In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.linear_model import LinearRegression
from statsmodels.formula.api import ols
miami = pd.read_csv('C:/Users/samik/Documents/econometrics/miami-housing.csv')#
miami
```

Out[1]:

	LATITUDE	LONGITUDE	PARCELNO	SALE_PRC	LND_SQFOOT	TOT_LVG_AREA	SPEC_FEA
0	25.891031	-80.160561	622280070620	440000.0	9375	1753	
1	25.891324	-80.153968	622280100460	349000.0	9375	1715	
2	25.891334	-80.153740	622280100470	800000.0	9375	2276	
3	25.891765	-80.152657	622280100530	988000.0	12450	2058	
4	25.891825	-80.154639	622280100200	755000.0	12800	1684	
...	...	...	...	...	...	...	
13927	25.783130	-80.259795	131320040990	275000.0	6780	967	
13928	25.783585	-80.260354	131320040910	340000.0	7500	1854	
13929	25.783793	-80.256126	131320040420	287500.0	8460	1271	
13930	25.784007	-80.257542	131320040330	315000.0	7500	1613	
13931	25.784387	-80.258901	131320040700	250000.0	8833	1867	

13932 rows × 17 columns

In [2]:
```
#very simple regression using python
simple = ols("SALE_PRC ~ CNTR_DIST", data = miami).fit()
print(simple.summary())
```

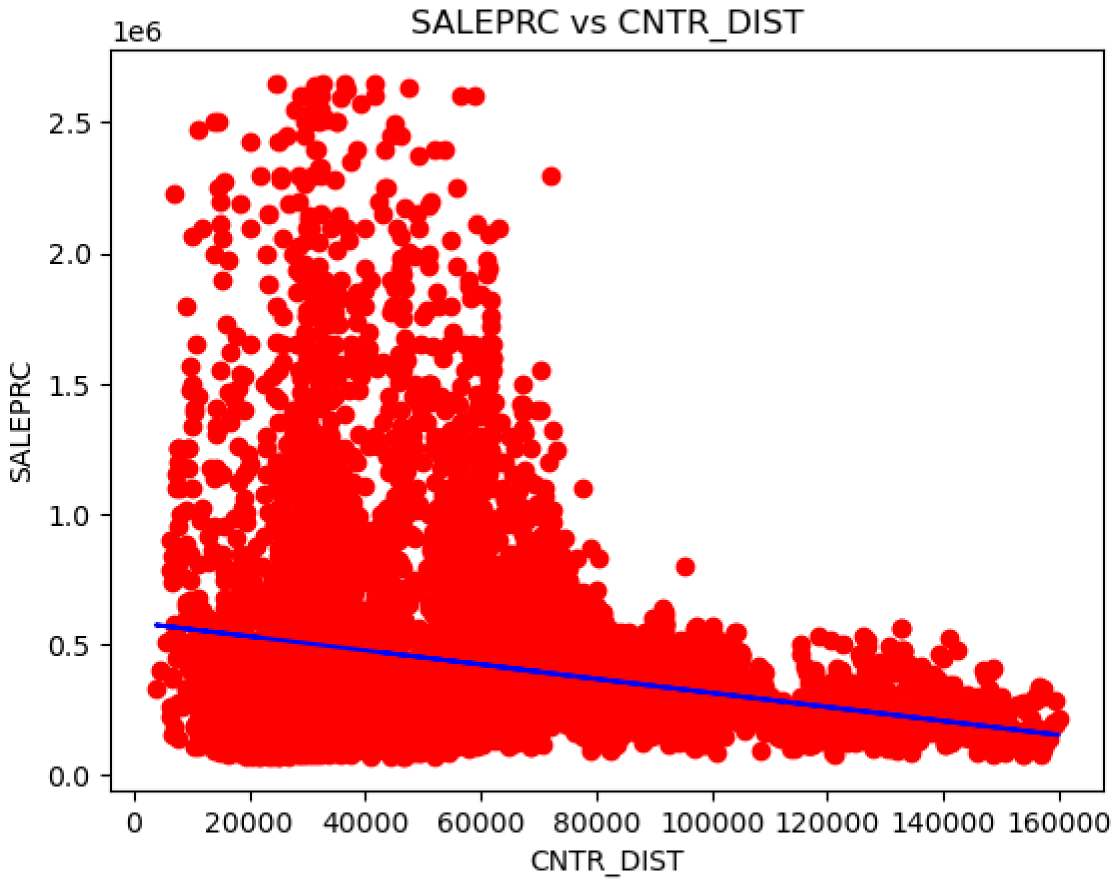
```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:               SALE_PRC   R-squared:                       0.074
Model:                            OLS   Adj. R-squared:                  0.074
Method:                 Least Squares   F-statistic:                     1108.
Date:                Wed, 22 Feb 2023   Prob (F-statistic):          8.20e-234
Time:                        22:55:28   Log-Likelihood:            -1.9572e+05
No. Observations:               13932   AIC:                         3.914e+05
Df Residuals:                   13930   BIC:                         3.915e+05
Df Model:                           1                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   5.842e+05   6109.678     95.615      0.000    5.72e+05    5.96e+05
CNTR_DIST     -2.6899      0.081    -33.285      0.000      -2.848      -2.532
==============================================================================
Omnibus:                     8857.556   Durbin-Watson:                   0.655
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           109677.981
Skew:                           2.923   Prob(JB):                         0.00
Kurtosis:                      15.440   Cond. No.                     1.79e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly s
pecified.
[2] The condition number is large, 1.79e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [3]:

```
X=miami[['CNTR_DIST']]
Y=miami[['SALE_PRC']]

regressor = LinearRegression()
regressor.fit(X,Y)
y_pred = regressor.predict(X)
plt.scatter(X, Y, color = 'red')
plt.plot(X, regressor.predict(X), color = 'blue')
plt.title('SALEPRC vs CNTR_DIST')
plt.xlabel('CNTR_DIST')
plt.ylabel('SALEPRC')
plt.show()
```



In [4]:
```
#very simple regression using python
simple = ols("SALE_PRC ~ CNTR_DIST + HWY_DIST + RAIL_DIST", data = miami).fit()
print(simple.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               SALE_PRC   R-squared:                       0.145
Model:                            OLS   Adj. R-squared:                  0.145
Method:                 Least Squares   F-statistic:                     788.5
Date:                Wed, 22 Feb 2023   Prob (F-statistic):               0.00
Time:                        22:55:28   Log-Likelihood:            -1.9516e+05
No. Observations:               13932   AIC:                         3.903e+05
Df Residuals:                   13928   BIC:                         3.904e+05
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   4.782e+05   6662.856     71.775      0.000    4.65e+05    4.91e+05
CNTR_DIST     -3.3164      0.087    -37.929      0.000      -3.488      -3.145
HWY_DIST      13.9236      0.415     33.562      0.000      13.110      14.737
RAIL_DIST      4.9484      0.454     10.909      0.000       4.059       5.838
==============================================================================
Omnibus:                     8720.077   Durbin-Watson:                   0.685
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           104813.028
Skew:                           2.872   Prob(JB):                         0.00
Kurtosis:                      15.148   Cond. No.                     2.05e+05
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly s
pecified.
[2] The condition number is large, 2.05e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
In [ ]:
```