Part 3:

Question: Is some part of city with more chain restaurants ?

1. **Which belong to chain restaurants? Rest\_rain.py**

If we want to know the distribution of restaurants in different citys, the first thing we need to do is pick chain restaurants. I only choose restaurants that have at least 20 restaurants. (McDonald's, White Spot, Tim Hortons, Starbucks, A&W, Subway) I combine them into one Dataframe for future analysis. Of course, those restaurants do not belong to chain restaurant also will be put into a Dataframe. I also remove some data from food categories, such as juice bars. In my view, they cannot be count as restaurants.

1. **Show chain restaurants on map: Rest\_rain.py**

地图

描述已自动生成We want to know and analysis the place of chain restaurants. The histogram is of course a good choice, but it is not intuitive enough. Visualizing these restaurants on the map will help to understand the relationship between the city and the chain restaurants. First, import folium and location of Vancouver to map we need to use. Then I marked different chain restaurants in different color in the map.

Through viewing this figure, we find those chain restaurants distribute in every city. However, the number of chain restaurants in Vancouver is far greater than those in other cities.

1. **Using cluster to show the density of chain restaurants and not chain restaurants in different city: Rest\_rain.py**

地图

描述已自动生成In order to better show the density of chain restaurants in different cities, I used Cluster in the map. It not only visualizes the density better but also displays the number of chain restaurants in different cities.

The orange color on Vancouver which means there has the highest density. Other city has the middle density. (yellow color) The green color shows rare density which means little chain restaurants between different cities. Through this picture, we have seen that there are more chain restaurants in some places. ( Vancouver)

地图

描述已自动生成I also create the cluster figure of not chain restaurants through the same way.

Similar, the Vancouver city also has the highest density. Moreover, the closer to the city center, the greater density of restaurants. From the above two figures, we can see that when a city has more chain restaurants, there are also more non-chain restaurants. Do they have the similar proportion in each city?

1. **Draw Pie graph to show the percentage : ChainVsNotChain.py**

图表, 饼图

描述已自动生成图表, 饼图

描述已自动生成In order to find out the proportion of chain restaurants and non-chain restaurants in different cities, I created pie chart for them respectively. My partner have create the way to show city of each restaurants. Therefore, I create Dataframe of restaurants with column[‘city’] firstly. Then split elements to Dataframe of chain restaurants and Dataframe of Not chain restaurants. Then created the Pie graph based on different cities.

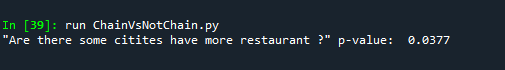
We can see that, Whether it is chain or non-chain, Vancouver has the largest proportion which followed by Not Vancouver area, Surrey, Burnaby, Richmond and Coquitlam. The two graphs can more intuitive to show that there are some city with more chain restaurants. As my team member Xu Bing described, Vancouver is the most prosperous city in the surrounding area. It has more amenities and traffic, so the density of chain restaurants are denser than other cities.

In addition, I also found an interesting phenomenon. Although there are more data for non-chain restaurants than chain restaurants, their proportions in different cities seem to be very close.

1. **The P-value of chi2\_test for Chain restaurants and not Chain restaurants in different cities. : ChainVsNotChain.py**

Moreover, I also used the p-value of chi2 test to show quantity of Chain restaurants and not Chain restaurants is different.



Take the number of Chain restaurants and not Chain restaurants in six city into chi2 test.

The p-value is 0.037 which is smaller than 0.5. It further means amount of restaurants are different in the six cities.

1. **Limitations**

* The limitation of data:

In my experiment, I discarded those restaurants with fewer than 20 , and chose some well-known chain restaurants（such as AW and Starbucks). Join those chain restaurants with few numbers, and there may be more discoveries.

* The relation between Chain restaurants and Not chain.

Just as I found through viewing Pie graph. The percentage of chain restaurants and not chain in six place seems to be similar. But I did not know how to use a suitable test to get the p-value to show their relationship.

1. **Accomplish Statement of Xiaohang Hu:**

* I learned how to visualize data point on the map in different color and cluster them.
* Learned how to create a Pie Graph.
* I found the a amazing thing the percentage of chain restaurants and not chain in six place seems to be similar. This will inspire me to explore further.