The Battle of Neighborhoods in Vancouver, BC, Canada

Introduction

For the past 10 years Canada has been the number one choice for immigrants all over the world. The stable economy, great educational system and the beautiful nature have been attracted people to apply for Canadian permanent residency. However, for anyone who decides to immigrate or just travel to Great White North, there is a lack of information about certain cities, specifically about their neighborhoods. Potential new immigrants could face a problem where to live in a certain town? Which neighborhoods have more parks nearby? Is there any coffee shops and yoga studious nearby? To address all these uncertainties, one could take a data of the city districts and its venues to find the most appropriate and suitable neighborhood for him. In this project, the neighborhoods of Vancouver were analyzed and as result the recommendations were made for new potential immigrants, who decided to travel to Vancouver.

Data description

Data sources

The data required for this project are the information about neighborhoods in Vancouver and the data about venues in the city.

The data about venues was taken via the Foursquare.com API, which returns the data about venues in the specified location. It is required to pass the user credential information, longitude and latitude of the interested area and the interested number of venues to the API. The data from the API is obtained in json file with the information about venue, address, category, price, users' comments and likes. For this project the data about venues categories was collected and analyzed.

The required information about city's districts boundaries was taken from the website *City Of Vancouver Open Data Portal (www.opendata.vancouver.ca).* The database *Local area boundary* contains the data about Postal codes, names of neighborhoods and its geo-coordinates.

Data cleaning and transforming

Local area boundary dataset was downloaded in csv format and transformed into table with the pandas library's function *read_csv().* Columns MAPID and Name were renamed for convenience.

```
df_vancouver=pd.read_csv('vancouver_dataset.csv')
df_vancouver.rename(columns={'MAPID': 'PostalCode', 'Name':'Neighborhood'},inplace=True)
df_vancouver
```

As the result the following table *df_vancouver* with the columns *PostalCode, Neighborhood, Latitude and Longitude* was created.

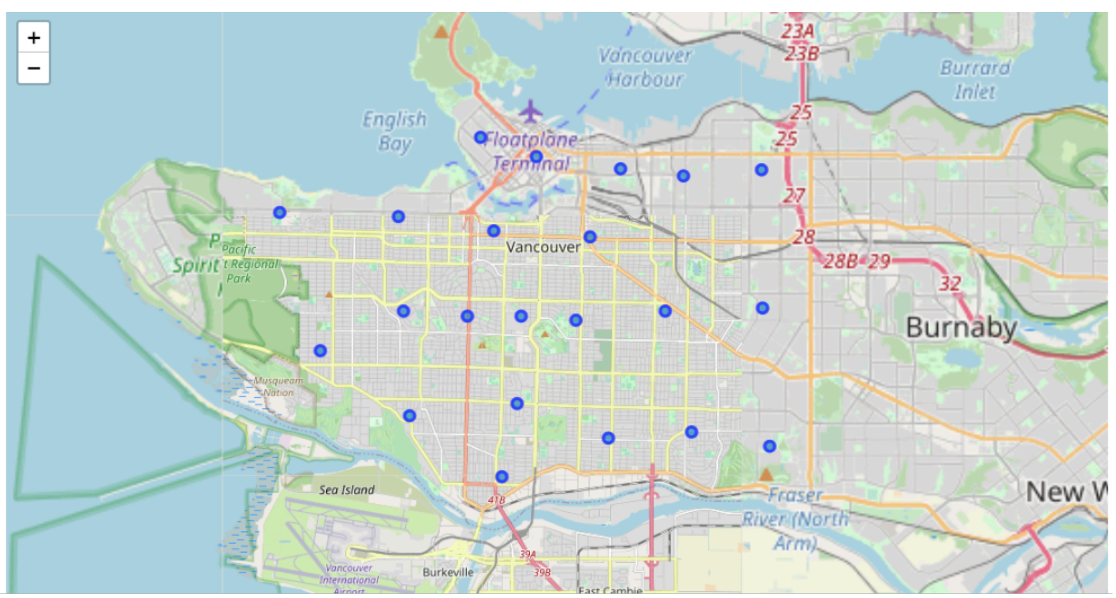
[2]:

	PostalCode	Neighborhood	Latitude	Longitude
0	AR	Arbutus-Ridge	49.246805	-123.161669
1	CBD	Downtown	49.280747	-123.116567
2	FAIR	Fairview	49.264540	-123.131049
3	GW	Grandview-Woodland	49.276440	-123.066728
4	HS	Hastings-Sunrise	49.277934	-123.040270
5	MARP	Marpole	49.210207	-123.128382
6	RP	Riley Park	49.244766	-123.103147
7	SHAU	Shaughnessy	49.245681	-123.139760
8	STR	Strathcona	49.278220	-123.088235
9	WE	West End	49.285011	-123.135438
10	DS	Dunbar-Southlands	49.237962	-123.189547
11	KERR	Kerrisdale	49.223655	-123.159576
12	KIL	Killarney	49.217022	-123.037647
13	KITS	Kitsilano	49.267540	-123.163295
14	SC	South Cambie	49.245556	-123.121801
15	VF	Victoria-Fraserview	49.220012	-123.064135
16	KC	Kensington-Cedar Cottage	49.246686	-123.072885
17	MP	Mount Pleasant	49.263065	-123.098513
18	OAK	Oakridge	49.226403	-123.123025
19	RC	Renfrew-Collingwood	49.247343	-123.040166
20	SUN	Sunset	49.218756	-123.092038
21	WPG	West Point Grey	49.268401	-123.203468

Create a map of Vancouver with neighborhoods superimposed on top.

For the illustration purposes the map of the Vancouver's neighborhoods with its names on top was created.

[4]:



Exploring neighborhoods in Vancouver with the Foursquare API

The acquisition process of data with the Foursquare API requires the following:

- url,
- CLIENT_ID,
- CLIENT_SECRET
- VERSION
- Neighborhood latitude,
- Neighborhood longitude,
- Radius,
- LIMIT

It is required to be a registered user at the Foursquare.com as a developer to receive the url, CLIENT_ID and CLIENT_SECRET.

```
CLIENT_ID = 'XNE1FN0GRLSRNRWJGKNJOGMV3CGXVQO525RVF1CGAIGBXFRI' # your Foursquare ID
CLIENT_SECRET = 'WAZCUGEY4E04HMLKSMISSD55VLLRVU4M5YATOG4IXONL4FRE' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

At this point all requested information for the Foursquare API is ready and the request can be send.

```
neighborhood_latitude = df_vancouver.loc[1, 'Latitude'] # neighborhood latitude value
neighborhood_longitude = df_vancouver.loc[1, 'Longitude'] # neighborhood longitude value

neighborhood_name = df_vancouver.loc[1, 'Neighborhood'] # neighborhood name

LIMIT = 100 # limit of number of venues returned by Foursquare API
radius = 500 # define radius

# create URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
url # display URL
```

The Foursquare API returns the results in json format file. Thus, to extract the required information for the further analysis of venues, the function *get_category_type* was defined.

```
# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']
```

Using the above function the results can be easily stored in a table format.

```
#recieving the reqsults from the Foursquae API

results = requests.get(url).json()

venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head() #show the first 5 venues
```

As the result the table *nearby_venus* was created. The table has the data about names of venues, categories and geo-coordinates of each venue.

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:7: FutureWarning: pandas.io.json.json_normalize is depr
ecated, use pandas.json_normalize instead
  import sys
```

	name	categories	lat	lng
0	L'Hermitage	Hotel	49.280139	-123.117480
1	Gotham Steakhouse & Cocktail Bar	Steakhouse	49.282830	-123.115865
2	Medina Café	Breakfast Spot	49.280565	-123.116859
3	JJ Bean	Coffee Shop	49.279382	-123.115181
4	Paramount Fine Foods	Lebanese Restaurant	49.280452	-123.118586

```
print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```

```
54 venues were returned by Foursquare.
```

Joining the data

For the further analysis, the data from both tables df_vancouver and nearby_venues was merged. As the results the table Vancouver_venues was created.

```
2]: vancouver_venues.head()
```

2]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arbutus-Ridge	49.246805	-123.161669	Arbutus Village Park	49.249486	-123.158716	Park
1	Arbutus-Ridge	49.246805	-123.161669	Total Corporate Learning Inc.	49.242441	-123.161609	Business Service
2	Arbutus-Ridge	49.246805	-123.161669	Triangle Park	49.245061	-123.167914	Park
3	Downtown	49.280747	-123.116567	L'Hermitage	49.280139	-123.117480	Hotel
4	Downtown	49.280747	-123.116567	Gotham Steakhouse & Cocktail Bar	49.282830	-123.115865	Steakhouse

```
3]: vancouver_venues.shape
```

```
3]: (453, 7)
```

For the sake of convenience, the data in the table Vancouver_venues was grouped and sorted.

```
vancouver_venues.groupby('Neighborhood').count().sort_values(by='Venue',ascending=False)
```

[36]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Mount Pleasant	79	79	79	79	79	79
Riley Park	57	57	57	57	57	57
Downtown	54	54	54	54	54	54
Kitsilano	51	51	51	51	51	51
Grandview-Woodland	37	37	37	37	37	37
Fairview	26	26	26	26	26	26
West End	24	24	24	24	24	24
Killarney	19	19	19	19	19	19
Kensington-Cedar Cottage	16	16	16	16	16	16
Hastings-Sunrise	14	14	14	14	14	14
South Cambie	12	12	12	12	12	12
Oakridge	11	11	11	11	11	11
Strathcona	10	10	10	10	10	10
Victoria-Fraserview	8	8	8	8	8	8
Marpole	7	7	7	7	7	7
West Point Grey	6	6	6	6	6	6
Dunbar-Southlands	6	6	6	6	6	6
Renfrew-Collingwood	5	5	5	5	5	5
Shaughnessy	4	4	4	4	4	4
Kerrisdale	4	4	4	4	4	4
Arbutus-Ridge	3	3	3	3	3	3

Methodology section

One Hot Encode

One Hot Encoding methodology is used in machine learning to analyze the categorical data. Thus, in this project this methodology should be applied due the collected data about venues is a categorical data, which should be converted into numbers for the further analysis.

```
# one hot encoding
vancouver_onehot = pd.get_dummies(vancouver_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
vancouver_onehot['Neighborhood'] = vancouver_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [vancouver_onehot.columns[-1]] + list(vancouver_onehot.columns[:-1])
vancouver_onehot = vancouver_onehot[fixed_columns]

vancouver_onehot.head()
```

	Neighborhood	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Stadium	Belgian Restaurant	Board Shop	Boat or Ferry	Bookstore	Breakfast Spot
0	Arbutus-Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Arbutus-Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Arbutus-Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Downtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Downtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Sorting the data

After the data was transformed into the required format, it was sorted and grouped for the next step. At this point rows were grouped by neighborhood and by taking the mean of the frequency of occurrence of each category.

```
vancouver_grouped = vancouver_onehot.groupby('Neighborhood').mean().reset_index()
vancouver_grouped
```

	Neighborhood	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Stadium	Beer Garden	Belgian Restaurant
0	Arbutus-Ridge	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Downtown	0.000000	0.018519	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.055556	0.000000	0.000000	0.018519
2	Dunbar-Southlands	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Fairview	0.000000	0.000000	0.038462	0.076923	0.000000	0.038462	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Grandview-Woodland	0.000000	0.000000	0.000000	0.027778	0.027778	0.027778	0.000000	0.027778	0.000000	0.000000	0.000000	0.000000	0.000000
5	Hastings-Sunrise	0.000000	0.000000	0.000000	0.000000	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000	0.000000	0.076923	0.000000
6	Kensington-Cedar Cottage	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.071429	0.000000	0.000000	0.000000	0.000000
7	Kerrisdale	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000	0.000000	0.000000
8	Killarney	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632	0.000000	0.000000	0.000000	0.000000
9	Kitsilano	0.020000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.040000	0.000000	0.000000	0.000000	0.000000
10	Marpole	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	Mount Pleasant	0.012658	0.012658	0.025316	0.000000	0.000000	0.000000	0.012658	0.012658	0.000000	0.025316	0.000000	0.000000	0.000000
12	Oakridge	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	Renfrew-Collingwood	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	Riley Park	0.000000	0.017544	0.017544	0.000000	0.017544	0.000000	0.000000	0.000000	0.017544	0.000000	0.017544	0.000000	0.000000
15	Shaughnessy	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	South Cambie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000

K-means Clustering

K-means clustering is one of the simplest and popular methodologies in machine learning to structure data. Clustering is the process of grouping similar data into subgroups (clusters) and separate different data into other subgroups. K-means clustering algorithm partitions data into non-overlapping datasets with the pre-defined K-number centroids. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid is at the minimum.

The library scikit-learn provides the great method KMeans that can easily handle the K-means clustering process with all complicated mathematical computations.

For this project the number of the K means centroids was K=5.

```
# set number of clusters
kclusters = 5

vancouver_grouped_clustering = vancouver_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(vancouver_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([2, 1, 4, 1, 1, 1, 1, 1, 1, 1], dtype=int32)
```

Next, it is required to add clusters into existing table.

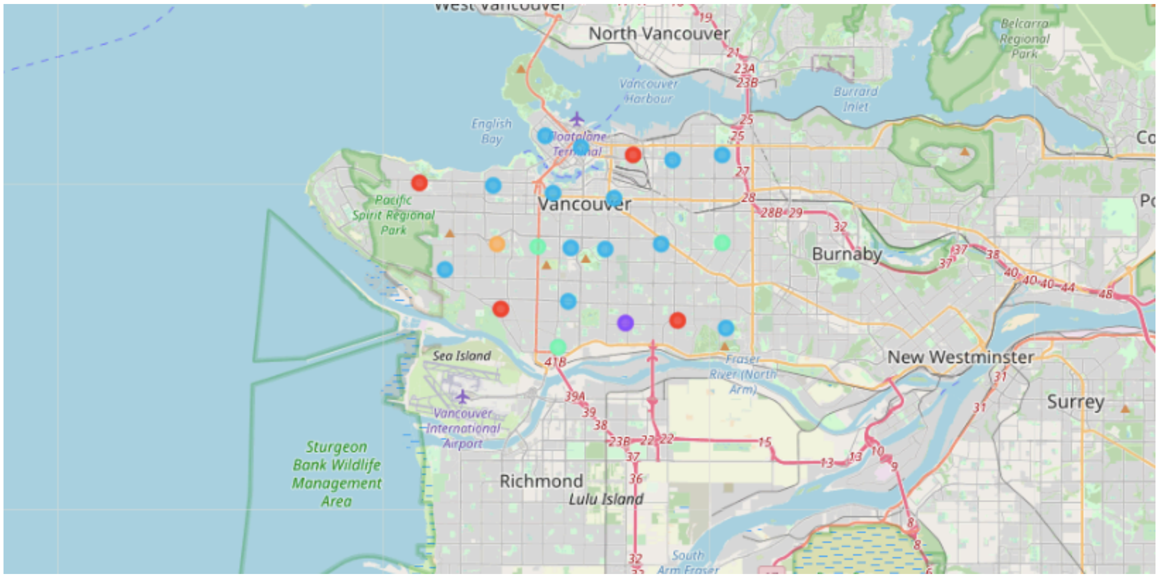
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

vancouver_merged = df_vancouver

# merge vancouver_grouped with vancouver_data to add latitude/longitude for each neighborhood
vancouver_merged = vancouver_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

vancouver_merged # check the last columns!
```

Results section

As the result for running the code with the defined Kmeans function and merging it with the existing table *df_vanc*ouver, the following dataframe was created.

	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels	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
0	AR	Arbutus-Ridge	49.246805	-123.161669	4	Park	Business Service	Fish & Chips Shop	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store
1	CBD	Downtown	49.280747	-123.116567	2	Hotel	Coffee Shop	Restaurant	Bar	Sandwich Place	Concert Hall	Vegetarian / Vegan Restaurant	Steakhouse
2	FAIR	Fairview	49.264540	-123.131049	2	Coffee Shop	Park	Asian Restaurant	Breakfast Spot	Japanese Restaurant	Furniture / Home Store	Camera Store	Spa
3	GW	Grandview-Woodland	49.276440	-123.066728	2	Coffee Shop	Pizza Place	Deli / Bodega	Vegetarian / Vegan Restaurant	Theater	Indian Restaurant	Cajun / Creole Restaurant	Café
4	HS	Hastings-Sunrise	49.277934	-123.040270	2	Bridal Shop	Theme Park Ride / Attraction	Café	Pizza Place	Beer Garden	Park	Portuguese Restaurant	Theater
5	MARP	Marpole	49.210207	-123.128382	3	Bus Stop	Pizza Place	Plaza	Bus Station	Japanese Restaurant	Taiwanese Restaurant	Yoga Studio	Electronics Store
6	RP	Riley Park	49.244766	-123.103147	2	Farmers Market	Vietnamese Restaurant	Japanese Restaurant	Coffee Shop	Sporting Goods Shop	Grocery Store	Lounge	Café
7	SHAU	Shaughnessy	49.245681	-123.139760	3	Bus Stop	Print Shop	Chocolate Shop	Park	Yoga Studio	Dive Bar	Falafel Restaurant	Event Space
8	STR	Strathcona	49.278220	-123.088235	0	Park	Food Truck	Arts & Crafts Store	Coffee Shop	Deli / Bodega	Pub	Cheese Shop	Soup Place
9	WE	West End	49.285011	-123.135438	2	Café	Farmers Market	Gay Bar	Sushi Restaurant	Sandwich Place	Falafel Restaurant	Restaurant	Spanish Restaurant
10	DS	Dunbar-Southlands	49.237962	-123.189547	2	Grocery Store	Bus Stop	Coffee Shop	Japanese Restaurant	Liquor Store	Yoga Studio	Electronics Store	Farmers Market
11	KERR	Kerrisdale	49.223655	-123.159576	0	Pool	Park	Café	Bar	Deli / Bodega	Dessert Shop	Diner	Disc Golf

Additionally, for the illustration purposes the map with the identified clusters was created.



Examining each cluster

The K-mean clustering algorithm with the number of centroids equal 5, partitioned the data about venues in each neighborhood of Vancouver into five clusters.

```
vancouver_merged1.loc[vancouver_merged1['Cluster Labels'] == 0, vancouver_merged1.columns[[1] + list(range(5, vancouver_merged1.shape[1]))]]
```

	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
8	Strathcona	Park	Food Truck	Arts & Crafts Store	Coffee Shop	Deli / Bodega	Pub	Cheese Shop	Soup Place	Brewery	Restaurant
11	Kerrisdale	Pool	Park	Café	Bar	Deli / Bodega	Dessert Shop	Diner	Disc Golf	Discount Store	Dance Studio
15	Victoria-Fraserview	Park	Pet Store	Fish Market	Noodle House	Motorcycle Shop	Middle Eastern Restaurant	Asian Restaurant	Convenience Store	Diner	Dessert Shop
21	West Point Grey	Harbor / Marina	Gym / Fitness Center	Gym	Disc Golf	Performing Arts Venue	Park	Yoga Studio	Dive Bar	Falafel Restaurant	Event Space

```
vancouver_merged1.loc[vancouver_merged1['Cluster Labels'] == 1, vancouver_merged1.columns[[1] + list(range(5, vancouver_merged1.shape[1]))]]
```

	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
20	Sunset	South Indian Restaurant	Yoga Studio	Fish & Chips Shop	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Donut Shop	Dive Bar

```
vancouver_merged1.loc[vancouver_merged1['Cluster Labels'] == 2, vancouver_merged1.columns[[1] + list(range(5, vancouver_merged1.shape[1]))]]
```

	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
1	Downtown	Hotel	Coffee Shop	Restaurant	Bar	Sandwich Place	Concert Hall	Vegetarian / Vegan Restaurant	Steakhouse	Seafood Restaurant	Plaza
2	Fairview	Coffee Shop	Park	Asian Restaurant	Breakfast Spot	Japanese Restaurant	Furniture / Home Store	Camera Store	Spa	Korean Restaurant	Nail Salon
3	Grandview-Woodland	Coffee Shop	Pizza Place	Deli / Bodega	Vegetarian / Vegan Restaurant	Theater	Indian Restaurant	Cajun / Creole Restaurant	Café	Record Shop	Scandinavian Restaurant
4	Hastings-Sunrise	Bridal Shop	Theme Park Ride / Attraction	Café	Pizza Place	Beer Garden	Park	Portuguese Restaurant	Theater	Gas Station	BBQ Joint
6	Riley Park	Farmers Market	Vietnamese Restaurant	Japanese Restaurant	Coffee Shop	Sporting Goods Shop	Grocery Store	Lounge	Café	Restaurant	Skating Rink
9	West End	Café	Farmers Market	Gay Bar	Sushi Restaurant	Sandwich Place	Falafel Restaurant	Restaurant	Spanish Restaurant	Lingerie Store	Pub
10	Dunbar-Southlands	Grocery Store	Bus Stop	Coffee Shop	Japanese Restaurant	Liquor Store	Yoga Studio	Electronics Store	Farmers Market	Falafel Restaurant	Event Space
12	Killarney	Bus Stop	Chinese Restaurant	Pharmacy	Juice Bar	Sandwich Place	Salon / Barbershop	Farmers Market	Fast Food Restaurant	Recreation Center	Sushi Restaurant
13	Kitsilano	Coffee Shop	Japanese Restaurant	Food Truck	Italian Restaurant	Wine Shop	Bakery	Bank	Pizza Place	Optical Shop	Grocery Store
14	South Cambie	Coffee Shop	Malay Restaurant	Cantonese Restaurant	Vietnamese Restaurant	Grocery Store	Café	Park	Bank	Liquor Store	Sushi Restaurant
16	Kensington-Cedar Cottage	Vietnamese Restaurant	Seafood Restaurant	Supermarket	Gym / Fitness Center	Indian Restaurant	Burger Joint	Grocery Store	Sandwich Place	Café	Bank
17	Mount Pleasant	Coffee Shop	Diner	Sandwich Place	Sushi Restaurant	Breakfast Spot	Lounge	Indian Restaurant	Record Shop	Brewery	Clothing Store
18	Oakridge	Sporting	Bubble Tea	Light Rail	Coffee Shop	Vietnamese	Sandwich	Sushi	Fast Food	Farmers	Falafel

```
vancouver_merged1.loc[vancouver_merged1['Cluster Labels'] == 3, vancouver_merged1.columns[[1] + list(range(5, vancouver_merged1.shape[1]))]]
```

	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
5	Marpole	Bus Stop	Pizza Place	Plaza	Bus Station	Japanese Restaurant	Taiwanese Restaurant	Yoga Studio	Electronics Store	Falafel Restaurant	Event Space
7	Shaughnessy	Bus Stop	Print Shop	Chocolate Shop	Park	Yoga Studio	Dive Bar	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store
19	Renfrew-Collingwood	Convenience Store	Pizza Place	Bus Station	Bus Stop	Park	Electronics Store	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant

```
vancouver_merged1.loc[vancouver_merged1['Cluster Labels'] == 4, vancouver_merged1.columns[[1] + list(range(5, vancouver_merged1.shape[1]))]]
```

	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	Arbutus-Ridge	Park	Business Service	Fish & Chips Shop	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Donut Shop	Yoga Studio

Discussion

From the Results section, one could observe that the neighborhoods in Vancouver mostly similar to each other. Mostly half of the neighborhoods fell into the third cluster, while the rest of the neighborhoods were classified into four different clusters.

The cluster 1 includes neighborhoods with the recreational venues, such as parks and pools. Thus, these neighborhoods could by considered by the families who are looking for the active sport life or just want to have nice parks nearby.

The cluster 2 has the only one neighborhood with the Indian restaurant as the most favorable places. Thus, the Sunset neighborhood in Vancouver could be populated with Indian people. It could be helpful for people who want to live in a Asian populated area.

The cluster 3 is the biggest one and includes mostly 60% of all neighborhoods. Those neighborhoods have a lot of cafes, restaurants and coffee shops. So, this area could be considered by people looking to have places to go nearby. These districts could be good for students or young professionals.

The cluster 4 includes neighborhoods that don't have a lot of places to go, thus it could be just suburbs with houses. These places could be considered by families with kids or retired people.

The cluster 5 has only one neighborhood which is probably the business center.

Conclusion

To sum up, in this project, neighborhoods of Vancouver were analyzed based on the types of venues located at these areas and recommendations were made. The analysis included the data preparation with the pandas library, data analysis with the one hot encoding and K-means clustering methodologies. The recommendations were made based on the results from the clustering process. Hopefully, this analysis could help new or potential immigrants to find a suitable place to live or stay in Vancouver, BC, Canada.