Capstone Project - The Battle of Neighborhoods - Taichung version

This project is a part of IBM data science; you will find in this post an overview of my final capstone project.

In this assignment, I will go through the problem description (**Introduction/Business Problem**), data set preparation and final to analysis and overview these data step by step. Detailed code is given in Github and link can be found at the end of the post.

1 Introduction/Business Problem

1.1 Background

Taiwan is a famous nation in the world. It was very small $(36,197 \text{ km}^2)$ but with more than 23,600,000 population. Taiwan with the high metropolitan GDP in the world and it belonged a developed country.

There was COVID-19 spread in the world but it seems not impact to Taiwan. Here with the best epidemic prevention concept and medical technology so the infection rate and mortality rate is the lowest in the world.

With estimated 11,840,000 overseas visitors every year. It's famous with delicious foods and beautiful landscapes. Many travelers needed to know some leisure and entertainment places when they first visit to Taiwan.

These characteristics also attract many immigrations come to Taiwan.

Taichung is a second big city in Taiwan, its' urban scale was smaller than Taipei only. Taichung retained the advantage of Taipei like: convenient transportation/convenient life/good job opportunity/medical system...etc. Many immigrations will choose settle down in Taichung.

1.2 Business Problem

Taichung with 29 different districts, some districts are commercial and some are convenient and some are leisure. These districts are so difference!!!

In this article will help immigrations to understand and overview the Taichung city. Let them to choose the district which they want to live or travel.

2 Data and Data Preparation

2.1 Data Clean and Preprocess:

2.1.1 Define CSV Process Class

2.1.2 Define Json Process Class

```
In [5]: class JsonProcess(object):

def __init__(self):
    #self.address = "C:\\Users\\Brian\\Desktop\\Ibm_coursera\\Coursera_Capstone\\week4\\"

self.address = "C:\\Users\\brian\\Desktop\\Coursera_Capstone\\week4\\"

def load_from_json(self, filename):
    json_file = open(self.address + filename, "r", encoding="utf-8")
    j = json_file.close()
    dict_data = json.loads(j)
    return dict_data

def dictionary_to_json(self, dictionary):
    json_data = json.dumps(dictionary)
    return json_data
```

2.1.3 Define Get Data from Web Class

```
In [6]:

class GetDateFromWeb(object):

def get_taichung_info(self):
    response = requests.get('https://zh.wikipedia.org/wiki/臺中市#人口')
    content = response.content
    df = pd.read.html(content, encoding='utf-8')
    taichung_info_df = df[11]

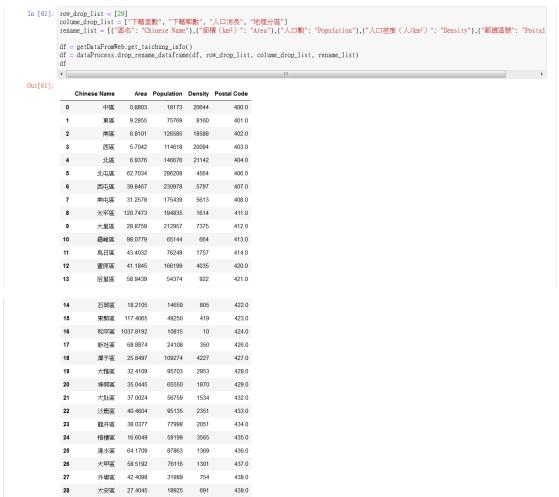
return taichung_info_df
```

2.1.4 Define DataFrame Process Class

2.2 Taichung wiki:

Data source: https://zh.wikipedia.org/wiki/臺中市#人口

Description: This data set included 29 districts in Taichung city. There are some fields we need like: Chinese Name, Area, Density and Postal Code.



2.3 English - Chinese glossary of districts in Taiwan:

Data source: http://gn.moi.gov.tw/geonames/Translation/Translation.aspx

Description: Government help to translate the city or districts name between English and Chinese

```
In [12]: taichung_english_name_df = csvProcess.load_from_csv("Taichung_English_Name.csv")
        print(taichung_english_name_df.shape)
taichung_english_name_df.head(30)
Out[12]:
            Chinese Name
         0 南屯區 Nantun District
                  西山區
         2
               北屯區
                          Beitun District
                    西區
          4 北區 North District
                  霧峰區 Wufeng District
              豐原區 Fengyuan District
          8 潭子區 Tanzi District
         10
              清水區 Qingshui District
                神岡區 Shengang District
         12
                  鳥日區
         14
                東勢區 Dongshi District
                  和平區 Heping District
                  和平區 Heping District
         16 沙鹿區 Shalu District
         17
                  后里區
         18
                 石岡區 Shigang District
         19
                  外埔區
                          Waipu District
                太平區 Taiping District
         21
                  大雅區
                           Daya District
         22
                 大里區
         23
                  大肚區
                          Dadu District
         25
                  大甲區
                           Daiia District
                  中區 Central District
                    東區
                            Fast District
```

2.4 Latitude and longitude in Taiwan:

Data source:

https://www.astrocode.net/%E5%8F%B0%E7%81%A3%E5%90%84%E7%B8%A3%E5%B8%82%E5%9C%B0%E5%8D%80%E7%B6%93%E7%B7%AF%E5%BA%A6/

Description: This is a XML file and we can transfer to JSON file. The file included all districts of latitude and longitude in Taiwan.

2.5 Venues in each neighborhood of Taichung City:

Data source: Foursquare APIs

Description: We will get all the venues in each neighborhood by using this API and filter

these venues to get restaurants and other data.