How is the similarity between Taipei and Kaohsiung?

Chia Wei Lee

MAR 25, 2019

1. Introduction
   * Background

Taipei and Kaohsiung are both well-known cities in Taiwan. Taipei locates in the north side, and Kaohsiung is the south side of Taiwan. This report is going to find the similarity and differences between the two cities.

* + Problem

But do we know how similar are those two cities? People in Taiwan always have a common concept that Taipei is more convenient and modem than Kaohsiung. But it is true or just a common misunderstanding? To find out the answer there are very features we can use to dig into.

* + Interest

In this report, we based on the diversity of point of interests to be the feature and compare those two cities and figure out if those two cities are similar or have huge differences. And expect to purpose a common method to compare not only those two cities, but also can apply to all well know cities all over the world.

1. Data acquisition and cleaning
   * Data Source
     + List of Taiwan district overall latitude and longitude

* Description: Information of Taiwan's overall latitude and longitude
* URL:

<https://data.gov.tw/dataset/25489>

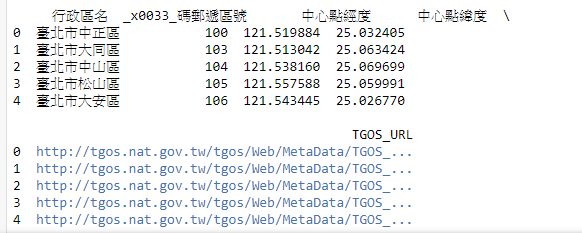
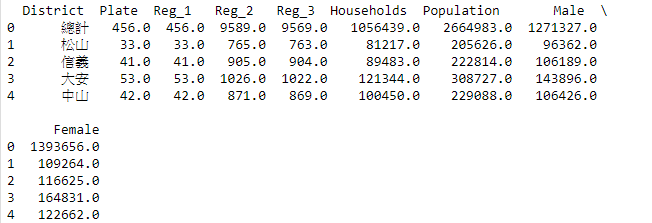
* + - Taipei government data
* Description: Taipei Government Dataset
* URL : <https://ca.gov.taipei/News_Content.aspx?n=8693DC9620A1AABF&sms=D19E9582624D83CB&s=EE7D5719108F4026>
* Kaohsiung government data
  + - Description: Kaohsiung Government Dataset
    - URL :

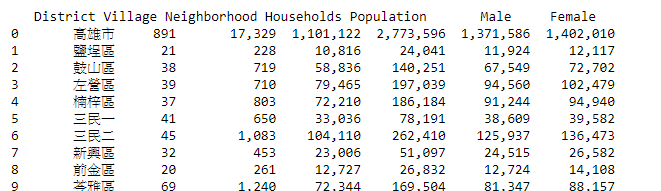
<http://cabu.kcg.gov.tw/Stat/StatRpts/StatRpt1.aspx>

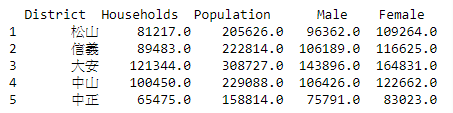
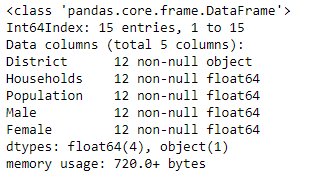
* + - Foursquare API
* Description: Retrieve the point of interests information.
* URL:

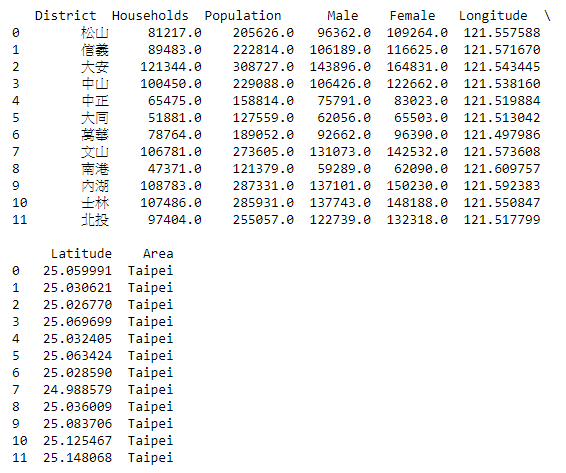
<https://developer.foursquare.com/docs/api/venues/details>

1. Data acquisition and cleaning
   * + Data Source

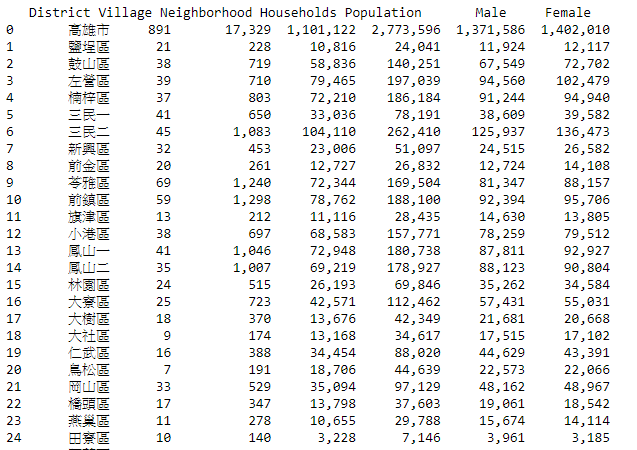
* Most data are from open information or Kaohsiung and Taipei government. Frist of all, fetch the Taiwan's overall latitude and longitude[[link](https://data.gov.tw/dataset/25489)] 
* Get the CVS format of district, households, population information of Taipei[[link](https://ca.gov.taipei/News_Content.aspx?n=8693DC9620A1AABF&sms=D19E9582624D83CB&s=EE7D5719108F4026)]. 
* Get the web site information of Kaohsiung[[link](http://cabu.kcg.gov.tw/Stat/StatRpts/StatRpt1.aspx)] by using BeautifulSoup library.



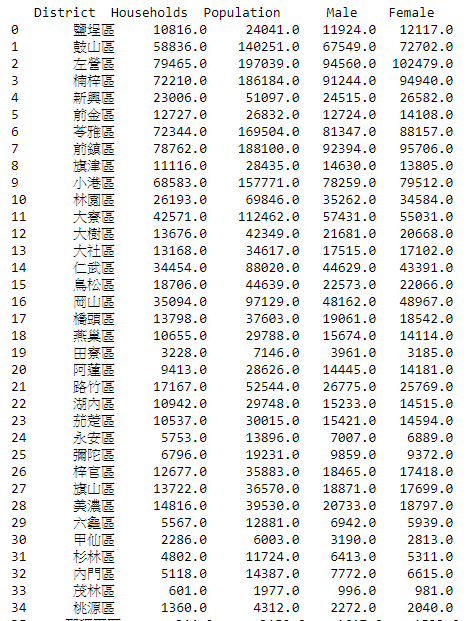
* + - Data Cleaning
* For Taiwan's overall latitude and longitude
  + Rename the column and only fetch the useful columns such as District, latitude and longitude.
  + Fetch only the Taipei and Kaohsiung’s information.
* For Taipei information
  + Rename the column and only fetch the useful columns such as District, Poulation, Male and Female. 
  + By checking the dataframe’s info. There is no empty and NAN values. 
  + Merge the Taipei District’s latitude and longitude and city information
  + And add addition column ‘Area’ for one hot encoding.



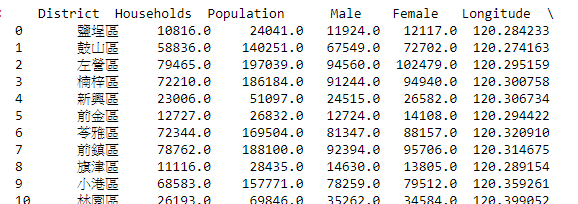
* For Kaohsiung
  + Retrieve the web site and set the information into dataframe.



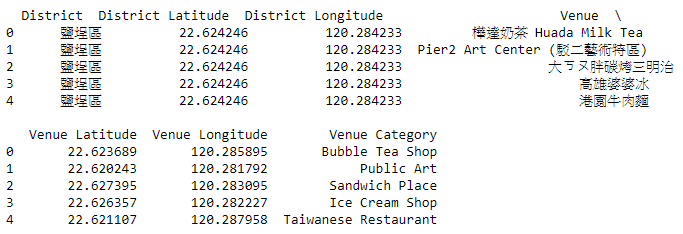
* + Get ride of useless columns, and merge the separate row data



* + Merge the Kaohsiung District’s latitude and longitude and city information
  + And add addition column ‘Area’ for one hot encoding.

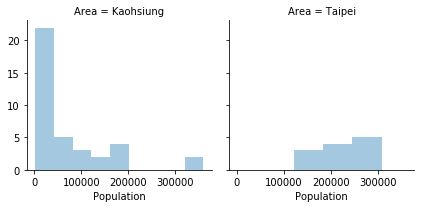


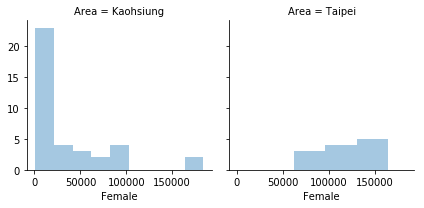
* Get venues data
  + By leveraging Foursquare API. We fit the longitude, latitude and radius 1500.



* + - Feature Selection
      * We only choose the population, Male, Female and venues data for feature.

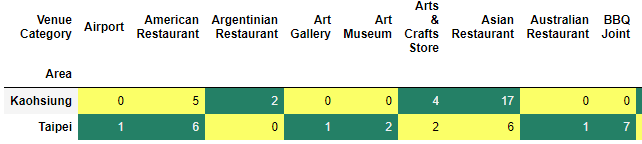
1. Exploratory Data Analysis
   * + By population



* + - By Female
    - By MaleA screenshot of a tree

      Description generated with high confidence
    - By HouseholdsA screenshot of a cell phone

      Description generated with very high confidence
    - Brief observation from basic information
      * Taipei city is more centralized. By the graph, the major population, male, female and households are distributed in the small range of districts, bit Kaohsiung have different situation. And Kaohsiung have lots of districts which have small population. Such as population differentials among the major district groups. And the survey reveals the reflect the same observation in male and female as well.
    - Venue Category
      * Taipei is 153 and Kaohsiung is 115



**Detail graph please check ipnb source**

* + - Brief observation from Venue Category
      * Taipei has more varieties of venues compare to Kaohsiung.

1. Conclusions
   * + According to the comparison of population, female, male, venues categories. We can say that districts in Taipei does not have huge gap. The population and related data distribute to over all districts, but in the other site. Kaohsiung have gap between districts.