**Praktikum 6 – Data Analitik**  
**ANALISIS DESKRIPTIF DAN REGRESI BERGANDA MENGGUNAKAN PYTHON**

**Catatan:**

**Data table houseprices : houseprices120.csv**

**Data table houseprices2 : houseprices2.csv**

**Install pymysql di cmd**

|  |
| --- |
| **pip install pymysql** |

**Install pymysql di jupyter notebook**

|  |
| --- |
| **import sys**  **!{sys.executable} -m pip install pymysql** |

|  |
| --- |
| import pymysql  db = pymysql.connect("localhost","root","","db\_da" )  cursor = db.cursor()  sql = "SELECT \* FROM houseprice120"  cursor.execute(sql)    rows = cursor.fetchall()  import pandas as pd  import pandas as pd  from pandas.tools import plotting  import matplotlib.pyplot as plt  import numpy as np  from sklearn.model\_selection import train\_test\_split, cross\_val\_score, KFold, GridSearchCV  from sklearn.linear\_model import LogisticRegression  from sklearn.tree import DecisionTreeClassifier,export\_graphviz  from sklearn.metrics import confusion\_matrix, accuracy\_score  from sklearn.ensemble import GradientBoostingClassifier, RandomForestClassifier |

|  |
| --- |
| house = pd.DataFrame( [[ij for ij in i] for i in rows] )  house.rename(columns={0: 'Price', 1: 'SqFt', 2: 'Bedrooms', 3: 'Bathrooms', 4:'Offers', 5:'Brick', 6:'Neighborhood'}, inplace=True);  house.head(5) |

|  |
| --- |
| #boxplot dari harga rumah/Price  col\_list=['Price']  numhouse = house[house.columns[house.columns.isin(col\_list)]]  plt.figure(figsize=(10,5))  numhouse.boxplot(sym='r\*', grid=False)  plt.show() |

|  |
| --- |
| #hist dari Price  plt.figure(figsize=(10,5))  plt.subplot(121)  house['Price'].plot.hist(bins=10, title='Price')  plt.show()  C:\Users\STEVEN\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\C61FFA8A.tmp |

|  |
| --- |
| col\_list=['Price', 'SqFt']  numhouse = house[house.columns[house.columns.isin(col\_list)]]  numhouse.plot.scatter(x='SqFt', y='Price') |

|  |
| --- |
| col\_list=['Price', 'Bedrooms']  numhouse = house[house.columns[house.columns.isin(col\_list)]]  plt.figure(figsize=(10,5))  numhouse.boxplot(by='Bedrooms')  plt.show() |

|  |
| --- |
| col\_list=['Price', 'Offers']  numhouse = house[house.columns[house.columns.isin(col\_list)]]  plt.figure(figsize=(10,5))  numhouse.boxplot(by='Offers')  plt.show() |

|  |
| --- |
| col\_list=['Price', 'Brick']  numhouse = house[house.columns[house.columns.isin(col\_list)]]  plt.figure(figsize=(10,5))  numhouse.boxplot(by='Brick')  plt.show() |

|  |
| --- |
| col\_list=['Price', 'Neighborhood']  numhouse = house[house.columns[house.columns.isin(col\_list)]]  plt.figure(figsize=(10,5))  numhouse.boxplot(by='Neighborhood')  plt.show() |

**Regresi Berganda Menggunakan data Housprices (Price, SqFt dan Brick)**

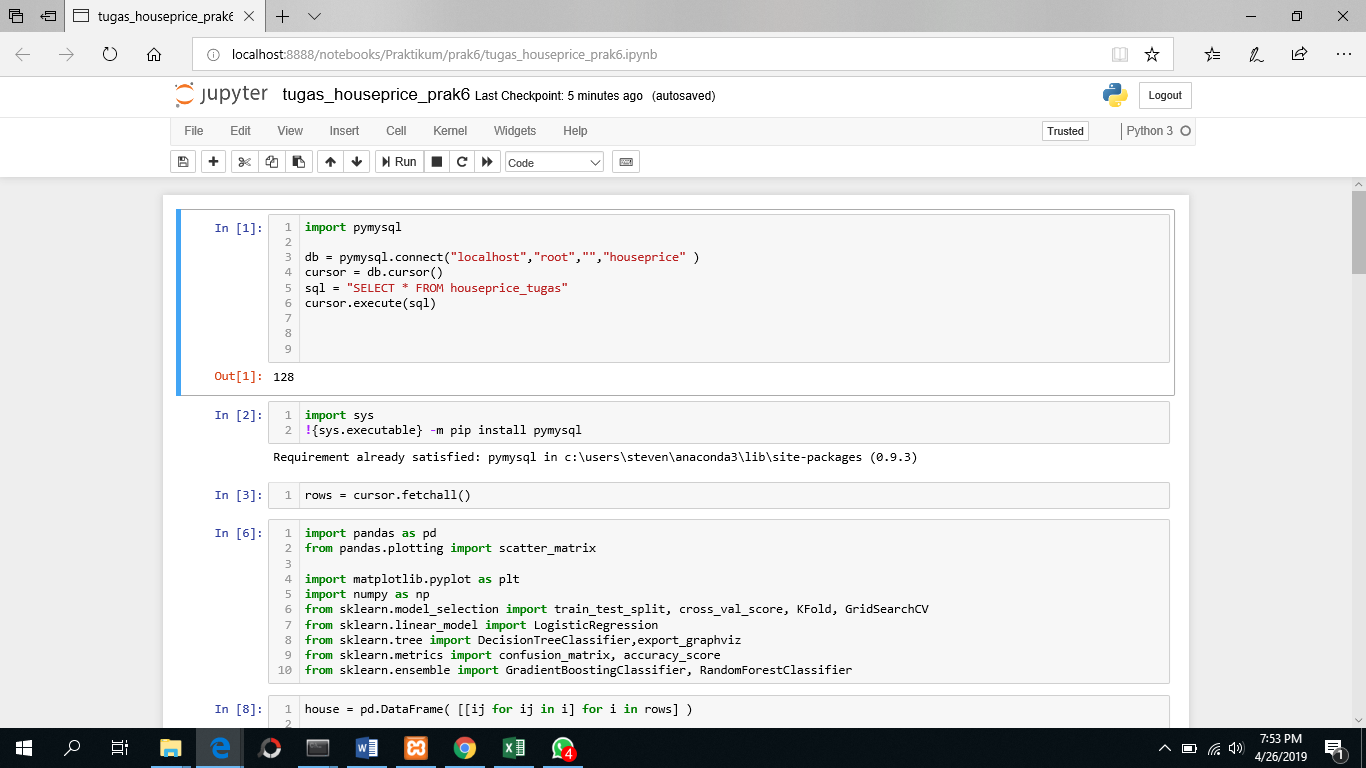
|  |
| --- |
| sql2 = "SELECT \* FROM houseprices2"  cursor.execute(sql2)  rows2 = cursor.fetchall()  house2 = pd.DataFrame( [[ij for ij in i] for i in rows2] )  house2.rename(columns={0: 'Price', 1: 'SqFt', 2: 'Bedrooms', 3: 'Bathrooms', 4:'Offers', 5:'Brick', 6:'Neighborhood'}, inplace=True);  house2.head(10) |

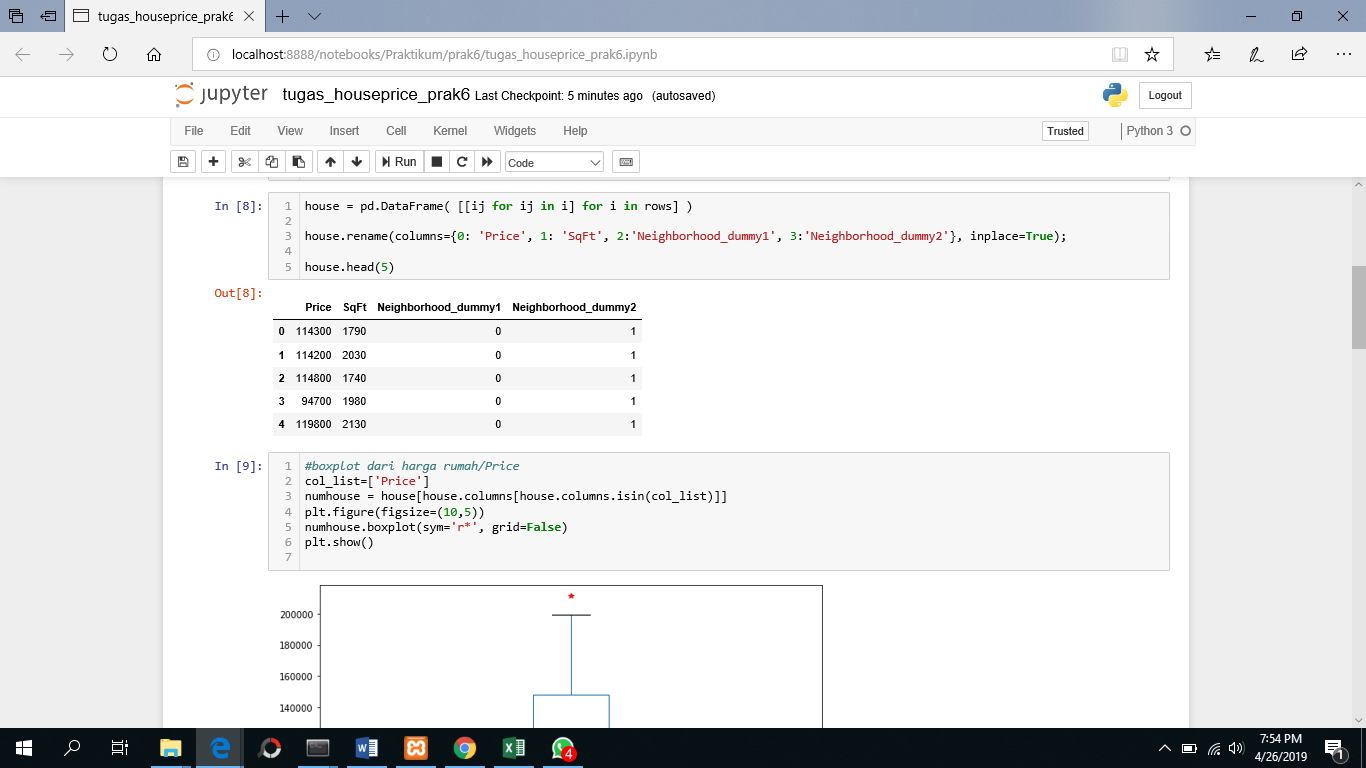
|  |
| --- |
| import statsmodels.api  import statsmodels.formula.api |

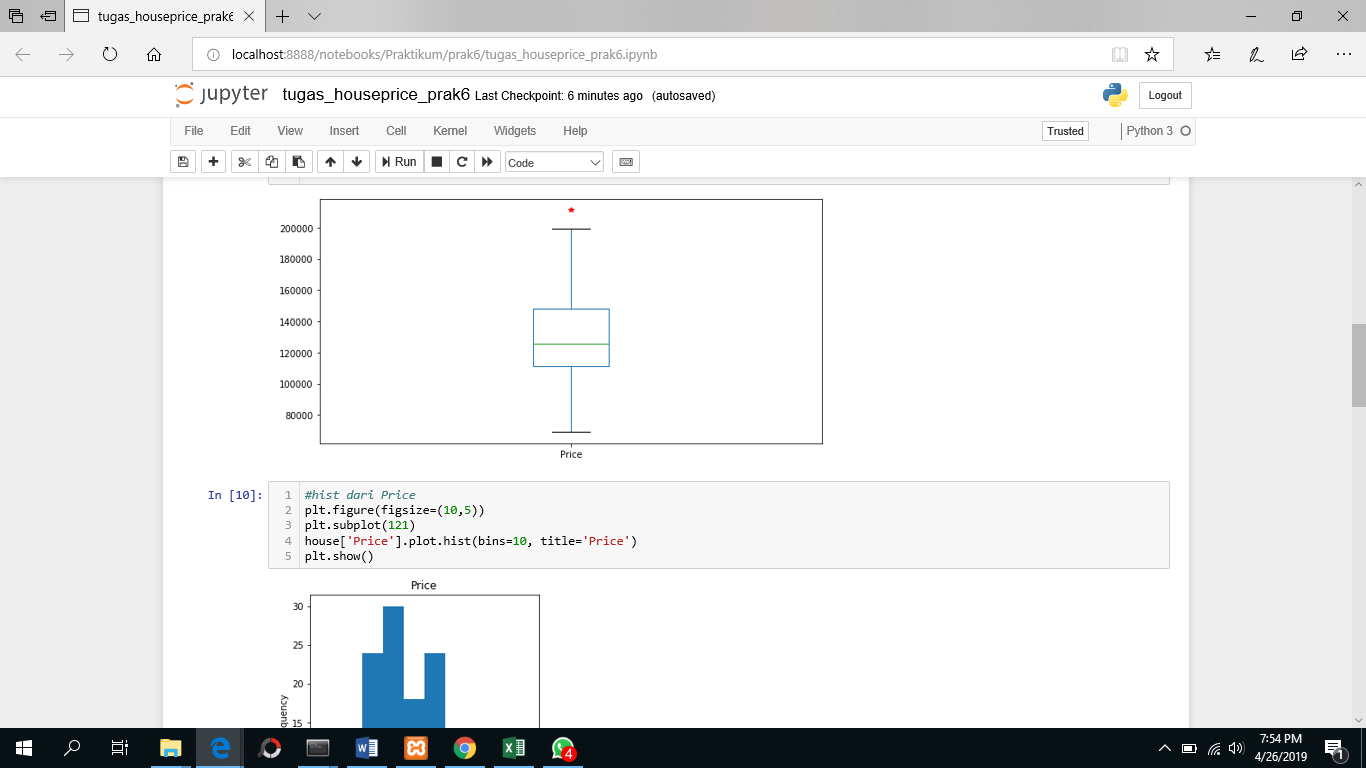
|  |
| --- |
| reg = statsmodels.formula.api.ols('Price~SqFt+Brick', data= house2).fit()  print(reg.summary()) |

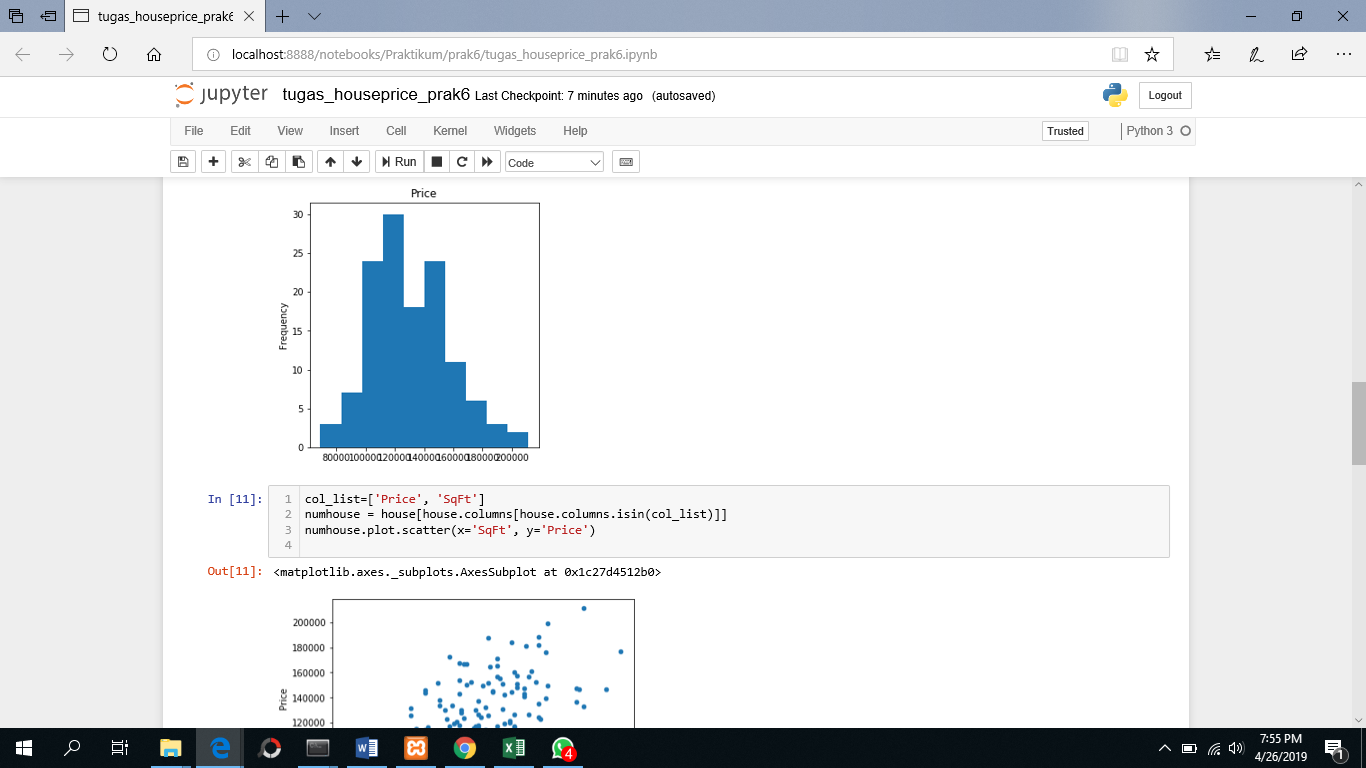
**Tugas:**

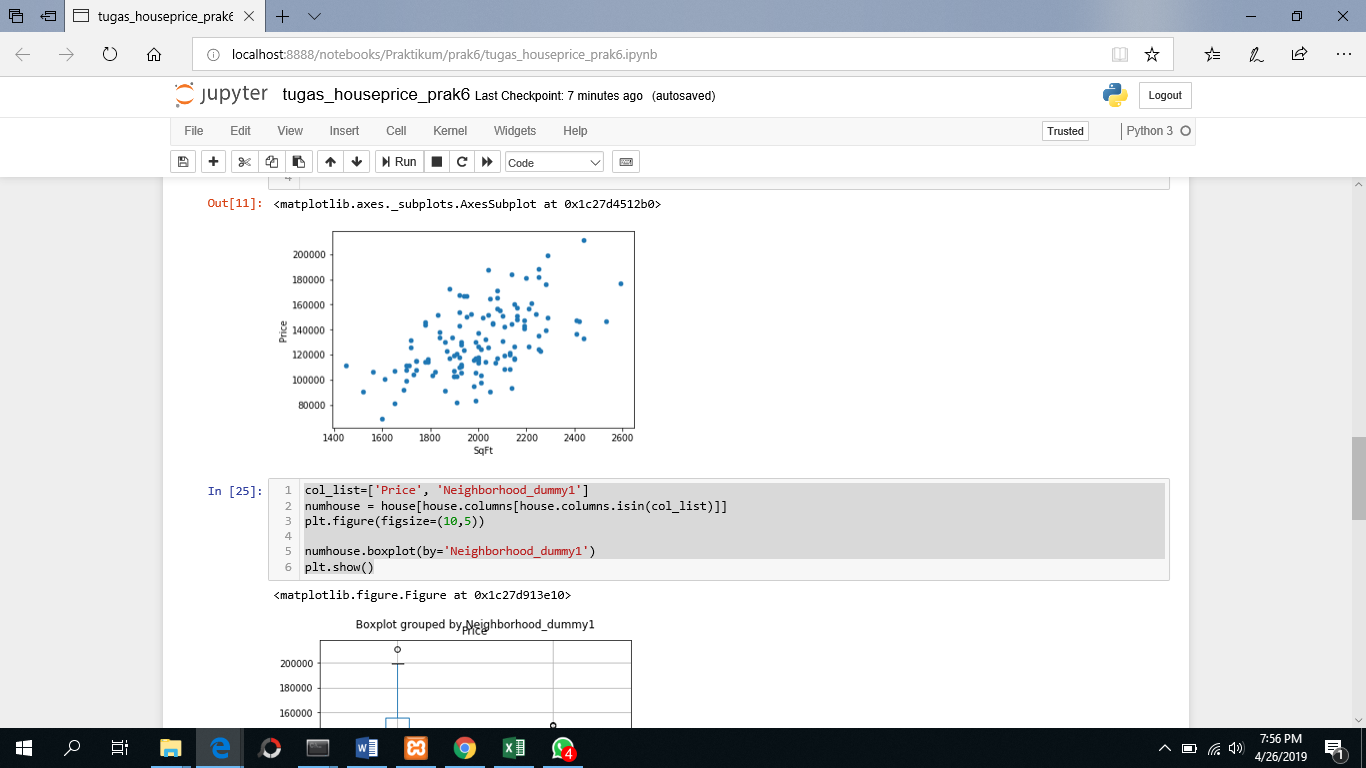
Buatlah regresi berganda menggunakan data houseprices antara variable Prices, SqFt dan Neighborhood. Terlebih dahulu, buatlah dummy variable dari Neighborhood.

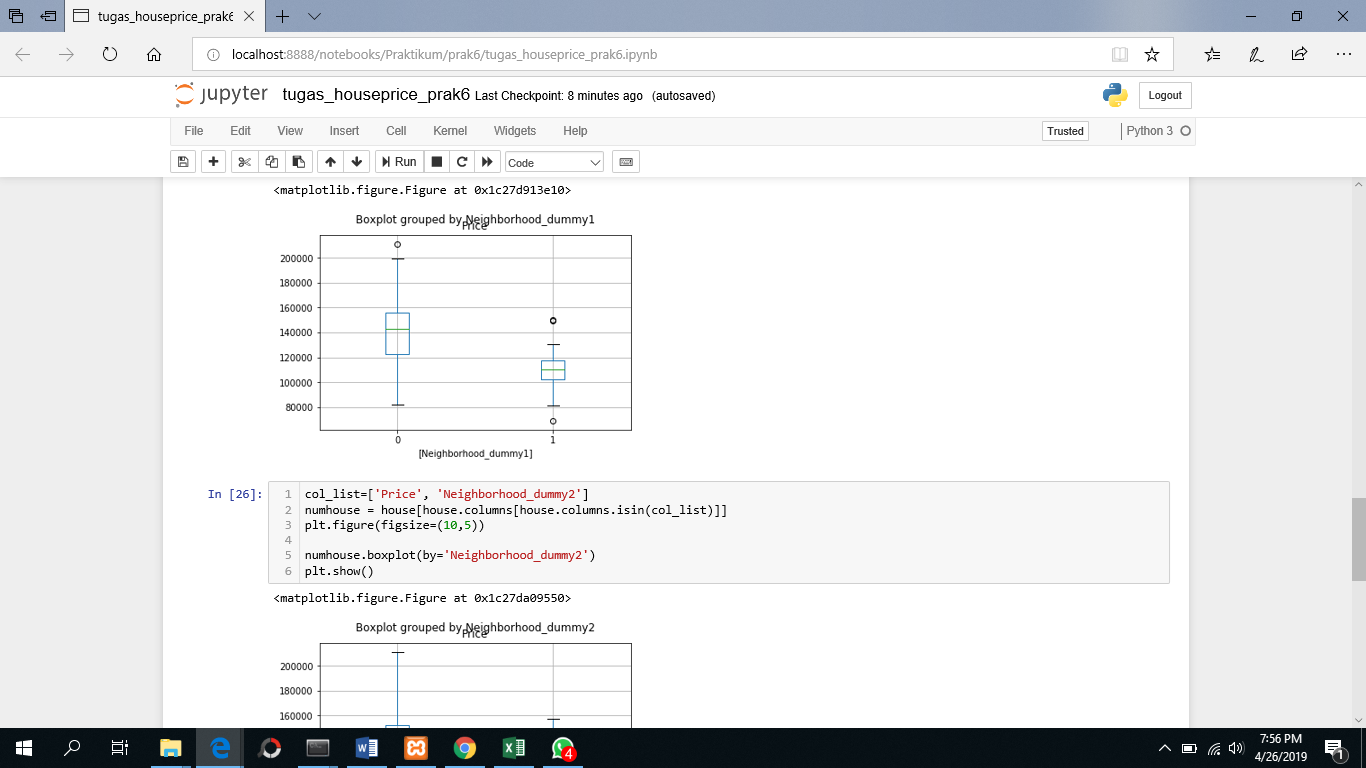


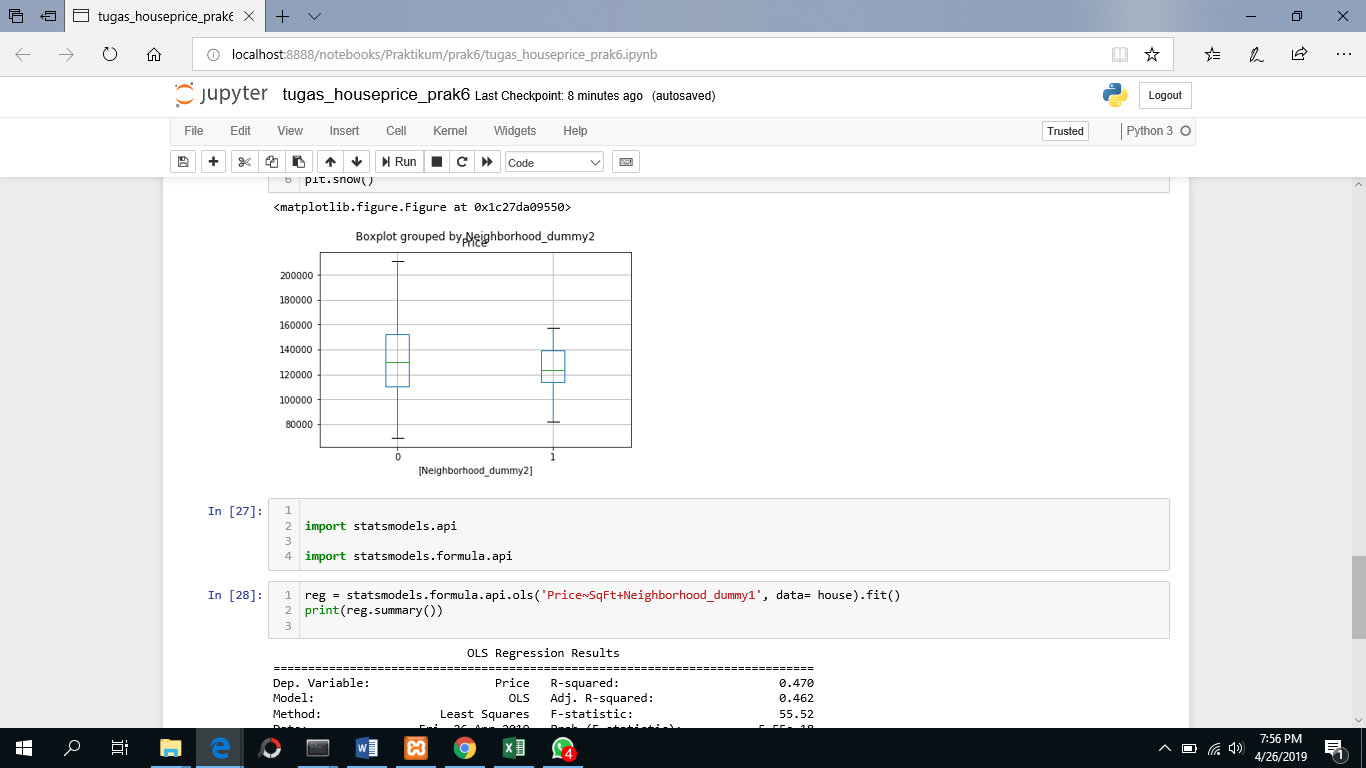


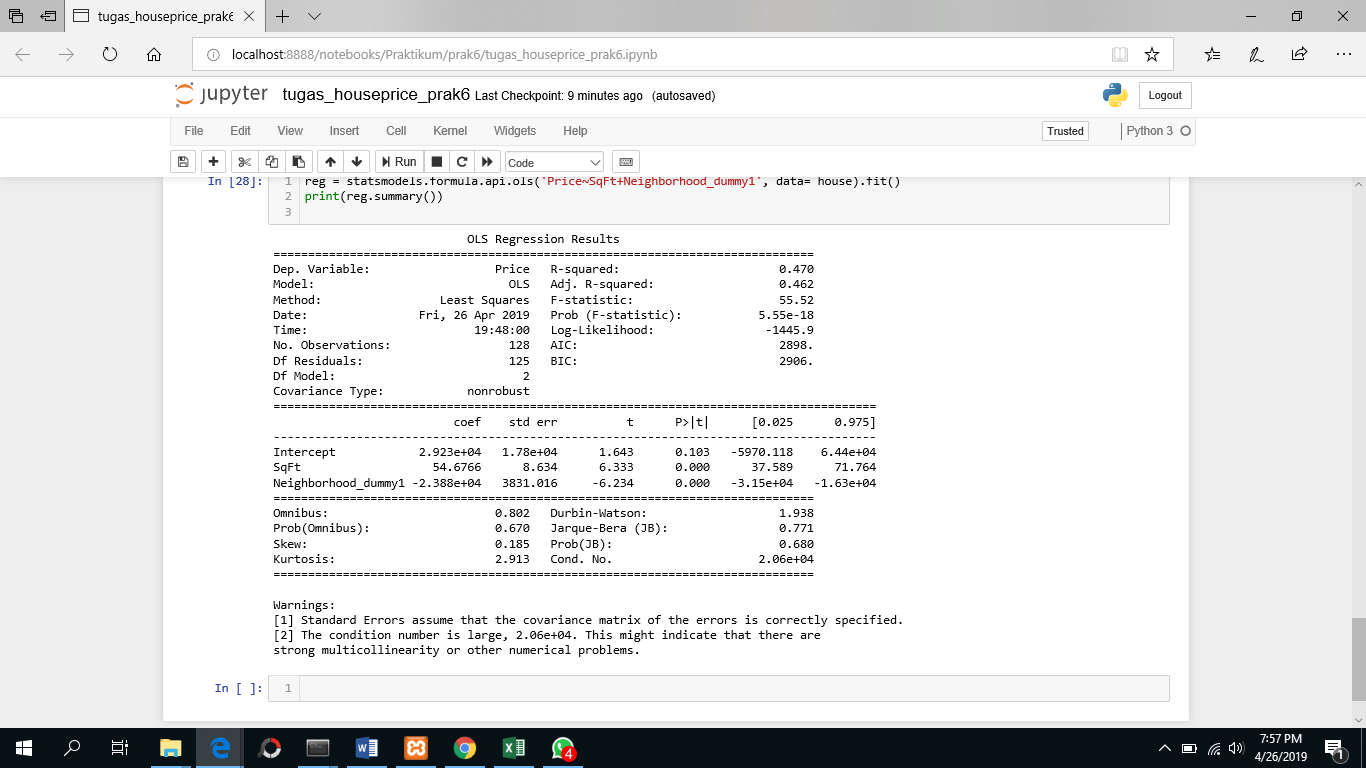












Sumber :

Modul Pelatihan Pemodelan Predictive Menggunakan Python, SPS IPB, 2017

<https://pandas.pydata.org/pandas-docs/stable/visualization.html>

<https://community.modeanalytics.com/python/tutorial/python-histograms-boxplots-and-distributions/>

<https://community.modeanalytics.com/python/tutorial/pandas-groupby-and-python-lambda-functions/>

<https://happygostacie.wordpress.com/2016/04/24/boxplot-fun-with-python/>