Property Price Analysis and Prediction

October 4, 2020

[1]:
```
# For library input
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import pandas.util.testing as tm
import statsmodels.api as sm
import numpy as np
```

[2]:
```
#Read CSV
data = pd.read_csv('full_data_20200920.csv')
```

[3]:
```
# Check if there is missed value
print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 190794 entries, 0 to 190793
Data columns (total 21 columns):
 #   Column                        Non-Null Count   Dtype
---  ------                        --------------   -----
 0   Unnamed: 0                    190794 non-null  int64
 1   Project Name                  190794 non-null  object
 2   Address                       190794 non-null  object
 3   No. of Units                  190794 non-null  int64
 4   Area (sqm)                    190794 non-null  int64
 5   Type of Area                  190794 non-null  object
 6   Transacted Price ($)          190794 non-null  int64
 7   Nett Price($)                 190794 non-null  object
 8   Unit Price ($ psm)            190794 non-null  int64
 9   Unit Price ($ psf)            190794 non-null  int64
 10  Sale Date                     190794 non-null  object
 11  Property Type                 190794 non-null  object
 12  Tenure                        190794 non-null  object
 13  Completion Date               190794 non-null  object
 14  Type of Sale                  190794 non-null  object
 15  Purchaser Address Indicator   190794 non-null  object
 16  Postal District               190794 non-null  int64
```

```
 17  Postal Sector               190794 non-null  int64
 18  Postal Code                 190794 non-null  int64
 19  Planning Region             190794 non-null  object
 20  Planning Area               190794 non-null  object
dtypes: int64(9), object(12)
memory usage: 30.6+ MB
None
```

0.0.1 Exploratory Data Analysis on Private Property

[4]:
```
# Show number of transcations in each distrcit
Trans_by_district = pd.DataFrame(data['Postal District'].value_counts().
 ↪sort_values())
Trans_by_district
```

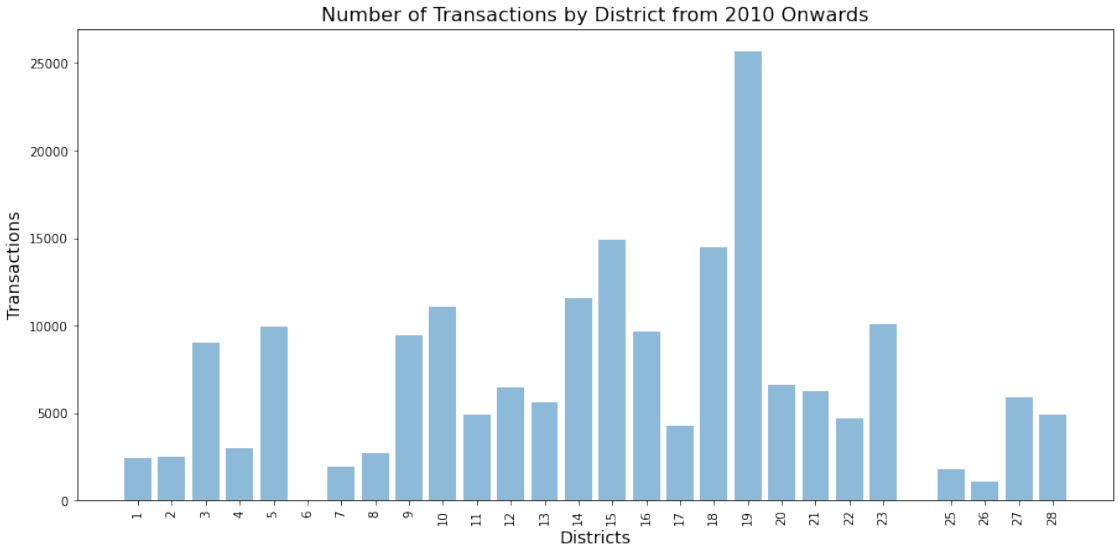
[4]:

	Postal District
6	23
26	1054
25	1795
7	1917
1	2432
2	2503
8	2703
4	3003
17	4281
22	4718
11	4861
28	4925
13	5634
27	5865
21	6233
12	6451
20	6593
3	8975
9	9443
16	9631
5	9964
23	10086
10	11088
14	11592
18	14434
15	14893
19	25697

[5]:
```
#Histogram to plot number of transcations in each distrcit
plt.figure(figsize=(15,7))
```

```
plt.bar(list(Trans_by_district.index), Trans_by_district['Postal District'],␣
 ↪align='center', alpha=0.5)
plt.xlabel('Districts', fontsize=14)
plt.xticks(list(Trans_by_district.index),rotation=90)
plt.ylabel('Transactions', fontsize=14)
plt.title('Number of Transactions by District from 2010 Onwards', fontsize=16)

plt.show()
```



0.0.2 Box plot

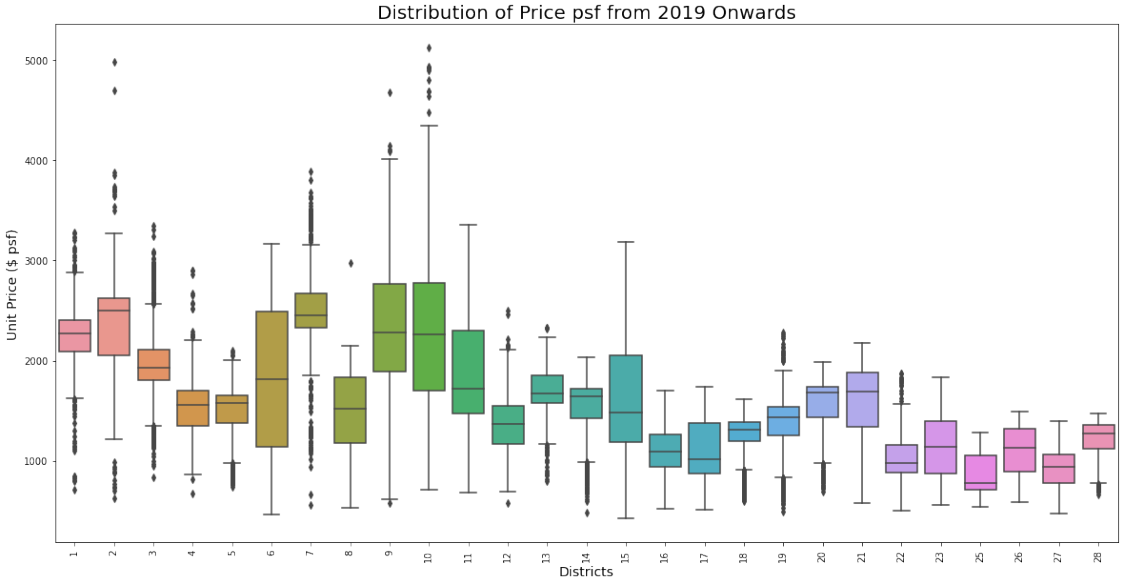
[6]:
```
#Amend the column of Sales Date into a datetime object (month_year)
data['Month_Year'] = pd.to_datetime(data['Sale Date']).dt.to_period('M')

#Drop the first column (Unnamed: 0)
data = data.drop('Unnamed: 0', axis=1)
```

[7]:
```
data_temp = data.loc[data['Month_Year'] > '2018-12']
data_temp['Month_Year']

plt.figure(figsize=(20,10))
sns.boxplot( x = data_temp['Postal District'], y = data_temp['Unit Price ($␣
 ↪psf)'], data=data)
plt.xlabel('Districts', fontsize=14)
plt.xticks(rotation=90)
plt.ylabel('Unit Price ($ psf)', fontsize=14)
plt.title('Distribution of Price psf from 2019 Onwards', fontsize=20)
```

[7]: Text(0.5, 1.0, 'Distribution of Price psf from 2019 Onwards')



We can see that District location could be a critical factor which determines the property price.

Data Pre-processing and ARIMA

[8]:
```
#Average_monthly for each district
Average_monthly = data.groupby(['Postal District', 'Month_Year'],␣
 ↪as_index=False)['Unit Price ($ psf)'].mean()
Average_monthly_df = pd.DataFrame(Average_monthly)
```

[9]:
```
Average_monthly_df.info()
Average_monthly_df['Postal District'].value_counts()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3154 entries, 0 to 3153
Data columns (total 3 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Postal District     3154 non-null   int64
 1   Month_Year          3154 non-null   period[M]
 2   Unit Price ($ psf)  3154 non-null   float64
dtypes: float64(1), int64(1), period[M](1)
memory usage: 98.6 KB
```

[9]:
```
27    121
3     121
```

```
4     121
8     121
10    121
12    121
14    121
16    121
18    121
20    121
22    121
28    121
1     121
5     121
9     121
11    121
13    121
15    121
17    121
19    121
21    121
23    121
2     121
25    120
26    120
7     116
6      15
Name: Postal District, dtype: int64
```

Our dataset period is from 2010-Sep to 2020-Sep, so there should be 121 datapoints.

District 6 has too little datapoint so we drop the analysis on district 6

For district 25,26, and 7, they have 116 and 120 points, therefore we will assume the missing monthly price is the same as previous monthly price

[10]:
```
# Drop District 6
Average_monthly_df_drop = Average_monthly_df[Average_monthly_df['Postal␣
 ↪District'] != 6]

#Input missing datapoint
# make it as a pivot table
pivot = Average_monthly_df_drop.pivot(index='Month_Year', columns = 'Postal␣
 ↪District', values = 'Unit Price ($ psf)')
#Fill in missing values
pivot.fillna(method='ffill', inplace=True)
#Back to an array
Average_monthly_df_filled = pivot.stack().reset_index()
Average_monthly_df_filled.sort_values(['Postal District', 'Month_Year'])
```

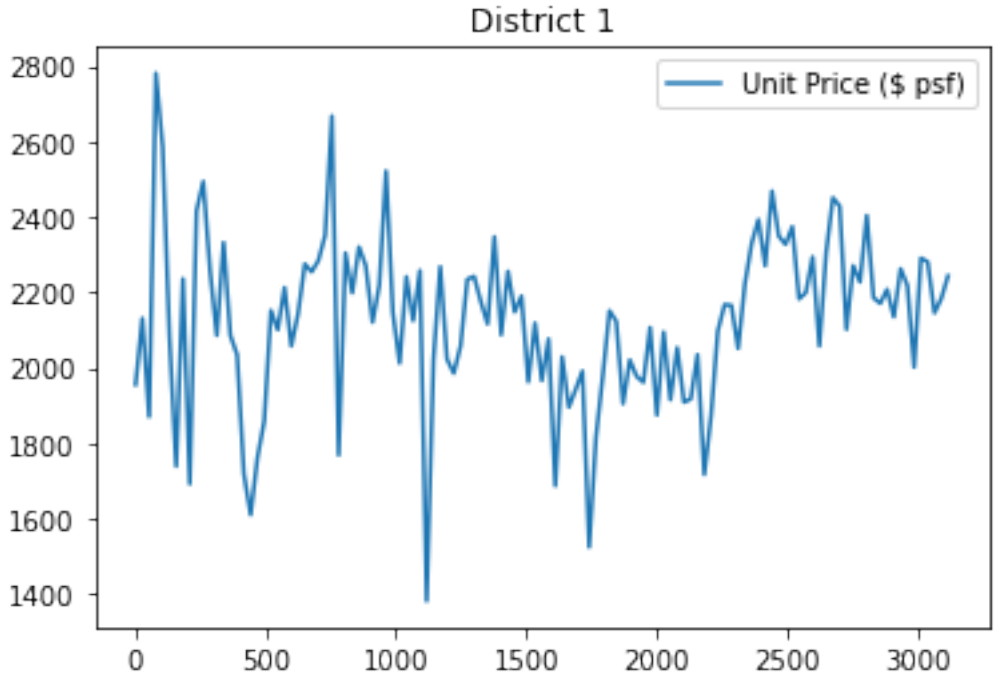
```
#To check
Average_monthly_df_filled['Postal District'].value_counts()
```

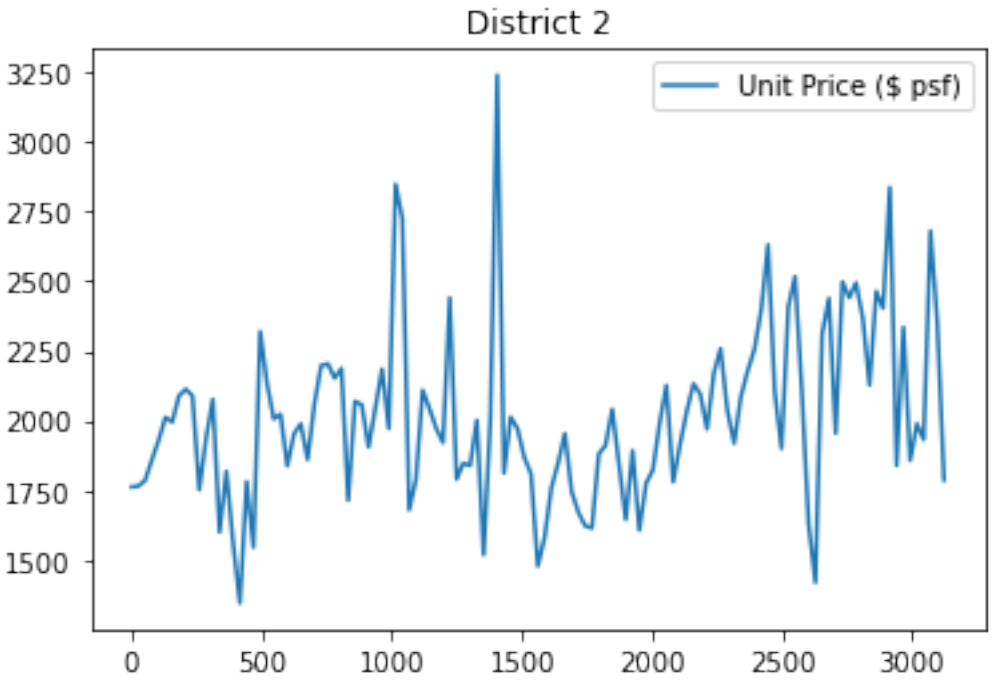
[10]:
```
27    121
25    121
4     121
8     121
10    121
12    121
14    121
16    121
18    121
20    121
22    121
26    121
28    121
1     121
3     121
5     121
7     121
9     121
11    121
13    121
15    121
17    121
19    121
21    121
23    121
2     121
Name: Postal District, dtype: int64
```

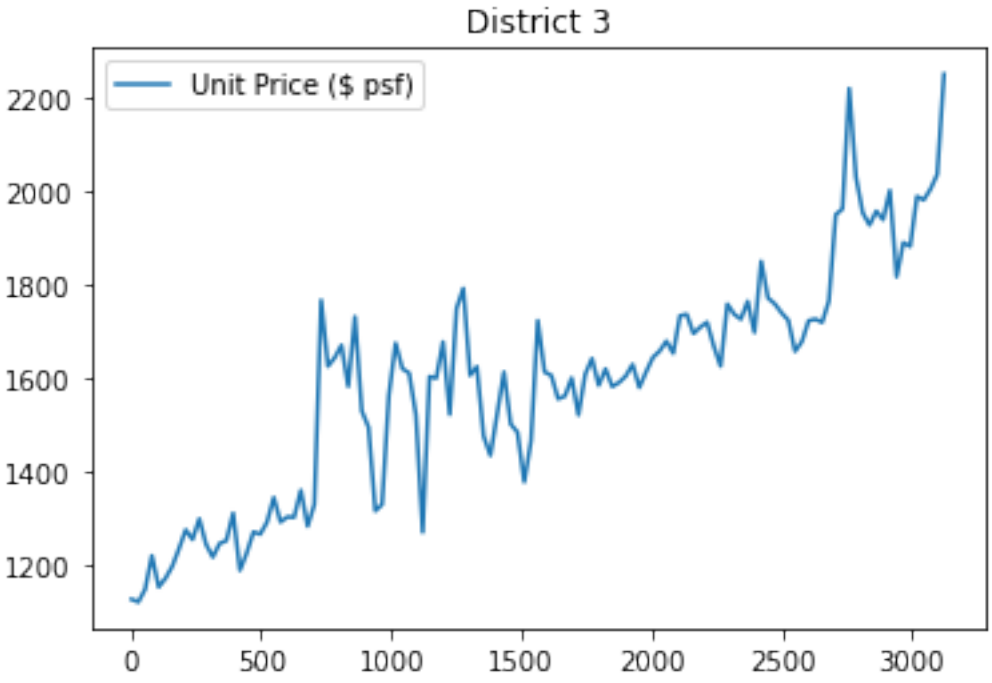
[11]:
```
Average_monthly_df_filled.rename(columns={0: "Unit Price ($ psf)"},␣
↪inplace=True)
Average_monthly_df_filled.info()
```

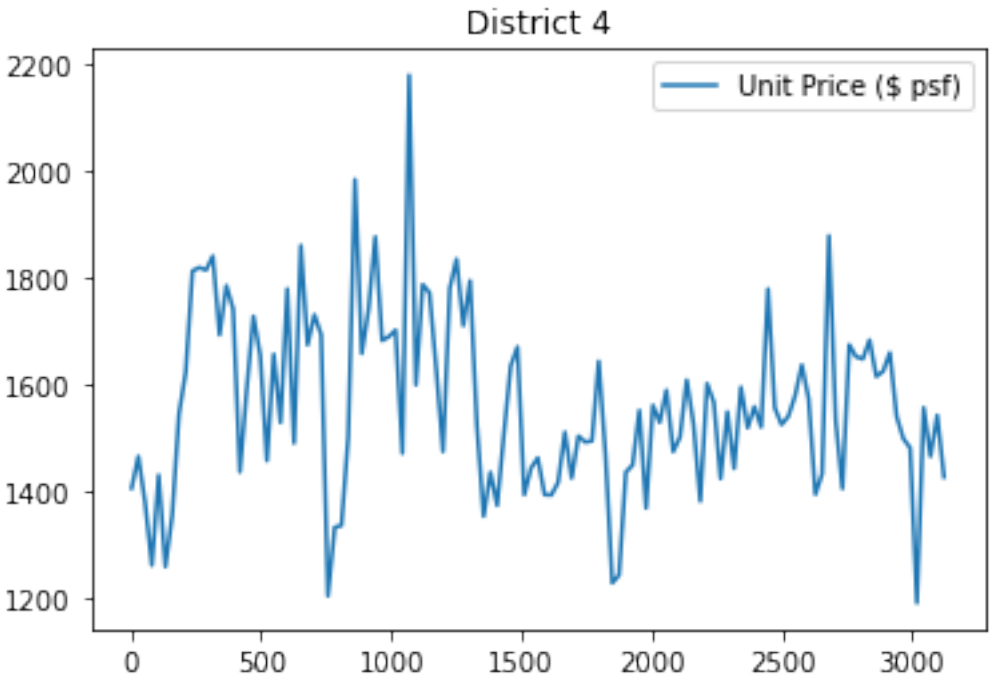
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3146 entries, 0 to 3145
Data columns (total 3 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Month_Year          3146 non-null   period[M]
 1   Postal District     3146 non-null   int64
 2   Unit Price ($ psf)  3146 non-null   float64
dtypes: float64(1), int64(1), period[M](1)
memory usage: 73.9 KB
```

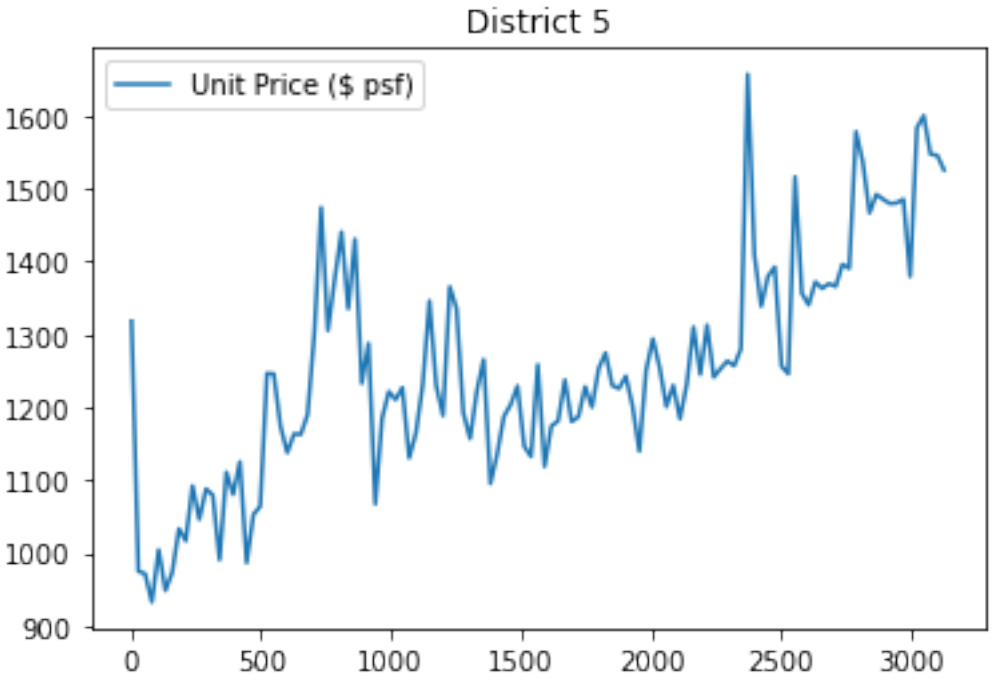
[12]:
```
# Plot monthly average unit price for each district for past ten years
for i in [x for x in range(1,29) if x != 6 and x !=24]:
    Average_monthly_df_filled.loc[(Average_monthly_df_filled['Postal District']␣
 ↪== i)].plot(y='Unit Price ($ psf)')
    plt.title('District ' + str(i))
    plt.show()
```

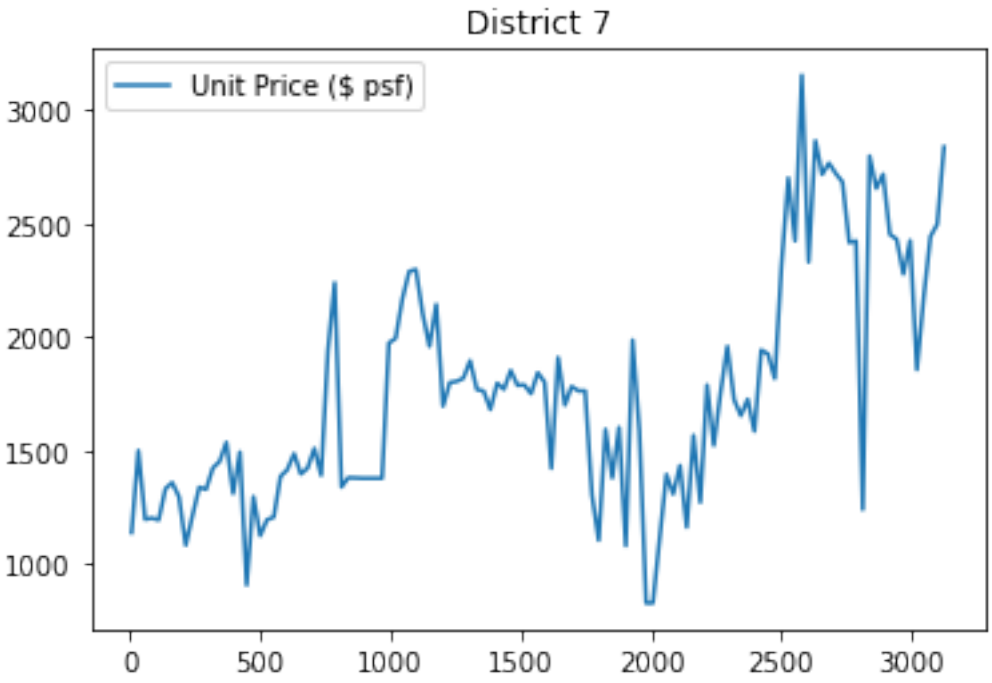


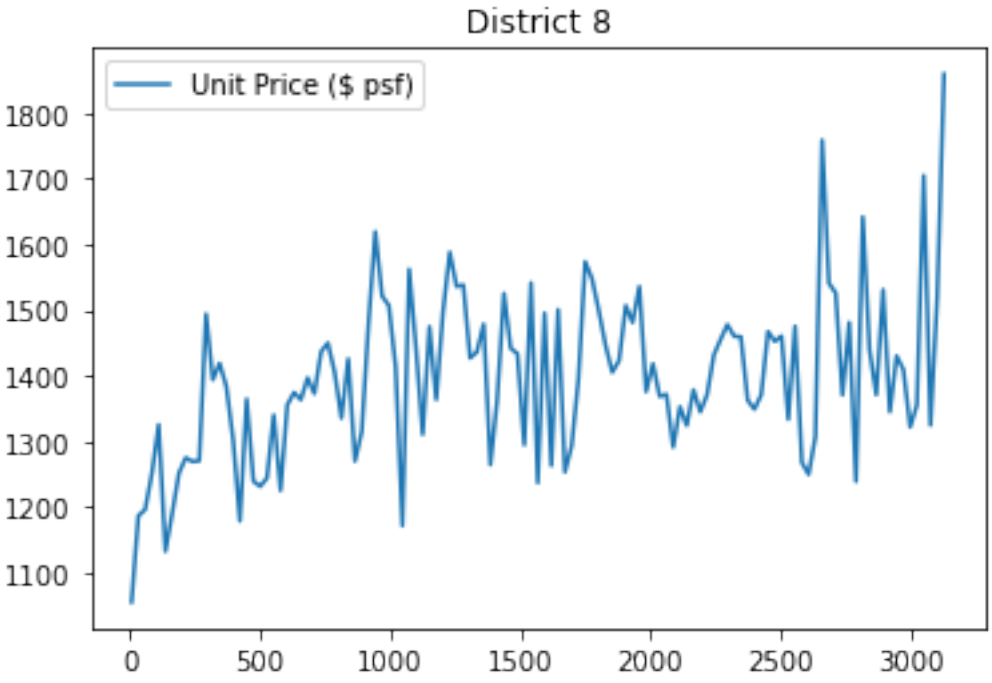

District 2
Unit Price ($ psf)
3250
3000
2750
2500
2250
2000
1750
1500
0
500
1000
1500
2000
2500
3000

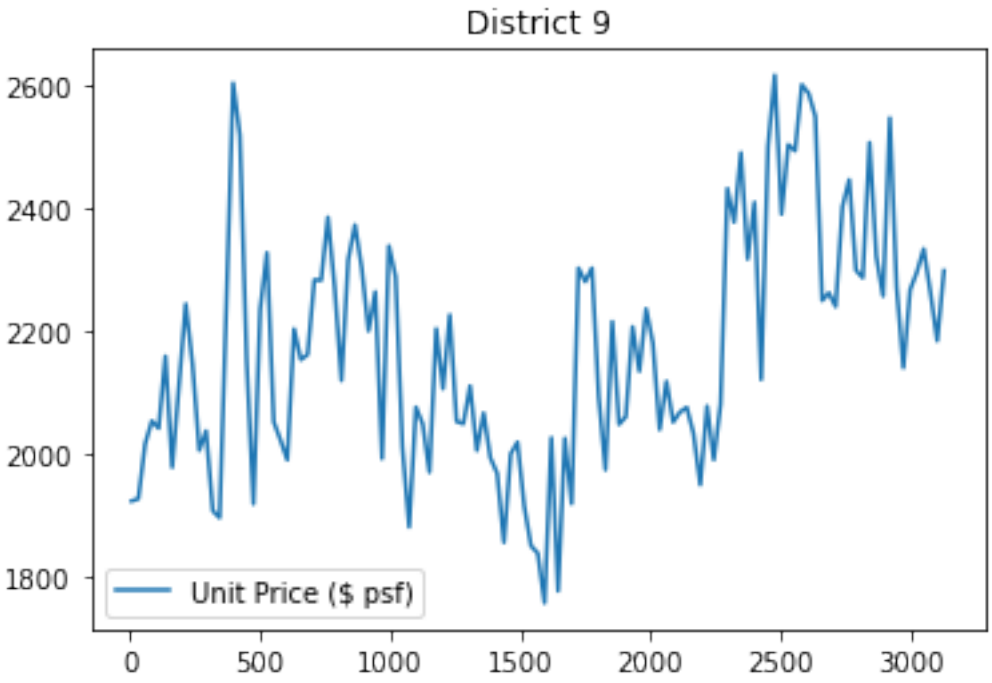

District 3
Unit Price ($ psf)
2200
2000
1800
1600
1400
1200
0
500
1000
1500
2000
2500
3000

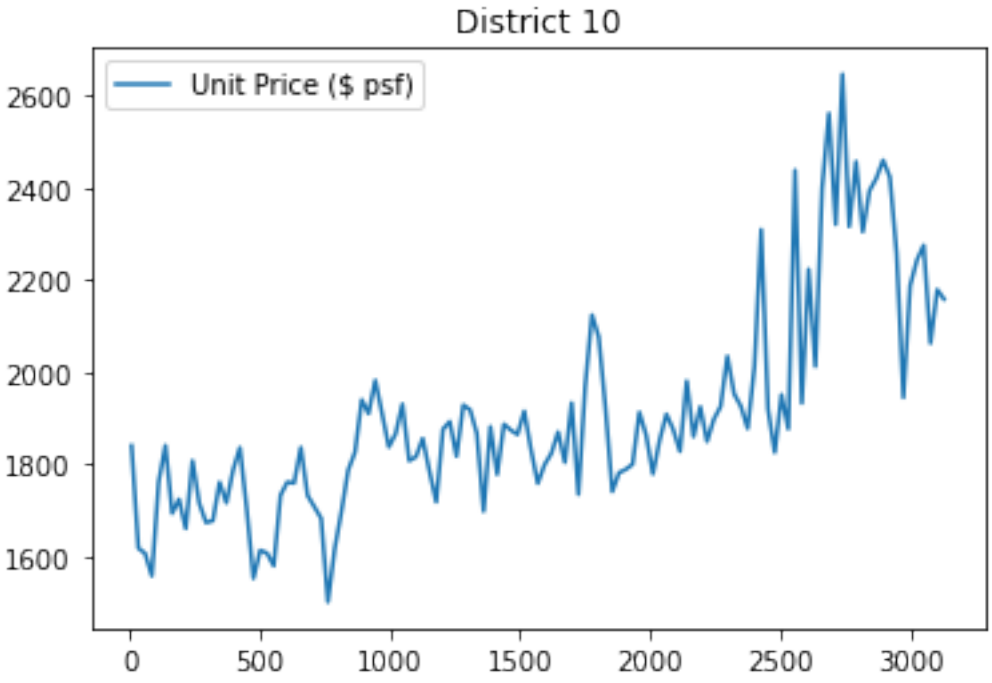

District 4
Unit Price ($ psf)
2200
2000
1800
1600
1400
1200
0
500
1000
1500
2000
2500
3000

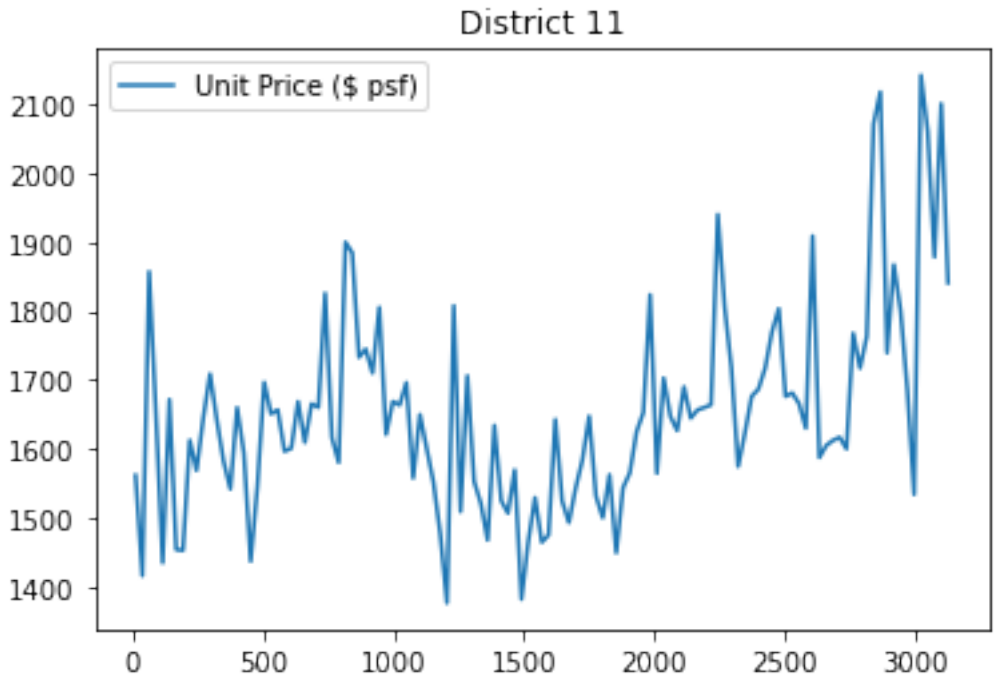

District 5
Unit Price ($ psf)
1600
1500
1400
1300
1200
1100
1000
900
0
500
1000
1500
2000
2500
3000

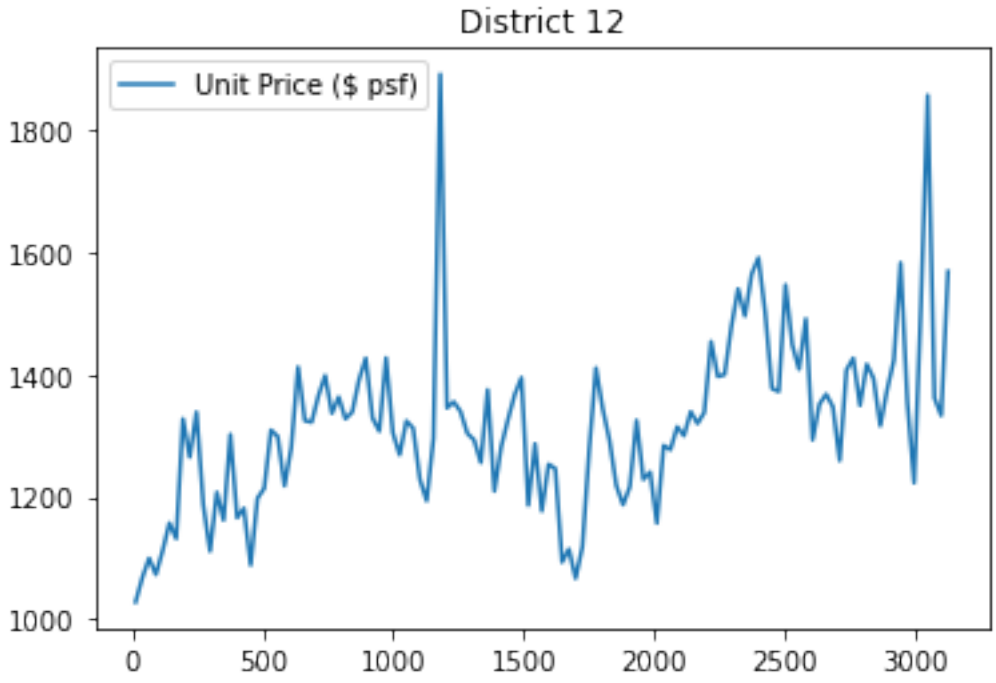

District 7
Unit Price ($ psf)
3000
2500
2000
1500
1000
0
500
1000
1500
2000
2500
3000

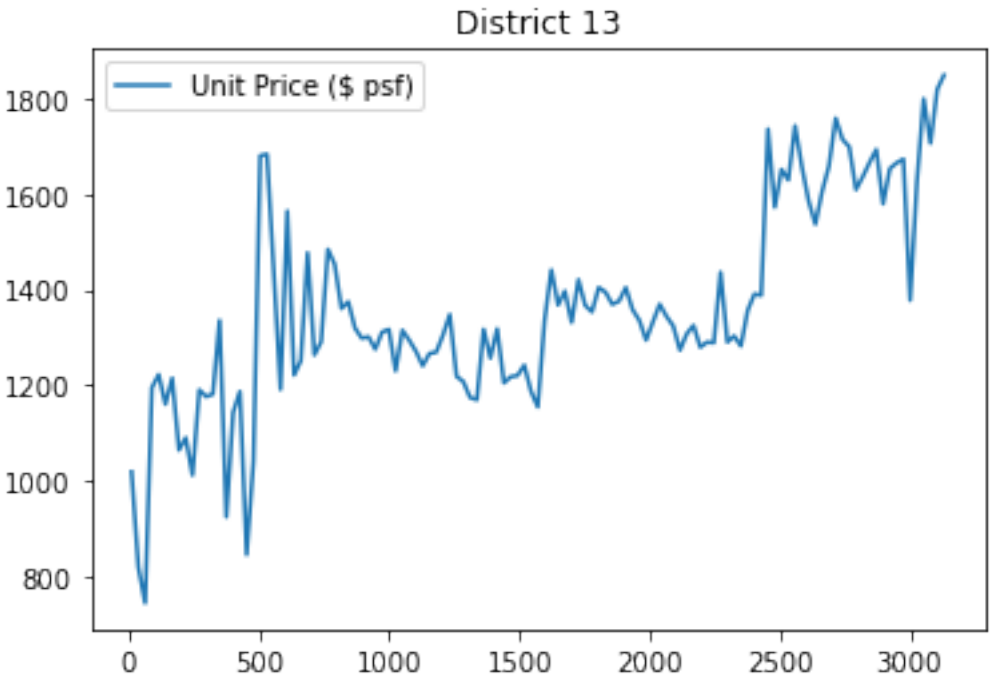

District 8
Unit Price ($ psf)
1800
1700
1600
1500
1400
1300
1200
1100
0
500
1000
1500
2000
2500
3000

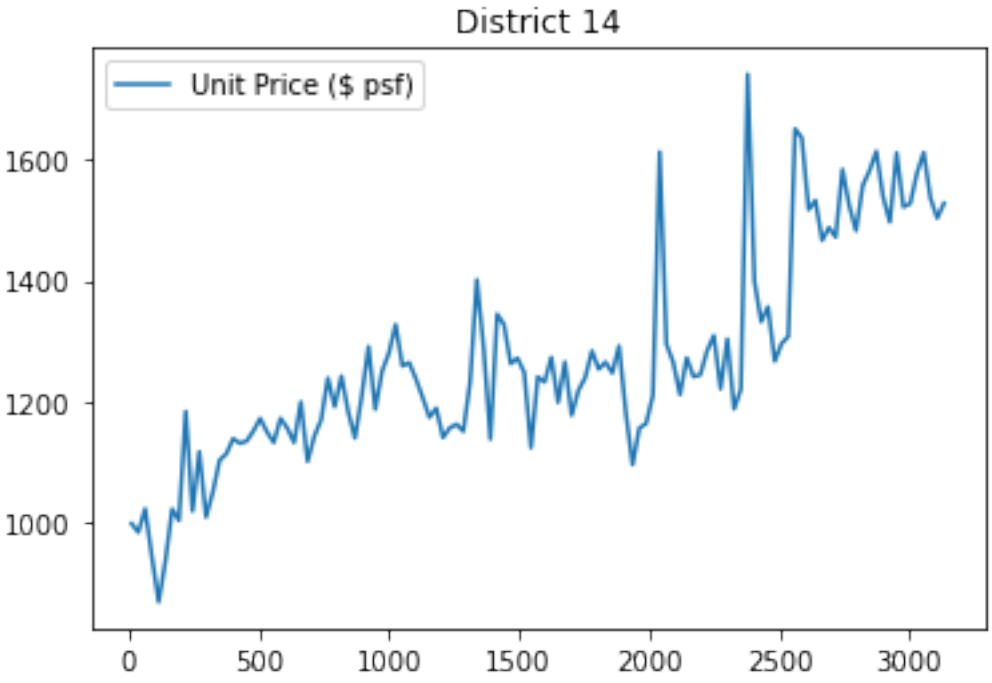

District 9
Unit Price ($ psf)
2600
2400
2200
2000
1800
0
500
1000
1500
2000
2500
3000

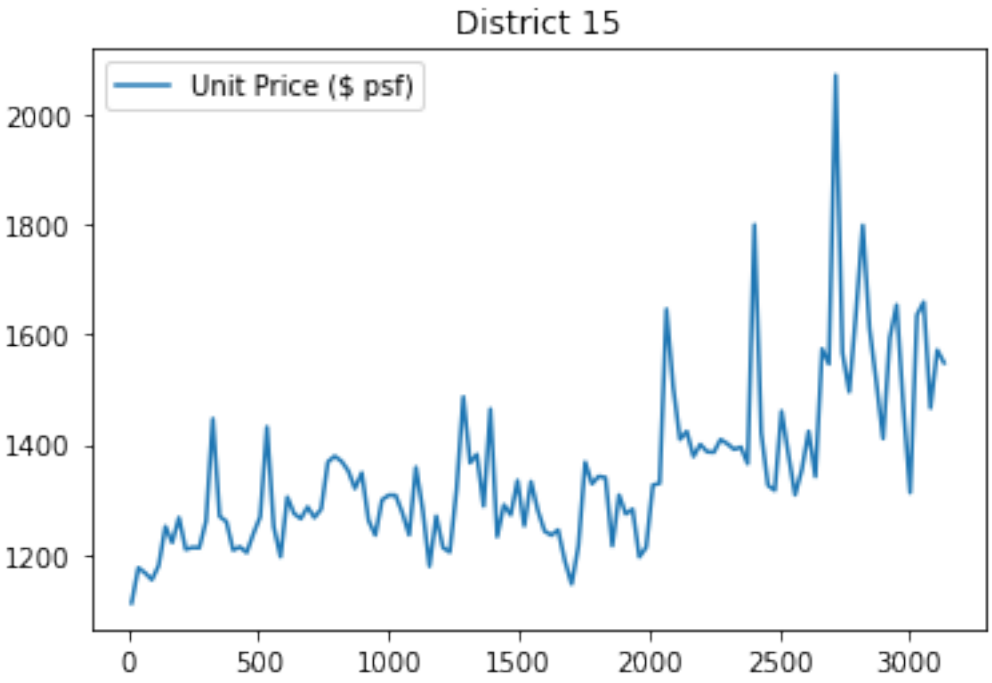

District 10
Unit Price ($ psf)
2600
2400
2200
2000
1800
1600
0
500
1000
1500
2000
2500
3000

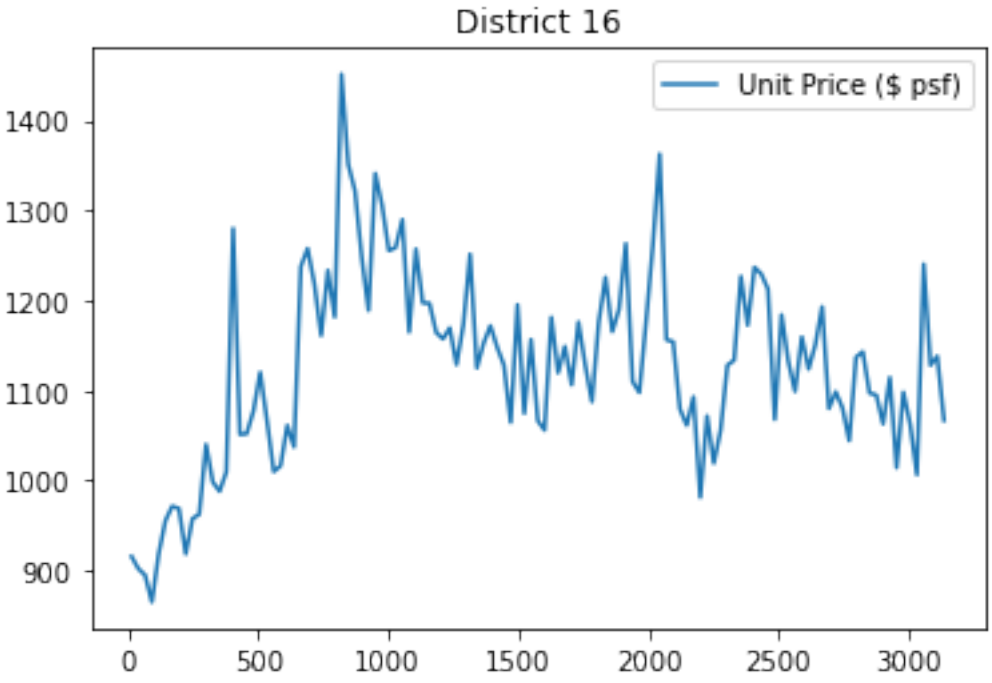

District 11
Unit Price ($ psf)
2100
2000
1900
1800
1700
1600
1500
1400
0
500
1000
1500
2000
2500
3000

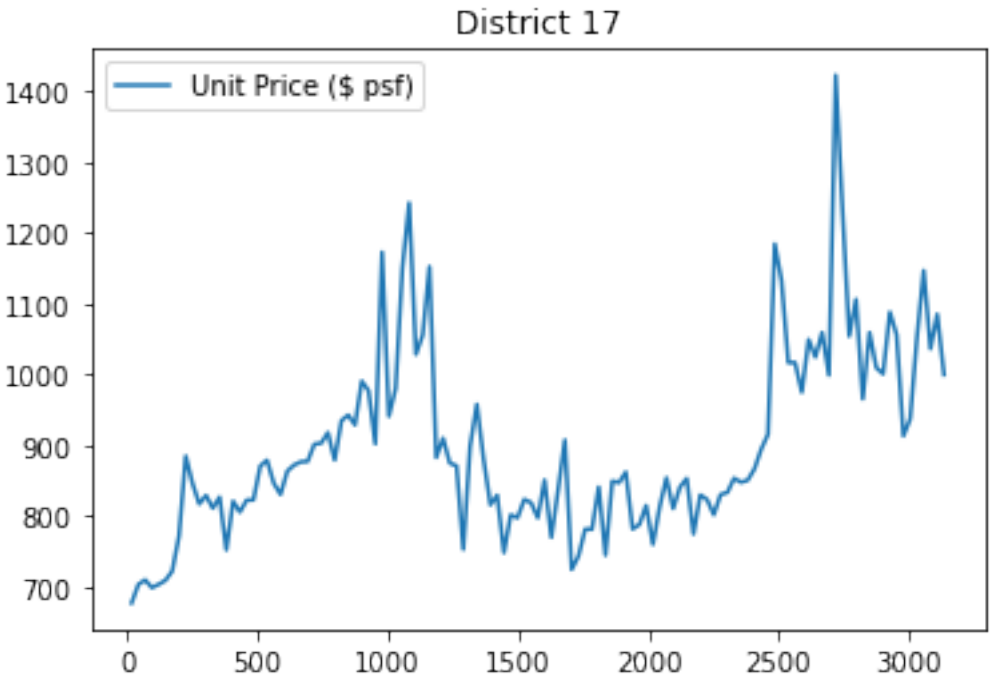

District 12
Unit Price ($ psf)
1800
1600
1400
1200
1000
0
500
1000
1500
2000
2500
3000

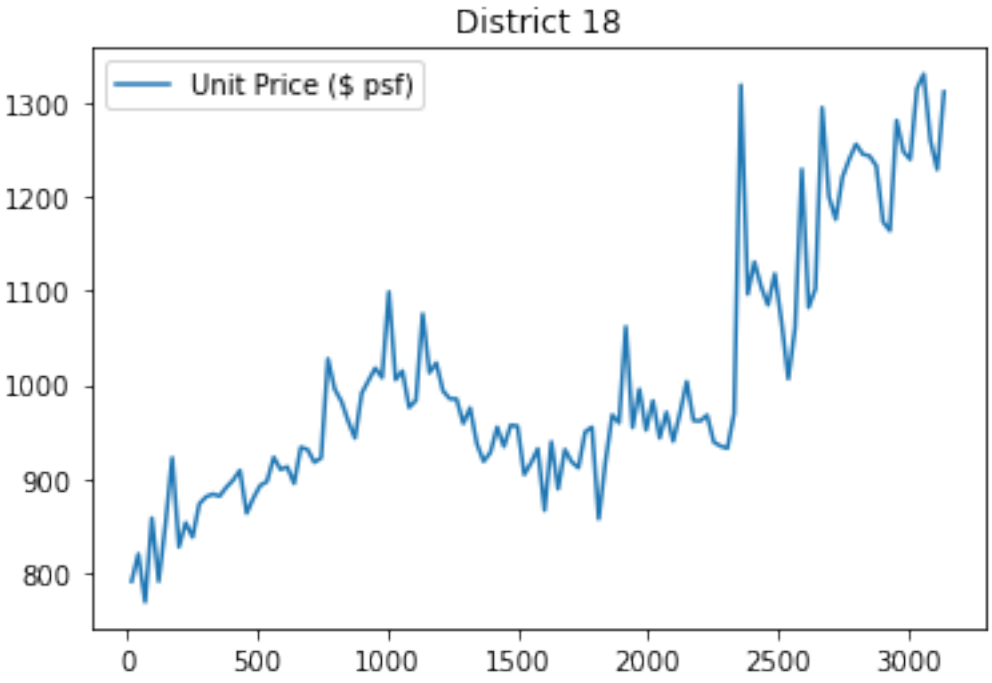

District 13
Unit Price ($ psf)
1800
1600
1400
1200
1000
800
0
500
1000
1500
2000
2500
3000

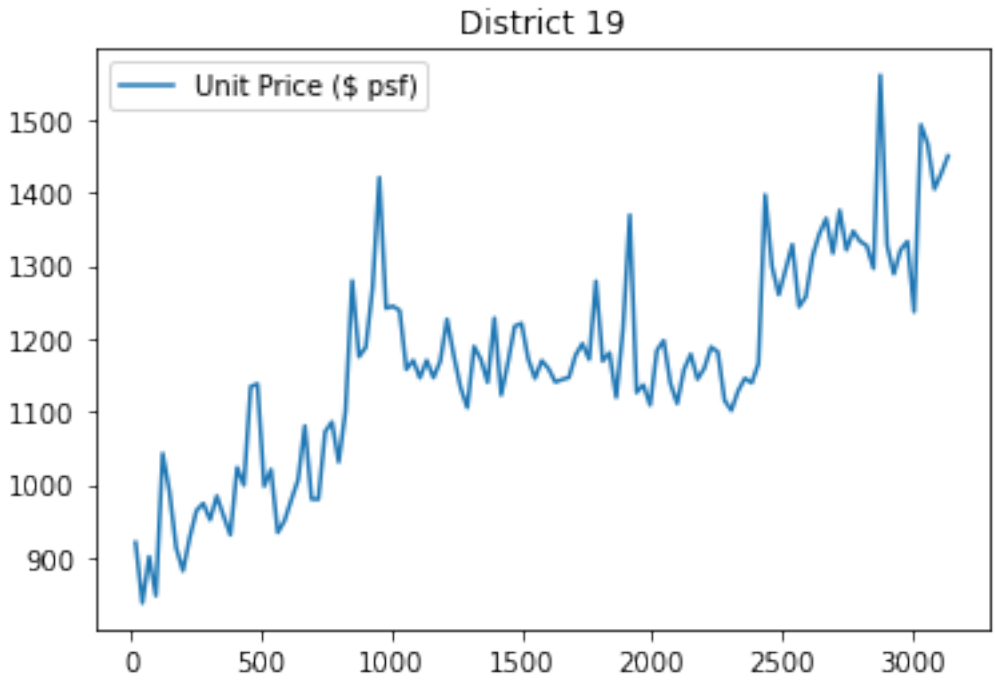

District 14
Unit Price ($ psf)
1600
1400
1200
1000
0
500
1000
1500
2000
2500
3000

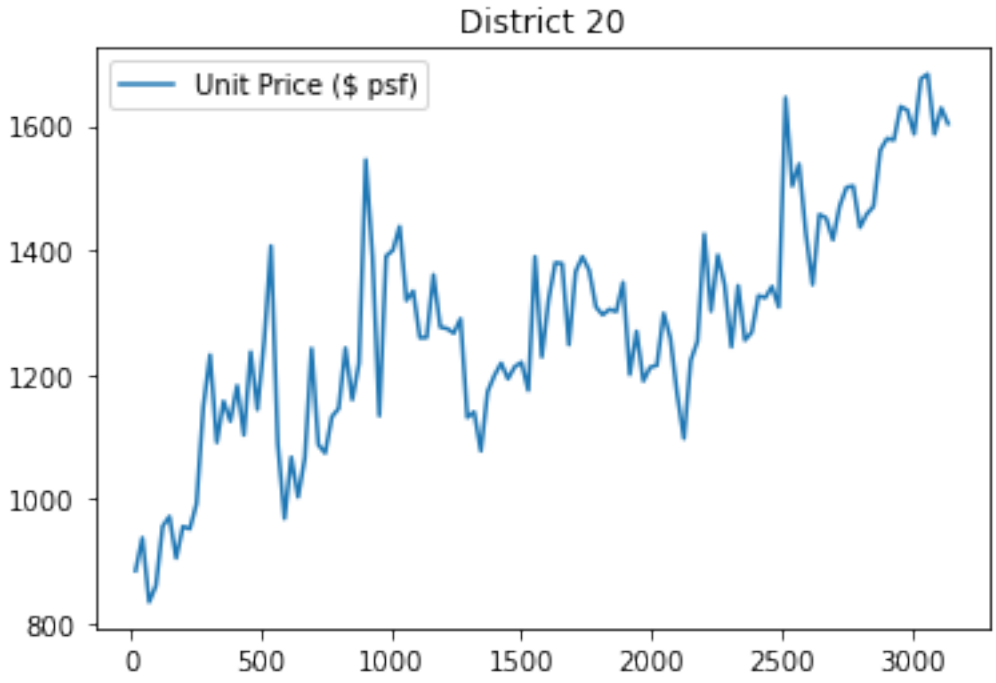

District 15
Unit Price ($ psf)
2000
1800
1600
1400
1200
0
500
1000
1500
2000
2500
3000

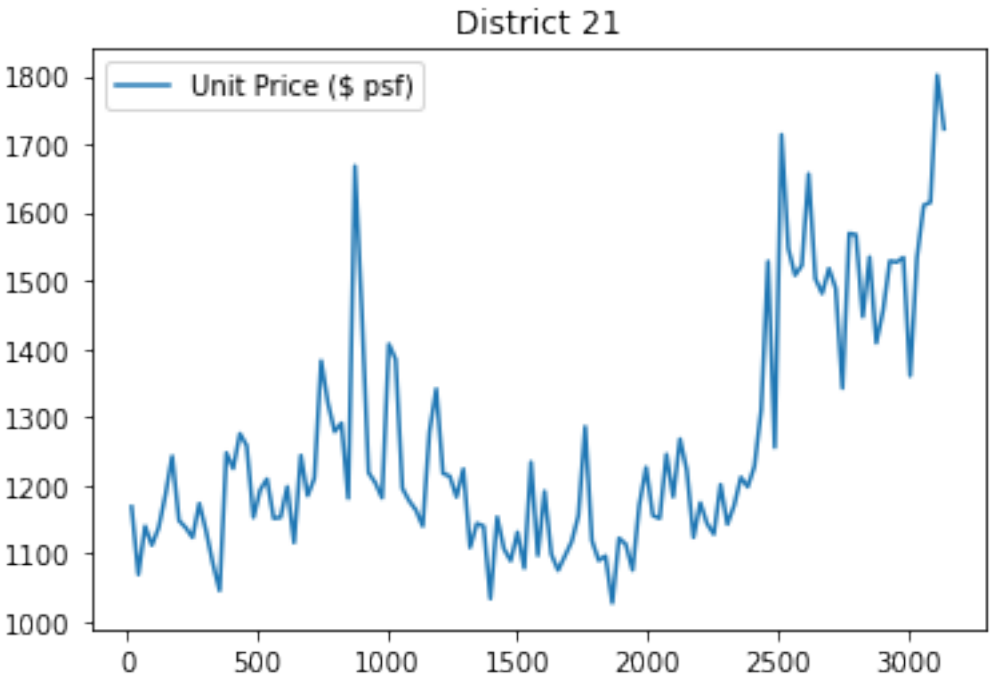

District 16
Unit Price ($ psf)
1400
1300
1200
1100
1000
900
0
500
1000
1500
2000
2500
3000

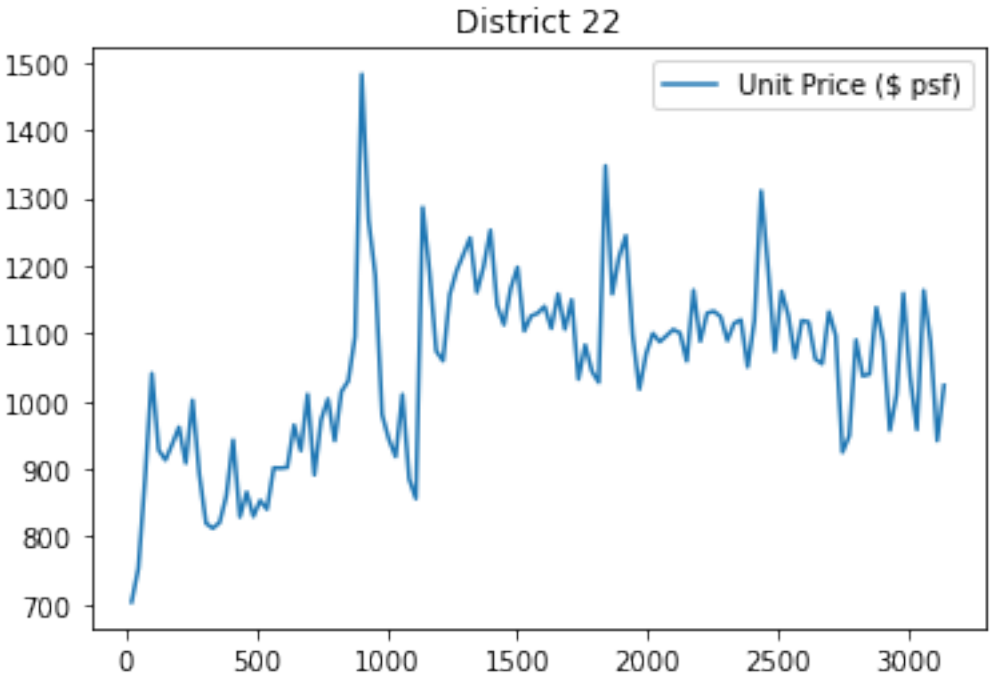

District 17
Unit Price ($ psf)
1400
1300
1200
1100
1000
900
800
700
0
500
1000
1500
2000
2500
3000

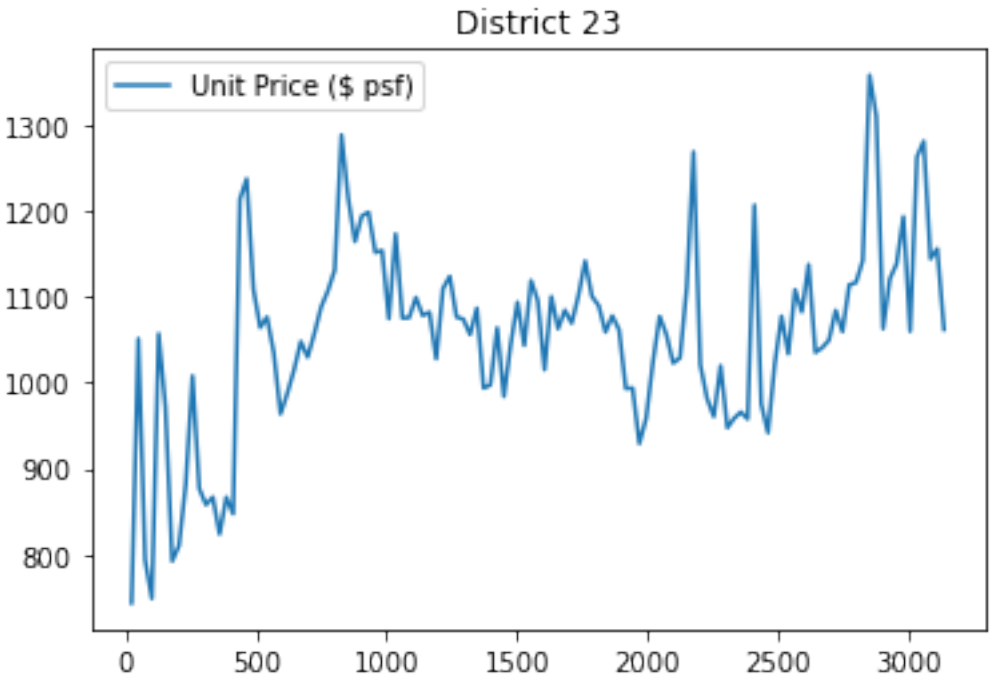

District 18
Unit Price ($ psf)
1300
1200
1100
1000
900
800
0
500
1000
1500
2000
2500
3000

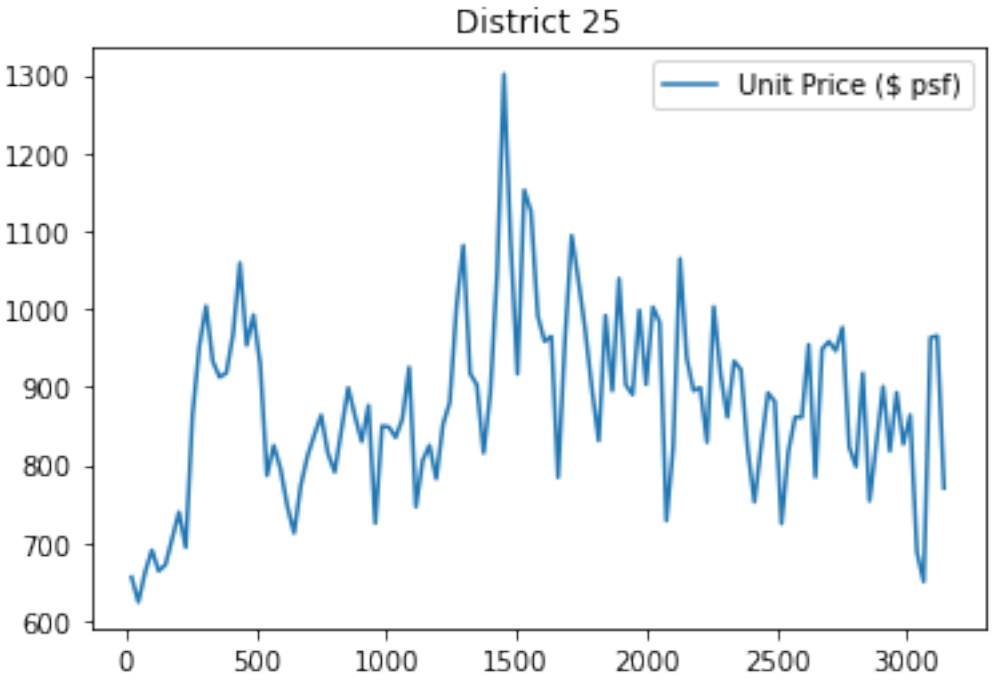

District 19
Unit Price ($ psf)
1500
1400
1300
1200
1100
1000
900
0
500
1000
1500
2000
2500
3000

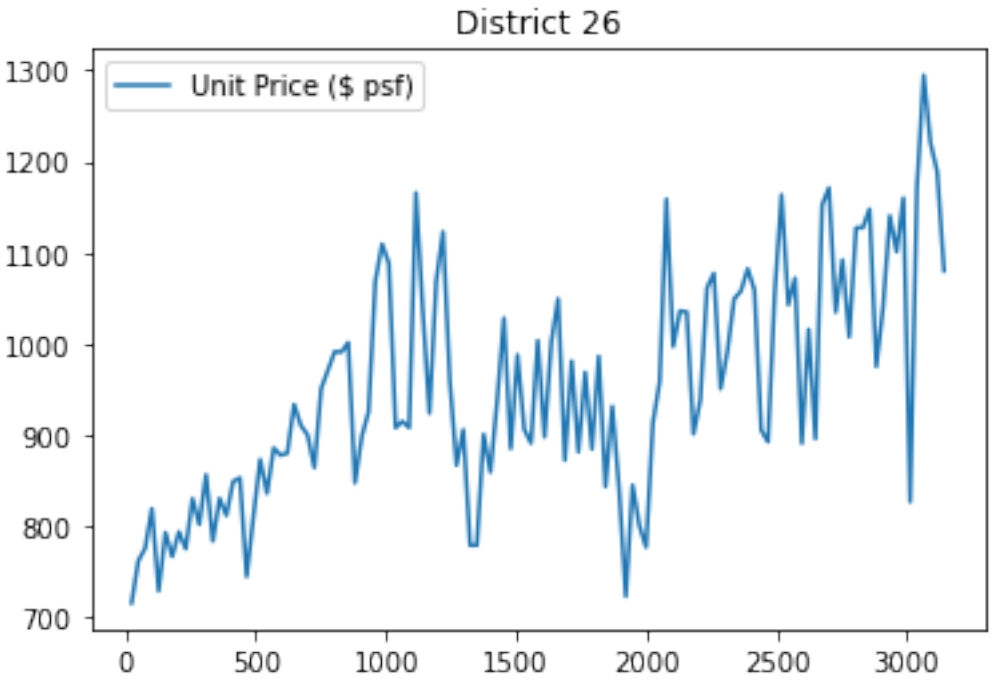

District 20
Unit Price ($ psf)
1600
1400
1200
1000
800
0
500
1000
1500
2000
2500
3000

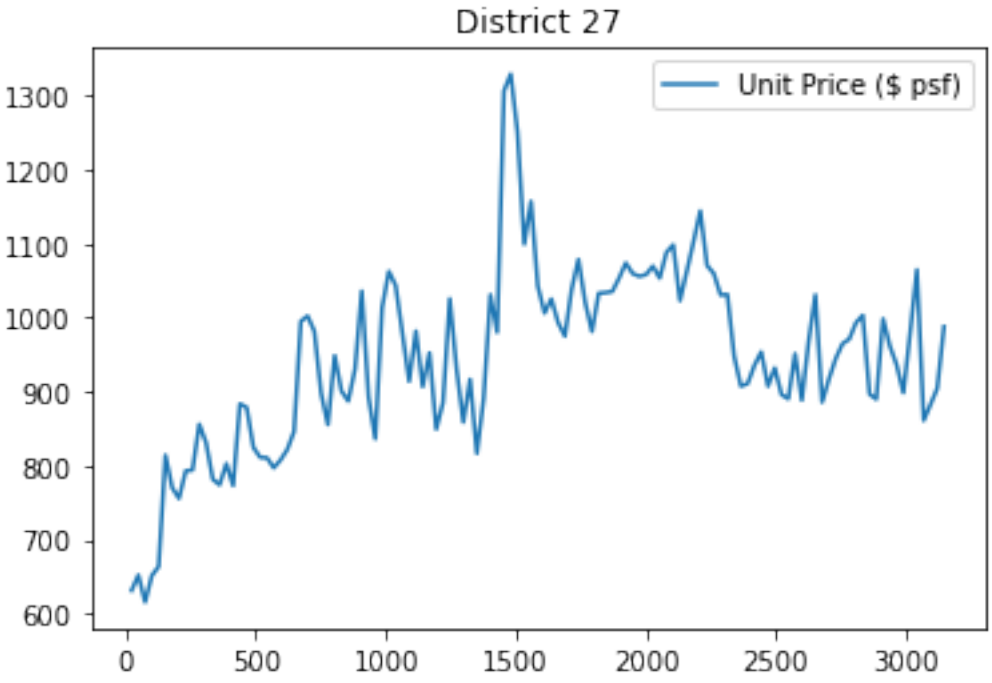

District 21
Unit Price ($ psf)
1800
1700
1600
1500
1400
1300
1200
1100
1000
0
500
1000
1500
2000
2500
3000

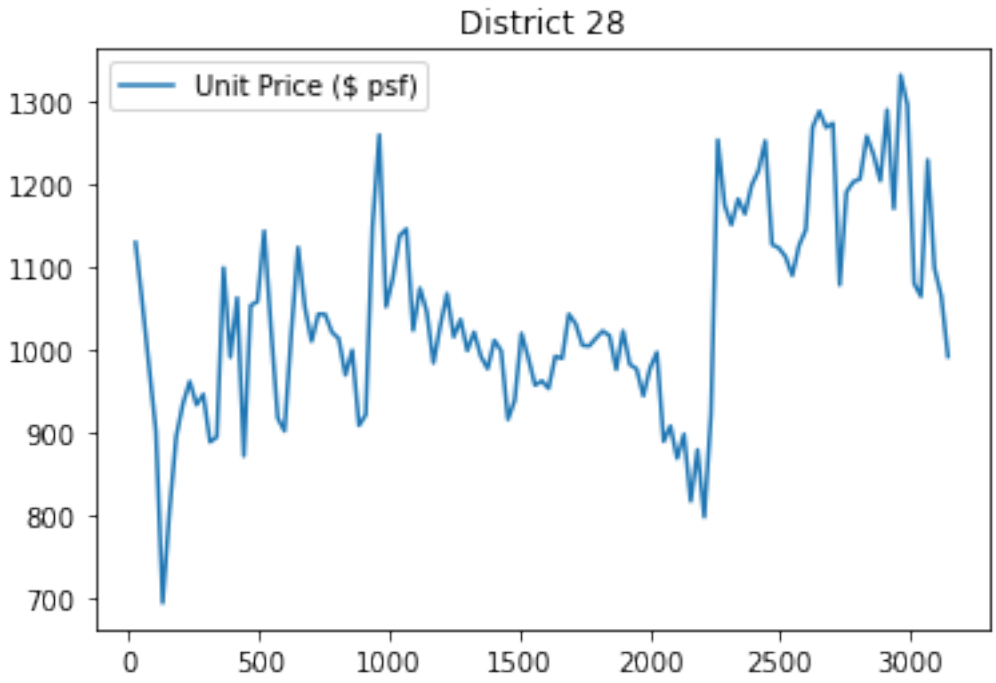

District 22
Unit Price ($ psf)
1500
1400
1300
1200
1100
1000
900
800
700
0
500
1000
1500
2000
2500
3000


District 23
Unit Price ($ psf)
1300
1200
1100
1000
900
800
0
500
1000
1500
2000
2500
3000


District 25
Unit Price ($ psf)
1300
1200
1100
1000
900
800
700
600
0
500
1000
1500
2000
2500
3000


District 26
Unit Price ($ psf)
1300
1200
1100
1000
900
800
700
0
500
1000
1500
2000
2500
3000


District 27
Unit Price ($ psf)
1300
1200
1100
1000
900
800
700
600
0
500
1000
1500
2000
2500
3000



Take log on Unit Price ($ psf)

[13]:
```
#Add a column to store log_psf
Average_monthly_df_filled.loc[ : , 'log_psf'] = np.
 ↪log(Average_monthly_df_filled['Unit Price ($ psf)'])
Average_monthly_df_filled.head()
```

[13]:

	Month_Year	Postal District	Unit Price ($ psf)	log_psf
0	2010-09	1	1955.500000	7.578401
1	2010-09	2	1762.523810	7.474502
2	2010-09	3	1129.108696	7.029184
3	2010-09	4	1404.771429	7.247630
4	2010-09	5	1318.385027	7.184163

Utilize ARIMA model and forecast for each district

[14]:
```
import pmdarima as pm
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import matplotlib.pyplot as plt
from math import e
```

[15]:
```
import numpy as np
def mape(y_true, y_pred):
    return np.mean(np.abs((y_pred - y_true) / y_true)) * 100
```

[16]:
```
# Print model summaries, plots, and prediction results for all districts except␣
 ↪Distrct 6 & 24
districts = []
MAPE = []
percentage_change = []
log_Predicted = []
log_Currentprice = []
for i in [x for x in range(1,29) if x != 6 and x !=24]:
        print('============================Results of District',␣
 ↪i,'============================')
        # train the model
        train_size = round(0.9 *␣
 ↪len(Average_monthly_df_filled[(Average_monthly_df_filled['Postal District']␣
 ↪== i)]))
        test_size = round(0.1 *␣
 ↪len(Average_monthly_df_filled[(Average_monthly_df_filled['Postal District']␣
 ↪== i)]))
        train = Average_monthly_df_filled[(Average_monthly_df_filled['Postal␣
 ↪District'] == i)][:train_size]
        test = Average_monthly_df_filled[(Average_monthly_df_filled['Postal␣
 ↪District'] == i)][train_size:]

        if len(train) >= 50:
            model = pm.auto_arima(train[(train['Postal District'] ==␣
 ↪i)]['log_psf'], suppress_warnings=True)
            #print(model.summary())
            temp = model.order
            p, d, q = temp[0], temp[1], temp[2]
            #print(model.plot_diagnostics())

            #model.resid()
            #plot_acf(model.resid(),lags=20)
            #plt.show()
            #plot_pacf(model.resid(),lags=20)
            #plt.show()

            pred = model.predict(test_size)
            Test = pd.DataFrame()
            Test['pred'] = list(pred)
            MAPE_ij = mape(list(test['log_psf']), Test['pred'])

            model_p = pm.
 ↪auto_arima(Average_monthly_df_filled[(Average_monthly_df_filled['Postal␣
 ↪District'] == i)]['log_psf'],suppress_warnings=True)
```

```
            #model_p = ARIMA(Average_monthly_df[(Average_monthly_df['Postal␣
↪District'] == i)&(Average_monthly_df['Tenure_dummy(>900)'] ==␣
↪j)]['log_psf'],order=(p, d, q))
            pred_p = model_p.predict(36)
            print(model_p.summary())
            print()
            # integrate the results
            districts.append(i)
            MAPE.append(MAPE_ij)
            percentage_change.append(float(100*(pred_p[-1:] -␣
↪float(Average_monthly_df_filled[(Average_monthly_df_filled['Postal␣
↪District'] == i)]['log_psf'][-1:]))))
            log_Predicted.append(float(pred_p[-1:]))
            log_Currentprice.
↪append(float(Average_monthly_df_filled[(Average_monthly_df_filled['Postal␣
↪District'] == i)]['log_psf'][-1:]))

        else:
            pass
```

```
==============================Results of District 1 ==============================
                               SARIMAX Results
==========================================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:                 SARIMAX(1, 0, 1)   Log Likelihood                 105.976
Date:                  Sun, 04 Oct 2020   AIC                           -203.952
Time:                          21:00:39   BIC                           -192.769
Sample:                               0   HQIC                          -199.410
                                  - 121
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      2.4818      1.535      1.617      0.106      -0.526       5.490
ar.L1          0.6757      0.200      3.370      0.001       0.283       1.069
ma.L1         -0.4160      0.241     -1.727      0.084      -0.888       0.056
sigma2         0.0100      0.001     11.310      0.000       0.008       0.012
==============================================================================
===
Ljung-Box (Q):                       39.24   Jarque-Bera (JB):
78.35
Prob(Q):                              0.50   Prob(JB):
0.00
Heteroskedasticity (H):               0.29   Skew:
-1.01
Prob(H) (two-sided):                  0.00   Kurtosis:
6.38
```

```
==========================================================================
===

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

============================Results of District 2 ============================
                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(0, 1, 2)   Log Likelihood                67.362
Date:                Sun, 04 Oct 2020   AIC                          -128.724
Time:                        21:00:42   BIC                          -120.361
Sample:                             0   HQIC                         -125.328
                                - 121
Covariance Type:                  opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
ma.L1         -0.6394      0.076     -8.372      0.000      -0.789      -0.490
ma.L2         -0.2300      0.078     -2.941      0.003      -0.383      -0.077
sigma2         0.0188      0.002      8.477      0.000       0.014       0.023
==========================================================================
===
Ljung-Box (Q):                       32.45   Jarque-Bera (JB):
3.44
Prob(Q):                              0.80   Prob(JB):
0.18
Heteroskedasticity (H):               1.37   Skew:
-0.05
Prob(H) (two-sided):                  0.32   Kurtosis:
3.82
==========================================================================
===

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

============================Results of District 3 ============================
                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(1, 1, 1)   Log Likelihood                170.434
Date:                Sun, 04 Oct 2020   AIC                          -332.869
Time:                        21:00:46   BIC                          -321.719
Sample:                             0   HQIC                         -328.341
```

```
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0022      0.001      3.725      0.000       0.001       0.003
ar.L1          0.5292      0.086      6.166      0.000       0.361       0.697
ma.L1         -0.9122      0.050    -18.261      0.000      -1.010      -0.814
sigma2         0.0034      0.000     12.425      0.000       0.003       0.004
==============================================================================
===
Ljung-Box (Q):                       34.74   Jarque-Bera (JB):
116.46
Prob(Q):                              0.71   Prob(JB):
0.00
Heteroskedasticity (H):               0.43   Skew:
0.87
Prob(H) (two-sided):                  0.01   Kurtosis:
7.50
==============================================================================
===

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

============================Results of District 4 ============================
                           SARIMAX Results
==============================================================================
Dep. Variable:                         y   No. Observations:                  121
Model:                 SARIMAX(2, 0, 0)   Log Likelihood                 113.097
Date:                  Sun, 04 Oct 2020   AIC                           -218.194
Time:                          21:00:50   BIC                           -207.011
Sample:                               0   HQIC                          -213.653
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      3.5320      0.838      4.214      0.000       1.889       5.175
ar.L1          0.3394      0.106      3.204      0.001       0.132       0.547
ar.L2          0.1798      0.091      1.976      0.048       0.001       0.358
sigma2         0.0090      0.001     10.170      0.000       0.007       0.011
==============================================================================
===
Ljung-Box (Q):                       25.24   Jarque-Bera (JB):
15.39
Prob(Q):                              0.97   Prob(JB):
```

```
0.00
Heteroskedasticity (H):               0.59   Skew:
0.17
Prob(H) (two-sided):                  0.10   Kurtosis:
4.71
================================================================================
===

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

============================Results of District 5 ============================
                              SARIMAX Results
==============================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:                SARIMAX(0, 1, 2)   Log Likelihood                 153.113
Date:                Sun, 04 Oct 2020   AIC                             -298.226
Time:                        21:00:54   BIC                             -287.076
Sample:                             0   HQIC                            -293.698
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0031      0.002      1.515      0.130      -0.001       0.007
ma.L1         -0.5433      0.105     -5.156      0.000      -0.750      -0.337
ma.L2         -0.1660      0.114     -1.455      0.146      -0.390       0.058
sigma2         0.0045      0.000     12.247      0.000       0.004       0.005
================================================================================
===
Ljung-Box (Q):                       31.99   Jarque-Bera (JB):
61.19
Prob(Q):                              0.81   Prob(JB):
0.00
Heteroskedasticity (H):               0.51   Skew:
-0.04
Prob(H) (two-sided):                  0.04   Kurtosis:
6.50
================================================================================
===

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

============================Results of District 7 ============================
                              SARIMAX Results
```

```
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                36.729
Date:                Sun, 04 Oct 2020   AIC                            -69.458
Time:                        21:00:57   BIC                            -63.883
Sample:                             0   HQIC                           -67.194
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5800      0.058     -9.950      0.000      -0.694      -0.466
sigma2         0.0316      0.003     11.808      0.000       0.026       0.037
==============================================================================
===
Ljung-Box (Q):                       32.31   Jarque-Bera (JB):
66.13
Prob(Q):                              0.80   Prob(JB):
0.00
Heteroskedasticity (H):               1.40   Skew:
-0.98
Prob(H) (two-sided):                  0.29   Kurtosis:
6.06
==============================================================================
===

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

============================Results of District 8 ============================
                           SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                131.956
Date:                Sun, 04 Oct 2020   AIC                           -257.912
Time:                        21:01:01   BIC                           -249.550
Sample:                             0   HQIC                          -254.516
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0018      0.001      1.713      0.087      -0.000       0.004
ma.L1         -0.8679      0.051    -17.087      0.000      -0.967      -0.768
sigma2         0.0064      0.001      7.913      0.000       0.005       0.008
==============================================================================
===
```

```
Ljung-Box (Q):                       19.28   Jarque-Bera (JB):
1.76
Prob(Q):                              1.00   Prob(JB):
0.42
Heteroskedasticity (H):               1.31   Skew:
0.18
Prob(H) (two-sided):                  0.39   Kurtosis:
3.47
===============================================================================
===

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

============================Results of District 9 ============================
                           SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(1, 1, 1)   Log Likelihood                156.644
Date:                Sun, 04 Oct 2020   AIC                         -307.287
Time:                        21:01:06   BIC                         -298.925
Sample:                             0   HQIC                        -303.891
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3820      0.109      3.503      0.000       0.168       0.596
ma.L1         -0.8323      0.072    -11.528      0.000      -0.974      -0.691
sigma2         0.0043      0.001      7.606      0.000       0.003       0.005
===============================================================================
===
Ljung-Box (Q):                       38.53   Jarque-Bera (JB):
2.74
Prob(Q):                              0.54   Prob(JB):
0.25
Heteroskedasticity (H):               0.74   Skew:
0.37
Prob(H) (two-sided):                  0.34   Kurtosis:
2.94
===============================================================================
===

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

```
============================Results of District 10 ============================
                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 159.569
Date:                Sun, 04 Oct 2020   AIC                           -315.138
Time:                        21:01:14   BIC                           -309.563
Sample:                             0   HQIC                          -312.874
                                - 121
Covariance Type:                  opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
ma.L1         -0.5986      0.074     -8.037      0.000      -0.745      -0.453
sigma2         0.0041      0.000     10.278      0.000       0.003       0.005
==========================================================================
===
Ljung-Box (Q):                       43.07   Jarque-Bera (JB):
13.14
Prob(Q):                              0.34   Prob(JB):
0.00
Heteroskedasticity (H):               1.91   Skew:
0.33
Prob(H) (two-sided):                  0.04   Kurtosis:
4.48
==========================================================================
===

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

============================Results of District 11 ============================
                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 146.156
Date:                Sun, 04 Oct 2020   AIC                           -288.313
Time:                        21:01:17   BIC                           -282.738
Sample:                             0   HQIC                          -286.049
                                - 121
Covariance Type:                  opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
ma.L1         -0.7666      0.051    -15.106      0.000      -0.866      -0.667
sigma2         0.0051      0.001      9.258      0.000       0.004       0.006
==========================================================================
```

```
===
Ljung-Box (Q):                       23.05   Jarque-Bera (JB):
9.66
Prob(Q):                              0.99   Prob(JB):
0.01
Heteroskedasticity (H):               1.53   Skew:
0.62
Prob(H) (two-sided):                  0.19   Kurtosis:
3.63
================================================================================
===

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

============================Results of District 12 ============================
                           SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 2)   Log Likelihood                136.967
Date:                Sun, 04 Oct 2020   AIC                         -267.935
Time:                        21:01:21   BIC                         -259.572
Sample:                             0   HQIC                        -264.539
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5030      0.074     -6.803      0.000      -0.648      -0.358
ma.L2         -0.2232      0.090     -2.488      0.013      -0.399      -0.047
sigma2         0.0059      0.001     11.475      0.000       0.005       0.007
==============================================================================
===
Ljung-Box (Q):                       31.17   Jarque-Bera (JB):
92.23
Prob(Q):                              0.84   Prob(JB):
0.00
Heteroskedasticity (H):               1.74   Skew:
0.95
Prob(H) (two-sided):                  0.08   Kurtosis:
6.86
================================================================================
===

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

```
==============================Results of District 13 ============================
                               SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:                SARIMAX(0, 1, 2)   Log Likelihood                110.734
Date:                 Sun, 04 Oct 2020   AIC                         -213.468
Time:                         21:01:25   BIC                         -202.318
Sample:                              0   HQIC                        -208.940
                                 - 121
Covariance Type:                   opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
intercept      0.0045      0.002      1.879      0.060      -0.000       0.009
ma.L1         -0.4113      0.073     -5.607      0.000      -0.555      -0.268
ma.L2         -0.3753      0.071     -5.278      0.000      -0.515      -0.236
sigma2         0.0092      0.001     10.665      0.000       0.007       0.011
==========================================================================
===
Ljung-Box (Q):                       31.06   Jarque-Bera (JB):
85.07
Prob(Q):                              0.84   Prob(JB):
0.00
Heteroskedasticity (H):               0.23   Skew:
0.63
Prob(H) (two-sided):                  0.00   Kurtosis:
6.93
==========================================================================
===

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

==============================Results of District 14 ============================
                               SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:                SARIMAX(0, 1, 2)   Log Likelihood                157.111
Date:                 Sun, 04 Oct 2020   AIC                         -306.221
Time:                         21:01:28   BIC                         -295.071
Sample:                              0   HQIC                        -301.693
                                 - 121
Covariance Type:                   opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
```

```
intercept      0.0039      0.002      2.395      0.017       0.001       0.007
ma.L1         -0.5987      0.105     -5.707      0.000      -0.804      -0.393
ma.L2         -0.2103      0.104     -2.014      0.044      -0.415      -0.006
sigma2         0.0042      0.000      9.493      0.000       0.003       0.005
==============================================================================
===
Ljung-Box (Q):                       32.39   Jarque-Bera (JB):
255.89
Prob(Q):                              0.80   Prob(JB):
0.00
Heteroskedasticity (H):               1.91   Skew:
1.78
Prob(H) (two-sided):                  0.04   Kurtosis:
9.21
==============================================================================
===

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

============================Results of District 15 ============================
                              SARIMAX Results
==============================================================================
Dep. Variable:                         y   No. Observations:                  121
Model:               SARIMAX(1, 1, 1)   Log Likelihood                 147.481
Date:                Sun, 04 Oct 2020   AIC                           -286.961
Time:                        21:01:34   BIC                           -275.811
Sample:                             0   HQIC                          -282.433
                                  - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0016      0.001      2.211      0.027       0.000       0.003
ar.L1          0.3094      0.103      2.994      0.003       0.107       0.512
ma.L1         -0.9200      0.043    -21.356      0.000      -1.004      -0.836
sigma2         0.0050      0.000     10.763      0.000       0.004       0.006
==============================================================================
===
Ljung-Box (Q):                       39.12   Jarque-Bera (JB):
156.84
Prob(Q):                              0.51   Prob(JB):
0.00
Heteroskedasticity (H):               4.15   Skew:
1.52
Prob(H) (two-sided):                  0.00   Kurtosis:
7.71
```

```
==============================================================================
===

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

============================Results of District 16 ============================
                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(1, 0, 1)   Log Likelihood                 164.671
Date:                Sun, 04 Oct 2020   AIC                           -321.341
Time:                        21:01:37   BIC                           -310.158
Sample:                             0   HQIC                          -316.799
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.8336      0.360      2.316      0.021       0.128       1.539
ar.L1          0.8811      0.051     17.131      0.000       0.780       0.982
ma.L1         -0.4039      0.111     -3.630      0.000      -0.622      -0.186
sigma2         0.0038      0.000      9.889      0.000       0.003       0.005
==============================================================================
===
Ljung-Box (Q):                       22.88   Jarque-Bera (JB):
25.08
Prob(Q):                              0.99   Prob(JB):
0.00
Heteroskedasticity (H):               0.57   Skew:
0.83
Prob(H) (two-sided):                  0.08   Kurtosis:
4.49
==============================================================================
===

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

============================Results of District 17 ============================
                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(1, 1, 2)   Log Likelihood                 133.732
Date:                Sun, 04 Oct 2020   AIC                           -259.463
Time:                        21:01:42   BIC                           -248.313
```

```
Sample:                             0   HQIC                          -254.935
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.8780      0.074    -11.888      0.000      -1.023      -0.733
ma.L1          0.4301      0.096      4.490      0.000       0.242       0.618
ma.L2         -0.5328      0.073     -7.275      0.000      -0.676      -0.389
sigma2         0.0063      0.001     12.214      0.000       0.005       0.007
==============================================================================
===
Ljung-Box (Q):                       27.92   Jarque-Bera (JB):
67.43
Prob(Q):                              0.93   Prob(JB):
0.00
Heteroskedasticity (H):               1.84   Skew:
0.80
Prob(H) (two-sided):                  0.06   Kurtosis:
6.31
==============================================================================
===

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

============================Results of District 18 ============================
                              SARIMAX Results
==============================================================================
Dep. Variable:                         y   No. Observations:                 121
Model:                SARIMAX(1, 1, 2)   Log Likelihood                 189.554
Date:                Sun, 04 Oct 2020   AIC                           -369.108
Time:                        21:01:45   BIC                           -355.170
Sample:                             0   HQIC                          -363.448
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0073      0.004      1.702      0.089      -0.001       0.016
ar.L1         -0.8480      0.130     -6.520      0.000      -1.103      -0.593
ma.L1          0.2895      0.132      2.188      0.029       0.030       0.549
ma.L2         -0.6348      0.105     -6.060      0.000      -0.840      -0.429
sigma2         0.0025      0.000     13.588      0.000       0.002       0.003
==============================================================================
===
Ljung-Box (Q):                       31.65   Jarque-Bera (JB):
```

```
674.40
Prob(Q):                                0.82   Prob(JB):
0.00
Heteroskedasticity (H):                 3.27   Skew:
2.14
Prob(H) (two-sided):                    0.00   Kurtosis:
13.79
================================================================================
===

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

============================Results of District 19 ============================
                              SARIMAX Results
================================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:                SARIMAX(0, 1, 2)   Log Likelihood                 171.619
Date:                 Sun, 04 Oct 2020   AIC                           -335.238
Time:                         21:01:49   BIC                           -324.088
Sample:                              0   HQIC                          -330.710
                                 - 121
Covariance Type:                   opg
================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
intercept      0.0038      0.002      1.784      0.074      -0.000       0.008
ma.L1         -0.5063      0.090     -5.612      0.000      -0.683      -0.329
ma.L2         -0.1774      0.108     -1.638      0.101      -0.390       0.035
sigma2         0.0033      0.000      8.591      0.000       0.003       0.004
================================================================================
===
Ljung-Box (Q):                       18.83   Jarque-Bera (JB):
34.90
Prob(Q):                                1.00   Prob(JB):
0.00
Heteroskedasticity (H):                 0.62   Skew:
1.03
Prob(H) (two-sided):                    0.13   Kurtosis:
4.66
================================================================================
===

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

```
============================Results of District 20 ============================
                               SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:                SARIMAX(0, 1, 2)   Log Likelihood                 146.586
Date:                Sun, 04 Oct 2020   AIC                           -285.173
Time:                        21:01:51   BIC                           -274.023
Sample:                             0   HQIC                          -280.645
                                - 121
Covariance Type:                  opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
intercept      0.0050      0.003      1.767      0.077      -0.001       0.011
ma.L1         -0.4389      0.084     -5.243      0.000      -0.603      -0.275
ma.L2         -0.1561      0.074     -2.099      0.036      -0.302      -0.010
sigma2         0.0051      0.001      9.367      0.000       0.004       0.006
==========================================================================
===
Ljung-Box (Q):                       49.42   Jarque-Bera (JB):
12.93
Prob(Q):                              0.15   Prob(JB):
0.00
Heteroskedasticity (H):               0.37   Skew:
0.36
Prob(H) (two-sided):                  0.00   Kurtosis:
4.43
==========================================================================
===

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

============================Results of District 21 ============================
                               SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:                SARIMAX(0, 1, 1)   Log Likelihood                 143.365
Date:                Sun, 04 Oct 2020   AIC                           -282.730
Time:                        21:01:54   BIC                           -277.155
Sample:                             0   HQIC                          -280.466
                                - 121
Covariance Type:                  opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
ma.L1         -0.5835      0.073     -7.997      0.000      -0.727      -0.441
```

```
sigma2         0.0053      0.000     11.546      0.000       0.004       0.006
==============================================================================
===
Ljung-Box (Q):                       25.16   Jarque-Bera (JB):
39.36
Prob(Q):                              0.97   Prob(JB):
0.00
Heteroskedasticity (H):               0.90   Skew:
0.93
Prob(H) (two-sided):                  0.73   Kurtosis:
5.11
==============================================================================
===

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

============================Results of District 22 ============================
                           SARIMAX Results
==============================================================================
Dep. Variable:                         y   No. Observations:                  121
Model:               SARIMAX(0, 1, 2)   Log Likelihood                 127.931
Date:                Sun, 04 Oct 2020   AIC                           -249.862
Time:                        21:01:59   BIC                           -241.500
Sample:                             0   HQIC                          -246.466
                                  - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.3572      0.098     -3.634      0.000      -0.550      -0.165
ma.L2         -0.3833      0.100     -3.835      0.000      -0.579      -0.187
sigma2         0.0069      0.001     11.679      0.000       0.006       0.008
==============================================================================
===
Ljung-Box (Q):                       31.46   Jarque-Bera (JB):
73.29
Prob(Q):                              0.83   Prob(JB):
0.00
Heteroskedasticity (H):               0.43   Skew:
1.10
Prob(H) (two-sided):                  0.01   Kurtosis:
6.13
==============================================================================
===

Warnings:
```

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

============================Results of District 23 ============================
                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(1, 1, 2)   Log Likelihood                 130.765
Date:                Sun, 04 Oct 2020   AIC                           -253.531
Time:                        21:02:05   BIC                           -242.381
Sample:                             0   HQIC                          -249.003
                                - 121
Covariance Type:                  opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
ar.L1         -0.4054      0.245     -1.657      0.097      -0.885       0.074
ma.L1         -0.0512      0.238     -0.215      0.830      -0.519       0.416
ma.L2         -0.5345      0.120     -4.443      0.000      -0.770      -0.299
sigma2         0.0066      0.001     11.739      0.000       0.005       0.008
==========================================================================
===
Ljung-Box (Q):                       25.00   Jarque-Bera (JB):
93.67
Prob(Q):                              0.97   Prob(JB):
0.00
Heteroskedasticity (H):               0.60   Skew:
1.31
Prob(H) (two-sided):                  0.11   Kurtosis:
6.44
==========================================================================
===

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

============================Results of District 25 ============================
                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(3, 0, 0)   Log Likelihood                 104.741
Date:                Sun, 04 Oct 2020   AIC                           -199.481
Time:                        21:02:11   BIC                           -185.502
Sample:                             0   HQIC                          -193.804
                                - 121
Covariance Type:                  opg
==========================================================================
```

```
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      1.7604      0.606      2.903      0.004       0.572       2.949
ar.L1          0.5781      0.088      6.534      0.000       0.405       0.752
ar.L2         -0.1072      0.112     -0.959      0.338      -0.326       0.112
ar.L3          0.2685      0.088      3.039      0.002       0.095       0.442
sigma2         0.0103      0.001      8.124      0.000       0.008       0.013
==============================================================================
===
Ljung-Box (Q):                       44.90   Jarque-Bera (JB):
0.36
Prob(Q):                              0.27   Prob(JB):
0.83
Heteroskedasticity (H):               1.55   Skew:
0.08
Prob(H) (two-sided):                  0.17   Kurtosis:
3.22
==============================================================================
===

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

============================Results of District 26 ============================
                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(1, 1, 1)   Log Likelihood                 118.088
Date:                Sun, 04 Oct 2020   AIC                           -228.175
Time:                        21:02:14   BIC                           -217.025
Sample:                             0   HQIC                          -223.647
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0024      0.001      1.972      0.049    1.42e-05       0.005
ar.L1          0.2627      0.101      2.606      0.009       0.065       0.460
ma.L1         -0.8728      0.055    -15.843      0.000      -0.981      -0.765
sigma2         0.0081      0.001      8.115      0.000       0.006       0.010
==============================================================================
===
Ljung-Box (Q):                       36.88   Jarque-Bera (JB):
7.10
Prob(Q):                              0.61   Prob(JB):
0.03
Heteroskedasticity (H):               2.48   Skew:
```

```
-0.53
Prob(H) (two-sided):                  0.01   Kurtosis:
3.55
================================================================================
===

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

============================Results of District 27 ============================
                              SARIMAX Results
================================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(0, 1, 2)   Log Likelihood                 148.552
Date:                Sun, 04 Oct 2020   AIC                           -291.104
Time:                        21:02:19   BIC                           -282.742
Sample:                             0   HQIC                          -287.708
                                - 121
Covariance Type:                  opg
================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
ma.L1         -0.2355      0.094     -2.499      0.012      -0.420      -0.051
ma.L2         -0.2633      0.093     -2.823      0.005      -0.446      -0.081
sigma2         0.0049      0.000     11.693      0.000       0.004       0.006
================================================================================
===
Ljung-Box (Q):                       35.06   Jarque-Bera (JB):
61.50
Prob(Q):                              0.69   Prob(JB):
0.00
Heteroskedasticity (H):               0.68   Skew:
0.85
Prob(H) (two-sided):                  0.22   Kurtosis:
6.07
================================================================================
===

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

============================Results of District 28 ============================
                              SARIMAX Results
================================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(1, 1, 1)   Log Likelihood                 136.384
```

```
Date:                Sun, 04 Oct 2020   AIC                           -266.768
Time:                        21:02:26   BIC                           -258.406
Sample:                             0   HQIC                          -263.372
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.6429      0.102      6.333      0.000       0.444       0.842
ma.L1         -0.9283      0.051    -18.116      0.000      -1.029      -0.828
sigma2         0.0060      0.001     10.241      0.000       0.005       0.007
===================================================================================
===
Ljung-Box (Q):                       24.53   Jarque-Bera (JB):
31.75
Prob(Q):                              0.97   Prob(JB):
0.00
Heteroskedasticity (H):               0.83   Skew:
0.11
Prob(H) (two-sided):                  0.56   Kurtosis:
5.51
===================================================================================
===

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

The Result table

[17]:
```
Result_1 = pd.DataFrame({'district': districts, 'MAPE': MAPE, 'log_Predicted':␣
 ↪log_Predicted, 'log_Currentprice': log_Currentprice})
Result_1['Predicted'] = np.exp(Result_1['log_Predicted'])
Result_1['Currentprice'] = np.exp(Result_1['log_Currentprice'])
Result_1['percentage_increase'] = Result_1['Predicted'] /␣
 ↪Result_1['Currentprice'] - 1
Result_1.sort_values('percentage_increase', ascending=False)
```

[17]:

	district	MAPE	log_Predicted	log_Currentprice	Predicted
18	20	0.710170	7.577187	7.380002	1953.127327
1	2	1.846413	7.675950	7.488453	2155.871823
12	14	0.410972	7.498114	7.331785	1804.635319
23	26	1.170766	7.147431	6.984716	1270.836826
16	18	0.441463	7.303875	7.179381	1486.046496
25	28	1.046103	7.024037	6.900328	1123.312170
17	19	0.715485	7.402667	7.279652	1640.353697
4	5	0.395657	7.450505	7.330332	1720.731237

22	25	1.685449	6.756545	6.646391	859.666818
11	13	0.703470	7.624110	7.522610	2046.956892
13	15	1.103909	7.434468	7.344958	1693.356315
3	4	0.781964	7.346149	7.262453	1550.214907
2	3	0.369448	7.794391	7.718808	2426.950447
21	23	1.151870	7.039318	6.967995	1140.609504
14	16	0.670500	7.008690	6.972271	1106.204694
9	11	2.047353	7.551950	7.518142	1904.452798
15	17	1.139743	6.938309	6.907073	1031.025002
20	22	0.875019	6.947774	6.930495	1040.830375
8	10	0.909143	7.680636	7.677465	2165.997183
7	9	0.565334	7.735511	7.739460	2288.177880
19	21	0.902176	7.430290	7.451450	1686.296517
24	27	0.939378	6.857823	6.895345	951.294145
0	1	0.529892	7.652918	7.715792	2106.783476
10	12	1.055623	7.291148	7.357769	1467.253584
5	7	4.266047	7.845551	7.950855	2554.344535
6	8	1.319041	7.380913	7.528108	1605.054631

	Currentprice	percentage_increase
18	1603.592593	0.217970
1	1787.285714	0.206227
12	1528.107143	0.180961
23	1080.000000	0.176701
16	1312.095238	0.132575
25	992.600000	0.131687
17	1450.483146	0.130902
4	1525.888889	0.127691
22	770.000000	0.116450
11	1849.387755	0.106829
13	1548.370370	0.093638
3	1425.750000	0.087298
2	2250.275862	0.078512
21	1062.090909	0.073928
14	1066.642857	0.037090
9	1841.142857	0.034386
15	999.318182	0.031728
20	1023.000000	0.017429
8	2159.138889	0.003176
7	2297.230769	-0.003941
19	1722.358209	-0.020937
24	987.666667	-0.036827
0	2243.500000	-0.060939
10	1568.333333	-0.064450
5	2838.000000	-0.099949
6	1859.583333	-0.136874

0.0.3 Now we want to want to focus on 'resale' property market

[18]:
```
#Check the values in feature ['Type of Sale']
data['Type of Sale'].unique()
```

[18]:
```
array(['Resale', 'New Sale'], dtype=object)
```

[19]:
```
#Subset the data
data_resale = data[data['Type of Sale']=='Resale']
data_resale.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 82389 entries, 0 to 190793
Data columns (total 21 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Project Name                  82389 non-null  object
 1   Address                       82389 non-null  object
 2   No. of Units                  82389 non-null  int64
 3   Area (sqm)                    82389 non-null  int64
 4   Type of Area                  82389 non-null  object
 5   Transacted Price ($)          82389 non-null  int64
 6   Nett Price($)                 82389 non-null  object
 7   Unit Price ($ psm)            82389 non-null  int64
 8   Unit Price ($ psf)            82389 non-null  int64
 9   Sale Date                     82389 non-null  object
 10  Property Type                 82389 non-null  object
 11  Tenure                        82389 non-null  object
 12  Completion Date               82389 non-null  object
 13  Type of Sale                  82389 non-null  object
 14  Purchaser Address Indicator   82389 non-null  object
 15  Postal District               82389 non-null  int64
 16  Postal Sector                 82389 non-null  int64
 17  Postal Code                   82389 non-null  int64
 18  Planning Region               82389 non-null  object
 19  Planning Area                 82389 non-null  object
 20  Month_Year                    82389 non-null  period[M]
dtypes: int64(8), object(12), period[M](1)
memory usage: 13.8+ MB
```

[20]:
```
#Average_monthly for each district
Average_monthly_resale = data_resale.groupby(['Postal District', 'Month_Year'],␣
 ↪as_index=False)['Unit Price ($ psf)'].mean()
Average_monthly_resale_df = pd.DataFrame(Average_monthly_resale)
```

[21]:
```
Average_monthly_resale_df.info()
Average_monthly_resale_df['Postal District'].value_counts()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3127 entries, 0 to 3126
Data columns (total 3 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Postal District     3127 non-null   int64
 1   Month_Year          3127 non-null   period[M]
 2   Unit Price ($ psf)  3127 non-null   float64
dtypes: float64(1), int64(1), period[M](1)
memory usage: 97.7 KB
```

[21]:
```
27    121
8     121
20    121
22    121
14    121
12    121
10    121
3     121
5     121
9     121
16    121
11    121
4     121
15    121
17    121
19    121
21    121
23    121
18    121
1     120
26    120
28    119
25    119
13    118
2     116
7     109
6       7
Name: Postal District, dtype: int64
```

Fill in missing value

[22]:
```
#Input missing datapoint (except district 6)
# Drop District 6
Average_monthly_df_drop_resale =␣
 ↪Average_monthly_resale_df[Average_monthly_resale_df['Postal District'] != 6]

#Input missing datapoint
```

```
# make it as a pivot table
pivot_2 = Average_monthly_df_drop_resale.pivot(index='Month_Year', columns =␣
↪'Postal District', values = 'Unit Price ($ psf)')
#Fill in missing values
pivot_2.fillna(method='ffill', inplace=True)
#Back to an array
Average_monthly_df_filled_resale = pivot_2.stack().reset_index()
Average_monthly_df_filled_resale.sort_values(['Month_Year', 'Postal District'])
#To check
Average_monthly_df_filled_resale['Postal District'].value_counts()
```

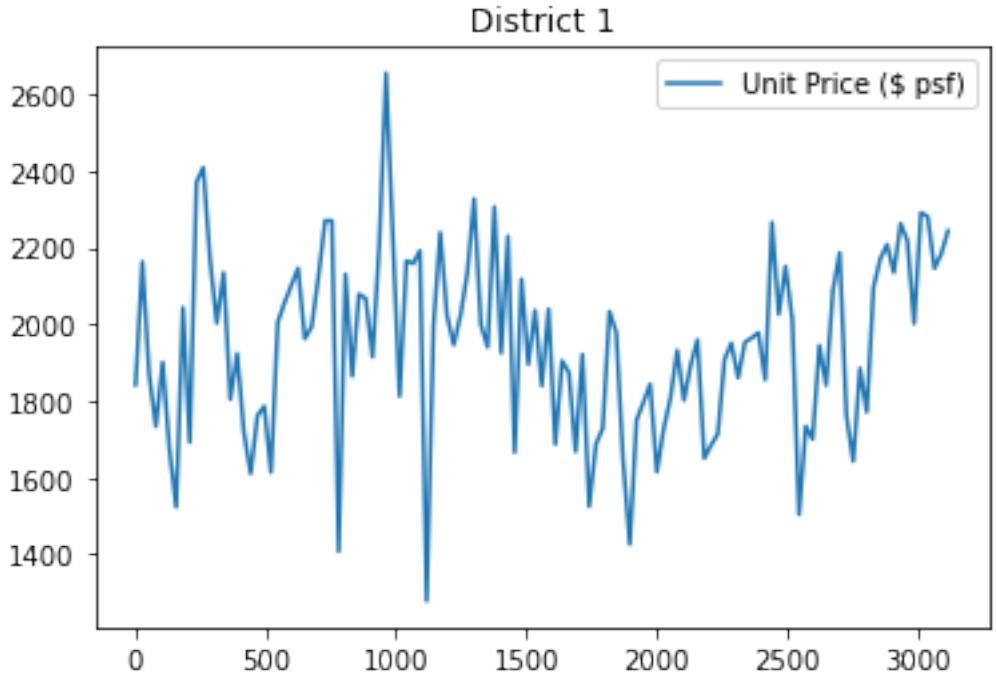
[22]:
```
27    121
25    121
4     121
8     121
10    121
12    121
14    121
16    121
18    121
20    121
22    121
26    121
28    121
1     121
3     121
5     121
7     121
9     121
11    121
13    121
15    121
17    121
19    121
21    121
23    121
2     121
Name: Postal District, dtype: int64
```

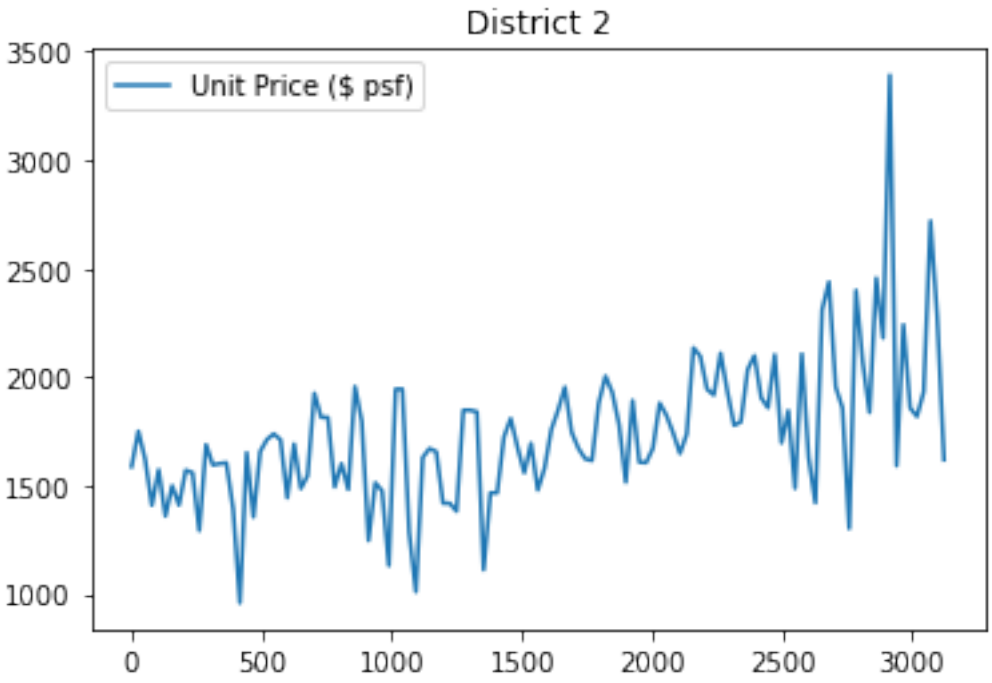
[23]:
```
Average_monthly_df_filled_resale.rename(columns={0: "Unit Price ($ psf)"},␣
↪inplace=True)
Average_monthly_df_filled.info()
```

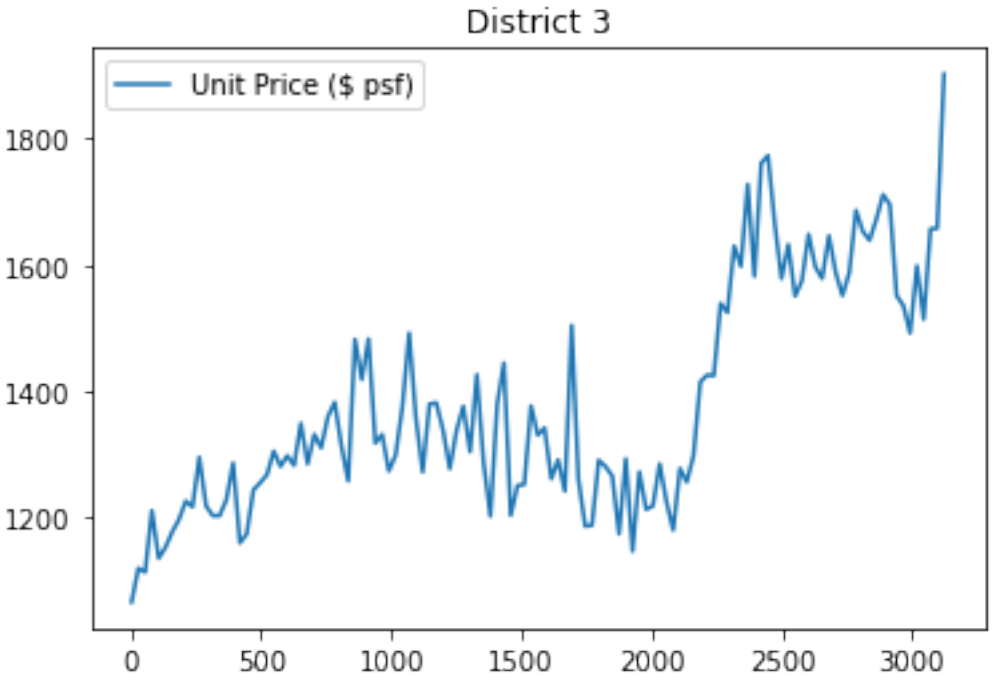
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3146 entries, 0 to 3145
Data columns (total 4 columns):
 #   Column              Non-Null Count  Dtype
```

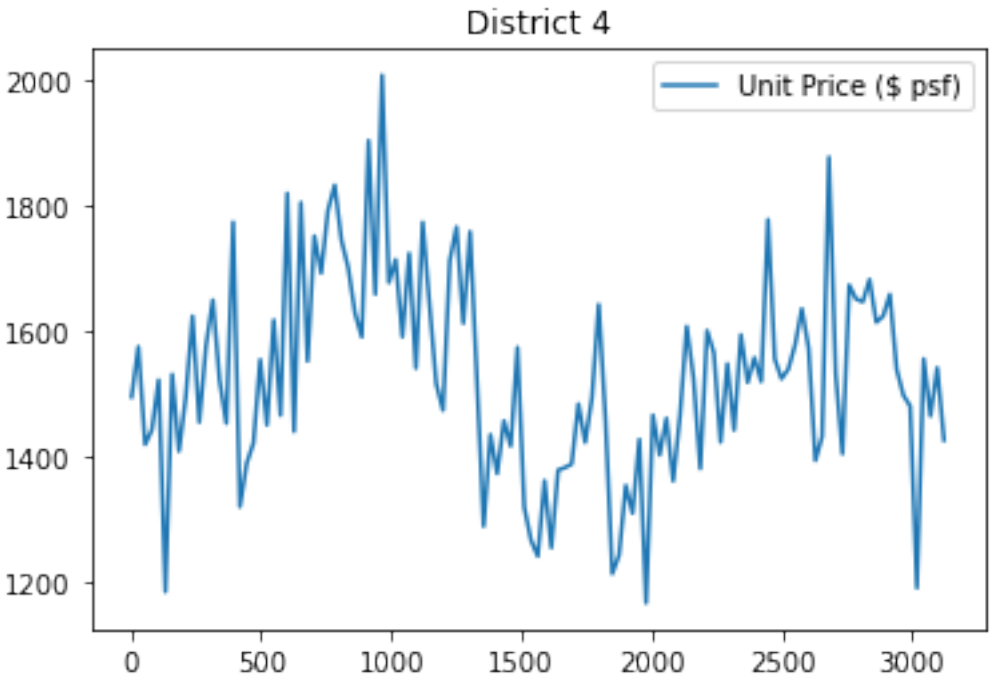
```
---  ------              --------------  -----
 0   Month_Year          3146 non-null   period[M]
 1   Postal District     3146 non-null   int64
 2   Unit Price ($ psf)  3146 non-null   float64
 3   log_psf             3146 non-null   float64
dtypes: float64(2), int64(1), period[M](1)
memory usage: 98.4 KB
```

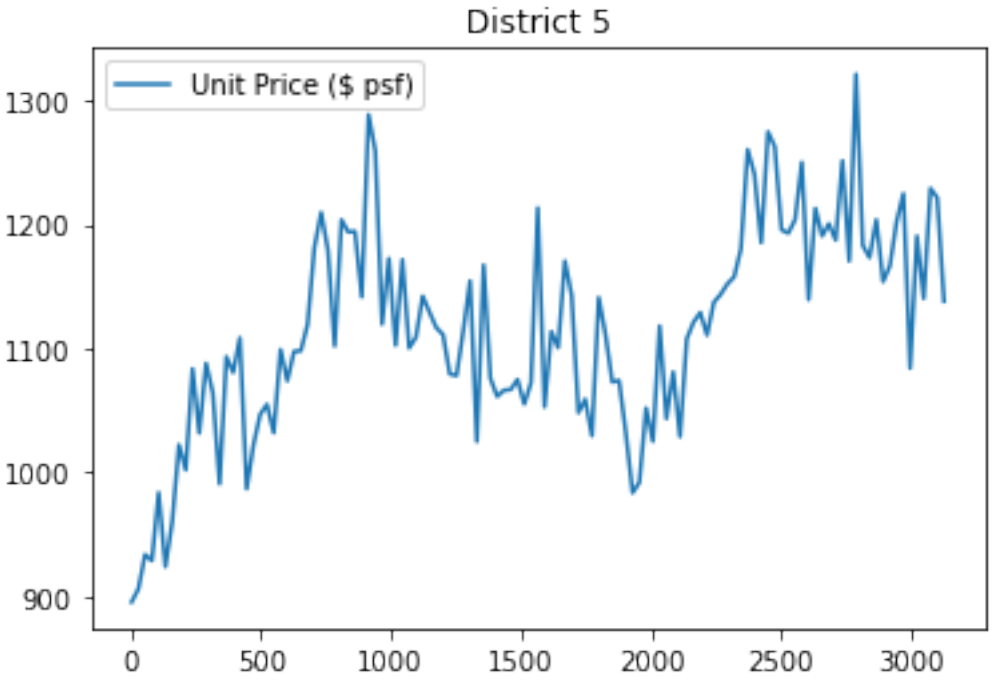
[24]:
```
# Plot monthly average unit price of resale propertiesfor each district for␣
↪past ten years
for i in [x for x in range(1,29) if x != 6 and x !=24]:
    Average_monthly_df_filled_resale.
↪loc[(Average_monthly_df_filled_resale['Postal District'] == i)].plot(y='Unit␣
↪Price ($ psf)')
    plt.title('District ' + str(i))
    plt.show()
```

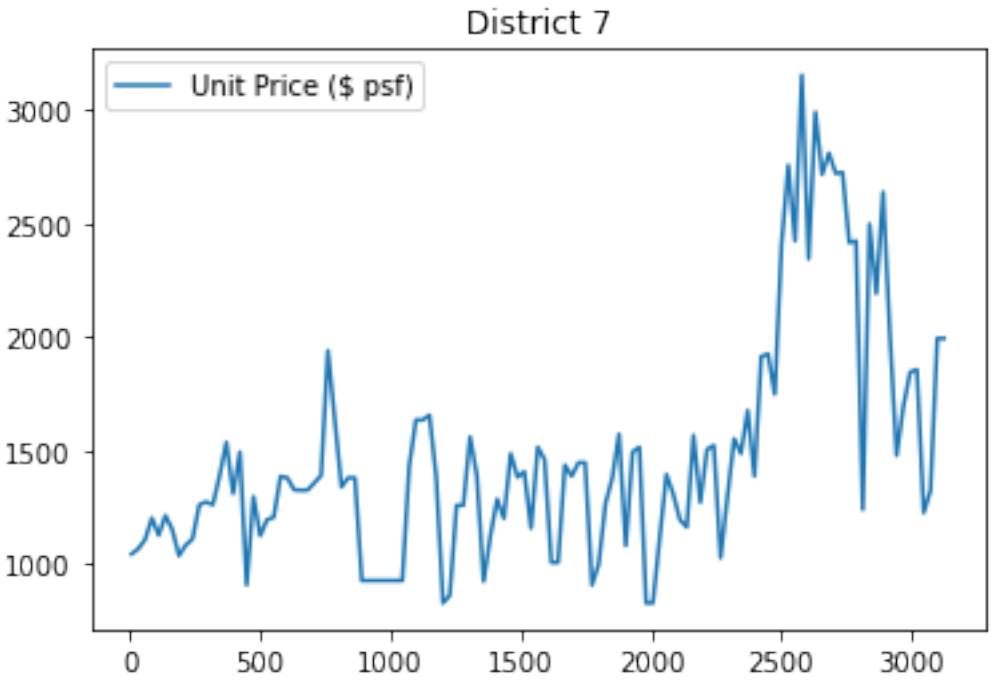


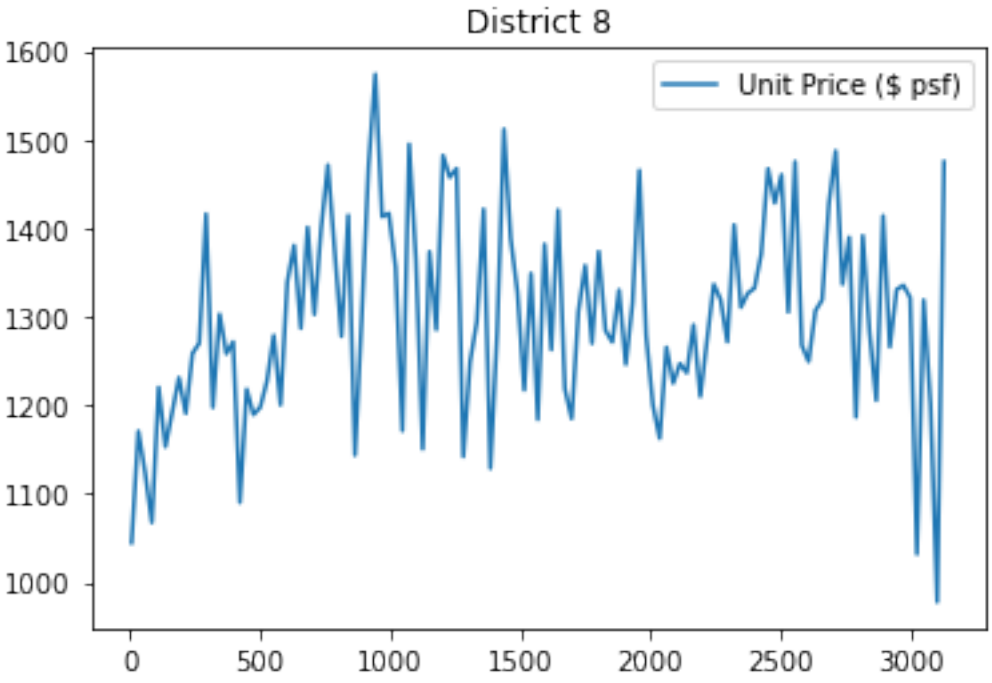

District 2
Unit Price ($ psf)
3500
3000
2500
2000
1500
1000
0
500
1000
1500
2000
2500
3000

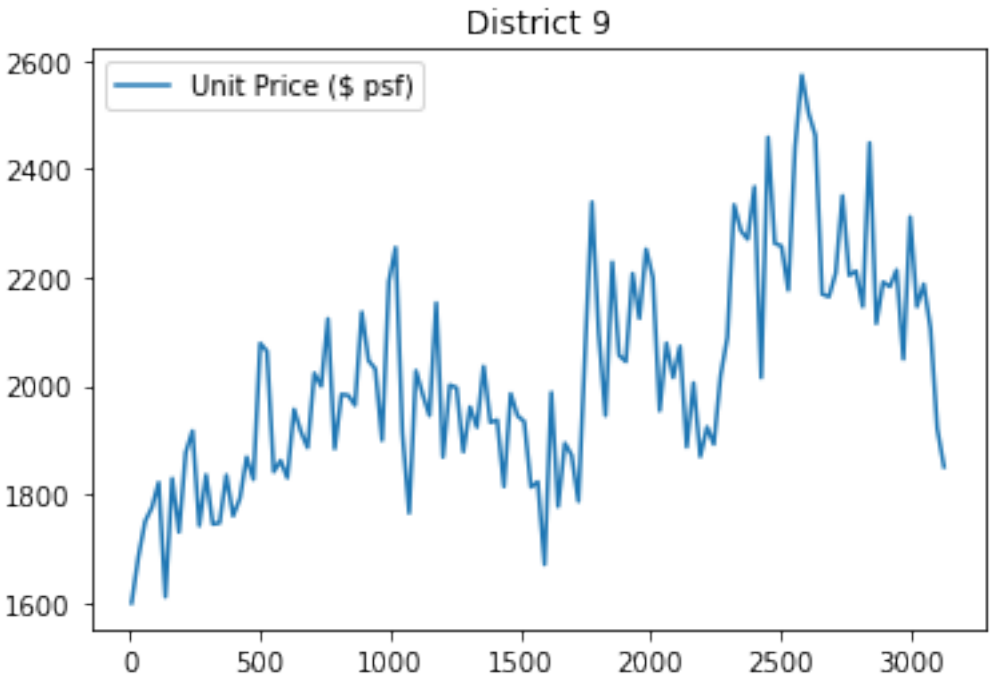

District 3
Unit Price ($ psf)
1800
1600
1400
1200
0
500
1000
1500
2000
2500
3000

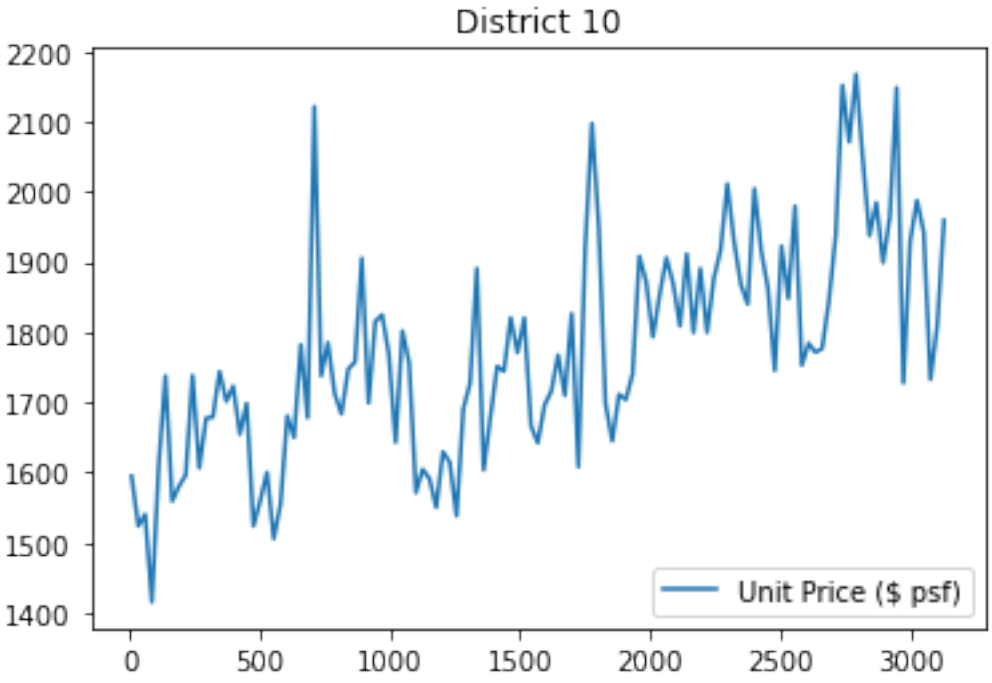

District 4
Unit Price ($ psf)
2000
1800
1600
1400
1200
0
500
1000
1500
2000
2500
3000

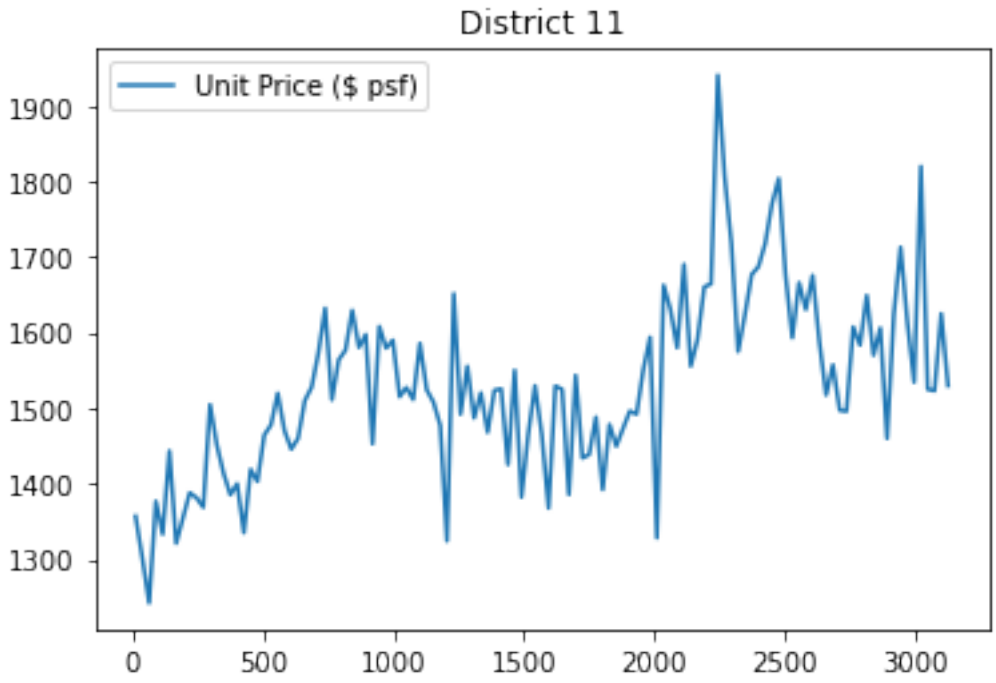

District 5
Unit Price ($ psf)
1300
1200
1100
1000
900
0
500
1000
1500
2000
2500
3000

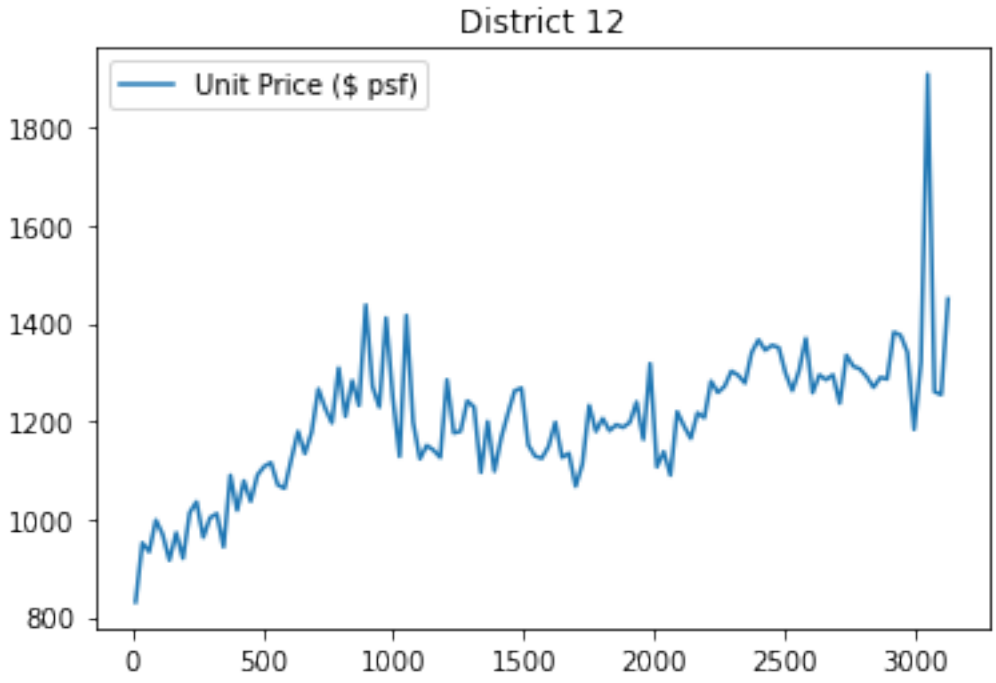

District 7
Unit Price ($ psf)
3000
2500
2000
1500
1000
0
500
1000
1500
2000
2500
3000

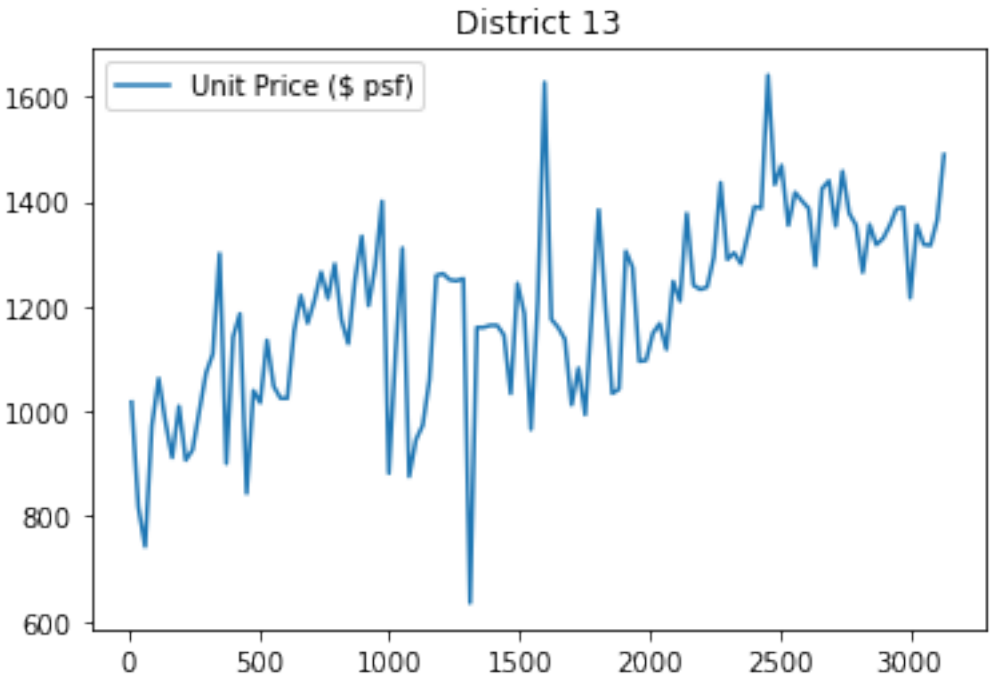

District 8
Unit Price ($ psf)
1600
1500
1400
1300
1200
1100
1000
0
500
1000
1500
2000
2500
3000

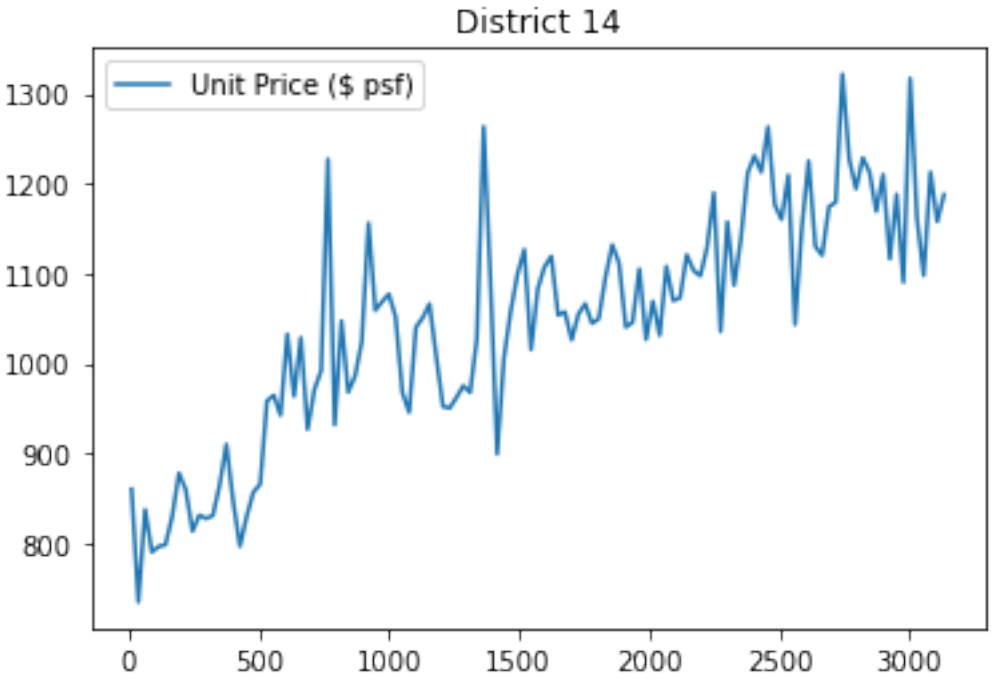

District 9
Unit Price ($ psf)
2600
2400
2200
2000
1800
1600
0
500
1000
1500
2000
2500
3000

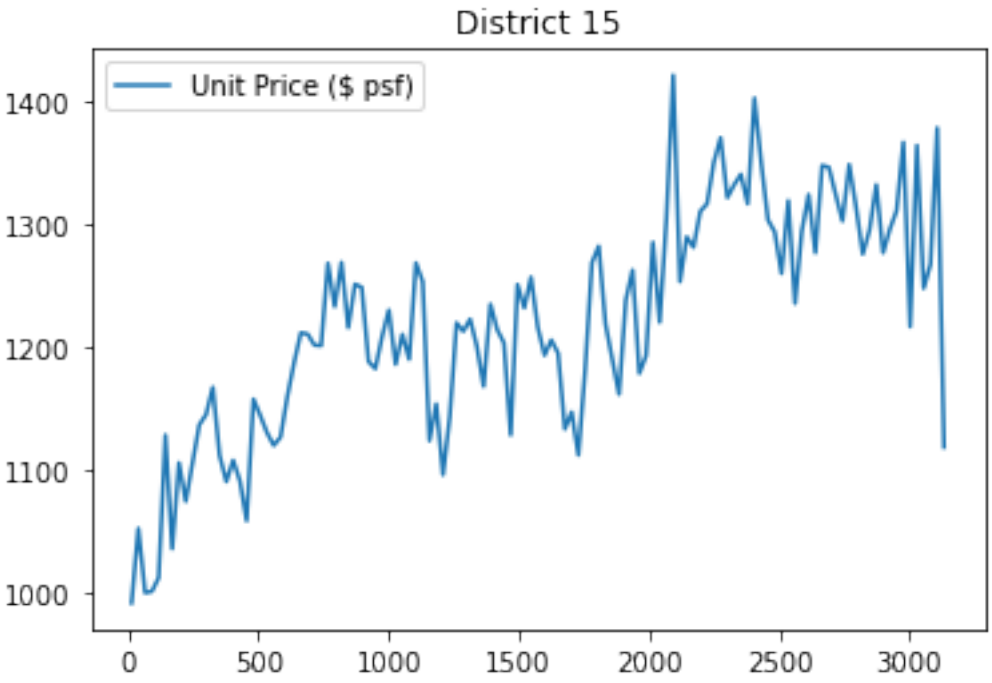

District 10
Unit Price ($ psf)
2200
2100
2000
1900
1800
1700
1600
1500
1400
0
500
1000
1500
2000
2500
3000

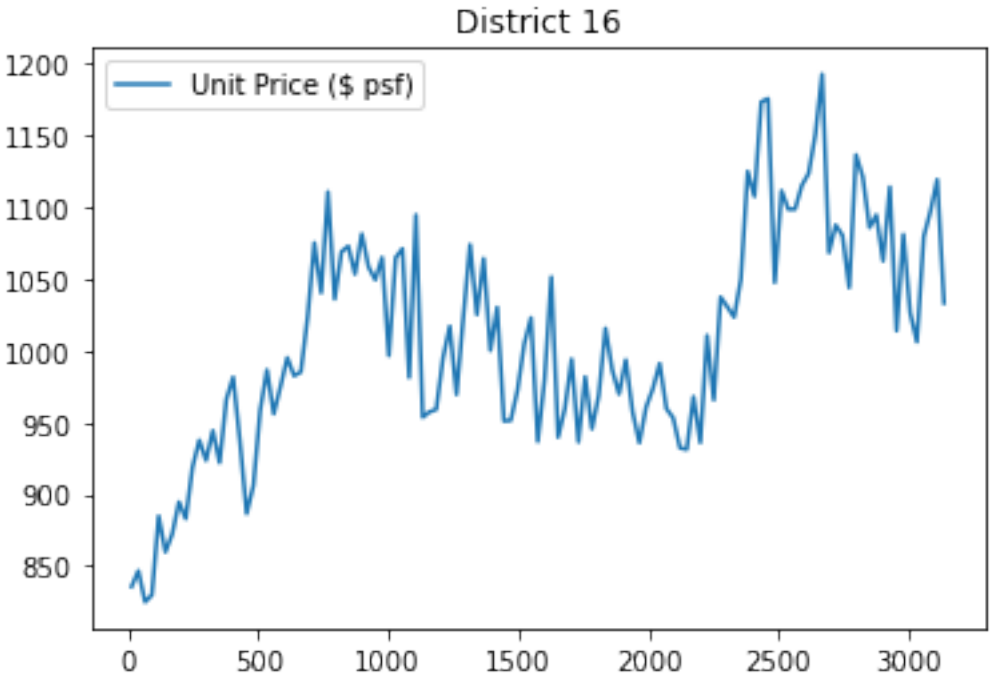

District 11
Unit Price ($ psf)
1900
1800
1700
1600
1500
1400
1300
0
500
1000
1500
2000
2500
3000

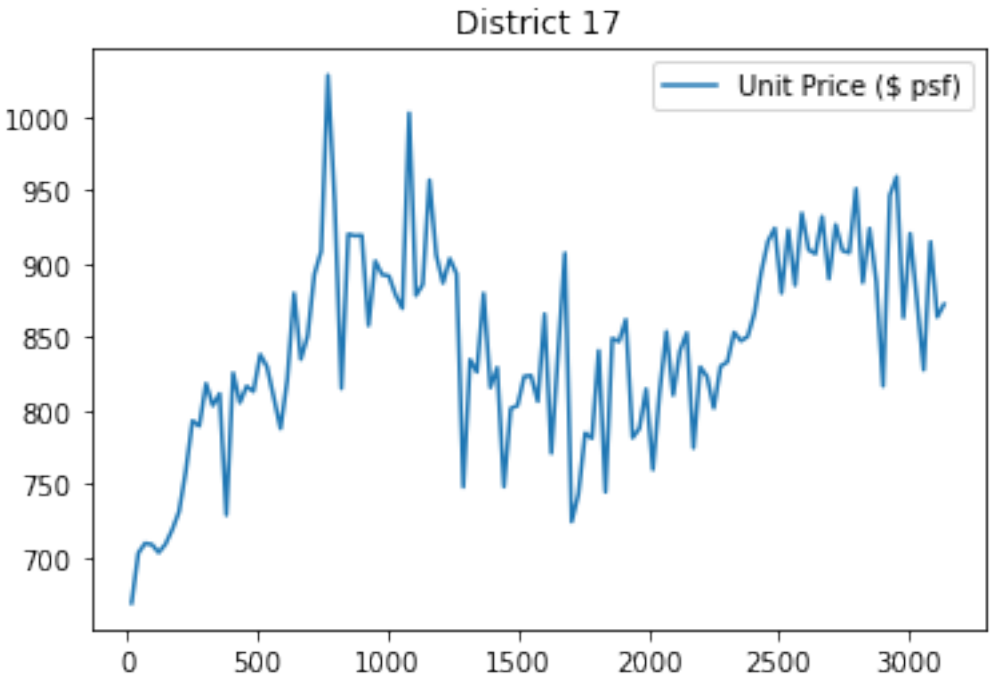

District 12
Unit Price ($ psf)
1800
1600
1400
1200
1000
800
0
500
1000
1500
2000
2500
3000

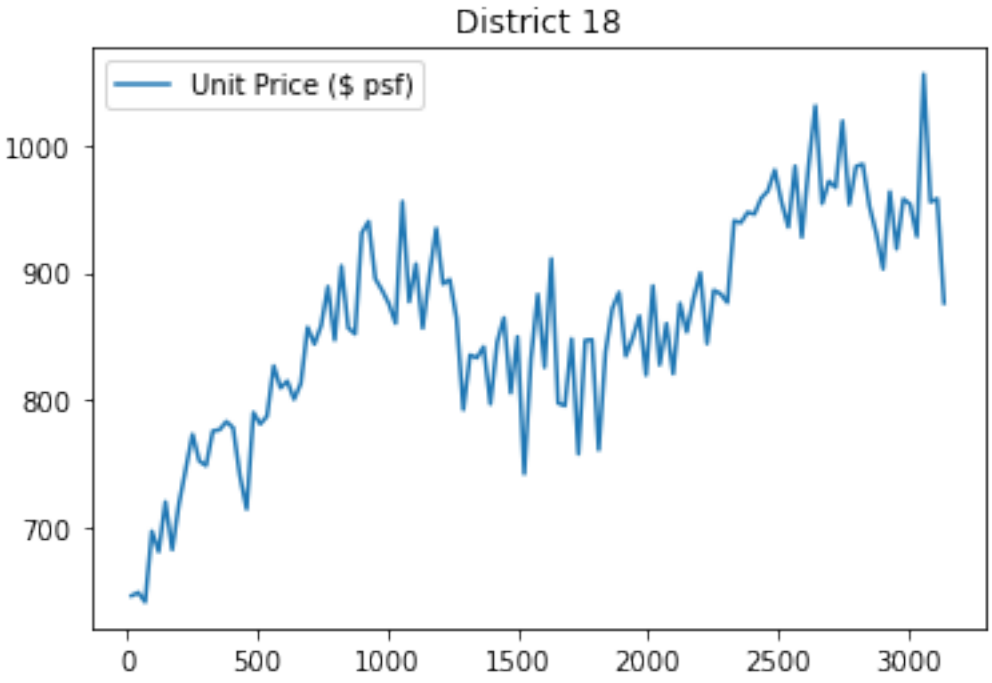

District 13
Unit Price ($ psf)
1600
1400
1200
1000
800
600
0
500
1000
1500
2000
2500
3000

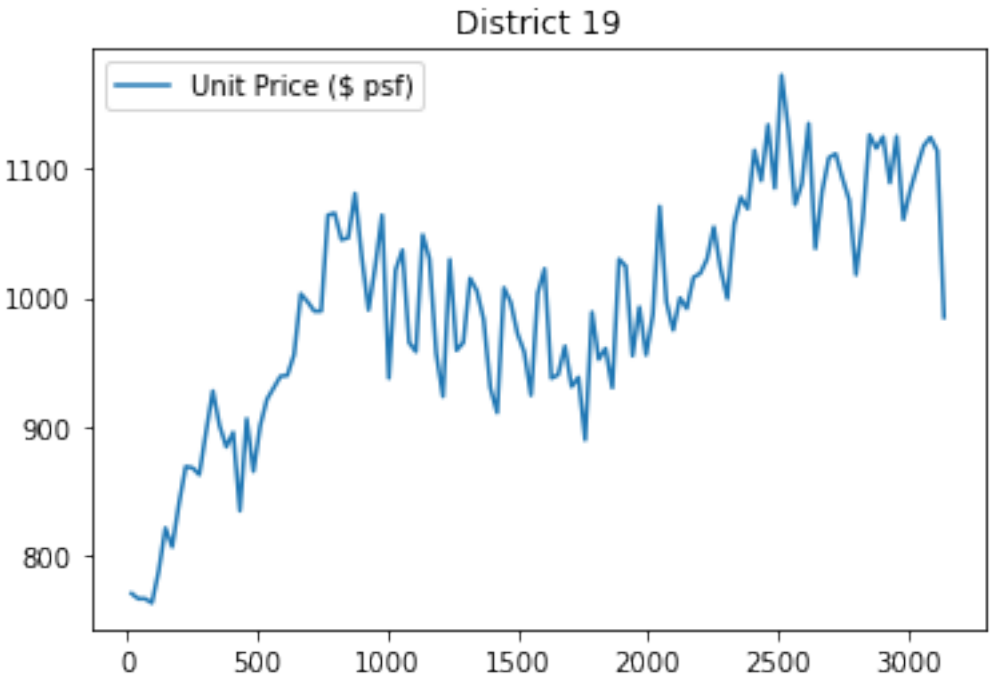

District 14
Unit Price ($ psf)
1300
1200
1100
1000
900
800
0
500
1000
1500
2000
2500
3000

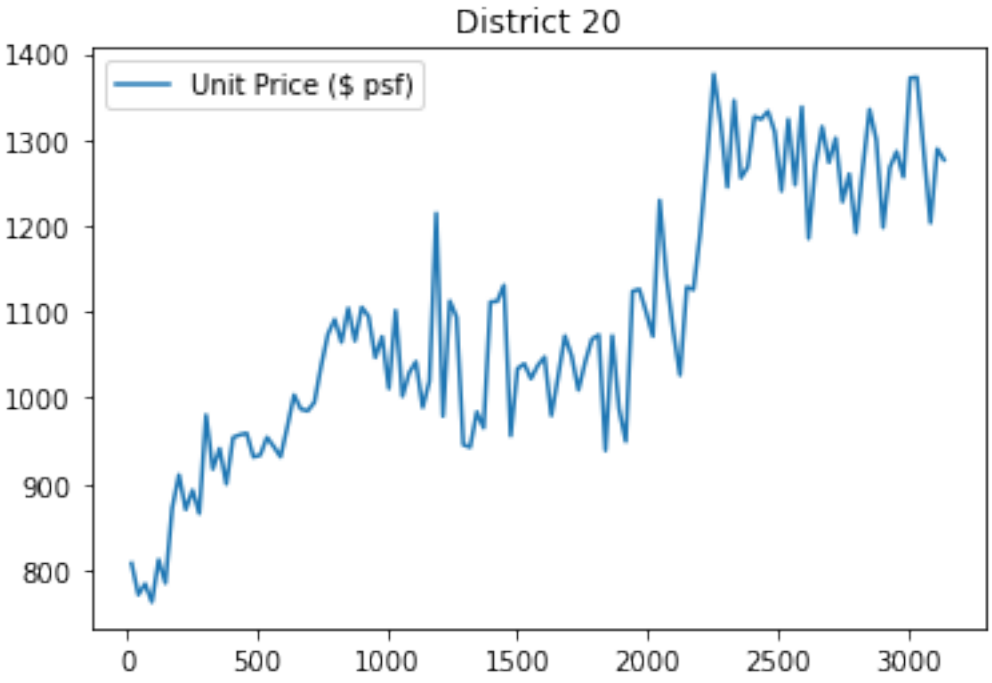

District 15
Unit Price ($ psf)
1400
1300
1200
1100
1000
0
500
1000
1500
2000
2500
3000

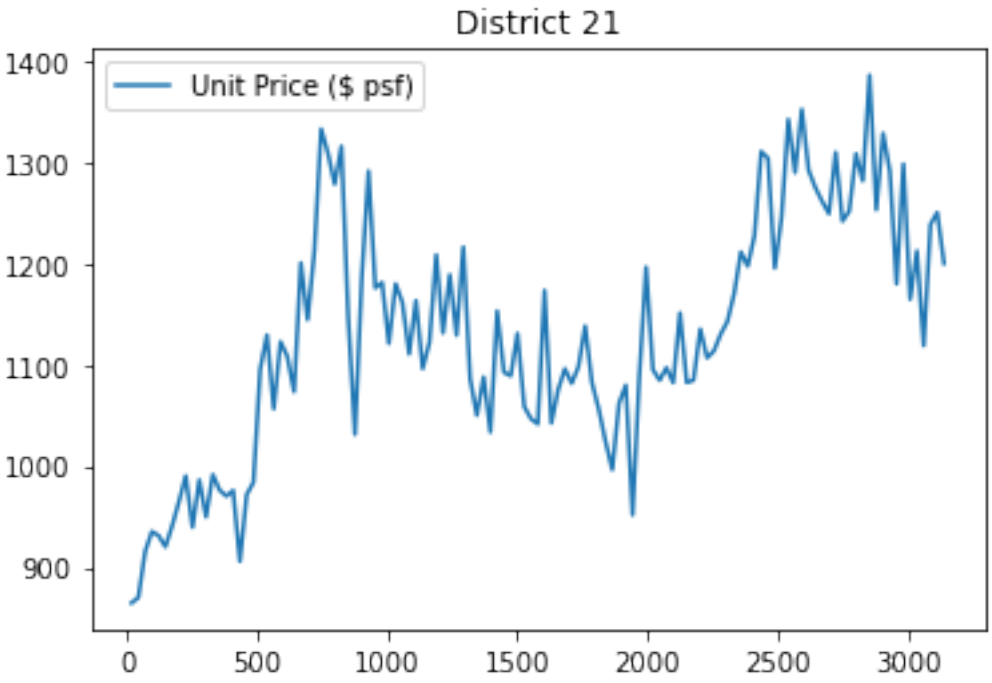

District 16
Unit Price ($ psf)
1200
1150
1100
1050
1000
950
900
850
0
500
1000
1500
2000
2500
3000

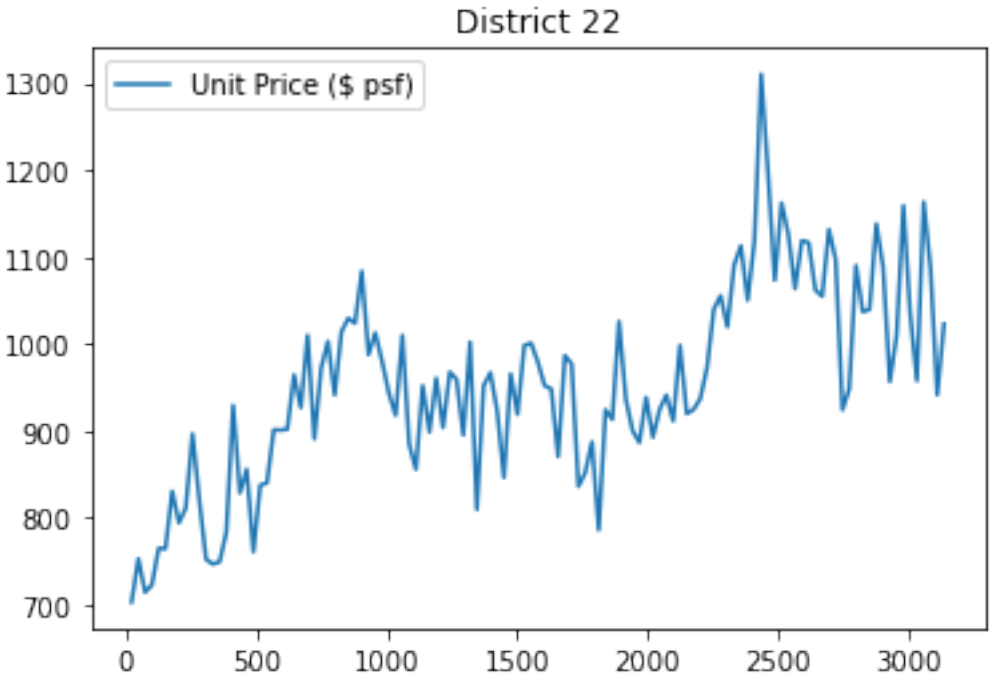

District 17
Unit Price ($ psf)
1000
950
900
850
800
750
700
0
500
1000
1500
2000
2500
3000

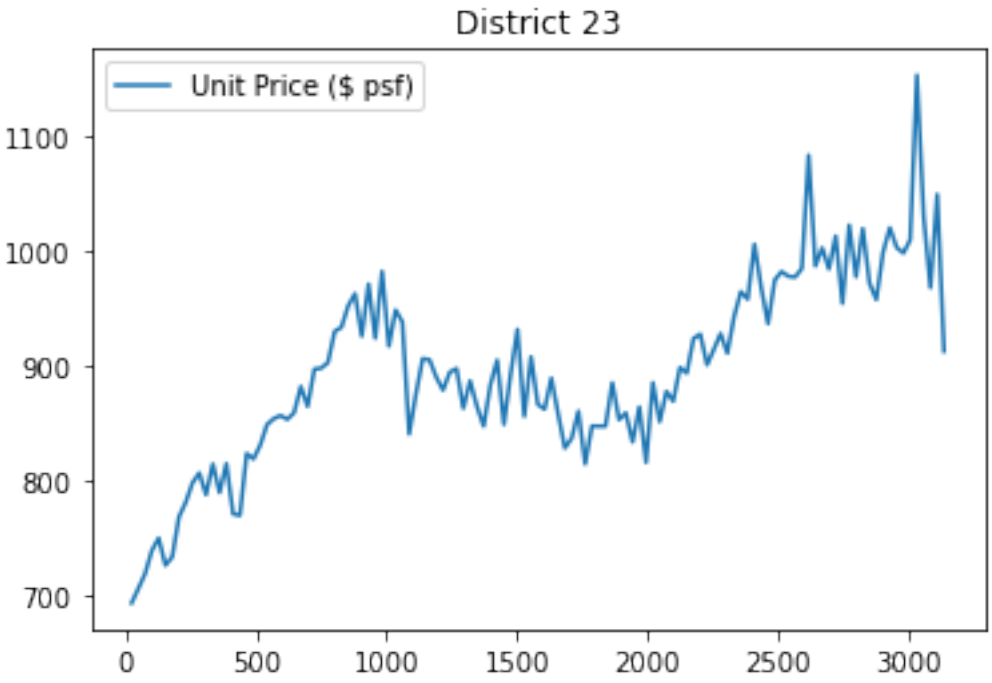

District 18
Unit Price ($ psf)
1000
900
800
700
0
500
1000
1500
2000
2500
3000

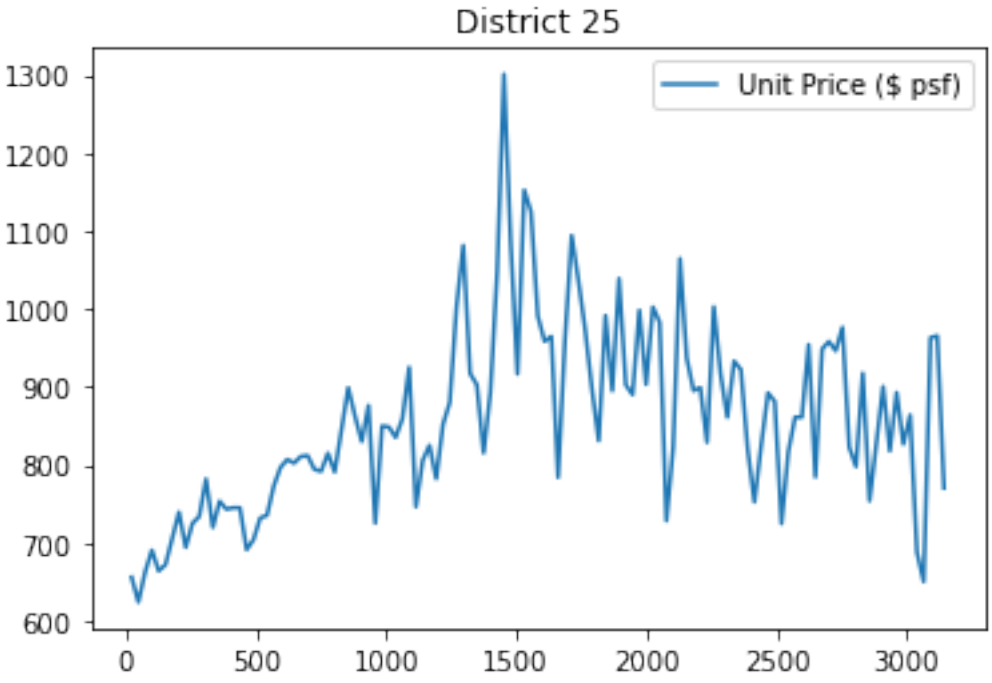

District 19
Unit Price ($ psf)
1100
1000
900
800
0
500
1000
1500
2000
2500
3000

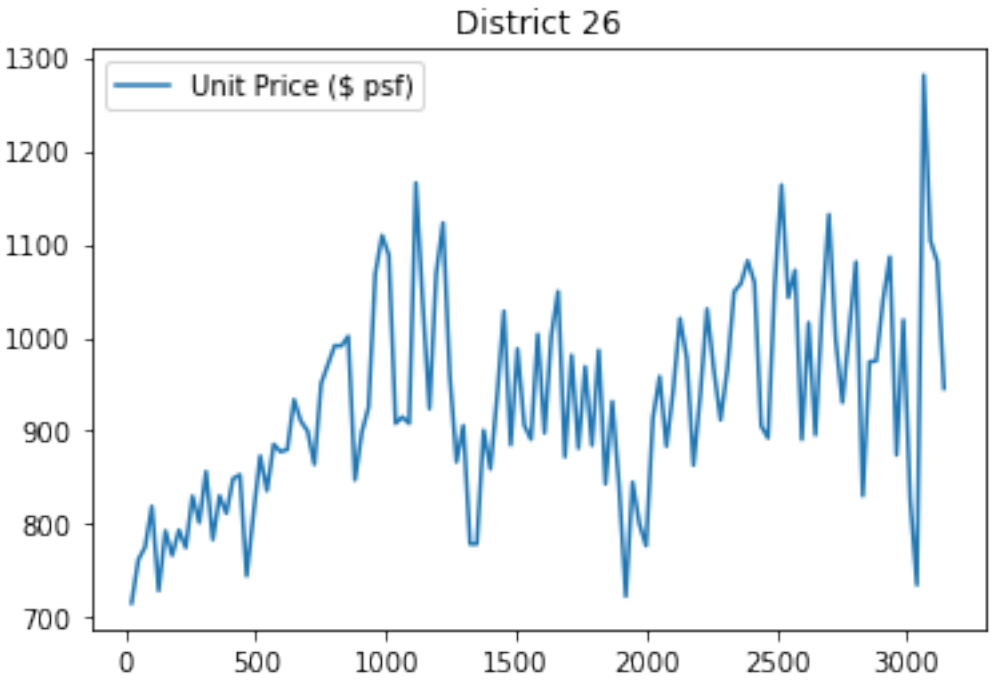

District 20
Unit Price ($ psf)
1400
1300
1200
1100
1000
900
800
0
500
1000
1500
2000
2500
3000

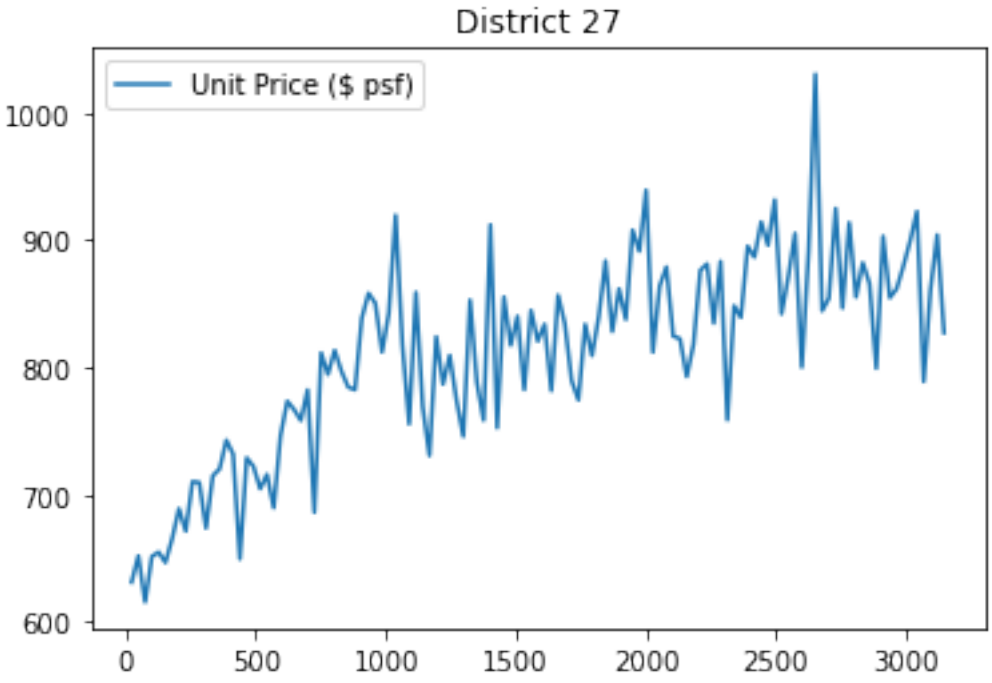

District 21
Unit Price ($ psf)
1400
1300
1200
1100
1000
900
0
500
1000
1500
2000
2500
3000

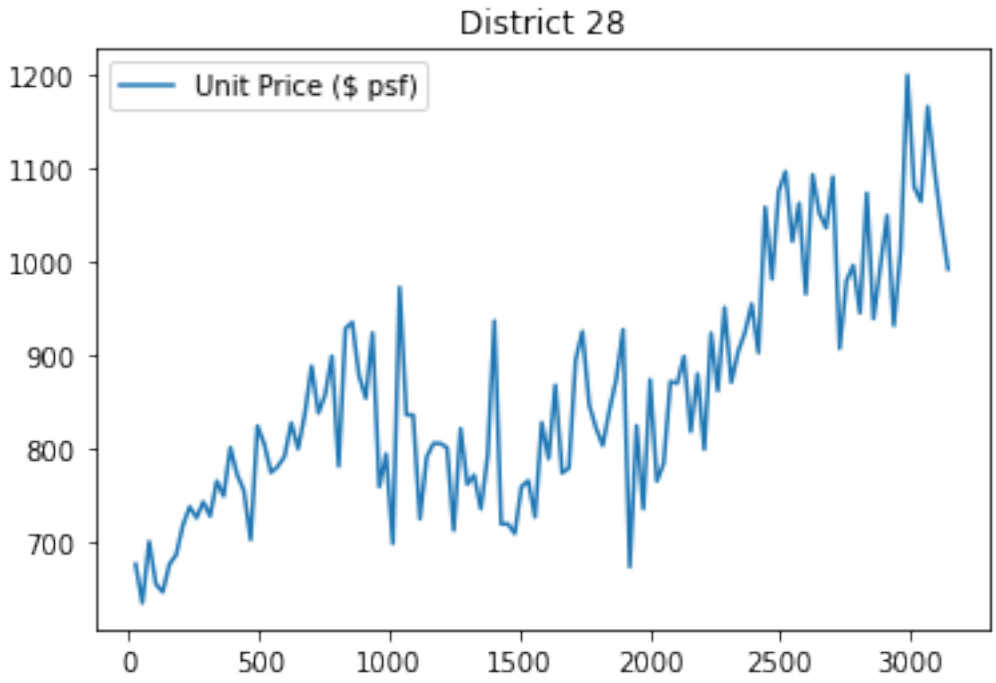

District 22
Unit Price ($ psf)
1300
1200
1100
1000
900
800
700
0
500
1000
1500
2000
2500
3000


District 23
Unit Price ($ psf)
1100
1000
900
800
700
0
500
1000
1500
2000
2500
3000


District 25
Unit Price ($ psf)
1300
1200
1100
1000
900
800
700
600
0
500
1000
1500
2000
2500
3000


District 26
Unit Price ($ psf)
1300
1200
1100
1000
900
800
700
0
500
1000
1500
2000
2500
3000


District 27
Unit Price ($ psf)
1000
900
800
700
600
0
500
1000
1500
2000
2500
3000



[25]:
```
#Add a column to store log_psf
Average_monthly_df_filled_resale.loc[ : , 'log_psf'] = np.
 ↪log(Average_monthly_df_filled_resale['Unit Price ($ psf)'])
Average_monthly_df_filled_resale.head()
```

[25]:

	Month_Year	Postal District	Unit Price ($ psf)	log_psf
0	2010-09	1	1842.100000	7.518662
1	2010-09	2	1586.833333	7.369496
2	2010-09	3	1066.400000	6.972044
3	2010-09	4	1494.769231	7.309727
4	2010-09	5	895.509434	6.797393

[26]:
```
# Print model summaries, plots, and prediction results for all districts except␣
 ↪Distrct 6 & 24
districts_2 = []
MAPE_2 = []
percentage_change_2 = []
log_Predicted_2 = []
log_Currentprice_2 = []
for i in [x for x in range(1,29) if x != 6 and x !=24]:
        print('============================Results of District',␣
 ↪i,'============================')
        # train the model
```

```
        train_size = round(0.9 *␣
↪len(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
↪District'] == i)]))
        test_size = round(0.1 *␣
↪len(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
↪District'] == i)]))
        train =␣
↪Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
↪District'] == i)][:train_size]
        test =␣
↪Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
↪District'] == i)][train_size:]

        if len(train) >= 50:
            model = pm.auto_arima(train[(train['Postal District'] ==␣
↪i)]['log_psf'], suppress_warnings=True)
            print()
            temp = model.order
            p, d, q = temp[0], temp[1], temp[2]
            #print(model.plot_diagnostics())

            #model.resid()
            #plot_acf(model.resid(),lags=20)
            #plt.show()
            #plot_pacf(model.resid(),lags=20)
            #plt.show()

            pred = model.predict(test_size)
            Test = pd.DataFrame()
            Test['pred'] = list(pred)
            MAPE_i = mape(list(test['log_psf']), Test['pred'])

            model_p_resale = pm.
↪auto_arima(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
↪District'] == i)]['log_psf'],suppress_warnings=True)
            #model_p = ARIMA(Average_monthly_df[(Average_monthly_df['Postal␣
↪District'] == i)&(Average_monthly_df['Tenure_dummy(>900)'] ==␣
↪j)]['log_psf'],order=(p, d, q))
            pred_p_resale = model_p_resale.predict(36)
            print(model_p_resale.summary())
            print()
            # integrate the results
            districts_2.append(i)
            MAPE_2.append(MAPE_i)
```

```
            percentage_change_2.append(float(100*(pred_p_resale[-1:] -␣
↪float(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
↪District'] == i)]['log_psf'][-1:])))  )
            log_Predicted_2.append(float(pred_p_resale[-1:]))
            log_Currentprice_2.
↪append(float(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
↪District'] == i)]['log_psf'][-1:]))

        else:
            pass
```

```
============================Results of District 1 ============================

                               SARIMAX Results
==========================================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:                 SARIMAX(1, 0, 1)   Log Likelihood                  88.288
Date:                  Sun, 04 Oct 2020   AIC                           -168.577
Time:                          21:02:34   BIC                           -157.394
Sample:                               0   HQIC                          -164.035
                                  - 121
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      2.7472      1.857      1.479      0.139      -0.893       6.388
ar.L1          0.6369      0.245      2.595      0.009       0.156       1.118
ma.L1         -0.3982      0.290     -1.371      0.170      -0.968       0.171
sigma2         0.0135      0.002      8.567      0.000       0.010       0.017
==============================================================================
===
Ljung-Box (Q):                       53.01   Jarque-Bera (JB):
30.06
Prob(Q):                              0.08   Prob(JB):
0.00
Heteroskedasticity (H):               0.54   Skew:
-0.86
Prob(H) (two-sided):                  0.05   Kurtosis:
4.74
==============================================================================
===

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

============================Results of District 2 ============================
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(3, 1, 1)   Log Likelihood                54.358
Date:                Sun, 04 Oct 2020   AIC                            -96.717
Time:                        21:02:42   BIC                            -79.992
Sample:                             0   HQIC                           -89.924
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0035      0.001      3.486      0.000       0.002       0.006
ar.L1          0.0299      0.093      0.323      0.747      -0.152       0.211
ar.L2         -0.1170      0.091     -1.290      0.197      -0.295       0.061
ar.L3         -0.1401      0.088     -1.594      0.111      -0.312       0.032
ma.L1         -0.9515      0.045    -21.313      0.000      -1.039      -0.864
sigma2         0.0232      0.002      9.996      0.000       0.019       0.028
===================================================================================
===
Ljung-Box (Q):                       19.35   Jarque-Bera (JB):
16.67
Prob(Q):                              1.00   Prob(JB):
0.00
Heteroskedasticity (H):               1.44   Skew:
-0.39
Prob(H) (two-sided):                  0.26   Kurtosis:
4.66
===================================================================================
===

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

============================Results of District 3 ============================

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(1, 1, 4)   Log Likelihood                180.344
Date:                Sun, 04 Oct 2020   AIC                           -346.689
Time:                        21:02:49   BIC                           -327.176
Sample:                             0   HQIC                          -338.765
                                - 121
Covariance Type:                  opg
==============================================================================
```

```
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0077      0.005      1.558      0.119      -0.002       0.017
ar.L1         -0.9288      0.075    -12.362      0.000      -1.076      -0.782
ma.L1          0.4376      0.124      3.533      0.000       0.195       0.680
ma.L2         -0.5756      0.101     -5.717      0.000      -0.773      -0.378
ma.L3         -0.0798      0.098     -0.812      0.417      -0.272       0.113
ma.L4          0.1936      0.099      1.965      0.049       0.001       0.387
sigma2         0.0029      0.000      6.701      0.000       0.002       0.004
==============================================================================
===
Ljung-Box (Q):                       34.18   Jarque-Bera (JB):
0.68
Prob(Q):                              0.73   Prob(JB):
0.71
Heteroskedasticity (H):               1.42   Skew:
0.12
Prob(H) (two-sided):                  0.27   Kurtosis:
2.73
==============================================================================
===

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

============================Results of District 4 ============================

                               SARIMAX Results
==============================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:                SARIMAX(2, 0, 2)   Log Likelihood                 125.729
Date:                Sun, 04 Oct 2020   AIC                           -239.457
Time:                        21:02:57   BIC                           -222.682
Sample:                             0   HQIC                          -232.644
                                  - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      1.3114      0.910      1.441      0.150      -0.473       3.095
ar.L1         -0.0481      0.090     -0.533      0.594      -0.225       0.129
ar.L2          0.8690      0.089      9.816      0.000       0.696       1.043
ma.L1          0.2745      0.134      2.044      0.041       0.011       0.538
ma.L2         -0.5633      0.145     -3.877      0.000      -0.848      -0.279
sigma2         0.0073      0.001      8.719      0.000       0.006       0.009
==============================================================================
===
```

```
Ljung-Box (Q):                       32.99   Jarque-Bera (JB):
5.16
Prob(Q):                              0.78   Prob(JB):
0.08
Heteroskedasticity (H):               0.85   Skew:
-0.33
Prob(H) (two-sided):                  0.60   Kurtosis:
3.77
================================================================================
===

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

============================Results of District 5 ============================

                              SARIMAX Results
================================================================================
Dep. Variable:                         y   No. Observations:                  121
Model:                 SARIMAX(0, 1, 2)   Log Likelihood                 200.605
Date:                  Sun, 04 Oct 2020   AIC                           -395.210
Time:                          21:03:00   BIC                           -386.848
Sample:                               0   HQIC                          -391.814
                                  - 121
Covariance Type:                    opg
================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
ma.L1         -0.6894      0.096     -7.146      0.000      -0.879      -0.500
ma.L2          0.1300      0.101      1.293      0.196      -0.067       0.327
sigma2         0.0021      0.000      7.100      0.000       0.001       0.003
================================================================================
===
Ljung-Box (Q):                       36.02   Jarque-Bera (JB):
0.39
Prob(Q):                              0.65   Prob(JB):
0.82
Heteroskedasticity (H):               0.74   Skew:
-0.10
Prob(H) (two-sided):                  0.33   Kurtosis:
2.80
================================================================================
===

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

```
============================Results of District 7 ============================

                               SARIMAX Results
==========================================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:                 SARIMAX(1, 1, 1)   Log Likelihood                22.647
Date:                  Sun, 04 Oct 2020   AIC                            -39.295
Time:                          21:03:04   BIC                            -30.932
Sample:                               0   HQIC                           -35.899
                                  - 121
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3957      0.123      3.220      0.001       0.155       0.637
ma.L1         -0.8126      0.082     -9.931      0.000      -0.973      -0.652
sigma2         0.0400      0.005      8.871      0.000       0.031       0.049
==============================================================================
===
Ljung-Box (Q):                       30.41   Jarque-Bera (JB):
19.88
Prob(Q):                              0.86   Prob(JB):
0.00
Heteroskedasticity (H):               2.61   Skew:
-0.81
Prob(H) (two-sided):                  0.00   Kurtosis:
4.17
==============================================================================
===

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

============================Results of District 8 ============================

                               SARIMAX Results
==========================================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:                 SARIMAX(3, 0, 0)   Log Likelihood                131.300
Date:                  Sun, 04 Oct 2020   AIC                           -252.601
Time:                          21:03:08   BIC                           -238.622
Sample:                               0   HQIC                          -246.924
                                  - 121
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
intercept      4.1012      0.885      4.635      0.000       2.367       5.836
ar.L1          0.1092      0.088      1.247      0.212      -0.062       0.281
ar.L2          0.1451      0.112      1.298      0.194      -0.074       0.364
ar.L3          0.1732      0.076      2.284      0.022       0.025       0.322
sigma2         0.0067      0.001      8.029      0.000       0.005       0.008
==============================================================================
===
Ljung-Box (Q):                       29.44   Jarque-Bera (JB):
3.79
Prob(Q):                              0.89   Prob(JB):
0.15
Heteroskedasticity (H):               1.06   Skew:
-0.37
Prob(H) (two-sided):                  0.86   Kurtosis:
3.46
==============================================================================
===

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

============================Results of District 9 ============================

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 158.255
Date:                Sun, 04 Oct 2020   AIC                           -312.510
Time:                        21:03:11   BIC                           -306.935
Sample:                             0   HQIC                          -310.246
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6214      0.070     -8.855      0.000      -0.759      -0.484
sigma2         0.0042      0.001      7.563      0.000       0.003       0.005
==============================================================================
===
Ljung-Box (Q):                       30.32   Jarque-Bera (JB):
0.43
Prob(Q):                              0.87   Prob(JB):
0.81
Heteroskedasticity (H):               1.44   Skew:
0.14
Prob(H) (two-sided):                  0.25   Kurtosis:
```

```
2.91
==============================================================================
===

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

============================Results of District 10 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(1, 1, 3)   Log Likelihood                 169.907
Date:                Sun, 04 Oct 2020   AIC                           -327.814
Time:                        21:03:17   BIC                           -311.089
Sample:                             0   HQIC                          -321.022
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0026      0.001      3.516      0.000       0.001       0.004
ar.L1         -0.5315      0.310     -1.714      0.087      -1.139       0.076
ma.L1         -0.0960      0.292     -0.328      0.743      -0.669       0.477
ma.L2         -0.5003      0.236     -2.120      0.034      -0.963      -0.038
ma.L3         -0.3654      0.110     -3.332      0.001      -0.580      -0.150
sigma2         0.0034      0.000      7.727      0.000       0.003       0.004
==============================================================================
===
Ljung-Box (Q):                       37.19   Jarque-Bera (JB):
10.56
Prob(Q):                              0.60   Prob(JB):
0.01
Heteroskedasticity (H):               0.82   Skew:
0.30
Prob(H) (two-sided):                  0.55   Kurtosis:
4.32
==============================================================================
===

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

============================Results of District 11 ============================

                              SARIMAX Results
```

```
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 182.739
Date:                Sun, 04 Oct 2020   AIC                           -361.478
Time:                        21:03:19   BIC                           -355.903
Sample:                             0   HQIC                          -359.214
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7035      0.057    -12.381      0.000      -0.815      -0.592
sigma2         0.0028      0.000      9.181      0.000       0.002       0.003
==============================================================================
===
Ljung-Box (Q):                       33.82   Jarque-Bera (JB):
3.32
Prob(Q):                              0.74   Prob(JB):
0.19
Heteroskedasticity (H):               2.14   Skew:
0.10
Prob(H) (two-sided):                  0.02   Kurtosis:
3.79
==============================================================================
===

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

============================Results of District 12 ============================

                           SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(2, 1, 1)   Log Likelihood                 158.495
Date:                Sun, 04 Oct 2020   AIC                           -306.991
Time:                        21:03:22   BIC                           -293.053
Sample:                             0   HQIC                          -301.331
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0056      0.003      1.672      0.095      -0.001       0.012
ar.L1         -0.2699      0.156     -1.730      0.084      -0.576       0.036
ar.L2         -0.3137      0.120     -2.620      0.009      -0.548      -0.079
ma.L1         -0.5278      0.168     -3.146      0.002      -0.857      -0.199
```

```
sigma2         0.0041      0.000     11.515      0.000       0.003       0.005
==============================================================================
===
Ljung-Box (Q):                       18.16   Jarque-Bera (JB):
188.18
Prob(Q):                              1.00   Prob(JB):
0.00
Heteroskedasticity (H):               1.36   Skew:
1.08
Prob(H) (two-sided):                  0.33   Kurtosis:
8.74
==============================================================================
===

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

============================Results of District 13 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                         y   No. Observations:                  121
Model:                 SARIMAX(0, 1, 1)   Log Likelihood                83.834
Date:                  Sun, 04 Oct 2020   AIC                           -161.668
Time:                          21:03:25   BIC                           -153.305
Sample:                               0   HQIC                          -158.272
                                  - 121
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0033      0.002      1.692      0.091      -0.001       0.007
ma.L1         -0.8556      0.055    -15.574      0.000      -0.963      -0.748
sigma2         0.0143      0.001     13.234      0.000       0.012       0.016
==============================================================================
===
Ljung-Box (Q):                       77.39   Jarque-Bera (JB):
241.18
Prob(Q):                              0.00   Prob(JB):
0.00
Heteroskedasticity (H):               0.29   Skew:
-1.41
Prob(H) (two-sided):                  0.00   Kurtosis:
9.34
==============================================================================
===
```

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

============================Results of District 14 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 163.720
Date:                Sun, 04 Oct 2020   AIC                           -321.441
Time:                        21:03:27   BIC                           -313.078
Sample:                             0   HQIC                          -318.045
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0032      0.001      2.236      0.025       0.000       0.006
ma.L1         -0.7698      0.055    -13.966      0.000      -0.878      -0.662
sigma2         0.0038      0.000     10.627      0.000       0.003       0.004
===================================================================================
===
Ljung-Box (Q):                       44.22   Jarque-Bera (JB):
33.82
Prob(Q):                              0.30   Prob(JB):
0.00
Heteroskedasticity (H):               0.66   Skew:
0.57
Prob(H) (two-sided):                  0.19   Kurtosis:
5.34
===================================================================================
===

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

============================Results of District 15 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 209.650
Date:                Sun, 04 Oct 2020   AIC                           -415.300
Time:                        21:03:30   BIC                           -409.725
Sample:                             0   HQIC                          -413.036
                                - 121
```

```
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6344      0.075     -8.416      0.000      -0.782      -0.487
sigma2         0.0018      0.000      9.919      0.000       0.001       0.002
==============================================================================
===
Ljung-Box (Q):                       33.45   Jarque-Bera (JB):
14.42
Prob(Q):                              0.76   Prob(JB):
0.00
Heteroskedasticity (H):               1.46   Skew:
-0.32
Prob(H) (two-sided):                  0.23   Kurtosis:
4.57
==============================================================================
===

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

==============================Results of District 16 ============================

                               SARIMAX Results
==============================================================================
Dep. Variable:                         y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 214.879
Date:                Sun, 04 Oct 2020   AIC                           -425.758
Time:                        21:03:32   BIC                           -420.183
Sample:                             0   HQIC                          -423.494
                                  - 121
Covariance Type:                   opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5302      0.074     -7.163      0.000      -0.675      -0.385
sigma2         0.0016      0.000      6.686      0.000       0.001       0.002
==============================================================================
===
Ljung-Box (Q):                       30.88   Jarque-Bera (JB):
3.36
Prob(Q):                              0.85   Prob(JB):
0.19
Heteroskedasticity (H):               1.71   Skew:
-0.39
Prob(H) (two-sided):                  0.09   Kurtosis:
```

```
2.74
==============================================================================
===

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

============================Results of District 17 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 181.820
Date:                Sun, 04 Oct 2020   AIC                           -359.639
Time:                        21:03:35   BIC                           -354.064
Sample:                             0   HQIC                          -357.375
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6503      0.061    -10.731      0.000      -0.769      -0.532
sigma2         0.0028      0.000     10.068      0.000       0.002       0.003
==============================================================================
===
Ljung-Box (Q):                       41.13   Jarque-Bera (JB):
24.69
Prob(Q):                              0.42   Prob(JB):
0.00
Heteroskedasticity (H):               0.64   Skew:
-0.67
Prob(H) (two-sided):                  0.16   Kurtosis:
4.77
==============================================================================
===

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

============================Results of District 18 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(2, 1, 3)   Log Likelihood                 206.638
Date:                Sun, 04 Oct 2020   AIC                           -399.276
```

```
Time:                        21:03:43   BIC                          -379.763
Sample:                             0   HQIC                         -391.352
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0091      0.006      1.641      0.101      -0.002       0.020
ar.L1         -1.4133      0.260     -5.445      0.000      -1.922      -0.905
ar.L2         -0.6232      0.248     -2.513      0.012      -1.109      -0.137
ma.L1          0.7598      0.313      2.424      0.015       0.145       1.374
ma.L2         -0.2630      0.170     -1.547      0.122      -0.596       0.070
ma.L3         -0.2158      0.219     -0.985      0.325      -0.645       0.214
sigma2         0.0019      0.000      7.425      0.000       0.001       0.002
===================================================================================
===
Ljung-Box (Q):                       36.06   Jarque-Bera (JB):
3.29
Prob(Q):                              0.65   Prob(JB):
0.19
Heteroskedasticity (H):               0.93   Skew:
-0.40
Prob(H) (two-sided):                  0.82   Kurtosis:
3.18
===================================================================================
===

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

============================Results of District 19 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(3, 1, 0)   Log Likelihood                217.363
Date:                Sun, 04 Oct 2020   AIC                         -424.726
Time:                        21:03:49   BIC                         -410.789
Sample:                             0   HQIC                        -419.066
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0060      0.004      1.533      0.125      -0.002       0.014
ar.L1         -0.5396      0.100     -5.406      0.000      -0.735      -0.344
ar.L2         -0.4901      0.108     -4.535      0.000      -0.702      -0.278
```

```
ar.L3         -0.1919      0.101     -1.907      0.056      -0.389       0.005
sigma2         0.0016      0.000      7.402      0.000       0.001       0.002
===================================================================================
===
Ljung-Box (Q):                       37.32   Jarque-Bera (JB):
6.74
Prob(Q):                              0.59   Prob(JB):
0.03
Heteroskedasticity (H):               1.06   Skew:
-0.57
Prob(H) (two-sided):                  0.85   Kurtosis:
3.20
===================================================================================
===

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

==============================Results of District 20 ============================

                              SARIMAX Results
==================================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 176.279
Date:                Sun, 04 Oct 2020   AIC                           -346.558
Time:                        21:03:51   BIC                           -338.195
Sample:                             0   HQIC                          -343.162
                                  - 121
Covariance Type:                  opg
==================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
intercept      0.0041      0.002      2.094      0.036       0.000       0.008
ma.L1         -0.6240      0.069     -9.096      0.000      -0.758      -0.490
sigma2         0.0031      0.000      8.621      0.000       0.002       0.004
===================================================================================
===
Ljung-Box (Q):                       39.58   Jarque-Bera (JB):
1.78
Prob(Q):                              0.49   Prob(JB):
0.41
Heteroskedasticity (H):               1.66   Skew:
0.06
Prob(H) (two-sided):                  0.11   Kurtosis:
3.58
===================================================================================
===
```

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

============================Results of District 21 ============================

                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 182.384
Date:                Sun, 04 Oct 2020   AIC                          -360.768
Time:                        21:03:56   BIC                          -355.193
Sample:                             0   HQIC                         -358.504
                                - 121
Covariance Type:                  opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
ma.L1         -0.5246      0.077     -6.855      0.000      -0.675      -0.375
sigma2         0.0028      0.000      8.328      0.000       0.002       0.003
==========================================================================
===
Ljung-Box (Q):                       40.23   Jarque-Bera (JB):
2.28
Prob(Q):                              0.46   Prob(JB):
0.32
Heteroskedasticity (H):               0.56   Skew:
-0.18
Prob(H) (two-sided):                  0.07   Kurtosis:
3.57
==========================================================================
===

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

============================Results of District 22 ============================

                              SARIMAX Results
==========================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 150.501
Date:                Sun, 04 Oct 2020   AIC                          -297.001
Time:                        21:03:58   BIC                          -291.426
Sample:                             0   HQIC                         -294.737
                                - 121
```

```
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6206      0.077     -8.028      0.000      -0.772      -0.469
sigma2         0.0047      0.001      7.834      0.000       0.004       0.006
==============================================================================
===
Ljung-Box (Q):                       47.10   Jarque-Bera (JB):
0.72
Prob(Q):                              0.20   Prob(JB):
0.70
Heteroskedasticity (H):               1.38   Skew:
-0.18
Prob(H) (two-sided):                  0.31   Kurtosis:
3.13
==============================================================================
===

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

============================Results of District 23 ============================

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(0, 1, 2)   Log Likelihood                 230.803
Date:                Sun, 04 Oct 2020   AIC                           -453.606
Time:                        21:04:01   BIC                           -442.456
Sample:                             0   HQIC                          -449.078
                                  - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0027      0.002      1.560      0.119      -0.001       0.006
ma.L1         -0.6725      0.118     -5.713      0.000      -0.903      -0.442
ma.L2          0.1966      0.090      2.196      0.028       0.021       0.372
sigma2         0.0012      0.000     11.311      0.000       0.001       0.001
==============================================================================
===
Ljung-Box (Q):                       21.77   Jarque-Bera (JB):
29.78
Prob(Q):                              0.99   Prob(JB):
0.00
Heteroskedasticity (H):               2.62   Skew:
```

```
-0.29
Prob(H) (two-sided):                  0.00   Kurtosis:
5.37
================================================================================
===

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

============================Results of District 25 ============================

                              SARIMAX Results
================================================================================
Dep. Variable:                        y   No. Observations:                  121
Model:                SARIMAX(0, 1, 3)   Log Likelihood                 115.271
Date:                 Sun, 04 Oct 2020   AIC                           -222.541
Time:                         21:04:08   BIC                           -211.391
Sample:                              0   HQIC                          -218.013
                                 - 121
Covariance Type:                   opg
================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
ma.L1         -0.5378      0.092     -5.819      0.000      -0.719      -0.357
ma.L2         -0.4150      0.105     -3.970      0.000      -0.620      -0.210
ma.L3          0.2110      0.086      2.455      0.014       0.043       0.380
sigma2         0.0085      0.001      9.614      0.000       0.007       0.010
================================================================================
===
Ljung-Box (Q):                       38.12   Jarque-Bera (JB):
9.78
Prob(Q):                              0.56   Prob(JB):
0.01
Heteroskedasticity (H):               3.24   Skew:
-0.28
Prob(H) (two-sided):                  0.00   Kurtosis:
4.29
================================================================================
===

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

============================Results of District 26 ============================

                              SARIMAX Results
```

```
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:                SARIMAX(0, 1, 2)   Log Likelihood                 112.023
Date:                Sun, 04 Oct 2020   AIC                           -218.045
Time:                        21:04:13   BIC                           -209.683
Sample:                             0   HQIC                          -214.649
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6355      0.082     -7.704      0.000      -0.797      -0.474
ma.L2         -0.1500      0.105     -1.430      0.153      -0.355       0.056
sigma2         0.0090      0.001      8.519      0.000       0.007       0.011
==============================================================================
===
Ljung-Box (Q):                       38.72   Jarque-Bera (JB):
8.42
Prob(Q):                              0.53   Prob(JB):
0.01
Heteroskedasticity (H):               3.00   Skew:
-0.11
Prob(H) (two-sided):                  0.00   Kurtosis:
4.28
==============================================================================
===

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

============================Results of District 27 ============================

                           SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:                SARIMAX(2, 1, 1)   Log Likelihood                 180.000
Date:                Sun, 04 Oct 2020   AIC                           -350.000
Time:                        21:04:17   BIC                           -336.063
Sample:                             0   HQIC                          -344.340
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0040      0.003      1.606      0.108      -0.001       0.009
ar.L1         -0.3712      0.172     -2.159      0.031      -0.708      -0.034
ar.L2         -0.2266      0.141     -1.611      0.107      -0.502       0.049
```

```
ma.L1         -0.5283      0.142     -3.727      0.000      -0.806      -0.250
sigma2         0.0029      0.000      7.183      0.000       0.002       0.004
==============================================================================
===
Ljung-Box (Q):                       44.10   Jarque-Bera (JB):
1.42
Prob(Q):                              0.30   Prob(JB):
0.49
Heteroskedasticity (H):               1.60   Skew:
-0.26
Prob(H) (two-sided):                  0.14   Kurtosis:
3.11
==============================================================================
===

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

============================Results of District 28 ============================

                              SARIMAX Results
==============================================================================
Dep. Variable:                          y   No. Observations:                  121
Model:               SARIMAX(0, 1, 1)   Log Likelihood                 139.515
Date:                Sun, 04 Oct 2020   AIC                           -273.030
Time:                        21:04:19   BIC                           -264.667
Sample:                             0   HQIC                          -269.634
                                - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0039      0.002      1.947      0.052   -2.69e-05       0.008
ma.L1         -0.7289      0.062    -11.735      0.000      -0.851      -0.607
sigma2         0.0057      0.001      8.470      0.000       0.004       0.007
==============================================================================
===
Ljung-Box (Q):                       44.93   Jarque-Bera (JB):
7.36
Prob(Q):                              0.27   Prob(JB):
0.03
Heteroskedasticity (H):               0.98   Skew:
-0.40
Prob(H) (two-sided):                  0.96   Kurtosis:
3.91
==============================================================================
===
```

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

[27]:
```
Result_resale = pd.DataFrame({'district': districts_2, 'MAPE_resale': MAPE_2,␣
↪'log_Predicted': log_Predicted_2, 'log_Currentprice': log_Currentprice_2})
Result_resale['Predicted'] = np.exp(Result_resale['log_Predicted'])
Result_resale['Currentprice'] = np.exp(Result_resale['log_Currentprice'])
Result_resale['percentage_increase'] = Result_resale['Predicted'] /␣
↪Result_resale['Currentprice'] - 1
```

[28]:
```
Result_resale.sort_values('percentage_increase',ascending=False)
```

[28]:

	district	MAPE_resale	log_Predicted	log_Currentprice	Predicted	\
1	2	2.356959	7.734650	7.389255	2286.208453	
25	28	0.816297	7.110641	6.900328	1224.933081	
17	19	0.510549	7.062730	6.892286	1167.628610	
16	18	1.013939	6.939979	6.775366	1032.748821	
18	20	0.518624	7.305020	7.151834	1487.749676	
21	23	0.617092	6.968269	6.816325	1062.382141	
24	27	0.747008	6.850653	6.718107	944.497907	
12	14	1.036005	7.197147	7.079858	1335.615194	
13	15	0.677026	7.122803	7.019535	1239.921091	
10	12	0.981370	7.363735	7.279112	1577.718035	
2	3	0.848960	7.620818	7.550661	2040.229982	
7	9	0.740020	7.590645	7.522980	1979.589485	
22	25	1.655379	6.712010	6.646391	822.221859	
8	10	0.593739	7.642121	7.580351	2084.159907	
3	4	0.803777	7.318726	7.262453	1508.280872	
23	26	1.689026	6.901489	6.851714	993.752897	
11	13	0.898123	7.353787	7.305994	1562.101342	
4	5	0.893910	7.071366	7.037028	1177.755859	
9	11	0.602799	7.363607	7.332762	1577.517042	
14	16	0.855107	6.971146	6.940595	1065.443383	
19	21	1.041617	7.100658	7.090077	1212.764806	
15	17	0.666870	6.775334	6.770933	875.971745	
20	22	0.861024	6.930941	6.930495	1023.456400	
5	7	2.840476	7.504553	7.597522	1816.292676	
6	8	1.317378	7.163555	7.296413	1291.493865	
0	1	1.771937	7.565580	7.715792	1930.587337	

	Currentprice	percentage_increase
1	1618.500000	0.412548
25	992.600000	0.234065
17	984.650000	0.185831

```
16    876.000000             0.178937
18   1276.444444             0.165542
21    912.625000             0.164095
24    827.250000             0.141732
12   1187.800000             0.124445
13   1118.266667             0.108788
10   1449.700000             0.088307
2    1902.000000             0.072676
7    1850.071429             0.070007
22    770.000000             0.067821
8    1959.315789             0.063718
3    1425.750000             0.057886
23    945.500000             0.051034
11   1489.200000             0.048953
4    1138.000000             0.034935
9    1529.600000             0.031327
14   1033.384615             0.031023
19   1200.000000             0.010637
15    872.125000             0.004411
20   1023.000000             0.000446
5    1993.250000            -0.088778
6    1475.000000            -0.124411
0    2243.500000            -0.139475
```

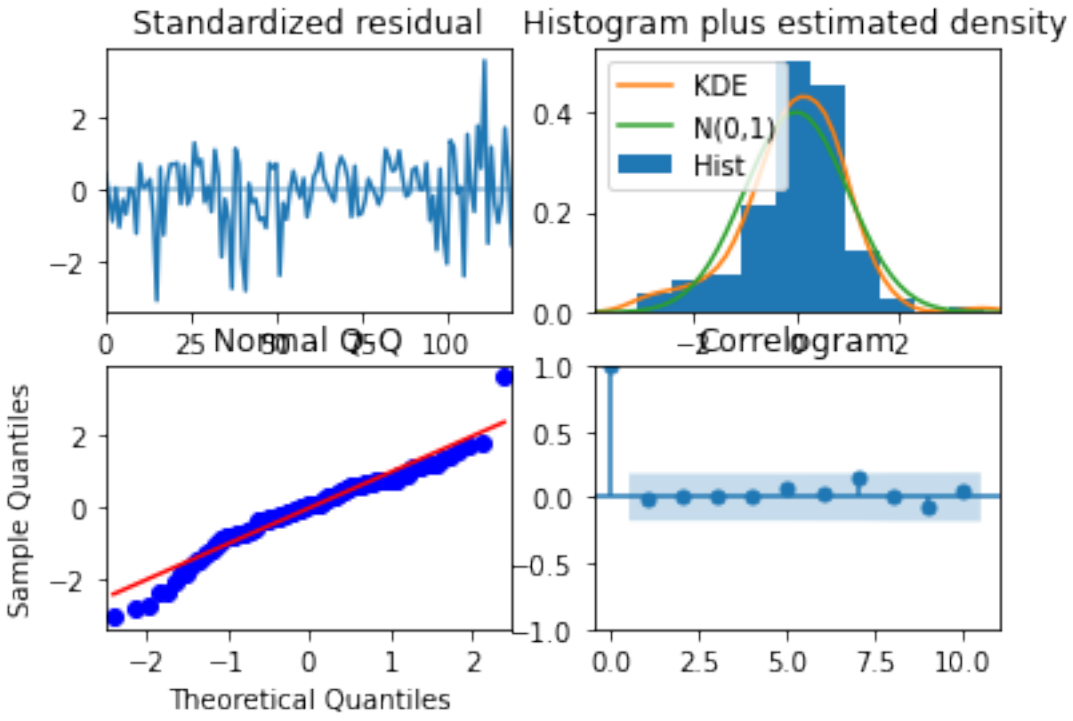
0.0.4 Distrcit 2_Resale gives the best result, now focus on District 2_Resale property

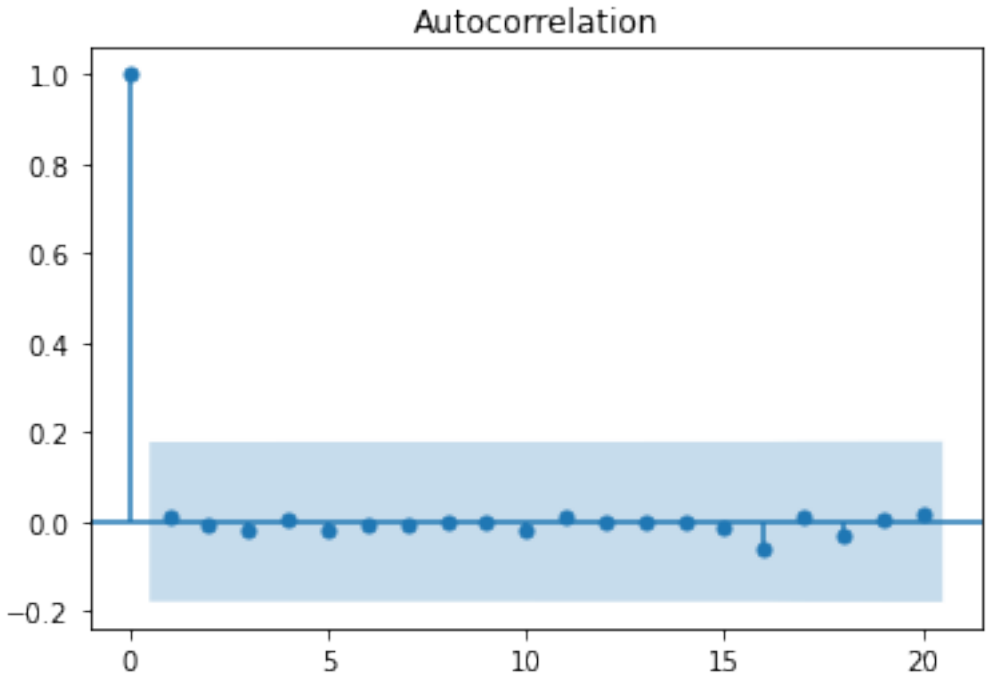
[29]:
```
model_p_resale_2 = pm.
 ↪auto_arima(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
 ↪District'] == 2)]['log_psf'],suppress_warnings=True)
print(model_p_resale_2.summary())
print(model_p_resale_2.plot_diagnostics())
# To check residual and any ARMA effect left
model_p_resale_2.resid()
plot_acf(model_p_resale_2.resid(),lags=20)
plt.show()
plot_pacf(model_p_resale_2.resid(),lags=20)
plt.show()
```

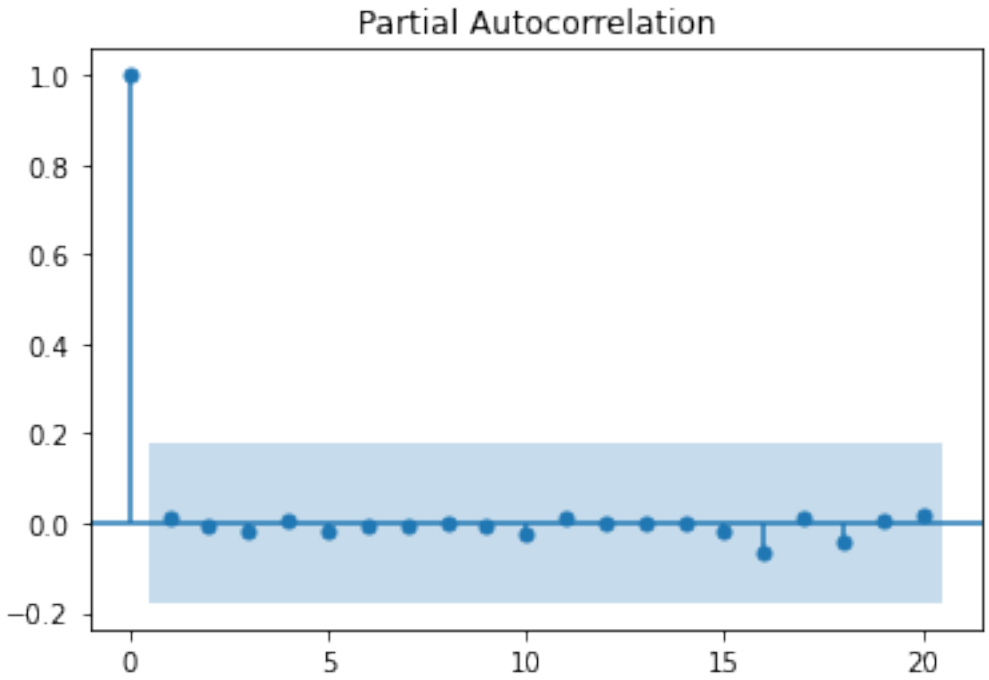
```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(3, 1, 1)   Log Likelihood                54.358
Date:                Sun, 04 Oct 2020   AIC                            -96.717
Time:                        21:04:22   BIC                            -79.992
Sample:                             0   HQIC                           -89.924
                                  - 121
Covariance Type:                  opg
```

```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0035      0.001      3.486      0.000       0.002       0.006
ar.L1          0.0299      0.093      0.323      0.747      -0.152       0.211
ar.L2         -0.1170      0.091     -1.290      0.197      -0.295       0.061
ar.L3         -0.1401      0.088     -1.594      0.111      -0.312       0.032
ma.L1         -0.9515      0.045    -21.313      0.000      -1.039      -0.864
sigma2         0.0232      0.002      9.996      0.000       0.019       0.028
==============================================================================
===
Ljung-Box (Q):                       19.35   Jarque-Bera (JB):
16.67
Prob(Q):                              1.00   Prob(JB):
0.00
Heteroskedasticity (H):               1.44   Skew:
-0.39
Prob(H) (two-sided):                  0.26   Kurtosis:
4.66
==============================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).
Figure(432x288)
```




Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20

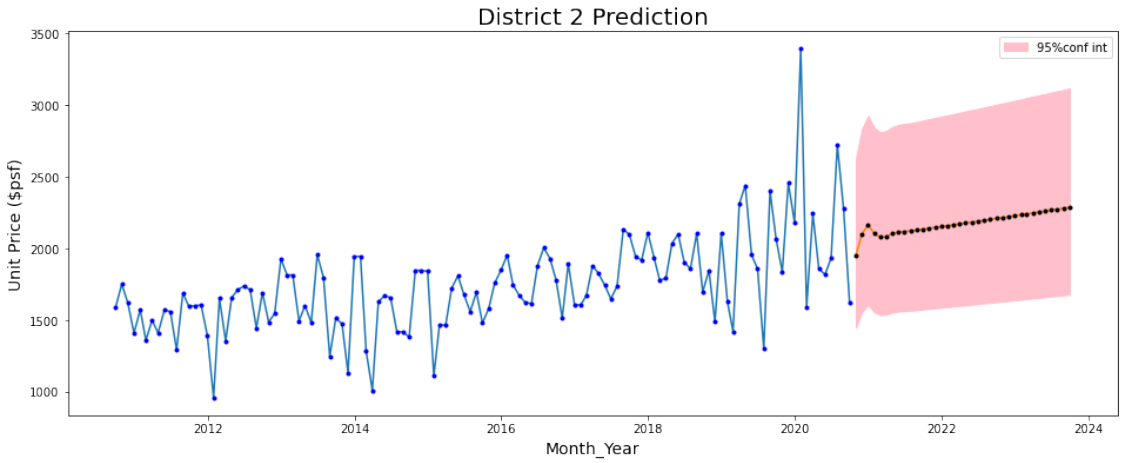

Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20

[30]:
```
# Plot line chart for District 2_Resale property with historical and predict
model_p_resale_2 = pm.
 ↪auto_arima(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
 ↪District'] == 2)]['log_psf'],suppress_warnings=True)
pred_p_resale_2, conf = model_p_resale_2.predict(36,␣
 ↪return_conf_int=True,alpha=0.05)
print(pred_p_resale_2)
print(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
 ↪District'] == 2)]['log_psf'][-1:])
```

```
[7.57538216 7.64931942 7.68107367 7.65081729 7.63936205 7.64163551
 7.65080584 7.65594158 7.65822633 7.6599319  7.66251906 7.66559973
 7.66867312 7.67156502 7.67438321 7.67722147 7.68009437 7.68297629
 7.68585161 7.68872082 7.69158936 7.69445952 7.69733067 7.70020174
 7.70307248 7.70594307 7.70881371 7.71168441 7.71455513 7.71742584
 7.72029654 7.72316723 7.72603793 7.72890863 7.73177933 7.73465003]
3121    7.389255
Name: log_psf, dtype: float64
```

[31]:
```
Test = pd.DataFrame()
Test['test_periods'] = pd.Series(pd.date_range('2020-10-01','2023-10-01', freq␣
 ↪= 'M'))
Test['lower_bounds'] = [i[0] for i in conf ]
Test['upper_bounds'] = [i[1] for i in conf ]
Test['prediction'] = list(pred_p_resale_2)

fig, ax1 = plt.subplots(figsize=(16, 6))
ax1.plot_date( pd.Series(pd.date_range('2010-09-01','2020-10-01', freq = 'M')),␣
 ↪Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
 ↪District'] == 2)]['Unit Price ($ psf)'],linestyle = 'solid',markeredgecolor␣
 ↪= 'blue',markerfacecolor = 'blue', markersize = 3)
ax1.plot_date( Test['test_periods'], np.exp(Test['prediction']), linestyle =␣
 ↪'solid',markeredgecolor = 'black',markerfacecolor = 'black', markersize = 3)
plt.fill_between(Test['test_periods'], np.exp(Test['lower_bounds']), np.
 ↪exp(Test['upper_bounds']), color='pink', label='95%conf int')
ax1.set_xlabel('Month_Year',fontsize=14)
ax1.set_ylabel('Unit Price ($psf)',fontsize=14)
plt.title('District 2 Prediction', size=20)
plt.legend()
plt.show()
```



0.0.5 We also want to know the transaction number in District 2 by full year

[32]:
```
data['Year'] = pd.to_datetime(data['Sale Date']).dt.to_period('Y')
```

[33]:
```
data['Year'].unique()
```

[33]:
```
<PeriodArray>
['2012', '2011', '2010', '2015', '2014', '2013', '2020', '2019', '2018',
 '2017', '2016']
Length: 11, dtype: period[A-DEC]
```

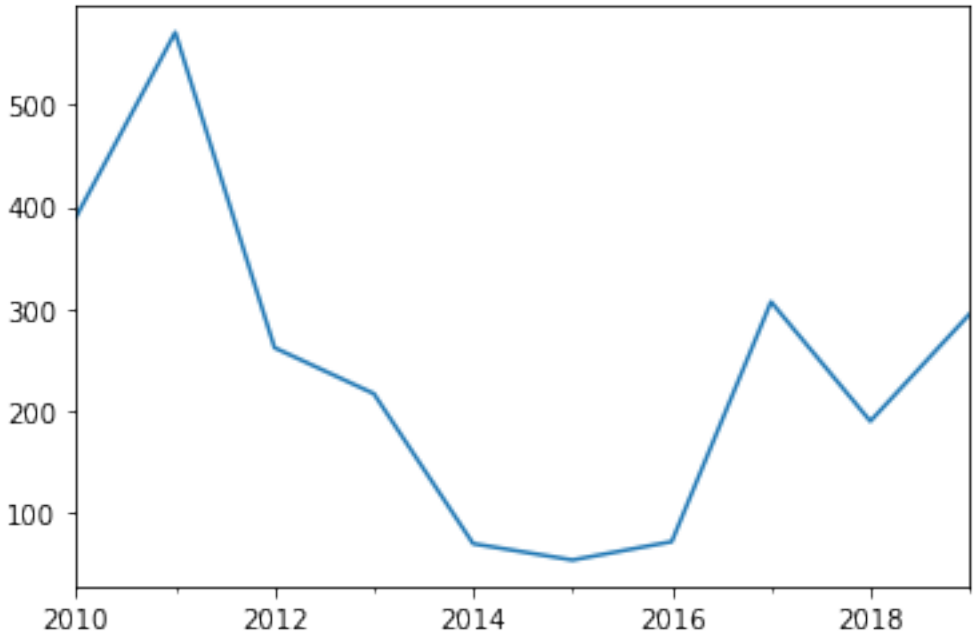
[34]:
```
#We only want District 2
district2 = data[data['Postal District'] == 2]
```

[35]:
```
#We do not need 2020 since it is not full year
district2 = district2[district2['Year'] != '2020']
```

[36]:
```
d2_by_year = district2['Year'].value_counts().sort_index()
```

[37]:
```
d2_by_year.plot()
```

[37]:
```
<AxesSubplot:>
```



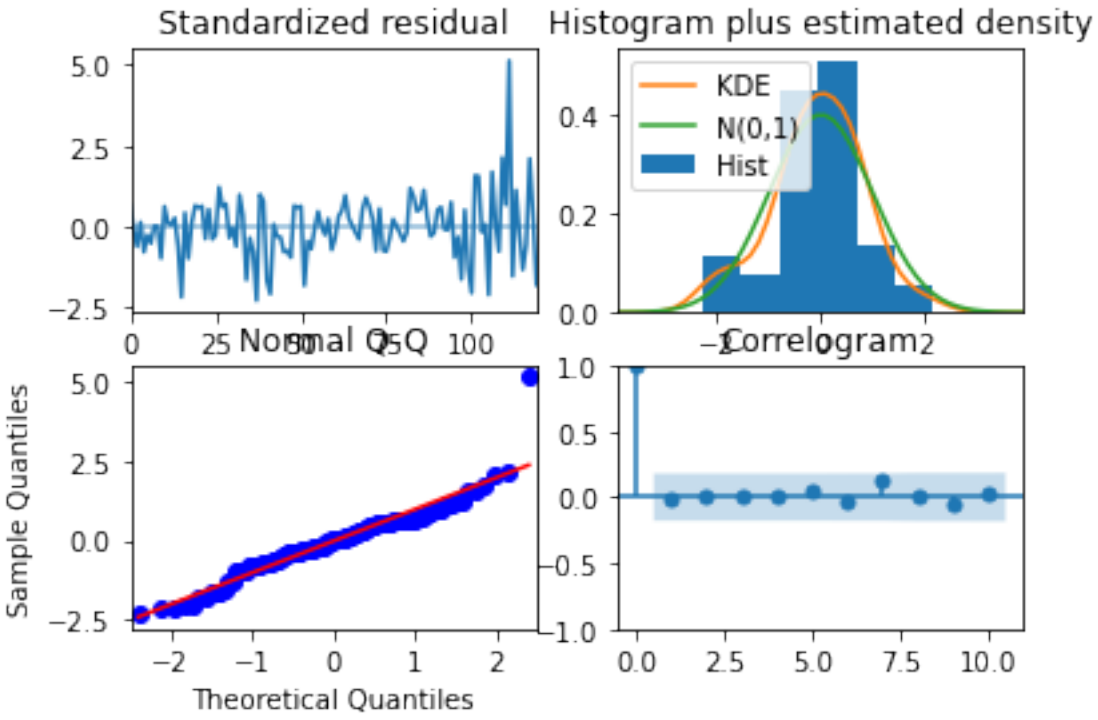
One thing more: try to run Arima model using the absolute price, not the log price

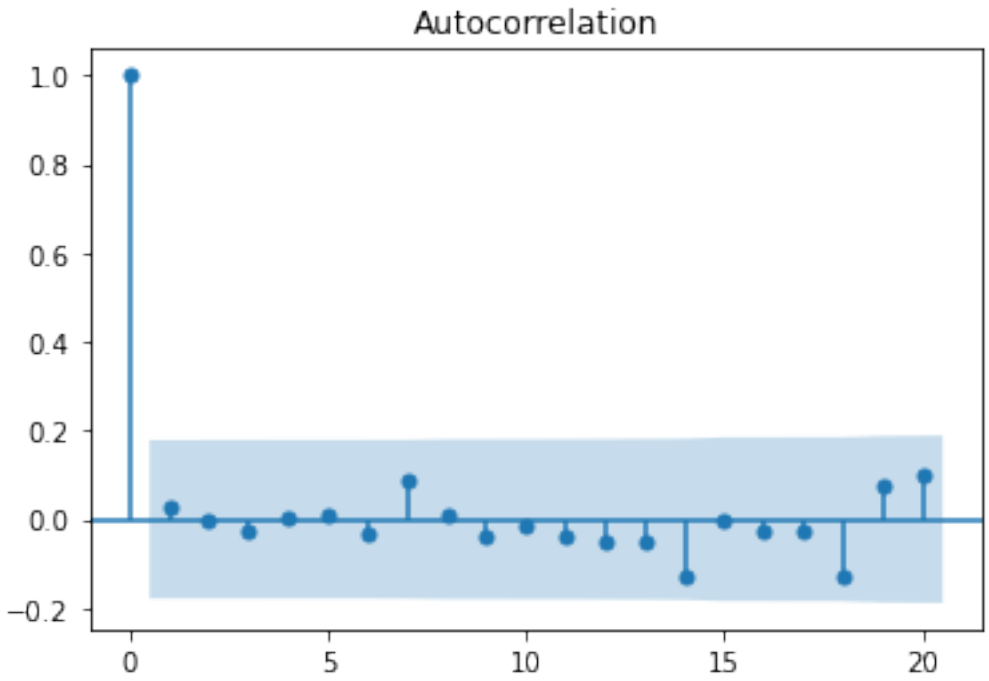
[47]:
```
model_p_resale_v2 = pm.
 ↪auto_arima(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
 ↪District'] == 2)]['Unit Price ($ psf)'],suppress_warnings=True)
print(model_p_resale_v2.summary())
print(model_p_resale_v2.plot_diagnostics())
# To check residual and any ARMA effect left
model_p_resale_v2.resid()
plot_acf(model_p_resale_v2.resid(),lags=20)
plt.show()
plot_pacf(model_p_resale_v2.resid(),lags=20)
plt.show()
```

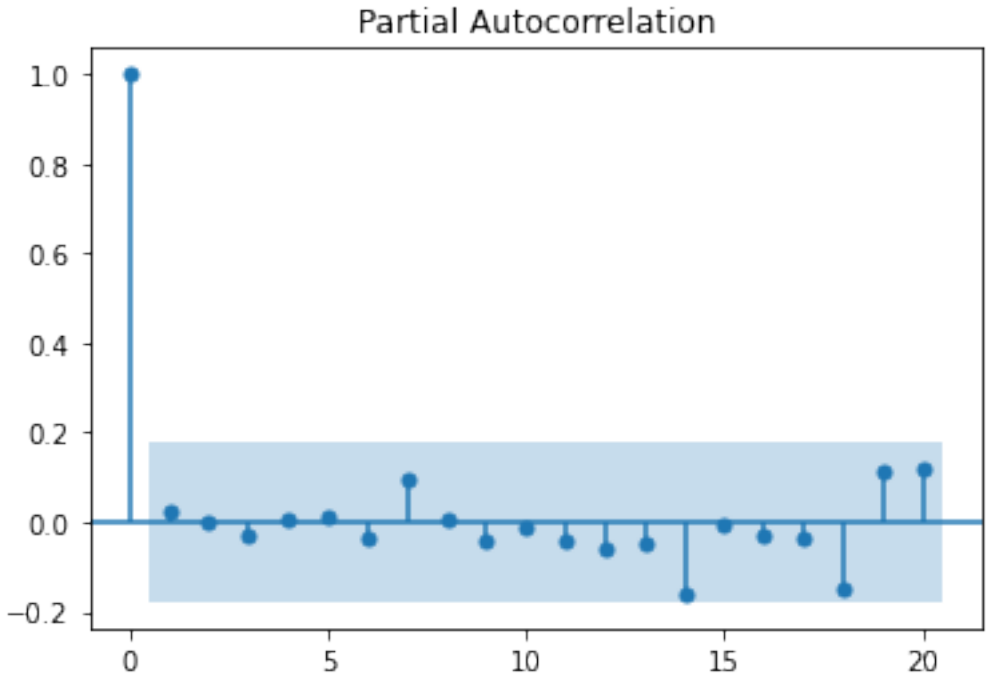
```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  121
Model:               SARIMAX(3, 1, 1)   Log Likelihood                -843.597
Date:                Sun, 04 Oct 2020   AIC                           1699.194
Time:                        21:10:13   BIC                           1715.919
Sample:                             0   HQIC                          1705.986
                                  - 121
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
intercept      6.8797      2.224      3.094      0.002       2.521      11.238
ar.L1         -0.0101      0.091     -0.111      0.912      -0.189       0.169
ar.L2         -0.0958      0.091     -1.058      0.290      -0.273       0.082
ar.L3         -0.1766      0.092     -1.916      0.055      -0.357       0.004
ma.L1         -0.9326      0.049    -19.142      0.000      -1.028      -0.837
sigma2      7.371e+04   6158.695     11.969      0.000    6.16e+04    8.58e+04
===================================================================================
===
Ljung-Box (Q):                       20.43   Jarque-Bera (JB):
136.74
Prob(Q):                              1.00   Prob(JB):
0.00
Heteroskedasticity (H):               2.88   Skew:
0.84
Prob(H) (two-sided):                  0.00   Kurtosis:
7.96
===================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).
Figure(432x288)
```







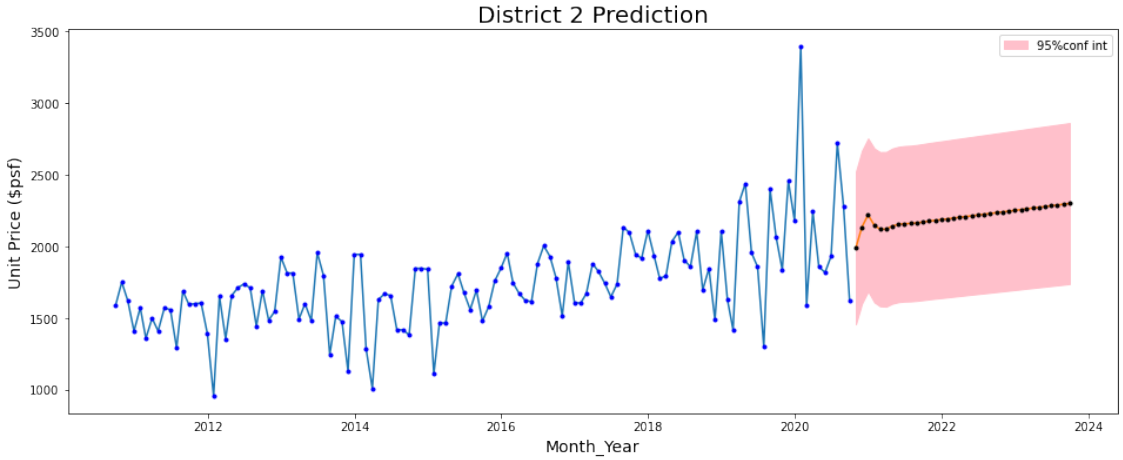
[48]:
```
pred_p_resale_v2, conf_v2 = model_p_resale_v2.predict(36,␣
↪return_conf_int=True,alpha=0.05)
```

```
print(pred_p_resale_v2)
print(Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
 ↪District'] == 2)]['Unit Price ($ psf)'][-1:])
```

```
[1989.289883   2133.77584362 2219.91546809 2146.58871294 2120.43539384
 2119.39354192 2141.74103869 2153.11370478 2157.92056228 2159.71470207
 2164.10691945 2169.92124248 2176.00427901 2181.48943784 2186.70372032
 2191.93058387 2197.28887662 2202.69247598 2208.08079884 2213.44172131
 2218.79638304 2224.15643227 2229.52186642 2234.88783553 2240.25233161
 2245.61584024 2250.97940555 2256.34332509 2261.70741002 2267.07144931
 2272.43541066 2277.79934795 2283.16330102 2288.52727    2293.89124155
 2299.25520877]
3121    1618.5
Name: Unit Price ($ psf), dtype: float64
```

[51]:
```
Test = pd.DataFrame()
Test['test_periods'] = pd.Series(pd.date_range('2020-10-01','2023-10-01', freq␣
 ↪= 'M'))
Test['lower_bounds_2'] = [i[0] for i in conf_v2 ]
Test['upper_bounds_2'] = [i[1] for i in conf_v2 ]
Test['prediction_2'] = list(pred_p_resale_v2)

fig, ax1 = plt.subplots(figsize=(16, 6))
ax1.plot_date( pd.Series(pd.date_range('2010-09-01','2020-10-01', freq = 'M')),␣
 ↪Average_monthly_df_filled_resale[(Average_monthly_df_filled_resale['Postal␣
 ↪District'] == 2)]['Unit Price ($ psf)'],linestyle = 'solid',markeredgecolor␣
 ↪= 'blue',markerfacecolor = 'blue', markersize = 3)
ax1.plot_date( Test['test_periods'], Test['prediction_2'], linestyle =␣
 ↪'solid',markeredgecolor = 'black',markerfacecolor = 'black', markersize = 3)
plt.fill_between(Test['test_periods'],␣
 ↪Test['lower_bounds_2'],Test['upper_bounds_2'], color='pink', label='95%conf␣
 ↪int')
ax1.set_xlabel('Month_Year',fontsize=14)
ax1.set_ylabel('Unit Price ($psf)',fontsize=14)
plt.title('District 2 Prediction', size=20)
plt.legend()
plt.show()
```


District 2 Prediction
95%conf int
Unit Price ($psf)
Month_Year
3500
3000
2500
2000
1500
1000
2012
2014
2016
2018
2020
2022
2024

[ ]: