COURSERA CAPSTONE

IBM Applied Data Science Capstone

Optimal Location for a new Bookstore in the city of Auckland, New Zealand

By: Rodrigo Castro

rodrigo20acc@gmail.com

June 2020

INDEX:

1. Introduction

The objective of this Data Science project is to determine an optimal location for a new bookstore in the city of Auckland, New Zealand. In this study, the data was extracted for different sources according to its nature. The information regarding the neighbourhoods in Auckland was retrieved from Wikipedia while the location data of venues was obtained through the Foursquare server. We used as well Machine Learning tools such as K-means in order to obtain data clusters of potential locations. Data wrangling and preprocessing was also necessary to structure the data properly and to facilitate the analysis. Finally, the mean by which we made use of the algorithms and handled the datasets was the programming language Python.

1.1. Background

Even if nowadays the e-commerce of electronic books and books sent by delivery is showing a remarkable growth, it's still very important for booklovers to find a place to take their time and interact with the objects of their affection. Even more so, if this place's surroundings enable the costumers to complement their experience providing them, for instance, with comfortable places to read or spaces of cultural interest to visit. Bookstores are appealing places in which readers can have unexpected encounters and be welcomed with enjoyable surprises. For that very reason, physical bookstores will always be on demand. That is why finding an optimal location is necessary for their success.

1.2. Business Understanding

In 2018, a study conducted in the USA reflected in figures an increase of 35% in the rise of indie bookstores in New Zealand. In fact, they determined the existence of an upward trend of new bookstores opening in the near future. Now, even in the midst of the Covid-19 crisis, bookstores sales are experiencing a considerable boom as an effort to reactivate the economy following the period of lockdown. According to an article from The Guardian, the reason behind this behavior is the people's support to local venues rather than to the bigger enterprises that can provide cheaper books.

It seems that booklovers in New Zealand have a strong presence and that more citizens are coming to understand the value of bookstores around their neighbourhoods. The fact that these bookstores bring support to the local community makes them more welcome.

Auckland, which is the most important metropolitan city in New Zealand and is home to its largest population, provides an attractive environment in which to open a new bookstore. In fact, Auckland is the leading center of New Zealand's business and economic development. Thus, is reasonable to select this city as the subject of study for this project.

What is curious about bookstores is that, unlike other types of stores, they can benefit from sharing the same environment with their competitors as they complement each other. Let’s bear in mind that each bookstore may have a unique profile and also be specialized in a different publishing house. For that reason, in different bookstores we can find the same title not only at a different price, but also with a different editorial work behind it, which gives a totally different value to the book. By consequence, having bookstores located near each other enrich the experience of the costumer and ensure that all of them are eventually visited. In this project, we will focus only in the areas where there already are other bookstores.

1.3. Target Audience

The findings of this capstone project are intended to be of use to any stakeholder who is interested in opening a new bookstore in the city of Auckland by providing useful information that can be of help in order to find an optimal location.

2. Data

The data used to obtain the optimal location for the new bookstore has been retrieved and processed from multiple sources. Now, let’s explain the data in detail.

2.1. Neighbourhoods

In the first place, data regarding the neighbourhoods in Auckland has been retrieved from Wikipedia. In the References section at the end of this document you will find the link to the database. This data contains only the name of each neighbourhood in the city, information such as coordinates has been obtained by other means.

2.2. Latitude and Longitude Coordinates

The coordinates for each neighbourhood in Auckland are determined by the geocoder library of Python given their names. The values obtained were merged then with the names of the neighbourhoods obtained from Wikipedia.

2.3. Venue Data

The location data used in this project to obtain the characteristics of the venues in the different neighbourhoods came from Foursquare by making an API call and passing the required parameters to the server. The values obtained were merged as well with the data above mentioned.

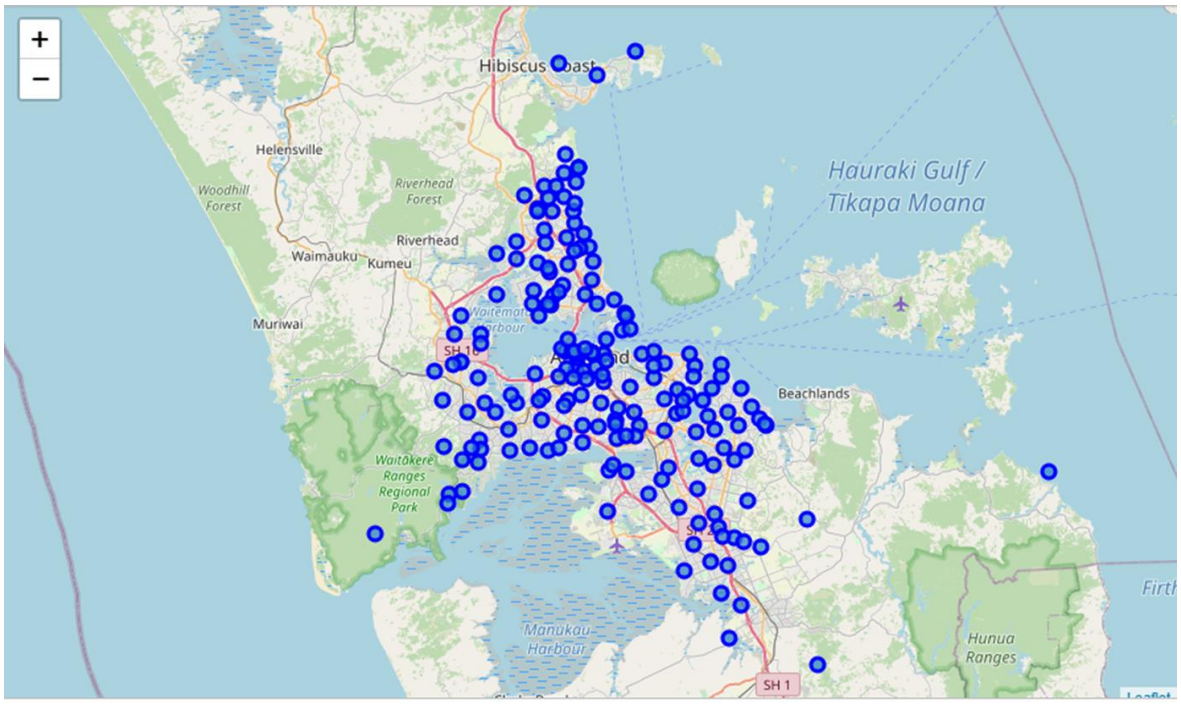
3. Methodology

3.1. Data Preparation

The initial step in the methodology section consists of preparing the raw data into a structured format that enables us to visualize the information more easily. The Beautiful Soup package was used to compile the neighbourhood data from Wikipedia. Then with the use of the geocoder library the coordinates were extracted. Here is how a portion of the initial dataframe, that contains the name and coordinates of the neighbourhoods of Auckland, looks like.

	Neighbourhood	Latitude	Longitude
1	Airport Oaks	-36.729160	174.708930
2	Albany, New Zealand	-37.010493	174.940852
3	Alfriston, New Zealand	-36.848399	174.764388
4	Algies Bay	-36.863020	174.748580
5	Arch Hill, New Zealand	-36.847862	174.737085

After that, using the Folium package enabled us to create a map of Auckland along with the marked coordinates of its neighbourhoods.



3.2. Foursquare API

As mentioned before, the location data was extracted from the Foursquare server inputting the credentials associated to a personal account. For the purpose of this project, only venues within a radius of 3 km around each coordinate were considered. In addition, only the hundred venues closest to said coordinates were considered. Thus, a new dataframe was created including the characteristics of the venues retrieved.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Airport Oaks	-36.72916	174.70893	K-Mart	-36.728644	174.709722	Department Store
1	Airport Oaks	-36.72916	174.70893	QBE Stadium	-36.726937	174.702059	Stadium
2	Airport Oaks	-36.72916	174.70893	The Merchant Bar & Kitchen	-36.727810	174.709088	Bar
3	Airport Oaks	-36.72916	174.70893	Event Cinemas	-36.728335	174.708643	Movie Theater
4	Airport Oaks	-36.72916	174.70893	Albany Mega Centre	-36.731220	174.706719	Shopping Mall

3.3. One-hot Encoding

To prepare the data for clustering, it was necessary to perform the one-hot encoding process. It consists of transforming the categorical variables of a dataset into numeric features so the Machine Learning algorithms can process said dataset more efficiently. To this effect, the unique items in the Venue Category column of the dataframe were one-hot encoded. Afterwards, the new dataframe was filtered considering only the neighbourhoods that contain bookstores, since they are the only ones that are of our interest.

	Neighbourhood	African Restaurant	Airport	Airport Lounge	Airport Service	American Restaurant	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	...	Vietnamese Restaurant
9	Avondale, Auckland	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	...	0.000000
19	Brookby	0.0	0.0	0.0	0.0	0.010000	0.000000	0.000000	0.0	0.000000	...	0.000000
25	Chatswood, New Zealand	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	...	0.000000
30	Cockle Bay, New Zealand	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	...	0.000000
33	Dannemora, New Zealand	0.0	0.0	0.0	0.0	0.000000	0.012195	0.000000	0.0	0.000000	...	0.000000
34	Devonport, New Zealand	0.0	0.0	0.0	0.0	0.000000	0.024390	0.000000	0.0	0.000000	...	0.000000
40	Ellerslie, New Zealand	0.0	0.0	0.0	0.0	0.010000	0.000000	0.000000	0.0	0.010000	...	0.010000

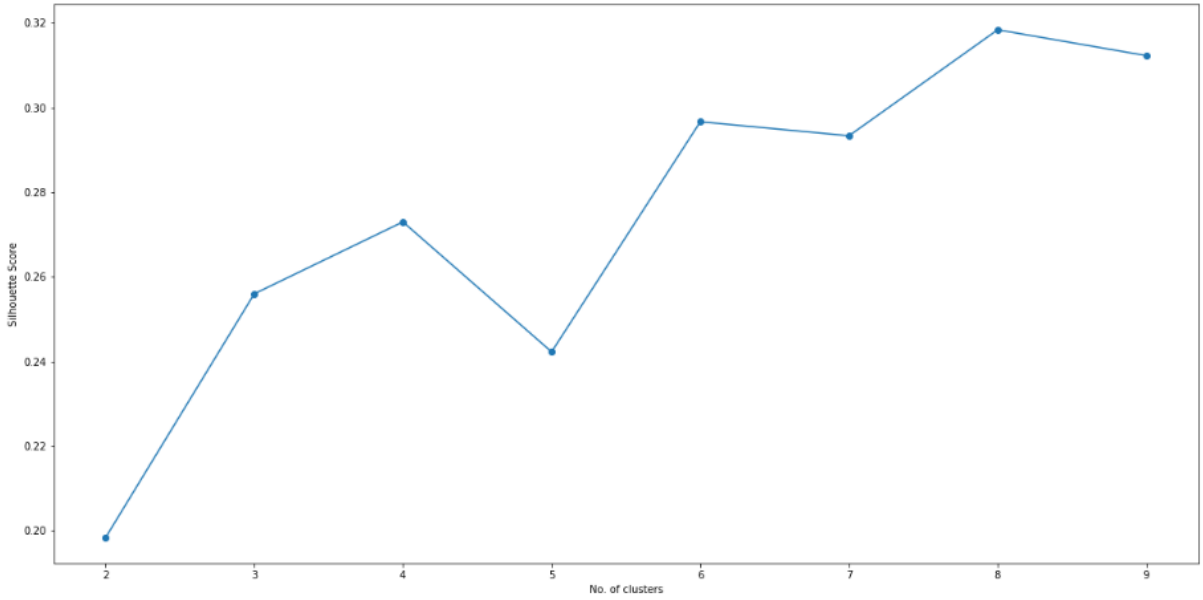
3.4. Filtering by most popular venues

For further understanding, for each of the selected neighbourhoods the 10 most popular venues were retrieved. This information helps us to comprehend the characteristics of those areas.

	Neighbourhood	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
9	Avondale, Auckland	Café	Chinese Restaurant	Bar	Coffee Shop	Bakery	Gym	Park	Indian Restaurant	Burger Joint	Asian Restaurant
19	Brookby	Café	Fast Food Restaurant	Beach	Coffee Shop	Neighborhood	Burger Joint	Japanese Restaurant	Supermarket	Indian Restaurant	Bar
25	Chatswood, New Zealand	Café	Scenic Lookout	Beach	Bar	History Museum	Bakery	Performing Arts Venue	Comfort Food Restaurant	Candy Store	Burger Joint
30	Cockle Bay, New Zealand	Café	Park	Clothing Store	Grocery Store	Fast Food Restaurant	Farm	Rental Car Location	Juice Bar	Sporting Goods Shop	Motel
33	Dannemora, New Zealand	Café	Hotel	Bar	Beach	Harbor / Marina	Restaurant	Coffee Shop	Pizza Place	Turkish Restaurant	Japanese Restaurant

3.5. Clustering

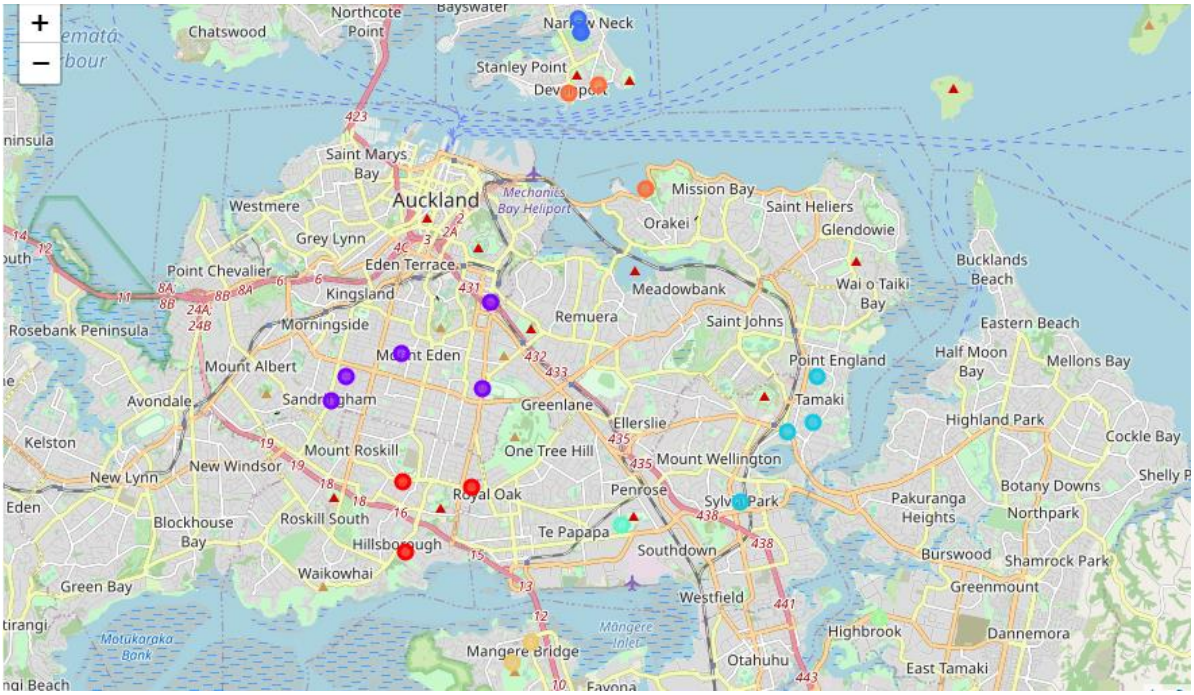
Now that the one-hot encoded dataframe was obtained, it is possible to proceed to cluster the neighbourhoods by similarity of venues. In this case, the K-mean algorithm, an unsupervised Machine Learning clustering method, was used since it is computationally faster when working with huge datasets. But before that, it was necessary to define an ideal number of clusters that ensured that the elements in each one are adequately similar and that the elements from different clusters are significantly different.



By running different clustering to the dataset and obtaining their respective scores, it was concluded that 8 is the ideal number of clusters for our dataset. Finally, it was time to perform the K-mean algorithm and assign the cluster labels to the neighbourhoods. The dataset obtained was merged along with the one that contained the 10 most popular venues, so in only one dataframe we can visualize the name of the neighbourhood, its coordinates, its number of cluster and its most common venues.

	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
161	Rothesay Bay	-36.910653	174.772330	0	Café	Fast Food Restaurant	Park	Bakery	Indian Restaurant
186	Te Papapa	-36.909430	174.754830	0	Café	Indian Restaurant	Park	Bakery	Chinese Restaurant
75	Hillpark, Auckland	-36.923898	174.755363	0	Café	Fast Food Restaurant	Park	Shopping Mall	Bakery
10	Avondale, Auckland	-36.888200	174.740190	1	Café	Chinese Restaurant	Bar	Coffee Shop	Bakery

Finally, with the Folium package, we can visualize one more time the map of Auckland. However, this time it shows only the neighbourhoods that according to our parameters have bookstores in them. The coordinates are colored according to the cluster they belong to.



4. Analysis

Once the clustering was performed, we obtained the 8 desired clusters of neighbourhoods. Each cluster represents a potential location to open a new bookstore since they are the areas that contain others as well. Now is the moment to define which cluster is the most appropriate to harbor the new bookstore and for that reason let's analyse them based on their

profiles. A suitable name was given to each of the clusters in a way that it summarizes their characteristics.

Cluster 1: Diverse Restaurants and Parks

	Neighbourhood	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
161	Rothesay Bay	Café	Fast Food Restaurant	Park	Bakery	Indian Restaurant	Japanese Restaurant	Gastropub	Chinese Restaurant	Shopping Mall	Planetarium
186	Te Papapa	Café	Indian Restaurant	Park	Bakery	Chinese Restaurant	Thai Restaurant	Fast Food Restaurant	Japanese Restaurant	Coffee Shop	Supermarket
75	Hillpark, Auckland	Café	Fast Food Restaurant	Park	Shopping Mall	Bakery	Grocery Store	Burger Joint	Indian Restaurant	Supermarket	Outdoor Supply Store

Cluster 2: Asian Cuisine Restaurants and Bars

	Neighbourhood	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
10	Avondale, Auckland	Café	Chinese Restaurant	Bar	Coffee Shop	Bakery	Gym	Park	Indian Restaurant	Burger Joint	Asian Restaurant
164	Saint Marys Bay, New Zealand	Café	Chinese Restaurant	Bar	Indian Restaurant	Coffee Shop	Bakery	Gym	Thai Restaurant	Asian Restaurant	Food & Drink Shop
98	Manukau	Café	Chinese Restaurant	Thai Restaurant	Gym	Japanese Restaurant	Bar	Park	Ice Cream Shop	Bakery	Burger Joint
114	Mount Albert, New Zealand	Café	Japanese Restaurant	Chinese Restaurant	Bar	Gym	Malay Restaurant	Thai Restaurant	Bakery	Coffee Shop	Park
42	Ellerslie, New Zealand	Café	Japanese Restaurant	Park	Coffee Shop	Restaurant	Dessert Shop	Bagel Shop	Thai Restaurant	Pizza Place	Bistro

Cluster 3: Scenic Lookout and Cultural Places

	Neighbourhood	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
117	Murrays Bay	Café	Bar	Beach	Scenic Lookout	Movie Theater	Fruit & Vegetable Store	Sushi Restaurant	Supermarket	Bookstore	History Museum
26	Chatswood, New Zealand	Café	Scenic Lookout	Beach	Bar	History Museum	Bakery	Performing Arts Venue	Comfort Food Restaurant	Candy Store	Burger Joint
192	Unsworth Heights	Café	Scenic Lookout	Beach	Bar	History Museum	Bakery	Performing Arts Venue	Comfort Food Restaurant	Candy Store	Burger Joint

Cluster 4: Fast Food Restaurants and Diverse Stores

	Neighbourhood	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
115	Mount Eden	Café	Fast Food Restaurant	Burger Joint	Department Store	Electronics Store	Grocery Store	Clothing Store	Malay Restaurant	Bar	Shopping Mall
140	Pakuranga Heights	Café	Fast Food Restaurant	Grocery Store	Department Store	Clothing Store	Pizza Place	Train Station	Supermarket	Burger Joint	Malay Restaurant
169	Southdown, New Zealand	Café	Fast Food Restaurant	Pizza Place	Grocery Store	Clothing Store	Electronics Store	Supermarket	Athletics & Sports	Train Station	Department Store
180	Takapuna	Café	Pizza Place	Park	Fast Food Restaurant	Burger Joint	Athletics & Sports	Train Station	Coffee Shop	Department Store	Supermarket

Cluster 5: Fast Food Restaurant Mostly

	Neighbourhood	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
20	Brookby	Café	Fast Food Restaurant	Beach	Coffee Shop	Neighborhood	Burger Joint	Japanese Restaurant	Supermarket	Indian Restaurant	Bar
168	Shelly Park	Café	Fast Food Restaurant	Burger Joint	Clothing Store	Grocery Store	Pizza Place	Electronics Store	Bar	Athletics & Sports	Furniture / Home Store

Cluster 6: Asian Cuisine Restaurants Mostly

	Neighbourhood	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
70	Herne Bay, New Zealand	Vietnamese Restaurant	Chinese Restaurant	Japanese Restaurant	Market	Café	Department Store	Fast Food Restaurant	Gym	Mongolian Restaurant	Basketball Court

Cluster 7: Parks and Diverse Stores

	Neighbourhood	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
95	Māngere	Café	Park	Fast Food Restaurant	Grocery Store	Farm	Bookstore	Train Station	Clothing Store	Outdoor Supply Store	Shopping Mall
31	Cockle Bay, New Zealand	Café	Park	Clothing Store	Grocery Store	Fast Food Restaurant	Farm	Rental Car Location	Juice Bar	Sporting Goods Shop	Motel

Cluster 8: Beach and Harbor

	Neighbourhood	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
35	Devonport, New Zealand	Café	Beach	Harbor / Marina	Bar	Scenic Lookout	Restaurant	Supermarket	Boat or Ferry	Bookstore	Sporting Goods Shop
34	Dannemora, New Zealand	Café	Hotel	Bar	Beach	Harbor / Marina	Restaurant	Coffee Shop	Pizza Place	Turkish Restaurant	Japanese Restaurant
129	Ōpaheke	Café	Park	Beach	Coffee Shop	Italian Restaurant	Fish & Chips Shop	Bar	Harbor / Marina	Comfort Food Restaurant	Deli / Bodega

5. Result

We got a glimpse of the possible locations for a new bookstore in Auckland and we got interesting information which may be useful to any stakeholder willing to take on this business.

For example, it seems that out of all the neighbourhoods in Auckland, in only 23 we can find the major density of bookstores. Let’s consider, of course, that we only imputed for each neighbourhood an analysis to an area of 3 km picking only the first hundred venues nearer to the coordinates.

Among the most common venues in Auckland, we can find that Cafés, Fast Food Restaurants and Asian Cuisine Restaurants are indisputably popular. Indeed, the food consumption business seems to be quite the profitable one.

6. Discussion

After conducting the K-mean clustering in Auckland's neighbourhoods and analysing each cluster, we can observe that Cluster 3 'Scenic Lookout and Cultural Places' seems to be the most appropriate place to open a new bookstore. The reason being that these neighbourhoods seem to be located in an attractive place to the kind of costumers who frequent bookstores given that it contains places of cultural interest and scenic lookouts where they may read. These elements enrich the experience of any booklover willing to walk through an entire neighbourhood visiting different bookstores since they complement each other.

Let’s remember that this analysis by no means guarantee the success of opening a new bookstore, it only provides a guide to the stakeholders to identify suitable regions for their business. However, the findings obtained in this project have a solid background behind them and are of great help at the moment of taking a decision.

7. Conclusion

The task of identifying an optimal location to open a bookstore in the city of Auckland has required that we used multiple Data Science Skills. From data collection and data preparation to modeling and evaluation, we went through some of the Data Science Methodology steps. The data analysis, data visualization and Machine Learning with Python had a key role in this project as well.

This project, despite using in detail the Machine Learning tools and the data extraction techniques, limits the scope of its analysis by deciding an optimal location based solely on the venues of a neighbourhood. Other factors such as socioeconomic statistics, population density, land price or supply chain analysis, just to name a few, could have been used to produce a more in-depth investigation. Nevertheless, the findings of this study represent a significant step towards investigating the bookstore business in the city of Auckland.

8. References

- https://www.booksellers.co.nz/news/new-zealand-bookshop-numbers-rise
- https://www.theguardian.com/world/2020/may/23/like-christmas-new-zealands-post-covid-books-boom
- https://en.wikipedia.org/wiki/Category:Suburbs_of_Auckland