In [43]:
```
import statsmodels.api as sm
import statsmodels.formula.api as smf
import pandas as pd
import numpy as np
```

In [44]:
```
civic = pd.read_csv("Civic-142A-Fall24.csv")
```

In [45]:
```
train = civic[civic["Year"] < 2020]
test = civic[civic["Year"] >= 2020]
train.head()
```

Out[45]:

	MonthNumeric	MonthFactor	Year	CivicSales	Unemployment	CivicQueries	CPIAll	C
0	1	January	2014	21824	6.6	66	235.288	
1	2	February	2014	21575	6.7	69	235.547	
2	3	March	2014	27697	6.7	72	236.028	
3	4	April	2014	27611	6.2	69	236.468	
4	5	May	2014	36089	6.3	69	236.918	

In [46]:
```
#list of vars to subset
xVars = ["CivicQueries","CPIAll"]

# filter taining and test into independent vars list
Xtrain = train[xVars]
Xtest = test[xVars]

#create target vectors to predict Sales
Ytrain = train[["CivicSales"]]
Ytest = test[["CivicSales"]]

#add Constants
Xtrain = sm.add_constant(Xtrain)
Xtest = sm.add_constant(Xtest)
```

In [47]:
```
model = sm.OLS(Ytrain, Xtrain)
```

In [48]:
```
results = model.fit()
```

In [49]:
```
print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             CivicSales   R-squared:                       0.395
Model:                            OLS   Adj. R-squared:                  0.377
Method:                 Least Squares   F-statistic:                     22.51
Date:                Mon, 27 Jan 2025   Prob (F-statistic):           2.98e-08
Time:                        21:26:57   Log-Likelihood:                -689.88
No. Observations:                  72   AIC:                             1386.
Df Residuals:                      69   BIC:                             1393.
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const         5.713e+04   1.45e+04      3.953      0.000    2.83e+04     8.6e+04
CivicQueries   359.5442     53.958      6.663      0.000     251.901     467.187
CPIAll        -230.2333     64.419     -3.574      0.001    -358.745    -101.722
==============================================================================
Omnibus:                        2.257   Durbin-Watson:                   1.622
Prob(Omnibus):                  0.323   Jarque-Bera (JB):                1.780
Skew:                           0.224   Prob(JB):                        0.411
Kurtosis:                       2.374   Cond. No.                     8.77e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 8.77e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [50]:
```
test.head(5)
```

Out[50]:

	MonthNumeric	MonthFactor	Year	CivicSales	Unemployment	CivicQueries	CPIAll
72	1	January	2020	20054	3.5	74	259.037
73	2	February	2020	25617	3.5	76	259.248
74	3	March	2020	18273	4.4	66	258.124
75	4	April	2020	13410	14.8	57	256.092
76	5	May	2020	27244	13.3	78	255.868

In [51]:
```
modelMonths = smf.ols("CivicSales ~ CivicQueries + CPIAll + CPIEnergy + Unemploymen
```

In [52]:
```
results2 = modelMonths.fit()
```

In [53]:
```
print(results2.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             CivicSales   R-squared:                       0.765
Model:                            OLS   Adj. R-squared:                  0.697
Method:                 Least Squares   F-statistic:                     11.19
Date:                Mon, 27 Jan 2025   Prob (F-statistic):           5.05e-12
Time:                        21:26:57   Log-Likelihood:                -655.83
No. Observations:                  72   AIC:                             1346.
Df Residuals:                      55   BIC:                             1384.
Df Model:                          16
Covariance Type:            nonrobust
=====================================================================================
===========
                                   coef    std err          t      P>|t|      [0.025
0.975]
-------------------------------------------------------------------------------------
-----------
Intercept                     2.597e+04   5.38e+04      0.482      0.631   -8.19e+04
1.34e+05
C(MonthFactor)[T.August]      2810.6266   1501.365      1.872      0.067    -198.177
5819.430
C(MonthFactor)[T.December]    2023.6110   1729.147      1.170      0.247   -1441.678
5488.900
C(MonthFactor)[T.February]   -4350.5616   1509.515     -2.882      0.006   -7375.697
-1325.426
C(MonthFactor)[T.January]    -5543.3141   1508.810     -3.674      0.001   -8567.038
-2519.591
C(MonthFactor)[T.July]         796.0087   1510.946      0.527      0.600   -2231.995
3824.012
C(MonthFactor)[T.June]         782.3073   1535.117      0.510      0.612   -2294.135
3858.750
C(MonthFactor)[T.March]       1043.1405   1521.718      0.686      0.496   -2006.451
4092.732
C(MonthFactor)[T.May]         4311.4869   1503.319      2.868      0.006    1298.768
7324.205
C(MonthFactor)[T.November]   -1676.6088   1606.548     -1.044      0.301   -4896.202
1542.985
C(MonthFactor)[T.October]    -2192.4334   1513.577     -1.449      0.153   -5225.709
840.842
C(MonthFactor)[T.September] -1153.0328   1517.026     -0.760      0.450   -4193.221
1887.155
CivicQueries                   257.9525     72.508      3.558      0.001     112.642
403.263
CPIAll                        -363.9977    207.430     -1.755      0.085    -779.697
51.702
CPIEnergy                       27.4736     38.325      0.717      0.476     -49.331
104.278
Unemployment                   467.6756   1741.508      0.269      0.789   -3022.385
3957.737
MilesTraveled                    0.2420      0.164      1.472      0.147      -0.087
0.571
==============================================================================
Omnibus:                        5.209   Durbin-Watson:                   1.308
Prob(Omnibus):                  0.074   Jarque-Bera (JB):                4.456
Skew:                           0.474   Prob(JB):                        0.108
Kurtosis:                       3.765   Cond. No.                     4.83e+07
```

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

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
cified.
[2] The condition number is large, 4.83e+07. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [54]:
```
model3 = smf.ols("CivicSales ~ CivicQueries + CPIAll + C(MonthFactor)", data = trai
```

In [55]:
```
print(model3.summary())
```

OLS Regression Results

Dep. Variable:	CivicSales	R-squared:	0.756
Model:	OLS	Adj. R-squared:	0.701
Method:	Least Squares	F-statistic:	13.79
Date:	Mon, 27 Jan 2025	Prob (F-statistic):	2.80e-13
Time:	21:26:57	Log-Likelihood:	-657.24
No. Observations:	72	AIC:	1342.
Df Residuals:	58	BIC:	1374.
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.078e+04	1.04e+04	5.838	0.000	3.99e+04	8.16e+04
C(MonthFactor)[T.August]	2297.4909	1441.994	1.593	0.117	-588.974	5183.956
C(MonthFactor)[T.December]	2255.3858	1559.828	1.446	0.154	-866.950	5377.722
C(MonthFactor)[T.February]	-4944.7181	1438.721	-3.437	0.001	-7824.632	-2064.804
C(MonthFactor)[T.January]	-5906.4114	1455.679	-4.057	0.000	-8820.269	-2992.554
C(MonthFactor)[T.July]	245.1886	1443.417	0.170	0.866	-2644.125	3134.502
C(MonthFactor)[T.June]	100.0323	1441.306	0.069	0.945	-2785.056	2985.120
C(MonthFactor)[T.March]	374.8885	1437.445	0.261	0.795	-2502.471	3252.248
C(MonthFactor)[T.May]	3743.5754	1437.176	2.605	0.012	866.755	6620.396
C(MonthFactor)[T.November]	-2158.3494	1500.261	-1.439	0.156	-5161.449	844.750
C(MonthFactor)[T.October]	-2545.2858	1456.460	-1.748	0.086	-5460.708	370.137
C(MonthFactor)[T.September]	-1581.3890	1452.733	-1.089	0.281	-4489.350	1326.572
CivicQueries	289.3001	48.340	5.985	0.000	192.536	386.064
CPIAll	-220.2553	49.162	-4.480	0.000	-318.663	-121.848

Omnibus:	5.392	Durbin-Watson:	1.293
Prob(Omnibus):	0.067	Jarque-Bera (JB):	4.627
Skew:	0.493	Prob(JB):	0.0989
Kurtosis:	3.755	Cond. No.	9.12e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
[2] The condition number is large, 9.12e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [56]:
```
test.reset_index(drop=True, inplace=True)
preds = model3.predict(test)
```

In [57]:
```
SSR = np.sum((test["CivicSales"] - preds) ** 2)
y_mean = np.mean(test["CivicSales"])
```

In [58]:
```
TSS = np.sum((test["CivicSales"] - y_mean) ** 2)
```

In [59]:
```
OSR2 = 1 - (SSR/TSS)
```

In [60]:
```
OSR2
```

Out[60]:
```
-0.09021353946711175
```

In [61]:
```
CCI = pd.read_csv("UMCSENT (1).csv")
```

In [70]:
```
civic2 = pd.concat([civic, CCI["UMCSENT"]],axis = 1)
trainCCI = civic2[civic2["Year"] < 2020]
testCCI = civic2[civic2["Year"] >= 2020]
CCI.head()
```

Out[70]:

	DATE	UMCSENT
0	2014-01-01	81.2
1	2014-02-01	81.6
2	2014-03-01	80.0
3	2014-04-01	84.1
4	2014-05-01	81.9

In [63]:
```
model4 = smf.ols("CivicSales ~ UMCSENT + C(MonthFactor) + CivicQueries + CPIAll", d
```

In [64]:
```
print(model4.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            CivicSales   R-squared:                       0.757
Model:                           OLS   Adj. R-squared:                  0.697
Method:                Least Squares   F-statistic:                     12.66
Date:               Mon, 27 Jan 2025   Prob (F-statistic):           9.61e-13
Time:                       21:26:57   Log-Likelihood:                -657.09
No. Observations:                 72   AIC:                             1344.
Df Residuals:                     57   BIC:                             1378.
Df Model:                         14
Covariance Type:           nonrobust
====================================================================================
===========
                                  coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------------
-----------
Intercept                    6.043e+04   1.05e+04      5.752      0.000    3.94e+04
8.15e+04
C(MonthFactor)[T.August]     2161.0249   1478.960      1.461      0.149    -800.541
5122.590
C(MonthFactor)[T.December]   2405.6354   1600.868      1.503      0.138    -800.048
5611.318
C(MonthFactor)[T.February]  -4931.0496   1448.613     -3.404      0.001   -7831.847
-2030.252
C(MonthFactor)[T.January]   -5876.2250   1466.743     -4.006      0.000   -8813.327
-2939.123
C(MonthFactor)[T.July]        158.8047   1464.064      0.108      0.914   -2772.933
3090.543
C(MonthFactor)[T.June]         74.9039   1451.873      0.052      0.959   -2832.422
2982.230
C(MonthFactor)[T.March]       362.0891   1447.295      0.250      0.803   -2536.068
3260.247
C(MonthFactor)[T.May]        3724.6814   1447.310      2.574      0.013     826.492
6622.871
C(MonthFactor)[T.November]  -2100.8033   1514.995     -1.387      0.171   -5134.528
932.921
C(MonthFactor)[T.October]   -2562.0591   1466.606     -1.747      0.086   -5498.887
374.769
C(MonthFactor)[T.September] -1640.4358   1467.558     -1.118      0.268   -4579.171
1298.299
UMCSENT                       -36.7838     76.282     -0.482      0.632    -189.537
115.969
CivicQueries                  296.8771     51.137      5.805      0.000     194.476
399.278
CPIAll                       -207.1039     56.507     -3.665      0.001    -320.258
-93.950
==============================================================================
Omnibus:                       5.601   Durbin-Watson:                   1.287
Prob(Omnibus):                 0.061   Jarque-Bera (JB):                4.879
Skew:                          0.499   Prob(JB):                       0.0872
Kurtosis:                      3.793   Cond. No.                     9.73e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spe
```

```
cified.
[2] The condition number is large, 9.73e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [65]:
```python
preds4 = model4.predict(testCCI)
```

In [66]:
```python
SSR_model4 = np.sum((testCCI["CivicSales"] - preds4) ** 2)
y_mean_model4 = np.mean(testCCI["CivicSales"])
TSS_model4 = np.sum((testCCI["CivicSales"]-y_mean_model4)**2)
```

In [67]:
```python
RSquared_model4 = 1 - (SSR_model4/TSS_model4)
```

In [71]:
```python
RSquared_model4
```

Out[71]:
```
-0.4579339533332556
```

In [ ]: