Importing all necessary pacakges

In [1]:
```
import pandas as pd
import requests
import html5lib
from bs4 import BeautifulSoup
from tabulate import tabulate
import numpy as np
```

In [2]:
```
url = "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"
dfs = pd.read_html(url)
```

In [3]:
```
df = dfs[0]
df.head()
```

Out[3]:

	Postal Code	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	Regent Park, Harbourfront

Cleaning data

In [4]:
```
df.columns=['PostalCode', 'Borough','Neighbourhood']
```

Not assigned in column get replace by nan with help of numpy library and then droping that row at last reset the index

In [5]:
```
df = df.replace('Not assigned',np.nan).dropna(axis=1,how='all')
df= df.dropna()
df=df.reset_index(drop=True)
df
```

Out[5]:

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto	Church and Wellesley
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

103 rows × 3 columns

In [6]:
```
df.head(12)
```

Out[6]:

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...

Shape of table

In [7]:
```
df.shape
```

Out[7]:
```
(103, 3)
```

Adding two columns of Latitude and Longitude in existing data

In [9]:
```
df1=pd.read_csv('Geospatial_Coordinates.csv')
merged = df.merge(df1, on= 'PostalCode')
merged
```

Out[9]:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
...	...	...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

103 rows × 5 columns

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: