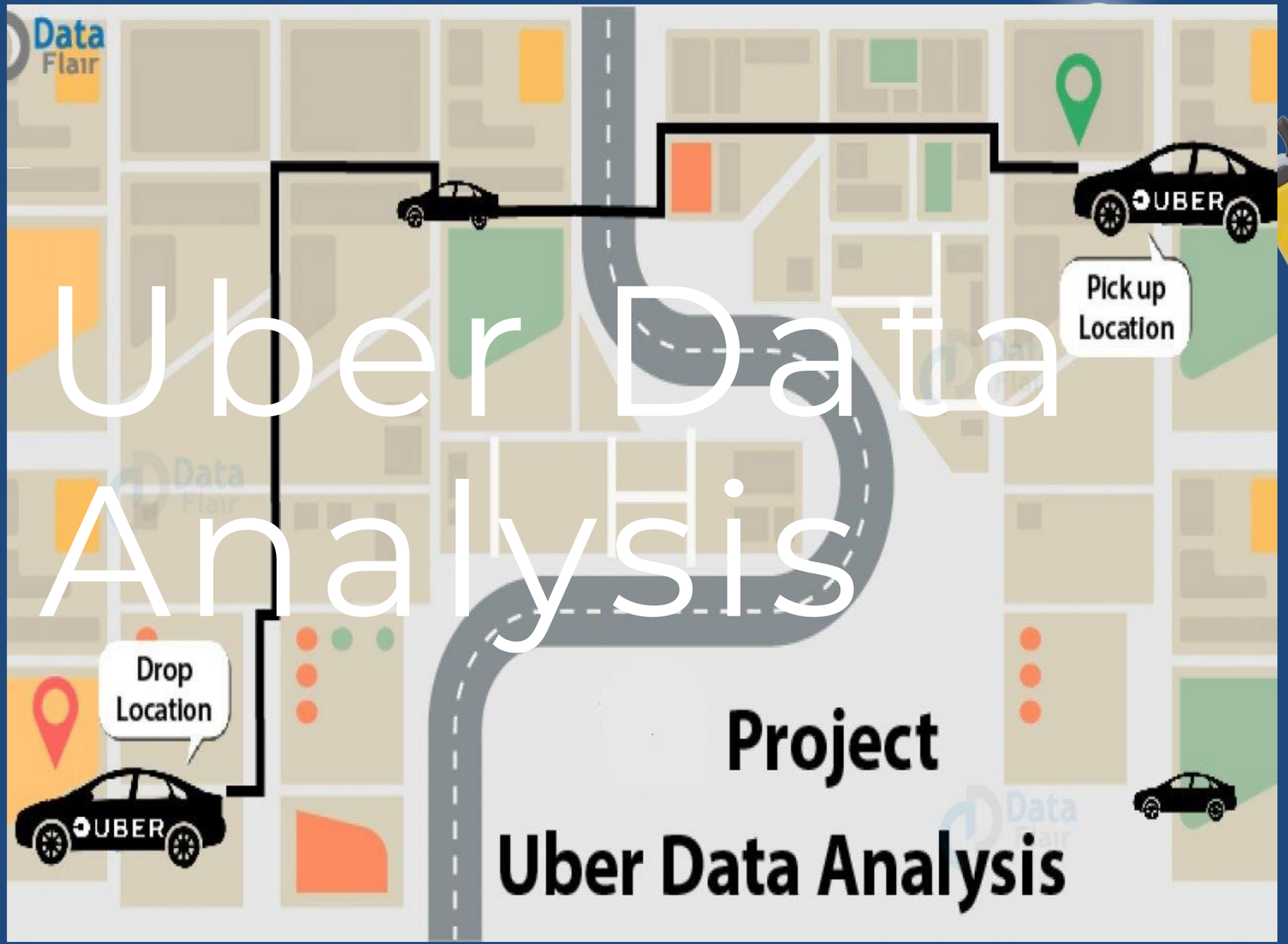
Uber Data Analysis



Group Members:-

1. Smit Gawand

2. Gaurav Yadav

3. Aayush Vishwakarma

4. VedanshuNikam

Introduction

In this analysis, we will be focusing on the use of machine learning algorithms to analyze Uber's data and identify patterns in the data that can help improve its services. The data includes information on Uber's trips, drivers, passengers, and ratings. The goal is to use machine learning algorithms to gain insights into the data that can help Uber optimize its operations, increase revenue, and improve customer satisfaction.

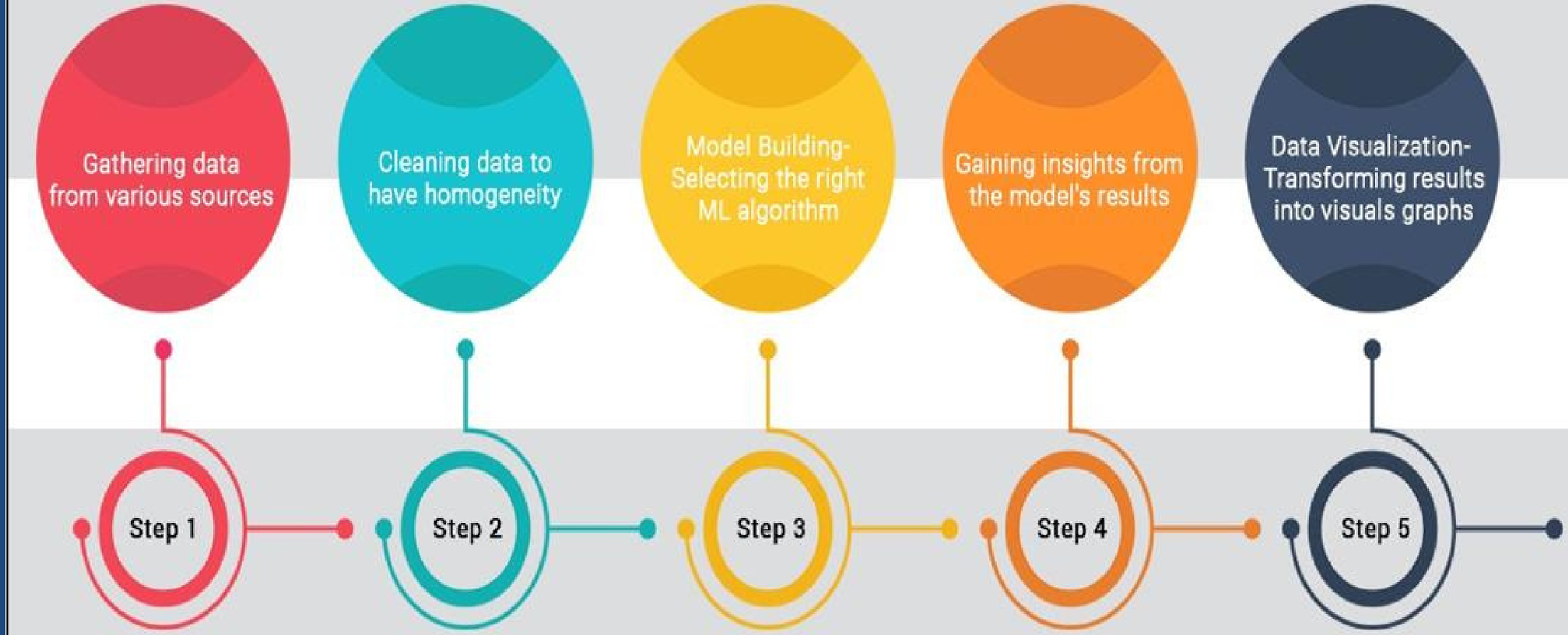
Objective

The objective is to first explore hidden or previously unknown information by applying exploratory data analytics on the dataset and to know the effect of each field on price with every other field of the dataset. Then we apply different machine learning models to complete the analysis. After this, the results of applied machine learning models were compared and analyzed on the basis of accuracy, and then the best performing model was suggested for further predictions of the label ‘Price’.

PROBLEM STATEMENT

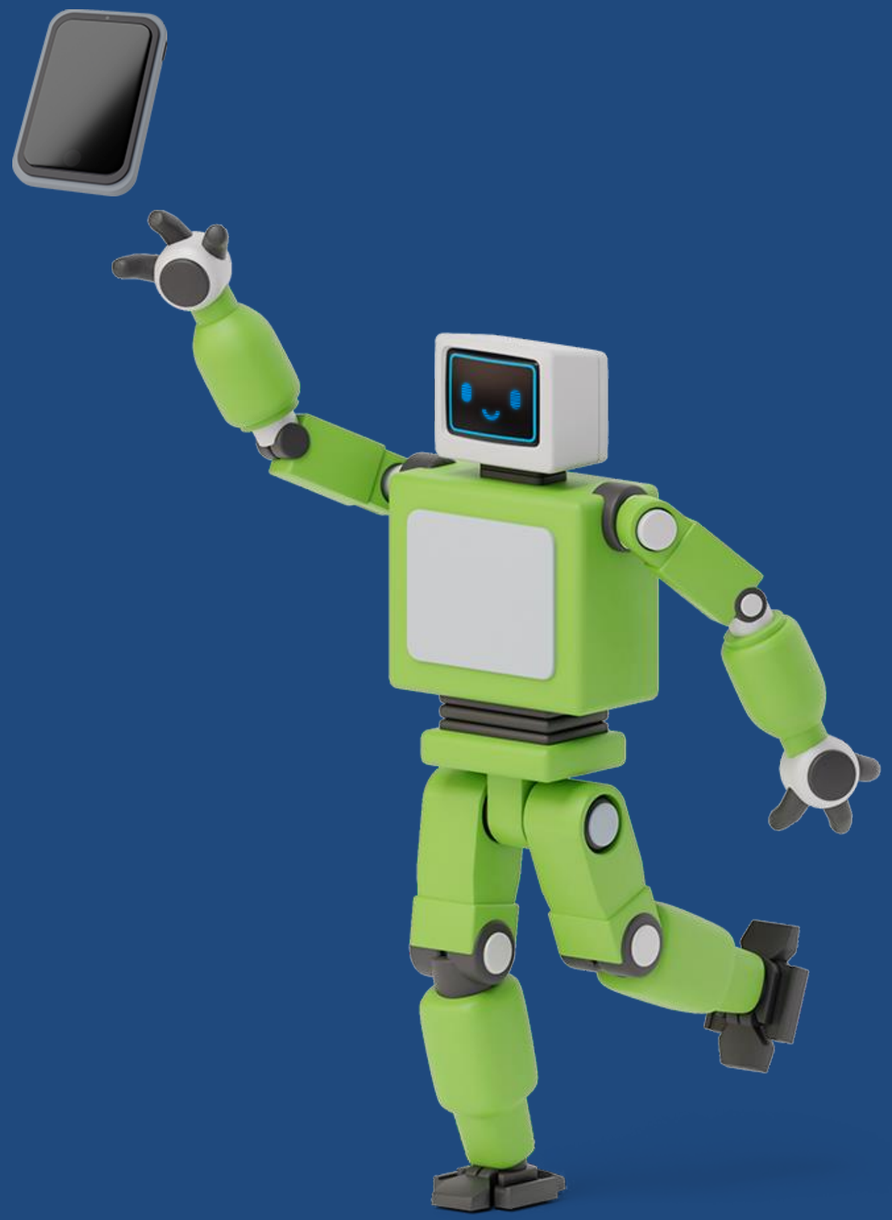
TO EXTRACT INSIGHTS AND MAKE PREDICTIONS FROM THE VAST AMOUNT OF DATA THAT UBER COLLECTS ON ITS USERS, DRIVERS, AND TRIPS. THIS DATA CAN BE USED TO ADDRESS VARIOUS BUSINESS CHALLENGES AND IMPROVE THE OVERALL EFFICIENCY AND EFFECTIVENESS OF UBER'S OPERATIONS

-THE MACHINE LEARNING PROCESS-

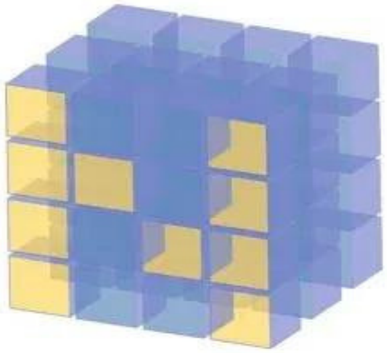


Algorithm :-

1. Installing The Libraries
2. Importing the dataset
3. Filling the Missing Values
4. Data Anaysis
5. Implementing Machine Learning Models
6. Predicting unseen Data
7. Concluding the Report



Libraries



NumPy

matplotlib

Pandas

scikit learn

Seaborn

Exploratory Data Analysis

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns to spot anomalies to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

It is a good practice to understand the data first and try to gather as many insights from it.

EDA is all about making sense of data in hand.

Applications

1. Prediting Demand
2. Driver Allocation
3. Pricing Optimization
4. Route Planning

Future scope

- Autonomous Vehicles
- Real Time Pricing
- Improvement Recommendation
- Enviromental Sustainability

Conclusion

we visualize the data by drawing various plots, due to which we understand that we don’t have any data for taxi’s price, also the price variations of other cabs and different types of weather. Other value count plots show the type and amount of data the dataset has. After this, we convert all categorical values into continuous data type and fill price Nan by the median of other values.

Thank You