Impacts of COVID-19 and "shelter-in-place" order on the Lyft biking user behavior

by Duy Pham

Investigation Overview

The overall goal is to understand the possible changes in the user behavior with the Lyft biking business in COVID-19. This analysis does not aim to produce a causal inference but focus on the potential consequences that we can observe in March, or more specifically, the later half of March 2020.

Dataset Overview

The dataset includes the following **13 variables**, with a total of **1,480,500 observations**.

- Trip Duration (seconds)
- Start Time and Date
- End Time and Date
- Start Station ID
- Start Station Name
- Start Station Latitude
- Start Station Longitude
- End Station ID
- End Station Name
- End Station Latitude
- End Station Longitude
- Bike ID
- User Type

In [1]:
```
# import all packages and set plots to be embedded inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb

%matplotlib inline

# suppress warnings from final output
import warnings
warnings.simplefilter("ignore")
```

In [2]:
```
# load in the dataset into a pandas dataframe
df = pd.read_csv ('data.csv')
```

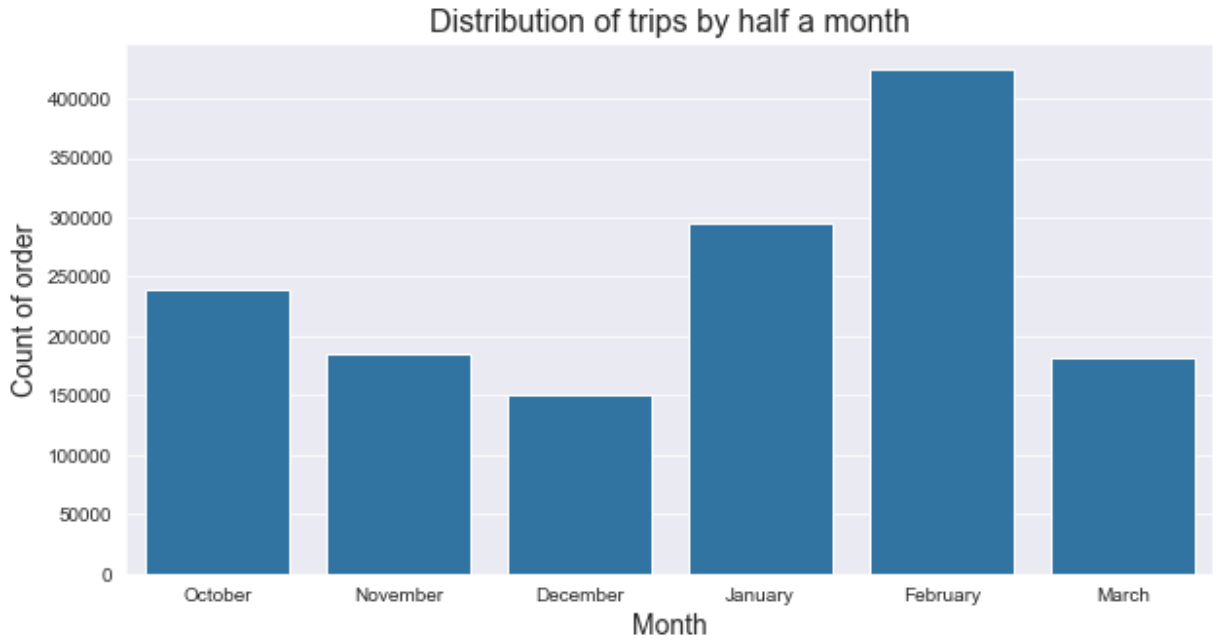
> Note that the above cells have been set as "Skip"-type slides. That means that when the notebook is rendered as http slides, those cells won't show up.

Surpising drop in the biking sessions in March

> In March, there has been a surprising drop in the number of biking sessions, which is opposite to the increasing trend since the beginning of the year. While the figure in March is not that different from October, November and December last year, it is a red flag about the potentially observable impact of COVID-19 and the "shelter-in-place" order.

In [3]:

```
plt.figure (figsize = (10,5))
base_color = sb.color_palette('Paired')[1]
sb.set_style('darkgrid')
sb.countplot(data=df, x='start_month', color=base_color)
plt.xlabel('Month', fontsize=14)
plt.ylabel('Count of order', fontsize=14)
plt.title("Distribution of trips by half a month", fontsize=16);
```

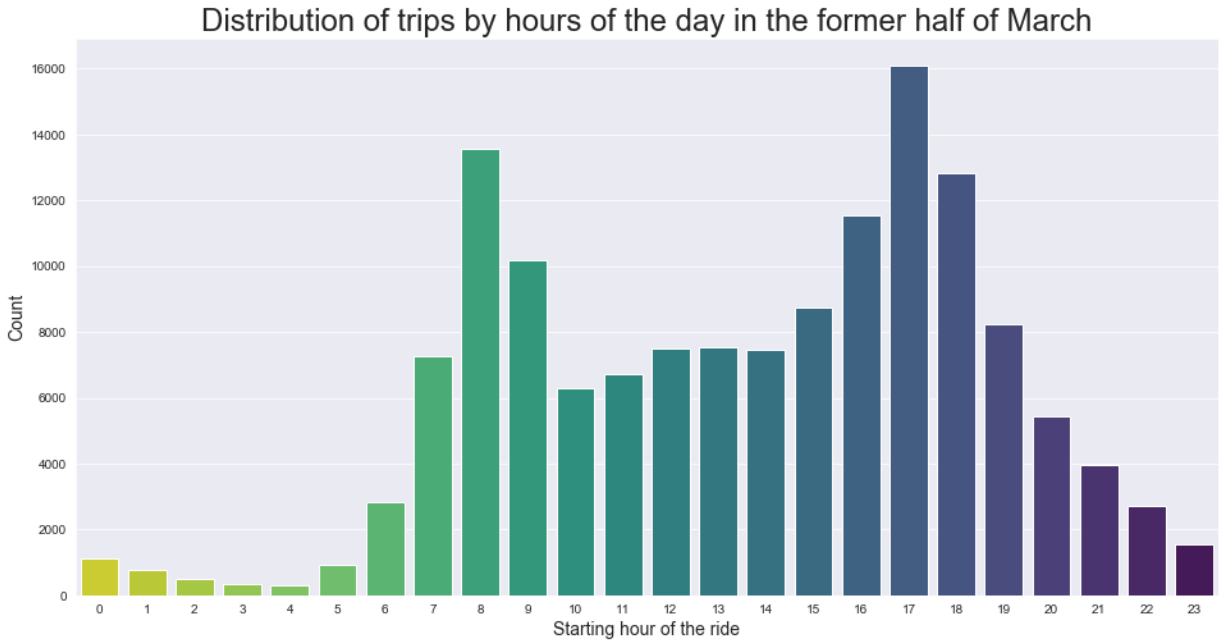


Changing behavior of starting time: The high number of morning session is no longer there

> Looking at the graph, morning (7am - 9am) and afternoon (4pm-6pm) werer the busiest time of the day in terms of bike riding pre-COVID-19. However, in March, there was a significantly visible drop in the distribution of people riding bike in the morning. It seems that fewer people going to work led to the drop in the number of biking sessions in the morning.

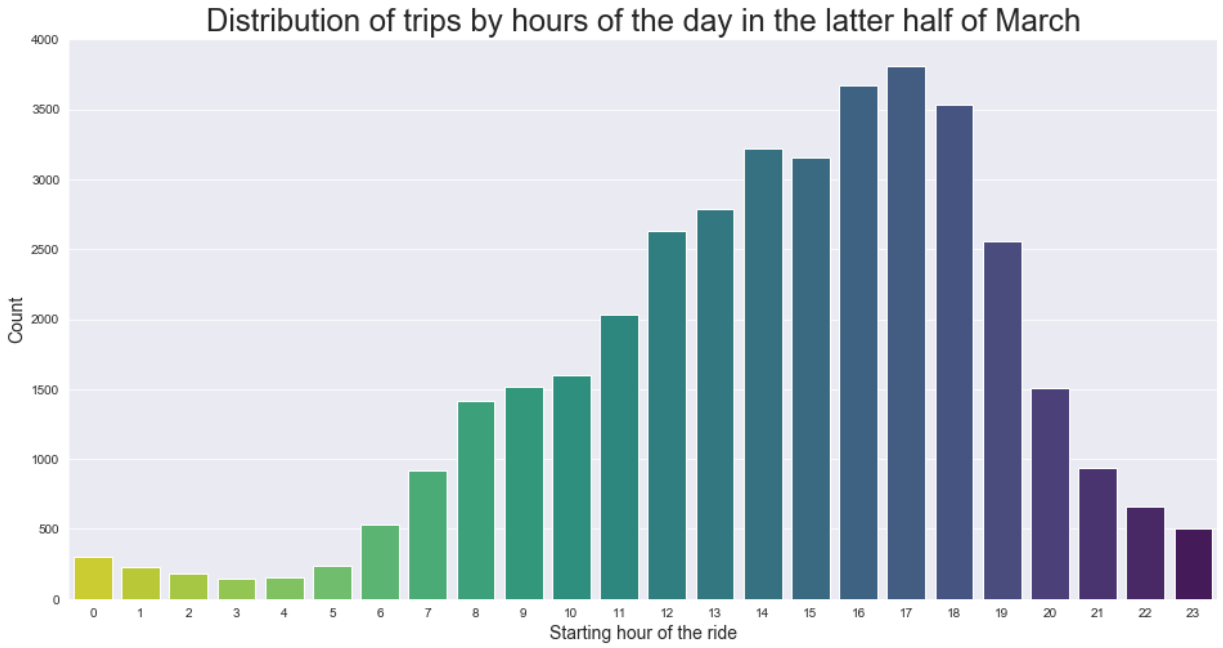
In [4]:

```
# Visualizing distribution of trips by hours
plt.figure (figsize = (16,8))
sb.countplot(data=df[df['half-month']=='March - first'], x='start_hour', pal
plt.xlabel('Starting hour of the ride', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.title("Distribution of trips by hours of the day in the former half of M
```



In [5]:

```
# Visualizing distribution of trips by hours
plt.figure (figsize = (16,8))
sb.countplot(data=df[df['half-month']=='March - second'], x='start_hour', pa
plt.xlabel('Starting hour of the ride', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.title("Distribution of trips by hours of the day in the latter half of M
```



As expected, at the latter half of March when the "shelter-in-place" took effect, there was a significant drop in the proportion of people riding a bike in the morning. The drop in the morning highly correlated with the working time. Below are two hypotheses for further user study:

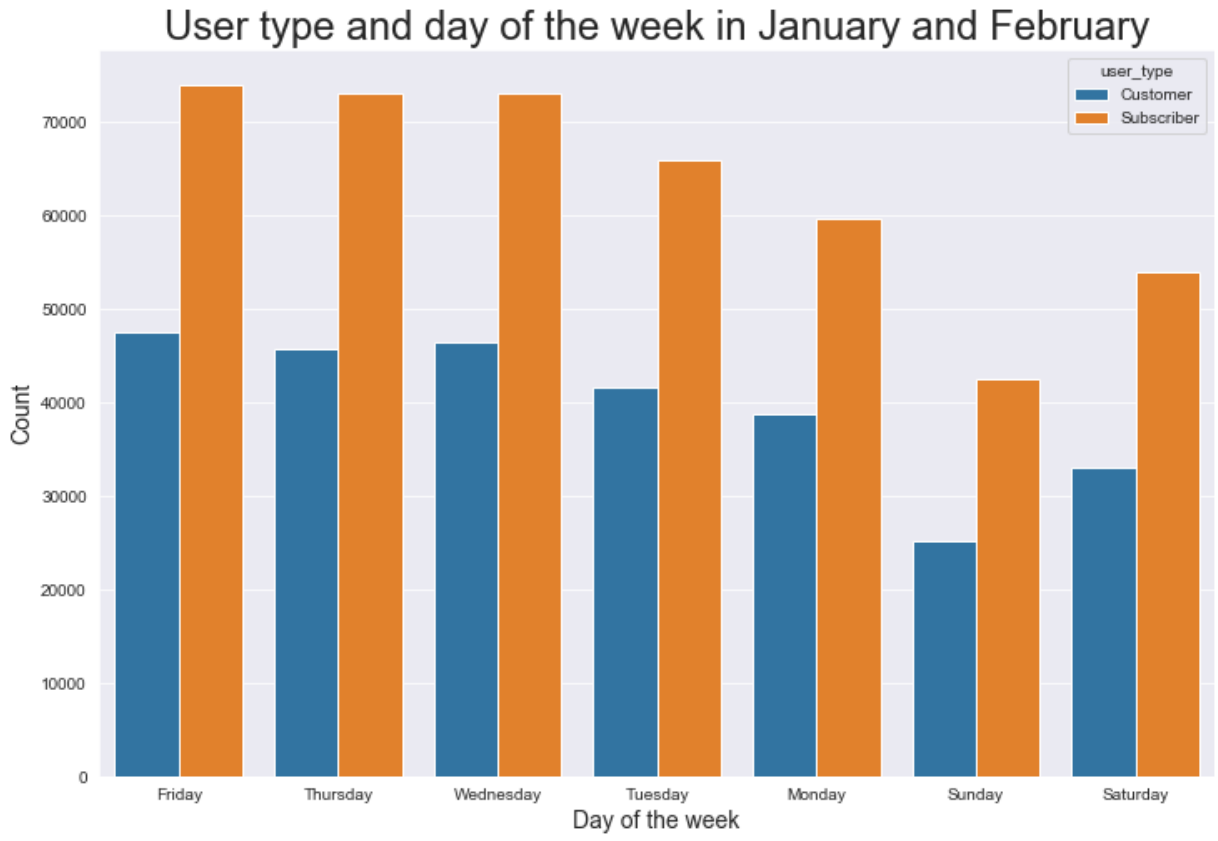
1. Once the "shelter-in-place" takes effect, people do not need to go to work. The assumption is that people often went to work by bike in the past. As long as they can work from home, they don't need to go to work with Lyft in the morning. Probably they walked back home in the past, too.
2. The biking is also for physical exercise purposes. When people can work from home, people basically have a more flexible schedule. For that reason,

instead of waking up super early and biking in the morning to go to work on time, people can basically exercise at any time of the day. Thus, people decide to exercise equally over the day. but the traditional routine of biking in the afternoon as an exercise still remains.

Changing behavior in days of the week: Sunday turns out to be the busiest day for Lyft biking session once "shelter-in-place" took effect.

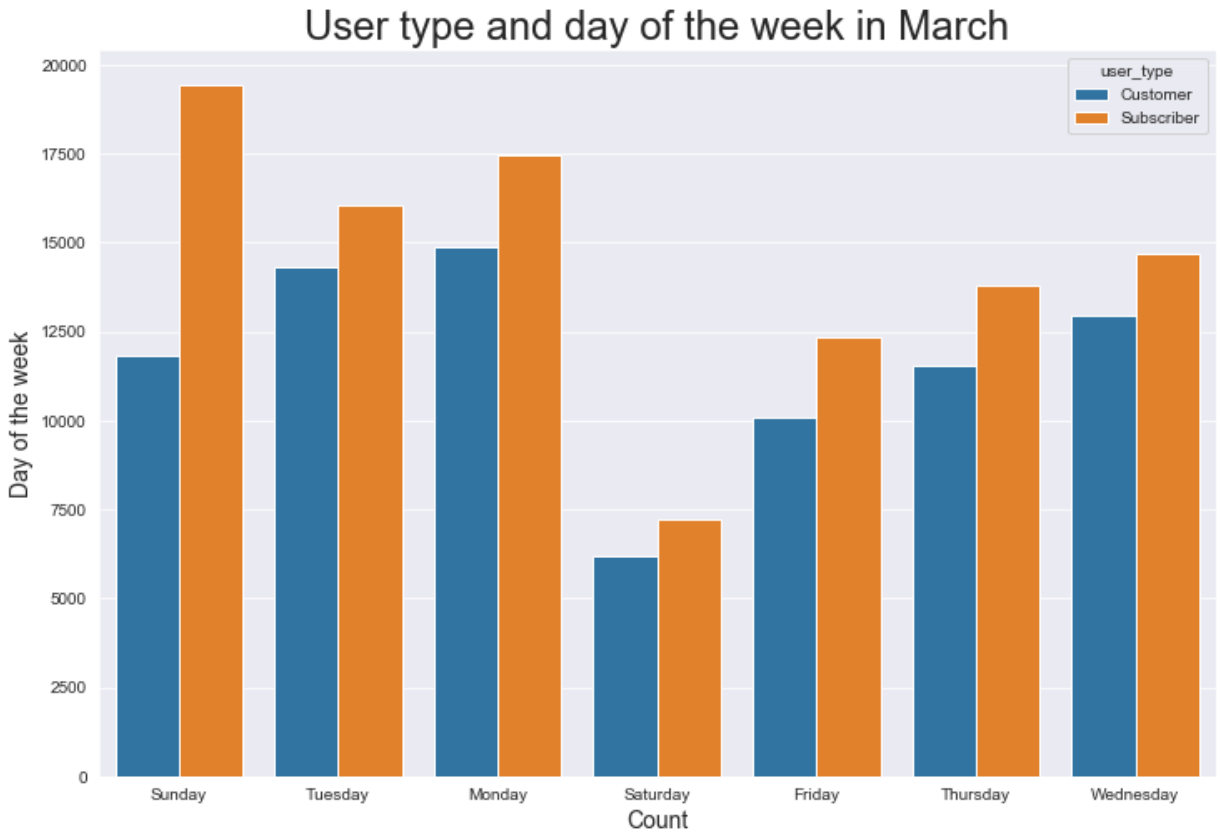
Before March, weekdays had the highest amount of ride requests. In March, Sunday, Tuesday and Monday had the highest amount of ride requests. For more specifically, there was a significant drop in the number of subscribers compared to customers. In other words, subscribers did not have to go to the office to work on weekdays, but a number of them still went for exercise on Sunday.

In [9]:
```
#In January and February
plt.figure (figsize = (12,8))
sb.countplot(data=df[(df['start_month']=='January')|(df['start_month']=='Feb
plt.xlabel('Day of the week', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.title("User type and day of the week in January and February", fontsize=
```



In [7]:

```
#In March
plt.figure (figsize = (12,8))
sb.countplot(data=df[(df['start_month']=='March')], x='start_day_of_week', h
plt.xlabel('Day of the week', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.title("User type and day of the week in March", fontsize=24);
```

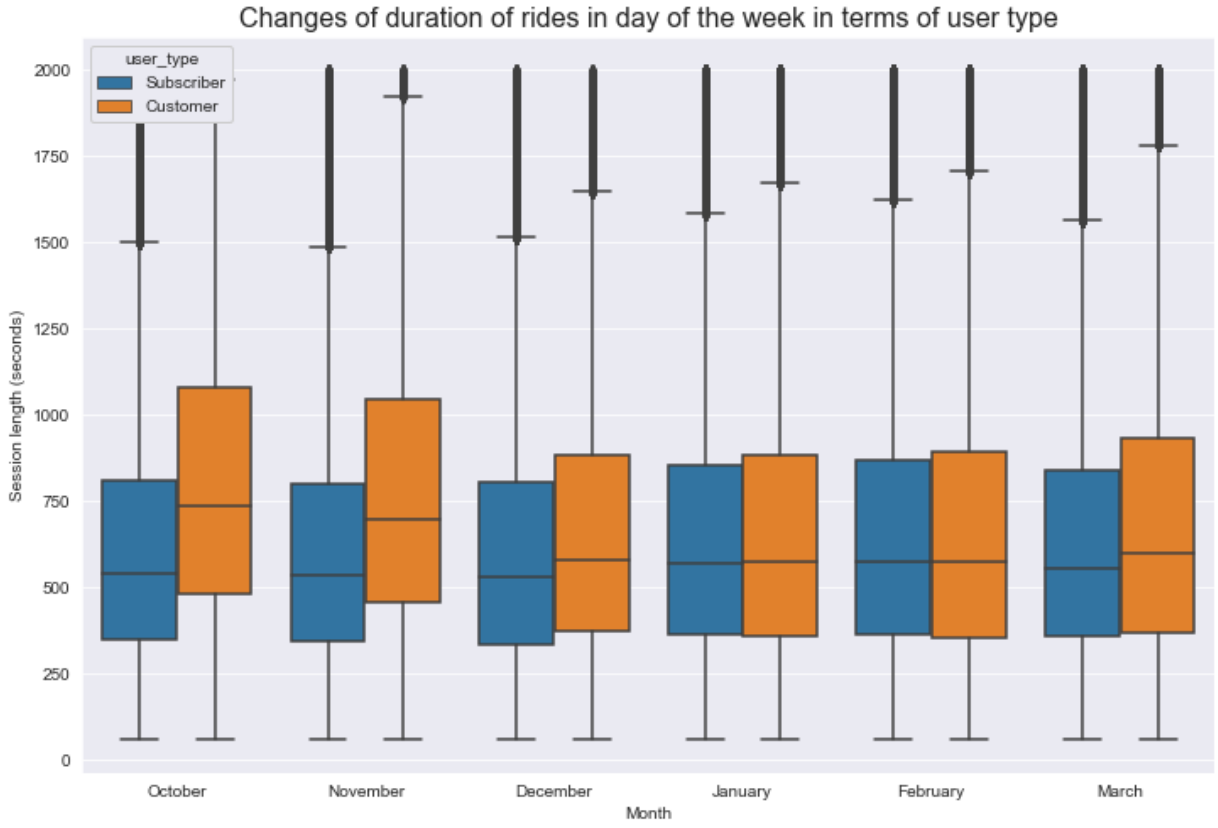


The average session length stays the same

> The session length does not seem to be affected by the COVID-19 and "shelter-in-place" order. This is an expected finding. I don't think people will cut short their exercise, or relocate their destination because of this pandemic. Once they are out from home, they will travel as usual, thus, session length still stays the same.

In [11]:

```
pplt.figure (figsize = (12,8))
sb.boxplot(data = df[df['duration_sec']<2000], x = 'start_month', y = 'durat
plt.legend(loc = 2, framealpha = 1, title = 'user_type')
plt.title('Changes of duration of rides in day of the week in terms of user
plt.ylabel('Session length (seconds)')
plt.xlabel('Month')
plt.show()
```



Conclusion

This finding is important for Lyft to identify user behavior in times of pandemic (and in normal time). By understanding that many subscribers use Lyft biking to go to work, Lyft can provide additional function in the app (of course, when the pandemic is over), such as, pick up breakfast on the way to work. Optimizing function according to user behavior will allow Lyft to come up with great ideas to support its customers.

> Once you're ready to finish your presentation, check your output by using nbconvert to export the notebook and set up a server for the slides. From the terminal or command line, use the following expression:
>
> ```
> jupyter nbconvert <file_name>.ipynb --to slides --post serve --template output_toggle
> ```

> This should open a tab in your web browser where you can scroll through your presentation. Sub-slides can be accessed by pressing 'down' when viewing its parent slide. Make sure you remove all of the quote-formatted guide notes like this one before you finish your presentation!

In [ ]:
```
1
```