Coursera Capstone Peergroup Project

June 12, 2019

In [3]: import numpy as np

```
import pandas as pd
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json

from geopy.geocoders import Nominatim

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

import matplotlib.cm as cm
import matplotlib.colors as colors

from sklearn.cluster import KMeans

import folium

print('Libraries imported.')
```

Libraries imported.

OBTAIN DATASET FROM TORONTO

In [4]: url='https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
```
df=pd.read_html(url, header=0)[0]
df.head()
```

Out[4]:

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

IGNORE CELLS WITH BOROUGHS AS "NOT ASSIGNED" BY DELETING AND RESET THE INDEX

```python
In [5]: df = df[df.Borough != 'Not assigned']
        df.reset_index(inplace = True)
        df.drop('index', axis=1,inplace=True)
        df.head()
```

```
Out[5]:   Postcode          Borough      Neighbourhood
      0     M3A         North York         Parkwoods
      1     M4A         North York  Victoria Village
      2     M5A  Downtown Toronto       Harbourfront
      3     M5A  Downtown Toronto        Regent Park
      4     M6A         North York  Lawrence Heights
```

LINKING TOGETHER CELLS BASED ON BOROUGH AND POSTCODE FOR NEIGHBORHOOD COLUMN

```python
In [7]: df_group = df.groupby(['Postcode','Borough'])['Neighbourhood'].apply(lambda x:', '.join(x))
        df_group2 = pd.DataFrame(df1)
        df_group2.reset_index(inplace = True)
        df_group2.head(12)
```

```
Out[7]:    Postcode      Borough                                   Neighbourhood
       0     M1B  Scarborough                                  Rouge, Malvern
       1     M1C  Scarborough          Highland Creek, Rouge Hill, Port Union
       2     M1E  Scarborough              Guildwood, Morningside, West Hill
       3     M1G  Scarborough                                          Woburn
       4     M1H  Scarborough                                       Cedarbrae
       5     M1J  Scarborough                            Scarborough Village
       6     M1K  Scarborough      East Birchmount Park, Ionview, Kennedy Park
       7     M1L  Scarborough               Clairlea, Golden Mile, Oakridge
       8     M1M  Scarborough    Cliffcrest, Cliffside, Scarborough Village West
       9     M1N  Scarborough                     Birch Cliff, Cliffside West
      10     M1P  Scarborough  Dorset Park, Scarborough Town Centre, Wexford ...
      11     M1R  Scarborough                             Maryvale, Wexford
```

```python
In [8]: df_group2.shape
```

```
Out[8]: (103, 3)
```

```python
In [ ]:
```