Analysis of Michelin restaurants.

Quang Tran Ngoc.

Feburary, 4, 2021.

1. Introduction:

Background:

Michelin restaurant is one of the most famous restaurants on the world. These restaurants is famous with food as well as luxurious. Currently, there are many fake Michelin restaurants with low quality. Therefore, we need to analysis data to group and classify these restaurant on each region. In addition, we can know which the retaurant is the best in the Michelin restaurants.

Problem statement:

Our data is related to Michelin restaurants that provide region, zip code, name of restaurant, latitude, and longitude. The project aims to group the restaurants of Michelin system.

2. Data acquisition:

Data sources:

The data of Michelin restaurant is from Kaggle. In the dataset, there are 3 kinds of Michelin restaurants, but we only choose 1st star restaurants.

Data cleaning:

There are many information in the dataset. Firstly, I download the data and put it into one table. There are one problem for downloading and loading dataset, I have to load dataset from a github link.

name	year	latitude	longitude	city	region	zipCode	cuisine	price	url
Kilian Stuba	2019	47.348580	10.17114	Kleinwalsertal	Austria	87568	Creative	$	https://guide.michelin.com/at/en/vorarlberg/kl...
Pfefferschiff	2019	47.837870	13.07917	Hallwang	Austria	5300	Classic cuisine	$	https://guide.michelin.com/at/en/salzburg-regi...
Esszimmer	2019	47.806850	13.03409	Salzburg	Austria	5020	Creative	$	https://guide.michelin.com/at/en/salzburg-regi...
Carpe Diem	2019	47.800010	13.04006	Salzburg	Austria	5020	Market cuisine	$	https://guide.michelin.com/at/en/salzburg-regi...
Edvard	2019	48.216503	16.36852	Wien	Austria	1010	Modern cuisine		https://guide.michelin.com/at/en/vienna/wien/r...

Figure 1: The data table of Michelin restaurants.

Note: The data table is a simple table and not full.

Feature selection:

To solve the problem, we need the name of restaurant, region, zip code,lattitude, and longitude in the Michelin dataset.

	name	zipCode	region	latitude	longitude
0	Kilian Stuba	87568	Austria	47.348580	10.17114
1	Pfefferschiff	5300	Austria	47.837870	13.07917
2	Esszimmer	5020	Austria	47.806850	13.03409
3	Carpe Diem	5020	Austria	47.800010	13.04006
4	Edvard	1010	Austria	48.216503	16.36852

Figure 2: The data table is used for the problem.

Note: The data table is a simple table and not full.

3. Exploratory data analysis:

Choosing target variable:

To solve the problem, I base on region data which include 5 labels. Due to the relationship between region and the name of restaurant, the region is used to classify the Mechelin restaurants.

	region	108	21212	28+	360°	8 1/2 Otto e Mezzo - Bombana	A. Wong	Adam's	Addison	Agern	...	Zero Complex	Zhejiang Heen	Zi Yat Heen	aend	atelier Amaro	formel B	jū-ni	logy	wilks	Épure
0	Austria	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.083333	0.0	0.0	0.000000	0.0	0.0	0.0
1	California	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.014493	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.014493	0.0	0.0	0.0
2	Chicago	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0
3	Croatia	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0
4	Czech Republic	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0

5 rows × 545 columns

Figure 3: the group table of data table about Michelin restaurant.

4. Predictive Modeling:

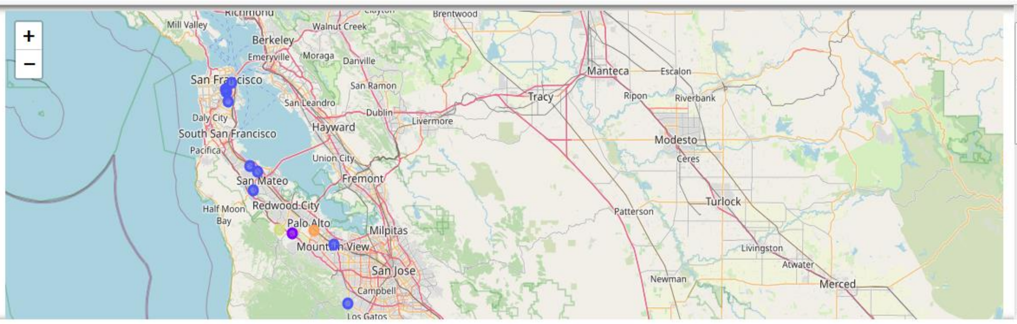
Solution

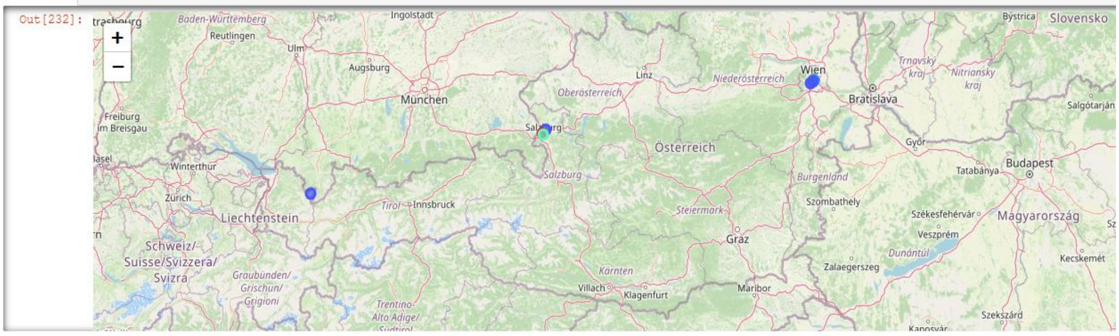
I applied K-means to solve the problem.

```
In [236]: from sklearn.cluster import KMeans
          # set number of clusters
          kclusters = 5

          ml_tab_grouped_clustering = ml_tab_grouped.drop('region', 1)

          # run k-means clustering
          kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(ml_tab_grouped_clustering)

          # check cluster labels generated for each row in the dataframe
          kmeans.labels_

Out[236]: array([0, 0, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 4, 3, 0, 0, 0, 0, 0, 0,
                 0, 0], dtype=int32)
```

Visualization for the problem

Out[232]:





Examination

To check the evaluation, we choose random lable from 0 to 4, we can visualize the name of restaurants in the region.

	zipCode	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	87568	2	Das Loft	Kilian Stuba	Edvard	SHIKI	Le Ciel by Toni Mörwald	Pfefferschiff	Walter Bauer	Pramerl & the Wolf	Tian	Esszimmer
1	5300	2	Das Loft	Kilian Stuba	Edvard	SHIKI	Le Ciel by Toni Mörwald	Pfefferschiff	Walter Bauer	Pramerl & the Wolf	Tian	Esszimmer
2	5020	2	Das Loft	Kilian Stuba	Edvard	SHIKI	Le Ciel by Toni Mörwald	Pfefferschiff	Walter Bauer	Pramerl & the Wolf	Tian	Esszimmer
3	5020	2	Das Loft	Kilian Stuba	Edvard	SHIKI	Le Ciel by Toni Mörwald	Pfefferschiff	Walter Bauer	Pramerl & the Wolf	Tian	Esszimmer
4	1010	2	Das Loft	Kilian Stuba	Edvard	SHIKI	Le Ciel by Toni Mörwald	Pfefferschiff	Walter Bauer	Pramerl & the Wolf	Tian	Esszimmer
...	...	...	...	...	...	...	...	...	...	...	...	...
544	CT5 4BP	2	Lyle's	Quilon	Roganic	Rogan & Co	River Café	Ritz Restaurant	Restaurant Tristan	Restaurant Hywel Jones by Lucknam Park	Red Lion Freehouse	Eipic
545	TN27 8AH	2	Lyle's	Quilon	Roganic	Rogan & Co	River Café	Ritz Restaurant	Restaurant Tristan	Restaurant Hywel Jones by Lucknam Park	Red Lion Freehouse	Eipic
546	CT2 0DB	2	Lyle's	Quilon	Roganic	Rogan & Co	River Café	Ritz Restaurant	Restaurant Tristan	Restaurant Hywel Jones by Lucknam Park	Red Lion Freehouse	Eipic
547	JE2 4TQ	2	Lyle's	Quilon	Roganic	Rogan & Co	River Café	Ritz Restaurant	Restaurant Tristan	Restaurant Hywel Jones by Lucknam Park	Red Lion Freehouse	Eipic
548	JE2 4UH	2	Lyle's	Quilon	Roganic	Rogan & Co	River Café	Ritz Restaurant	Restaurant Tristan	Restaurant Hywel Jones by Lucknam Park	Red Lion Freehouse	Eipic