In [1]:
```
import pandas as pd
from bs4 import BeautifulSoup
import requests
import numpy as np
from geopy.geocoders import Nominatim # convert an address into latitude and lo
from pandas.io.json import json_normalize  # tranform JSON file into a pandas d
```

In [ ]:
```
```

In [4]:
```
from sklearn.cluster import KMeans

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
```

In [5]:
```
source = requests.get("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Ca
soup = BeautifulSoup(source, 'lxml')

table = soup.find("table")
table_rows = table.tbody.find_all("tr")

res = []
for tr in table_rows:
    td = tr.find_all("td")
    row = [tr.text for tr in td]

    # Only process the cells that have an assigned borough. Ignore cells with a
    if row != [] and row[1] != "Not assigned":
        # If a cell has a borough but a "Not assigned" neighborhood, then the n
        if "Not assigned" in row[2]:
            row[2] = row[1]
        res.append(row)

# Dataframe with 3 columns
df = pd.DataFrame(res, columns = ["PostalCode", "Borough", "Neighborhood"])
df.head()
```

Out[5]:

	PostalCode	Borough	Neighborhood
0	M1A\n	Not assigned\n	Not assigned\n
1	M2A\n	Not assigned\n	Not assigned\n
2	M3A\n	North York\n	Parkwoods\n
3	M4A\n	North York\n	Victoria Village\n