Exploring Bay Wheels' Bike Share trip data

by Mayukh Chakravartti

This dataset contains the bike trip details for Bay Wheel's bike sharing program

Import and Constants

```
# 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
import glob
import os

%matplotlib inline
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=

Enter your authorization code:
··········
Mounted at /content/drive
```

```
# Constants
data_folder = '/content/drive/My Drive/Colab Notebooks/Data Visualization/Project/Data
base_color = sb.color_palette()[0]
```

Load and Cleanup Dataset

```
# Load the data
all_files = glob.glob(os.path.join(data_folder, '*.csv'))
df_bikedata = pd.concat((pd.read_csv(f, low_memory=False) for f in all_files), sort=Fa
df_bikedata.reset_index(drop=True, inplace=True)
```

```
# Structure of the data
df_bikedata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4256681 entries, 0 to 4256680
Data columns (total 17 columns):
duration_sec               int64
start_time                 object
end_time                   object
start_station_id           float64
start_station_name         object
start_station_latitude     float64
start_station_longitude    float64
end_station_id             float64
end_station_name           object
end_station_latitude       float64
end_station_longitude      float64
bike_id                    int64
user_type                  object
member_birth_year          float64
member_gender              object
bike_share_for_all_trip    object
rental_access_method       object
dtypes: float64(7), int64(2), object(8)
memory usage: 552.1+ MB
```

```
# Some sample rows
df_bikedata.head()
```

tation_name	end_station_latitude	end_station_longitude	bike_id	user_type	membe
ɔde Island St at 17th St	37.764478	-122.402570	1035	Subscriber	
Union Square ll St at Post St)	37.788300	-122.408531	1673	Customer	
Francisco Ferry Building (Harry Bridges Pl...	37.795392	-122.394203	3498	Customer	
Francisco Ferry Building (Harry Bridges Pl...	37.795392	-122.394203	3129	Customer	
t St at Grant St	37.322980	-121.887931	1839	Subscriber	

```
# Data Cleanup 1 Start Time and End Time are objects, lets change that to Datetime typ
# Check for any nulls
df_bikedata.end_time.isnull().sum(), df_bikedata.start_time.isnull().sum()
```

```
(0, 0)
```

```
# Change column type to DatTime
df_bikedata.start_time = pd.to_datetime(df_bikedata.start_time)
df_bikedata.end_time = pd.to_datetime(df_bikedata.end_time)
# Data Cleanup 2 Change Birth Year to an integer instead of a int64
df_bikedata.member_birth_year = df_bikedata.member_birth_year.fillna(0).astype(int)
# Data Cleanup 3 Cleanup the column member_gender
df_bikedata.member_gender.unique()
```

```
array(['Male', 'Female', nan, 'Other', 'M', '?', 'F', 'O'], dtype=object)
```

```
di = {'M': 'Male',
      'F': 'Female',
      'O': 'Other',
      'Male': 'Male',
      'Female': 'Female',
      'Other': 'Other'}
df_bikedata.member_gender = df_bikedata.member_gender.map(di)


# Data Cleanup 4 Change columns user_type, member_gender, rental_access_method to cate
df_bikedata.user_type = df_bikedata.user_type.astype('category')
df_bikedata.member_gender = df_bikedata.member_gender.astype('category')
df_bikedata.rental_access_method = df_bikedata.rental_access_method.astype('category')
# Lets add a minute column for Duration
df_bikedata['duration_min'] = round(df_bikedata.duration_sec/60).astype(int)


# Some additional columns to be used for Data Visualization
df_bikedata['Start_Year_Month']=df_bikedata.start_time.dt.strftime('%Y-%m')
df_bikedata['member_age'] =  df_bikedata.start_time.dt.year - df_bikedata.member_birtl
df_bikedata['Start_Time_Hour'] = df_bikedata.start_time.dt.hour
df_bikedata['Start_Day_Of_Week'] = df_bikedata.start_time.dt.weekday_name


day_type = {'Monday':'Weekday', 'Tuesday':'Weekday', 'Wednesday':'Weekday', 'Thursday'
df_bikedata['Start_Day_Type'] = df_bikedata.Start_Day_Of_Week.apply(lambda x: day_type


df_bikedata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4256681 entries, 0 to 4256680
Data columns (total 23 columns):
duration_sec               int64
start_time                 datetime64[ns]
end_time                   datetime64[ns]
start_station_id           float64
start_station_name         object
start_station_latitude     float64
start_station_longitude    float64
end_station_id             float64
end_station_name           object
end_station_latitude       float64
end_station_longitude      float64
bike_id                    int64
user_type                  category
member_birth_year          int64
member_gender              category
bike_share_for_all_trip    object
rental_access_method       category
duration_min               int64
Start_Year_Month           object
member_age                 int64
Start_Time_Hour            int64
Start_Day_Of_Week          object
Start_Day_Type             object
dtypes: category(3), datetime64[ns](2), float64(6), int64(6), object(6)
memory usage: 661.7+ MB
```

```
df_bikedata.head()
```

de	bike_id	user_type	member_birth_year	member_gender	bike_share_for_all_trip
570	1035	Subscriber	1988	Male	No
531	1673	Customer	1987	Male	No
203	3498	Customer	1986	Female	No
203	3129	Customer	1981	Male	No
931	1839	Subscriber	1976	Female	Yes

Areas of Interest

What is the structure of your dataset?

- The dataset contains the trip details of Bay Wheel/Ford Go's Bike Rental trip details. It contains s end location details, information regarding the rider's general information for sex and year of bir

What is/are the main feature(s) of interest in your dataset?

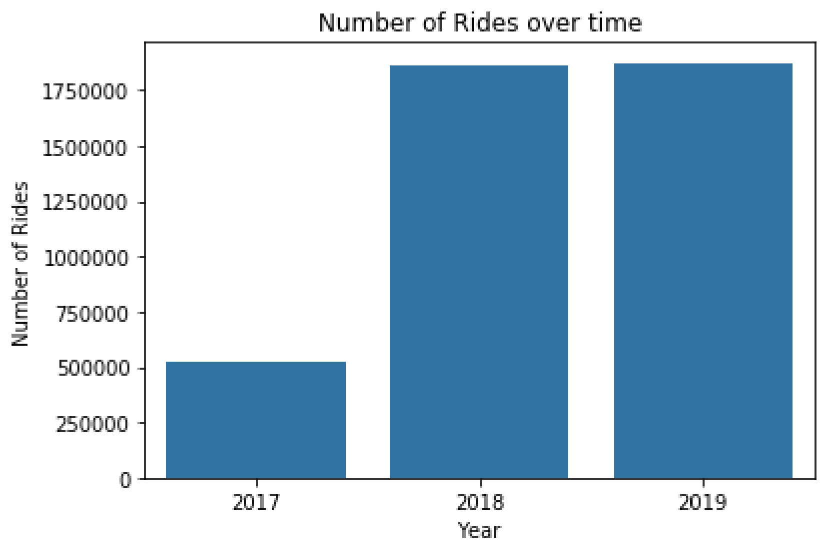
- Number of Rides
  - Trend in the number of rides over time - year/month
  - How does the number of rides vary over weekdays vs weekends
  - How has the number of rides on weekdays vs weekends varied over time
  - What are the number of rides by gender
  - What are the average number of rides by Rental Access Type
- Duration
  - What are the most common trip durations
  - How has the trip duration trended over time
  - How does the trip duration vary by most popular starting stations
  - How does the trip duration vary by gender
  - How has the trip duration varied by gender over time
  - How has the trip duration varied by age
  - How has the trip duration varied by the time of day
  - How does the trip duration vary by the gender and start time of day
  - How has the trip duration varied by the time of day over time
- Stations
  - What are the most popular starting and ending stations
  - What is the gender breakup for the rides from the most popular stations
  - How has the popularity of the top 5 most popular stations trended over time
  - How has the number of starting/ending stations affected the total number of bike rides
  - How has the average ride duration impacted by the number of stations

Univariate Section

Trend in the number of rides over time - year/month/weeks

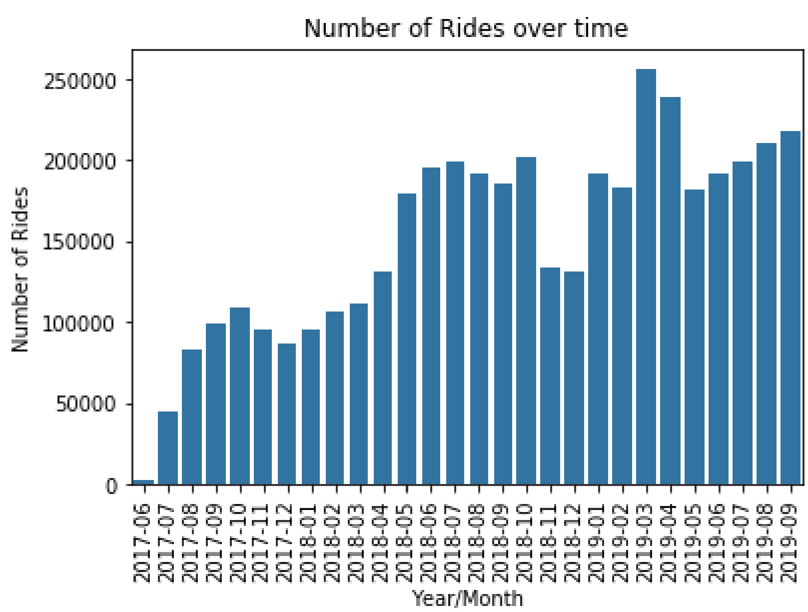
```
# Number of Bikerides over time - Year
sb.countplot(data=df_bikedata, x=df_bikedata.start_time.dt.year, color=base_color);
plt.xlabel('Year');
```

```
plt.ylabel('Number of Rides');
plt.title('Number of Rides over time');
```



- Number of bike rides have increased year on year and seems to be increasing even for the curre

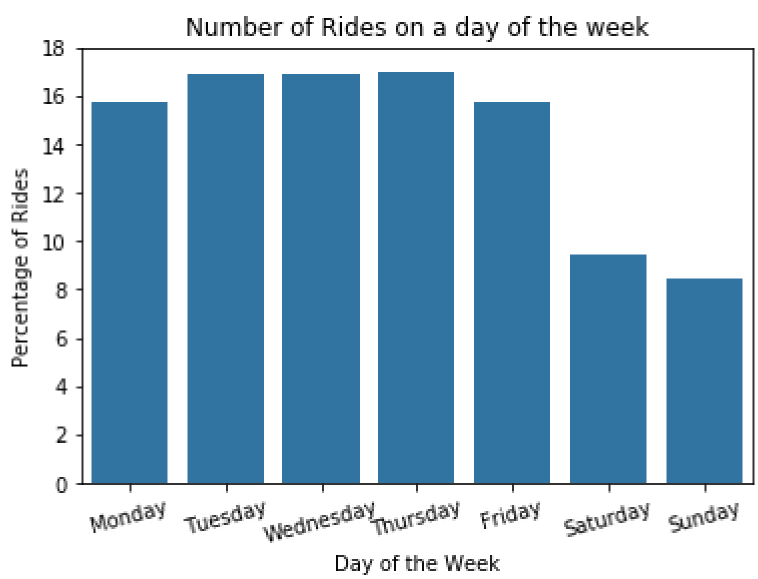
```
# Number of Bikerides over time - Month/Year
order = np.array(df_bikedata.Start_Year_Month.sort_values().unique())
sb.countplot(data=df_bikedata, x='Start_Year_Month', color=base_color, order=order);
plt.xlabel('Year/Month');
plt.ylabel('Number of Rides');
plt.xticks(rotation=90);
plt.title('Number of Rides over time');
```



- Monthly usage seems to have increased month over month with a dip in November and Decemb season when people are travelling or working from home and there are lot of holidays. Susbsequ rides jump back to normal from January and are trending upwards

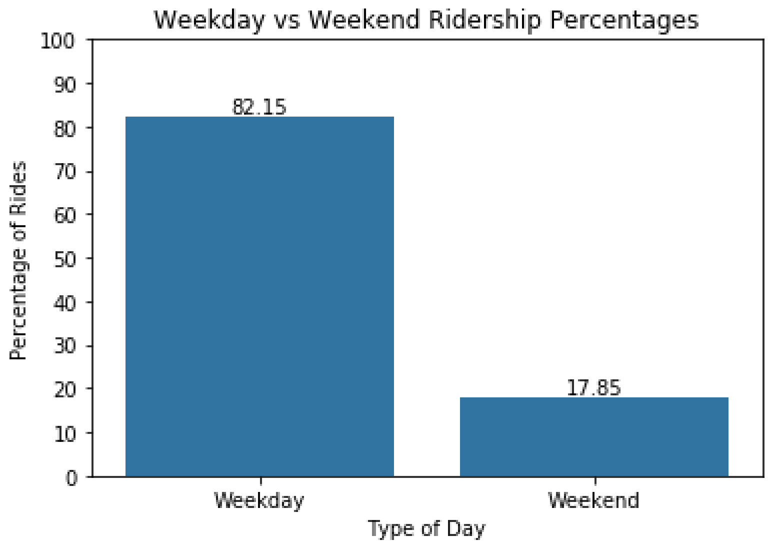
How does the number of rides vary over weekdays vs weekends

```
# Number of Bike Rides for the day of the week as a percentage
total_rides = df_bikedata.shape[0]
max_day_of_week_count = df_bikedata.Start_Day_Of_Week.value_counts().max()
max_prop = max_day_of_week_count / total_rides
ytick_values = np.arange(0, max_prop + 0.02, 0.02)
yticks_labels = ['{:0.0f}'.format(v*100) for v in ytick_values]


order = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'
sb.countplot(data=df_bikedata, x='Start_Day_Of_Week', color=base_color, order=order);
plt.yticks(ytick_values*total_rides, yticks_labels);
plt.xlabel('Day of the Week');
plt.ylabel('Percentage of Rides');
plt.xticks(rotation=15);
plt.title('Number of Rides on a day of the week');
```




```
# Breakup of rides between Weekdays vs Weekends
total_rides = df_bikedata.shape[0]
max_typeofday_count = df_bikedata.Start_Day_Type.value_counts().max()
max_prop = max_typeofday_count / total_rides
ytick_values = np.arange(0, max_prop + 0.2, 0.1)
yticks_labels = ['{:0.0f}'.format(v*100) for v in ytick_values]


ax = sb.countplot(data=df_bikedata, x='Start_Day_Type', color=base_color);
plt.xlabel('Type of Day');
plt.ylabel('Percentage of Rides');
plt.yticks(ytick_values*total_rides, yticks_labels);
total = df_bikedata.shape[0]
for p in ax.patches:
```

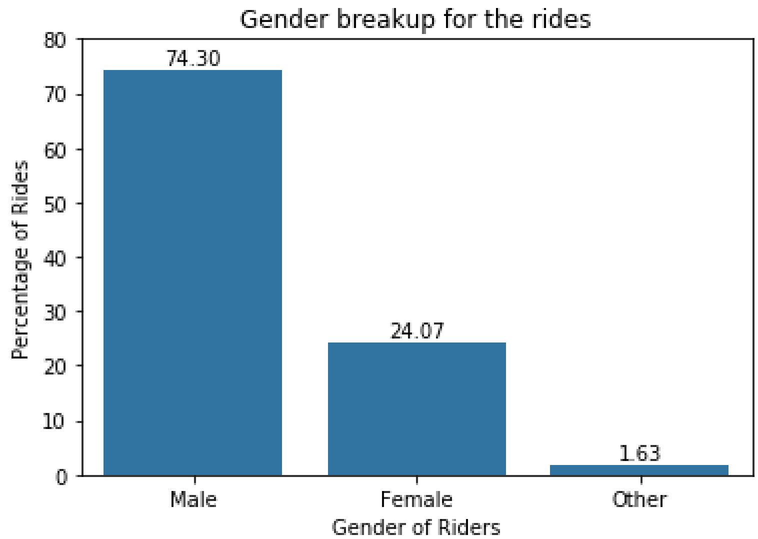
```
    height = p.get_height()
    ax.text(p.get_x()+p.get_width()/2,
            height + 40000,
            '{:1.2f}'.format(100*height/total),
            ha="center")
plt.title('Weekday vs Weekend Ridership Percentages');
```



- Most of the rides are during weekdays, with over 82% happening during weekdays as against arc

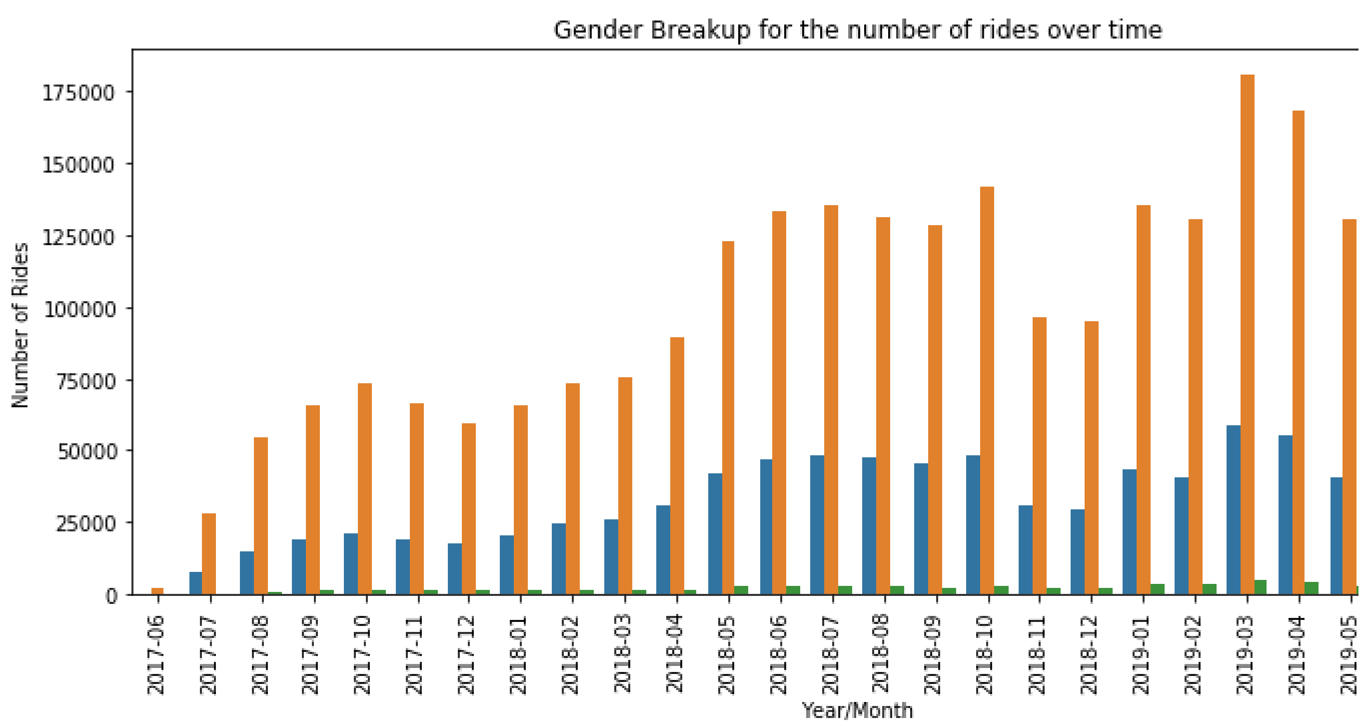
What are the number of rides by gender

```
df_bikedata_tmp = df_bikedata.dropna(subset=['member_gender'])
total_rides = df_bikedata_tmp.shape[0]
max_gender_count = df_bikedata_tmp.member_gender.value_counts().max()
max_prop = max_gender_count / total_rides
ytick_values = np.arange(0, max_prop + 0.1, 0.1)
yticks_labels = ['{:0.0f}'.format(100*v) for v in ytick_values]


order_gender_plot = df_bikedata_tmp.member_gender.value_counts().index
ax = sb.countplot(data=df_bikedata_tmp, x='member_gender', color=base_color, order=orc
plt.yticks(ytick_values*total_rides, yticks_labels);
plt.xlabel('Gender of Riders')
plt.ylabel('Percentage of Rides');
total = df_bikedata_tmp.shape[0]
for p in ax.patches:
    height = p.get_height()
    ax.text(p.get_x()+p.get_width()/2,
            height+40000,
            '{:1.2f}'.format(100*height/total),
            ha="center")
plt.title('Gender breakup for the rides');
```



- Almost 75% of riders are men while women make up 24.07% of the total rides.

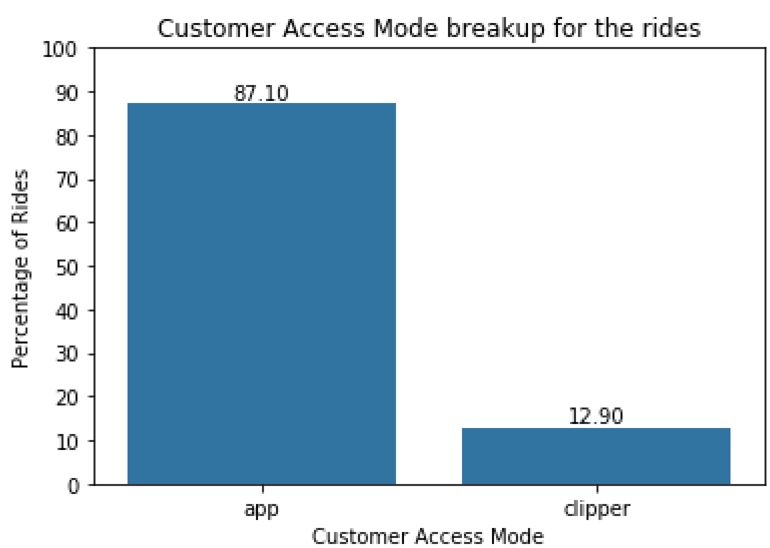
```
order = np.array(df_bikedata_tmp.Start_Year_Month.sort_values().unique())
plt.figure(figsize=(13,5));
sb.countplot(data=df_bikedata_tmp, x='Start_Year_Month', hue='member_gender', order=or
plt.xlabel('Year/Month');
plt.ylabel('Number of Rides');
plt.xticks(rotation=90);
plt.legend(title = 'Gender of Rider', bbox_to_anchor=(1,1));
plt.title('Gender Breakup for the number of rides over time');
```



- There has been major spike in the number of male riders through time, however, the number of r stabilised in the last 4 to 5 months with a similar distribution as seen as a whole

What are the average number of rides by Rental Access Type

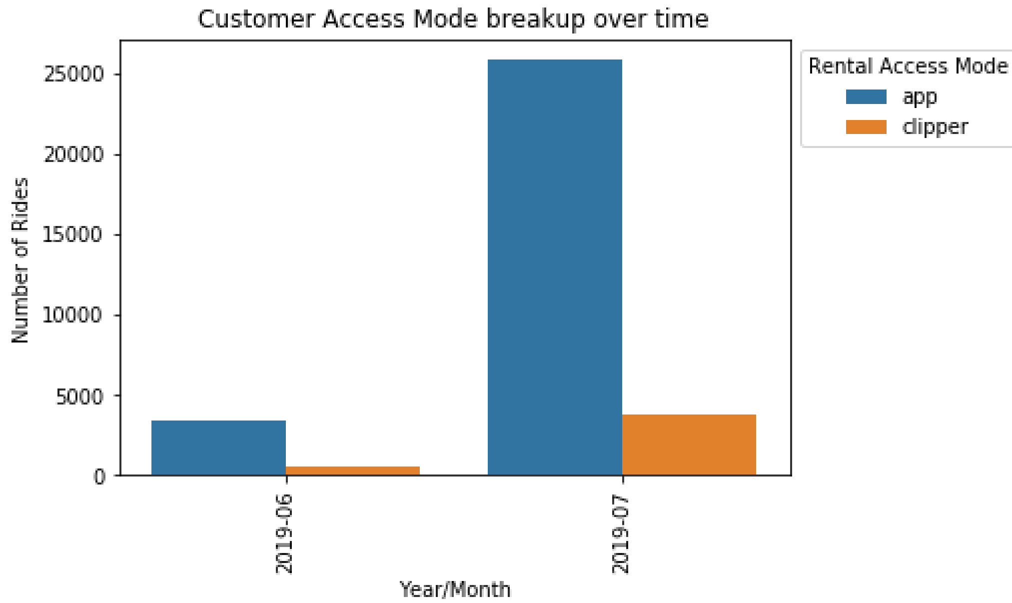
```
df_bikedata_tmp = df_bikedata.dropna(subset=['rental_access_method'])
total_rides = df_bikedata_tmp.shape[0]
max_accesstype_count = df_bikedata_tmp.rental_access_method.value_counts().max()
max_prop = max_accesstype_count / total_rides
ytick_values = np.arange(0, max_prop + 0.2, 0.1)
yticks_labels = ['{:0.0f}'.format(100*v) for v in ytick_values]


ax = sb.countplot(data=df_bikedata_tmp, x='rental_access_method', color=base_color)
plt.yticks(ytick_values*total_rides, yticks_labels);
plt.xlabel('Customer Access Mode')
plt.ylabel('Percentage of Rides');
total = df_bikedata_tmp.shape[0]
for p in ax.patches:
  height = p.get_height()
  ax.text(p.get_x()+p.get_width()/2,
          height+400,
          '{:1.2f}'.format(100*height/total),
          ha="center")
plt.title('Customer Access Mode breakup for the rides');
```



- The majority of rides 87% are coming via the app and only 13% coming via clipper

```
order = np.array(df_bikedata_tmp.Start_Year_Month.sort_values().unique())
sb.countplot(data=df_bikedata_tmp, x='Start_Year_Month', hue='rental_access_method', c
plt.xlabel('Year/Month');
```

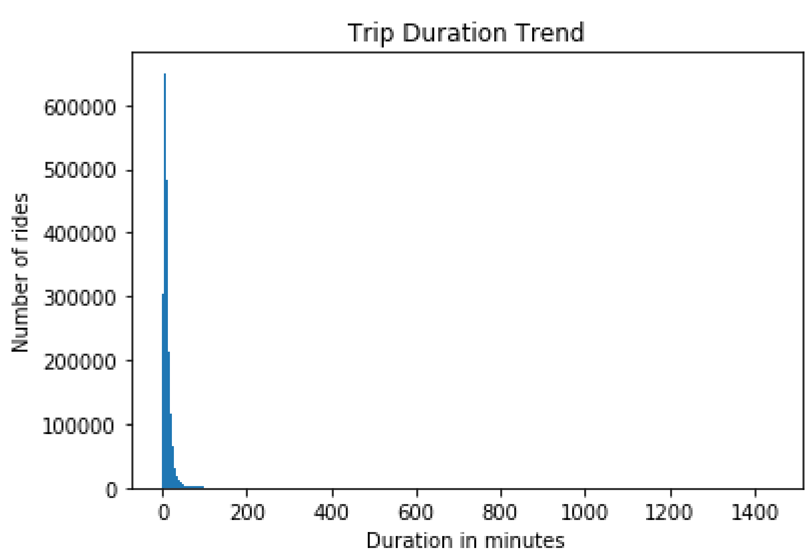
```
plt.ylabel('Number of Rides');
plt.xticks(rotation=90);
plt.legend(title = 'Rental Access Mode', bbox_to_anchor=(1,1));
plt.title('Customer Access Mode breakup over time');
```



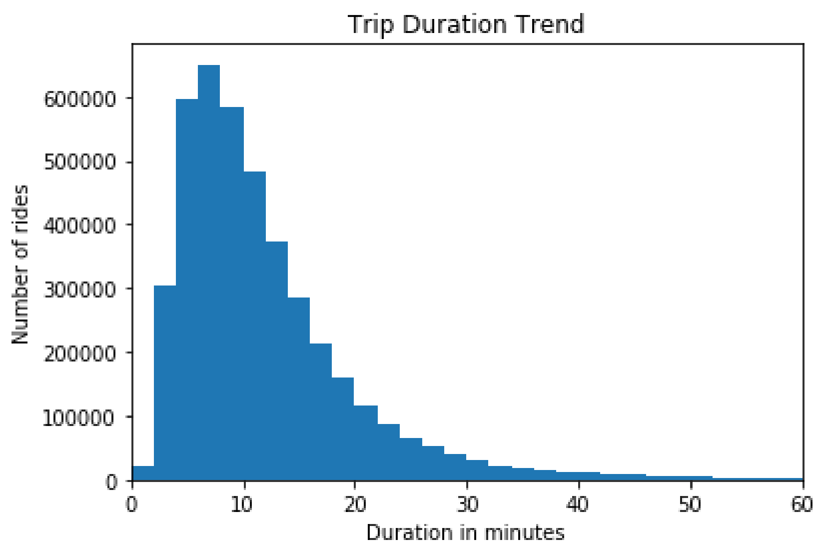
- Looks like this data is available only for the last two months. However shows that the app is defi rides

What are the most common trip durations

```
bins = np.arange(0, df_bikedata.duration_min.max()+2, 2)
plt.hist(data=df_bikedata, x='duration_min', bins=bins);
plt.xlabel('Duration in minutes');
plt.ylabel('Number of rides');
plt.title('Trip Duration Trend');
```



```
# Looks like most bike trips are much lesser than highest, lets zoom into that area
plt.hist(data=df_bikedata, x='duration_min', bins=bins);
plt.xlim((0, 60));
plt.xlabel('Duration in minutes');
plt.ylabel('Number of rides');
plt.title('Trip Duration Trend');
```

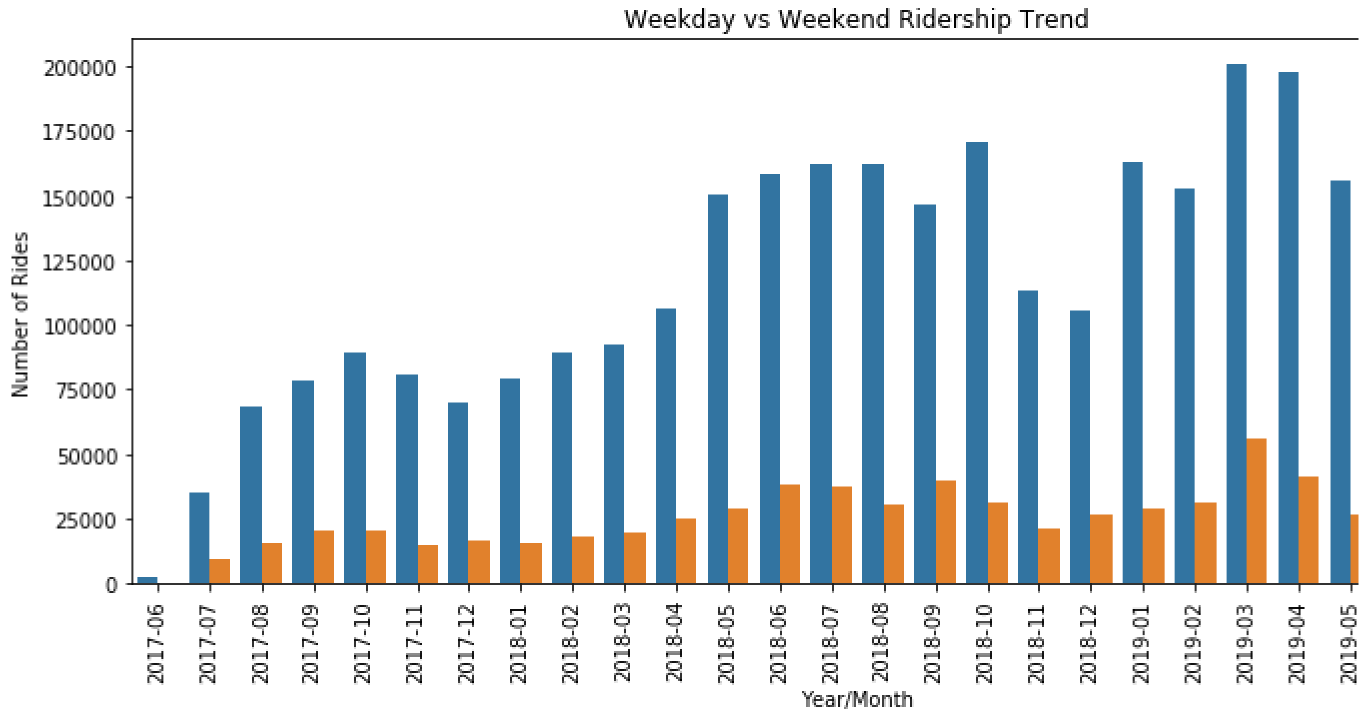


- Looks like most bike rides are around 7 and 11 minutes

Bivariate Section

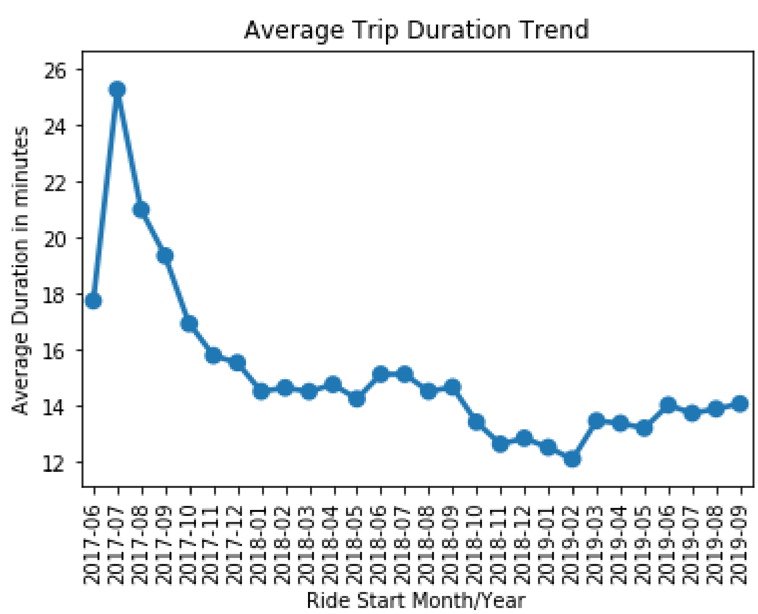
How has the number of rides on weekdays vs weekends varied over time

```
order = np.array(df_bikedata.Start_Year_Month.sort_values().unique())
plt.figure(figsize=(13,5));
sb.countplot(data=df_bikedata, x='Start_Year_Month', hue='Start_Day_Type', order=order
plt.xlabel('Year/Month');
plt.ylabel('Number of Rides');
plt.xticks(rotation=90);
plt.legend(title = 'Type of Day', bbox_to_anchor=(1,1));
plt.title('Weekday vs Weekend Ridership Trend');
```



How has the trip duration trended over time

```
order = np.array(df_bikedata.Start_Year_Month.sort_values().unique())
sb.pointplot(data=df_bikedata, x='Start_Year_Month', y='duration_min', color=base_colo
plt.xticks(rotation=90);
plt.ylabel('Average Duration in minutes');
plt.xlabel('Ride Start Month/Year');
plt.title('Average Trip Duration Trend');
```

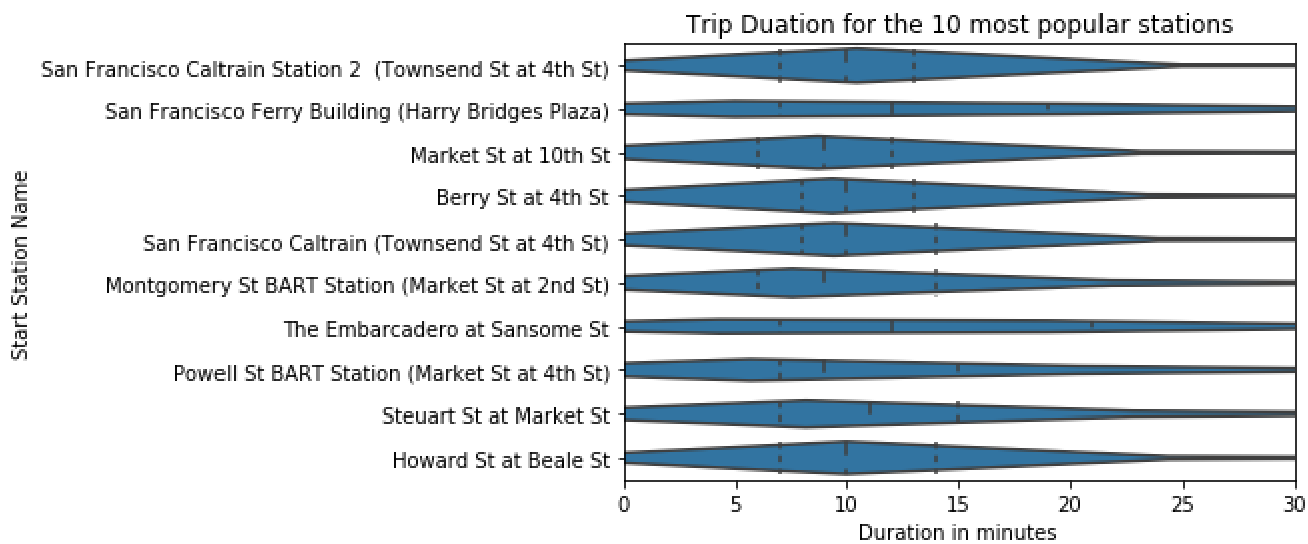


- The average ride duration seems to have started off with a bang initially and then tapered off

How does the trip duration vary by most popular starting stations

- I noticed that there is a mismatch/inaccuracies in the station id and names. I did explore the dat take quite some time to resolve all of them. Thats why i am going with station names instead

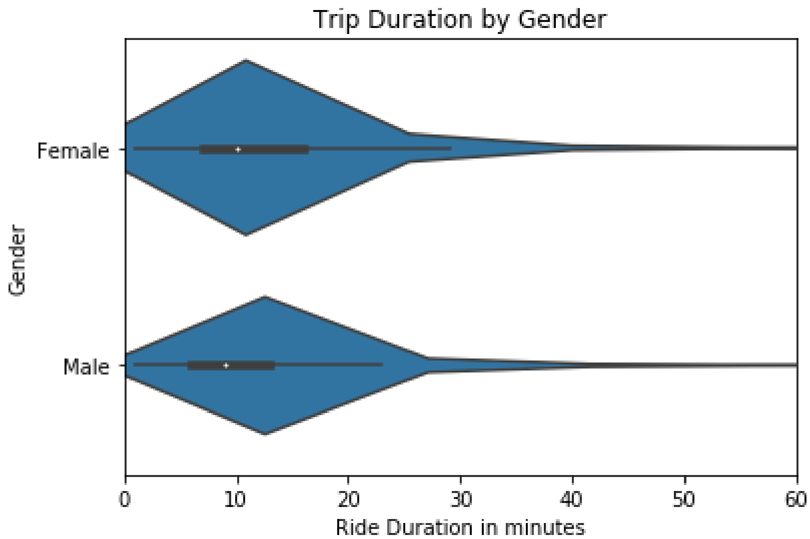
```
df_bikedata_wo_na = df_bikedata.dropna(subset=['start_station_name'])
order=df_bikedata.start_station_name.value_counts()[:10].index
sb.violinplot(data=df_bikedata_wo_na, y='start_station_name', x='duration_min', order=
plt.xlim((0, 30)); # Zooming in on the bulk of the data
plt.xlabel('Duration in minutes');
plt.ylabel('Start Station Name');
plt.title('Trip Duation for the 10 most popular stations');
```



- Most of the Mean times for these top 10 stations vary between 8 and 13 mins. Interestingly, for The Embercado station, its almost an equal spread in terms of the number of rides by duration t

How does the trip duration vary by gender

```
# Excluding gender type Other
df_bikedata_tmp = df_bikedata[df_bikedata.member_gender.isin(['Male', 'Female'])]
sb.violinplot(data=df_bikedata_tmp, y='member_gender', x='duration_min', color=base_cc
plt.ylabel('Gender');
plt.xlabel('Ride Duration in minutes');
plt.xlim((0, 60)); # Restricting to the bulk of the data
plt.title('Trip Duration by Gender');
```



- The mean ride time for the female riders is higher than men

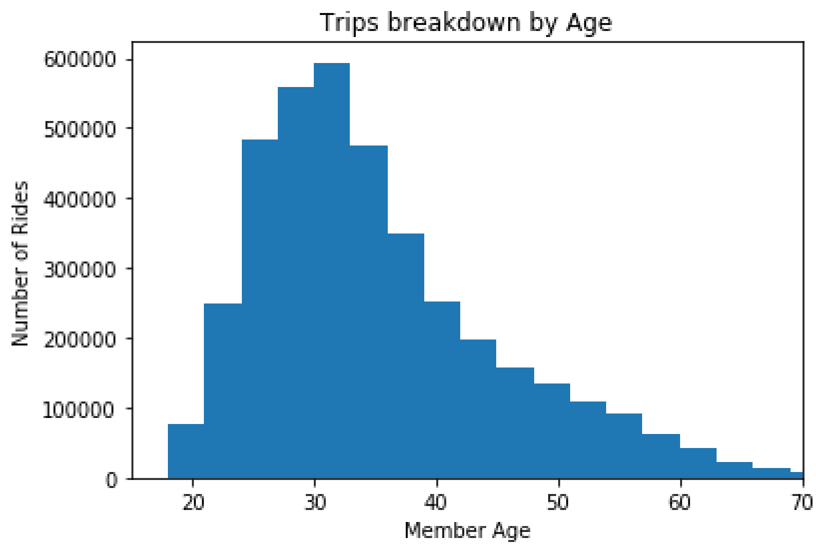
How has the trip duration varied by age

```
# Lets add a column for Member age
df_bikedata_tmp = df_bikedata[df_bikedata.member_age<100] # exlcuding the records whe

df_bikedata_tmp[['member_age', 'duration_min']].describe()
```

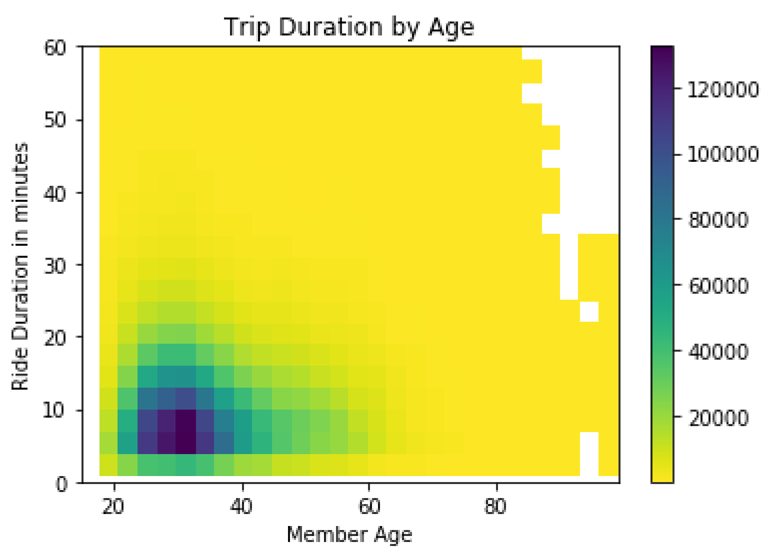
	member_age	duration_min
count	3.888425e+06	3.888425e+06
mean	3.474792e+01	1.278701e+01
std	1.015223e+01	3.180762e+01
min	1.800000e+01	1.000000e+00
25%	2.700000e+01	6.000000e+00
50%	3.200000e+01	9.000000e+00
75%	4.000000e+01	1.400000e+01
max	9.900000e+01	1.438000e+03

```
bins_x = np.arange(df_bikedata_tmp.member_age.min()-3, df_bikedata_tmp.member_age.max
bins_y = np.arange(df_bikedata_tmp.duration_min.min()-3, df_bikedata_tmp.duration_min.

plt.hist(data=df_bikedata_tmp, x='member_age', bins=bins_x);
plt.xlabel('Member Age');
```

```
plt.ylabel('Number of Rides');
plt.xlim((15, 70)); # Limiting it to the bulk of the data
plt.title('Trips breakdown by Age');
```




```
plt.hist2d(data=df_bikedata_tmp, x='member_age', y='duration_min', cmin=0.5, cmap='vi
plt.colorbar();
plt.xlabel('Member Age');
plt.ylabel('Ride Duration in minutes');
plt.ylim((0, 60)); # Limiting it to the bulk of the data
plt.title('Trip Duration by Age');
```

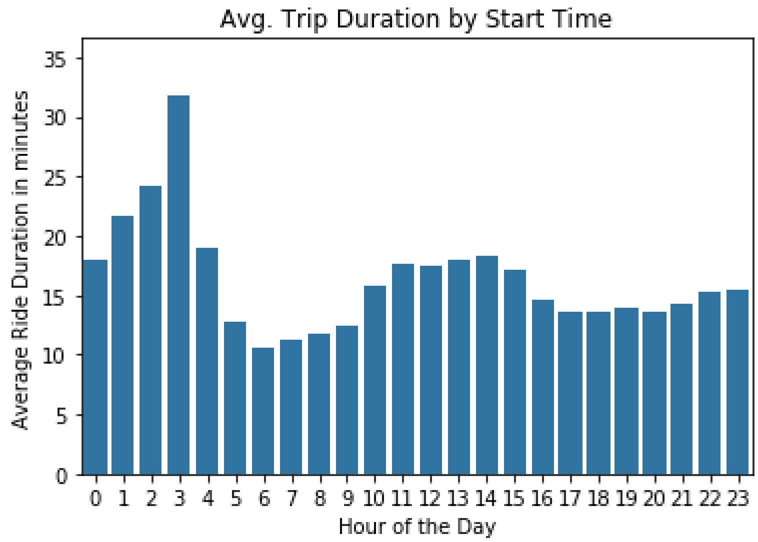


- Most people are in the age range of 25 to 35 riding between 8 and 15 mins

How has the trip duration varied by the time of day

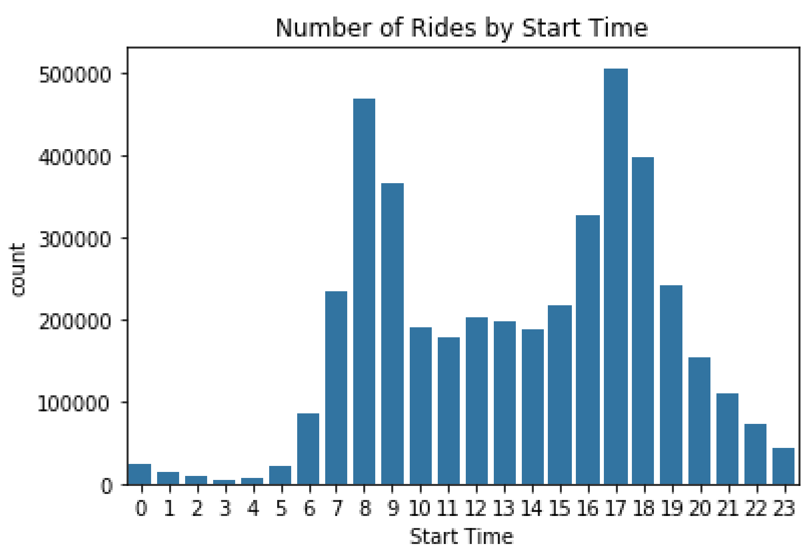
```
sb.barplot(data=df_bikedata, x=df_bikedata.start_time.dt.hour, y='duration_min', colo
plt.ylabel('Average Ride Duration in minutes');
```

```
plt.xlabel('Hour of the Day');
plt.title('Avg. Trip Duration by Start Time');
```



- Interestingly, the maximum average ride duration are for the ones that start at 3 am in the morni because there are fewer rides happening at this time causing the spike

```
sb.countplot(data=df_bikedata, x=df_bikedata.start_time.dt.hour, color=base_color)
plt.xlabel('Start Time');
plt.title('Number of Rides by Start Time');
```

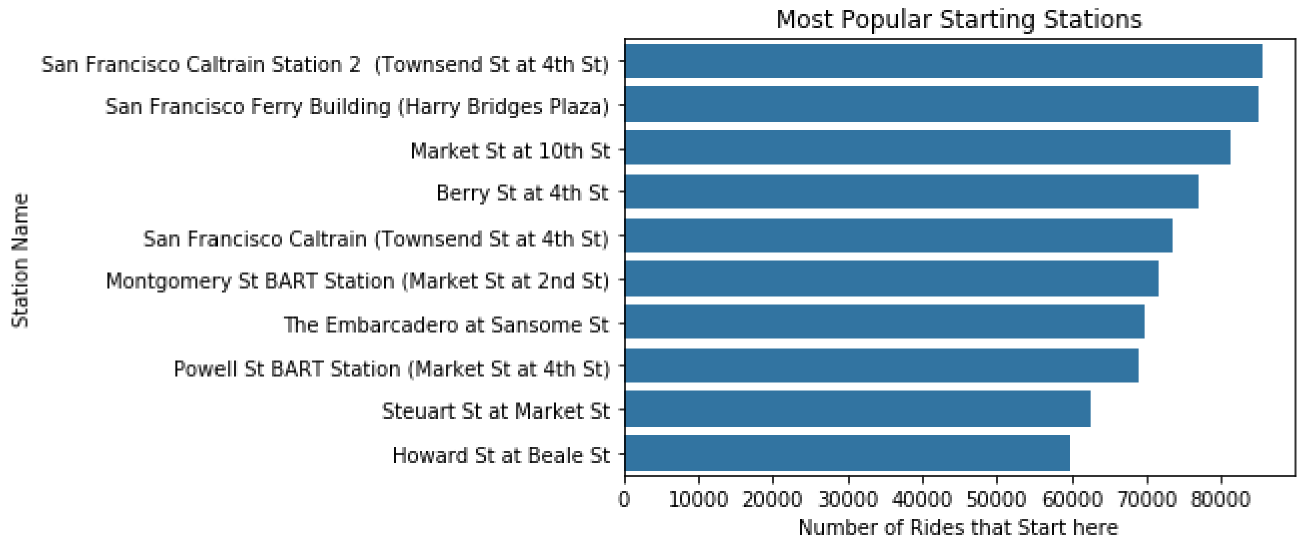


- The above plot shows how many rides were done at the hour of the day and as it shows, 3 am ha

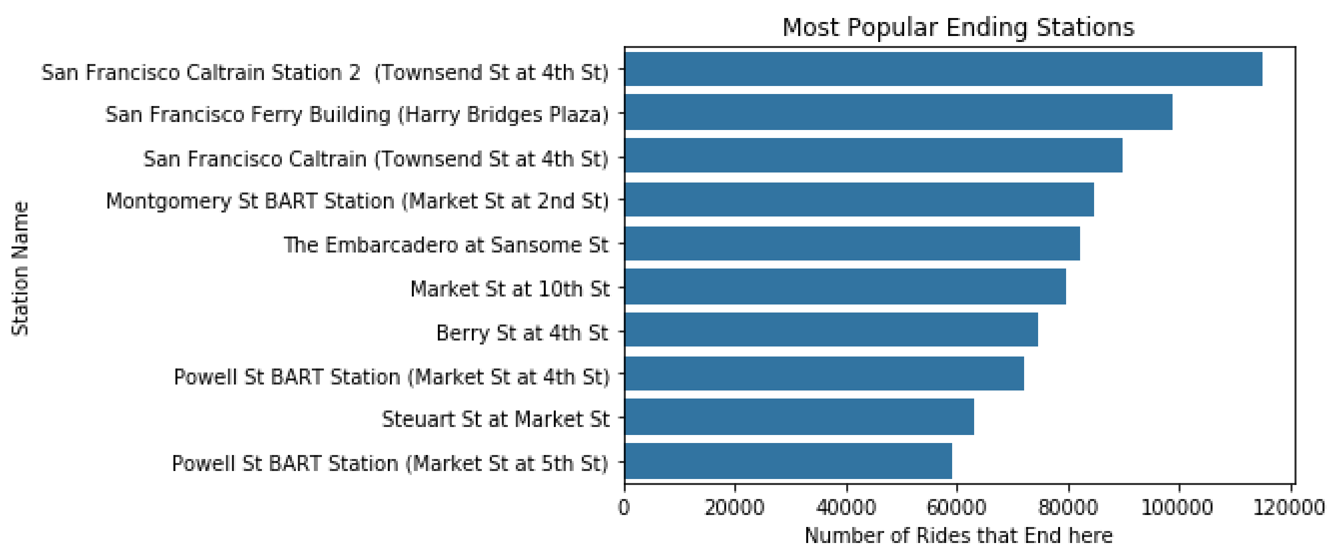
What are the most popular stations

```
# Top 10 starting stations
order = df_bikedata.start_station_name.value_counts()[:10].index
```

```
sb.countplot(data=df_bikedata, y='start_station_name', color=base_color, order=order);
plt.xlabel('Number of Rides that Start here');
plt.ylabel('Station Name');
plt.title('Most Popular Starting Stations');
```




```
# Top 10 ending stations
sb.countplot(data=df_bikedata, y='end_station_name', color=base_color, order=df_biked
plt.xlabel('Number of Rides that End here');
plt.ylabel('Station Name');
plt.title('Most Popular Ending Stations');
```

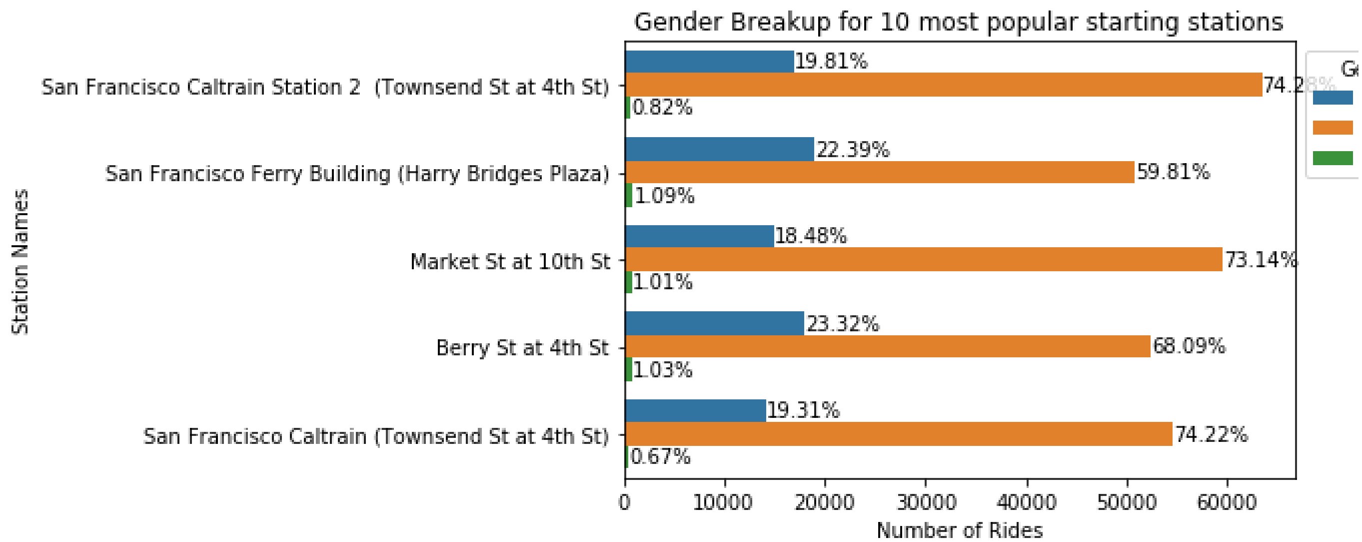


- Looks like the 'San Francisco Caltrain Station 2' is both the most popular starting and end points

What is the gender breakup for the rides from the most popular stations

```
station_counts = df_bikedata.start_station_name.value_counts()[:5]
order = station_counts.index
```

```
cplot = sb.countplot(data=df_bikedata, y='start_station_name', hue='member_gender', or
plt.xlabel('Number of Rides');
plt.ylabel('Station Names');
plt.legend(title = 'Gender', bbox_to_anchor=(1,1));
plt.title('Gender Breakup for 10 most popular starting stations');

# Adding percentages instead of absolute counts
i = 0
for p in cplot.patches:
  max_count = df_bikedata.start_station_name.value_counts()[i]
  str_pct = '{:1.2f}%'.format(100*p.get_width()/max_count)
  cplot.text(p.get_x() + p.get_width(), p.get_y()+0.2, str_pct)
  i = i + 1
  if(i%5==0):
    i = 0
```



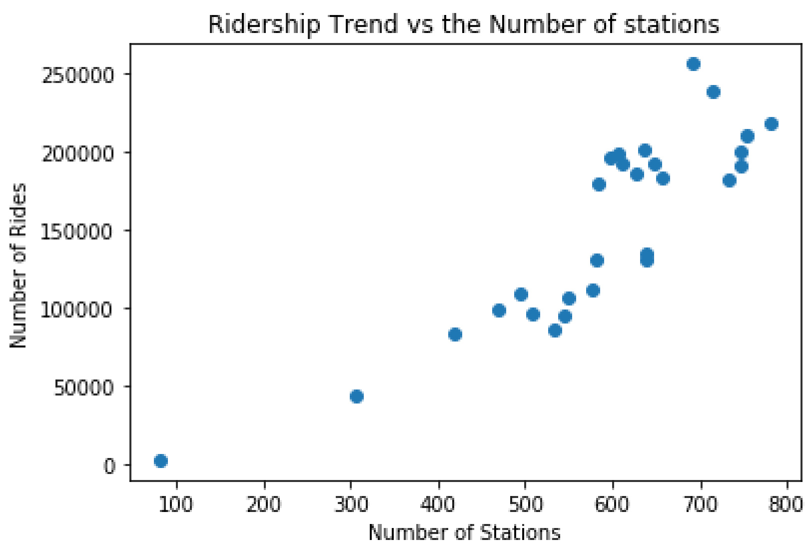
- The percentage of women riders hovers between 18% and 23% even in the most popular startin

How has the number of starting/ending stations over time affected the total num

```
df_temp = df_bikedata.groupby(['Start_Year_Month'], as_index=False).agg({'start_statio
df_temp.rename(columns={'start_station_name': 'number_of_start_stations', 'end_station
df_temp['total_stations'] = df_temp.number_of_start_stations + df_temp.number_of_end_s
df_temp.sort_values(['Start_Year_Month'], ascending=[True], inplace=True)
df_temp.head()
```

	Start_Year_Month	number_of_start_stations	number_of_end_stations	number_of_
0	2017-06	41	41	
1	2017-07	153	153	
2	2017-08	209	210	
3	2017-09	235	235	
4	2017-10	247	247	1

```
plt.scatter(data=df_temp, x='total_stations', y='number_of_rides');
plt.xlabel('Number of Stations');
plt.ylabel('Number of Rides');
plt.title('Ridership Trend vs the Number of stations');
```



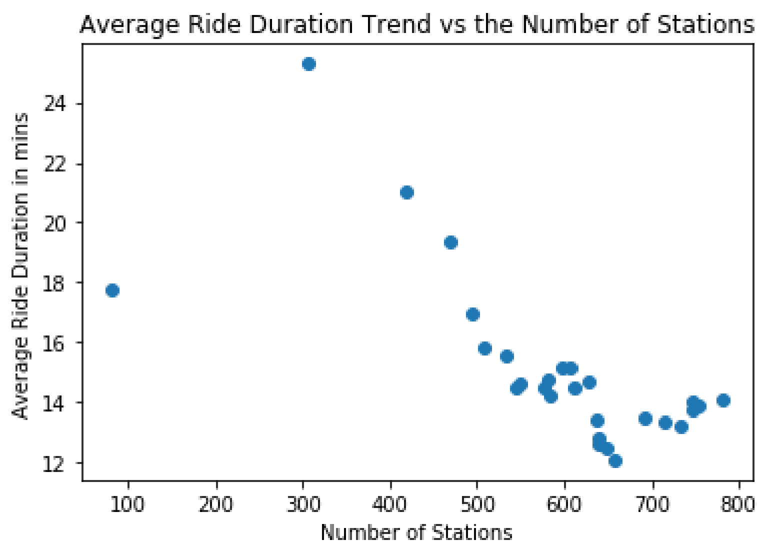
- This graph shows that the number of rides have increased with addition of stations. The critical number of stations went from 500 to 600 stations where the number of rides increased from 80

How has the average ride duration been impacted by the number of stations

```
df_temp = df_bikedata.groupby(['Start_Year_Month'], as_index=False).agg({'start_static
df_temp.rename(columns={'start_station_name': 'number_of_start_stations', 'end_statior
df_temp['total_stations'] = df_temp.number_of_start_stations + df_temp.number_of_end_s
df_temp.sort_values(['total_stations'], ascending=[True], inplace=True)
df_temp.head()
```

	Start_Year_Month	number_of_start_stations	number_of_end_stations	avg_durati
0	2017-06	41	41	17.
1	2017-07	153	153	25.
2	2017-08	209	210	21
3	2017-09	235	235	19.
4	2017-10	247	247	16.

```
plt.scatter(data=df_temp, x='total_stations', y='avg_duration_min');
plt.xlabel('Number of Stations');
plt.ylabel('Average Ride Duration in mins');
plt.title('Average Ride Duration Trend vs the Number of Stations');
```



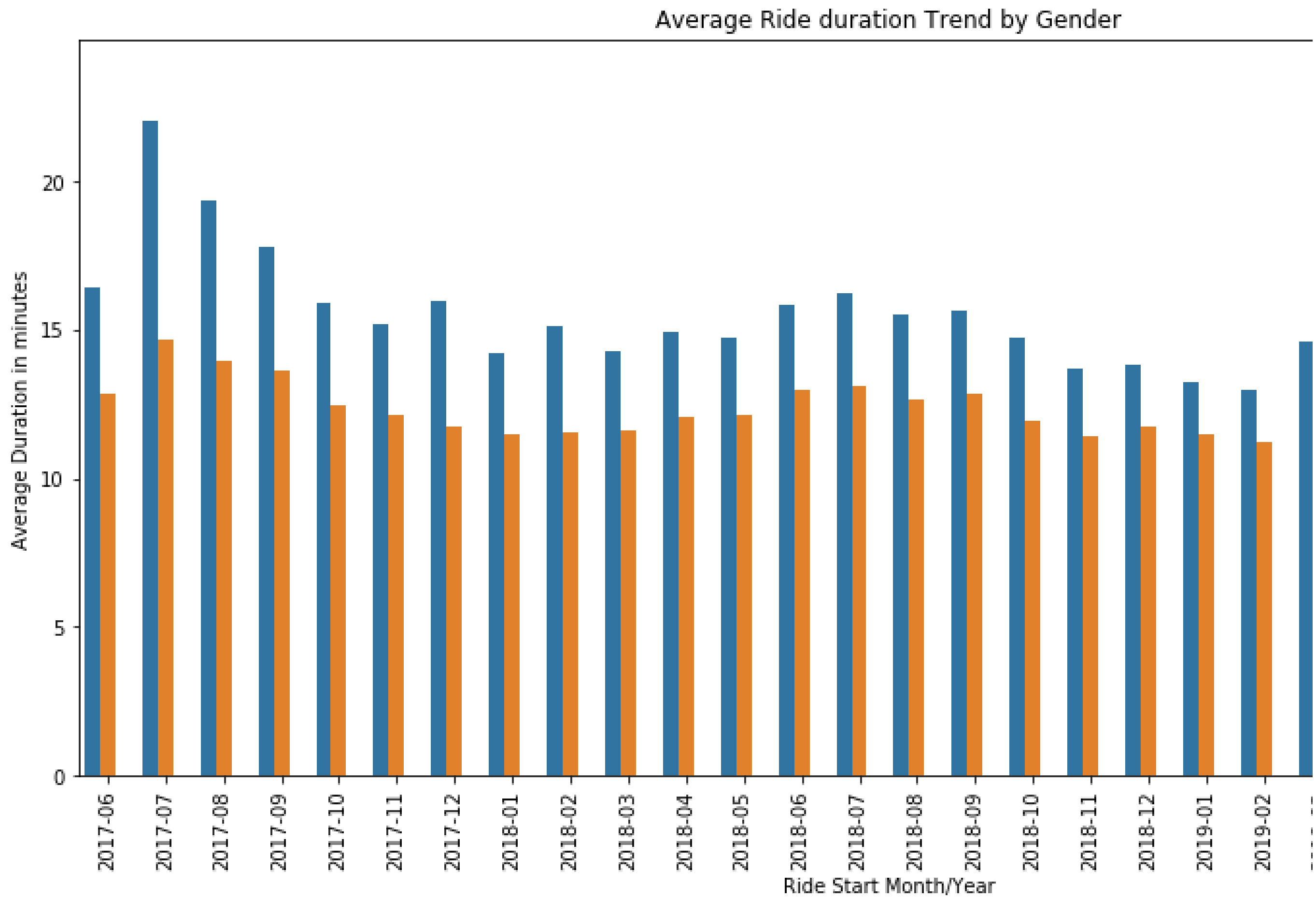
- As expected, as the number of stations increased, the number of rides increased, however due t average ride duration has decreased

Multivariate Section

How does the trip duration vary by gender over time

```
df_bikedata_tmp = df_bikedata[df_bikedata.member_gender.isin(['Male', 'Female'])]
order = np.array(df_bikedata.Start_Year_Month.sort_values().unique())
plt.figure(figsize=(15,7));
sb.barplot(data=df_bikedata_tmp, x='Start_Year_Month', y='duration_min', hue='member_g
plt.xticks(rotation=90);
plt.ylabel('Average Duration in minutes');
```

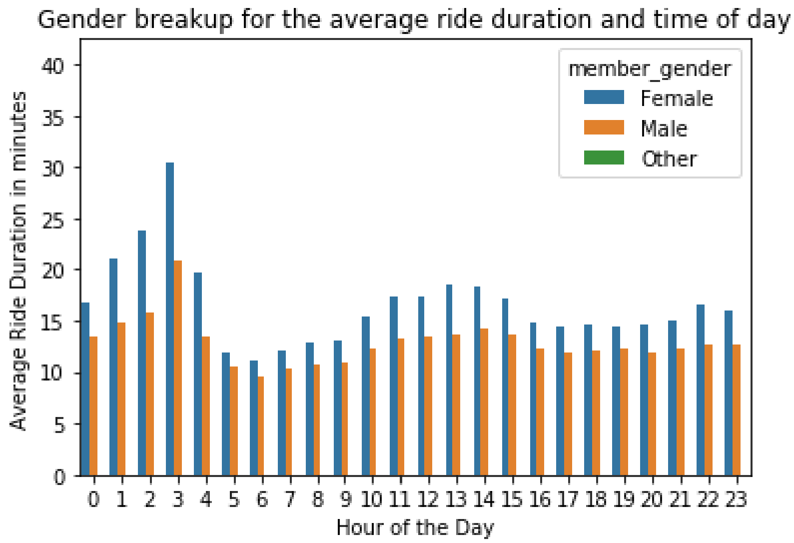
```
plt.xlabel('Ride Start Month/Year');
plt.legend(title = 'Gender', bbox_to_anchor=(1,1));
plt.title('Average Ride duration Trend by Gender');
```



- Female riders have had consistently higher average ride times than their male counterparts. The excluded as they could be Other or just incorrect data

How does the trip duration vary by the gender and start time of day

```
df_bikedata_tmp = df_bikedata[df_bikedata.member_gender.isin(['Male', 'Female'])]
sb.barplot(data=df_bikedata_tmp, x=df_bikedata.start_time.dt.hour, y='duration_min', e
plt.ylabel('Average Ride Duration in minutes');
plt.xlabel('Hour of the Day');
plt.title('Gender breakup for the average ride duration and time of day');
```

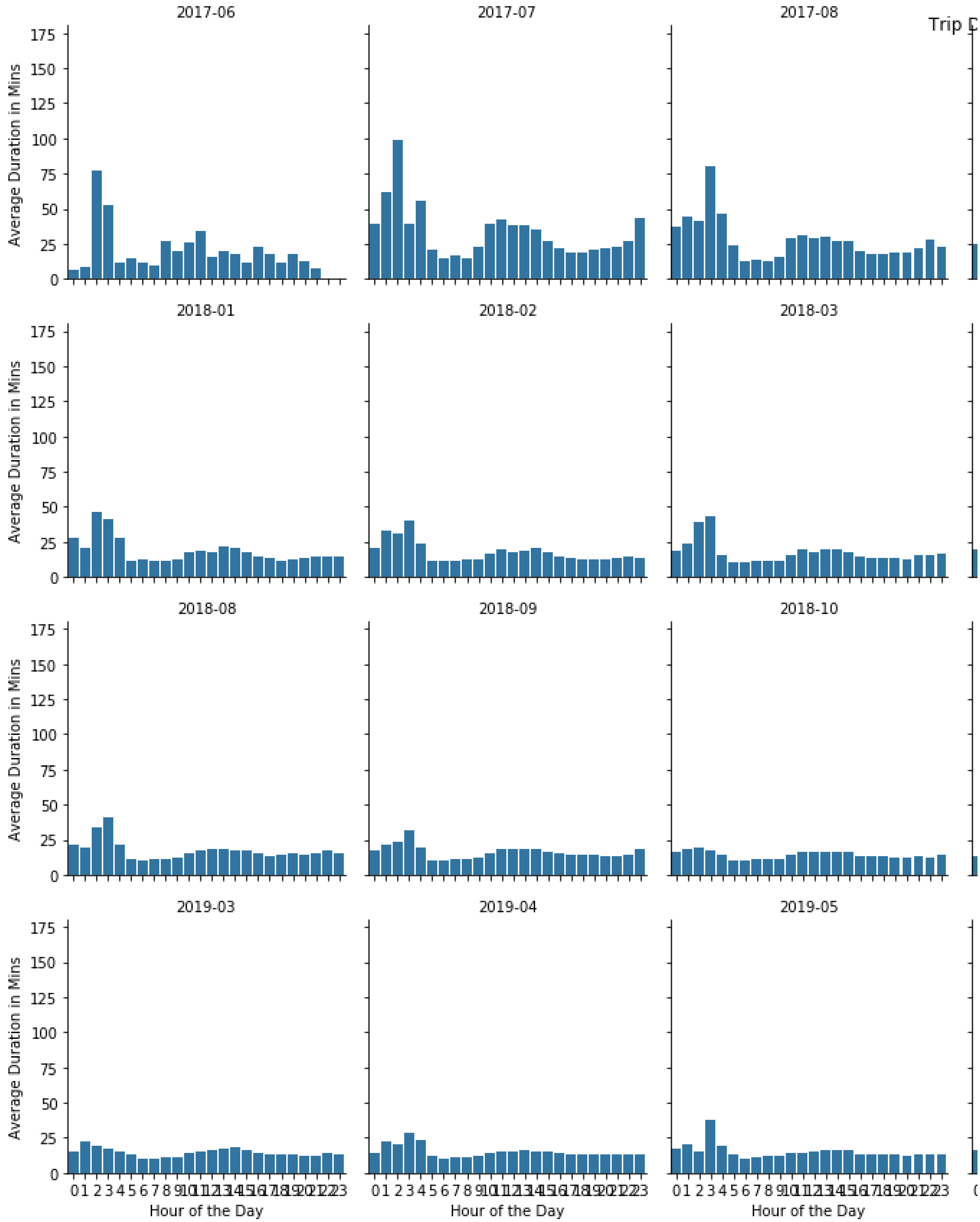


- Again, we notice that the ride duration for Female riders are higher than their male counterparts

How has the trip duration varied by the time of day over time

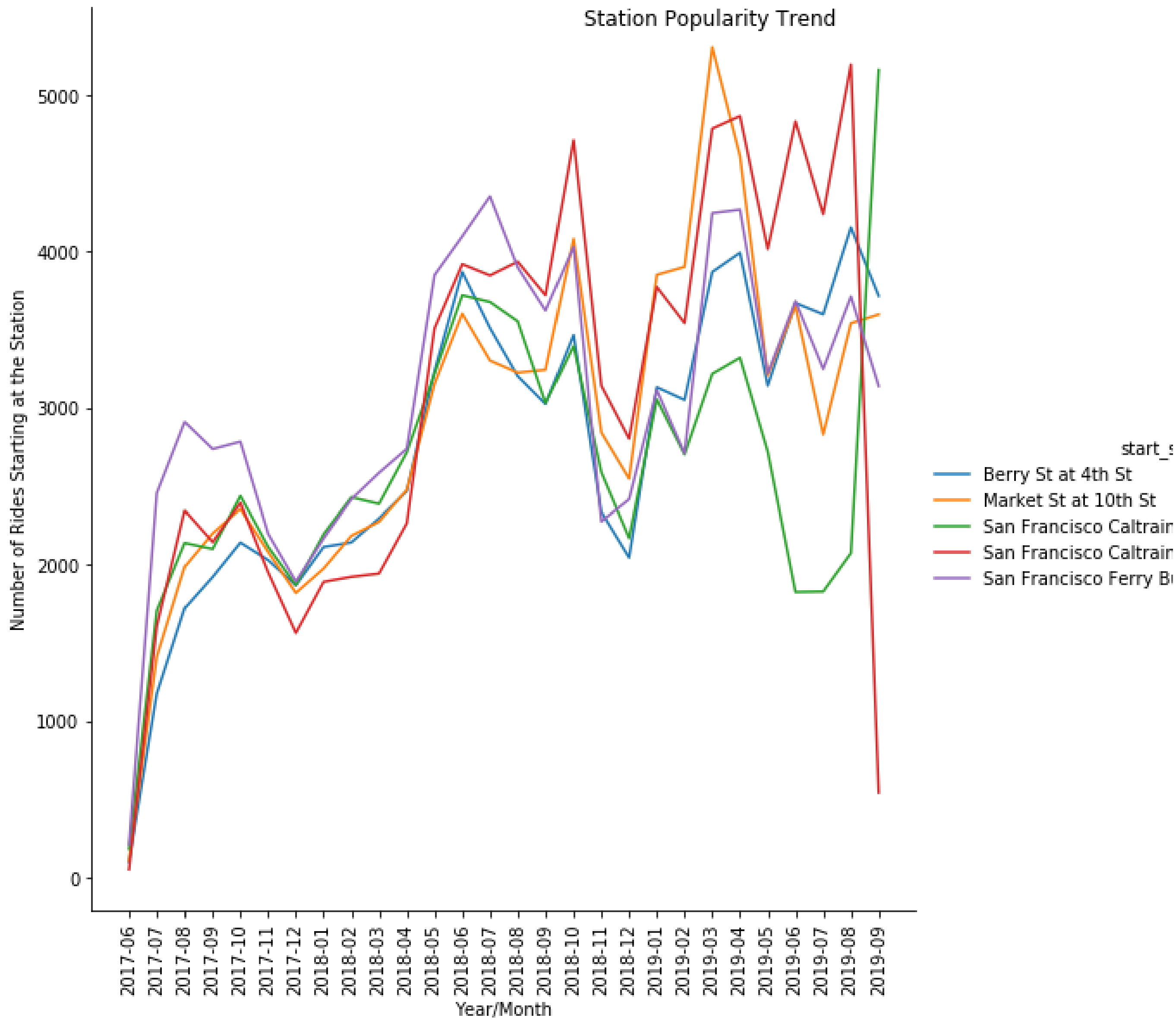
```
order = np.array(df_bikedata.Start_Year_Month.sort_values().unique())
g = sb.FacetGrid(data = df_bikedata, col = 'Start_Year_Month', col_order=order, col_wı
g.map(sb.barplot, 'Start_Time_Hour', 'duration_min', errwidth=0);
g.set_titles('{col_name}');
g.set_axis_labels("Hour of the Day", "Average Duration in Mins");
g.fig.suptitle('Trip Duration by Hour of Day Trend');
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/axisgrid.py:715: UserWarning: Usin
  warnings.warn(warning)
```



Which stations have become more popular over time

```
# Get the dataset for only top 5 starting stations
df_temp = df_bikedata[df_bikedata.start_station_name.isin(df_bikedata.start_station_na

# Group the data to get the number of rides by month and station
df_temp = df_temp.groupby(['start_station_name', 'Start_Year_Month'], as_index=False).
df_temp.rename(columns={'duration_min': 'number_of_rides'}, inplace=True)
df_temp.sort_values(['start_station_name', 'Start_Year_Month'], ascending=[True, True]

# Plot a map to show how the stations popularity has trended over time
g = sb.FacetGrid(data=df_temp, hue='start_station_name', height=8);
g.map(plt.plot, 'Start_Year_Month', 'number_of_rides');
g.set_titles('{col_name}');
g.add_legend();
g.set_axis_labels("Year/Month", "Number of Rides Starting at the Station");
for ax in g.axes.flat:
  for label in ax.get_xticklabels():
    label.set_rotation(90)
g.fig.suptitle('Station Popularity Trend');
```



- Couple of interesting insights here:
    - This plot matches with the number of rides plot over time showing the dips at the end of tł
    - For the last month of September 2019, there is a drastic drop in the number of rides in San St at 4th Street) and a similar drastic increase in the number of rides for San Francisco Cal not the same stations since they have different station id's. So there maybe some access i latter station
    - Every other stations shows a dip in the numbers other than Market St at 10th St for the mo
    - San Francisco Ferry Building station started out as the most popular station for more than overtaken by the Caltrain Station 2
    - Market St at 10th Station briefly took the crown of the most popular starting station betwe

Exploring the Number of Rides data of the Bike sharing information

Key Insights - Number of Trips

- The number of bike rides are trending upwards every year, however, the increase we see betwee service started midway between 2017 as against the whole of 2018
- Almost 82% of the rides are done during the weekdays as against only 18% over the weekend
- This trend has remained more or less the same since the start of the ride services
- Around 74% of the rides have been taken by men and only 24% by women. Somehow, the ride ha which the stations are available do not have enough women who would consider this service? T the same with some spikes over time
- The rental access mode data has been available only for the last couple of months however cus prefering the app to be the way to access the ride

Exploring the Duration aspect of the Bike Sharing data

Key Insights - Duration

- Most bike rides are between 7 and 11 mins
- Initially when the service started, the average ride duration (by month) was as high as 25 mins ir tapered down and has in the last couple of months started to rise up again
- Looking at the ride durations from the 10 most popular starting stations, the average ride durati
- On an average, female riders ride for longer durations compared to their male counterparts. This time this service was introduced.
- Looking at the age and the ride durations, most people are in the age range of 25 to 35 riding be
- Average ride durations are higher during the afternoon and later in the night
- When looking at the average ride duration trend over time for the given hour of the day, it is cons interesting point is that the average ride duration at 3 or 4 am seems unusually high. This is prob number of rides and for the people that do use them at this hour are riding for a longer period

Exploring the Station aspect of the Bike Sharing data

Key Insights - Stations

- 'San Francisco Caltrain Station 2' is both the most popular starting and end points
- The percentage of women riders hovers between 18% and 23% even in the most popular startin
- The number of rides have increased with addition of stations. The critical jump in ridership happ from 500 to 600 stations where the number of rides increased from 80,000 to almost 200,000

- As the number of stations increased, the number of rides increased, however due to the availabi duration has decreased
- For the last month of September 2019, there is a drastic drop in the number of rides in San Franc 4th Street) and a similar drastic increase in the number of rides for San Francisco Caltrain (Towr same stations since they have different station id's. So there maybe some access issues which
- Every other stations shows a dip in the numbers other than Market St at 10th St for the month o
- San Francisco Ferry Building station started out as the most popular station for more than a yea the Caltrain Station 2
- Market St at 10th Station briefly took the crown of the most popular starting station between Ja