



# UBER VS LYFT

Comparison Between Rivals in Boston, MA

A large, white, L-shaped graphic is positioned on the left side of the slide. It consists of a vertical bar and a horizontal bar that meet at a right angle, forming an 'L' shape. The background is a solid dark gray, and the text is in a clean, white, sans-serif font.

# WHERE DO I FIND THE DATA SOURCE ?

---

Find the data source from Kaggle, and they had a great source that predicts the cab prices for Uber vs. Lyft in Boston, MA.

```
In [3]: # Clean up the dataframe
# Dropping all NaN values
# Renaming the columns
# Converting time_stamp to date format

clean_prices = cab_prices_df.dropna(how='any')
clean_prices_df = pd.DataFrame(clean_prices)
clean_prices_df.head()

clean_prices_rename = clean_prices_df.rename(columns={"distance": "Distance", "cab_type": "Type of Cab",
"timestamp": "Date", "source": "Source",
"price": "Price", "surge_multiplier": "Surge Multiplier",
"id": "User ID", "product_id": "Service Type", "name": "Vehicle Type"})

clean_prices_rename_df = pd.DataFrame(clean_prices_rename)

clean_prices_rename_df['Date'] = pd.to_datetime(clean_prices_rename_df['Date']/1000, unit='s')
clean_prices_rename_df.head()
```

Out[3]:

	Distance	Type of Cab	Date	destination	Source	Price	Surge Multiplier	User ID	Service Type	Vehicle Type
0	0.44	Lyft	2018-12-16 09:30:07.890000105	North Station	Haymarket Square	5.0	1.0	424553bb-7174-41ea-aeb4-fe06d4f4b9d7	lyft_line	Shared
1	0.44	Lyft	2018-11-27 02:00:23.677000046	North Station	Haymarket Square	11.0	1.0	4bd23055-6827-41c6-b23b-3c491f24e74d	lyft_premier	Lux
2	0.44	Lyft	2018-11-28 01:00:22.197999954	North Station	Haymarket Square	7.0	1.0	981a3613-77af-4620-a42a-0c0866077d1e	lyft	Lyft
3	0.44	Lyft	2018-11-30 04:53:02.749000072	North Station	Haymarket Square	26.0	1.0	c2d88af2-d278-4bfc-a8d0-29ca770cc5512	lyft_luxsuv	Lux Black XL
4	0.44	Lyft	2018-11-29 03:49:20.223000050	North Station	Haymarket Square	9.0	1.0	e0126e1f-8ca9-4f2e-82b3-50505a09db9a	lyft_plus	Lyft XL

```
In [4]: # Clean up the dataframe
# Dropping all NaN values
# Renaming the columns
# Converting time_stamp to date format

clean_weather = weather_df.dropna(how='any')
clean_weather_df = pd.DataFrame(clean_weather)

clean_weather_rename = clean_weather_df.rename(columns={'temp': 'Temperature in Fahrenheit', 'location': 'Location',
'clouds': 'Cloudiness', 'pressure': 'Pressure', 'rain': 'Rainfall in inches',
'time_stamp': 'Date', 'humidity': 'Humidity', 'wind': 'Wind Speed'})

clean_weather_rename_df = pd.DataFrame(clean_weather_rename)

clean_weather_rename_df['Date'] = pd.to_datetime(clean_weather_rename_df['Date'], unit='s')
clean_weather_rename_df.head()
```

Out[4]:

	Temperature in Fahrenheit	Location	Cloudiness	Pressure	Rainfall in inches	Date	Humidity	Wind Speed
0	42.42	Back Bay	1.0	1012.14	0.1228	2018-12-16 23:45:01	0.77	11.25
1	42.43	Beacon Hill	1.0	1012.15	0.1846	2018-12-16 23:45:01	0.76	11.32
2	42.50	Boston University	1.0	1012.15	0.1089	2018-12-16 23:45:01	0.76	11.07
3	42.11	Fenway	1.0	1012.13	0.0969	2018-12-16 23:45:01	0.77	11.09
4	43.13	Financial District	1.0	1012.14	0.1786	2018-12-16 23:45:01	0.75	11.49

# CLEANING UP THE DATA

```
In [9]: # merge the datasets to refelect same time for a Location, Source, and Date in both two data
clean_prices_rename_df['merge_date'] = clean_prices_rename_df.Source.astype(str) + " - " + clean_prices_rename_df.Date.dt.date.astype(str)
clean_weather_rename_df['merge_date'] = clean_weather_rename_df.Location.astype(str) + " - " + clean_weather_rename_df.Date.dt.date.astype(str)
```

```
In [10]: # join two data structure
clean_weather_rename_df.index = clean_weather_rename_df['merge_date']

merged_df = clean_prices_rename_df.join(clean_weather_rename_df, on = ['merge_date'], rsuffix = '_w')
merged_df
```

Out[10]:

	Distance	Type of Cab	Date	destination	Source	Price	Surge Multiplier	User ID	Service Type	Vehicle Type	merge_date	Temperature in Fahrenheit	Location
0	0.44	Lyft	2018-12-16 09:30:07.890000105	North Station	Haymarket Square	5.0	1.0	424553bb-7174-41ea-ab4-fe06d4f4b9d7	lyft_line	Shared	Haymarket Square - 2018-12-16 - 9	NaN	Na
1	0.44	Lyft	2018-11-27 02:00:23.677000046	North Station	Haymarket Square	11.0	1.0	4bd23055-6827-41c6-b23b-3c491f24e74d	lyft_premier	Lux	Haymarket Square - 2018-11-27 - 2	44.31	Haymarket Square
1	0.44	Lyft	2018-11-27 02:00:23.677000046	North Station	Haymarket Square	11.0	1.0	4bd23055-6827-41c6-b23b-3c491f24e74d	lyft_premier	Lux	Haymarket Square - 2018-11-27 - 2	43.82	Haymarket Square
2	0.44	Lyft	2018-11-28 01:00:22.197999954	North Station	Haymarket Square	7.0	1.0	981a3613-77af-4620-a42a-0c086077d1e	lyft	Lyft	Haymarket Square - 2018-11-28 - 1	NaN	Na
3	0.44	Lyft	2018-11-30 04:53:02.749000072	North Station	Haymarket Square	26.0	1.0	c2d88af2-d278-4cfd-a8d0-29ca77cc5512	lyft_luxsuv	Lux Black XL	Haymarket Square - 2018-11-30 - 4	NaN	Na
...	...	...	...	...	...	...	...	...	...	...	...	...	...
693065	1.00	Uber	2018-12-01 23:53:05.533999919	North End	West End	9.5	1.0	353e6566-b272-479e-a9c6-98bd6cb23f25	9a0e7b09-b92b-4c41-9779-ccc7a75f6969	WAV	West End - 2018-12-01 - 23	NaN	Na
693066	1.00	Uber	2018-12-01 23:53:05.533999919	North End	West End	13.0	1.0	618d3611-1820-450a-9645-a9ff304a4842	6f72dfc5-27f1-42e8-84db-ccc7a75f6969	UberXL	West End - 2018-12-01 - 23	NaN	Na
693067	1.00	Uber	2018-12-01 23:53:05.533999919	North End	West End	9.5	1.0	633a3fc3-1f86-4b9e-9d48-2b7132112341	55c66225-fbe7-4fd5-9072-eab1ece5e23e	UberX	West End - 2018-12-01 - 23	NaN	Na
693069	1.00	Uber	2018-12-01 23:53:05.533999919	North End	West End	27.0	1.0	727e5f07-a96b-4ad1-a2c7-9abc3ad55b4e	6d318bcc-22a3-4af6-bddd-b409f0ce1546	Black SUV	West End - 2018-12-01 - 23	NaN	Na

```
In [13]: merged_df[merged_df['User ID'] == '81ffc3d0-b1e4-4f08-bfdd-b4b79035f9c9'].iloc[:, 10:20]
```

Out[13]:

	merge_date	Temperature in Fahrenheit	Location	Cloudiness	Pressure	Rainfall in inches	Date_w	Humidity	Wind Speed	merge_date_w
686386	Boston University - 2018-11-27 - 0	43.98	Boston University	1.00	1006.25	0.0272	2018-11-27 00:17:14	0.90	9.63	Boston University - 2018-11-27 - 0
686386	Boston University - 2018-11-27 - 0	43.89	Boston University	0.98	1005.89	0.2112	2018-11-27 00:42:13	0.91	10.64	Boston University - 2018-11-27 - 0
686386	Boston University - 2018-11-27 - 0	43.79	Boston University	0.99	1006.08	0.1622	2018-11-27 00:45:21	0.91	9.59	Boston University - 2018-11-27 - 0
686386	Boston University - 2018-11-27 - 0	43.88	Boston University	0.98	1005.88	0.2072	2018-11-27 00:42:27	0.91	10.66	Boston University - 2018-11-27 - 0
686386	Boston University - 2018-11-27 - 0	43.98	Boston University	1.00	1006.35	0.0318	2018-11-27 00:12:13	0.90	9.31	Boston University - 2018-11-27 - 0
686386	Boston University - 2018-11-27 - 0	43.88	Boston University	0.98	1005.85	0.1887	2018-11-27 00:43:37	0.91	10.73	Boston University - 2018-11-27 - 0
686386	Boston University - 2018-11-27 - 0	43.99	Boston University	0.98	1005.93	0.1662	2018-11-27 00:58:48	0.90	9.82	Boston University - 2018-11-27 - 0
686386	Boston University - 2018-11-27 - 0	43.99	Boston University	0.98	1005.92	0.1675	2018-11-27 00:59:13	0.90	9.84	Boston University - 2018-11-27 - 0

# MERGING DATA TOGETHER

```
In [1]: # Import dependencies
import csv
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import requests
import scipy.stats as st
from datetime import datetime
```

```
In [2]: # Import csv files and dependencies
cab_prices_file = "uber_lyft_data/cab_rides.csv"
weather_file = "uber_lyft_data/weather.csv"

cab_prices = pd.read_csv(cab_prices_file)
weather_file = pd.read_csv(weather_file)

cab_prices_df = pd.DataFrame(cab_prices)
weather_df = pd.DataFrame(weather_file)
```

```
In [7]: # Getting total number of services by uber and lyft
service_count = clean_prices_df["name"].value_counts()
service_count
```

```
Out[7]: UberXL      55096
Black SUV      55096
WAV            55096
Black          55095
UberX          55094
UberPool       55091
Lux            51235
Lux Black XL   51235
Lyft           51235
Lux Black      51235
Lyft XL        51235
Shared         51233
Name: name, dtype: int64
```

```
In [11]: # show an info for total number of indexes.
# It added extra data in the weather data, so index number increased to 719,135.
merged_df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 719135 entries, 0 to 693070
Data columns (total 20 columns):
#   Column              Non-Null Count  Dtype
---  --
0   Distance             719135 non-null  float64
1   Type of Cab          719135 non-null  object
2   Date                 719135 non-null  datetime64[ns]
3   destination          719135 non-null  object
4   Source               719135 non-null  object
5   Price                719135 non-null  float64
6   Surge Multiplier     719135 non-null  float64
7   User ID              719135 non-null  object
8   Service Type         719135 non-null  object
9   Vehicle Type         719135 non-null  object
10  merge_date           719135 non-null  object
11  Temperature in Fahrenheit  190311 non-null  float64
12  Location             190311 non-null  object
13  Cloudiness           190311 non-null  float64
14  Pressure             190311 non-null  float64
15  Rainfall in inches   190311 non-null  float64
16  Date_u               190311 non-null  datetime64[ns]
17  Humidity             190311 non-null  float64
18  Wind Speed           190311 non-null  float64
19  merge_date_w         190311 non-null  object
dtypes: datetime64[ns](2), float64(9), object(9)
memory usage: 115.2+ MB
```

```
In [12]: # show the value counts of the id, in order to see total number of the id
merged_df["user ID"].value_counts()

Out[12]: 5b85b6df-567b-4a48-b8c0-02ec07c1e4b0      8
3c78cd0f-f2d7-4a4e-8f4d-ea55f09565f5      8
4d45a67b-ba7e-41e4-8b59-b84fc0f99fa      8
a0698e3b-3a7d-42a6-be55-6365c79977df      8
cdc1ba5b-4a56-433c-bbc3-5ec7a7b3c83a      8
...
bed8bfcf-7524-4021-b6e2-1b5ba725b219      1
55fa4f4d-796d-4f3d-8b8b-ad22ced57652      1
1470959d-88f2-4b0c-998a-f1bb487c7c6      1
8512566e-a302-45f1-8005-48ec48f22ed      1
079f486f-e748-4fbd-ea53-fa8e5edf7ade      1
Name: user ID, Length: 637976, dtype: int64
```

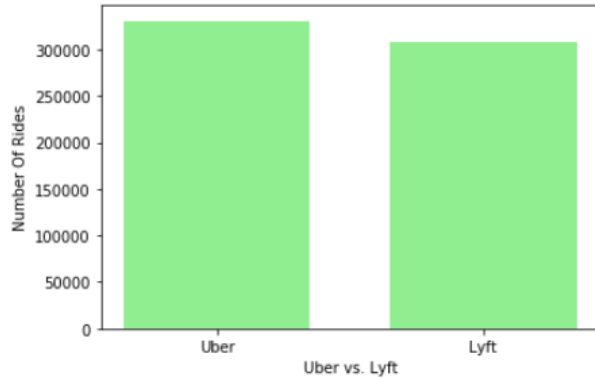
# Importing and Finding Objects...

- Importing the dependencies and files
- Finding necessary services for Uber and Lyft
- Finding the total values, after merging the data together

```
In [6]: # bar chart for comparing uber and lyft rides
labels_x = ['Uber', 'Lyft']
counts_y = [clean_prices_df.cab_type[(clean_prices_df.cab_type) == 'Uber'].count(),
            clean_prices_df.cab_type[(clean_prices_df.cab_type) == 'Lyft'].count()]

plt.bar(labels_x, counts_y, color="lightgreen", align="center", width=0.5)

plt.xlabel("Uber vs. Lyft")
plt.ylabel("Number Of Rides")
plt.savefig("uber_vs_lyft_number_of_rides.png")
plt.show()
```



```
In [5]: # Uber vs Lyft Usage Comparison in pie chart
cab_counts = [clean_prices_df.cab_type[(clean_prices_df.cab_type) == 'Lyft'].count(),
              clean_prices_df.cab_type[(clean_prices_df.cab_type) == 'Uber'].count()]

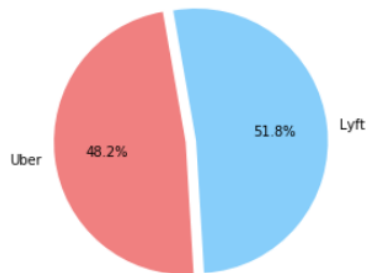
explode = (0.08, 0)

cab_types = ['Uber', 'Lyft']

colors = ["lightcoral", "lightskyblue"]

plt.pie(cab_counts, explode=explode, labels=cab_types, colors=colors,
        autopct="%1.1f%%", shadow=False, startangle=100)

plt.axis("equal")
plt.savefig("uber_vs_lyft_usage_comparison.png")
plt.show()
```



## Questions: Is there more Uber users than Lyft users?

- Yes, there are more rides with Uber than Lyft rides.
- However, I find out that Lyft has silently more users than Uber users in percentages in the pie chart.

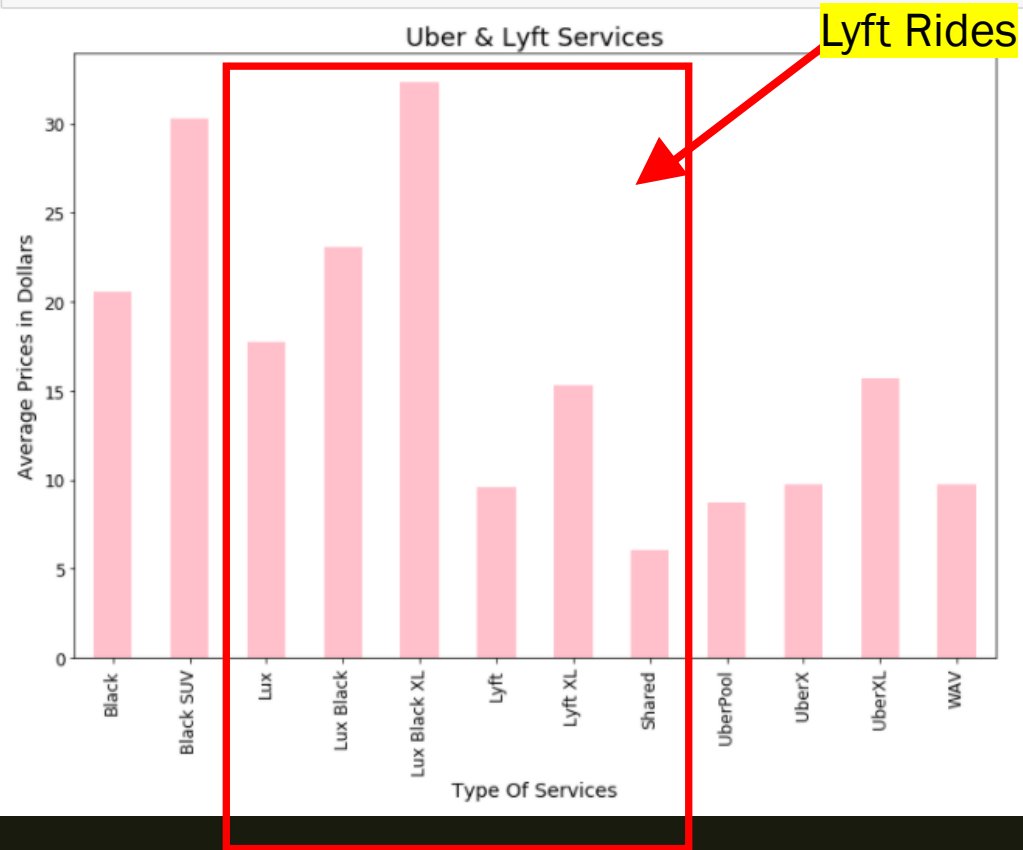
```
In [8]: # Uber vs Lyft Price Comparison bar chart
```

```
average_price = clean_prices_df.groupby('name')['price'].mean()

bar_chart_price = average_price.plot.bar(x="name", y="price",
                                         color="pink", figsize=(10,8), fontsize= 12)

bar_chart_price

plt.title("Uber & Lyft Services", fontsize = 18)
plt.xlabel("Type Of Services", fontsize = 14)
plt.ylabel("Average Prices in Dollars", fontsize = 14)
plt.tight_layout()
plt.savefig("uber_vs_lyft_avg_price_comparison_by_service_type.png")
plt.show()
```



Questions: How many average prices are there in each Uber and Lyft services?

- Lyft Shared has approx. 5-6 dollars in average prices.
- Lyft Lux Black XL has slightly more than 30 dollars in average prices.
- Lyft Shared has a lowest average prices in dollars, while Lyft Lux Black XL has the highest average prices in dollars.
- Lyft Rides are in the red box.
- Uber Rides are outside of the box.

```
In [71]: # Average Price Vs Distance traveled

# Collect Uber and Lyft in the data
uber_df = merged_df[merged_df['Type of Cab'] == 'Uber']
lyft_df = merged_df[merged_df['Type of Cab'] == 'Lyft']

# Find the averages in Uber and Lyft prices
uber_avgprice = uber_df.groupby('Distance')['Price'].mean()
lyft_avgprice = lyft_df.groupby('Distance')['Price'].mean()

# Plot the charts and apply some styling
fig1, ax1 = plt.subplots(figsize=(10,8))

plt.plot(uber_avgprice, label='Uber')
plt.plot(lyft_avgprice, label='Lyft')

plt.title('Average Price in Dollars VS Distance Traveled', fontsize=16)
plt.xlabel('Distance Traveled in Miles', fontsize=16)
plt.ylabel('Average Price in Dollars', fontsize=16)
plt.legend()
plt.savefig('Average_Price_vs_Distance_Traveled')
plt.show()
```



Questions: Are there any differences between average price of the data and distance traveled?

- Lyft has the highest amount of average price than Uber as the distance increased.
- Both average prices are increasing as the distance traveled increases.



# Summary Analysis

- **Is there more Uber users than Lyft users?**
  - *Yes, there are more rides with Uber than Lyft rides.*
  - *However, I find out that Lyft has silently more users than Uber users in percentages in the pie chart.*
- **How many average prices are there in each Uber and Lyft services?**
  - *Lyft Shared has approx. 5-6 dollars in average prices.*
  - *Lyft Lux Black XL has slightly more than 30 dollars in average prices.*
  - *Lyft Shared has a lowest average prices in dollars, while Lyft Lux Black XL has the highest average prices in dollars.*
  - *Lyft Rides are in the red box.*
  - *Uber Rides are outside of the box.*
- **Questions: Are there any differences between average price of the data and distance traveled?**
  - *Lyft has the highest amount of average price than Uber as the distance increased.*
  - *Both average prices are increasing as the distance traveled increases.*

# Resources

- <https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city>
- <https://www.kaggle.com/ravi72munde/uber-lyft-cab-prices>