SegmtClstringNhdsToronto

Segmenting and Clustering Neighborhoods in Toronto week 3 assignment

VERSION AUTHOR: hari kumar
LAST UPDATED: 12 Sep 2019, 6:37 AM
LANGUAGE: Python 3.6

Assignment IBM - Segmenting and Clustering Neighborhoods in Toronto

To scrape the given Wikipedia page and create a Dataframe. Transform the data on Wiki page into pandas dataframe

Explore and cluster the neighborhoods in Toronto

Import Libraries ¶

In [1]:
```
# !conda install -c anaconda beautifulsoup4 -y
from bs4 import BeautifulSoup
import requests
import pandas as pd
```

Scrap List of postal codes of Canada wiki page content by using BeautifulSoup

In [2]:
```
# download url data from internet
url = "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"
source = requests.get(url).text
Canada_data = BeautifulSoup(source, 'lxml')
#Canada_data
```

Convert content of PostalCode HTML table as dataframe

In [3]:
```
# creat a new Dataframe
column_names = ['Postalcode','Borough','Neighborhood']
toronto = pd.DataFrame(columns = column_names)

# loop through to find postcode, borough, neighborhood
content = Canada_data.find('div', class_='mw-parser-output')
table = content.table.tbody
postcode = 0
borough = 0
```