

**Edmonton Residential Property Market: Potential ways to Increase Profits**

# Executive summary

Our recommendation to increase profit by flipping properties is to find the least expensive properties in the most expensive neighborhood. For example, we predicted that Windsor Park is the most expensive neighborhood in Ward 8 (see page 8). We expect that buying the cheapest properties in this Neighborhood will have highest returns on investment (ROI). We also demonstrated it by visualizing the Assessed Value of properties (see page 6). We have identified Ward 6,8,9, 10, and 5 (see page 7) as the most valuable Wards. We further explore the Neighborhoods in Ward 8 and Ward 6. Our explanatory variable was able to seems to estimate the Assessed Values of properties. R-squared measures the strength of the relationship between our model and the dependent variable on a convenient 0 – 100% scale.

Our model determine which Wards have higher Accessed Value of properties in Edmonton.

Assessed\_Value =β0 +β1Ward + β2Zoning + β3Garage + β4Lot\_Size + β5Age\_of\_Property+α

A R-squared: 0.685 shows that our explanatory variables were able to determine the Assessed Value of properties . It confirms that the Ward a property is located as well as the Zoning by law of the location can determine the value of the properties. In the same way, Garage, Lot Size, and Age of Property contribute to the Accessed Value of property. We were able to determine that Wards 6, 8, 9,10 and 5 are well priced than Ward 1 which is our reference, Ward. We also determine that properties in Ward 1 are pricier than Ward 7,4,3,2, 11 and 12

We used the model for Neighborhoods in Ward 8 and 6 to determine the Assessed Value of property:

Assessed\_Value =β0 + β1Neighborhood + β2Garage + β3Lot\_Size + β4Age\_of\_Property + α

We imported statsmodels library in pandas for our model

model = smf.ols('Assessed\_Value ~ C(Neighborhood)+ Lot\_Size +Age\_Property + C(Garage)', data=X). While ‘X’ is our data.

For Ward 8 model the R-squared: 0.708, while Ward 6 model the R-squared: 0.603

Windsor Park seems to be the most priced Neighborhood in Ward 8; we are 95% confident that this Neighborhood is highly priced than Belgravia which is our reference Neighborhood. We are also 95% confident that Belgravia is the second most priced Neighborhood in Ward 8.

For Ward 6, Our reference Neighborhood for this prediction is Boyle Street. We have 95% confidence that the Assessed Value of properties in Rossdale, Glenora, Downtown and Oliver is higher than Boyle Street. While Grovenor, North Glenora, Queen Mary Park, Westmount, Central McDougall, McCauley are valued lower than Boyle Street.

RECOMMENDATION:

Potential customers of the finished buildings constructed by the firm:

* Buy cheaper properties in blue zones closer to the green on the map in Ward(6,8,9,10 and 5)
* Understand the Average Assessed Value of properties in Neighborhood
* Check every single piece of information about any property using the Account Number

Potential renters of the building the firm retains ownership:

* Rent cheaper properties in blue zones closer to the green on the map in Ward(6,8,9,10 and 5)

Investors and lenders that provide financing for the firm’s activities:

* Inflation Hedging and understand historical prices : Home prices tend to rise along with inflation. This is because homebuilders' costs rise with inflation
* Invest in Real Estate Mutual Funds

# METHODOLOGY

Data Source: The Property Assessment Data (Historical) was sourced from City of Edmonton Open Data site. Available at: <https://data.edmonton.ca/City-Administration/Property-Assessment-Data-2012-2019-/qi6a-xuwt>. Accessed on: 2022-11-19.

Graphical user interface, application, table

Description automatically generated

Also, we accessed the Lab Assignment2 which contains 384,217 rows of (Account Number, Suite, House Number, Street Name, Garage, Neighbourhood ID, Neighbourhood, Ward, Assessed Value, Latitude and Longitude)

Subsequently, we merged the two files to have a data frame with ‘Ward’ and Neighbourhood ID . We considered the Wards as groups of different Neighborhoods. “Wards are neighbouring groups of residential, commercial, industrial, and other areas that are represented by a City Councillor. Wards are an essential part of Edmonton’s local government system. Voters in each Ward elect a Councillor to deal with local issues and make decisions that impact all residents”. (Ward Boundary Review, City of Edmonton).

Edmonton has 12 wards consisting of different Neighborhoods. We also considered Zoning as a strong variable for our model. Zoning by law is for the purpose of land development, the City of Edmonton is divided into zones. The zone that a particular property is in determines what can be built on that property.

# EXPLORATORY DATA ANALYSIS

We removed all missing values from the entire dataset and used the last assessment date to determine the Assessed Value of property. We considered only ‘RESIDENTIAL’ under “Assessment Class 1” that have:

RESIDENTIAL |323264

COMMERCIAL |12901

OTHER RESIDENTIAL |3168

FARMLAND |10

NONRES MUNICIPAL/RES EDUCATION |3

DESIGNATED IND PROPERTIES |1

We removed all outliers from the dataset, and our variable for prediction and analysis are:

Account Number

Assessment Year

Latitude

Longitude

Neighbourhood

Actual Year Built

Garage

Zoning

Lot Size

Assessed Value

Age of Property

Assessment Class 1

Ward

Neighbourhood ID

Note: The Age of Property was calculated by subtracting Actual Year Built from current year(2022).

# PREDICTIVE ANALYTICS

Model

Assessed\_Value =β0 +β1Ward + β2Zoning + β3Garage + β4Lot\_Size + β5Age\_of\_Property+α

We imported statsmodels library in pandas for our model

model = smf.ols('Assessed\_Value ~ C(Ward) +C(Zoning) + Lot\_Size +Age\_Property + C(Garage)', data=X)

While ‘X’ is our data.

We also limited our explanatory variable Zoning to most frequent types:

RF1 108355

RSL 41385

RA7 31853

RF5 24650

RF4 18761

RF3 18156

RPL 15258

DC2 9450

DC1 7473

RA9 6083

RMD 5780

RA8 3820

RF6 2858

RF2 2755

HA 1202

HDR 1168

Visualization

We imported the shapefile of Edmonton from the City of Edmonton’s website and plot the coordinates( Longitude and Latitude) using library geo pandas to create GeoPandas data frame.

# RESULT

Map

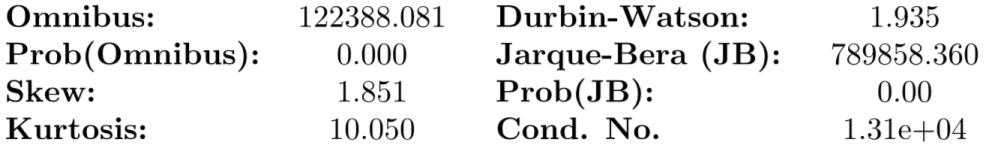
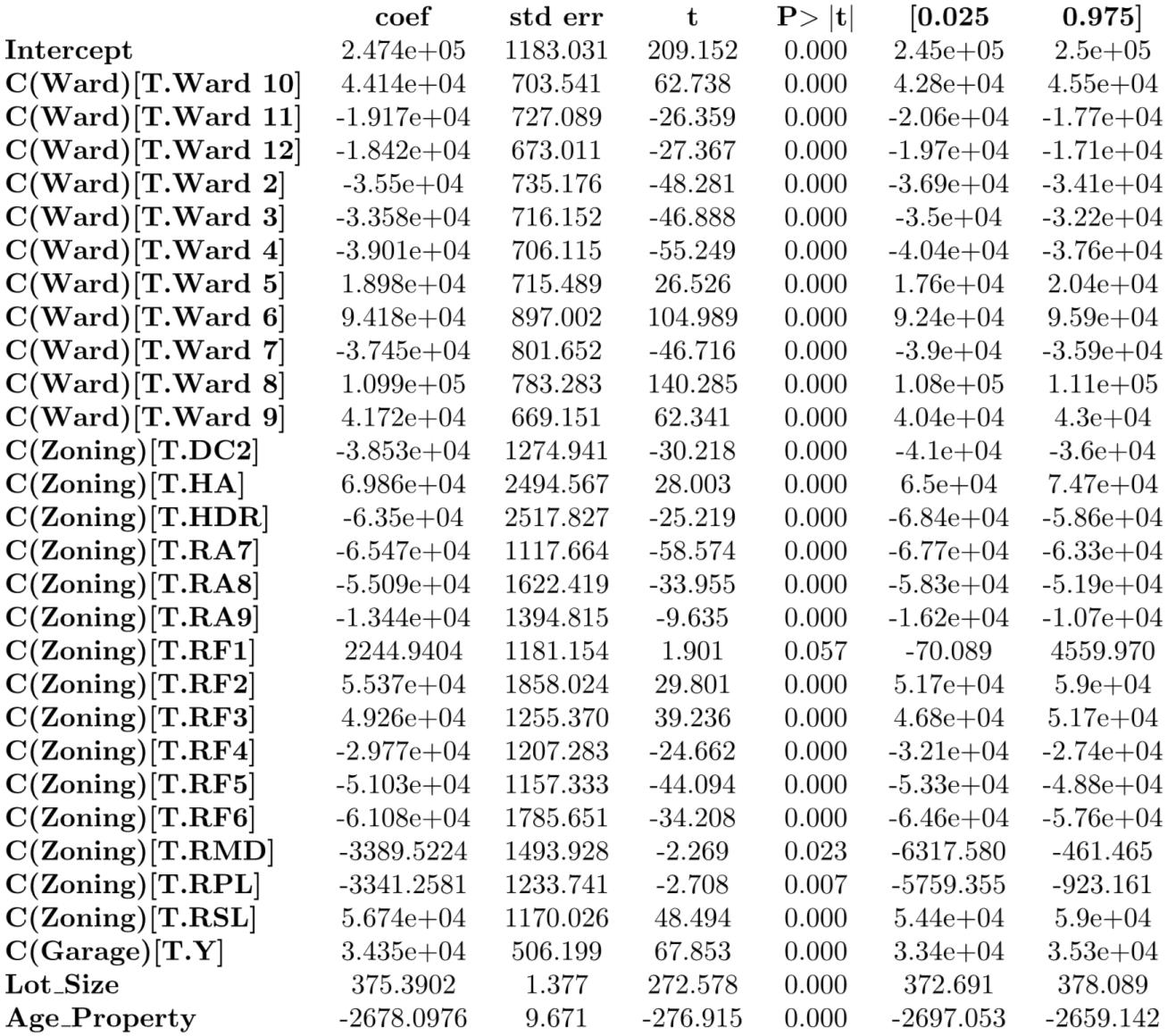
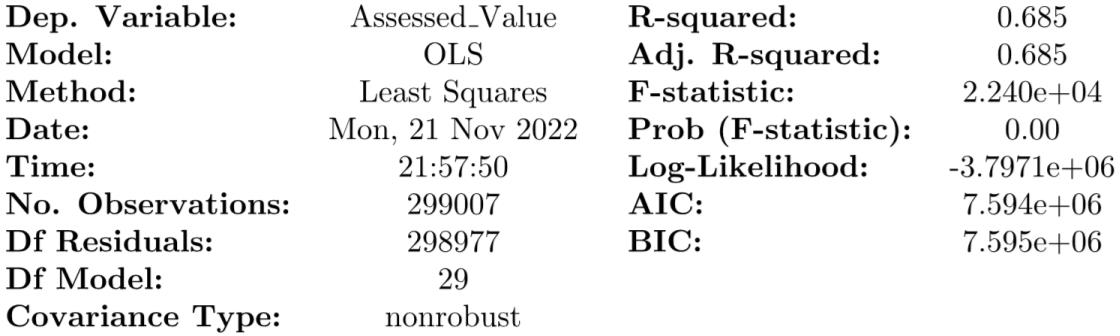
Description automatically generated

Inferring from the map, we expect Residential Properties in Ward 6,8,9, and 10 to be highly valued than other Wards. Also, we expect to see a flipping of the blue region of around the green zones of these wards in the nearest future. These are properties that are most likely below-market price.

Result

Inferring from the map, we expect Residential Properties in Ward 6,8,9, and 10 to be highly valued than other Wards. Also, we expect to see a flipping of the blue region of around the green zones of these wards in the nearest future. These are properties that are most likely below-market price.

# REGRESSION RESULT:



Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly

specified.

[2] The condition number is large, 1.31e+04. This might indicate that there are

strong multicollinearity or other numerical problems.

We looked at the 95% confidence intervals for the coefficients. Those Wards whose CI are all negative we have considerable evidence that they tend to be priced lower than the reference Ward (Ward1). Those will all positive values tend to be priced higher on average. Also, the F-statistics tells us if we have sufficient evidence that the mean Assessed Value differs based on the Ward (as that is proxy explanatory variable Neighborhood). The small p-value here tells us to reject this null hypothesis here. Apparently, the location of property does matter.

We also looked at the boxplot:

Chart, box and whisker chart

Description automatically generated Ward:

Zoning:

The frequent Zoning bylaw in Edmonton from our research shows that only

Chart, box and whisker chart

Description automatically generatedRF1, RF2, RF3, RSL, HA, are highly priced than DC1 except RF4, RF5, RF6, RMD RPL, DC2, RA9, RA8, HDR, and RA7.

We plot the Assessed\_Value against the Age of property. From the regression output, we are 95% confident that newer properties tend to be priced higher than old houses.

New property tends to be pricier than Assessed Value

Chart, line chart

Description automatically generated

We used the model for Neighborhoods in Ward 8 to determine the Assessed \_Value of property:

Assessed\_Value =β0 + β1Neighborhood + β2Garage + β3Lot\_Size + β4Age\_of\_Property + α

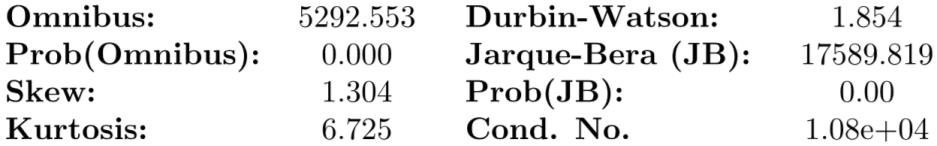
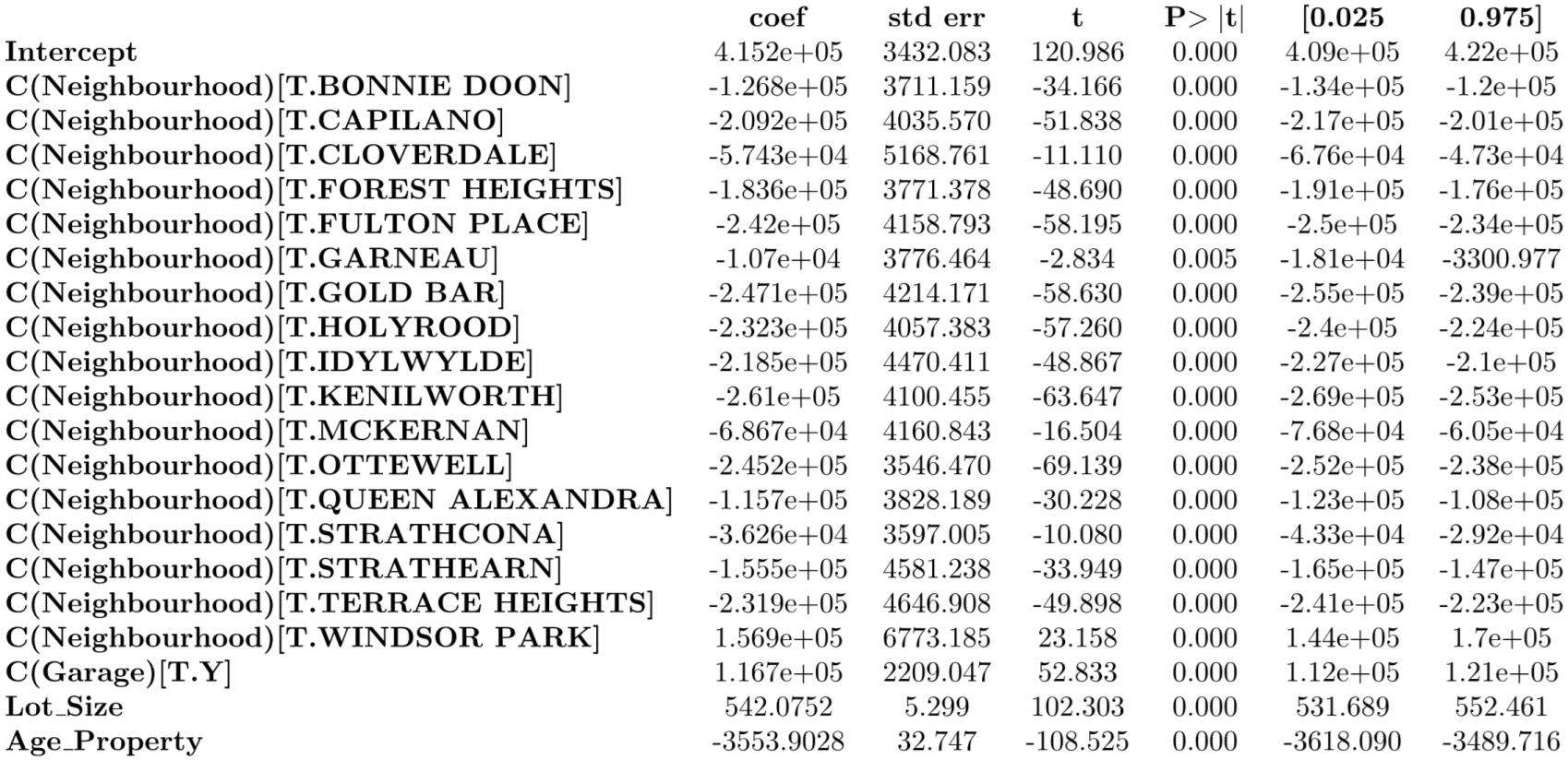
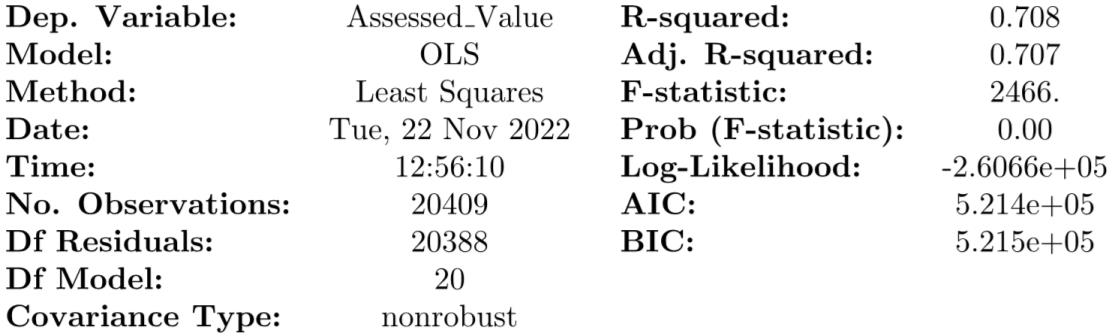
We imported statsmodels library in pandas for our model

model = smf.ols('Assessed\_Value ~ C(Neighborhood)+ Lot\_Size +Age\_Property + C(Garage)', data=X)

While ‘X’ is our data.

## RESULT WARD 8 PREDICTION:

Windsor Park seems to be the most priced Neighborhood in Ward 8; we are 95% confident that this Neighborhood is highly priced than Belgravia which is our reference Neighborhood. We are also confident that Belgravia is the second most priced Neighborhood in Ward 8. Also, we are 95% confident that all the Neighborhood CI with negative range are priced below Belgravia



Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly

specified.

[2] The condition number is large, 1.08e+04. This might indicate that there are

strong multicollinearity or other numerical problems.

Chart, box and whisker chart

Description automatically generated

We used the model for Neighborhoods in Ward 6:

Assessed\_Value =β0 + β1Neighborhood + β2Garage + β3Lot\_Size + β4Age\_of\_Property+α

We imported statsmodels library in pandas for our model

β0 = Intercept

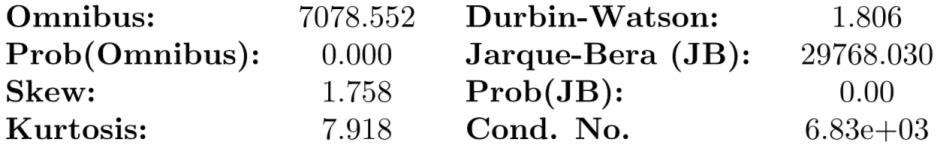
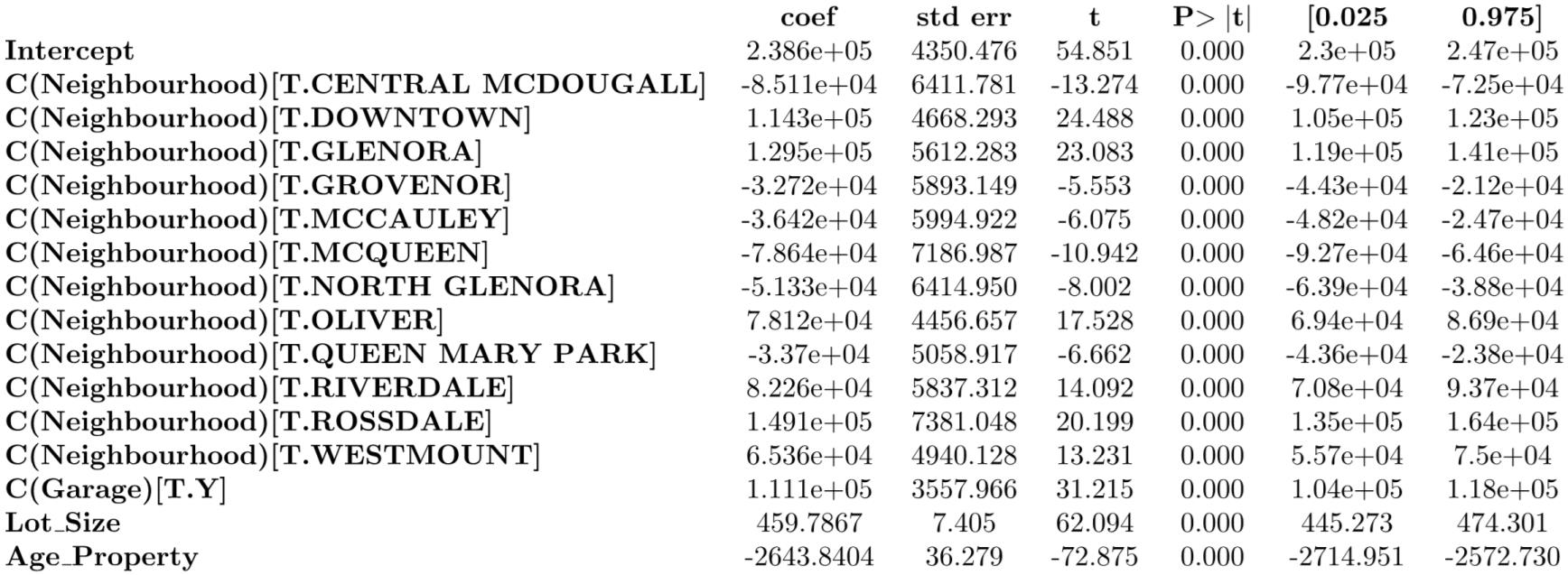
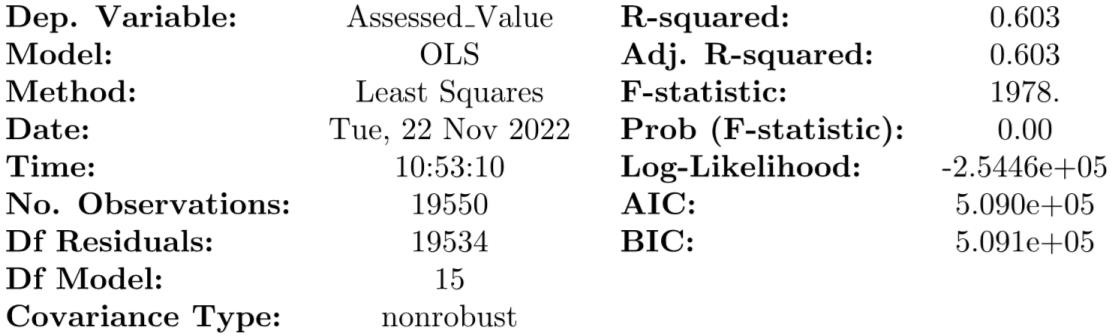
α = Error term

model = smf.ols('Assessed\_Value ~ C(Neighborhood)+ Lot\_Size +Age\_Property + C(Garage)', data=X1)

While ‘X1’ is our data.

## RESULT WARD 6 PREDICTION:

Our reference Neighborhood for this prediction is Boyle Street. We have 95% confidence that the Assessed Value of properties in Rossdale, Glenora, Downtown and Oliver is higher than the than Boyle Street. While Grovenor, North Glenora, Queen Mary Park, Westmount, Central McDougall, McCauley are valued lower than Boyle Street.



Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly

specified.

[2] The condition number is large, 6.83e+03. This might indicate that there are

strong multicollinearity or other numerical problems.