

In [10]:

```
pip install beautifulsoup4
```

Requirement already satisfied: beautifulsoup4 in /srv/conda/envs/notebook/lib/python3.6/site-packages (4.9.1)
Requirement already satisfied: soupsieve>1.2 in /srv/conda/envs/notebook/lib/python3.6/site-packages (from beautifulsoup4) (2.0.1)
Note: you may need to restart the kernel to use updated packages.

In [11]:

```
pip install html5lib
```

Collecting html5lib
 Downloading html5lib-1.1-py2.py3-none-any.whl (112 kB)
 |██████████████████████████████████████| 112 kB 3.3 MB/s eta 0:00:01
Requirement already satisfied: webencodings in /srv/conda/envs/notebook/lib/python3.6/site-packages (from html5lib) (0.5.1)
Requirement already satisfied: six>=1.9 in /srv/conda/envs/notebook/lib/python3.6/site-packages (from html5lib) (1.15.0)
Installing collected packages: html5lib
Successfully installed html5lib-1.1
Note: you may need to restart the kernel to use updated packages.

In [1]:

```
import numpy as np # library to handle data in a vectorized manner
import pandas as pd # library for data analysis
import requests
from bs4 import BeautifulSoup

print('Libraries imported.')
```

Libraries imported.

In [12]:

```
URL = "https://www.wikizeroo.org/index.php?q=aHR0cHM6Ly91bi53aWtpcGVkawEub3JnL3dpa2kvTG1zdF9vZl9wb3N0YWxfY29kZXNfb2ZfQ2FuYWRhO19N"
r = requests.get(URL)

soup = BeautifulSoup(r.content, 'html5lib')
table = soup.find('div', attrs = {'id':'container'})

# print(soup.prettify())
print('Page Scrapped.')
```

Page Scrapped.

In [8]:

```
postalCodes = [];
boroughs= [];
neighborhoods = [];
columnNum = 1;
passVal = False

for row in soup.find_all('td'):
    for cell in row:
        if cell.string and cell.string[0].isalpha() and len(cell.string) > 2:
            passVal = False
            if columnNum == 1:
                if passVal == False and cell.string[1].isdigit():
                    postalCodes.append(cell.string);
                    columnNum = 2
            else:
                continue
        elif columnNum == 2 :
            if cell.string == 'Not assigned':
                passVal = True
                del postalCodes[-1]
                columnNum = 1
                continue
            else:
                boroughs.append(cell.string);
                columnNum = 3
        elif columnNum == 3 :
            if cell.string == 'Not assigned\n':
                neighborhoods.append(boroughs[-1])
            else:
                neighborhoods.append(cell.string);
                columnNum = 1

print('Data Collected.')
```

Data Collected.

In [14]:

```
# define the dataframe columns
column_names = ['PostalCode', 'Borough', 'Neighborhood']

# instantiate the dataframe
neighbors = pd.DataFrame(columns=column_names)

neighbors
```

Out[14]:

PostalCode	Borough	Neighborhood
------------	---------	--------------

In [15]:

```
for data in range(len(neighborhoods)):
    code = postalCodes[data]
    borough = boroughs[data]
    neighborhood_name = neighborhoods[data]

    neighbors = neighbors.append({ 'PostalCode': code,
                                   'Borough': borough,
                                   'Neighborhood': neighborhood_name}, ignore_index=True)
```

In [17]:

```
neighbors
```

Out[17]:

PostalCode	Borough	Neighborhood
------------	---------	--------------

In [18]:

```
neighbors.shape
```

Out[18]:

```
(0, 3)
```

In []: