**Applied Data Science Capstone**

Capstone Project

– Fining a good neighborhood in Scarborough, Canada

1. **Introduction / Define a problem**

People want to live in good neighborhoods. One can say that the only thing he concerns is the place itself, but after few days of moving, he will soon get care about the infrastructure, safety and other considerations. As we can notice from Maslow's Hierarchy of Needs, seeking for a safe and secure place is human's basic needs. It also implies once these desires are not fulfilled, feeling an accomplishment, and obtaining your prestige would be behind the back.

And there are various positive effects when people live in a great neighborhood. First, good places promote intimacy. In a good neighborhood, it is more likely to talk into people you know, and you want to bring your friends and family to your place. Secondly, if you raise a child, neighborhood quality has significant and long-term effects on child and adolescent problem behaviors according to Johns Hopkins School. With all the other reasons, people want to live near the main shopping street, a park, a library, playground etc.

From this project, I would like to help those who want to move to Scarborough by looking up median housing price and crime rates of the area, weather conditions, good management of emergency, road connectivity and so on. This will be a good indicator to lead a good neighborhood.

1. **Data Description**

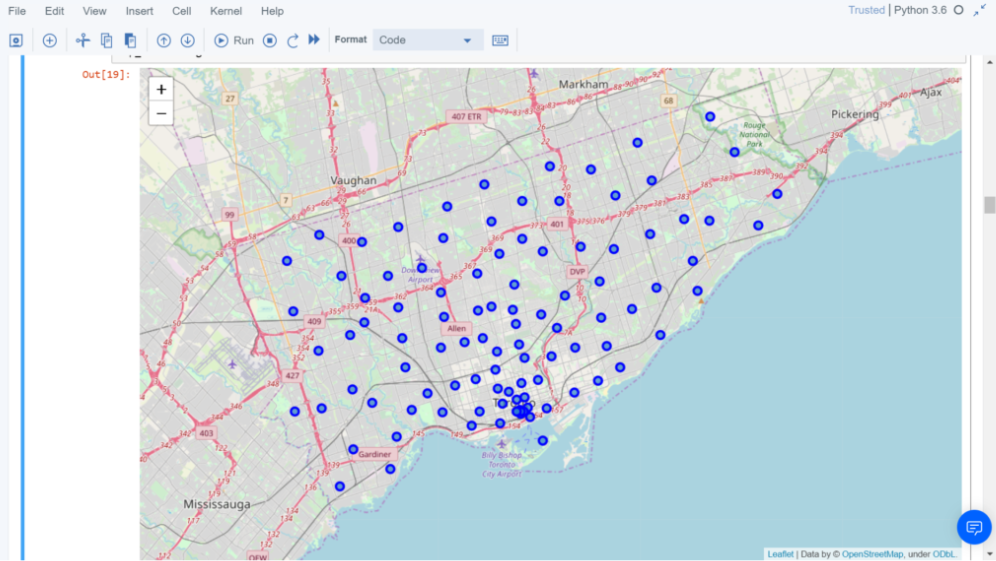
<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> This is the data I used for this project. Specially, I am going to use the Scarborough, Ontario dataset.

In order to get information about the area I'm looking for, I will use Foursquare location information. Foursquare is the most trusted, independent location data platform for understanding how people move through the real world. From this platform, longitude and latitude of the place can be obtained as well as venue, name.

1. **Methodology**

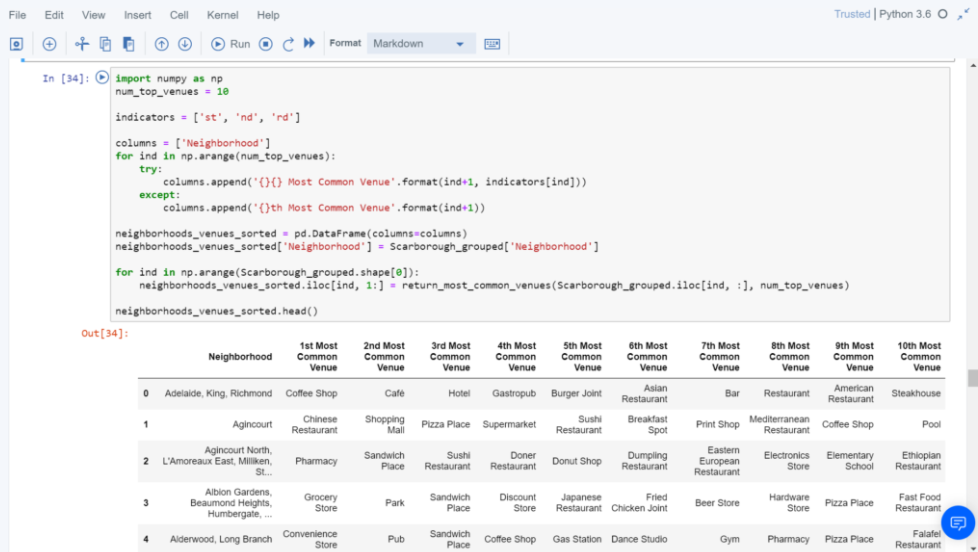
I am going to use Python to analyze the data in this project. Additionally, Foursquare, Numpy and Pandas are used to solve the problem.

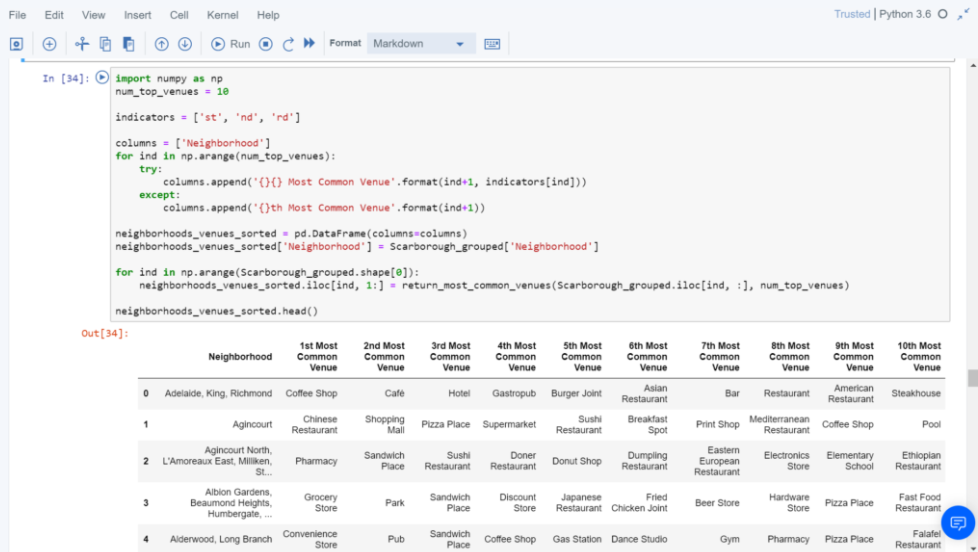
Following photo is a map of Scarborough



First step is to cluster the neighborhoods. We can collect the related information from the Wikipedia mentioned earlier. K-means help to make clusters. With this function, we can extract only needed information. Latitude, Longitude, and labeled clusters. The 10 neighborhoods can be shown using Numpy. Among 10 results, the first row of the result is neighborhood Adelaide. 1st Most Common Venue is Coffee Shop, 2nd is Café, 3rd is Hotel and so on.

Following shows clustering and k-means



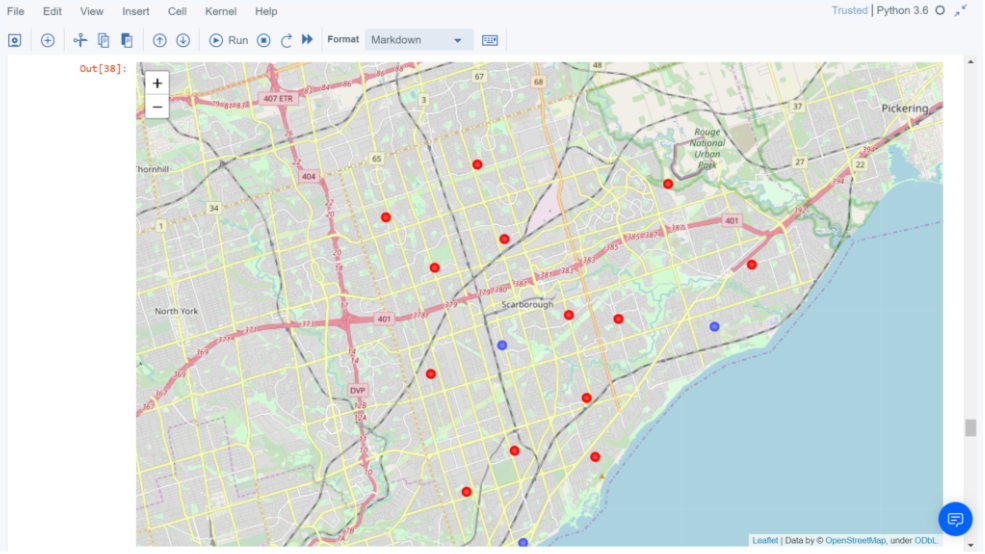


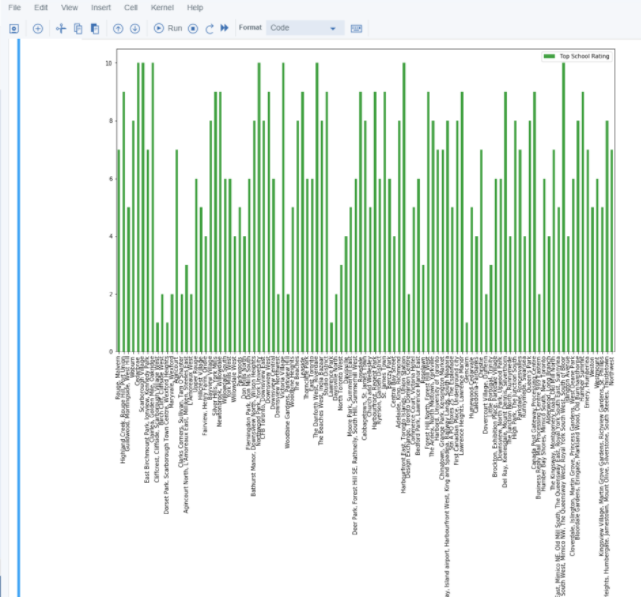
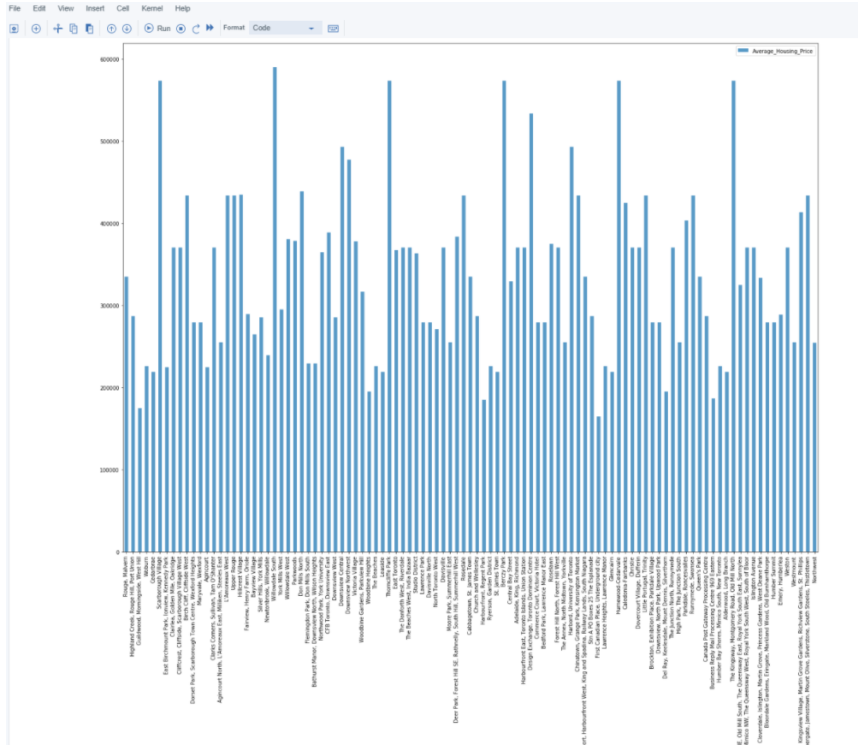
1. **Conclusion**

Using Foursquare API as the data gathering source since the number of the data is tremendous. And from this, it is possible to search a location, location sharing and details about the business as well as the review.

Scarborough is one of the popular cities where many immigrants in Canada to reside. Therefore, when someone wants to find a place with multi- culture, Scarborough is recommended to move. On the other hand, those people who prefer to avoid diversity, even though it is hard to avoid, the region would not fit. For the family with kids, they can decide based on the result of the school rating in Scaraborough.

Below photos are about clusters in Scarborough, average housing price and school ratings.





1. **Future work**

With more variables, such as the existence of the park, road congestion, and so on, we can get more accurate result.

It is proven that finding a real-world solution is possible from this project. Likewise, when it comes to find a conclusion in a huge dataset. Specially, the Data tools learned from the course will be useful.