Uber and Lyft Price Forecast

Vishesh Singh Thakur
50322513, vthakur@buffalo.edu

Ashwin Nair
50321006, anair3@buffalo.edu

Xuanyu Lu
50164188, xuanyulu@buffalo.edu

Group Number 9
Department of Industrial and Systems Engineering, University at Buffalo
May 2020

Abstract

The cab industry has risen to a new era in recent years. They have become a part of our day to day commutes and are pretty affordable for most of us. There are various factors that rule how much the will be the cost of a ride, some of the factors maybe the time, distance, cab type, vehicle size and the weather conditions. Here we analyse the time, distance and cab type factors to forecast the ride cost. The forecasting methods used are Moving average, Exponential Smoothing and Holt's double exponential smoothing. The results for all three forecasting methods are almost similar, hence we cannot conclude on any used forecasting method being better than the other one.

1 Introduction

The words like Uber and Lyft have become a part of the day to day life. The whole online cab booking industry has seen an up rise in stocks after the evolution of smart phones. The prices of Ubers and Lyfts are pretty affordable and they can be booked by only few clicks. Even if the prices of the cabs are very cheap but they are not at all constant. The prices vary a lot. The factors at which the cab fare depends on are the time of the day, distance of the ride, the cab type ordered, the vehicle size, weather conditions and number of rides available in the local area.

Using the above-mentioned factors, the cost of the ride can be forecast if the past data is provided to us. In this paper only three factors are considered. Distance of the ride, Cab type, and time of the day. All other constraints are relaxed and only these factors are analysed. The results after forecasting can gives an idea of how much the price of ride depends on the factors that are being used in the forecasting.

The Structure of the project is as follows:

- Data pre-processing and cleaning
- Creating different datasets for Uber and Lyft
- Applying the forecasting method

2 Data collection and processing

The data for this project is collected from Kaggle [1]. The dataset contains cab fares for Uber and Lyft for different distances and at different time stamps. The first task for the data cleaning was to convert the time stamp to date and time format. Since, there was no use for the date, the date time was converted to only time. Once, the time was sorted, the next task was to convert the whole data set into two parts. Uber dataset and Lyft dataset. This task was performed in Python using the Pandas library. After getting the two datasets, each data set was again divided into four datasets. The four datasets differentiated with respect to time. Each of those were datasets were considered as a period. Again, each period was divided into 7(in case of Lyft) or 8(in case of Uber) parts. Those divisions were done on the basis of travel distance. The figure 1 illustrates on what basis the dataset division is performed

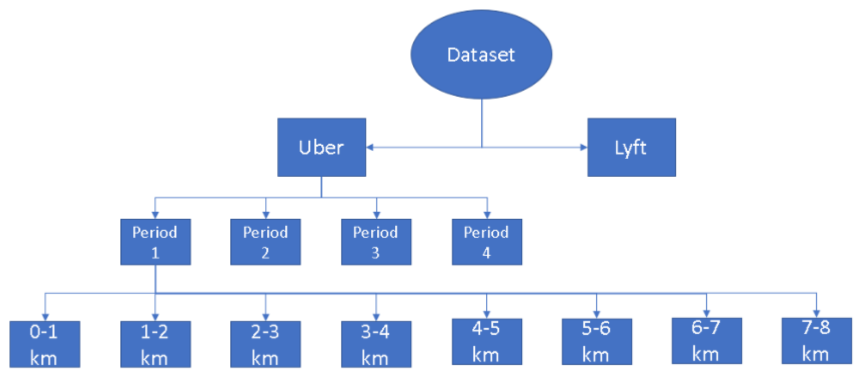

Figure 1: Dataset Distribution

3 Research Methodology

The factors used for the forecasting are as follows:

- Cab type
- Time at which the cab was booked
- Distance of the ride

Since different time periods have different effects on the average price values for a specific season and a specific period, the need to be deseasonlised to use for future forecasting. It is done by calculating the Seasonality Indices for each season and then dividing it by the known value for that season. The seasonality index can be calculated by [2]:

$$SI = \frac{Seasonal \quad average \quad for \quad a \quad particular \quad distance}{Overall \quad Average}$$

3.1 Moving Average Method

A moving average is a technique to get an overall idea of the trends in a data set; it is an average of any subset of numbers. The moving average is extremely useful for forecasting long-term trends. It is used to calculate it for any period of time. This is a widely used indicator in technical analysis that helps smooth out price action by filtering out the "noise" from random short-term price fluctuations. The following formula is generally used to calculate the Moving Average forecast. Here n represents the number of periods, m represents the number of periods for which the moving average forecast is to be calculated, F_{n+1} represents the forecast for the $(n+1)^{th}$ period, D_n represents the known Demand values. The following formula is [2] [3]

$$F_{n+1} = \frac{D_n + D_{n-1} + D_{n-2} + ... + D_{n-m+1}}{m}$$

For this project n is the period for which the forecast is calculated, m is taken as 2, i.e., the moving average is calculated for 2 periods, D is deseasonalised demand values. From observation the initial forecast values for the first two periods is taken as 20 and 20

3.2 Exponential Smoothing Method

Exponential smoothing is a time series forecasting method for univariate data, where the prediction is a weighted linear sum of recent past observations or lags.

Exponential smoothing forecasting methods are similar in that a prediction is a weighted sum of past observations, but the model explicitly uses an exponentially decreasing weight for past observations. Specifically, past observations are weighted with a geometrically decreasing ratio.

Single Exponential Smoothing requires a single parameter, called alpha α, also called the smoothing factor or smoothing coefficient. This parameter controls the rate at which the influence of the observations at prior time steps decay exponentially. Alpha is often set to a value between 0 and 1. Large values mean that the model pays attention mainly to the most recent past observations, whereas smaller values mean more of the history is taken into account when making a prediction. For this project, the α value was taken as 0.2. The following formula is used to calculate the forecast for single exponential smoothing [2] [4]:

$$F_{n+1} = \alpha D_n + \alpha(1-\alpha)D_{n-1} + \alpha(1-\alpha)^2 D_{n-2} + ... \qquad \alpha \in (0,1]$$

3.3 Holt's Double Exponential Smoothing Method

Holt, in 1957, extended simple exponential smoothing to allow the forecasting of data with a trend. This method involves a forecast equation and two smoothing equations (one for the level and one for the trend). This method tries to estimate a trend (i.e., a line). At each step, the newest observation is used to improve the estimate of the estimate of slope and intercept. The term G_n represents the slope at observation n, S_n represents the forecast for period n in period n-1, α represents the smoothing parameter for the level, and β represents the smoothing parameter for the trend. The initial values for S_0 and G_0 are calculated by regression analysis. F_n represents the forecast at period n. The following formulas are used for the calculating the Holt's forecast [2] [4]:

$$F_n = S_n + G_n$$

$$S_n = \alpha D_n + (1-\alpha)(S_{n-1} + G_{n-1} \qquad \alpha \in (0,1]$$

$$[G_n = \beta(S_n - S_{n-1}) + (1-\beta)G_{n-1} \qquad \beta \in (0,1]$$

Here, the initial S_0 and G_0 values were calculated using regression analysis, α value is taken as 0.2, and β value is taken as 0.2

4 Results

The figures 2 and figure 3 are the finals tables that represent the results for Uber and Lyft respectively. The first column represents i.e. the time frame.

Time frame is defined as follows:

- Season 1: The time period between 12:00 AM to 6:00 AM
- Season 2: The time period between 6:00 AM to 12:00 PM
- Season 3: The time period between 12:00 PM to 6:00 PM
- Season 4: The time period between 6:00 PM to 12:00 AM

The next column is the distance column. Distance 1 is the all the rides between distances 0 to 1 kilometres. Same way distance 2 is the collection of all the rides between 0 to 2 kilometres. Thus there are 8 distances for Uber and 7 distances for the Lyft. The next column defines average price of each period or season. Using the overall average, the seasonality index of each season is found. The next three columns are the forecasting which are also seasonalised. Judging from figures 2 and 3 that the moving average forecasting and exponential smoothing forecasting give out great results, but as Holt's double exponential smoothing forecasting gives the closest forecast value to the average prices for any time period of day corresponding to the travel distance. It was expected as Holt's forecasting follows the trend of the graph, and from the dataset, it can be observed that the for any particular season, the prices gradually increase as the distance increases, i.e., following a linear trend.

This forecasting model is implementable as is, given constraints that the data used for input and the forecasting is done for a specific weather condition (fall season, or spring season), with a single vehicle type for transportation, and equal availability of the cabs in any part of the city from which the data is collected. The solution can be considered realistic as the predicted values are very close to the expected values. Some of the outliers were expected as the dataset includes the prices for different weather and cab types.

	Season	Distance	Price	Periodic Average	Seasonality Index	Actual Forecast (Exponential)	Moving Average Forecast	Holts Forecast Actual
0	1	1	12.546811	12.472748	0.621859	12.437181	12.437181	12.508547
1	1	2	13.838275	13.786760	0.687372	13.771683	13.747447	13.833929
2	1	3	16.481474	16.517083	0.823499	16.514975	16.596980	16.573811
3	1	4	18.564800	18.566796	0.925693	18.556895	18.581470	18.604727
4	1	5	21.747914	21.752195	1.084508	21.742447	21.727578	21.779458
5	1	6	25.697973	25.570770	1.274893	25.560596	25.566879	25.584727
6	1	7	25.284983	24.208505	1.206974	24.224885	24.266336	24.237343
7	1	8	27.094382	27.582699	1.375202	27.842933	28.264564	27.895406
8	2	1	12.473332	12.472748	0.621859	12.522726	12.639653	12.545812
9	2	2	13.717776	13.786760	0.687372	13.831084	13.665045	13.852831
10	2	3	16.523272	16.517083	0.823499	16.543035	16.476147	16.559001
11	2	4	18.610081	18.566796	0.925693	18.591526	18.523824	18.598796
12	2	5	21.724947	21.752195	1.084508	21.785515	21.781625	21.784553
13	2	6	25.301379	25.570770	1.274893	25.595699	25.584561	25.582852
14	2	7	24.004687	24.208505	1.206974	24.176378	24.065823	24.144682
15	2	8	27.507042	27.582699	1.375202	27.506971	27.321293	27.446672
16	3	1	12.442161	12.472748	0.621859	12.438510	12.403136	12.403586
17	3	2	13.776237	13.786760	0.687372	13.749723	13.750948	13.706054
18	3	3	16.482886	16.517083	0.823499	16.479063	16.490527	16.425255
19	3	4	18.623376	18.566796	0.925693	18.524918	18.540490	18.465682
20	3	5	21.612814	21.752195	1.084508	21.726202	21.762820	21.665364
21	3	6	25.654535	25.570770	1.274893	25.513555	25.527807	25.447618
22	3	7	24.250633	24.208505	1.206974	24.181032	24.170596	24.130670
23	3	8	27.591522	27.582699	1.375202	27.567258	27.651877	27.526338
24	4	1	12.428687	12.472748	0.621859	12.467960	12.485595	12.456590
25	4	2	13.814753	13.786760	0.687372	13.772786	13.764614	13.765294
26	4	3	16.580699	16.517083	0.823499	16.510396	16.504677	16.508655
27	4	4	18.468929	18.566796	0.925693	18.575085	18.621401	18.582873
28	4	5	21.923104	21.752195	1.084508	21.737032	21.736756	21.749949
29	4	6	25.629193	25.570770	1.274893	25.596693	25.603833	25.623588
30	4	7	23.293716	24.208505	1.206974	24.239200	24.331264	24.273743
31	4	8	28.137850	27.582699	1.375202	27.402220	27.093063	27.405190

Figure 2: The forecast of Uber prices

	Season	Distance	Price	Periodically Average	Seasonality Index	Actual Forecast	Moving Average Forecast	Holts Forecast Actual
0	1	1	12.881078	12.793480	0.601537	12.030744	12.030744	13.184456
1	1	2	14.637059	14.587945	0.685911	13.912145	13.718224	14.917713
2	1	3	18.352055	18.356293	0.863095	17.688356	18.450037	18.627525
3	1	4	20.856117	20.970526	0.986014	20.359108	21.003407	21.121356
4	1	5	25.337677	25.250590	1.187259	24.634072	25.178796	25.239854
5	1	6	27.250277	27.025346	1.270706	26.516108	26.998229	26.901477
6	1	7	34.166667	29.891667	1.405478	29.490825	30.067607	29.699793
7	2	1	12.750291	12.793480	0.601537	13.022169	13.761559	13.113684
8	2	2	14.587304	14.587945	0.685911	14.786708	15.606479	14.876381
9	2	3	18.351651	18.356293	0.863095	18.556217	18.324905	18.639728
10	2	4	20.961738	20.970526	0.986014	21.152183	20.967414	21.207620
11	2	5	25.377536	25.250590	1.187259	25.423461	25.242107	25.439870
12	2	6	27.010326	27.025346	1.270706	27.200537	27.087618	27.172327
13	2	7	32.500000	29.891667	1.405478	30.043361	29.958499	29.964490
14	3	1	12.908812	12.793480	0.601537	13.068690	13.348102	13.062019
15	3	2	14.582103	14.587945	0.685911	14.865296	15.290168	14.875410
16	3	3	18.344070	18.356293	0.863095	18.634020	18.435357	18.649821
17	3	4	20.976610	20.970526	0.986014	21.221557	20.959345	21.228447
18	3	5	24.958853	25.250590	1.187259	25.493869	25.245846	25.479284
19	3	6	27.350096	27.025346	1.270706	27.171200	26.873146	27.113718
20	3	7	27.766667	29.891667	1.405478	30.092563	29.898584	30.002393
21	4	1	12.633740	12.793480	0.601537	12.680369	12.415602	12.594396
22	4	2	14.545313	14.587945	0.685911	14.448334	13.978343	14.308876
23	4	3	18.377395	18.356293	0.863095	18.205024	18.214872	17.999740
24	4	4	21.087639	20.970526	0.986014	20.837098	20.951938	20.592599
25	4	5	25.328295	25.250590	1.187259	25.150265	25.335612	24.870095
26	4	6	26.490687	27.025346	1.270706	26.956078	27.142393	26.688017
27	4	7	25.133333	29.891667	1.405478	29.712102	29.641977	29.435018

Figure 3: The forecast of Lyft prices

5 Conclusion and Future work

In the end, it can be concluded that all the forecasting methods that have been used give a pretty similar result. But there are a certain number of relaxations that have been used hence, making the results very compromised. So, in the future work, the relaxed constraints can be considered and make the forecasting model even more efficient and realistic. Also, for future work rather than dividing the dataset into time periods a time series model can be implemented, and for a better data source an API can be used for constantly updating data so that if there is any kind of change in the market prices the model can adjust accordingly.

References

[1] Uber and lyft cab prices, 2019.

[2] A Ravi Ravindran and Donald P Warsing Jr. *Supply chain engineering: Models and applications*. CRC Press, 2012.

[3] Wikipedia contributors. Moving average — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Moving_average&oldid=952740050`, 2020. [Online; accessed 14-May-2020].

[4] Wikipedia contributors. Exponential smoothing — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Exponential_smoothing&oldid=954453668`, 2020. [Online; accessed 14-May-2020].