**The Battle of the Neighbourhoods Project Report**

**Contents**

1. **Business Problem**
2. **Target Audience**
3. **Data**
4. **Methodology**
5. **Results and Outcomes**
6. **Conclusion**
7. Business Problem

In large multicultural cities one of the benefits for consumers is the amount of choice offered to them, Toronto is one of these cities. In densely populated cities like Toronto entrepreneurs face the difficulty of market saturation. In our example the problem is concerning coffee shops. For entrepreneurs who want to open a coffee shop in Toronto knowing what neighbourhood to open their coffee shop business in can be the difference between a coffee shop succeeding of failing.

This is the objective of this project is to help those looking to open a coffee shop in Toronto. This will be done by using Foursquare location data.

1. Target Audience

Given the nature of the project the audience can be any individual or organization looking to open a coffee shop in the Toronto area.

1. Data

Toronto Neighbourhood Data:

* Data Source: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* The link above contains information regarding the neighbourhoods in Toronto.

Location data:

* Data Source: <https://cocl.us/Geospatial_data>
* This data provides us with the coordinates of the corresponding neighbourhoods found in the Wikipedia page.

Foursquare API:

* Data Source: <https://foursquare.com/developers/apps>
* The Foursquare API provides us with the information regarding the venues. This will allow us to analyse the coffee shops in Toronto.

1. Methodology

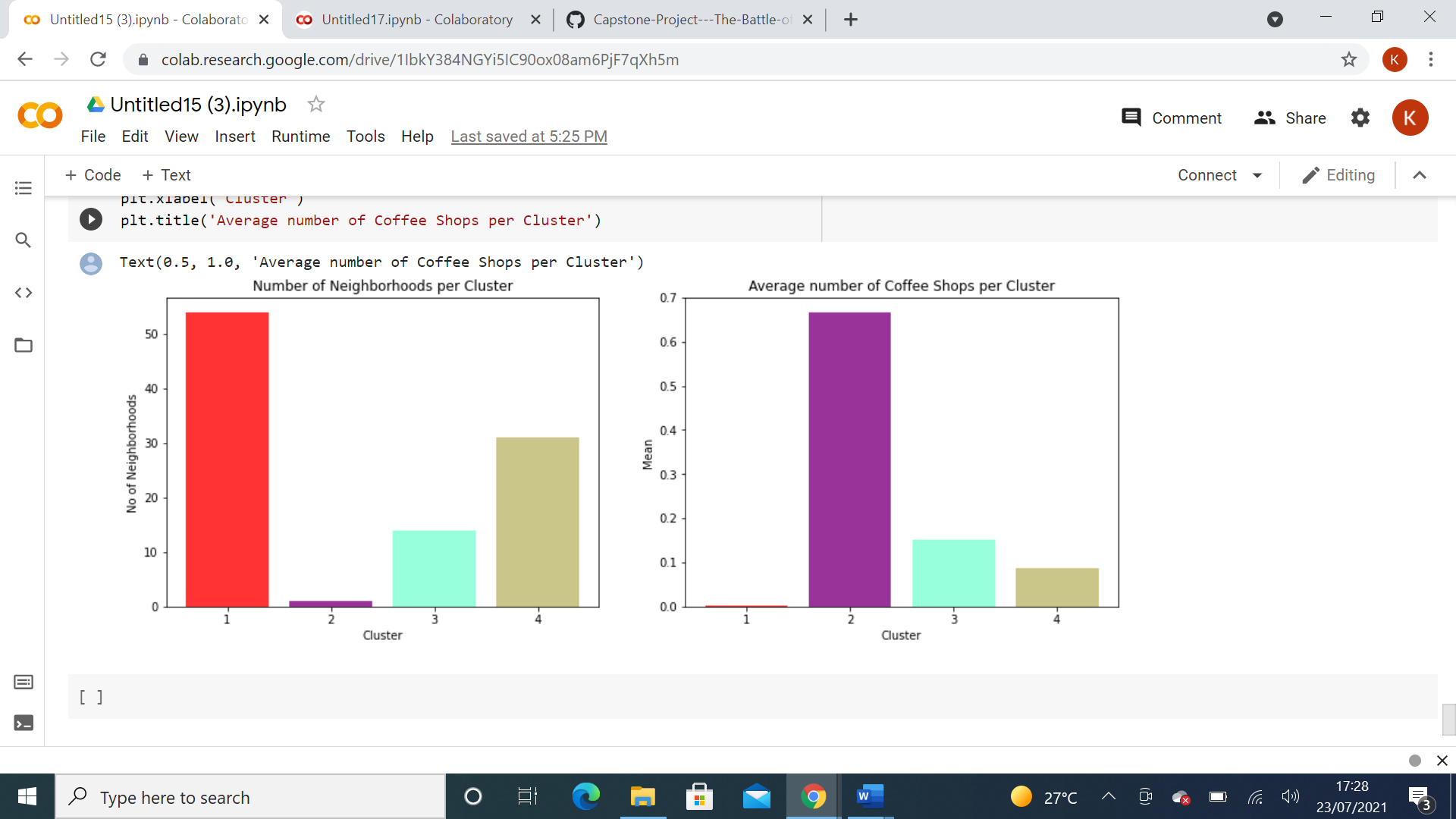
After cleaning and compiling the data into a format in which it could be analysed K-Means clustering was used on the neighbourhoods based on the frequency of coffee shops in Toronto. In order to execute this a K-Means clustering algorithm was used.

When using K-Means it is important to not overfit or underfit the model, to achieve this we would need an optimal K. To find this optimal we used the elbow method as can be seen in the code. The optimal value of K from the dataset was 4, thus the dataset will be clustered into 4 clusters.

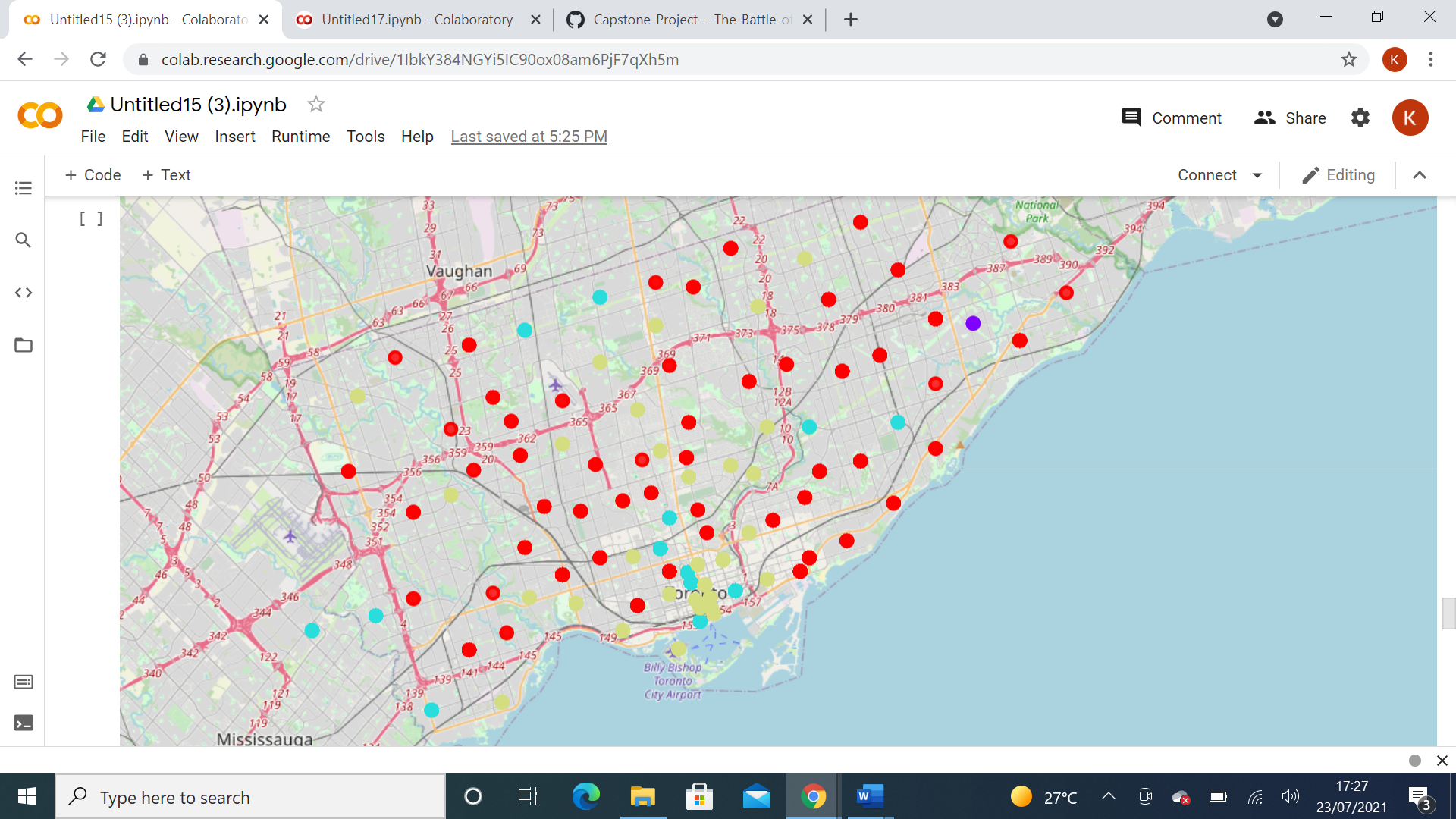
1. Results and Outcomes

The graphs below present the results of the project. Figure 1 shows the number of neighbourhoods per cluster and also the average number of coffee shops per cluster. While Figure 2 demonstrates the map of the clusters.

From the graphs one could deduce that opening a coffee shop in the red clusters would be best practice given the lower density of coffee shops in these areas, thus less competition for a new business.



Figure



Figure

1. Conclusion

To conclude, in the city of Toronto there are many coffee shops. For anyone wanting to open a coffee shop they may have been confused about where to open so they could be successful. The difficulty however is knowing which neighbourhood to open a coffee shop in.

Using Toronto location data from Foursquare the K-Means method could be applied to cluster these neighbourhoods. From this one could then work out the neighbourhoods with a lower density of coffee shops and therefore be able to know where is best to open a coffee shop to face less competition.