Edmonton Neighbourhoods Income Distribution Analysis

Part of
Applied Data Science Capstone Project

Date: 27.07.2021

1.Background / Introduction

Understanding your target market is undoubtedly one of the most important steps for any new or aspiring entrepreneur. Various demographic factors such as age, gender, race, income etc. play an important role in customer segmentation.
Among them, a customer's income gives a good measure of
the buying power of your audience.
When you know the income range of consumers in a neighbourhood,
you can design appropriate marketing strategies and find matching products for this population.

---

1. Problem Statement

Edmonton is the capital city of the province Alberta in Canada. This
project aims to analyze the neighbourhoods in Edmonton,
based on their income category and then predict the most preferred neighbourhood for starting a particular category of business, let's say a luxury business such as a spa or a yoga studio. The first part of the project aims to find the borough with most high-income neighbourhoods.
We also use Foursquare API and other data science tools to analyse these
wealthy boroughs and find out the common venues in these neighbourhoods.
We will then use k-means clustering
algorithm to segment the neighbourhoods and try to predict what would be the best category of business to start in any affluent neighbourhood.

---

2. Data used:

1.Edmonton -2016 Census - Population by Household Income (Neighbourhood/Ward)
.I have used the dataset that is publicly available on the website data.edmonton.ca
.Source : https://data.edmonton.ca/Census/2016-Census-Population-by-Household-Income-Neighbo/jkjx-2hix

Columns in this dataset and datatype

1 City Ward Number	Plain Text
2.Neighbourhood Number	Number
3.Neighbourhood Name	Plain Text
4.Less than $30,000	Number
5.$30,000 to less than 60,000	Number
6.$60,000 to less than 100,000	Number
7.$100,000 to less than 125,000	Number
8.$125,000 to less than 150,000	Number
9.$150,000 to less than 200,000	Number
10.$200,000 to less than 250,000	Number
11.$250,000 or more	Number
12.No Response	Number

2. List of centroids for current neighbourhood boundaries found within the boundaries of the City of Edmonton. Latitude and longitude data of the neighbourhoods have been obtained from the dataset available on https://data.edmonton.ca/City-Administration/City-of-Edmonton-Neighbourhoods-Centroid-Point-/3b6m-fezs/data

Note – this includes both residential and industrial neighbourhoods. Also, due to the configuration of a neighbourhood, the centroid may not necessarily lie properly within a neighbourhood boundary.

3. For the information about the Boroughs to which the neighbourhoods belong, I have used Wiki

I have created a CSV file of Boroughs based on the info available on the page

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Edmonton

> 3.1 Data Cleaning
>
> For simplicity. I have categorized the income into 3 classes; Low: Income till 60,000, Medium: Income 60,000-150,000, High – 150,000 and above. All Nan values were also removed for easier computation.
>
> Next step is to get information about the Boroughs to which the neighbourhoods belong from Wiki
>
> I have created a CSV file of Boroughs based on the info available on the page
>
> https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Edmonton

---

4.Methodology

We will first look at the census data of Edmonton which provides insights on the income distribution within Edmonton. We will then add the location data of each neighbourhood to the census data and create a new dataframe. Using basic exploratory data analysis, we will look at this dataset for common trends and patterns.

In the second part, using Foursquare API,
we will get details about the common venues within each neighbourhood. For simplicity purposes, we will only analyze the most affluent borough.

Finally, we will use k means algorithm to create clusters of neighbourhoods within the most affluent borough to analyze which are the most common categories of business within each neighbourhood.

---

5. Exploratory Data Analysis

I have categorized the income distribution into three brackets: Low income: Less than 60,000, Middle income: $60,000 to 150,000 and High income: Above $150,000 . I have also added two new columns 'Total Respondents' and 'Total Population'.

Dataframe containing income data was merged with the dataframe containing borough information for the purpose of analysis.

Top 2 High Income Neighbourhoods in each Borough

Central: Oliver, Downtown
North: Canossa, McLeod
North East: Hollick Kenyon, Klarvatten
North West: Inglewood , WestMount
South East: Summerside, Ottewell
South West: Rutherford, Twin Brooks
West: The Hamptons, Parkview

Income distribution by Boroughs – Southwest Edmonton has the largest number of high-income families.

Borough	Neighbourhood Number	Less than $30,000	$30,000 to less than $60,000	$60,000 to less than $100,000	$100,000 to less than $125,000	$125,000 to less than $150,000	$150,000 to less than $200,000	$2
Central	164336	8824	8832	7592	2930	1728	1559	
East	6380	0	0	0	0	0	0	
North	37606	482	711	774	352	217	173	
North East	166439	4840	7947	7826	3684	2170	1861	
North West	200545	3423	5598	5850	2619	1827	1648	
South	17439	0	0	0	0	0	0	
South East	453788	4396	8230	9711	4561	2771	2550	
South West	386535	4677	8222	9742	4771	3271	3730	
West	268264	3266	6571	6504	3333	2191	2201	

Southwest Edmonton has 73 neighbourhoods in total. We have chosen to analyse only this borough in details for our project.

Top 5 High Income Neighbourhoods in South West Edmonton: Rutherford, Twin Brooks, Windermere, Haddow, Terwillegar Towne.

Among this **Windermere** has the largest number of families earning 250,000 or more which shows that there is a small class of ultra rich people in this neighbourhood. So, Windermere is a good choice for starting a luxury business.

	Neighbourhood Number	Neighbourhood Name	Less than $30,000	$30,000 to less than $60,000	$60,000 to less than $100,000	$100,000 to less than $125,000	$125,000 to less than $150,000	$150,000 to less than $200,000	$200,000 to less than $250,000
0	5454	RUTHERFORD	130	368	621	334	255	273	99
5	5511	TWIN BROOKS	73	198	316	201	151	200	9
25	5570	WINDERMERE	37	144	271	137	151	150	74
47	5610	HADDOW	49	97	171	129	93	146	72
54	5640	TERWILLEGAR TOWNE	52	147	265	175	135	154	73

Using Foursquare API to further study Southwest Edmonton, we found the most common venues in each neighbourhood in this borough. We then used K Means clustering to create five clusters of neighbourhoods with similar behaviour

---

6.Results and Discussion

The objective of the business problem was to help new entrepreneurs study the income patterns of neighbourhoods in Edmonton and provide insights about the buying power to new entrepreneurs who wish to start a new business in Edmonton. This has been achieved by making use of Edmonton's census data from 2016.We also tried to identify the most affluent borough in Edmonton for starting a business such as a luxury spa or yoga studio. After selecting the borough, using Foursquare API we studied the neighbourhoods in this borough to identify the common business categories already present. We achieved this by grouping the neighborhoods into clusters using k means algorithm and identifying commonalities within each cluster.

7.Observations and Conclusions

1 South West Edmonton has the largest population of affluent class population and can be chosen to start any luxury business.

2.Central zone is not ideal for starting a luxury business since many of the people there fall in low-income category.

3.Cluster 2 has many yoga studios and gyms.

4.Neighbourhoods in cluster 4 seems to be saturated with construction and landscaping business and are not ideal locations for such categories of business.

5.Similarly cluster 3 already seems to be saturated with many restaurants and food joints and does not seem to be preferred for anyone wishing to start a restaurant business.

We have explored the census and income data of Edmonton to understand the neighbourhood of Edmonton and later categorized them into different borough. On analyzing, it has been found that Southwest Edmonton is the most affluent borough in Edmonton and Windermere and Rutherford are the most favourable locations for any new entrepreneur who wishes to start a luxury business.