CSDA 1050 Sprint 2

August 6, 2019

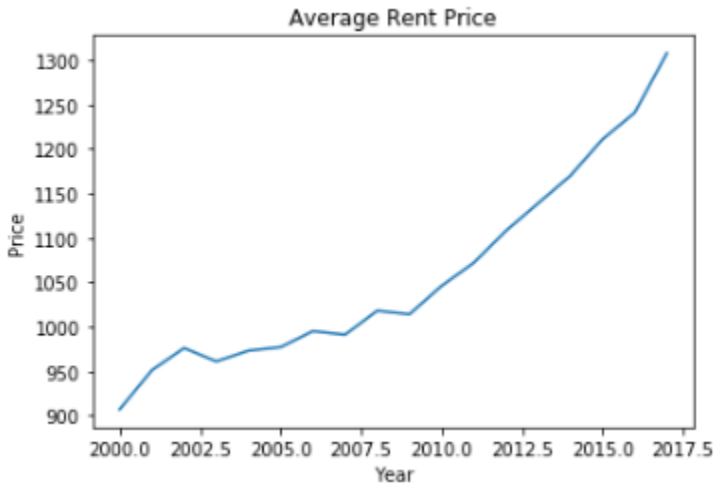
Nickolas Lal

Overview: 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.

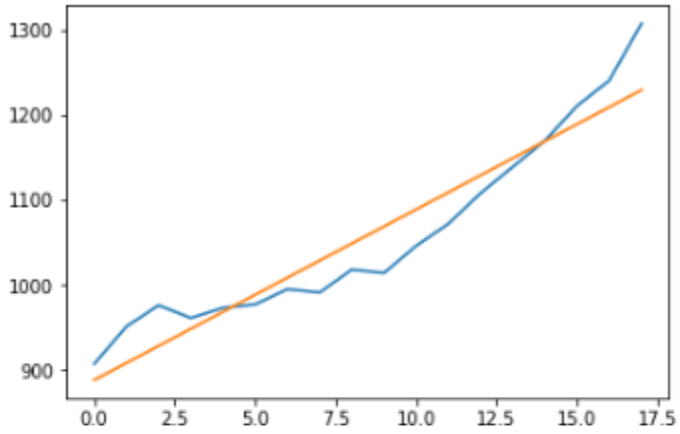
Here we are trying to forecast future average price using ARIMA. First, we create a new subset for only year and City Average price and view it to make sure.

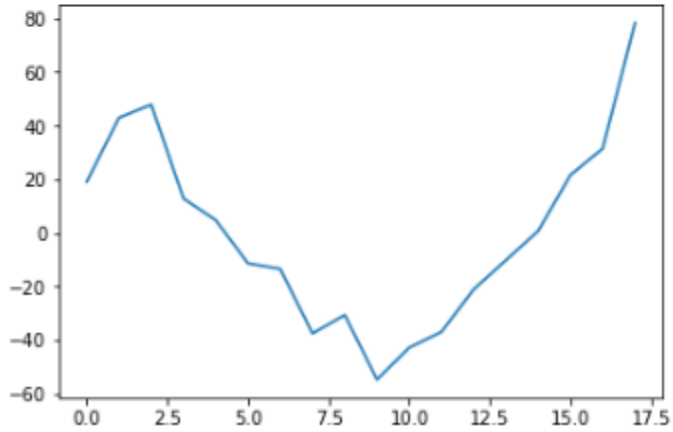
```
      CityAverage
Year
2000          907
2001          951
2002          976
2003          961
2004          973
2005          977
2006          995
2007          991
2008         1018
2009         1014
2010         1046
2011         1072
2012         1108
2013         1139
2014         1170
2015         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

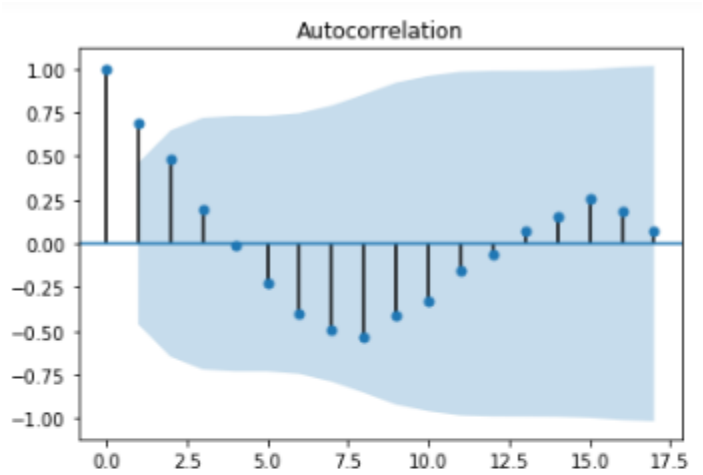


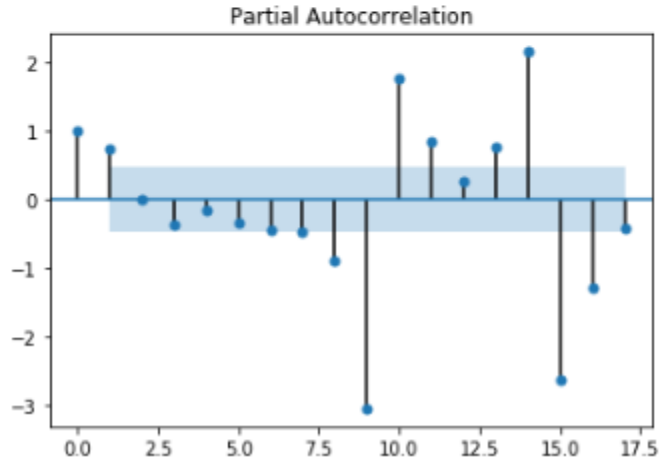
Next we try to detrend by model fitting using a linear model. We calculate trend, plot it, try to detrend, and then plot that as well, detrended being the bottom graph.





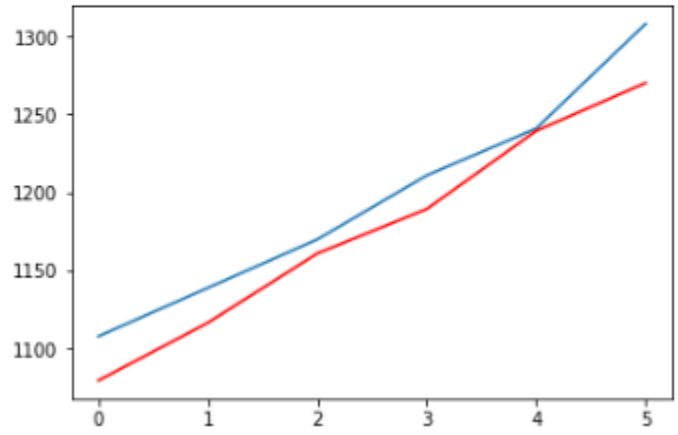
Next we plot the ACF and PACF with 95% confidence interval to help show us the lag.





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.

Next steps:

- Complete notebook
- Complete report