Calgary: Sites and Hoods (A guide for newcomers)

Business Idea:

For my data science Capstone Project, I am using the Foursquare API and the openly available Wikipedia location data, to explore the City of Calgary, Alberta, within which I reside. As a newcomer to the City of Calgary, I was unclear about the best residential locations with access to relevant facilities. By performing this project, I am creating a platform for other newcomers to have a high level perspective of the distribution of the neighborhoods and locations of interests within Calgary, to help with smooth settlement.

Target Audience:

- Newcomers to Calgary

Reference Data and usage:

- List of Neighborhoods in Calgary – Wikipedia
  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T
- Foursquare API
  https://foursquare.com/developers/apps

For this project, the Wikipedia page provides a list of all neighborhoods and postal codes within Calgary, including their coordinates, while the Foursquare API provides the surrounding venues and venue categories for these coordinates. I intend to use this data to create neighborhood clusters based on the frequency of venue categories in each neighborhood.

Methodology:

The following work flow has been applied for this project:

- Extraction, exploration and clean-up of neighborhood data from Wikipedia
- Obtaining the surrounding venues from Foursquare API
- Analyzing the venue distribution for each neighborhood
- Apply K-Means to group neighborhoods into clusters
- Visualizing clusters on Map with Folium

Results:

An exploration of the Wikipedia data produced a total of 33 neighborhoods within Calgary.

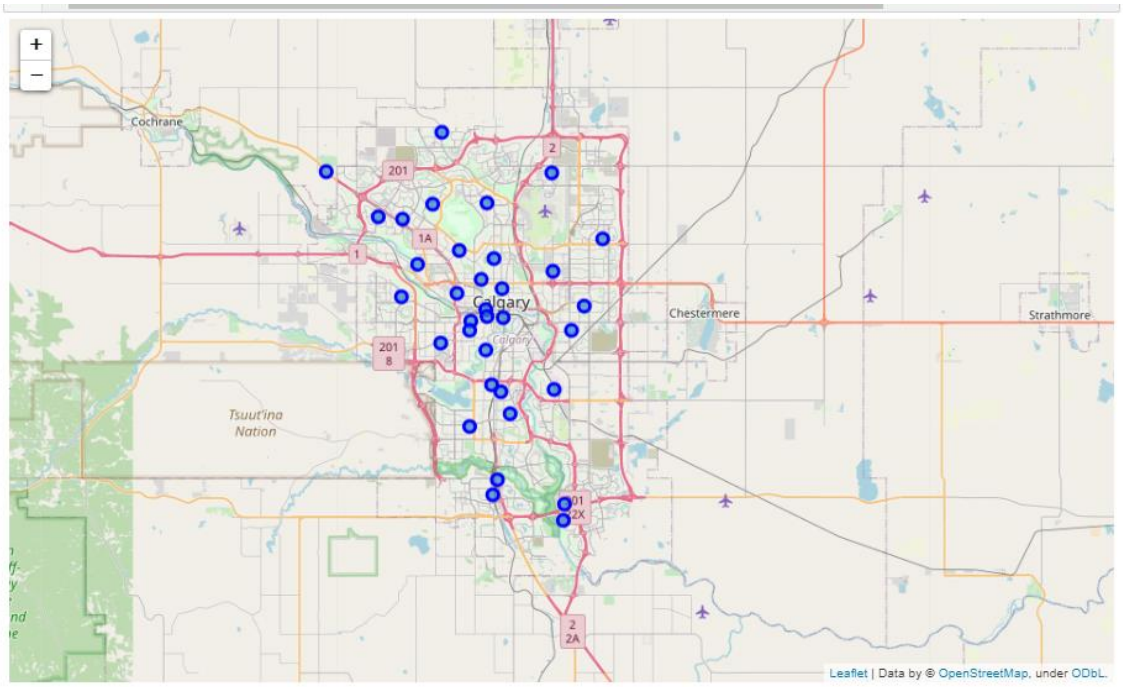
```
In [40]:
1 calgary_data.reset_index(drop=True)
```

Out[40]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
1	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158
2	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786
3	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.0809	-114.1616
4	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.9878	-114.0001
5	T3C	Calgary	Rosscarrock, Westgate, Wildwood, Shaganappi, S...	51.0388	-114.098
6	T2E	Calgary	Bridgeland, Greenview, Zoo, YYC	51.0632	-114.0614
7	T3E	Calgary	Lakeview, Glendale, Killarney, Glamorgan	51.0227	-114.1342
8	T2G	Calgary	Inglewood, Burnsland, Chinatown, East Victoria...	51.0415	-114.0599
9	T3G	Calgary	Hawkwood, Arbour Lake, Citadel, Ranchlands, Ro...	51.1147	-114.1796
10	T2H	Calgary	Highfield, Burns Industrial	50.9857	-114.0631
11	T3H	Calgary	Discovery Ridge, Signal Hill, West Springs, Ch...	51.0566	-114.1815
12	T2J	Calgary	Queensland, Lake Bonavista, Willow Park, Acadia	50.9693	-114.0514
13	T3J	Calgary	Martindale, Taradale, Falconridge, Saddle Ridge	51.0999	-113.9422
14	T2K	Calgary	Thorncliffe, Tuxedo Park	51.0857	-114.0714
15	T3K	Calgary	Sandstone, MacEwan Glen, Beddington, Harvest H...	51.127	-114.0787
16	T2L	Calgary	Brentwood, Collingwood, Nose Hill	51.0917	-114.1127

As visualized on a map:

Out[93]:



Using the Foursquare API produced the Venue and Venue Category for each Neighborhood, based on their coordinates.

```
In [96]:
#Size of venues dataframe
print(calgary_venues.shape)
calgary_venues.head()
```

```
(332, 7)
```

Out[96]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Penbrooke Meadows, Marlborough	51.049680	-113.964320	Bearcat General Contracting	51.047779	-113.968599	Construction & Landscaping
1	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158	Petro-Canada	51.128068	-114.138057	Gas Station
2	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158	Edgemont City	51.126473	-114.138997	Asian Restaurant
3	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158	Friends Cappuccino Bar & Bake Shop	51.126370	-114.138676	Café
4	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158	Mac's	51.128309	-114.137902	Convenience Store

Analyzing each neighborhood further produces the list of top 5 venues per neighborhood, ranked by frequency.

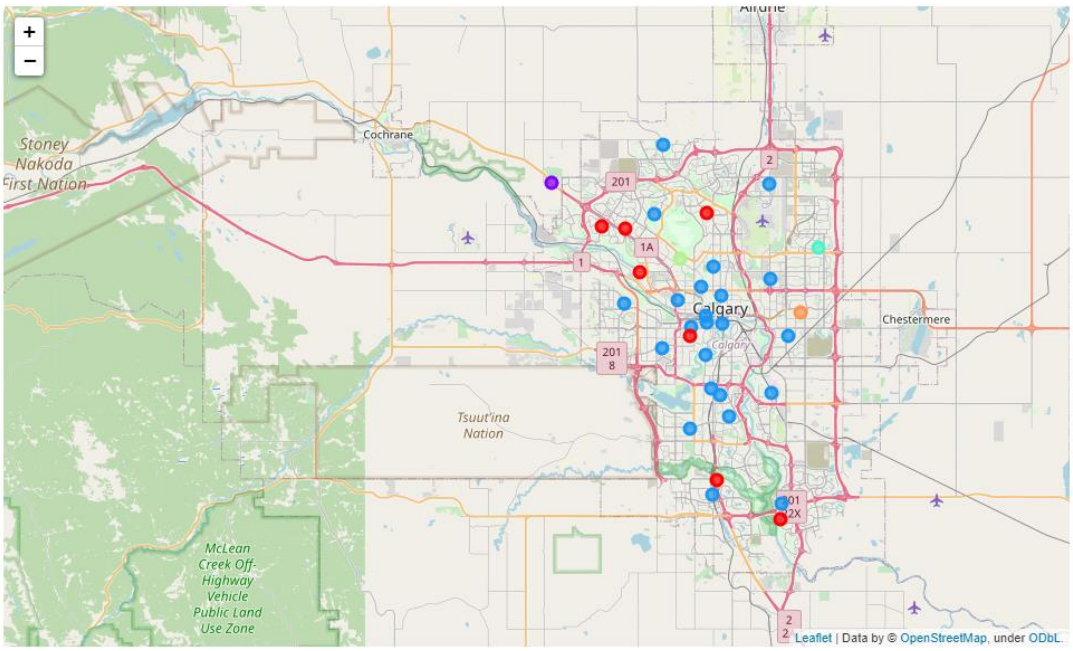
```
----Braeside, Cedarbrae, Woodbine----
               venue  freq
0        Hockey Rink  0.11
1           Pharmacy  0.11
2  Convenience Store  0.11
3     Ice Cream Shop  0.11
4        Pizza Place  0.11


----Brentwood, Collingwood, Nose Hill----
                     venue  freq
0         Asian Restaurant   1.0
1      American Restaurant   0.0
2        Korean Restaurant   0.0
3                Nightclub   0.0
4  New American Restaurant   0.0


----Bridgeland, Greenview, Zoo, YYC----
                      venue  freq
```

With this, it was able to further extract into a dataframe, the top 10 most common Venues for each 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	9th Most Common Venue	10th Most Common Venue
51.049680	-113.964320	5	Construction & Landscaping	Convenience Store	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Donut Shop	Dry Cleaner
51.126060	-114.143158	2	Convenience Store	Asian Restaurant	Gas Station	Café	Flea Market	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store
51.0318	-113.9786	2	Smoke Shop	Bar	Convenience Store	Food Court	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run
51.0809	-114.1616	0	Food Court	Coffee Shop	Bank	Scenic Lookout	Flea Market	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run
50.9878	-114.0001	2	Convenience Store	Pizza Place	Diner	Clothing Store	Yoga Studio	Deli / Bodega	Department Store	Dim Sum Restaurant	Discount Store	Dog Run

By applying K-Means clustering, 6 neighborhood clusters were created by grouping neighborhoods according to the kind of surrounding venues they had, based on their frequency scores. Plotting this on a map produces:

Out[175]:



The individual clusters are further detailed below:

Cluster 1

In [176]:

```
cluster1_df = calgary_merged.loc[calgary_merged['Cluster Labels'] == 0, calgary_merged.columns[[2] + list(range(5, calgar
cluster1_df
```

	Neighborhood	Cluster Labels	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue	Common Venue
3	Montgomery, Bowness, Silver Springs, Greenwood	0	Food Court	Coffee Shop	Bank	Scenic Lookout	Flea Market	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run
15	Sandstone, MacEwan Glen, Beddington, Harvest H...	0	Pizza Place	Grocery Store	Italian Restaurant	Liquor Store	Pharmacy	Bank	Coffee Shop	Dog Run	Dry Cleaner	Donut Shop
17	Tuscany, Scenic Acres	0	Pizza Place	Pharmacy	Liquor Store	Pub	Chinese Restaurant	Convenience Store	Video Store	Diner	Dim Sum Restaurant	Eastern European Restaurant
19	Cranston, Auburn Bay, Mahogany	0	Cosmetics Shop	Pizza Place	Liquor Store	Fast Food Restaurant	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run
28	South Calgary (Altadore / Bankview / Richmond)	0	Convenience Store	Pizza Place	Coffee Shop	Liquor Store	Yoga Studio	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store

Cluster 2

In [177]:
```
cluster2_df = calgary_merged.loc[calgary_merged['Cluster Labels'] == 1, calgary_merged.columns[[2] + list(range(5, calgar
cluster2_df
```

Out[177]:

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
25	Northwest Calgary	1	Flea Market	Yoga Studio	Convenience Store	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Donut Shop

Cluster 3

In [178]:
```
cluster3_df = calgary_merged.loc[calgary_merged['Cluster Labels'] == 2, calgary_merged.columns[[2] + list(range(5, calgar
cluster3_df
```

Out[178]:

	Neighborhood	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	9th Most Common Venue	10 C
1	Dalhousie, Edgemont, Hamptons, Hidden Valley	2	Convenience Store	Asian Restaurant	Gas Station	Café	Flea Market	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	D
2	Forest Lawn, Dover, Erin Woods	2	Smoke Shop	Bar	Convenience Store	Food Court	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	D
4	Lynnwood Ridge, Ogden, Foothills Industrial, G...	2	Convenience Store	Pizza Place	Diner	Clothing Store	Yoga Studio	Deli / Bodega	Department Store	Dim Sum Restaurant	Discount Store	D
5	Rosscarrock, Westgate, Wildwood, Shaganappi, S...	2	Sandwich Place	Pub	Indian Restaurant	Mexican Restaurant	Gas Station	Sports Bar	Pizza Place	Vietnamese Restaurant	Fast Food Restaurant	B

Cluster 4

In [179]:
```
cluster4_df = calgary_merged.loc[calgary_merged['Cluster Labels'] == 3, calgary_merged.columns[[2] + list(range(5, calgar
cluster4_df
```

Out[179]:

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
13	Martindale, Taradale, Falconridge, Saddle Ridge	3	Dog Run	Yoga Studio	Convenience Store	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dry Cleaner

Cluster 5

In [180]:
```
cluster5_df = calgary_merged.loc[calgary_merged['Cluster Labels'] == 4, calgary_merged.columns[[2] + list(range(5, calgar
cluster5_df
```

Out[180]:

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
16	Brentwood, Collingwood, Nose Hill	4	Asian Restaurant	Yoga Studio	Food Court	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Donut Shop

Cluster 6

In [185]:
```
cluster6_df = calgary_merged.loc[calgary_merged['Cluster Labels'] == 5, calgary_merged.columns[[2] + list(range(5, calgar
cluster6_df
```

Out[185]:

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
0	Penbrooke Meadows, Marlborough	5	Construction & Landscaping	Convenience Store	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Donut Shop	Dry Cleaner

Conclusion:

With this detailed analysis, any newcomer to Calgary can have a high level view at the distribution of neighborhoods (as well as their similarities), within the city.