Part_II_slide

November 17, 2022

1 Part II - Explanatory Data Visualization for Ford GoBike System Data

1.1 by Sarah Ali

1.2 Investigation Overview

This project has two parts that demonstrate the importance and value of data visualization techniques in the data analysis process which is as Capstone project part of the Udacity Data Analysis Nanodegree course.

1.3 Dataset Overview

This dataset consisted of 10 features of the bike trip in the first quater of 2019,This data set includes ride of 870,174 individuals. The features include the This includes the start and end time of each bike trip, the gender of riders, the age of riders, duration of each trip, the location each trip started and the location it ended. With Missing values in member_age and member-gender, this is a very high number and can not be dropped.

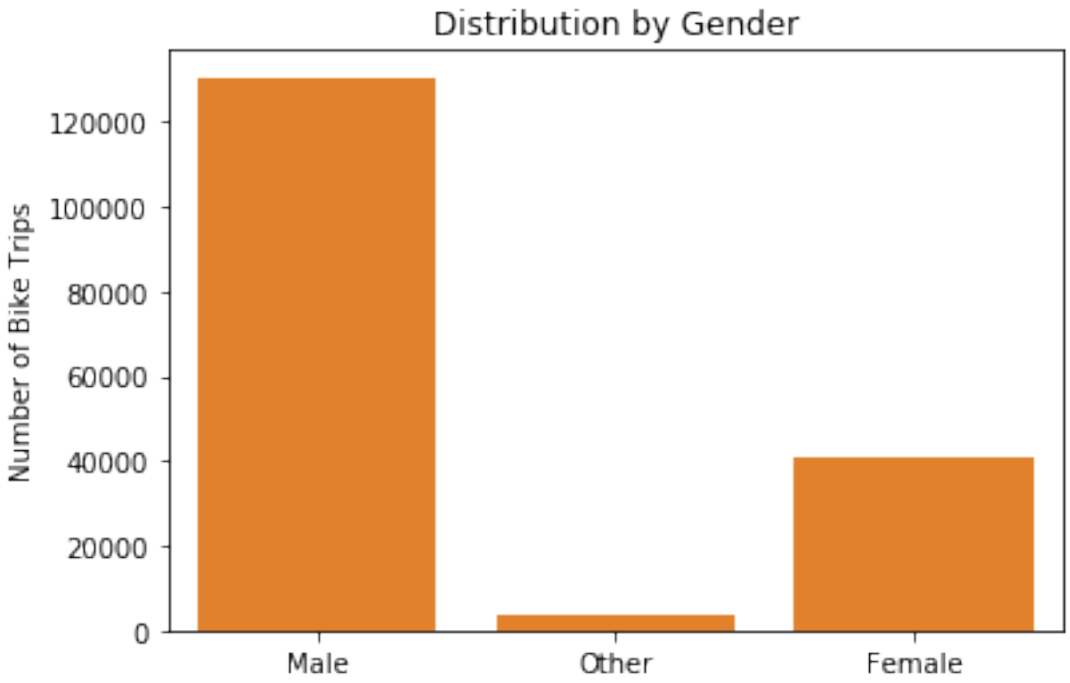
```
In [1]: # import all packages and set plots to be embedded inline
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sb

        %matplotlib inline

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

```
In [2]: # load in the dataset into a pandas dataframe
        slide = pd.read_csv('ford_cleandata1.csv')
```

1.4 Visualization 1: Distribution of riders gender

This countplot shows the distribtion of gender and how many bike trips each gender were on, it shows that Male take the highest number of trips, next to the female gender, a portion of people did not indicate what gender they may be associated with.

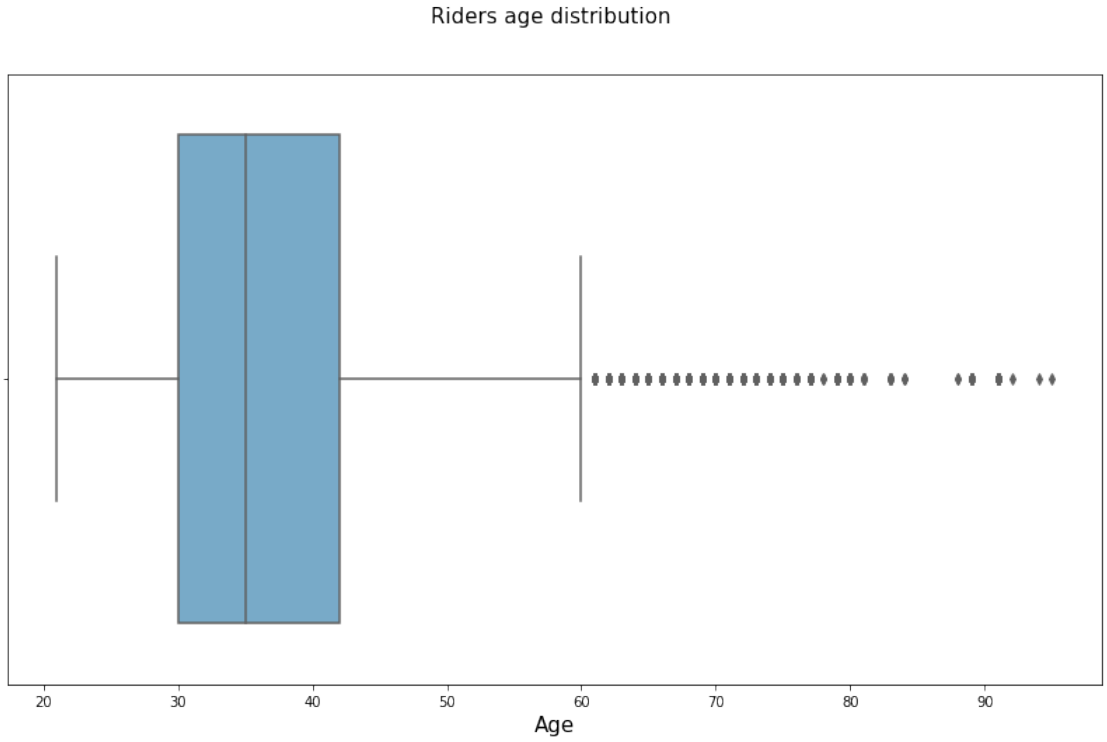
```
In [3]: base_color = sb.color_palette()[1]
        plot = sb.countplot(data=slide, x='member_gender',color=base_color)
        plt.xlabel('')
        plt.ylabel('Number of Bike Trips')
        plt.title('Distribution by Gender');
        #show the plot
        plt.show()
```



1.5 Visualization 2: Distribution of Riders Age

A boxplot from pythons seaborn library shows the way the age of riders are spread, outliers have been removed for abnormal ages

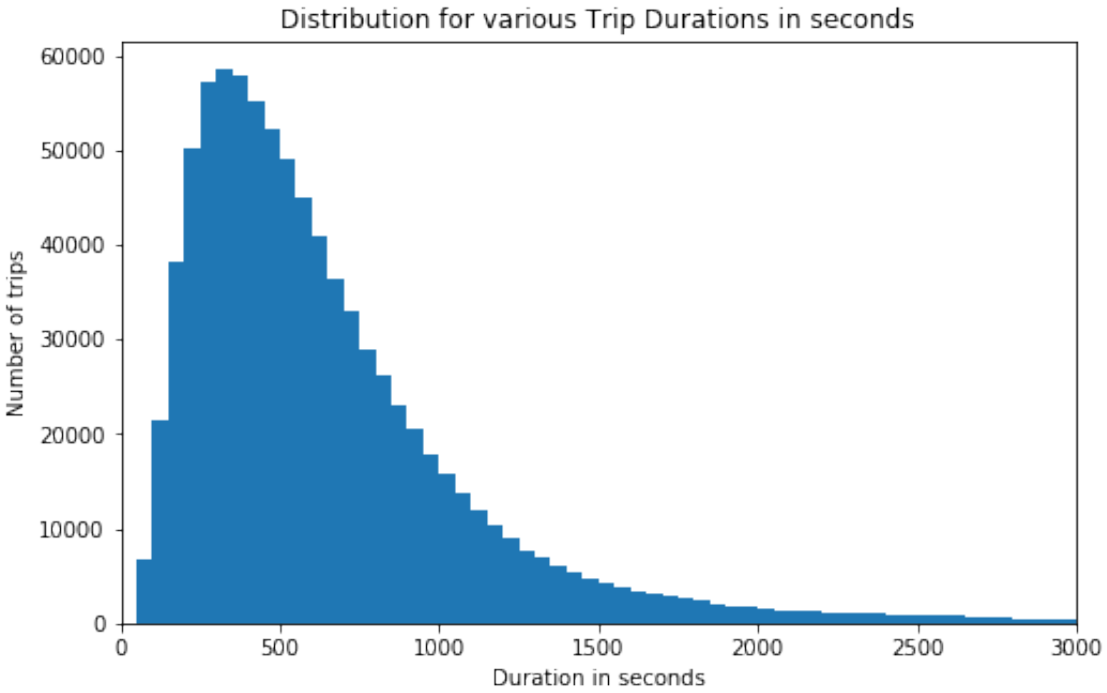
```
In [9]: #recheck with a graph to show age distribution
        plt.figure(figsize=(14,8))
        sb.boxplot(x='member_age', data=slide[slide['member_age'] <= 100], palette='Blues')
        plt.title("Riders age distribution ", fontsize=15, y=1.07)
        plt.xlabel("Age", fontsize=15);
```



1.6 Visualization 3: Trip duration (time)

This histogram shows the distribution of time suring each ride, it shows that most of the rides are peaking at the 500 seconds mark.

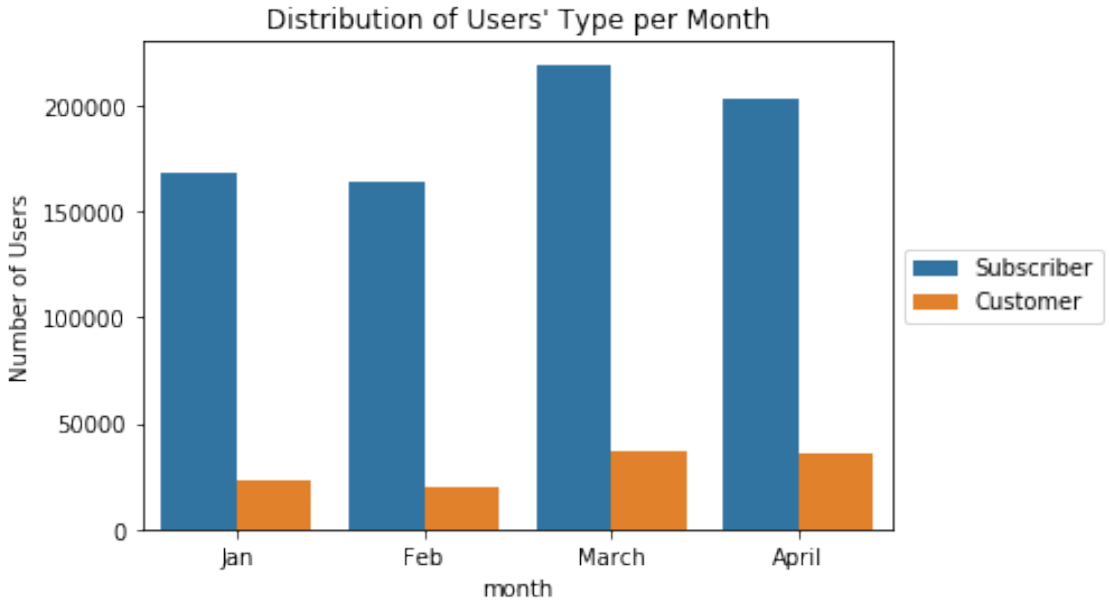
```
In [4]: # Plot a histogram representing time spent on each race in seconds
        binsize = 50
        bins = np.arange(50, slide['duration_sec'].max()+binsize, binsize)

        plt.figure(figsize=[8, 5])
        plt.hist(data = slide, x = 'duration_sec', bins=bins)
        plt.title('Distribution for various Trip Durations in seconds')
        plt.xlabel('Duration in seconds')
        plt.ylabel('Number of trips')
        plt.xlim([0, 3000]);
```



1.7 Visualization 4: Relationship between user type and month distribution

this shows that Subscribers are generally higher than customers all through the months, customer count remains relatively low

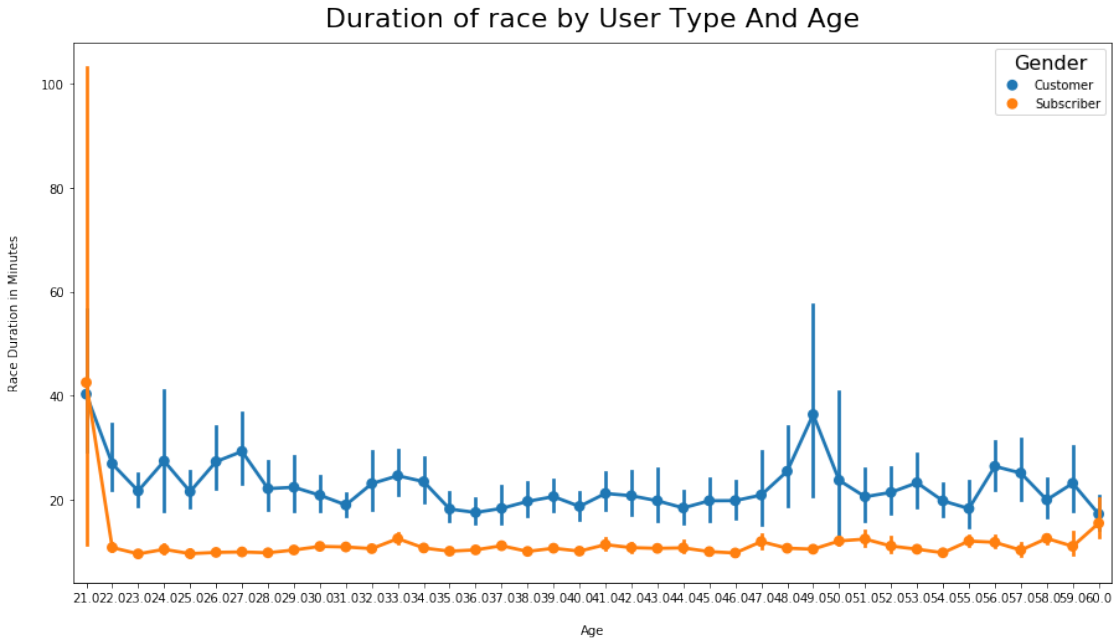
In [11]:
```
#a countplot to show the relationship between user types and the month of each trip

g = sb.countplot(data=slide, x='month', hue='user_type')
g.legend(loc='center left', bbox_to_anchor=(1, 0.5))
plt.xticks([0, 1, 2, 3], ['Jan', 'Feb', 'March', 'April']);
#plt.xlabel('Months')
plt.ylabel('Number of Users')
plt.title("Distribution of Users' Type per Month");
```



1.8 Visualization 5: Relationship between user type and duration of trips

The pointplot shows that customers had longer trips than subscribers., the age of younger subsciber had a high peak compared to customers, but in general regardless of age, the customers went on much more longer rides than subscribers.

```
In [5]: plt.figure(figsize=(15,8))
        ax = sb.pointplot(x='member_age', y='race_mins', hue='user_type',data=slide.query('membe
        plt.title('Duration of race by User Type And Age', fontsize=22, y=1.015)
        plt.xlabel('Age', labelpad=16)
        plt.ylabel('Race Duration in Minutes', labelpad=16)
        leg = ax.legend()
        leg.set_title('Gender',prop={'size':16})
        ax = plt.gca();
```



1.8.1 Generate Slideshow

Once you're ready to generate your slideshow, use the `jupyter nbconvert` command to generate the HTML slide show.

```
In [ ]: # Use this command if you are running this file in local
        !jupyter nbconvert "Part_II_slide.ipynb" --to slides --post serve --no-input --no-prompt
```

```
[NbConvertApp] Converting notebook Part_II_slide.ipynb to slides
[NbConvertApp] Writing 383873 bytes to Part_II_slide.slides.html
[NbConvertApp] Redirecting reveal.js requests to https://cdnjs.cloudflare.com/ajax/libs/reveal.j
Serving your slides at http://127.0.0.1:8000/Part_II_slide.slides.html
Use Control-C to stop this server
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: x-www-browser: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: firefox: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: iceweasel: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: seamonkey: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: mozilla: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: epiphany: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: konqueror: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: chromium-browser: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: google-chrome: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: www-browser: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: links2: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: elinks: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: links: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: lynx: not found
```

```
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: w3m: not found
xdg-open: no method available for opening 'http://127.0.0.1:8000/Part_II_slide.slides.html'
```

1.8.2 Recommendation:

Ford go bike should make conscious effort to make more users subscribers. Marketing strategy should be put in place to get new subscribers or convert customers to subscribers.

In [ ]:

In [ ]: