**Business Requirements Specification(BRS)**

1. **Background of the project**

The project is about the world's largest taxi company Uber inc. Uber delivers service to lakhs of customers daily. Now it becomes really important to manage their data properly to come up with new business ideas to get best results. Eventually, it becomes really important to estimate the fare prices accurately.

**Products:**

* Our aim is to analyze the prices of these ride-sharing apps and try to figure out what factors are driving the demand.
* We have provided a large dataset as well as a mechanism to collect more data.
* With no public data of rides/prices shared by any entity, we tried to collect real-time data using Uber & Lyft api queries and corresponding weather conditions.
* We queried cab ride estimates every 5 mins and weather data every 1 hr.

1. **Goals of project**

* The Cab ride data covers various types of cabs for Uber & Lyft and their price for the given location. You can also find if there was any surge in the price during that time.
* Weather data contains weather attributes like temperature, rain, cloud, etc for all the locations taken into consideration.
* Reduce extra driving.
* Minimum waiting time

1. **Problem Statement**

Uber and Lyft's ride prices are not constant like public transport. They are greatly affected by the demand and supply of rides at a given time. So the demand depends onthe time of the day; times around 9 am and 5 pm should see the highest surges on account of people commuting to work/home. And the weather; rain/snow should cause more people to take rides.

NOTE: The date is for simulated rides with real prices i.e. how much would the ride cost IF someone actually took it. Uber/Lyft DO NOT make this data public and nor is the case in this dataset

1. **Solution / Proposed System**

**1. Data Preparation**

The data we used for our project was provided on the www.kaggle.com website. The original dataset contains 693071 rows and 57 columns which contain the data of both Uber and Lyft. But for our analysis, we just need the Uber data so we filter out the data according to our purpose and got a new dataset that has 322844 rows and 56 columns. The dataset has many fields that describe us about the time, geographic location, and climatic conditions when the different Uber cabs opted.

**2. Data Visualization**

Data visualization is a graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.For the same purpose, we have to import matplotlib and seaborn library and plot different types of charts like strip plot, scatter plot, and bar chart.

**3. Feature Engineering**

Feature engineering is the most important part of the data analytics process. It deals with, selecting the features that are used in training and making predictions. All machine learning algorithms use some input data to create outputs. This input data comprise features, which are usually in the form of structured columns. Algorithms require features with some specific characteristics to work properly. A bad feature selection may lead to a less accurate or poor predictive model. To filters out all the unused or redundant features, the need for feature engineering arises. It has mainly two goals:

* Preparing the proper input dataset, compatible with the machine learning algorithm requirements.
* Improving the performance of machine learning models.

**4. Modeling**

The process of modeling means training a machine-learning algorithm to predict the labels from the features, tuning it for the business needs, and validating it on holdout data. When you train an algorithm with data it will become a model. One important aspect of all machine learning models is to determine their accuracy. Now to determine their accuracy, one can train the model using the given dataset and then predict the response values for the same dataset using that model and hence, find the accuracy of the model.

* In this project, we use Scikit-Learn to rapidly implement a few models such as Linear Regression, Decision Tree, Random Forest.

1. **Users of project**
2. **Supply Service (for cabs)**
3. **Demand Service (for riders)**

**d) Requirements**

**Functional Requirements**

1. **Management of data**

We consider travel distance and travel time as the main features for predicting dynamic prices by analyzingthe associations between dynamic prices and different data features. A linear regression model is trained to make predictions of the dynamic data price.

Linear Regression is one of the machine learning algorithms where the outcome is estimated by the use of known parameters that are associated with output. Instead of trying to classify values into different groups, it is used to predict values within a continuous range. The known parameters are used to predict the unknown parameter or the result. When known and unknown parameters are plotted on the X and Y axes, it forms a continuous and steady slope. Our trained model predicts dynamic price (trip fare) with good precision and efficiency (i.e., 93.40%) by considering the features (travel distance & travel time) which are strongly correlated to dynamic price (trip fare).

1. **Problem Handling**

* To understand details of the Cab rides Dataset.
* To implement EDA on the Cab rides dataset.
* To examine Data Cleaning, Data Pre-Processing, and Data Visualization.
* To develop Machine Learning Algorithms and to check the accuracy for the Uber cab price prediction.

1. **MARKETING FULFILLMENT RESPONSE**

Uber service that uses dynamic pricing to balance supply and demand in an attempt to increase service quality. The research focuses primarily on exploratory data analysis (EDA), metric calculation, and development of model.

1. **Selling**

* Selling processes are responsible for managing car price and weather dataset for satisfaction of the customer.

Key Factors :

* Acquire Customer Data
* Customer preference

**Non-Functional Requirements**

1. **Performance Requirements**

* Comparison between Uber and same type of cabs.
* Variation of the car prices based on monsoon of the day.
* Variation of the car prices based on time.