2021

Analysis of the commercial retail, hairdressing, Cosmetology, Beauty sectors in Chicago

Clarence Mgutshini

IBM CAPSTONE PROJECT

8/25/2021

1. Introduction

1.1 Background

Chicago is the third largest city in the USA in terms of population. Chicago is considered a world- class city with a highly diversified economy. The economic activities consist of financial services, commerce, industry, education, hairdressing, cosmetology, beauty, telecommunications and transportation. Chicago city is a vibrant city to consider for careful and targeted development of either individual stores or shopping centres for commercial retail, hairdressing, cosmetology and beauty.

1.2 Problem Statement

The increased rise in the rate of urbanisation requires appropriate planning in the development of commercial retail and strategies in establishment and development of new retail, hairdressing, cosmetology and beauty stores. There is a need to find the areas that are highly competitive in terms of commercial retail, hairdressing, cosmetology and beauty stores in the city of Chicago and to identify areas with a low number of these stores in order to identify areas for potential development.

1.3 Interest

Property developers of new shopping centres consisting of commercial retail, hairdressing, cosmetology and beauty stores will be interested in the location information of areas with high concentrations of these stores for the purpose of competitive analysis and to establish the key reasons and factors for these areas being desirable for development of these stores. Furthermore, market researchers and strategists will be

interested in areas with a low concentration of these stores for the purpose of conducting research into opportunities of possible growth in these under- represented areas.

2. Data Acquisition and Processing

The data for the list of the neighbourhoods in Chicago was web-scrapped from the Wikipedia page with the url given by https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago . The beautiful soup package in python was used for web-scrapping the table data. There were 247 Neighbourhoods in Chicago. The output of the web-scrapping process is given in the table below:

[5]:

	Neighborhood	Community area
0	Albany Park	Albany Park
1	Altgeld Gardens	Riverdale
2	Andersonville	Edgewater
3	Arcadia Terrace	West Ridge
4	Archer Heights	Archer Heights
...	...	...
242	Wildwood	Forest Glen
243	Woodlawn	Woodlawn
244	Wrightwood	Ashburn
245	Wrightwood Neighbors	Lincoln Park
246	Wrigleyville	Lake View

247 rows × 2 columns

Foursquare location data through its API, is used to acquire the 100 most common venues in the city of Chicago and to get venues near each neighbourhood. The geocoder package was then used for scrapping the data of all the neighbourhoods in Chicago. Finally, the data was combined in a single data-frame which is given by the output table below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Albany Park	41.96829	-87.72338	Lawrence Fish Market	41.968280	-87.726250	Seafood Restaurant
1	Albany Park	41.96829	-87.72338	Chicago Kalbi Korean BBQ	41.968314	-87.722771	Korean Restaurant
2	Albany Park	41.96829	-87.72338	Starbucks	41.968911	-87.728817	Coffee Shop
3	Albany Park	41.96829	-87.72338	El Gallo Bravo #6	41.968324	-87.721338	Mexican Restaurant
4	Albany Park	41.96829	-87.72338	Rojo Gusano	41.968425	-87.724549	Taco Place
...	...	...	...	...	...	...	...
6340	Wrigleyville	41.94725	-87.65320	CTA Bus Stop 1825	41.944617	-87.654612	Bus Station
6341	Wrigleyville	41.94725	-87.65320	Chase ATM	41.944031	-87.653679	ATM
6342	Wrigleyville	41.94725	-87.65320	CTA Bus Stop 5361	41.947426	-87.647445	Bus Station
6343	Wrigleyville	41.94725	-87.65320	CTA #152 Bus Stop at Addison & Racine	41.947506	-87.658993	Bus Station
6344	Wrigleyville	41.94725	-87.65320	Newport Harbor	41.944195	-87.657616	Harbor / Marina

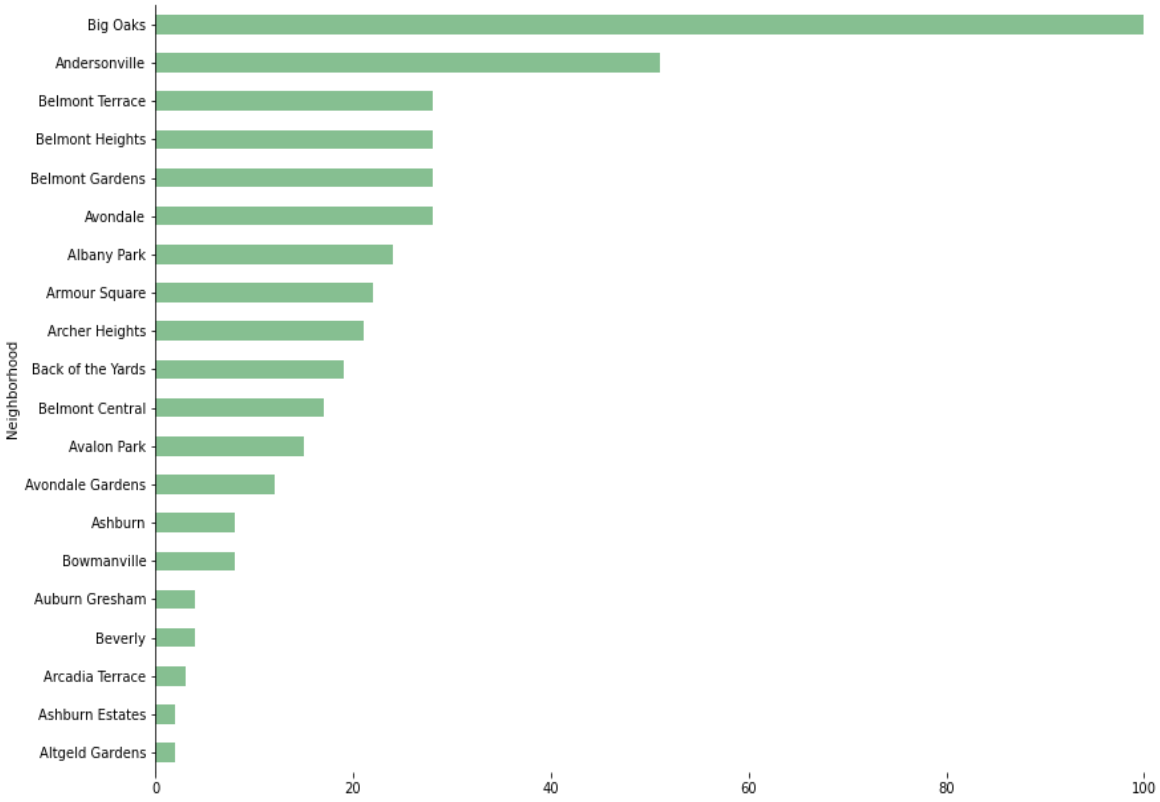
6345 rows × 7 columns

3. Methodology

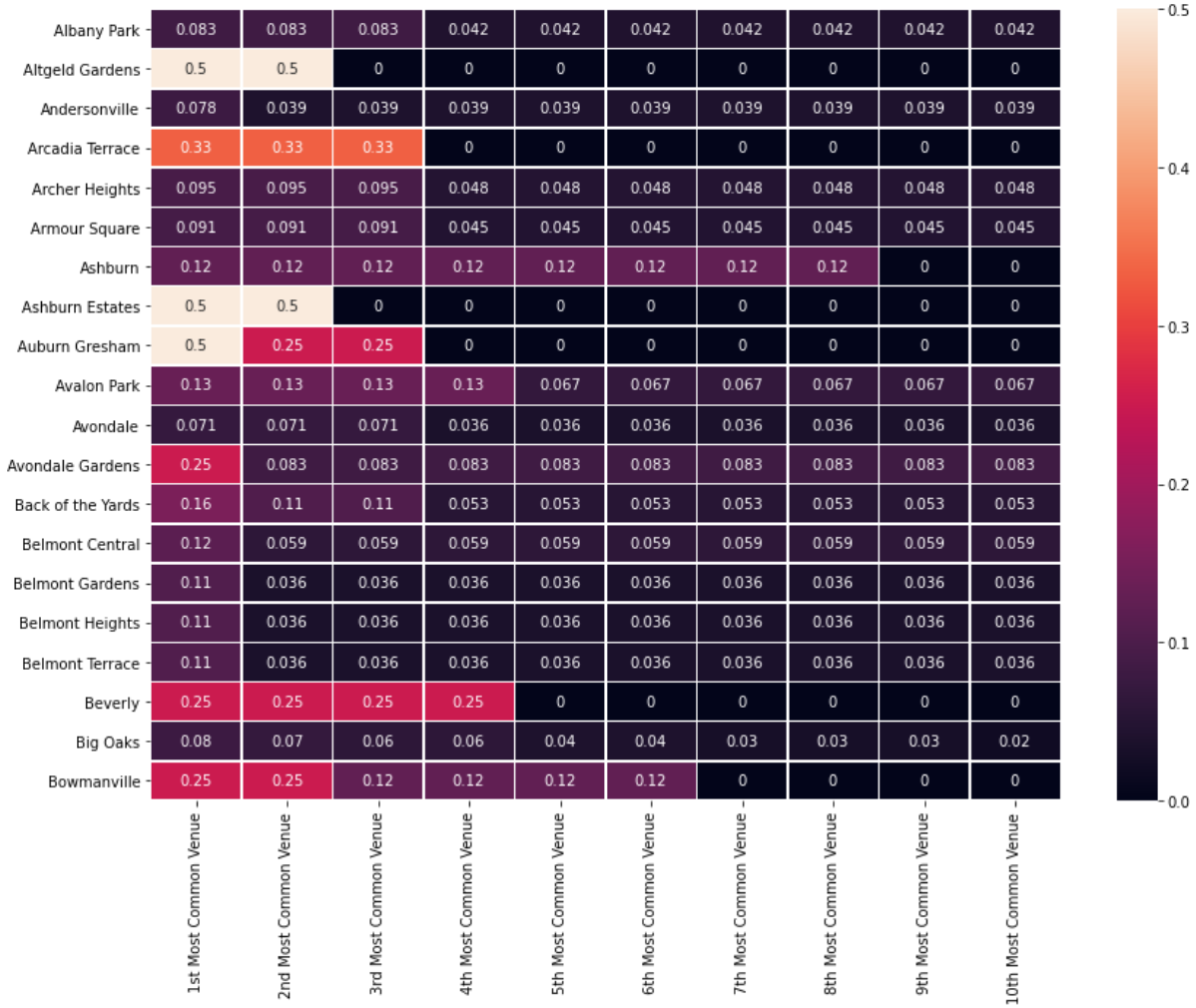
The database for storage and sharing of data in the case study was Git Hub. The code implemented in the study was shared through Git Hub.

3.1 Exploratory Data Analysis

The top 20 neighbourhoods in Chicago with the highest number of venue categories were considered for exploration. The bar graph below shows the 20 neighbourhoods in Chicago with the highest number of venue categories:



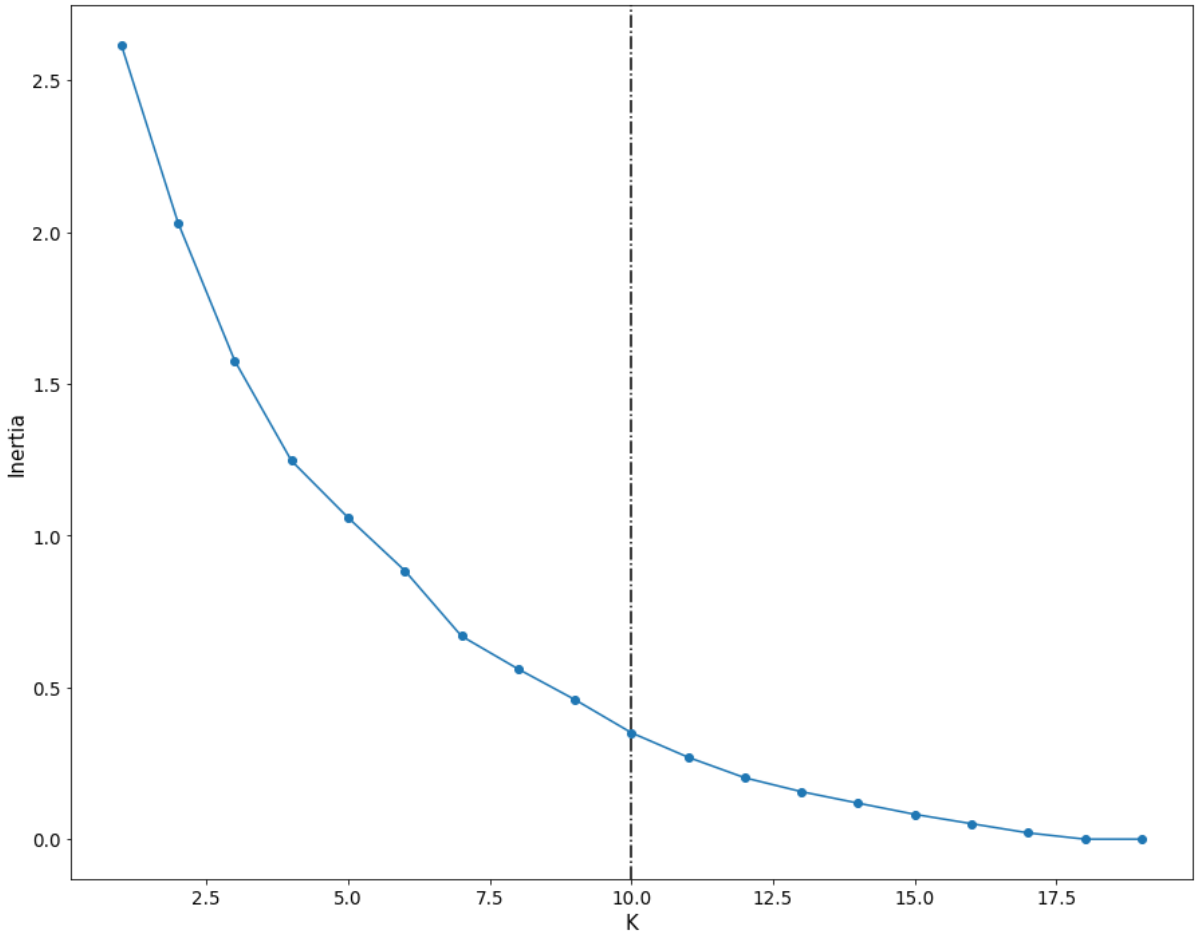
The Big Oaks neighbourhood area had the highest number of venue categories with 100 venue categories followed by Andersonville with 51 venue categories and Belmont Terrace with 28 venue categories. In the top 20 neighbourhoods by number of category venues Ashburn Estates and Altgeld gardens had the lowest number of venue categories with only 2 venue categories. The heat map below shows the distribution of the ten most common venue categories in each neighbourhood.



The heat-map confirms that the distribution of the venue categories was skewed towards to the most commonly occurring venues.

3.2 Cluster Analysis

K-means clustering, which is a classification machine learning algorithm was used to classify the top 20 neighbourhoods by venue categories in Chicago city into clusters based on the ten most commonly occurring venue categories. The number of clusters used for the K-means clustering was determined using an elbow chart that is shown below:



From the chart is can be seen that the inertia starts gradually slowing down at a number of clusters of 10. Therefore, a number of 10 clusters was used for the k-means classification model for the neighbourhoods in Chicago being investigated.