Segmenting and Clustering Neighbourhoods in Toronto

In [1]:
```
import pandas as pd
import numpy as np
import requests
```

In [8]:
```
wiki = 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_
M'
wiki_page = requests.get(wiki)

df = pd.read_html(wiki_page.content, header = 0)[0]
df = df[wiki_raw.Neighbourhood != 'Not assigned']
df.reset_index(inplace = True)
df.head()
```

Out[8]:

	index	Postal Code	Borough	Neighbourhood
0	2	M3A	North York	Parkwoods
1	3	M4A	North York	Victoria Village
2	4	M5A	Downtown Toronto	Regent Park, Harbourfront
3	5	M6A	North York	Lawrence Manor, Lawrence Heights
4	6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

In [9]:
```
df.groupby(['Postal Code']).first()
```

Out[9]:

Postal Code	index	Borough	Neighbourhood

Postal Code	index	Borough	Neighbourhood
M1B	9	Scarborough	Malvern, Rouge
M1C	18	Scarborough	Rouge Hill, Port Union, Highland Creek
M1E	27	Scarborough	Guildwood, Morningside, West Hill
M1G	36	Scarborough	Woburn
M1H	45	Scarborough	Cedarbrae
...	...	...	...
M9N	98	York	Weston
M9P	107	Etobicoke	Westmount
M9R	116	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...
M9V	143	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
M9W	152	Etobicoke	Northwest, West Humber - Clairville

103 rows × 3 columns

In [11]:
```
len(df['Postal Code'].unique())
```

Out[11]: 103

In [12]:
```
df[df['Borough'] == 'Not assigned']
```

Out[12]:

index	Postal Code	Borough	Neighbourhood

In [13]:
```
df.shape
```

Out[13]: (103, 4)

In [ ]:
```

```