

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 modern 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.

2. 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>

3. 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[\[link\]](#)

| | 行政區名 | _x0033_碼郵遞區號 | 中心點經度 | 中心點緯度 | \ |
|----------|---|--------------|------------|-----------|---|
| 0 | 臺北市中正區 | 100 | 121.519884 | 25.032405 | |
| 1 | 臺北市大同區 | 103 | 121.513042 | 25.063424 | |
| 2 | 臺北市中山區 | 104 | 121.538160 | 25.069699 | |
| 3 | 臺北市松山區 | 105 | 121.557588 | 25.059991 | |
| 4 | 臺北市大安區 | 106 | 121.543445 | 25.026770 | |
| TGOS_URL | | | | | |
| 0 | http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_... | | | | |
| 1 | http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_... | | | | |
| 2 | http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_... | | | | |
| 3 | http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_... | | | | |
| 4 | http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_... | | | | |

- Get the CVS format of district, households, population information of Taipei[\[link\]](#).

| | District | Plate | Reg_1 | Reg_2 | Reg_3 | Households | Population | Male | \ |
|--------|-----------|-------|-------|--------|--------|------------|------------|-----------|---|
| 0 | 總計 | 456.0 | 456.0 | 9589.0 | 9569.0 | 1056439.0 | 2664983.0 | 1271327.0 | |
| 1 | 松山 | 33.0 | 33.0 | 765.0 | 763.0 | 81217.0 | 205626.0 | 96362.0 | |
| 2 | 信義 | 41.0 | 41.0 | 905.0 | 904.0 | 89483.0 | 222814.0 | 106189.0 | |
| 3 | 大安 | 53.0 | 53.0 | 1026.0 | 1022.0 | 121344.0 | 308727.0 | 143896.0 | |
| 4 | 中山 | 42.0 | 42.0 | 871.0 | 869.0 | 100450.0 | 229088.0 | 106426.0 | |
| Female | | | | | | | | | |
| 0 | 1393656.0 | | | | | | | | |
| 1 | 109264.0 | | | | | | | | |
| 2 | 116625.0 | | | | | | | | |
| 3 | 164831.0 | | | | | | | | |
| 4 | 122662.0 | | | | | | | | |

- Get the web site information of Kaohsiung[\[link\]](#) by using BeautifulSoup library.

| | District | Village | Neighborhood | Households | Population | Male | Female |
|---|----------|---------|--------------|------------|------------|-----------|-----------|
| 0 | 高雄市 | 891 | 17,329 | 1,101,122 | 2,773,596 | 1,371,586 | 1,402,010 |
| 1 | 鹽埕區 | 21 | 228 | 10,816 | 24,041 | 11,924 | 12,117 |
| 2 | 鼓山區 | 38 | 719 | 58,836 | 140,251 | 67,549 | 72,702 |
| 3 | 左營區 | 39 | 710 | 79,465 | 197,039 | 94,560 | 102,479 |
| 4 | 楠梓區 | 37 | 803 | 72,210 | 186,184 | 91,244 | 94,940 |
| 5 | 三民一 | 41 | 650 | 33,036 | 78,191 | 38,609 | 39,582 |
| 6 | 三民二 | 45 | 1,083 | 104,110 | 262,410 | 125,937 | 136,473 |
| 7 | 新興區 | 32 | 453 | 23,006 | 51,097 | 24,515 | 26,582 |
| 8 | 前金區 | 20 | 261 | 12,727 | 26,832 | 12,724 | 14,108 |
| 9 | 苓雅區 | 69 | 1,240 | 72,344 | 169,504 | 81,347 | 88,157 |

- 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.

| | District | Households | Population | Male | Female |
|---|----------|------------|------------|----------|----------|
| 1 | 松山 | 81217.0 | 205626.0 | 96362.0 | 109264.0 |
| 2 | 信義 | 89483.0 | 222814.0 | 106189.0 | 116625.0 |
| 3 | 大安 | 121344.0 | 308727.0 | 143896.0 | 164831.0 |
| 4 | 中山 | 100450.0 | 229088.0 | 106426.0 | 122662.0 |
| 5 | 中正 | 65475.0 | 158814.0 | 75791.0 | 83023.0 |

- By checking the dataframe's info. There is no empty and NAN values.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 15 entries, 1 to 15
Data columns (total 5 columns):
District      12 non-null object
Households    12 non-null float64
Population    12 non-null float64
Male          12 non-null float64
Female        12 non-null float64
dtypes: float64(4), object(1)
memory usage: 720.0+ bytes
```

- Merge the Taipei District's latitude and longitude and city information
- And add addition column 'Area' for one hot encoding.

| | District | Households | Population | Male | Female | Longitude \ |
|----|----------|------------|------------|----------|----------|-------------|
| 0 | 松山 | 81217.0 | 205626.0 | 96362.0 | 109264.0 | 121.557588 |
| 1 | 信義 | 89483.0 | 222814.0 | 106189.0 | 116625.0 | 121.571670 |
| 2 | 大安 | 121344.0 | 308727.0 | 143896.0 | 164831.0 | 121.543445 |
| 3 | 中山 | 100450.0 | 229088.0 | 106426.0 | 122662.0 | 121.538160 |
| 4 | 中正 | 65475.0 | 158814.0 | 75791.0 | 83023.0 | 121.519884 |
| 5 | 大同 | 51881.0 | 127559.0 | 62056.0 | 65503.0 | 121.513042 |
| 6 | 萬華 | 78764.0 | 189052.0 | 92662.0 | 96390.0 | 121.497986 |
| 7 | 文山 | 106781.0 | 273605.0 | 131073.0 | 142532.0 | 121.573608 |
| 8 | 南港 | 47371.0 | 121379.0 | 59289.0 | 62090.0 | 121.609757 |
| 9 | 內湖 | 108783.0 | 287331.0 | 137101.0 | 150230.0 | 121.592383 |
| 10 | 士林 | 107486.0 | 285931.0 | 137743.0 | 148188.0 | 121.550847 |
| 11 | 北投 | 97404.0 | 255057.0 | 122739.0 | 132318.0 | 121.517799 |

| | Latitude | Area |
|----|-----------|--------|
| 0 | 25.059991 | Taipei |
| 1 | 25.030621 | Taipei |
| 2 | 25.026770 | Taipei |
| 3 | 25.069699 | Taipei |
| 4 | 25.032405 | Taipei |
| 5 | 25.063424 | Taipei |
| 6 | 25.028590 | Taipei |
| 7 | 24.988579 | Taipei |
| 8 | 25.036009 | Taipei |
| 9 | 25.083706 | Taipei |
| 10 | 25.125467 | Taipei |
| 11 | 25.148068 | Taipei |

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

| | District | Village | Neighborhood | Households | Population | Male | Female |
|----|----------|---------|--------------|------------|------------|-----------|-----------|
| 0 | 高雄市 | 891 | 17,329 | 1,101,122 | 2,773,596 | 1,371,586 | 1,402,010 |
| 1 | 鹽埕區 | 21 | 228 | 10,816 | 24,041 | 11,924 | 12,117 |
| 2 | 鼓山區 | 38 | 719 | 58,836 | 140,251 | 67,549 | 72,702 |
| 3 | 左營區 | 39 | 710 | 79,465 | 197,039 | 94,560 | 102,479 |
| 4 | 楠梓區 | 37 | 803 | 72,210 | 186,184 | 91,244 | 94,940 |
| 5 | 三民一 | 41 | 650 | 33,036 | 78,191 | 38,609 | 39,582 |
| 6 | 三民二 | 45 | 1,083 | 104,110 | 262,410 | 125,937 | 136,473 |
| 7 | 新興區 | 32 | 453 | 23,006 | 51,097 | 24,515 | 26,582 |
| 8 | 前金區 | 20 | 261 | 12,727 | 26,832 | 12,724 | 14,108 |
| 9 | 苓雅區 | 69 | 1,240 | 72,344 | 169,504 | 81,347 | 88,157 |
| 10 | 前鎮區 | 59 | 1,298 | 78,762 | 188,100 | 92,394 | 95,706 |
| 11 | 旗津區 | 13 | 212 | 11,116 | 28,435 | 14,630 | 13,805 |
| 12 | 小港區 | 38 | 697 | 68,583 | 157,771 | 78,259 | 79,512 |
| 13 | 鳳山一 | 41 | 1,046 | 72,948 | 180,738 | 87,811 | 92,927 |
| 14 | 鳳山二 | 35 | 1,007 | 69,219 | 178,927 | 88,123 | 90,804 |
| 15 | 林園區 | 24 | 515 | 26,193 | 69,846 | 35,262 | 34,584 |
| 16 | 大寮區 | 25 | 723 | 42,571 | 112,462 | 57,431 | 55,031 |
| 17 | 大樹區 | 18 | 370 | 13,676 | 42,349 | 21,681 | 20,668 |
| 18 | 大社區 | 9 | 174 | 13,168 | 34,617 | 17,515 | 17,102 |
| 19 | 仁武區 | 16 | 388 | 34,454 | 88,020 | 44,629 | 43,391 |
| 20 | 鳥松區 | 7 | 191 | 18,706 | 44,639 | 22,573 | 22,066 |
| 21 | 岡山區 | 33 | 529 | 35,094 | 97,129 | 48,162 | 48,967 |
| 22 | 橋頭區 | 17 | 347 | 13,798 | 37,603 | 19,061 | 18,542 |
| 23 | 燕巢區 | 11 | 278 | 10,655 | 29,788 | 15,674 | 14,114 |
| 24 | 田寮區 | 10 | 140 | 3,228 | 7,146 | 3,961 | 3,185 |

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

| | District | Households | Population | Male | Female |
|----|----------|------------|------------|---------|----------|
| 0 | 鹽埕區 | 10816.0 | 24041.0 | 11924.0 | 12117.0 |
| 1 | 鼓山區 | 58836.0 | 140251.0 | 67549.0 | 72702.0 |
| 2 | 左營區 | 79465.0 | 197039.0 | 94560.0 | 102479.0 |
| 3 | 楠梓區 | 72210.0 | 186184.0 | 91244.0 | 94940.0 |
| 4 | 新興區 | 23006.0 | 51097.0 | 24515.0 | 26582.0 |
| 5 | 前金區 | 12727.0 | 26832.0 | 12724.0 | 14108.0 |
| 6 | 苓雅區 | 72344.0 | 169504.0 | 81347.0 | 88157.0 |
| 7 | 前鎮區 | 78762.0 | 188100.0 | 92394.0 | 95706.0 |
| 8 | 旗津區 | 11116.0 | 28435.0 | 14630.0 | 13805.0 |
| 9 | 小港區 | 68583.0 | 157771.0 | 78259.0 | 79512.0 |
| 10 | 林園區 | 26193.0 | 69846.0 | 35262.0 | 34584.0 |
| 11 | 大寮區 | 42571.0 | 112462.0 | 57431.0 | 55031.0 |
| 12 | 大樹區 | 13676.0 | 42349.0 | 21681.0 | 20668.0 |
| 13 | 大社區 | 13168.0 | 34617.0 | 17515.0 | 17102.0 |
| 14 | 仁武區 | 34454.0 | 88020.0 | 44629.0 | 43391.0 |
| 15 | 鳥松區 | 18706.0 | 44639.0 | 22573.0 | 22066.0 |
| 16 | 岡山區 | 35094.0 | 97129.0 | 48162.0 | 48967.0 |
| 17 | 橋頭區 | 13798.0 | 37603.0 | 19061.0 | 18542.0 |
| 18 | 燕巢區 | 10655.0 | 29788.0 | 15674.0 | 14114.0 |
| 19 | 田寮區 | 3228.0 | 7146.0 | 3961.0 | 3185.0 |
| 20 | 阿蓮區 | 9413.0 | 28626.0 | 14445.0 | 14181.0 |
| 21 | 路竹區 | 17167.0 | 52544.0 | 26775.0 | 25769.0 |
| 22 | 湖內區 | 10942.0 | 29748.0 | 15233.0 | 14515.0 |
| 23 | 茄萣區 | 10537.0 | 30015.0 | 15421.0 | 14594.0 |
| 24 | 永安區 | 5753.0 | 13896.0 | 7007.0 | 6889.0 |
| 25 | 彌陀區 | 6796.0 | 19231.0 | 9859.0 | 9372.0 |
| 26 | 梓官區 | 12677.0 | 35883.0 | 18465.0 | 17418.0 |
| 27 | 旗山區 | 13722.0 | 36570.0 | 18871.0 | 17699.0 |
| 28 | 美濃區 | 14816.0 | 39530.0 | 20733.0 | 18797.0 |
| 29 | 六龜區 | 5567.0 | 12881.0 | 6942.0 | 5939.0 |
| 30 | 甲仙區 | 2286.0 | 6003.0 | 3190.0 | 2813.0 |
| 31 | 杉林區 | 4802.0 | 11724.0 | 6413.0 | 5311.0 |
| 32 | 內門區 | 5118.0 | 14387.0 | 7772.0 | 6615.0 |
| 33 | 茂林區 | 601.0 | 1977.0 | 996.0 | 981.0 |
| 34 | 桃源區 | 1360.0 | 4312.0 | 2272.0 | 2040.0 |

- Merge the Kaohsiung District's latitude and longitude and city information
- And add addition column 'Area' for one hot encoding.

| | District | Households | Population | Male | Female | Longitude \ |
|----|----------|------------|------------|---------|----------|-------------|
| 0 | 鹽埕區 | 10816.0 | 24041.0 | 11924.0 | 12117.0 | 120.284233 |
| 1 | 鼓山區 | 58836.0 | 140251.0 | 67549.0 | 72702.0 | 120.274163 |
| 2 | 左營區 | 79465.0 | 197039.0 | 94560.0 | 102479.0 | 120.295159 |
| 3 | 楠梓區 | 72210.0 | 186184.0 | 91244.0 | 94940.0 | 120.300758 |
| 4 | 新興區 | 23006.0 | 51097.0 | 24515.0 | 26582.0 | 120.306734 |
| 5 | 前金區 | 12727.0 | 26832.0 | 12724.0 | 14108.0 | 120.294422 |
| 6 | 苓雅區 | 72344.0 | 169504.0 | 81347.0 | 88157.0 | 120.320910 |
| 7 | 前鎮區 | 78762.0 | 188100.0 | 92394.0 | 95706.0 | 120.314675 |
| 8 | 旗津區 | 11116.0 | 28435.0 | 14630.0 | 13805.0 | 120.289154 |
| 9 | 小港區 | 68583.0 | 157771.0 | 78259.0 | 79512.0 | 120.359261 |
| 10 | 林園區 | 26193.0 | 69846.0 | 35262.0 | 34584.0 | 120.399052 |

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

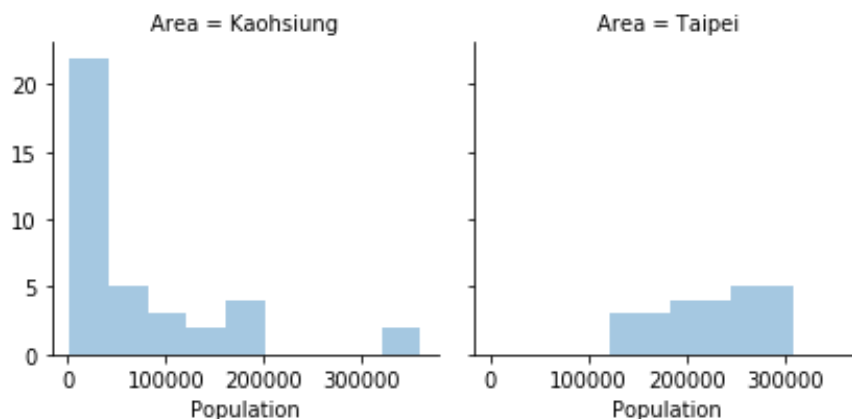
| District | District | Latitude | District | Longitude | Venue \ |
|----------|----------|-----------|------------|------------------|----------------|
| 0 | 鹽埕區 | 22.624246 | 120.284233 | 樺達奶茶 | Huada Milk Tea |
| 1 | 鹽埕區 | 22.624246 | 120.284233 | Pier2 Art Center | (駁二藝術特區) |
| 2 | 鹽埕區 | 22.624246 | 120.284233 | 大ㄚ又胖碳烤三明治 | |
| 3 | 鹽埕區 | 22.624246 | 120.284233 | 高雄婆婆冰 | |
| 4 | 鹽埕區 | 22.624246 | 120.284233 | 港園牛肉麵 | |

| Venue | Latitude | Venue | Longitude | Venue | Category |
|-------|-----------|------------|----------------------|-------|----------|
| 0 | 22.623689 | 120.285895 | Bubble Tea Shop | | |
| 1 | 22.620243 | 120.281792 | Public Art | | |
| 2 | 22.627395 | 120.283095 | Sandwich Place | | |
| 3 | 22.626357 | 120.282227 | Ice Cream Shop | | |
| 4 | 22.621107 | 120.287958 | Taiwanese Restaurant | | |

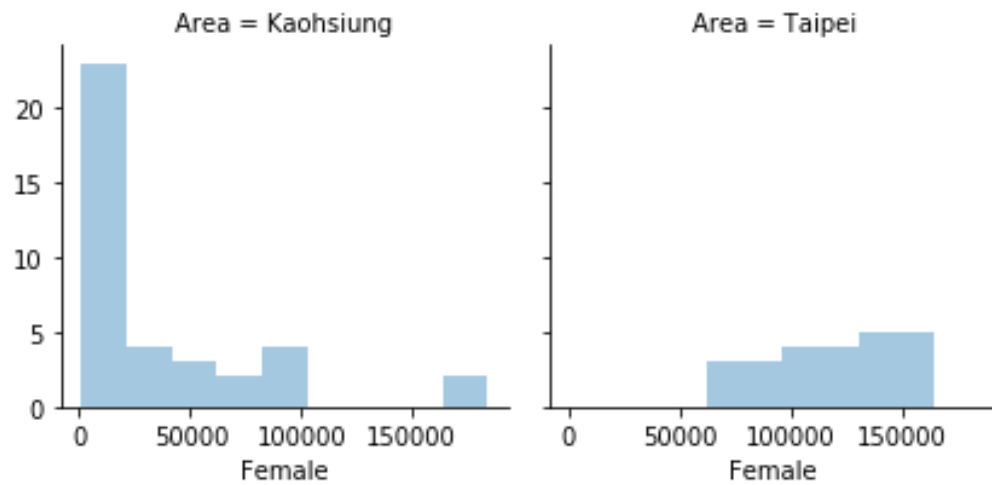
- Feature Selection
 - We only choose the population, Male, Female and venues data for feature.

4. Exploratory Data Analysis

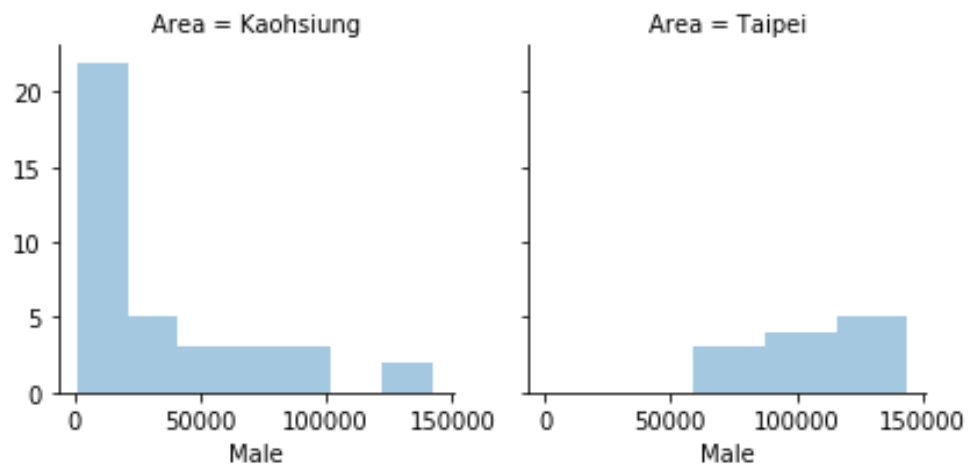
- By population



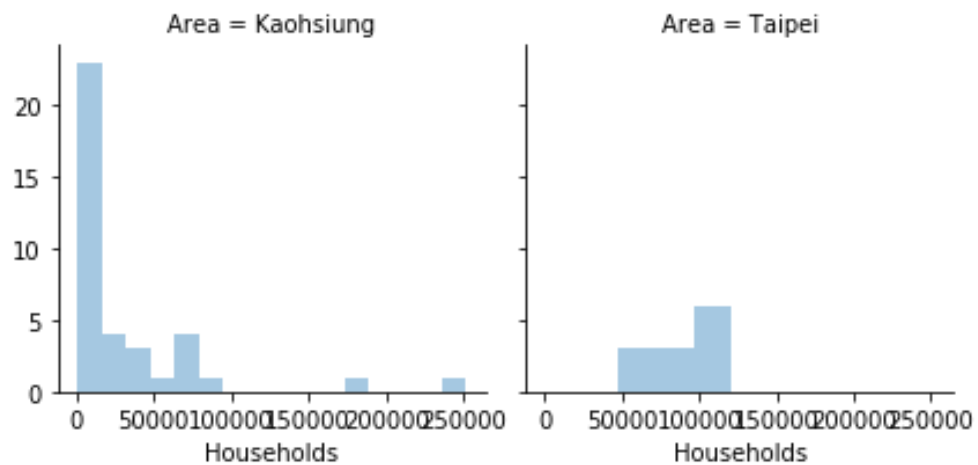
- By Female



- By Male



- By Households



- 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, but 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

| Venue Category | Airport | American Restaurant | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Australian Restaurant | BBQ Joint |
|----------------|---------|---------------------|------------------------|-------------|------------|---------------------|------------------|-----------------------|-----------|
| Area | | | | | | | | | |
| Kaohsiung | 0 | 5 | 2 | 0 | 0 | 4 | 17 | 0 | 0 |
| Taipei | 1 | 6 | 0 | 1 | 2 | 2 | 6 | 1 | 7 |

Detail graph please check ipnb source

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

5. 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.