



# Capstone project

The best place to open a new pub

# Introduction I



1. Everyone loves beer, there is not question about that. But due to the current pandemic situation, many pubs are endangered by the governmental restrictions.
2. Limitations of pubs include: lowered capacity due to safety distance measures, reduced opening time, or in extreme cases closure.
3. We can expect that many pubs will go bankrupt and will never reopen.

# Introduction II



1. This, on the other hand, provides an opportunity for brave entrepreneurs, who could open a new pub after the pandemic will have been vanished from the world, and thus fill the empty space in the pub and beer industry.
2. We are here to help the potential pub-owners to choose the best location for their new pub!
3. Since we already have some background knowledge of the Toronto area, we will focus on this location.



# Data I

A close-up photograph of a bartender's hands pouring beer from a stainless steel tap into a clear glass. The bartender is wearing a dark shirt and a brown leather apron. A red and yellow wristband is visible on their left wrist. The background is blurred, showing other glasses and bar equipment.

We will use 3 data sources:

1. Toronto neighbourhoods dataset

2. Geocoordinates

3. Foursquare Database

## Data II



1. We will scrap the Toronto dataset from a Wikipedia page using pandas. It comprises of 3 columns: Postal Code, Borough, and Neighbourhood.
2. The geocoordinates dataset will be loaded from a geojson file and contains longitude and latitude values for each postal code.
3. Lastly, Foursquare database is a huge database containing lots of data about various places on the planet. We will need only information about venues. Each venue contains information such as: Name, Location, and Category. We will access the database via REST API.

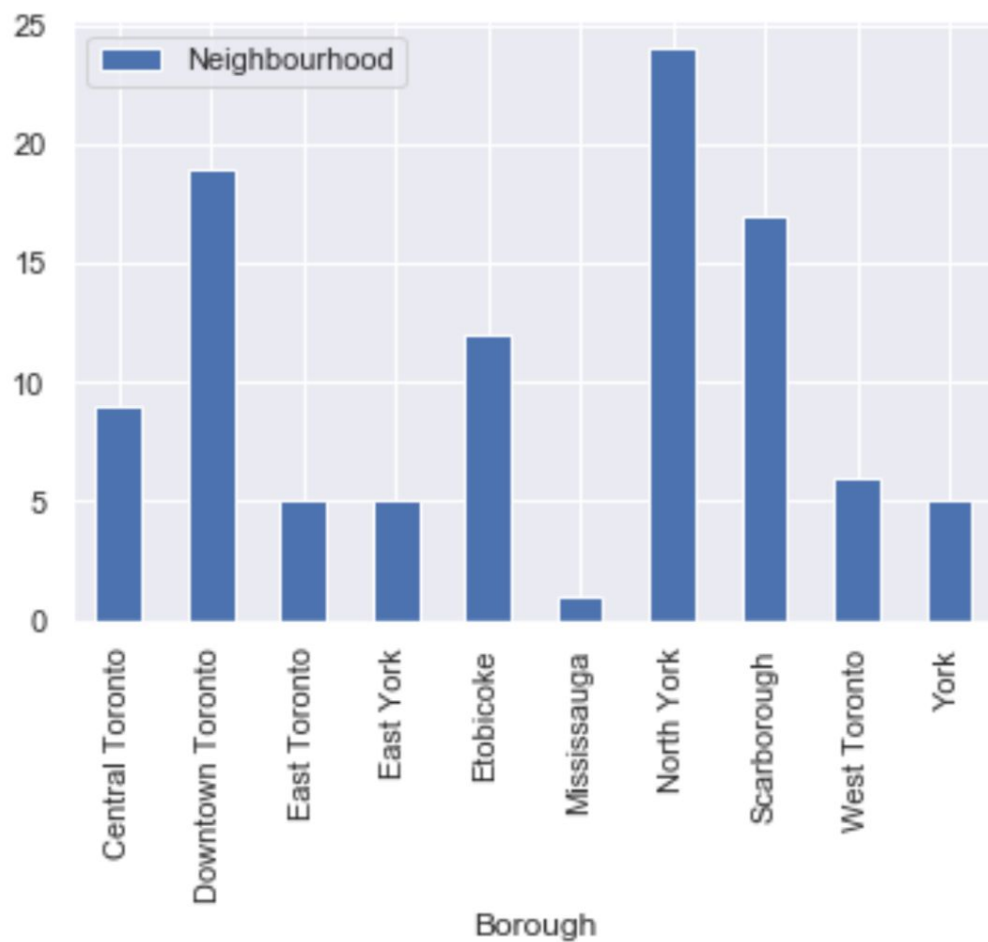
# Data III

A background image showing a close-up of a barista's hands pouring beer from a chrome tap into a clear glass. The barista is wearing a red wristband and a dark apron. The scene is set in a warm, dimly lit bar environment.

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
160	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
165	M4Y	Downtown Toronto	Church and Wellesley
168	M7Y	East Toronto	Business reply mail Processing Centre, South C...
169	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...



# Data IV







# K-means I

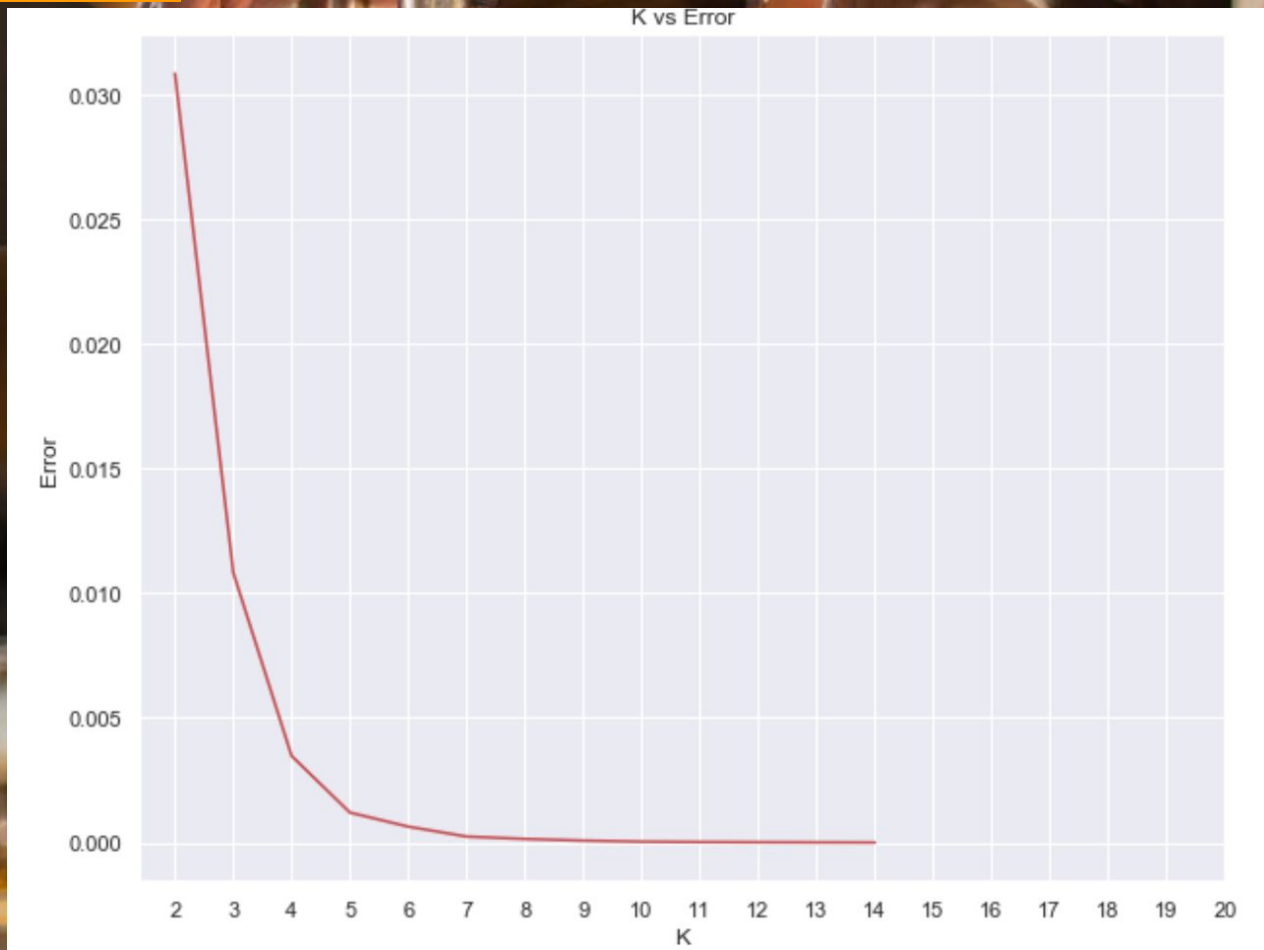
A close-up photograph of a bartender's hands pouring beer from a chrome tap into a clear glass. The bartender is wearing a dark shirt and a brown leather apron. A red and yellow wristband is visible on their left wrist. The background is blurred, showing other parts of the bar.

1. I used standard clustering algorithm K-means

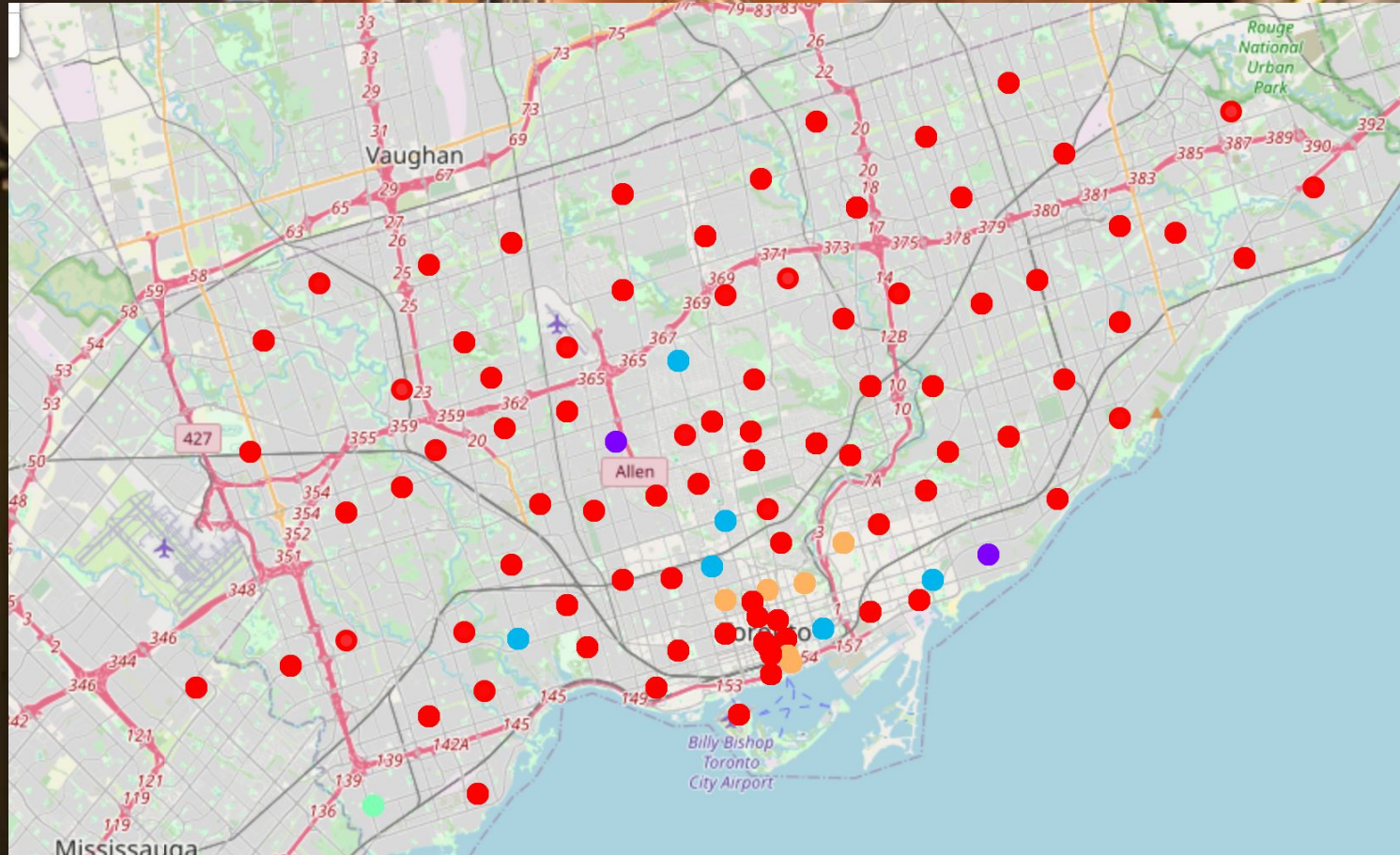
2. The average number of venues in a neighbourhood will serve as our clustering metric.

3. To choose the optimal number of cluster, I used the Elbow Method - which showed that it was 5

# K-means II

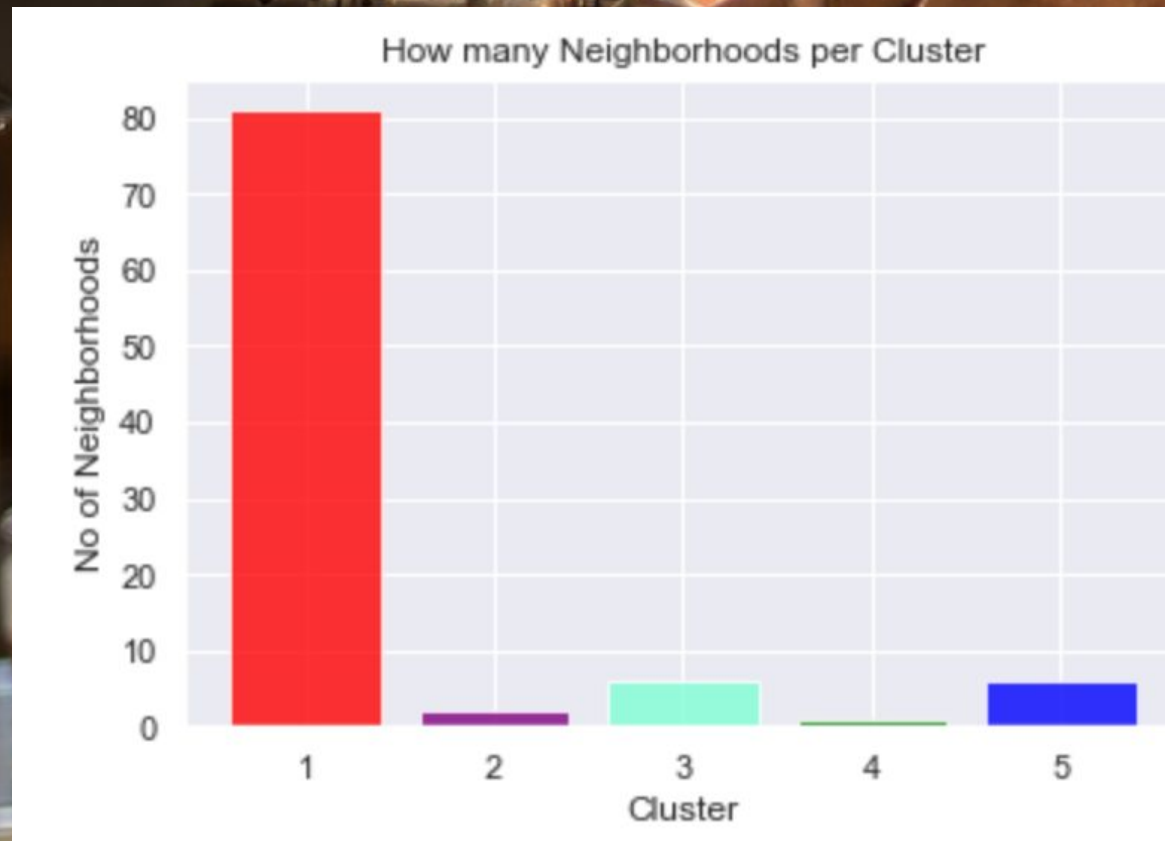


# Results I - Map with clusters

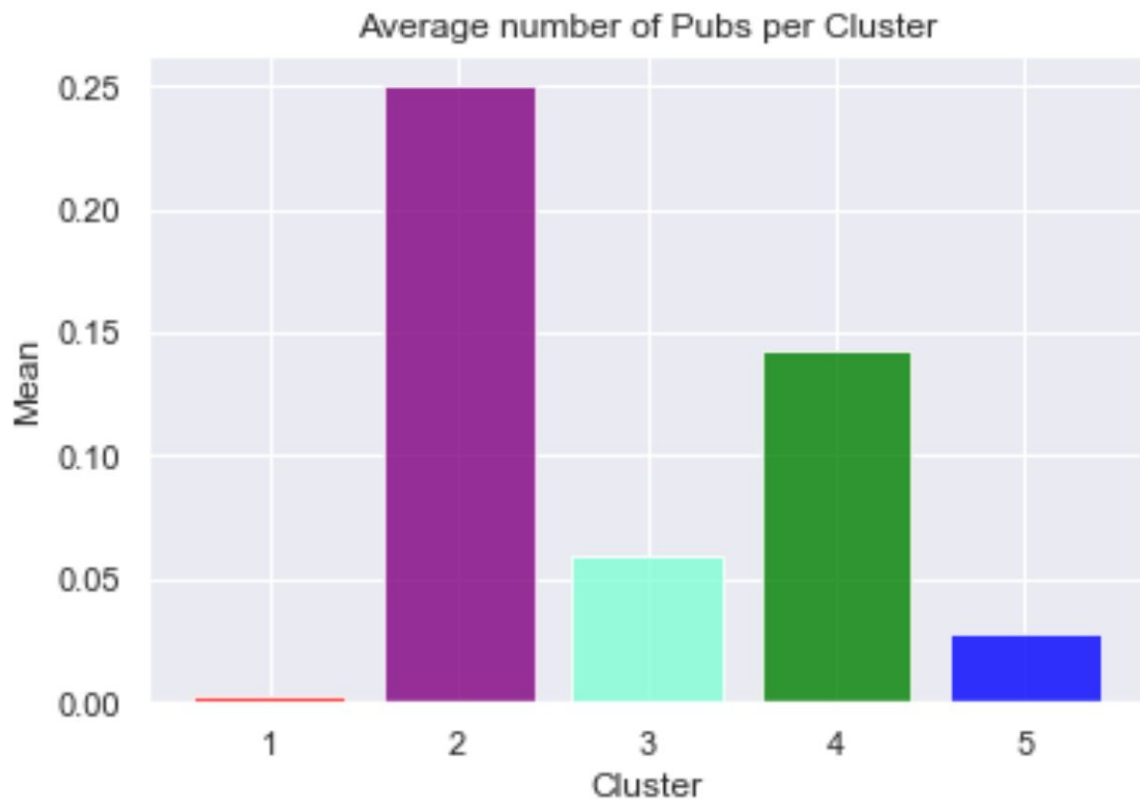




# Results II



# Results III



# Results IV - Summary

A close-up photograph of a bartender's hands pouring beer from a chrome tap into a clear glass. The bartender is wearing a dark shirt and a brown leather apron. A red wristband is visible on their left wrist. The background is blurred, showing other glasses and bar equipment.

**1. 5 Clusters with very different number of neighbourhoods**

**2. Clusters 2 and 4 have high number of average pubs**

**3. Clusters 3 and 6 have smaller number of average pubs**

**4. Cluster 1 has the smallest number**



A close-up photograph of a bartender's hands pouring beer from a metal tap into a clear glass. The bartender is wearing a dark shirt and a red wristband. The background is blurred, showing other parts of the bar and another person.

# Conclusion

1. I recommend choosing the 1. or 5. Cluster because of the results summarized on the previous slide
2. Newly set up pubs could prove highly profitable
3. The drawbacks of this project could be that data from Foursquare database might not represent the real situation in Toronto
4. Have a good beer and cheers! :)