CSDA 1050 Sprint 3

August 27, 2019

Nickolas Lal

Introduction: The Canadian housing market has always been an interesting topic of discussion, especially the market in Toronto. It is one of Canada's most popular and most diverse cities and it attracts people from all over the world however many people find affordability an issue, both in terms of buying and renting in Toronto. There are many individuals and organizations who study and research all aspects of the market all year long. The focus of this assignment will be on the residential rental side, specifically apartments.

In Toronto there are many regulations and guidelines in place when it comes to increasing and reducing rent, but are these policies working, and are they having the desired effect? There are many questions that can be attempted to be answered but the scope of this project is to look at the marginal difference in rent price between tenured and new renters. Although many other questions may be raised along the process, and they will attempt to be answered, the focus will be to try to identify this difference and provide some explanation as to what is causing it.

Research Question: What is the marginal difference in rent price between tenured renters and new renters in Toronto?

Per the above at the start of the project the goal was to try to identify the marginal difference in rent price between tenured renters and new renters in Toronto, specifically looking at apartments. During the completion of the previous two Sprints many gaps were identified, and this became increasingly more difficult, so the goal then became to forecast the average apartment rent price in Toronto as this can be very useful for prospective renters, landlords, and the City of Toronto. The report contains some of the code, output, rationale, and conclusions.

Sprint 1 Portion: Sprint 1 was the exploratory analysis phase of the project. It consisted of data collection and cleaning along with some preliminary analysis and data visualization with the findings outlined below. The full code and details are in the Sprint 1 notebook.

Data collection and cleaning:

Toronto Income data from 2013-2017- The below Statistics Canada link includes income of individuals in 2017 constant dollars from 2013-2017. There are options to filter the data by age group, sex, income source and by geographic location including Canada and provinces and selected census metropolitan areas. It was filtered for total income of all residents ages 16 and over in Toronto. The dataset was then further cleaned to only include the necessary information which is the average and median income per year.

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110023901&pickMembers%5B0%5D=1.17&pickMembers%5B1%5D=2.1&pickMembers%5B2%5D=3.1&pickMembers%5B3%5D=4.1

Annual Inflation: The annual inflation rate in Canada, comparing year over year, December to December, from 1984-2018 was taken from the below. It was then input into a CSV file so it could be used for analysis.

https://www.inflation.eu/inflation-rates/canada/historic-inflation/cpi-inflation-canada.aspx

Ontario rent increase guidelines- The full details on how the law and calculation works along with the historical numbers dating back to 1991 was available on the below government of Ontario website. It was then combined with the inflation numbers, starting from 1991, and then input into a CSV file so it could be used for analysis. A new column has also been added to help better showcase the difference in Ontario's rental guidelines and inflation. The column is simply the calculation of the guidelines rate minus the inflation rate.
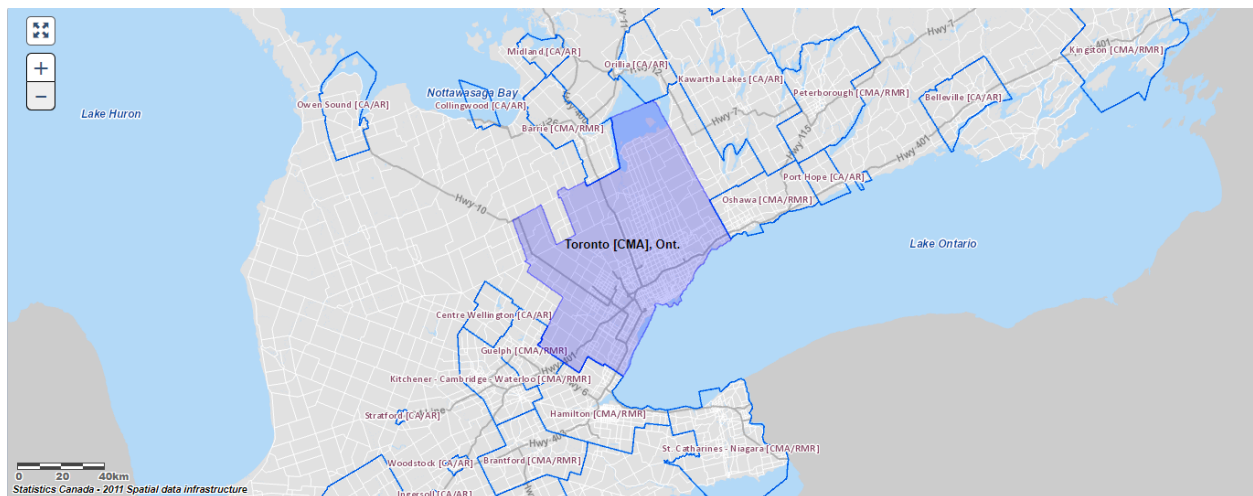
https://www.ontario.ca/page/rent-increase-guideline

Historical Toronto Rental Data

https://www.torontorentals.com/blog/average-rent-in-toronto-since-2000
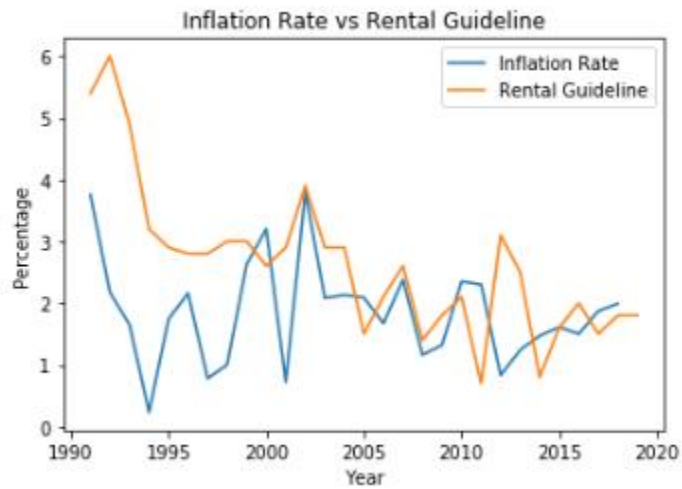

Preliminary Exploration and Analysis:

City of Toronto boundaries

The first graph below compares the Inflation rate with rental increase guidelines. Here we can see that generally the rent increase is always larger than inflation although there have been a few occasions where it was smaller.


A better visualization of the gap in the rental increase guidelines vs the Inflation rate is displayed by plotting the difference, Guideline-Inflation rate. Here we can better see that the rental guideline rate is usually higher, and occasionally lower or equal to the inflation rate.
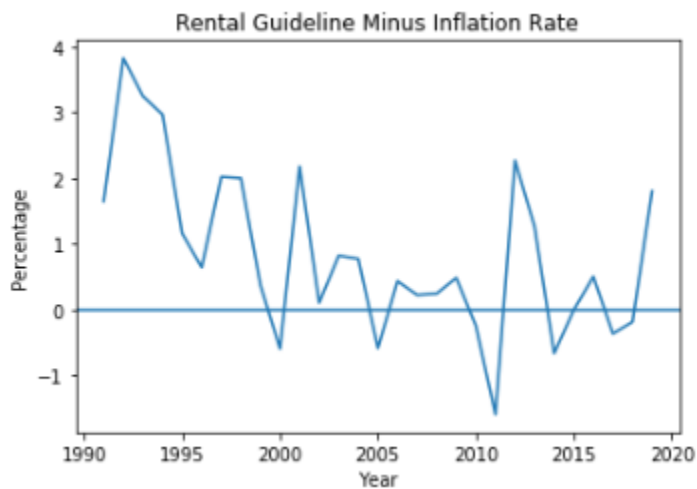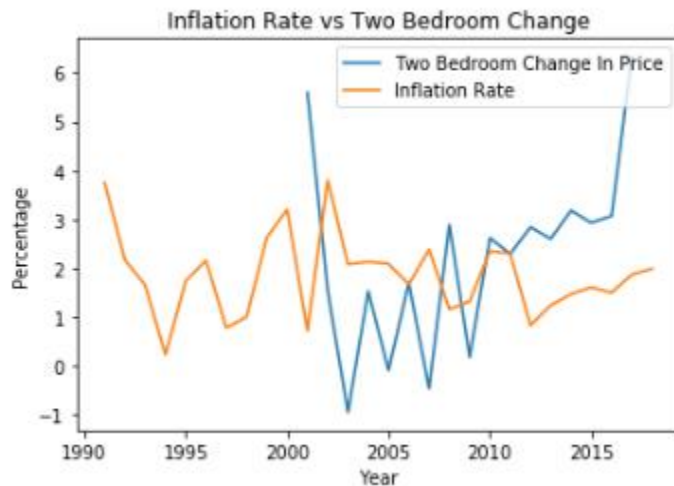
Next we compare the Inflation rate with the change in price in each type of dwelling, starting with the change in the price of Bachelor apartments. Here we can see that the Bachelor change in price is almost always greater than the rate of Inflation aside from a few exceptions where Inflation is equal to or lower than the Bachelor change. We see this at roughly 2002-early 2006, late 2007-2008, and 2009-early 2010.
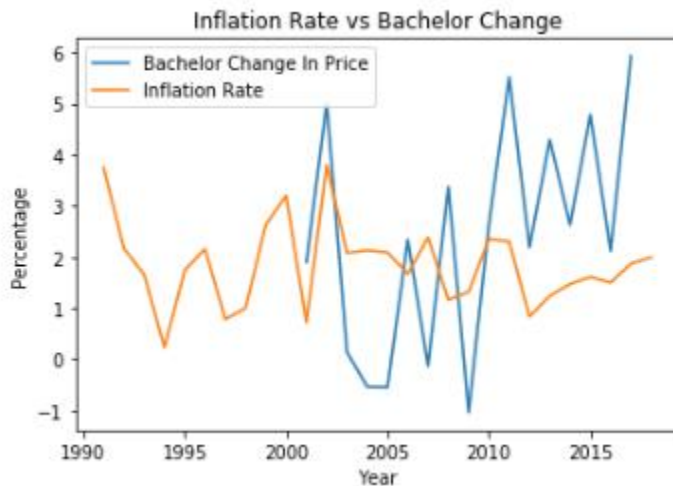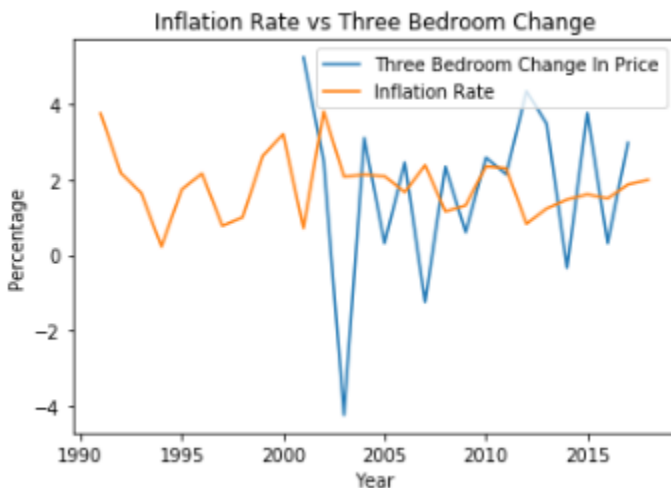

Below we are comparing the Inflation Rate with the change in price of One Bedroom apartments. Here we can see that the Inflation Rate was greater from about 2001-2008 and 2009-2010 and it was smaller otherwise. There is also a sharp increase in the change in price from about 2015 onwards.
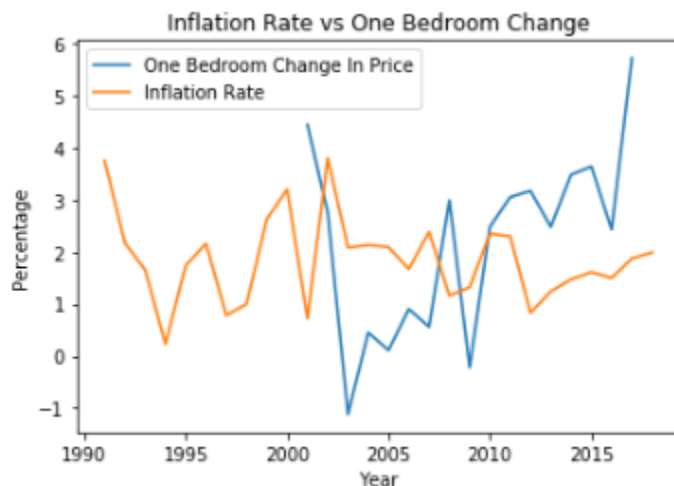
Below we are comparing the Inflation Rate with the change in price of Two-bedroom apartments. Here we can see there were many sharp increases and decreases in the change of price between 2001 and 2010. We can also see the Inflation Rate was greater from about 2001-2008 and 2009-2010 and it was smaller otherwise. There is also a sharp increase in the change in price from about 2015 onwards.


Below we are comparing the Inflation Rate with the change in price of Three-bedroom apartments. Here we can see there were many sharp increases and decreases in the change of price throughout, the biggest being the drop in 2003. Aside from a few exceptions we can see the Inflation Rate was greater or roughly the same from about 2003-2010. Aside from a few exceptions from about 2010 onwards the Inflation rate was smaller.

Below we are comparing the Inflation Rate with the change in price of the city average. Here we can see there were some sharp fluctuations in the change of price between 2001 and 2011. We can also see the Inflation Rate was greater from about 2001-2007 and 2009-2010, it was smaller otherwise. There is also a steep increase in the change in price from about 2015 onwards.
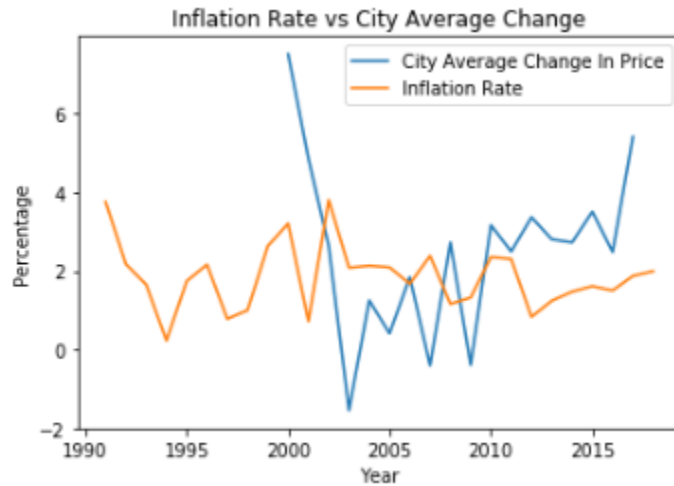
Finally, we are now plotting them all together to compare and get a visual on each individual apartment type alongside the city average, guideline, and inflation.
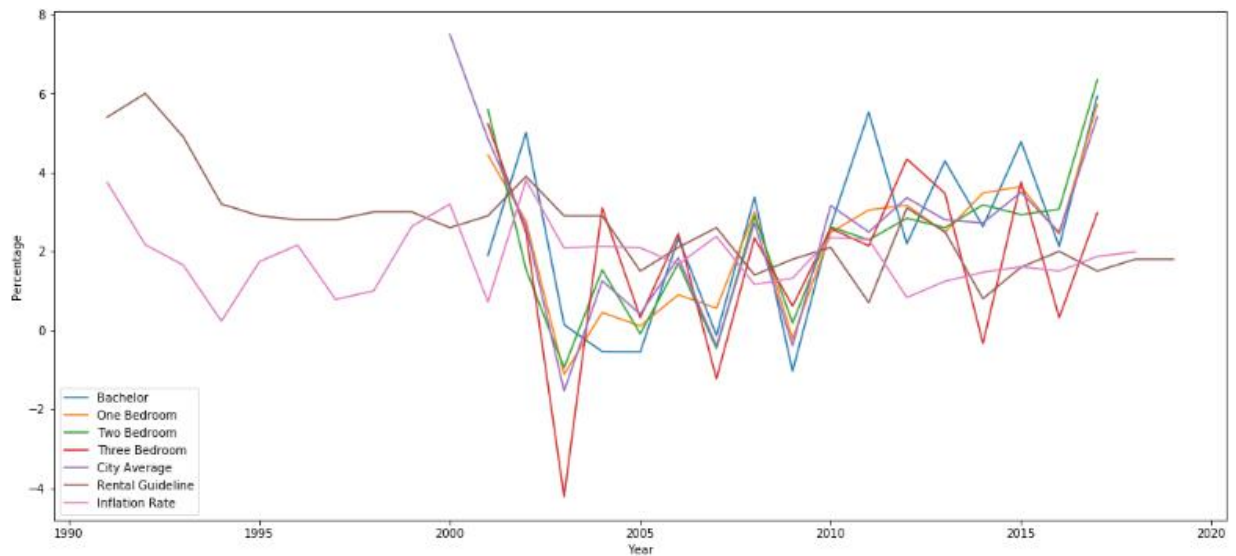
Sprint 2 Portion: Sprint 2 contains the modelling/analysis phase of the project and it continues from where Sprint 1 ended. At the start of the project the goal was to try to identify the marginal difference in rent price between tenured renters and new renters in Toronto, specifically looking at apartments. During the completion of the Sprint and project many gaps were identified, and this became increasingly more difficult. The Sprint contains an attempt at forecasting the average apartment rent price in Toronto. This can be very useful for prospective renters, landlords, and the City of Toronto. The code and rationale are in the notebook and the report.

Here we are trying to forecast future average price using ARIMA. First, we create a new subset for only year and City Average price and view it to make sure.

```
        CityAverage
Year
2000            907
2001            951
2002            976
2003            961
2004            973
2005            977
2006            995
2007            991
2008           1018
2009           1014
2010           1046
2011           1072
2012           1108
2013           1139
2014           1170
2015           1211
2016           1241
2017           1308
Int64Index([2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
            2011, 2012, 2013, 2014, 2015, 2016, 2017],
           dtype='int64', name='Year')
```

Before creating a forecasting model, we need to check for trends. Below we are plotting the Average Rent Price and we can see a general upward trend
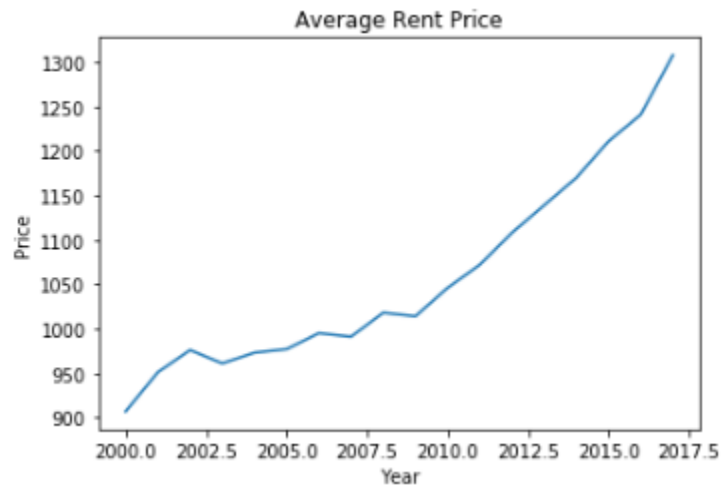
Next we try to detrend by model fitting using a linear model. We calculate trend, plot it, try to detrend, and then plot that as well, detrended being the bottom graph.
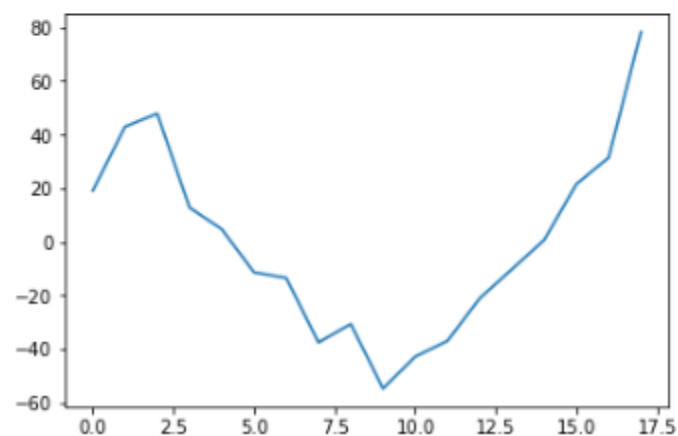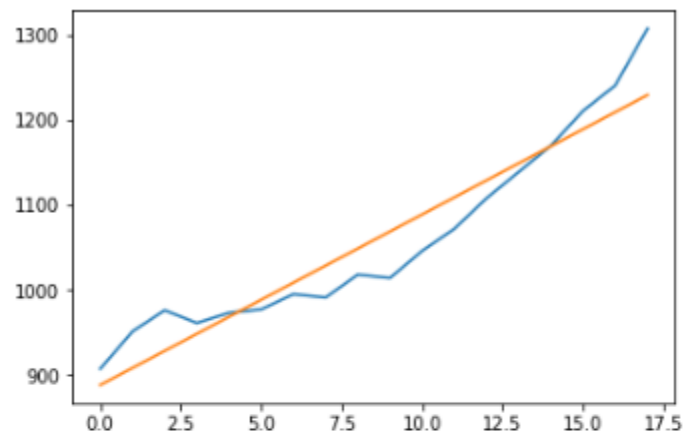
Next we plot the ACF and PACF with 95% confidence interval to help show us the lag.

Autocorrelation

Partial Autocorrelation
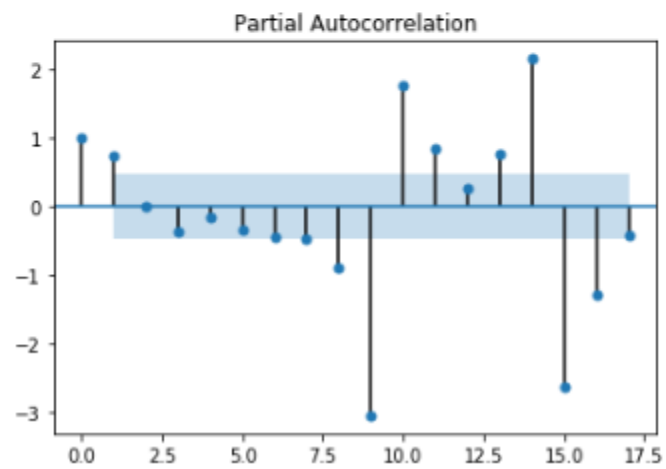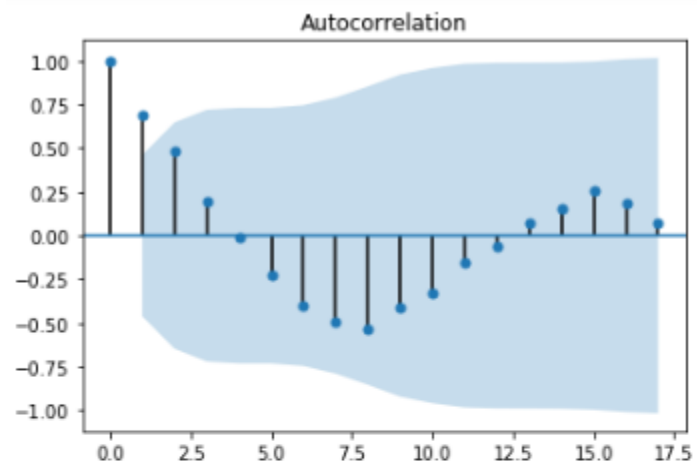
Finally, we create our ARIMA model and use it to predict. We use 70% of the data to train, and 30% to test the model. We also print the MSE to see how good the model is along with plotting the predictions against the actual. The prediction is red and the actual is blue, we can see that they intersect only once and are somewhat close, but not close enough.

```
Actual=1108.000000, Predicted=1079.699218
Actual=1139.000000, Predicted=1116.656325
Actual=1170.000000, Predicted=1161.014917
Actual=1211.000000, Predicted=1189.456927
Actual=1241.000000, Predicted=1239.477368
Actual=1308.000000, Predicted=1270.205296
Test Mean Squared Error: 545.961
```
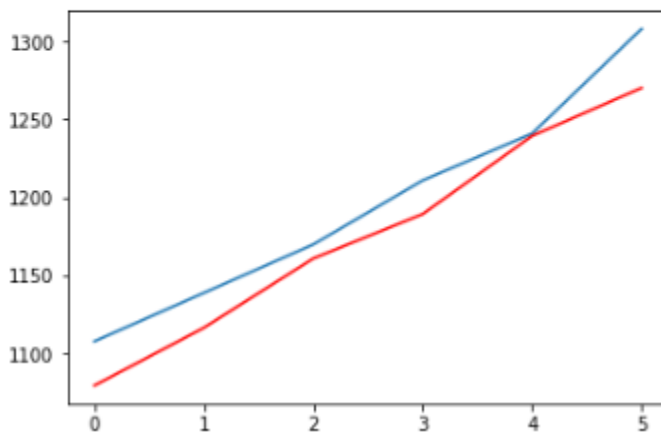

We can see the MSE is 545.96 dollars meaning the model is over or underestimating the average rent price by 545.96 dollars. This model is not the best as that is too large a number and should be improved before it is used.

-Perhaps there are stationary issues and the model is not fully stationary and needs to be better transformed, maybe stabilizing the variance, this can all be tested
-Perhaps multiple models with different AR and MA terms
-Perhaps other modelling techniques
-Perhaps more/better data

Although ARIMA does not consider economic and political conditions, or correlations of all factors related to rent prices perhaps after the example above is cleaned up it can be useful for forecasting the average apartment rent price in Toronto.

Final Conclusions and Recommendations:

Per the above the model created should not be currently used as is. The MSE is too high and the model is not very useful and outlined above were some potential solutions to help improve it along with why it may not be very useful.

Throughout the completion of the project there were many issues, obstacles, and challenges that were met along the way. Not all were completely resolved, and these were the key contributors to the many ways the project could be improved in the future (aside from the above issues with the actual model). These will be outlined below. Upon reflection it appears that the number one and biggest problem was the actual data that was collected and cleaned for analysis. The data collection and cleaning are vital as they lay the foundation for the project. If not done correctly it can cause major problems down the road like with this project.

- Generally, more data overall as well was more complete data.
- Higher quantity more complete/rich income data, perhaps with locations, education, etc and for more years.
- Data on the actual amount spent on rent, or actual rent expenditure, ideally in a dataset that includes geographical information, properties of dwelling, duration of stay, etc along with changes seen in the above characteristics. This was the key missing piece that was needed and could not be located and loaded to be properly used for analysis.
- Geographical data to better understand, identify and visualize areas as location is one of the biggest factors in housing prices.
- Generally, more practice with Geopandas and forecasting to help all the above.

Although the original research question was changed along the way and the final model was not particularly useful and although the final product was not ideal or what was expected **a lot** was learned! The above project was a unique and very rich learning experience! It encompassed a complete data analysis project from the very beginning to very end along with all the challenges that come with it.

Sources:

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110023901&pickMembers%5B0%5D=1.17&pickMembers%5B1%5D=2.1&pickMembers%5B2%5D=3.1&pickMembers%5B3%5D=4.1

https://www.inflation.eu/inflation-rates/canada/historic-inflation/cpi-inflation-canada.aspx

https://www.ontario.ca/page/rent-increase-guideline

https://www.torontorentals.com/blog/average-rent-in-toronto-since-2000#2008

http://geopandas.org/io.html#reading-spatial-data