

ML - Exp 6 - Portland Oregon Riders Monthly Data

```
#importing necessary libraries
import numpy as np
import pandas as pd
import warnings
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn import metrics
import matplotlib.pyplot as plt
warnings.filterwarnings('ignore')

from google.colab import drive
drive.mount("/content/gdrive")

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call drive.mount("/content/gdrive")

riders = pd.read_csv('/content/gdrive/My Drive/datasets/portland.csv',encoding= 'unicode_escape')

riders.head()
```

	Month	Portland Oregon average monthly bus ridership (/100) January 1973 through June 1982, n=114
0	1960-01	648
1	1960-02	646
2	1960-03	639

```
riders.shape

(115, 2)

riders = riders.rename(columns={"Portland Oregon average monthly bus ridership (/100) January 1973 through June 1982, n=114": "Ridership (/100)"})

riders.head()
```

	Month	bus ridership
0	1960-01	648

riders.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 0 to 114
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Month           115 non-null   object
1   bus ridership    115 non-null   object
dtypes: object(2)
memory usage: 1.9+ KB
```

riders.describe()

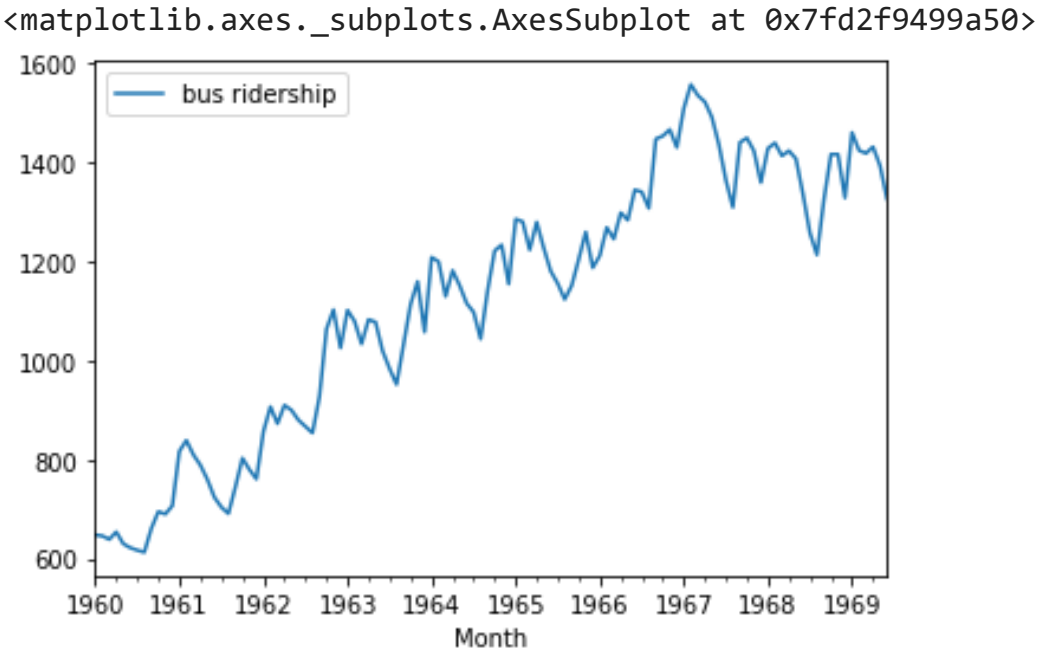
	Month	bus ridership
count	115	115
unique	115	112
top	1962-12	1417
freq	1	2

riders = riders[:-1]

riders['Month'] = pd.to_datetime(riders['Month'],format='%Y/%m')

riders['bus ridership'] = riders['bus ridership'].astype(int)

riders.plot.line(x='Month',y='bus ridership')



we can observe that average number of riders is increasing most of time

df1=riders

mon = df1['Month']

```
print(mon.shape)
print(mon.head(2))
```

```
print(mon.head(2))

(114,)
0    1960-01-01
1    1960-02-01
Name: Month, dtype: datetime64[ns]
```

```
temp=pd.DatetimeIndex(mon)
```

```
temp

DatetimeIndex(['1960-01-01', '1960-02-01', '1960-03-01', '1960-04-01',
              '1960-05-01', '1960-06-01', '1960-07-01', '1960-08-01',
              '1960-09-01', '1960-10-01',
              ...
              '1968-09-01', '1968-10-01', '1968-11-01', '1968-12-01',
              '1969-01-01', '1969-02-01', '1969-03-01', '1969-04-01',
              '1969-05-01', '1969-06-01'],
              dtype='datetime64[ns]', name='Month', length=114, freq=None)
```

```
month = pd.Series(temp.month)
```

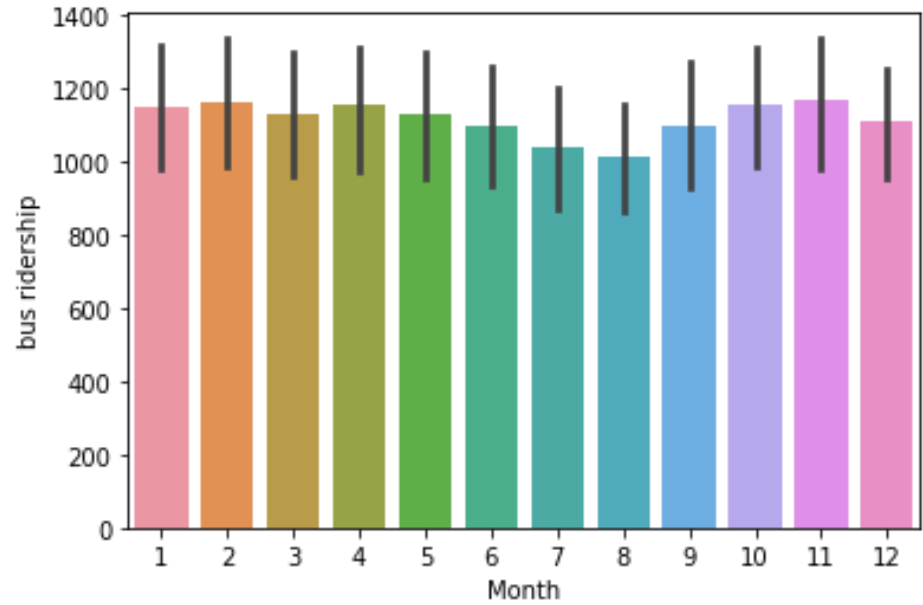
```
df1=df1.drop(['Month'],axis=1)
```

```
df1=df1.join(month)
```

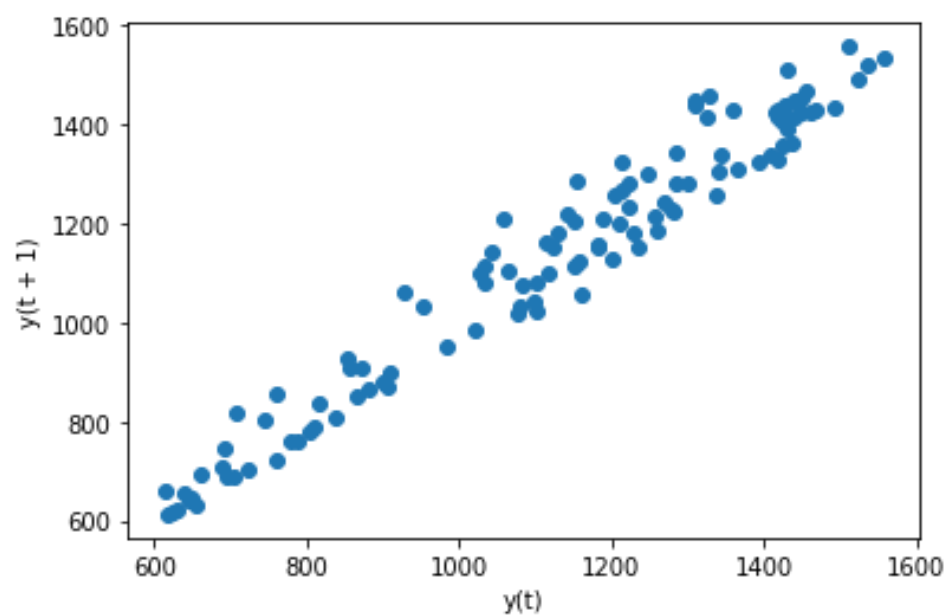
```
df1.head()
```

	bus ridership	Month
0	648	1
1	646	2
2	639	3
3	654	4
4	630	5

```
sns.barplot(x='Month',y='bus ridership',data=df1)
plt.show()
```



```
pd.plotting.lag_plot(riders['bus ridership'])
plt.show()
```



```
riders = riders.set_index('Month')
```

```
import statsmodels.api as sm
mod = sm.tsa.SARIMAX(riders['bus ridership'], trend='n', order=(0,1,0), seasonal_order=(1,1,1,12))
results = mod.fit()
print(results.summary())
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tsa/base/tsa_model.py:165: ValueWarning: No frequency in
% freq, ValueWarning)
```

Statespace Model Results

```
=====
Dep. Variable:          bus ridership      No. Observations:          114
Model:                SARIMAX(0, 1, 0)x(1, 1, 1, 12)  Log Likelihood            -501.340
Date:                  Fri, 05 Mar 2021      AIC                       1008.680
Time:                  13:46:09              BIC                       1016.526
Sample:                01-01-1960           HQIC                      1011.856
                   - 06-01-1969
```

```
Covariance Type:          opg
```

```
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.S.L12      0.3236      0.186      1.739      0.082      -0.041      0.688
ma.S.L12     -0.9991     48.597     -0.021      0.984     -96.247     94.249
sigma2      984.6342    4.77e+04      0.021      0.984    -9.26e+04    9.45e+04
=====
```

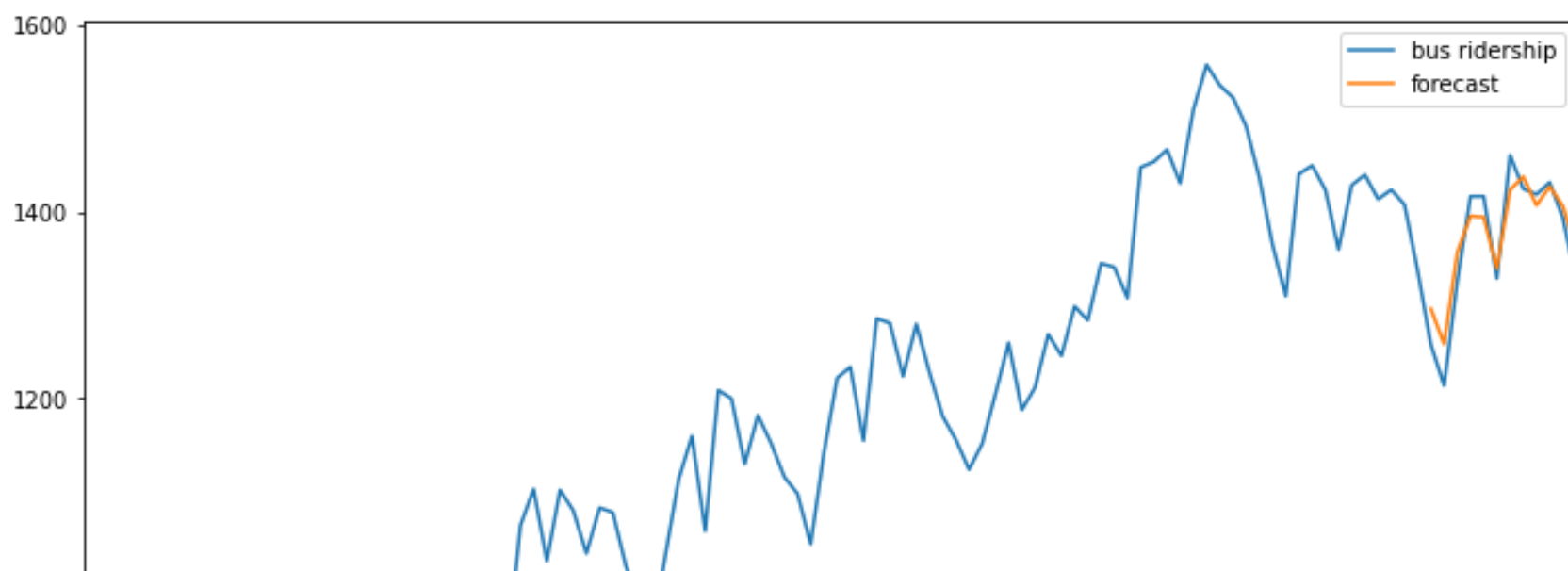
```
=====
Ljung-Box (Q):          36.56   Jarque-Bera (JB):          4.81
Prob(Q):                0.63   Prob(JB):                0.09
Heteroskedasticity (H):  1.48   Skew:                0.38
Prob(H) (two-sided):    0.26   Kurtosis:            3.75
=====
```

```
Warnings:
```

```
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

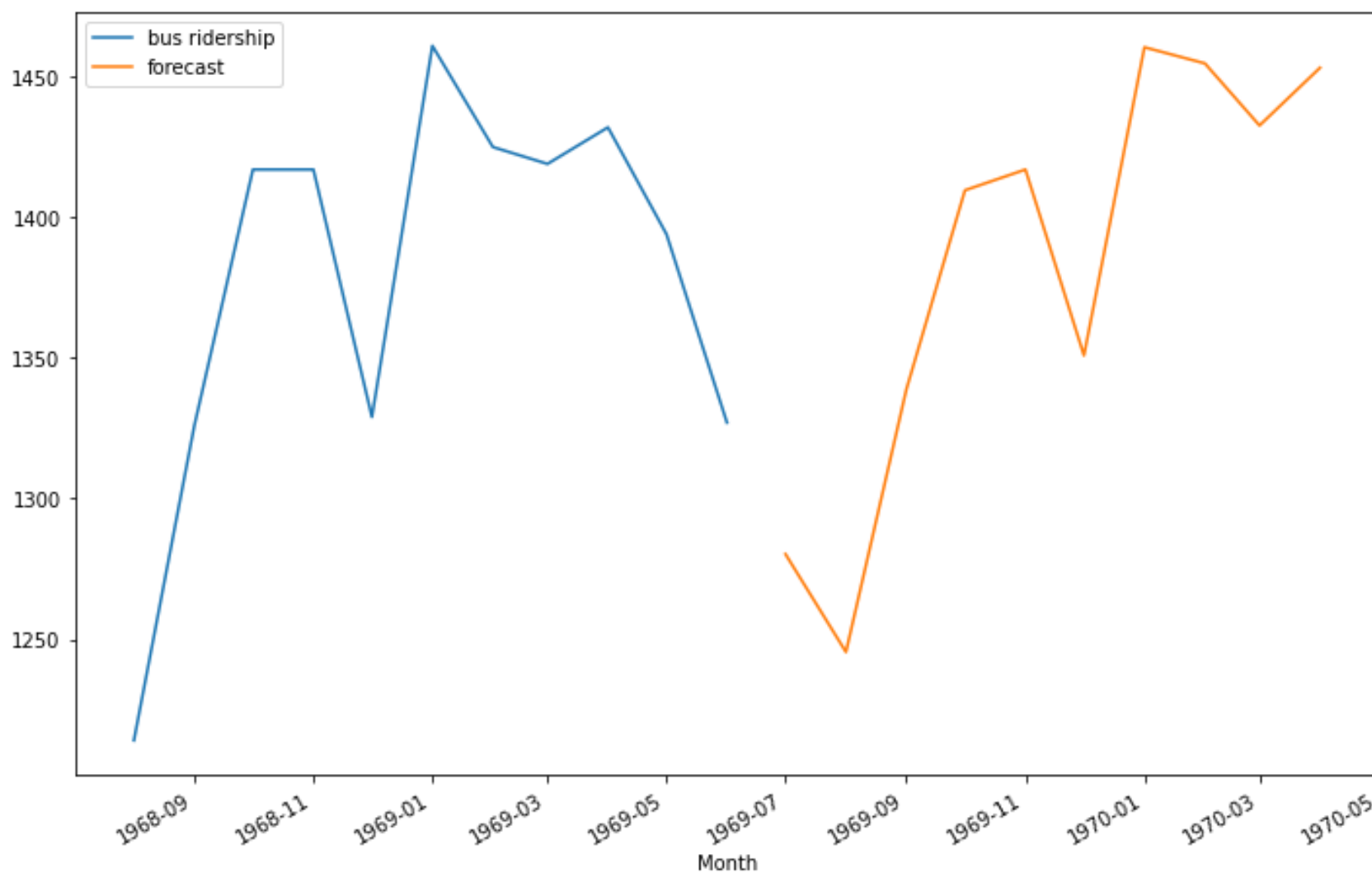


```
riders['forecast'] = results.predict(start = 102, end= 120, dynamic= True)
riders[['bus ridership', 'forecast']].plot(figsize=(12, 8))
plt.show()
```



```
def forecasting_future_months(df, no_of_months):
    df_perdict = riders.reset_index()
    mon = df_perdict['Month']
    mon = mon + pd.DateOffset(months = no_of_months)
    future_dates = mon[-no_of_months -1:]
    df_perdict = df_perdict.set_index('Month')
    future = pd.DataFrame(index=future_dates, columns= df_perdict.columns)
    df_perdict = pd.concat([df_perdict, future])
    df_perdict['forecast'] = results.predict(start = 114, end = 125, dynamic= True)
    df_perdict[['bus ridership', 'forecast']].iloc[-no_of_months - 12:].plot(figsize=(12, 8))
    plt.show()
    return df_perdict[-no_of_months:]
```

```
predicted = forecasting_future_months(riders,10)
```



```
riders.tail()
```

	bus ridership	forecast
1969-02-01	1425	1438.093754
1969-03-01	1419	1407.278872
1969-04-01	1432	1427.493545
1969-05-01	1394	1406.616299
1969-06-01	1327	1362.733941