

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.formula.api import ols
from scipy import stats
import statsmodels.api as sm
from datetime import datetime
```

```
In [2]: pip install pmdarima
```

Requirement already satisfied: pmdarima in c:\users\forbi carine\anaconda3\lib\site-packages (1.8.5)
Requirement already satisfied: numpy>=1.19.3 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (1.20.3)
Requirement already satisfied: joblib>=0.11 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (1.1.0)
Requirement already satisfied: Cython!=0.29.18,>=0.29 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (0.29.24)
Requirement already satisfied: scipy>=1.3.2 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (1.7.1)
Requirement already satisfied: pandas>=0.19 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (1.3.4)
Requirement already satisfied: urllib3 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (1.26.7)
Requirement already satisfied: statsmodels!=0.12.0,>=0.11 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (0.12.2)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (58.0.4)
Requirement already satisfied: scikit-learn>=0.22 in c:\users\forbi carine\anaconda3\lib\site-packages (from pmdarima) (0.24.2)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\forbi carine\anaconda3\lib\site-packages (from pandas>=0.19->pmdarima) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in c:\users\forbi carine\anaconda3\lib\site-packages (from pandas>=0.19->pmdarima) (2021.3)
Requirement already satisfied: six>=1.5 in c:\users\forbi carine\anaconda3\lib\site-packages (from python-dateutil>=2.7.3->pandas>=0.19->pmdarima) (1.16.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\forbi carine\anaconda3\lib\site-packages (from scikit-learn>=0.22->pmdarima) (2.2.0)
Requirement already satisfied: patsy>=0.5 in c:\users\forbi carine\anaconda3\lib\site-packages (from statsmodels!=0.12.0,>=0.11->pmdarima) (0.5.2)
Note: you may need to restart the kernel to use updated packages.

```
In [3]: df = pd.read_excel (r'C:\Users\FORBI CARINE\Desktop\Coursera\Cleaned hospital 1.xlsx')
df.tail()
```

```
Out[3]:
```

	appointment_date	clinic_name	patient_id	practitioner_id	appointment_duration_min	revenue
4056	2021-06-28	clinic_2	42155	748	30	
4057	2021-06-28	clinic_2	42146	748	30	
4058	2021-06-28	clinic_2	42789	741	30	
4059	2021-06-28	clinic_2	42157	741	30	
4060	2021-06-28	clinic_2	42774	741	15	

```
In [4]:
```

```
df.shape
```

Out[4]: (4061, 6)

```
In [5]: df.dtypes
```

Out[5]: appointment_date datetime64[ns]
clinic_name object
patient_id int64
practitioner_id int64
appointment_duration_min int64
revenues_from_appointment float64
dtype: object

```
In [6]: df_copy = df.copy()
```

```
In [7]: df["appointment_date"] = pd.to_datetime(df.appointment_date, errors='coerce')
```

```
In [8]: df.dtypes
```

Out[8]: appointment_date datetime64[ns]
clinic_name object
patient_id int64
practitioner_id int64
appointment_duration_min int64
revenues_from_appointment float64
dtype: object

```
In [9]: df.set_index('appointment_date')
```

Out[9]:

	clinic_name	patient_id	practitioner_id	appointment_duration_min	revenues_from
appointment_date					
2021-12-30	clinic_1	43859	756	30	
2021-12-30	clinic_1	53516	756	60	
2021-12-30	clinic_1	43323	756	90	
2021-12-30	clinic_2	70058	1030	30	
2021-12-30	clinic_1	44924	756	30	
...	
2021-06-28	clinic_2	42155	748	30	
2021-06-28	clinic_2	42146	748	30	
2021-06-28	clinic_2	42789	741	30	
2021-06-28	clinic_2	42157	741	30	
2021-06-28	clinic_2	42774	741	15	

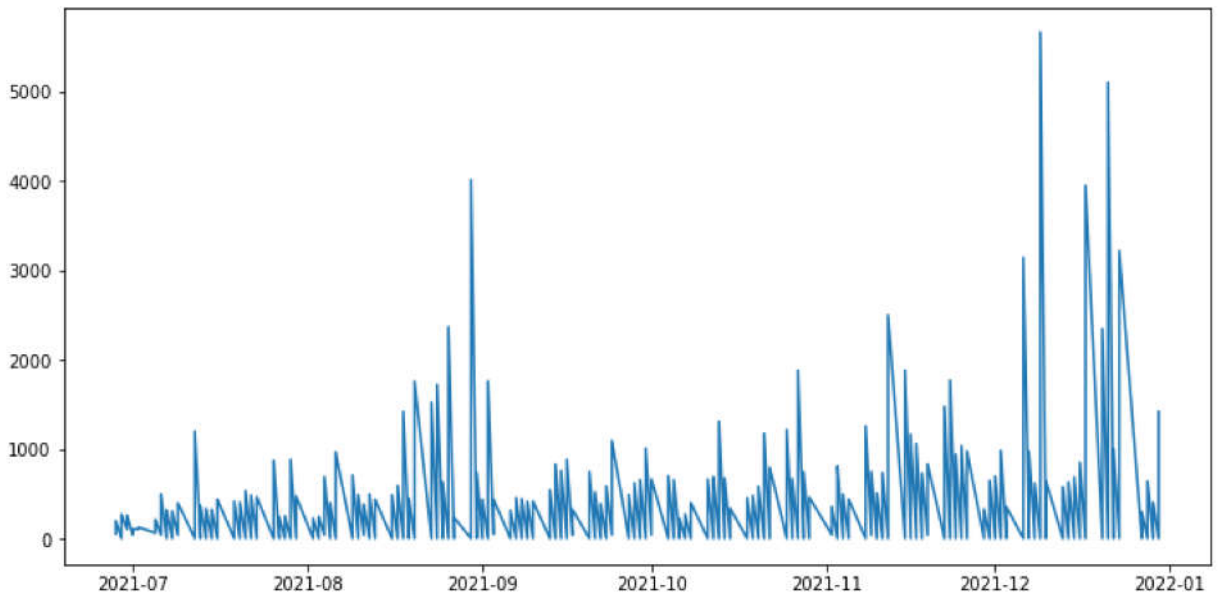
4061 rows × 5 columns



```
In [10]:
```

```
plt.figure(figsize = (12, 6))
plt.plot(df["appointment_date"], df["revenues_from_appointment"])
```

Out[10]: [`<matplotlib.lines.Line2D at 0x23b5ec5eee0>`]



In [11]: `from statsmodels.tsa.stattools import adfuller`

In [12]: `test_results = adfuller(df["revenues_from_appointment"])`

In [13]: `#Null hyothesis; The series is non stationary`
`def adfuller_test(revenue):`
 `result = adfuller(revenue)`
 `labels = ['ADF Test Statistic', 'p-value', 'lags used', 'num oservations']`
 `for value, label in zip(result, labels):`
 `print(label+' : '+str(value))`
 `if result[1] <= 0.05:`
 `print('reject null')`
 `else:`
 `print('weak evidence against null')`

In [14]: `adfuller_test(df['revenues_from_appointment'])`

```
ADF Test Statistic : -28.996587895200864
p-value : 0.0
lags used : 2
num oservations : 4058
reject null
```

In [15]: `from pmdarima.arma.utils import ndiffs`
`ndiffs(df.revenues_from_appointment, test = 'adf')`

Out[15]: 0

In [16]: `from pmdarima import auto_arima`
`import warnings`
`warnings.filterwarnings("ignore")`

```
In [17]: arima_models = auto_arima(df["revenues_from_appointment"], trace = True, suppress_warnings=True)
         arima_models.summary()
```

Performing stepwise search to minimize aic

```
ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=inf, Time=2.72 sec
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=57167.060, Time=0.11 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=56493.734, Time=0.34 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=55950.219, Time=1.07 sec
ARIMA(0,1,0)(0,0,0)[0]          : AIC=57165.066, Time=0.06 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=inf, Time=1.85 sec
ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=inf, Time=1.86 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=inf, Time=3.10 sec
ARIMA(0,1,1)(0,0,0)[0]          : AIC=55948.233, Time=0.38 sec
ARIMA(1,1,1)(0,0,0)[0]          : AIC=inf, Time=0.85 sec
ARIMA(0,1,2)(0,0,0)[0]          : AIC=inf, Time=0.95 sec
ARIMA(1,1,0)(0,0,0)[0]          : AIC=56491.742, Time=0.19 sec
ARIMA(1,1,2)(0,0,0)[0]          : AIC=inf, Time=1.60 sec
```

Best model: ARIMA(0,1,1)(0,0,0)[0]

Total fit time: 15.081 seconds

Out[17]:

SARIMAX Results

Dep. Variable:	y	No. Observations:	4061
Model:	SARIMAX(0, 1, 1)	Log Likelihood	-27972.117
Date:	Sun, 05 Jun 2022	AIC	55948.233
Time:	17:19:18	BIC	55960.851
Sample:	0	HQIC	55952.703
	- 4061		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.7427	0.006	-124.640	0.000	-0.754	-0.731
sigma2	5.648e+04	126.767	445.545	0.000	5.62e+04	5.67e+04

Ljung-Box (L1) (Q):	75.93	Jarque-Bera (JB):	6655544.28
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.36	Skew:	11.34
Prob(H) (two-sided):	0.00	Kurtosis:	200.05

Warnings:

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

```
In [18]: from statsmodels.tsa.arima_model import ARIMA
```

```
In [19]: model = ARIMA(df["revenues_from_appointment"], order = (1, 0, 1))
         Actual_model = model.fit(dispatch = 0)
         print(Actual_model.summary())
```

ARMA Model Results

```

=====
=
Dep. Variable:    revenues_from_appointment    No. Observations:    406
1
Model:            ARMA(1, 1)    Log Likelihood        -27749.80
1
Method:          css-mle    S.D. of innovations    224.59
3
Date:            Sun, 05 Jun 2022    AIC                    55507.60
2
Time:            17:19:19    BIC                    55532.83
9
Sample:          0    HQIC                    55516.54
1

```

```

=====
=====
                                coef    std err          z      P>|z|      [0.
025      0.975]
-----
const                146.1682      5.953     24.553      0.000     134.
500      157.836
ar.l1.revenues_from_appointment    0.5647      0.036     15.672      0.000      0.
494      0.635
ma.l1.revenues_from_appointment    -0.2646      0.042     -6.309      0.000     -0.
347      -0.182

Roots
=====
                                Real      Imaginary      Modulus      Frequency
-----
AR.1                1.7708      +0.0000j      1.7708      0.0000
MA.1                3.7794      +0.0000j      3.7794      0.0000
-----

```

In [20]:

```

future_date = pd.date_range(start = '2021-12-30', end = '2022-06-28')
print(future_date)

```

```

DatetimeIndex(['2021-12-30', '2021-12-31', '2022-01-01', '2022-01-02',
               '2022-01-03', '2022-01-04', '2022-01-05', '2022-01-06',
               '2022-01-07', '2022-01-08',
               ...,
               '2022-06-19', '2022-06-20', '2022-06-21', '2022-06-22',
               '2022-06-23', '2022-06-24', '2022-06-25', '2022-06-26',
               '2022-06-27', '2022-06-28'],
              dtype='datetime64[ns]', length=181, freq='D')

```

In [25]:

```

pred= Actual_model.predict(start = len(df), end = len(df)+180, typ = 'levels').renam
print(pred)

```

```

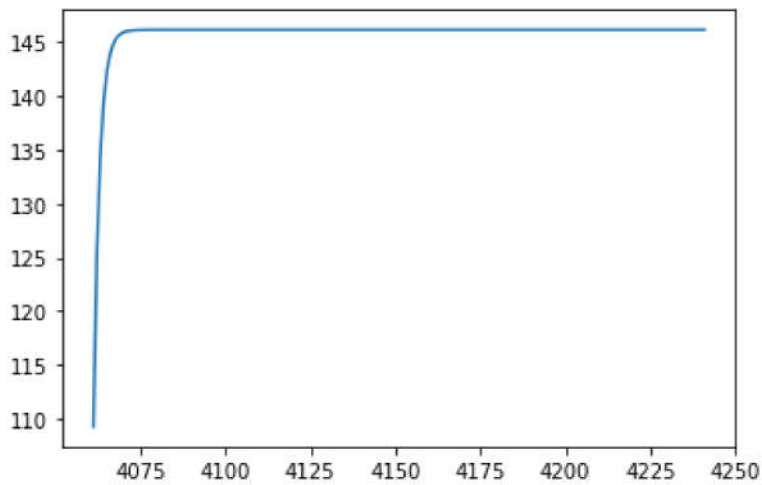
4061    109.238084
4062    125.312814
4063    134.390625
4064    139.517096
4065    142.412145
...
4237    146.168185
4238    146.168185
4239    146.168185
4240    146.168185

```


4241 146.168185
 Name: ARIMA Predictions, Length: 181, dtype: float64

In [29]: `pred.plot()`

Out[29]: <AxesSubplot:>



In [26]: `pred_df = pd.concat([df, pred])`
`pred_df`

Out[26]:

	0	appointment_date	appointment_duration_min	clinic_name	patient_id	practitione
0	NaN	2021-12-30	30.0	clinic_1	43859.0	7!
1	NaN	2021-12-30	60.0	clinic_1	53516.0	7!
2	NaN	2021-12-30	90.0	clinic_1	43323.0	7!
3	NaN	2021-12-30	30.0	clinic_2	70058.0	10!
4	NaN	2021-12-30	30.0	clinic_1	44924.0	7!
...
4237	146.168185	NaT	NaN	NaN	NaN	↑
4238	146.168185	NaT	NaN	NaN	NaN	↑
4239	146.168185	NaT	NaN	NaN	NaN	↑
4240	146.168185	NaT	NaN	NaN	NaN	↑
4241	146.168185	NaT	NaN	NaN	NaN	↑

4242 rows × 7 columns



In [27]: `pred_df.columns = ['Forecast', 'appointment_date', 'appointment_duration_min', 'clin`

In [28]: `pred_df`

Out[28]:

	Forecast	appointment_date	appointment_duration_min	clinic_name	patient_id	practitione
0	NaN	2021-12-30	30.0	clinic_1	43859.0	7!

	Forecast	appointment_date	appointment_duration_min	clinic_name	patient_id	practitione
1	NaN	2021-12-30	60.0	clinic_1	53516.0	7!
2	NaN	2021-12-30	90.0	clinic_1	43323.0	7!
3	NaN	2021-12-30	30.0	clinic_2	70058.0	10!
4	NaN	2021-12-30	30.0	clinic_1	44924.0	7!
...
4237	146.168185	NaT	NaN	NaN	NaN	↑
4238	146.168185	NaT	NaN	NaN	NaN	↑
4239	146.168185	NaT	NaN	NaN	NaN	↑
4240	146.168185	NaT	NaN	NaN	NaN	↑
4241	146.168185	NaT	NaN	NaN	NaN	↑

4242 rows × 7 columns

