

Coursera Capstone

IBM Applied Data Science Capstone

Real Estate Prices & Venues Data Analysis of Winnipeg City

By: Lilivett Ovalle Borrego

November 2020



Problem Statement

Winnipeg is the capital and largest city of the province of Manitoba in Canada. It is centred on the confluence of the Red and Assiniboine rivers, near the longitudinal centre of North America. The region was a trading centre for Indigenous peoples long before the arrival of Europeans. French traders built the first fort on the site in 1738. As of the Canada 2016 Census there were 705,244 people living in Winnipeg proper, with approximately 778,489 living in the Winnipeg Census Metropolitan Area (CMA). Thus, Winnipeg is Manitoba's largest city and Canada's seventh largest. Furthermore, it represents 54.9% of the population of the province of Manitoba, the highest population concentration in one city of any province in Canada.[1]

This boom means a strong interest by buyers and investors in the real estate market. Real estate investors look for locations where house prices are low and facilities (shops, restaurants, parks, etc.) and social spaces are close by. For an individual to find these locations in such a big city, and to gather such information, is a difficult task.

Considering these problems, we seek to create a map and an information chart of the real estate index of Winnipeg where each of its districts are clustered according to venue density.

Data Collection

To solve the problem, we will need the following data:

1. List of areas of Winnipeg with its neighborhoods and postcodes from www.geonames.org .[2]
2. Properties data such as address, price, postcode from the website www.point2homes.com [3]
3. Foursquare API to explore the boroughs of Winnipeg and segment them. [4]

Data Preprocessing

For the first table, the data scraped from www.geonames.org was cleaned and formatted.

	Neighborhood	PostCode
16	Winnipeg (Transcona)	R2C
20	Winnipeg (River East North)	R2G
22	Winnipeg (St. Boniface NW)	R2H
24	Winnipeg (St. Boniface NE)	R2J
26	Winnipeg (River East Central)	R2K

I removed all the columns with not relevant information and deleted the special characters. I only selected the location and the postcode columns.

	Neighborhood	PostCode
16	Transcona	R2C
20	River East North	R2G
22	St. Boniface NW	R2H
24	St. Boniface NE	R2J
26	River East Central	R2K

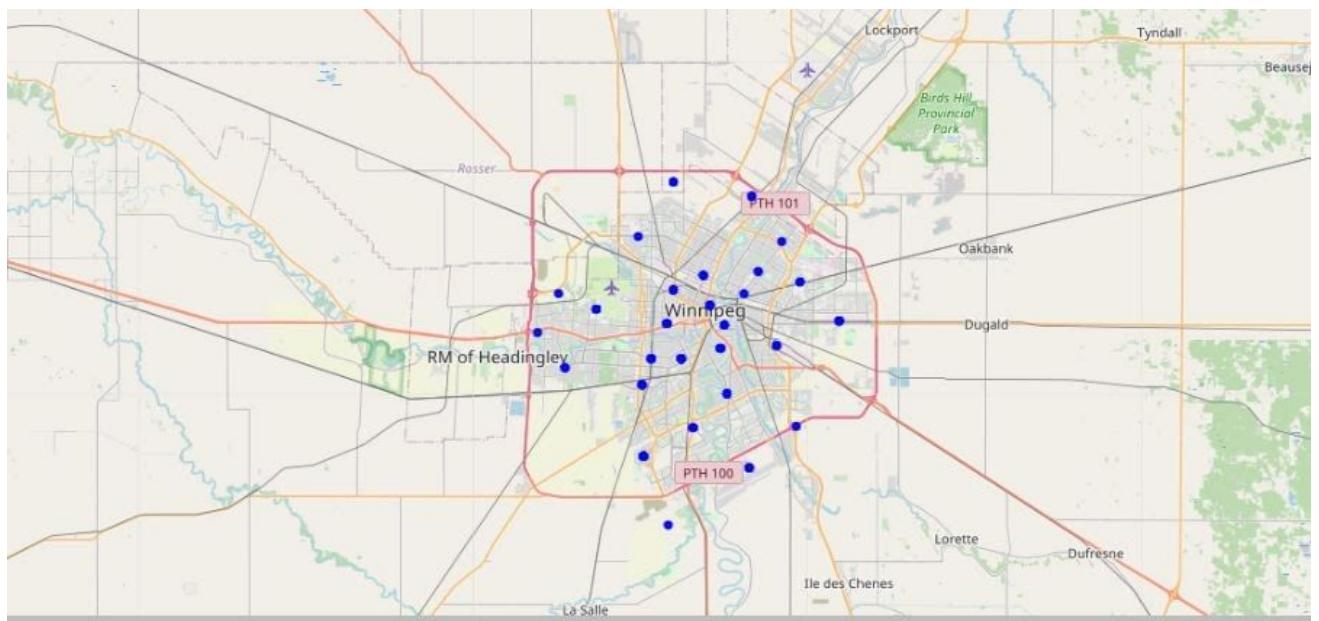
For the second table, the data was scraped from www.point2homes.com. I removed all null values and only kept the PostCode and Avg. Price columns. The Avg Price column contained strings so it was processed to convert to an integer by removing the '\$' special character and comma.

	AvePrice	PostCode
0	283873	R2H
1	271628	R3M
2	283873	R2H
3	271628	R3L
4	257606	R2M

After cleaning these two tables, I merged the information. Then by using the geocoder library I found the Longitudes and Latitudes of each Location and added new columns for each in my dataframe.

	Neighborhood	PostCode	AvePrice	Latitude	Longitude
0	Transcona	R2C	232143	49.88818	-96.99897
1	Transcona	R2C	384000	49.88818	-96.99897
2	Transcona	R2C	379900	49.88818	-96.99897
3	Transcona	R2C	264900	49.88818	-96.99897
4	Transcona	R2C	279900	49.88818	-96.99897

I used the **folium** python library to visualize geographic details from Winnipeg city and its boroughs, and I created a map of Winnipeg with the boroughs superimposed on top. I used latitude and longitude values to get the visual as below:



I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 10 venues. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

	name	categories	lat	lng
0	Royal George Hotel	Dive Bar	49.895275	-97.004188
1	Dairy Queen	Ice Cream Shop	49.895770	-97.002308
2	7-Eleven	Convenience Store	49.894793	-97.001976
3	Petro-Canada	Gas Station	49.893552	-97.003200
4	Silver Spike Saloon	Bar	49.894328	-97.005470
5	Rookies Sports Bar	Bar	49.894524	-96.999922
6	Pandora Inn	Bar	49.893598	-97.005659
7	Transcona Legion	Lounge	49.895187	-97.001252
8	East End Arena	Hockey Arena	49.894146	-96.991560
9	Whistle Pig	Diner	49.895675	-97.002378

I repeat the process for all location and venue count of each locations:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0 Transcona	49.88818	-96.99897	Transcona main shops	49.889102	-97.005703	Train Station
1 Transcona	49.88818	-96.99897	Transcona main shops	49.889102	-97.005703	Train Station
2 Transcona	49.88818	-96.99897	Transcona main shops	49.889102	-97.005703	Train Station
3 Transcona	49.88818	-96.99897	Transcona main shops	49.889102	-97.005703	Train Station
4 Transcona	49.88818	-96.99897	Transcona main shops	49.889102	-97.005703	Train Station

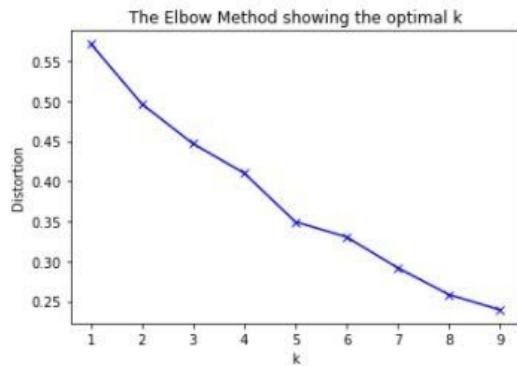
Finally, using the Foursquare API in conjunction with the created datasets, a new table with the most common visited venues in Winnipeg neighborhoods is generated.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Assiniboine South	Pizza Place	Gas Station	Massage Studio	Fast Food Restaurant	Falafel Restaurant	Dance Studio	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner
1 Chinatown	Coffee Shop	Dim Sum Restaurant	Bakery	Nightclub	Café	Hotel	French Restaurant	Sushi Restaurant	Diner	Historic Site
2 Fort Garry NE	Fast Food Restaurant	Chinese Restaurant	American Restaurant	Shopping Mall	Gas Station	Furniture / Home Store	Food Truck	Korean Restaurant	Diner	Department Store
3 Fort Garry NW	Dance Studio	Golf Course	Shoe Store	Sporting Goods Shop	Wine Shop	Falafel Restaurant	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner
4 Fort Garry West	Grocery/ Store	Pet Store	Coffee Shop	Park	Wine Shop	Falafel Restaurant	Department Store	Dim Sum Restaurant	Diner	Discount Store

Machine Learning

We have some common venue categories in boroughs. I used K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

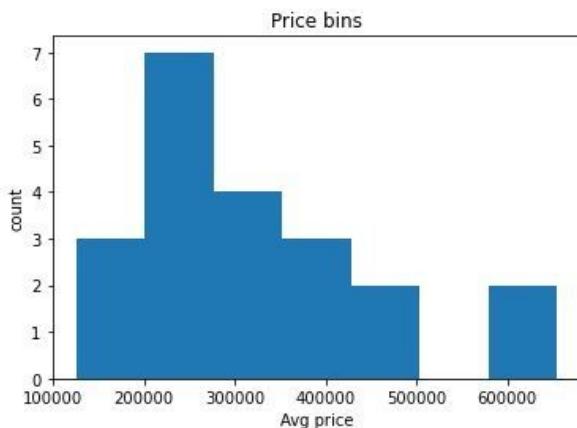
First, I will run K-Means to cluster the boroughs into 6 clusters because when I analyze the K-Means with elbow method it ensured me the 6 degree for optimum k of the K-Means.



Then, I merged tables with cluster labels for each borough. After examining each cluster I label each cluster as follows:

1. Mixed Social Venues
2. Hotels and Social Venues
3. Stores and seafood restaurants
4. Pubs and Historic places
5. Sports and Athletics
6. Restaurants and Bars

I also visualize house prices of each location:



As it seems in above histogram, we can define the ranges as below:

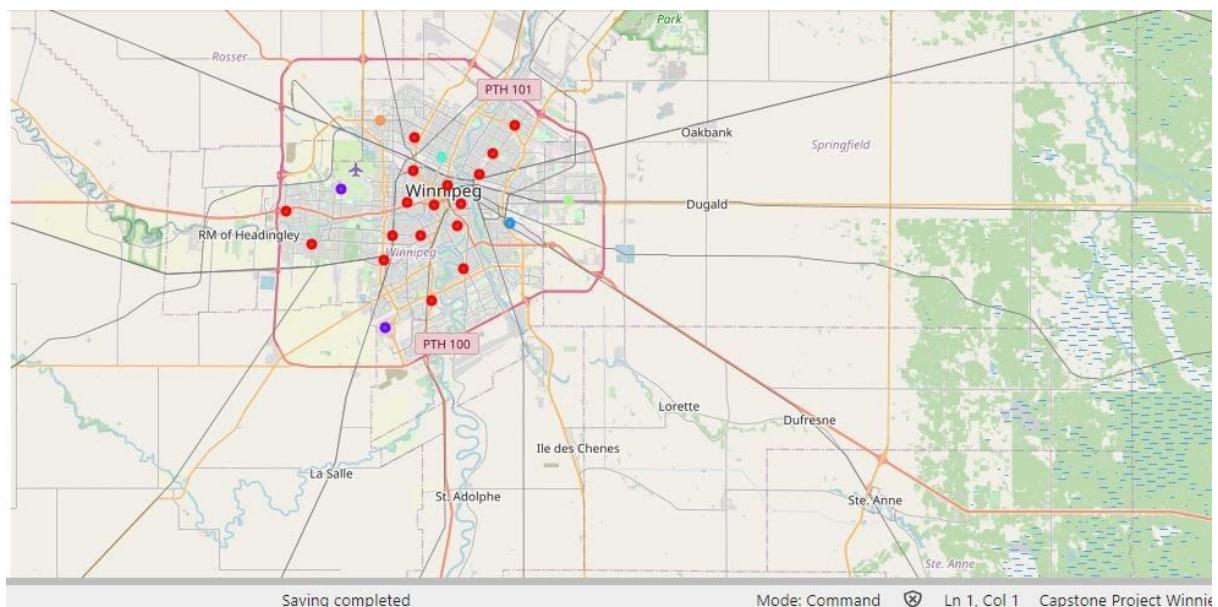
- <125000 : “Low Level 1”
- 125000–200700 : “Low Level 2”
- 200700–276400 : “Average Level 1”
- 276400–352100 : “Average Level 2”
- 352100–427800: “Above Average”
- 427800–503500 : “High Level 1”
- >579200 : “High Level 2”

Resulting table:

	Neighborhood	PostCode	AvePrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Price-Categories	Cluster-Category
0	Transcona	R2C	232143	49.88818	-96.99897	5	Train Station	Wine Shop	Fast Food Restaurant	Dance Studio	Deli / Bodega	Low level 2	Restaurants and Bars
1	River East North	R2G	184900	49.94218	-97.05900	0	Grocery Store	Ice Cream Shop	Juice Bar	Chinese Restaurant	Electronics Store	Low level 1	Mixed Social Venues
2	St. Boniface NW	R2H	283873	49.88606	-97.11885	0	Park	Pharmacy	Discount Store	Grocery Store	Coffee Shop	Average level 1	Mixed Social Venues
3	St. Boniface NE	R2J	249900	49.87144	-97.06417	2	Gym	Wine Shop	Fast Food Restaurant	Deli / Bodega	Department Store	Low level 2	Stores and seafood restaurants
4	River East Central	R2K	284900	49.92240	-97.08298	0	Gas Station	Business Service	Massage Studio	Sporting Goods Shop	Wine Shop	Average level 1	Mixed Social Venues

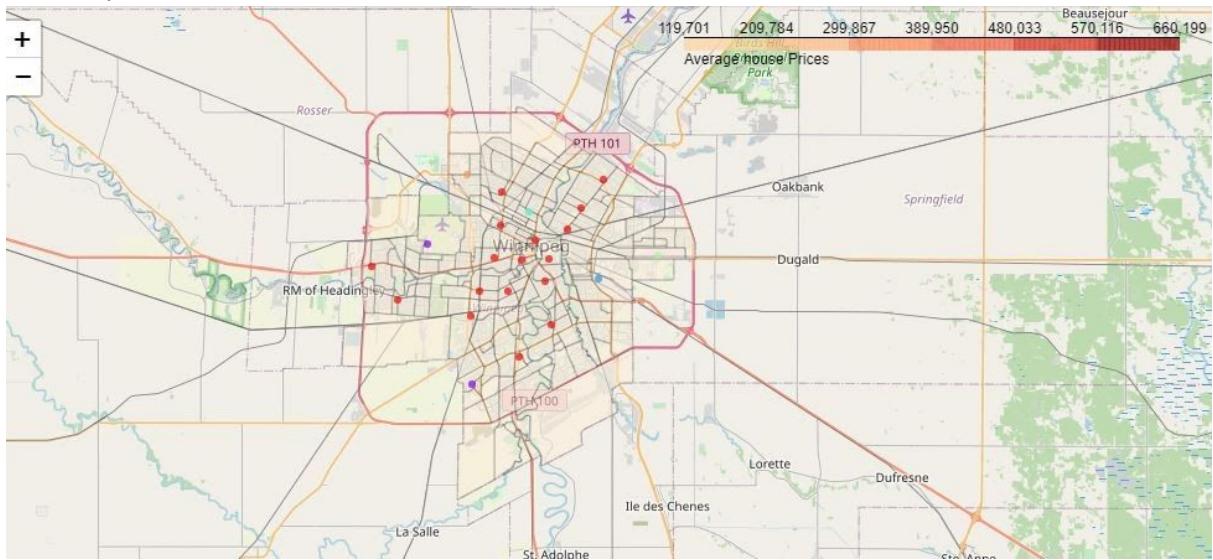
Visualization

First I visualize the cluster and you can see the clustered map below:



1. Red(cluster 0)
2. Purple(cluster 1)
3. Blue(cluster 2)
4. Cyan(cluster 3)
5. Green(cluster 4)
6. Orange(cluster 5)

Next I visualize the Average house pricing on the map by using Folium Chloropleth map and clusters on the top:



Result

As a final result we can see that the average home price in Winnipeg ranges from 200,000-3500,000 Canadian dollars. Some boroughs such as River Park South, Charleswood and Riverview have very high prices and are located mainly away from downtown and count with easy access to venues. While Transcona, Seven Oaks East, and Sargent Park boroughs have low property prices and are mainly north of downtown and somewhat close to venues.

Conclusion

In this project we explored, manipulated and analyzed the data that was scattered on the Internet about the city of Winnipeg. We were able to reach results that would help real estate investors to make strategic decisions, optimize their processes and maximize their profits.

References

[1] <https://en.wikipedia.org/wiki/Winnipeg>

[2] <https://www.geonames.org/postal-codes/CA/MB/manitoba.html>

[3] <https://www.point2homes.com/CA/Real-Estate-Listings/MB/Winnipeg.html>

[4] <https://developer.foursquare.com>