Explanatory

August 15, 2022

1 Ford GoBike System Data

1.1 by Salomey Bezua

1.2 Investigation Overview

The aim of the presentation is to provide insights into the Ford GoBike System Data

1.3 Dataset Overview

The project uses Ford GoBike System data from the year 2018 (January to December). The dataset has 1736572 rows and 22 columns.

[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")
```

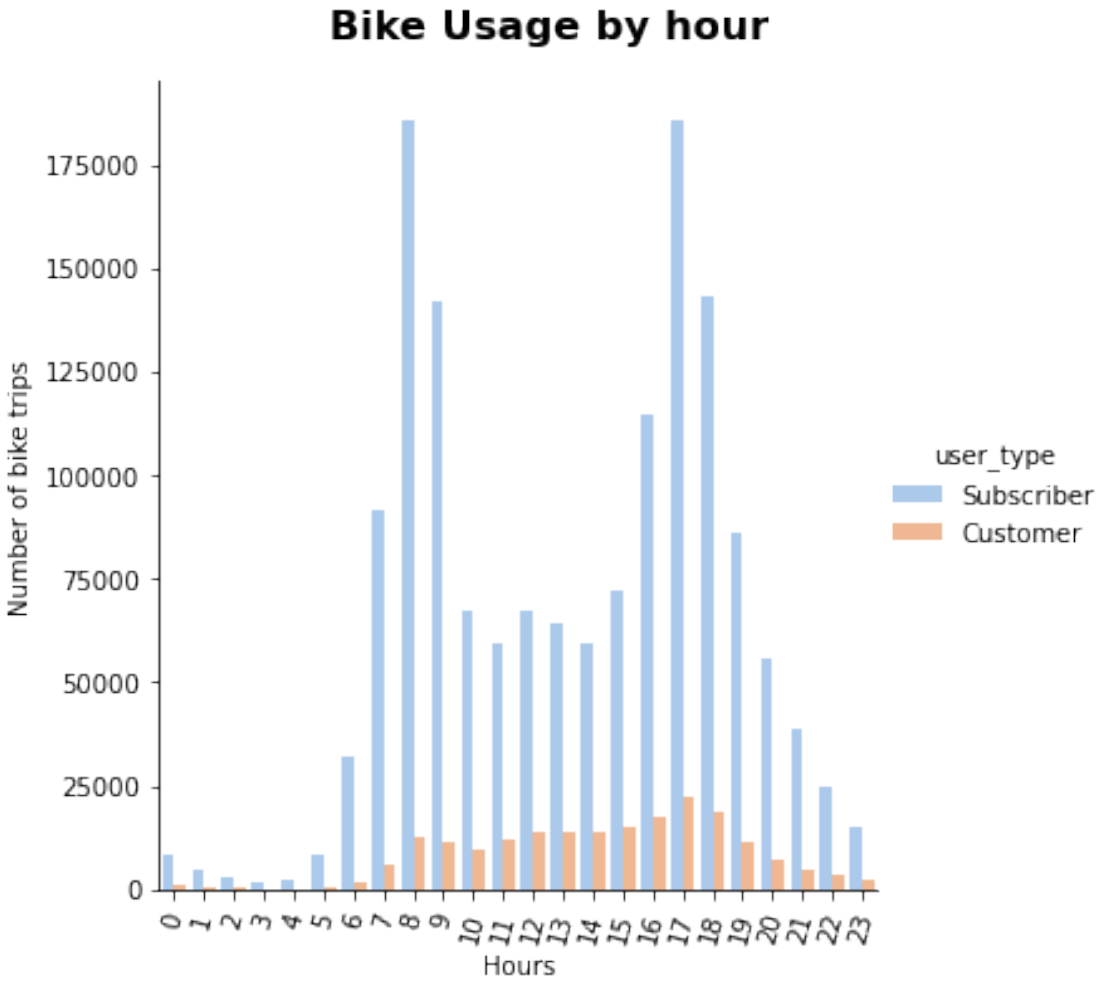
[2]:
```
# load in the dataset into a pandas dataframe
df = pd.read_csv('fordgobike_master_clean.csv')
```

1.4 (Visualization 1)

The figure below depicts the distribution of bike rides during the day. Subcribers use the system more than the customers. There are two peaks representing the two busiest time(Rush hour) of the day. Subscribers usage increases at 8 a.m and 5p.m .

[4]:
```
h = sb.
 ↪catplot(data=df,x='start_time_hour',kind='count',hue='user_type',palette='pastel')
h.set_axis_labels('Hours','Number of bike trips')
h.fig.suptitle('Bike Usage by hour',y=1.05,fontsize=16,fontweight='bold')
h.set_xticklabels(rotation=75)
```

[4]: <seaborn.axisgrid.FacetGrid at 0x7fd65fd195b0>

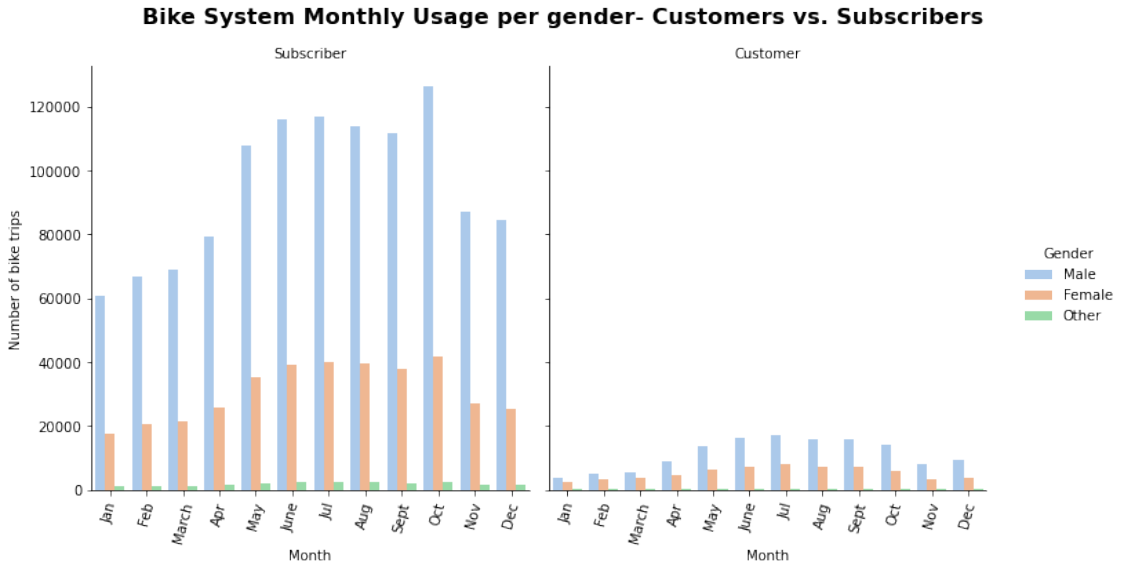


1.5 (Visualization 2)

It is observed that for subscribers, there is alot of demand from May to October and a decline in November through to March of the following year. This can be attributed to the Seasonal change. Similarly, this trend is observed for customers. The ratio of Males to Females is quite high for Subscibers as compared to the ratio for Customers

[5]:
```
g = sb.
 ↪catplot(data=df,x='start_time_month',col='user_type',hue='member_gender',kind='count',palet
ticks = [0,1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
months = ['Jan', 'Feb', 'March', 'Apr', 'May', 'June', 'Jul', 'Aug', 'Sept',␣
 ↪'Oct', 'Nov','Dec']
plt.xticks(ticks, months)
g.set_axis_labels('Month','Number of bike trips')
```

```
g.set_titles("{col_name}")
g._legend.set_title('Gender')
plt.suptitle('Bike System Monthly Usage per gender- Customers vs. Subscribers',␣
 ↪y=1.05, fontsize=16, fontweight='bold');
g.set_xticklabels(rotation=75);
```



1.6 (Visualization 3)

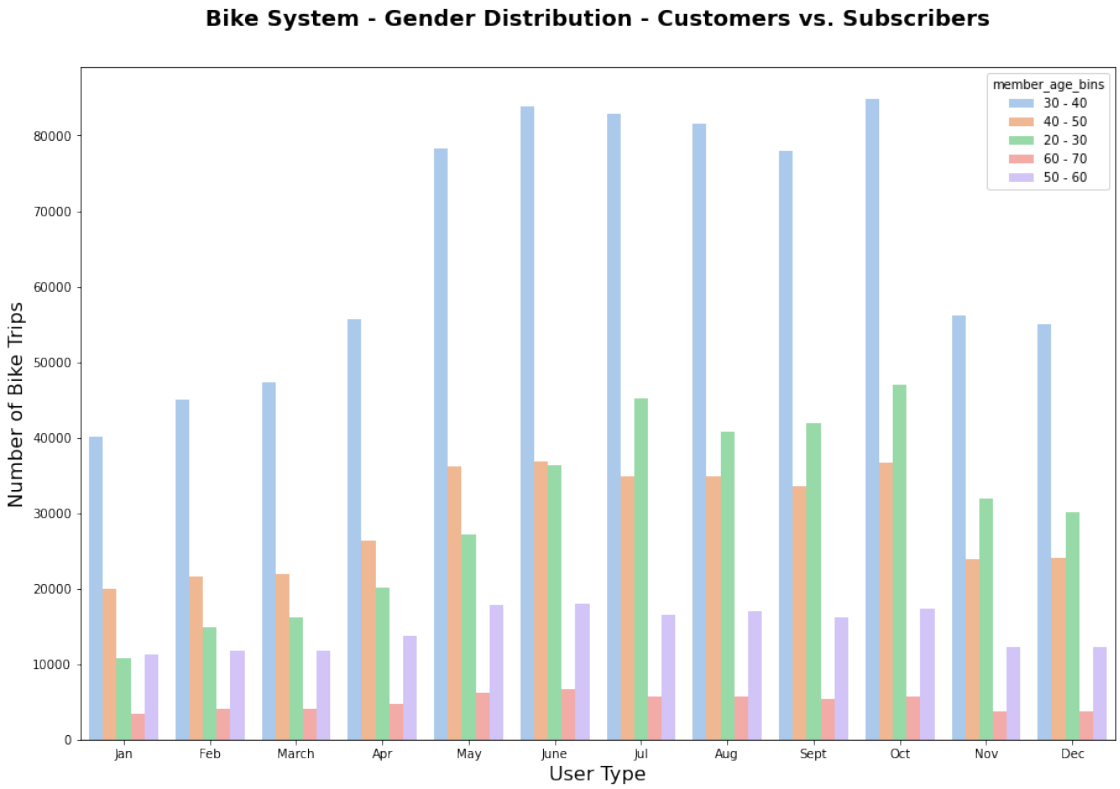
We can clearly see (30-40) age group uses the system often. It is observed that age group(40-50) usage increases from January to May and declines from July to November. Surprisingly, age group (20-30) Increases in the summertime with a peak in October.

[6]:
```
df['member_age_bins'] = df['member_age'].apply(lambda x: '10 - 20' if 10<x<=20
                                                else '20 - 30' if 20<x<=30
                                                else '30 - 40' if 30<x<=40
                                                else '40 - 50' if 40<x<=50
                                                else '50 - 60' if 50<x<=60
                                                else '60 - 70' if 60<x<=70
                                                else '70 - 80' if 70<x<=80
                                                else x)
```

[7]:
```
plt.figure(figsize = [15, 10])

g = sb.
 ↪countplot(data=df,x='start_time_month',hue='member_age_bins',palette='pastel')
ticks = [0,1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
months = ['Jan', 'Feb', 'March', 'Apr', 'May', 'June', 'Jul', 'Aug', 'Sept',␣
 ↪'Oct', 'Nov','Dec']
```

```
plt.xticks(ticks, months)
g.set_xlabel('User Type',fontsize=16)
g.set_ylabel('Number of Bike Trips',fontsize=16);
plt.title('Bike System - Gender Distribution - Customers vs. Subscribers', y=1.
 ↪05, fontsize=18, fontweight='bold');
```



Generate Slideshow: Once you're ready to generate your slideshow, use the `jupyter nbconvert` command to generate the HTML slide show. . From the terminal or command line, use the following expression.

[9]:
```
!jupyter nbconvert <Explanatory>.ipynb --to slides --post serve --no-input␣
 ↪--no-prompt
```

```
zsh:1: no such file or directory: Explanatory
```

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! At last, you can stop the Kernel.

[ ]:
```

```