```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

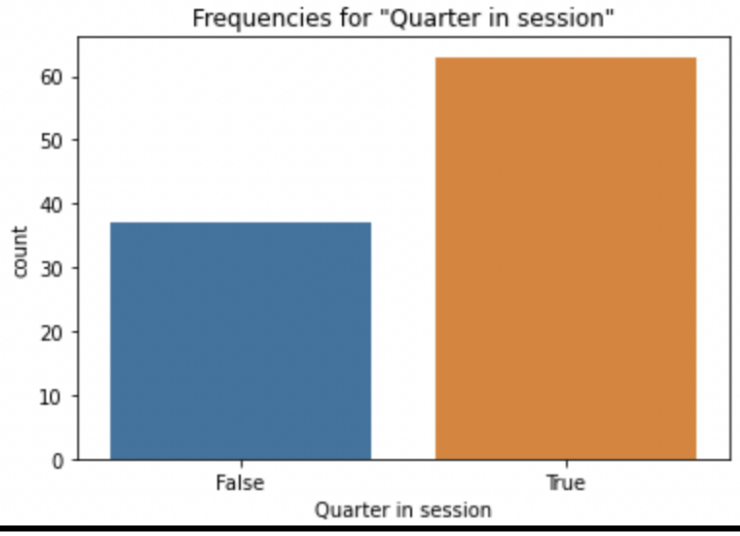
Load in Data

```
df= pd.read_csv('/Users/briannagrissom/downloads/data.csv')
# load in data

df.head()
```

	Full date	Day of week	In Person school in session	Quarter in session	Temp (avg)	Temp (high)	Temp (low)	Day Length	Bicyclist Count
0	6/9/21	Wed	False	True	66.0	78	54	14:24	603
1	6/16/21	Wed	False	False	70.0	77	63	14:27	486
2	6/23/21	Wed	False	False	68.0	76	60	14:27	464
3	6/30/21	Wed	False	False	68.0	74	62	14:25	600
4	7/7/21	Wed	False	False	67.0	75	59	14:21	654

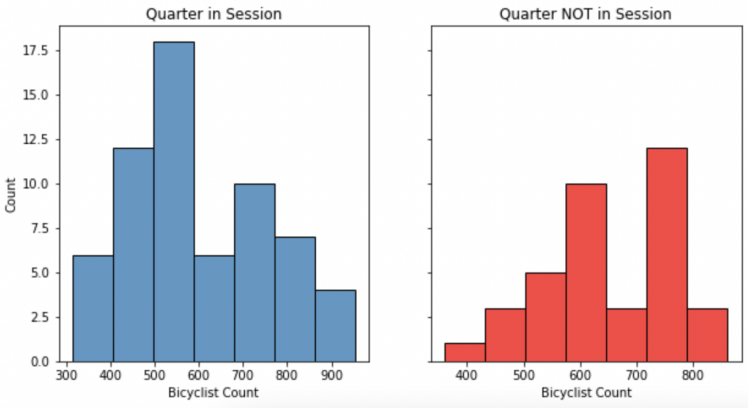
Counts Recorded for Quarter in Session & Quarter NOT in Session

```
sns.countplot(x='Quarter in session',data=df)
plt.title('Frequencies for "Quarter in session"')
plt.show()
```



Histogram of Quarter in Session vs. Quarter NOT in Session

```
fig, axes = plt.subplots(1, 2, figsize=(10,5), sharey=True)

sns.histplot(ax= axes[0], x= df_insession['Bicyclist Count'])
axes[0].set_title('Quarter in Session')

sns.histplot(ax= axes[1], x= df_notinsession['Bicyclist Count'], color='red')
axes[1].set_title('Quarter NOT in Session')
plt.show()
```

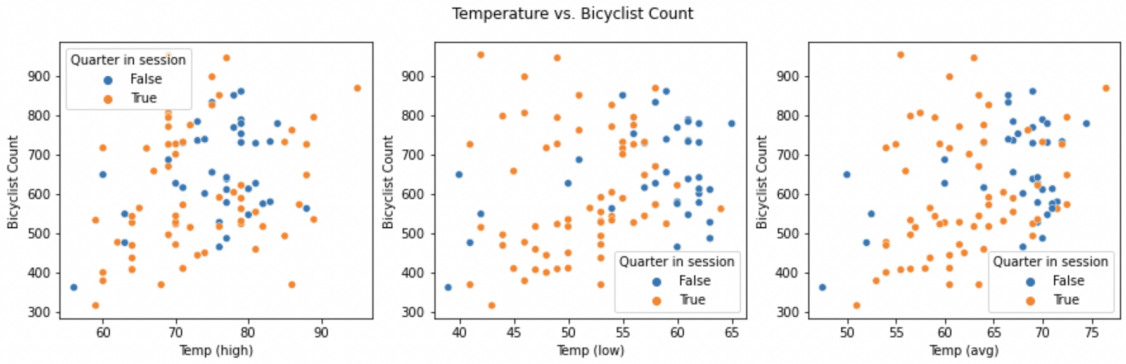


Barplot of Mean Bicyclist Count by Quarter in Session

```
]: sns.catplot(x='Quarter in session',y='Bicyclist Count',data= df,kind='bar',order= [True,False],palette=['blue','red'])
plt.title('Mean Bicyclist Count by "Quarter in session"')
plt.show()
```

High, Low, & Average Temperatures vs. Bicycle Count

```
import matplotlib.pyplot as plt
import seaborn as sns

fig, axes = plt.subplots(1,3, figsize=(15,4))
fig.suptitle('Temperature vs. Bicyclist Count')
sns.scatterplot(x= 'Temp (high)', y= 'Bicyclist Count', hue='Quarter in session', data=df, ax= axes[0])
sns.scatterplot(x= 'Temp (low)', y= 'Bicyclist Count', hue='Quarter in session', data=df, ax=axes[1])
sns.scatterplot(x= 'Temp (avg)', y= 'Bicyclist Count',hue='Quarter in session', data=df, ax= axes[2])
plt.show()
```



Mean, Minimum, and Maximum Bicyclist Counts per Month

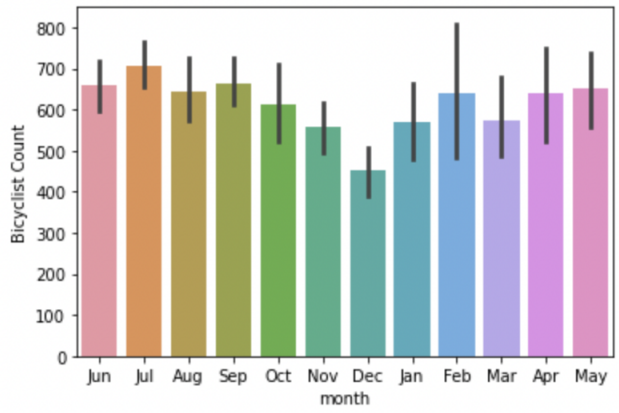
```
df['month'] = pd.DatetimeIndex(df['Full date']).month

monthdict= {1:'Jan',2:'Feb',3:'Mar',4:'Apr',5:'May',6:'Jun',7:'Jul',8:'Aug',9:'Sep',10:'Oct',11:'Nov',12:'Dec'}

df['month'] = df['month'].map(monthdict)
df.head()
df.groupby(df['month'])['Bicyclist Count'].agg(['mean','min','max']).sort_values('mean', ascending=False)

# the mean, min, max of bicyclist counts by month
```

month	mean	min	max
Jul	707.666667	579	832
Sep	664.777778	553	850
Jun	660.230769	464	783
May	651.000000	470	825
Aug	643.428571	527	860
Feb	640.625000	315	953
Apr	637.888889	368	868
Oct	613.600000	449	761
Mar	572.888889	368	897
Jan	567.777778	399	805
Nov	556.875000	410	731
Dec	450.333333	361	548

Visualization of Mean Bicyclist Counts per Month

```
sns.barplot(x='month', y='Bicyclist Count',data=df)

# to visualize the mean bicyclist counts by month
```

```
<AxesSubplot:xlabel='month', ylabel='Bicyclist Count'>
```



Mean Temperatures and Bicyclist Counts Grouped by Month

```
df.groupby(df['month'])[['Temp (avg)','Temp (high)', 'Temp (low)','Bicyclist Count']].mean().sort_values('Bicyclist Cou
# the mean temperatures and bicyclist count grouped by month
```

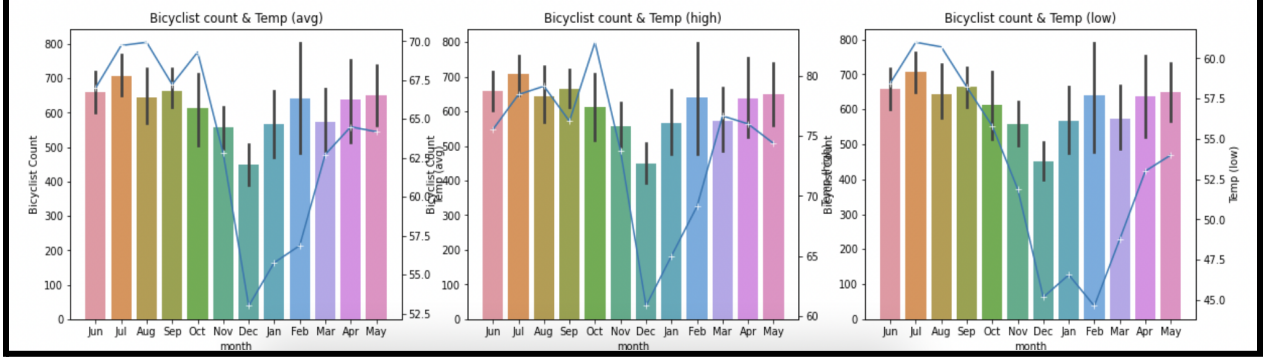
month	Temp (avg)	Temp (high)	Temp (low)	Bicyclist Count
Jul	69.722222	78.444444	61.000000	707.666667
Sep	67.222222	76.222222	58.222222	664.777778
Jun	67.000000	75.538462	58.461538	660.230769
May	64.187500	74.375000	54.000000	651.000000
Aug	69.928571	79.142857	60.714286	643.428571
Feb	56.875000	69.125000	44.625000	640.625000
Apr	64.500000	76.000000	53.000000	637.888889
Oct	69.300000	82.800000	55.800000	613.600000
Mar	62.722222	76.666667	48.777778	572.888889
Jan	55.777778	65.000000	46.555556	567.777778
Nov	62.812500	73.750000	51.875000	556.875000
Dec	53.000000	60.833333	45.166667	450.333333

Get Mean Temperatures (average, high, low) by Month

```
temp_by_month_avg = df.groupby(df['month']).mean()['Temp (avg)'].reset_index()

temp_by_month_high = df.groupby(df['month']).mean()['Temp (high)'].reset_index()
temp_by_month_low = df.groupby(df['month']).mean()['Temp (low)'].reset_index()
# get the mean tempeatures by month


temp_by_month_avg = temp_by_month_avg.reindex([6,5,1,11,10,9,2,4,3,7,0,8])
temp_by_month_high = temp_by_month_high.reindex([6,5,1,11,10,9,2,4,3,7,0,8])
temp_by_month_low = temp_by_month_low.reindex([6,5,1,11, 10,9,2,4,3,7,0,8])
# to make sure the months are in the right order to plot with the bicyclist counts
```

Visualization of Mean Bicyclist Count and Mean Temperature by Month

```
fig, axes = plt.subplots(1,3,figsize=(20,5))

sns.barplot(x='month', y='Bicyclist Count',data=df, ax=axes[0])
axes0twin= axes[0].twinx()
sns.lineplot(x='month',y='Temp (avg)', ax=axes0twin, data=temp_by_month_avg,  marker='+')
axes[0].set_title('Bicyclist count & Temp (avg)')

sns.barplot(x='month', y='Bicyclist Count',data=df, ax=axes[1])
axes1twin= axes[1].twinx()
sns.lineplot(x='month',y='Temp (high)', data=temp_by_month_high, ax=axes1twin, marker='+')
axes[1].set_title('Bicyclist count & Temp (high)')

sns.barplot(x='month', y='Bicyclist Count',data=df, ax=axes[2])
axes2twin= axes[2].twinx()
sns.lineplot(x='month',y='Temp (low)',data=temp_by_month_low, ax=axes2twin, marker='+')
axes[2].set_title('Bicyclist count & Temp (low)')



plt.show()

# seems like there's at least some correlation between bicyclist count and temperature
```



Mean Average, Low, and High Temperatures by Quarter In Session

```
df.loc[df['Quarter in session']==True,'Temp (avg)'].mean()
df.loc[df['Quarter in session']==True,'Temp (high)'].mean()
df.loc[df['Quarter in session']==True,'Temp (low)'].mean()
# 62, 73, 51 mean avg, low, and high temp respectively, & 599 mean bicyclist count when quarter in session

df.loc[df['Quarter in session']==False,'Temp (avg)'].mean()
df.loc[df['Quarter in session']==False,'Temp (high)'].mean()
df.loc[df['Quarter in session']==False,'Temp (low)'].mean()
# 67, 76, 57  mean avg, low, and high temp respectively, & 655 mean bicyclist count when quarter NOT in session
```

Quarter in Session: 62, 73, 51, avg, low, and high temperature respectively. 599 mean bicyclist count.

Quarter NOT in Session: 67, 76, 57, avg, low, and high temperature respectively. 655 mean bicyclist count.

Visualization of Day Length vs. Bicyclist Count

```
sns.scatterplot(x='Day Length',y='Bicyclist Count',data= df, hue='Quarter in session', hue_order=[True,False], palette=
plt.title('Day Length vs. Bicyclist Count')
```

```
Text(0.5, 1.0, 'Day Length vs. Bicyclist Count')
```