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. **Background**

**Taiwan is a famous nation in the world. It was very small (36,197** km2 **) 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. **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.**

1. **Data and Data Preparation**
   1. Data Clean and Preprocess:
      1. Define CSV Process Class



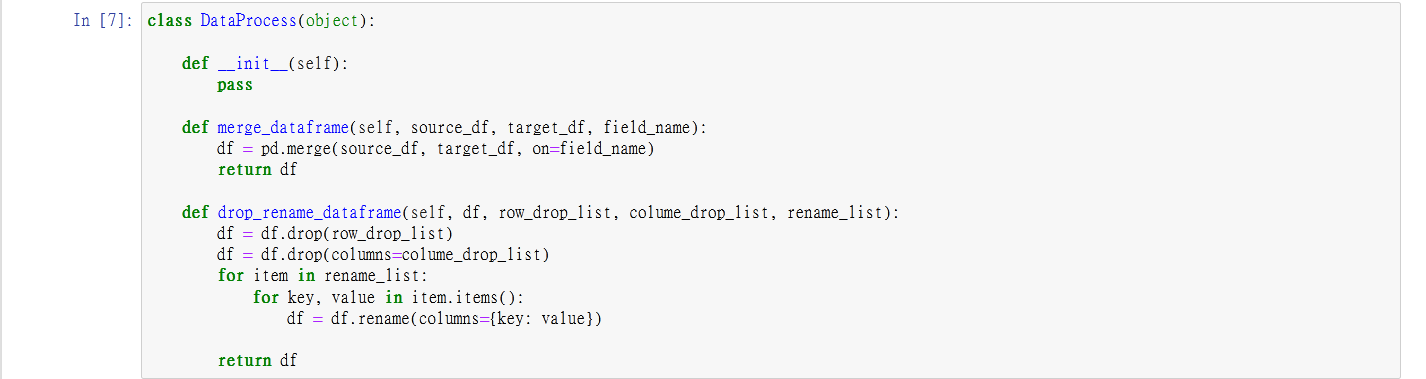
* + 1. Define Json Process Class



* + 1. Define Get Data from Web Class



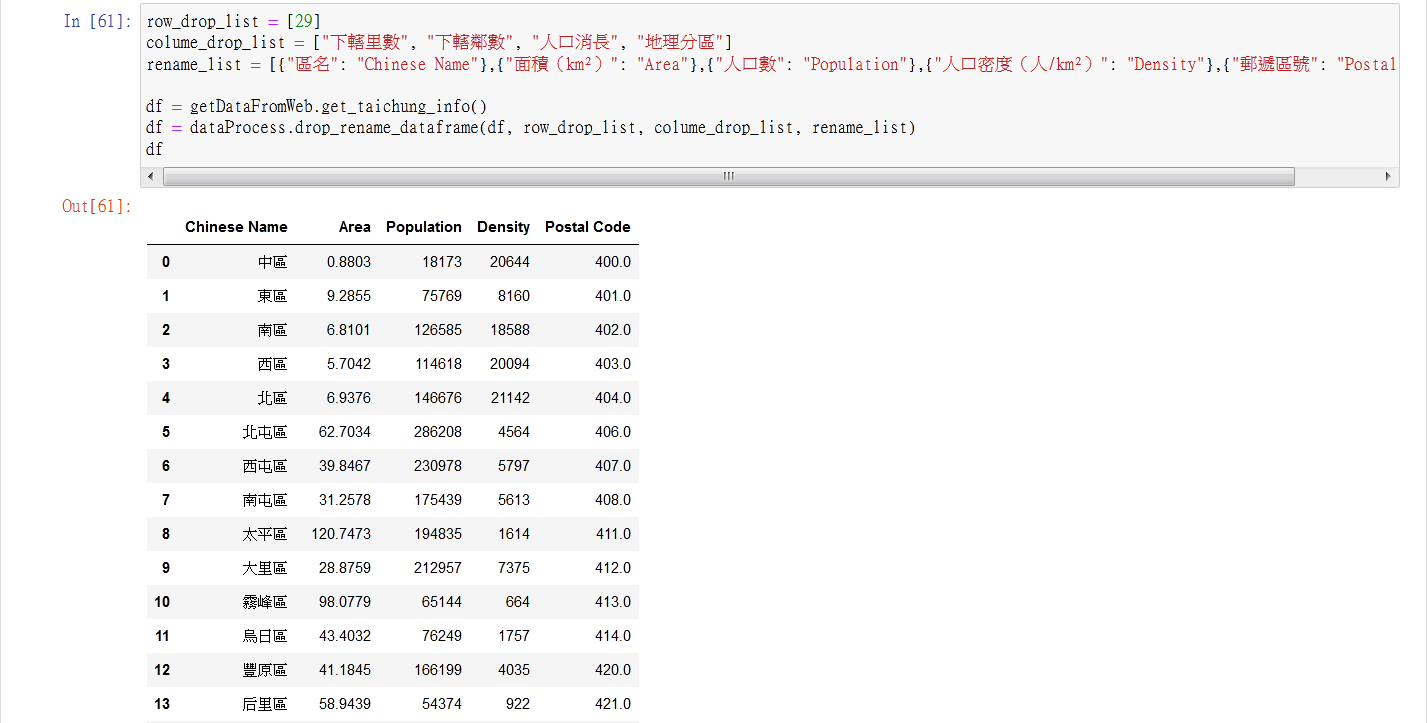
* + 1. Define DataFrame Process Class

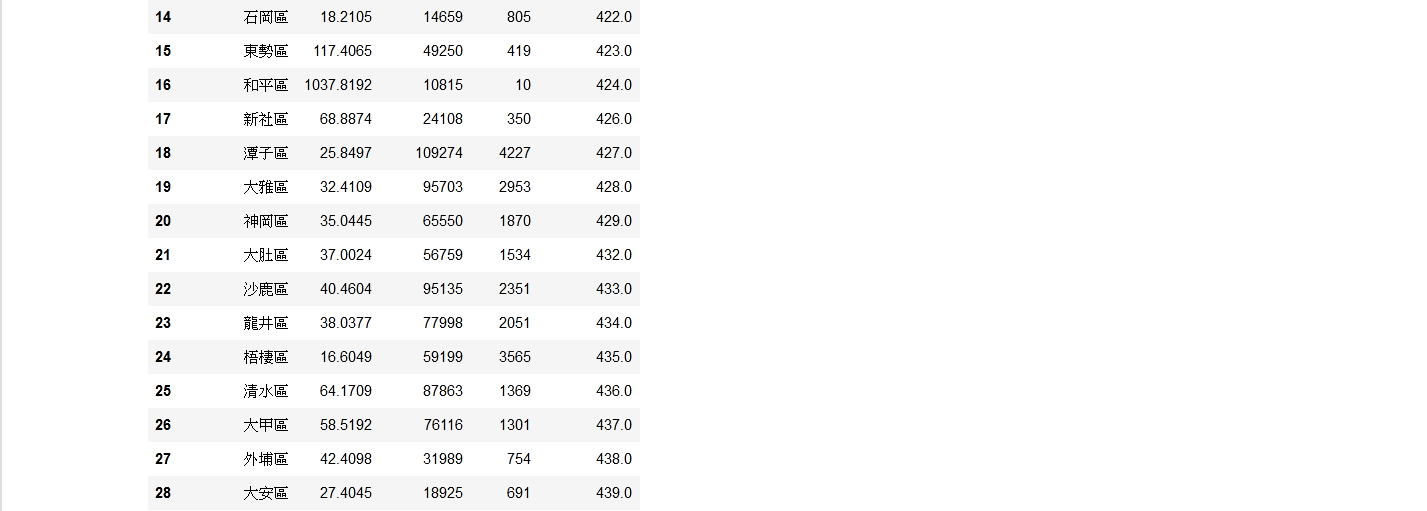


* 1. Taichung wiki:

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

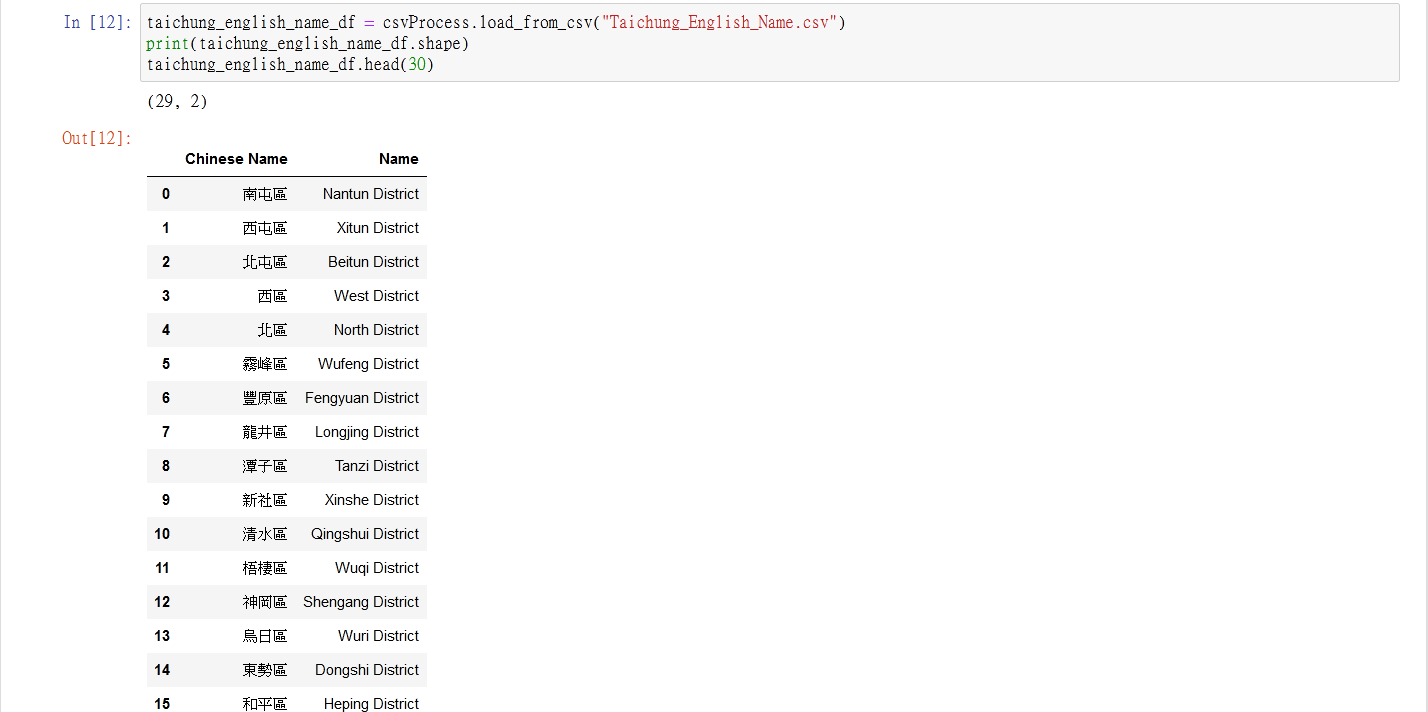
Description: This data set included 29 **districts in Taichung city. There are some** fields we need like: Chinese Name, Area, Density and Postal Code.





* 1. 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





* 1. 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.



* 1. 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.