taipei_restaurant_distribution

October 11, 2020

1 Taipei Restaurant Distribution

We would also consider the competitor or other restaurant in each area. The data that we would use is:

- **Use the travel network data from gov site as sample**. The data contains 308 restaurant which we would use to simulate the real / population data.

From the data, we would try to predict the real / population data, which we will do: - **Simulate based on all number of Restaurant in Taipei**. The detail of the data and statistical formula would is explained after this. - **Aggregate the data per village level**. We would have the number of restaurant per area in the end

1.1 Simulate real / population restaurant data in Taipei

We would build assumption and data based on:

- **Real / population restaurant number.** From the data, we would know how much we should simulate the sample data.
- **Sampling simulation formula.** We would make some probability formula that would fit our use case.

Therefore what we would do would be:

- **Get the sample-to-population ratio.**
- **Set some probability simulation formula**.
- **Simulate real / population data**

1.1.1 Used initial data

The real data contains more information, but for this analysis we heuristically would use restaurant coordinate data

here is the example of the table:

	name	longitude	latitude
0	OGGI 歐奇窯烤披薩 市府店	121.567	25.0423
1	OGGI 歐奇窯烤披薩 敦南店	121.549	25.0285
2	DOORS 門片咖啡	121.529	25.1208
3	亞尼克夢想村陽明山店	121.544	25.1375
4	上古藝術當代畫廊	121.537	25.0411

1.1.2 Setting sample-to-population multiplier

Some fact that we would use are: - World Cities Culture Forum show us in 2017 Taipei number of restaurant per 100,000 population is 307.6 - Goverment data show that Taipei population in 2016 was 2,695,704 people

Therefore we would use this assumption: - Current sample data distribution represent all of the population data (this is a bit heuristic and not really accurate, but for this case we would like to have a ballpark)

Therefore we would use this sample-to-population multiplier formula

$$sample\text{-}to\text{-}population\ multiplier = \left(\frac{Taipei\ population}{100\,000} * 307.6 \right) \Big/ Total\ sample\ data \tag{1}$$

$$\tag{2}$$

The calculation above would **26.92**. So we would make the sample data **27** times larger to simulate the population data.

1.1.3 Make the distribution formula

As our data sample is a bit specific and might not represent the real data, we try to create wider distribution. In this case, we would use radius limit of 5 km (1 point in lat/long coordinate is about 111 km), it would be rarely pass 5 km limit although it is possible. Therefore the distribution formula that we would use is exponential, which is:

$$f(x; \frac{1}{\beta}) = \left(\frac{1}{\beta} \exp(-\frac{x}{\beta}) \Big/ 5 \right) * max\ radius\ level \tag{3}$$

with: - x > 0 - $\beta = 1.5$ - make possibility into two identitcal side (negative / positive)

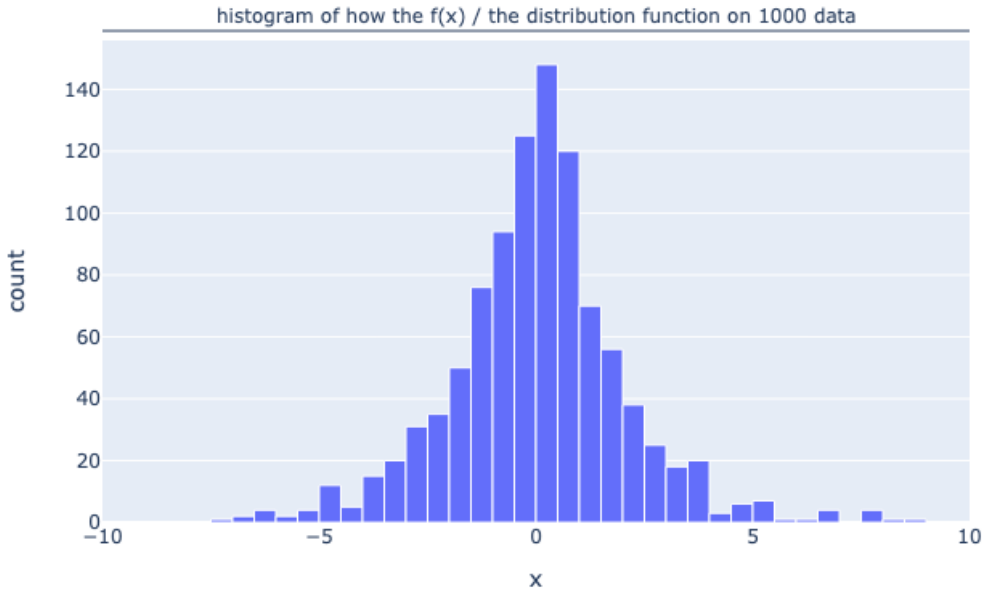
Therefore, the final formula would be

$$new\ data_{longitude} = (sample\ data_{longitude}) + \frac{f(x)}{111} \tag{4}$$

$$new\ data_{latitude} = (sample\ data_{latitude}) + \frac{f(x)}{111} \tag{5}$$

$$\tag{6}$$

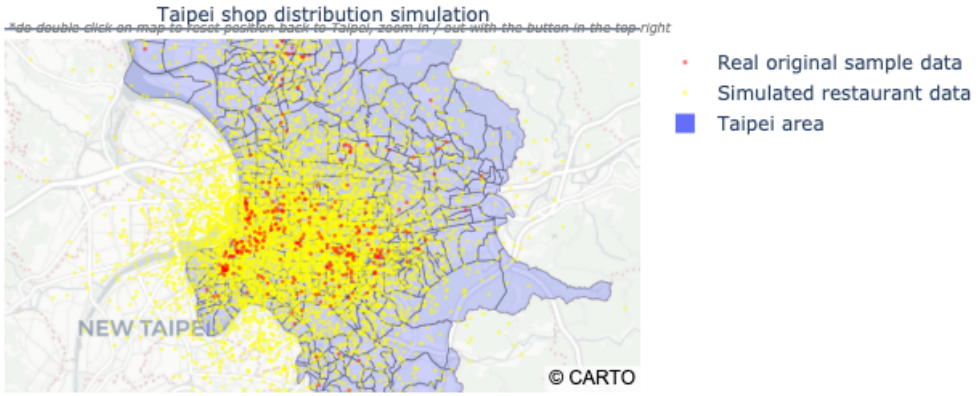
And we would make duplicate the data by 27 times using the formula



1.1.4 Data visualization

With processing the input data with the above formula, we would have the simulation like this.

Most restaurant is on mid-west side of Taipei



1.2 Get per area data

To make the clearer analysis, we would aggregate the restaurant into per village area.

1.2.1 Map data inside the village area

We would fit the coordinate of the restaurant data into the our geo coder. Our geo coder use the mapping coordinate system that is the same with the government data.

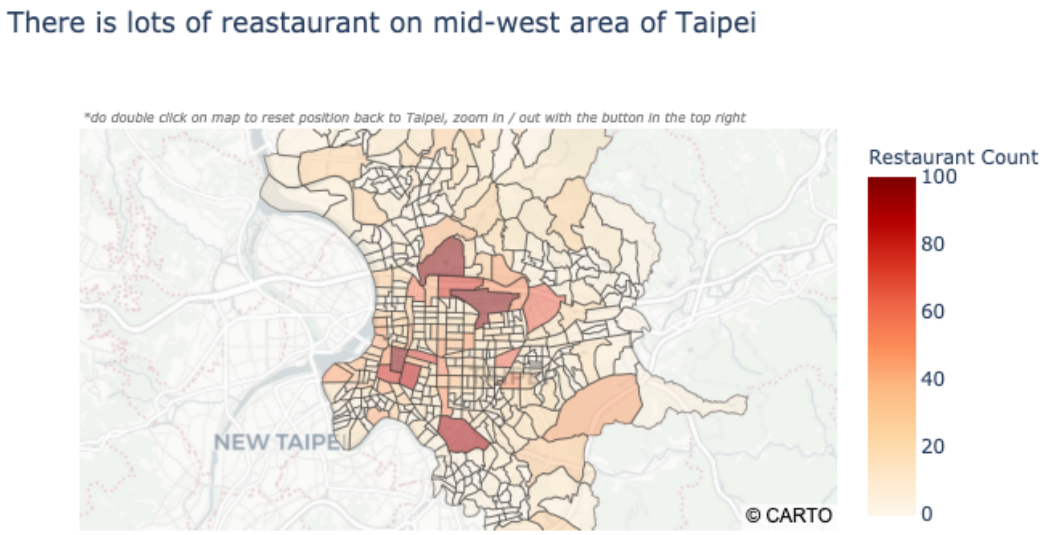
1.2.2 Data table

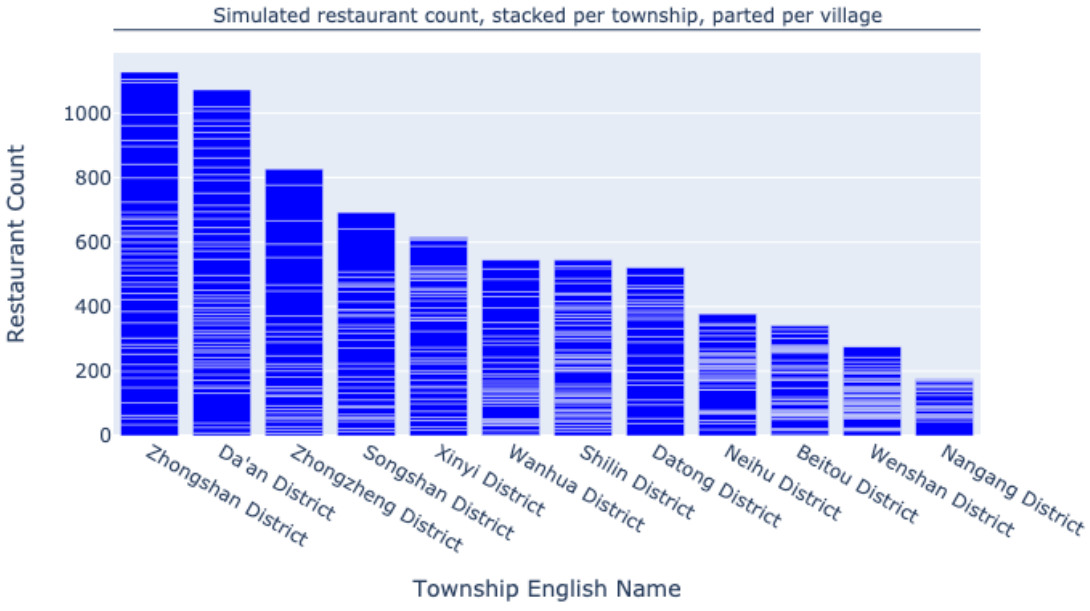
here is the example of the table:

	village_code	village_english_name	township_english_name	restaurant_count
0	63000080031	Zhangxin Vil.	Wenshan District	0
1	63000080037	Laoquan Vil.	Wenshan District	16
2	63000080032	Zhangjiao Vil.	Wenshan District	2
3	63000080041	Zhangwen Vil.	Wenshan District	3
4	63000080043	Zhangshu Vil.	Wenshan District	1

1.2.3 Data Visualization

Here how the data looks like





1.3 Analysis

Analysis about the restaurant in Taipei are:

- **The competition in Taipei mid-west area is pretty rough.** On average there are twice as much as restaurant compare to the other area.