**Project Title**

**UBER AND LYFT CAB PRICES**



**Under the guidance of**

**Prof Shilpa Balan Ph.D.**

**By,**

**Sowmya Mareedu**

**Data Set URL**

<https://www.kaggle.com/ravi72munde/uber-lyft-cab-prices?select=cab_rides.csv>

**About the dataset:**

* This is a dataset hosted by the city of New York.
* Dataset of cab rides collected for a week in Nov - Dec '18. Collected at a regular interval of 5 mins. Dataset of cab rides collected for a week in Nov - Dec '18. Collected at a regular interval of 5 mins.

**Dataset details:**

|  |  |
| --- | --- |
| Size | 84.65 MB |
| Number of columns | 18 |
| Number of rows | 6094 |
| Original file format | CSV |

**Data fields Table:**

|  |  |  |
| --- | --- | --- |
| COLUMN NAMES | COLUMN DETAILS | |
| DISTANCE | Distance of travel from source to destination | |
| CAB\_TYPE | Type of the cab UBER and LYFT | |
| TIME\_STAMP | Time taken or time of ride | |
| DESTINATION | End of the ride | |
| SOURCE | Where Ride started from | |
| PRICE | Cost of the ride | |
| SURGE MULTIPLIER | How much your base fare will be multiplied.  So, a fare that is usually 10$ would be 18$ if surge is 1.8x. | |
| PRODUCT\_ID | A unique ID of each ride | |
| NAME | Name of the cab service type | |
| TEMPERATURE | Temperature at the time of ride | |
| LOCATION | Location of the ride | |
| CLOUDS | Clouds that produce rain and snow fall into this category. | |
| PRESSURE | A level in atmosphere | |
| RAIN | Rainfall like 10 MM, 20 MM.., |

**Data Visualizations**

1. **What is the average price to the average distance ?**

A screenshot of a video game

Description automatically generated

A close up of a map

Description automatically generated

**Categories used:**

* Dual Axis chart
* Groups

**Analysis:**

The above visualization clearly shows that the Average price and Average distance using the Dual axis chart to the particular destination. As the dual axis is synchronized, the readability is clearer. In this I have created a group with only one value Boston University and There were around 11 “Other” destinations such as “Black-Bay”, “Bacon-hill” etc. In the above visual I have kept a filter of Destination(group) with including other So that we can see all the destinations. If I keep the destination(group) filter to only one group then you could see only Boston university. These properties were grouped under one and named as “Others”.

1. **What are the TOP 5 destinations having the costliest price?**

A screenshot of a computer

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Categories used:**

* Ranks
* Reference line

**Analysis:**

The above bar chart visualization shows the top 5 destinations which are having the highest price followed by other destinations. Only the top 12 ranks are being shown which have the highest to lowest price. A table calculation has been done to assign rank to the destination based on average price . A reference “Average” line is also being displayed which is closed to 17$. That average line is reference line. The average price to go to any destination is 16.4$.

1. **What is the Avg price of Uber and Lyft in rain?**

A screenshot of a computer

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Categories used:**

* Box and Whisker plots

**Analysis:**

The above visualization shows the average price of the cab type based on rain. The pattern is very erratic. Lyft price is more when compared to uber even when the rain is high. The above Box and Whisker plot provides a clear statistic of the price in rain. The median of the average price is $16.5. The median for the rain is 25.7mm. The circle indicates the Uber & Lyft in a sets model.

1. **What is average surge multiplier of each cab type ?**

A screenshot of a computer

Description automatically generated

A screenshot of text

Description automatically generated

**Categories used:**

* Histogram
* Reference Line

**Analysis:**

The above chart is a histogram. It shows the percentile of average surge multiplier on uber and lyft. The bins size used in the chart is 0.372 The chart clearly indicates that whenever the surge multiplier is increasing the price of cab is also increasing. Uber doesn’t do surge pricing frequently so I showed the chart only for lyft. The color gradient shows the average price which is between $16.58- $39.05.The average lines shows the average percentile of surge multiplier which is 50%.

1. **How many number of rides were done by uber and lyft based on cab category ?**

A screenshot of a computer

Description automatically generated

A screenshot of a social media post

Description automatically generated

**Categories used:**

* Sets

**Analysis:**

The above bar chart clearly shows the sets. I have created two sets one is uber set and the other one is lyft set then I created a combined set with both uber & lyft. There are 2 check boxes for IN and OUT. The OUT value is the maximum value and and IN is the inner values of data.

1. **What is the average price of cab based on the average distance to the location at particular date and time ?**

**A screenshot of a computer

Description automatically generated**

A screenshot of a cell phone

Description automatically generated

**Categories used:**

* Parameters [Date & Time]

**Analysis:**

The above chart shows the average price of cab based on average distance to the location based on particular destination. But, I want to see the price at particular date and time so for this, I have attached parameters to the existing chart which are Date and Time. This date and time shows the range between minimum to maximum. These cab rides was between Nov 2018- Dec 2018.

**Dashboard**

**A screenshot of a cell phone

Description automatically generated**

**Categories used:**

|  |  |  |
| --- | --- | --- |
| Visualization Type | Visualization Title | Categories Used |
| Already used visuaalizations | Average price of cab to average distance. | * Dual axis chart * Groups |
| Top 5 destinations having the costliest price. | * Rank * Reference line |
| Average price of uber and Lyft in rain. | * Box-and-whisker plots |
|  | Average surge multiplier of each cab type. | * Histogram |
|  | Number of rides based on cab category. | * Sets |
| New Visualization | Average price at particular date and time. | * Parameters(Date & Time) |

**Analysis:**

A dashboard is a consolidated display of many worksheets and related information in a single place. It is used to compare and monitor a variety of data simultaneously. The different data views are displayed all at once.

The above dashboard provides a summary of the Uber and lyft prices.

In the above dashboard, these six visualizations have already been used and explained in the “Data Visualizations” section:

* Average Price based on distance and rain
* Average surge multiplier of the cab type
* Top 5 costliest destinations.
* Number of rides based on cab type

The new visualization added to the dashboard is:

* Average price at particular date and time

For this visualization, I added the parameters (Date and time ) the date has been in Automatic format “Month Date Year”. The visualization shows the first and last dates and time of the cab ride period. It can be seen that maximum to minimum cab prices in particular time period which is from Nov 2018- Dec 2018.

**Story Telling**

The main objective of the project is to understand and visualize prices of Uber and Lyft based on several factors, 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 what exactly drives this demand? The first guess would be the time of the day; times around 9 am and 5 pm should see the highest surges on account of people commuting to work/home. Another guess would be the weather; rain/snow should cause more people to take rides.

The data is approx. for a month of Nov '18 (I didn't consider this as a time-series problem so did not worry about regular interval. The chosen interval was to query as much as data possible without unnecessary redundancy. So, data can go from end week of Nov to few in Dec)

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.  
I did data cleaning process and combined cab ride data and Weather data which contains weather attributes like temperature, rain, cloud, etc for all the locations taken into consideration.

**Content:** Dataset has one file with 13 columns and more than 30000+ rows. This data file includes Our aim was to try to analyze the prices of these ride-sharing apps and try to figure out what factors are driving the demand.

Some of the attributes we are going to use for analysis are Price, temperature, Cab Name, Distance, Time stamp, Destination, source, price, Cab Type.

A quick glance at the data shows that there are:

* Around 7000 unique listings of pricing
* Over 1 million reviews have been written by riders since then.

1. Price and Distance -

A screenshot of a cell phone

Description automatically generated

Average price and average distance to the particular destination. In most cities, your cost is calculated up front, before you confirm your ride. In others, you will see an estimated fare range.

Here are some fees and factors that can affect your price:

### Base rate:

The base rate is determined by the time and distance of a trip.

### Booking feel:

In your city, a flat fee might be added to each trip. It helps support operational, regulatory, and safety costs.

### Busy times and areas:

When there are more riders than available drivers, prices may temporarily increase until the marketplace is rebalanced.[1]

1. Expensive destinations-

A screenshot of a cell phone

Description automatically generated

So, after the average price we are going to see the highest - lowest prices to particular destinations. The analysis is showing the top 12 ranks to particular destinations.

1. Average surge multiplier-

A screenshot of a social media post

Description automatically generated

The above analysis shows the surge pricing. The average surge percentile of lyft is 50%.

The average price bar is between $16.58-$39.05.

How surge pricing works is:

1. **Demand for rides increases -**

There are times when so many people are requesting rides that there aren’t enough cars on the road to help take them all. Bad weather, rush hour, and special events, for instance, may cause unusually large numbers of people to want to request a ride with Uber all at the same time.

1. **Prices go up -**

In these cases of very high demand, prices may increase to help ensure that those who need a ride can get one. This system is called surge pricing, and it lets the Uber app continue to be a reliable choice.

1. **Riders pay more or wait -**

Whenever rates are raised due to surge pricing, the Uber app lets riders know. Some riders will choose to pay, while some will choose to wait a few minutes to see if the rates go back down.

Surge prices are calculated as,

When prices are surging, you’ll see a multiplier to the standard rates on the map. For example, you might see surge at 1.8x or 2.5x. This is how much your base fare will be multiplied by, so a fare that is usually $10 would be $18 when it’s at 1.8x surge. Uber’s fee percentage does not change during surge pricing.[3]

Because rates are updated based on the demand in real time, surge can change quickly. Surge pricing is also specific to different areas in a city, so some neighborhoods may have surge pricing at the same time that other neighborhoods do not.

Surge pricing on the map shows the price that will apply to riders using the Uber app in that area. The rider’s location determines the amount of surge pricing on a trip, not the driver’s location.

With surge, or so-called "prime" price hikes, each company tackles the issue differently.[3]

* Uber uses a multiplier surge pricing model, which simply adds the surge rate to the price of the ride, often doubling the price of the ride.
* Lyft uses a surge pricing model that uses a percentage-based formula. For example, in rush hour conditions Lyft passengers may see their ride cost boosted by 50%, making a $20 ride cost $30, instead.

1. Prices in rain-

A screenshot of a cell phone

Description automatically generated

Using all Lyft and Uber rides in New York City, we show that the number of Uber and Lyft rides is significantly correlated with whether it rained. The number of Uber (Lyft) rides per hour is about 22 (19)% higher when it is raining, while the number of Uber rides per hour increases by 15% in rainy hours-suggesting that surge pricing (prime time) encourages an increase in supply.

We show that while the number of taxi rides, passengers and fare income all significantly decreased after Uber entered the market in May 2011, taxis do not respond differently to increased demand in rainy hours than non-rainy hours since the entrance of Uber. Last, we test whether Lyft’s entry in the market affected Uber. [2]

Our estimates suggest that Uber was still growing after Lyft entered the market, but that Uber rides during rainy hours decreased by about 6%. Our findings suggest that dynamic pricing make Lyft and Uber drivers compete for rides when demand suddenly increases, i.e., during rainy hours

1. Total cab rides-

A screenshot of a computer

Description automatically generated

This analysis shows cab type names with In and out records of cab names. Here are few cab names explained below,

* **Uber X** is the cheapest and most common platform offered. Can take up to 4 passengers and the type of vehicle may vary anything from a Toyota Prius to a Ford Explorer and everything in between.
* **Uber XL** is the next level up and is used when you need more room, must be a vehicle that can seat up to 6 passengers, usually a larger SUV or van of some type. This obviously costs more than an Uber X. Used mostly for groups headed out to party or a family headed to the airport or just anyone who wants more room.
* **Uber Select** - Nicer, newer vehicles that don't have to be black but must have leather seats. Costs more than Uber X but less than Black car. Above Uber X but less expensive than Black Car.
* **Uber SUV** - Newer SUV's, must be black and have leather seats, drivers dress more professionally. Cadillac Escalade, Toyota Sequoia, Lincoln Navigator, GMC Yukon, etc.
* **Lyft Lux**: High-end, four seats, the perfect way to add a little luxury and arrive in style
* **Lux Black:** Premium black car service limited to only the most luxurious makes and models
* **Lux Black XL:** Premium black SUV service with space for parties up to six passengers[1]

1. Average price in time period-

A screenshot of a cell phone

Description automatically generated

Average prices of uber and lyft in between particular date and time.Here are the notable differences:

* Uber includes a button to let you split the cost of the fare with multiple riders, which is handy when you’re out with a group. Lyft discontinued its fare-splitting feature last year and said it planned to release an improved version this year.
* In addition to UberPool, Uber includes a car-sharing feature called Express Pool, where you agree to walk a short distance from your actual location, like to a street corner, and be dropped off somewhere close to your destination. In other words, you save money by walking a bit to make the route more efficient for the driver. Lyft is currently testing a similar feature in three cities. (This isn’t a game-changing feature: In my tests, Uber’s Express Pool option was not available in many locations.)
* Most important, Uber lets people request wheelchair-accessible cars in about 15 of its top cities. In my tests in New York, wheelchair-accessible cars were available in many areas. Lyft, which declined to comment for this column, has a very limited wheelchair-accessibility program.I wasn’t able to summon one here.
* In terms of features, reward programs and availability, Uber has an edge over Lyft.
* Even if you prefer Lyft in your hometown for ethical reasons, it may be practical to keep the Uber app on your phone in case you travel somewhere like London, Paris or Tokyo.[3]

**References:**

1. Uber VS Lyft [ Edelstein, K. (2019, February 07)]

Web site name- UBER

Retrieved from, <https://www.uber.com/us/en/drive/uber-vs-lyft/>

1. Uber and Lyft drivers pricing wages [John Edward ( 15 January 2018)]

Website name – The Guardian

Retrievedfrom,<https://www.theguardian.com/technology/2019/apr/18/uber-lyft-drivers-surge-pricing-wages>

1. Uber/Lyft surge pricing

Website name- LYFT

Retrieved from, <https://help.lyft.com/hc/en-us/articles/115012923147-Lyft-Lux-Lux-Black-and-Lux-Black-XL-rides-for-drivers>