***D Dynamic Pricing Prediction For Cabs Using***

***IBM Watson***

# Abstract

Many organizations do not have a direct role in travel and tourism but offer related products and services. Some examples would be offering travel insurance, parking facilities at airports, theatre and event tickets, car hire, and travel by rail or coach to airports, etc. at competitive rates.

There are various different forms of dynamic pricing:

1. Peak Pricing – This is a strategy that is common in transportation businesses. Airlines are a good example. Airlines often charge a higher price to travel during rush hour mostly on weekdays and sometimes on weekends.

2. Surge Pricing – Companies such as Uber respond dynamically to changes in supply and demand in order to price their services differently. Like most of us have noticed, this frequently happens on stormy evenings and nights when more people request for cabs. Taxify also not so long ago introduced dynamic pricing to ensure the drivers are encouraged to go online and offer services when the demand is high.

# INTRODUCTION

* 1. Overview

As explained in the abstract the workings of uber dynamic model and about the price prediction model to predict the price of the ride from given source to destination we use the data such as the distance between the source and destination. As weather plays a very important role in deciding the surge in the price of the cab, we take the weather report given for the respective day and by using this weather data we are able to predict the actual price for the given ride at a certain period of time. And with the help of linear and logistic regression we are able to visualize the data into pictures or graphs for better understanding and visualizing the estimation of the pricing with various factors.

The traﬃc also plays a major role in calculating the surge of price of the

ride with increase of the traﬃc the availability of the cabs becomes limited and when the demand for the cabs start to increase the service provider will not be able to provide the cabs this causes the surge in the price during the peak hours. The peak hours are usually calculated as the time when there is a large number of requests for the cabs and the price is increased during the peak hour. With the help of the driver’s data set as well as the customers dataset we are able to calculate the peak hour for each day.

* 1. Purpose

With the growth of combining the concept of ‘Big Data and Machine-Learning’ and introducing machine learning model for data analysis and the importance of data produced by the cabs on daily basis and how this data can be used by the machine learning to tell the consumer about the exact price of their ride before

starting the ride. This provides the consumer to make better choice of cab based on the price predicted by the Machine-Learning model. The proposed system uses the cab dataset and weather dataset to make predictions for each ride booked

by the customer.

# LITERATURE SURVEY

* 1. Existing problem

The author tells us the popularity of uber in the recent years and about the urban citizens who are beneﬁted by the uber. Later the author compares the diﬀerence between the competitive taxis and uber and deﬁnes new way of calling and also the new way of paying for cabs, the author also tells us about the importance of data produced by the cabs daily and also about the visualization and analysis of data. Aher that the author tells how the diﬀerent time and diﬀerent environments will have an eﬀect on passengers to make diﬀerent choices.

* 1. Proposed solution

The proposed system helps us for better predictions of cab fare from source the destination using the methods of Linear and Logistic regression. The model uses machine learning technique (Supervised learning) which helps to train the machine with labeled data that is already tagged with some predeﬁned class.

Then we test our model with some new unknown set of data and predict the price for them.

# THEORETICAL ANALYSIS

* 1. Project Flow

* + 1. Install Required Libraries.
    2. Data Collection.
       1. Collect the dataset or Create the dataset
    3. Data Pre- processing.
       1. Import the Libraries.
       2. Reading the dataset.
       3. Understanding Data Type and Summary of features.
       4. Take care of missing data & create columns.
       5. Data Visualization.
       6. Observing Target, Numerical and Categorical Columns
       7. Label Encoding
       8. Splitting Dataset into Train Set and Test Set
    4. ModelBuilding
       1. Training and testing the model Using Random Forest Regression
       2. Model Evaluvation
       3. Saving the Model
    5. Build Flask Application

* + - 1. Flask Structure
      2. Importing Libraries
      3. Loading flas and Assigning the model variable
      4. Routing the html page
      5. Final ui

1. Train The model in IBM

Hardware / Sofware designin

All the requirements of the project are listed below:

1. Google Colaboratory
2. Anaconda 3
3. Sklearn (Machine-Learning model)
4. Flask
5. IBM Watson Studio Python libraries:
6. Pandas
7. NumPy
8. Matplotlib.pyplot
9. Seaborn

# EXPERIMENTAL INVESTIGATIONS

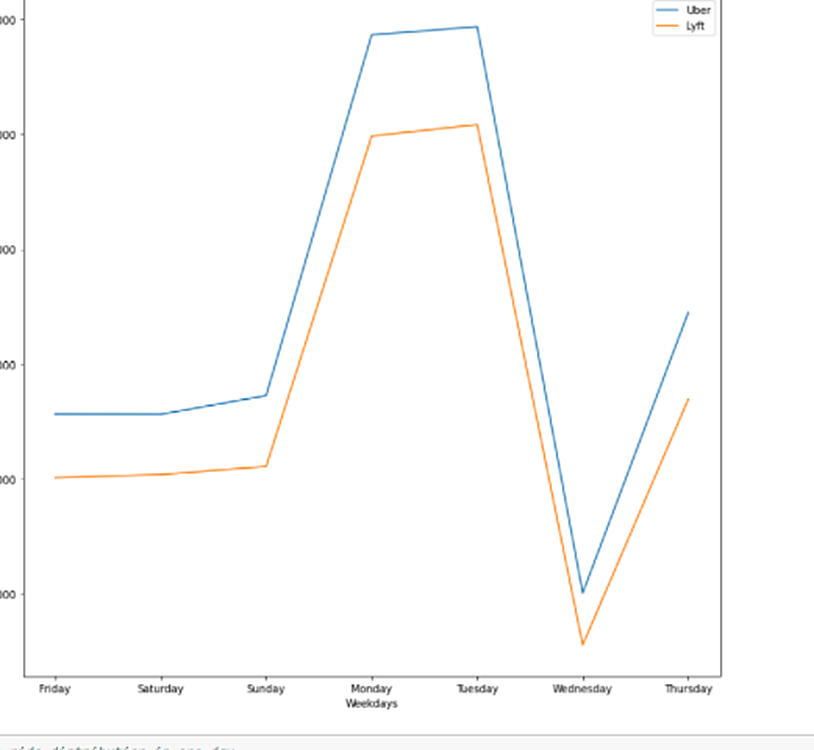
The code for the entire model is done in python the project contains of two parts. First part contains cleaning of data as well as data acquisition and the second part contains exploratory data analysis and implementing machine-learning algorithm. In the ﬁrst part we will load the data into the google colaboratory using the ‘wget’ command both the cab data and weather data is stored in the dropbox so that the data can be accessed from anywhere and at anytime once the data gets loaded into the program the size of the data is checked. Since the dataset contains large number of data i.e., around six lakh data we use memory reduce function to reduce the size of the data.

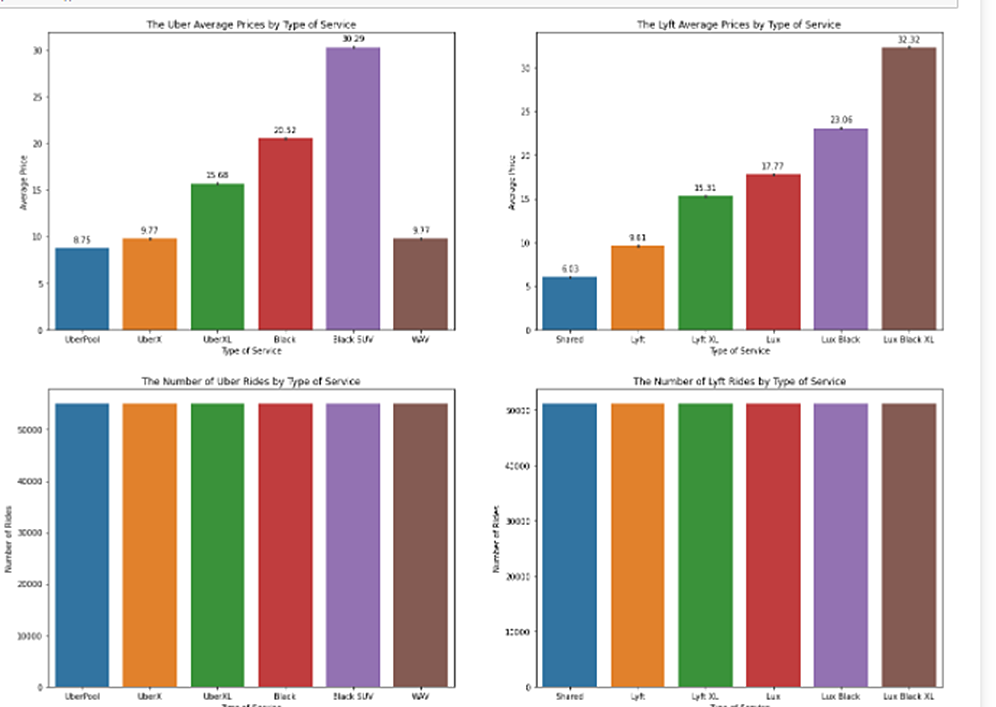
The memory reduce function helps us to reduce the size of the total dataset by compressing the original dataset, since the original dataset is compressed the operations carried on the dataset can happen much faster as the size of the dataset is now small. Next, we have to do the data cleaning for the compressed dataset. The data cleaning helps us to detect and correct the data which is inaccurate or null within the data set we can choose between replacing, modifying or cleaning the inaccurate or corrupt data in our dataset.

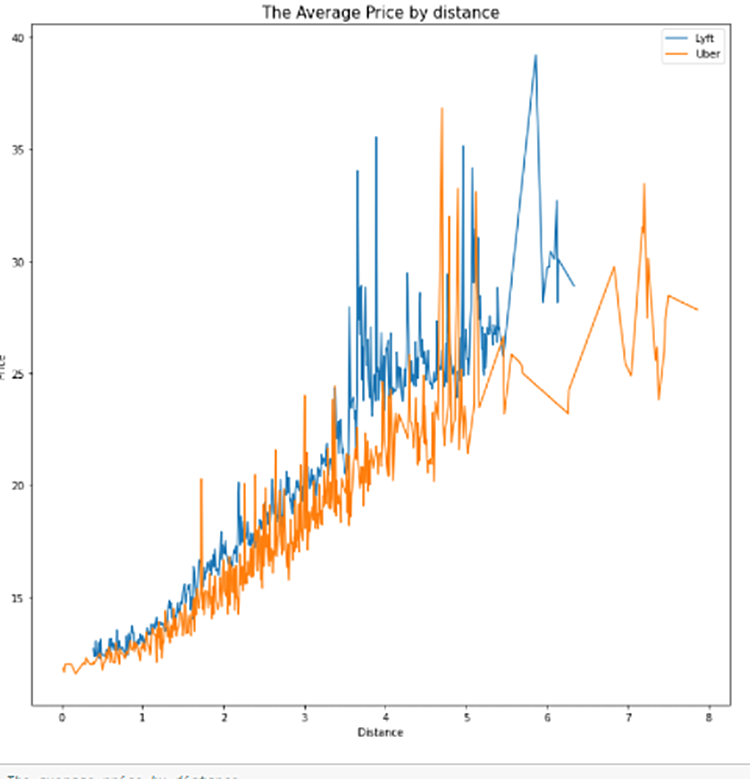
For example, in some of the entries we have distance travelled in the ride

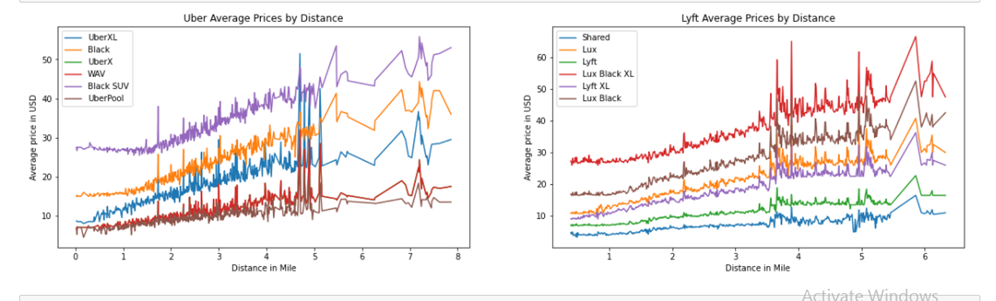
is zero this leads to confusion for machine-learning model in order to predict the accurate price for the cab ride so all such type of inaccuracies in the dataset are detected and has to be removed before going to the next process which is expl

* 1. Graphs









# FLOWCHART

|  |  |
| --- | --- |
|  |  |
|  |  |

1. **RESULT**

|  |  |
| --- | --- |
|  |  |
|  |  |

1. **ADVANTAGES & DISADVANTAGES**

This project can potentially be proﬁtable in the real market.

If a taxi driver could know in advance (and with precision) which boroughs or areas are going to have the biggest demand, he could optimise his workday by driving only around those areas. He could choose whether to earn more money in the same time or save that time for his family/personal life. Either way, it will improve his life.

# APPLICATIONS

Solving an existing problem.

There has been a lot of debate in the past regarding how Uber is literally eating the traditional street hail taxi market. Taxi drivers are afraid that they cannot compete with the kind of on-demand fare-adjusted service Uber provides, based in cufing-edge technology. I think that traditional taxi drivers could also make use of advance technology like this machine learning app in order to improve their service and proﬁtability.

# CONCLUSION

This project gives us basic understanding of how we can use machine learning in order to predict the cab fare from given source to destination before starting the cab ride. The model created is able to give us the predictions which are not exactly equal to the actual the price ﬂuctuation is around the diﬀerence of ten to twenty rupees compared to the actual price. Since the model is good but not the best, we can improve the predictions of the model by using the Fine-tuning technique. If ﬁne tuning is applied to the existing model, we are able to get higher accuracy

than the proposed model.

# BIBLIOGRAPHY

* **[www.wikipedia.org](http://www.wikipedia.org)**
* **[www.google.com](http://www.google.com)**
* **[www.github.com](http://www.github.com)**