

Find Which restaurant type should be open in Columbia, MO

Introduction: Business Problem

Columbia, Missouri is the city I spend most of 5 years to complete degree. Columbia is university city where there are not many types of restaurants or not many types of entertainment area. For Applied Data science project, If I were entrepreneur, I would like to investigate if there any shop should I plan or recommend opening that could help to city. My initial idea finds similar district nearby as reference point and compare to Columbia and suggest what kind of restaurant should it be open.

Data

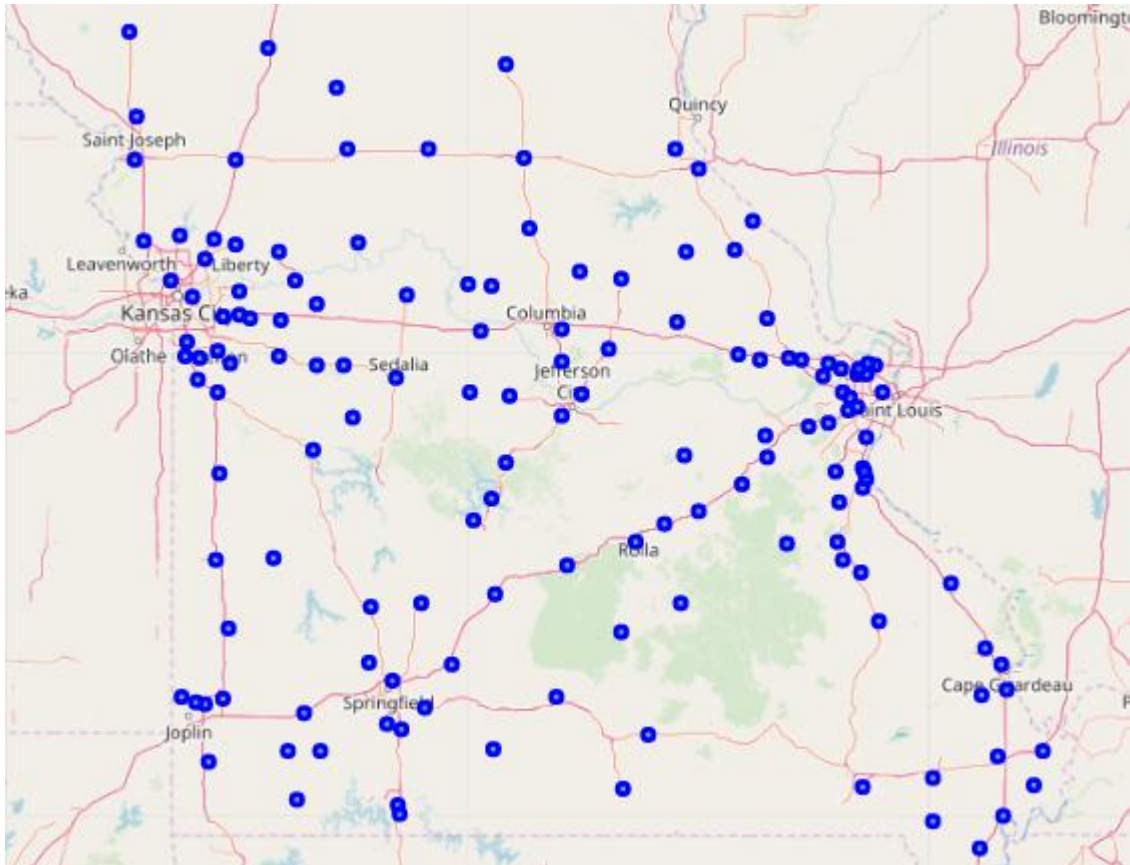
The mainly sources would be data gathering for foursquare API, Missouri State consensus data, and Wikipedia of City list in Missouri. Foursquare API mainly source is to attract restaurant data. While other data is to extract as other feature data for clustering process. This is an initial data sources that would be use. These are links that project was used for based line.

- <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/>
- https://en.wikipedia.org/wiki/List_of_municipalities_in_Missouri

Note Upon Discovery on consensus data. The detail of Missouri, City Demographic are available for small number. Thus data will limited

In Summary there are three main sources of data this project will be used as following:

- [City names in Missouri State along with population Consensus](#)
- [City list along with latitude and longitude](#)
- [Four Square API](#)



Using Folium plot mapping of city in Missouri State

Methodology

For the methodology approach, is to clustering city into similarity group to find City similarity pattern on **restaurant type**

For Data Source We combine between List of Cities and City with list of Longitude and Latitude. As the base line for searching in Foursquare API. There are Cleaning and limited data as following :

List of cities

- Cleaning text name such as trimming and converting abbreviation for matching another source
- Data extract only Type of "City" (filter out village or town type)
- Screen only top 200 city that have most population for seeking pattern

List city that have Longitude and Latitude

- There multiple ZIP Code that stay in city so grouping same city and find average value as a centroid of that city
- Using City name to merge with previous data source

```
missouri_data.head(5)
```

	index	City	Type	2010 U.S. Census[1]	County	Latitude	Longitude
0	485	Kansas City	City	459787	Jackson	39.095394	-94.497847
1	876	Saint Louis	City	319294	Independent City	38.643335	-90.289811
2	864	Springfield	City	159498	Greene	37.253954	-93.269458
3	462	Independence	City	116830	Jackson	38.812911	-93.964082
4	205	Columbia	City	108500	Boone	38.940322	-92.236863

Sample of City Raw data include Latitude, Longitude, and Consensus

After that, we use info (centroid of each city) as a data search for Foursquare API.

Since the free-tier account Foursquare data, the project limited criteria as following:

- Clustering City is involved only in Missouri State with top 200 most populated city
- Restaurants need to be fetched within 5km from central of each city
- Also for clustering model will be use K-mean model using restaurant type ratio in each city
 - For determine K cluster, the project use Elbow method to find sensible K for clustering process

After Clustering result, we will be focus clustering group model that have Columbia city lied on to find group behavior of restaurant type. Then compare with Columbia data to find the restaurant type that Columbia lack as a recommended restaurant type that should be open in Columbia.

Results and Discussions

For City in Missouri State top 168 populated City, There are 5681 restaurants with 75 unique category Venues. The common venue are Fast Food restaurants, American Restaurants, and pizza place.

Venue Category	
Fast Food Restaurant	812
Pizza Place	763
American Restaurant	541
Sandwich Place	502
Mexican Restaurant	469
Chinese Restaurant	230
Italian Restaurant	224
BBQ Joint	189
Breakfast Spot	183
Steakhouse	139
Bakery	138
Burger Joint	130
Restaurant	129
Fried Chicken Joint	105

Above is the summary of data

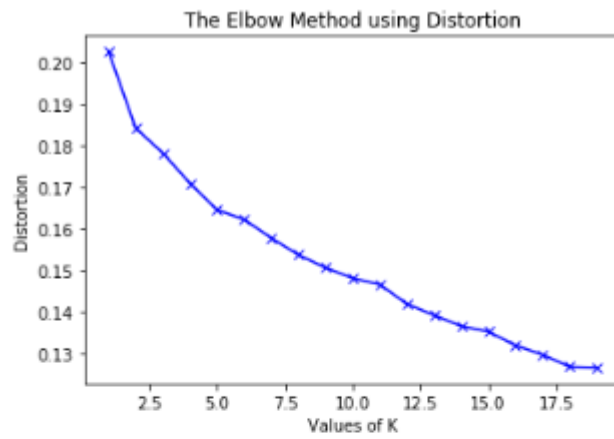
With data scarping from four squared API we are limited search due to this criteria

- Searching is limited to any food related store
- The search are based on centroid point from ZIP code that have same city name
- radius for searching grid within 10 kilometers from centroid
- in each city, the limited results would be around 400 max search data per city

Here are some samples

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kansas City	39.095394	-94.497847	Vivilore	39.087602	-94.449800	American Restaurant
1	Kansas City	39.095394	-94.497847	Steam Bowl	39.105027	-94.522495	Chinese Restaurant
2	Kansas City	39.095394	-94.497847	Elvira's Pasteleria	39.106096	-94.535963	Mexican Restaurant
3	Kansas City	39.095394	-94.497847	Happy Gillis Cafe & Hangout	39.108916	-94.570382	Sandwich Place
4	Kansas City	39.095394	-94.497847	Garozzo's Ristorante	39.109492	-94.571955	Italian Restaurant
5	Kansas City	39.095394	-94.497847	Gates Bar-B-Q	39.098435	-94.555813	BBQ Joint
6	Kansas City	39.095394	-94.497847	Cafe Verona	39.092208	-94.417490	Italian Restaurant
7	Kansas City	39.095394	-94.497847	A Little BBQ Joint	39.101690	-94.428947	BBQ Joint
8	Kansas City	39.095394	-94.497847	Vietnam Cafe	39.109306	-94.573173	Vietnamese Restaurant
9	Kansas City	39.095394	-94.497847	Dixon's Chili	39.059371	-94.471833	American Restaurant

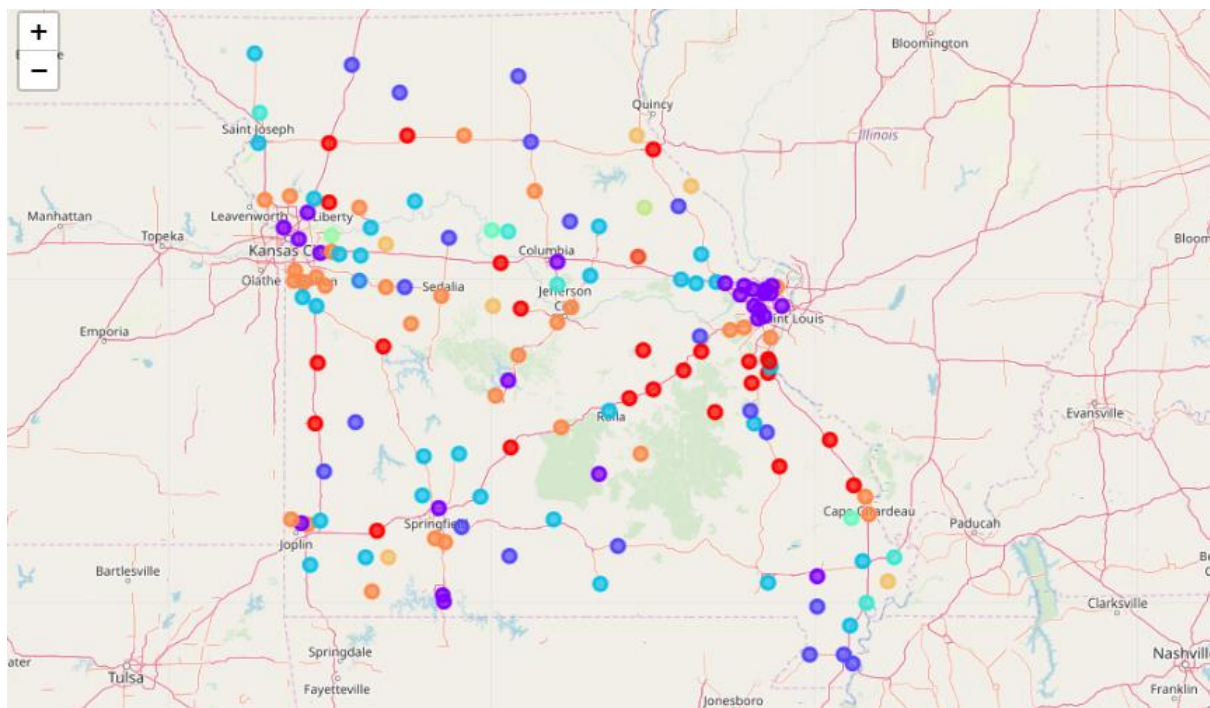
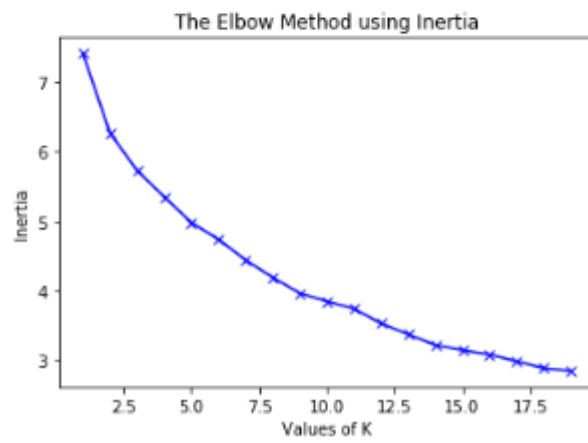
The result of analysis shows that for K-mean model. By using Elbow method to determine suitable number K for clustering algorithm, the suitable number is 12. When apply the k-mean we found out there are 25 city they have similar trait like Columbia city.



```

plt.plot(K, inertias, 'bx-')
plt.xlabel('Values of K')
plt.ylabel('Inertia')
plt.title('The Elbow Method using Inertia')
plt.show()

```



Clustering plot

City	Type	2010 U.S. Census[1]	County	Latitude	Longitude	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
Kansas City	City	459787	Jackson	39.095394	-94.497847	1	American Restaurant	BBQ Joint	Mexican Restaurant	Italian Restaurant	Burger Joint	Sandwich Place	New American Restaurant
Saint Louis	City	319294	Independent City	38.643335	-90.289811	1	Italian Restaurant	American Restaurant	Pizza Place	Sandwich Place	Café	Taco Place	Seafood Restaurant
Springfield	City	159498	Greene	37.253954	-93.269458	1	Fast Food Restaurant	Sandwich Place	Mexican Restaurant	Pizza Place	American Restaurant	Breakfast Spot	Chinese Restaurant
Columbia	City	108500	Boone	38.940322	-92.236863	1	Pizza Place	American Restaurant	Sandwich Place	Mexican Restaurant	Fast Food Restaurant	BBQ Joint	Burger Joint
Saint Charles	City	65794	St. Charles	38.720464	-90.643052	1	Fast Food Restaurant	American Restaurant	Sandwich Place	Mexican Restaurant	Italian Restaurant	Pizza Place	Burger Joint
Blue Springs	City	52575	Jackson	38.999563	-94.308882	1	Pizza Place	Mexican Restaurant	American Restaurant	Sandwich Place	Breakfast Spot	Fast Food Restaurant	Chinese Restaurant
Saint Peters	City	52575	St. Charles	38.778579	-90.620450	1	American Restaurant	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	Breakfast Spot	Italian Restaurant
Chesterfield	City	47484	St. Louis	38.644107	-90.532616	1	Mexican Restaurant	Sandwich Place	Italian Restaurant	American Restaurant	Thai Restaurant	Breakfast Spot	Pizza Place
Ballwin	City	30404	St. Louis	38.613573	-90.484374	1	Italian Restaurant	American Restaurant	Pizza Place	Sandwich Place	Steakhouse	Mexican Restaurant	Chinese Restaurant
Liberty	City	29149	Clay	39.270639	-94.411875	1	Pizza Place	Breakfast Spot	Sandwich Place	Mexican Restaurant	American Restaurant	Burger Joint	Chinese Restaurant
Maryland Heights	City	27472	St. Louis	38.725331	-90.444510	1	Sandwich Place	Bakery	Mexican Restaurant	Italian Restaurant	Pizza Place	Breakfast Spot	Korean Restaurant

Data Sample the clustering that have Columbia city

We also discover that the common restaurant are usually open in Columbia Pizza Place, American Restaurant and Sandwich Place. Since there are many competitors are in Columbia, These type of restaurant are should be avoid.

As we compare between normalized of Columbia group cluster data, there are some recommend that type restaurant shop open such as Brazilian, Caribbean, Asian or French restaurants. Since the data indicated that Columbia these type of restaurant are lacking

However, upon comparing the result there are multiple topics in order to improve data quality and improve model as approach as following

- 1. Foursquare API - instead using latitude and longitude as the criteria searching, there are might need to searching via searching area instead. Also there are limited search, these could be increase maximum search and increase radius of city
- 2. Modeling - the data for modeling use only restaurant type ratio as a feature for clustering model. For improvement, it could include data such as city demographic, restaurant reviews, or include other type of building beside restaurant to find additional factors

Conclusion

Purpose of this project was to find which type of restaurant that likely could open the new one. By grouping pattern restaurant type compare with nearby city in the state. it found out there are around 18 city that similar to Columbia. When upon find detail of the group there are some interesting shop that Columbia didn't open but there are available to other similar city such as Brazilian Caribbean shop. Thus, this would are guideline to help exploration by stakeholders to found new venture.

However, this project could be improve since the main factor analysis is mainly on the restaurant type. it could expand in numerous way such as finding entertain venue pattern as well or expand search radius to nearby State to find interesting pattern. Also, there could be more topics to find data for analysis