Classifying a Good BBQ Restaurant

Christopher Kang

The Overall Problem

- Restaurant industries are competitive and newly opened restaurants are less likely to succeed.
- Typically a good restaurant will only succeed if it has good service, food, and price.
- Is there a way to prioritize on what a restaurant needs in order to be successful?

The Project Goal

- This analysis will attempt to address on how valid the prior statements claims are.
- Investors and small restaurant owners will have a better idea in looking for specific data/trend
- Knowing this data/trend can then help them know if the restaurant is succeeding or not

Data Analysis Breakdown

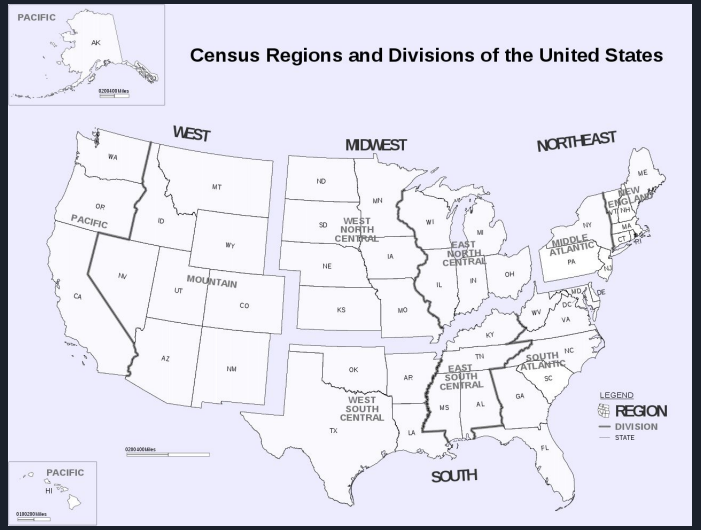
- Data Obtaining
- Data Cleaning
- Data Modeling
- Results
- Conclusion

Data Obtaining

- Restaurants data was obtained from the FourSquare website and API calls.
- Menu items, price of each item, rating, number of rating, etc were collected

Data Obtaining

- To factor location as a possible feature, 3 states were chosen from each division.
- There are a total of 9 divisions in the USA

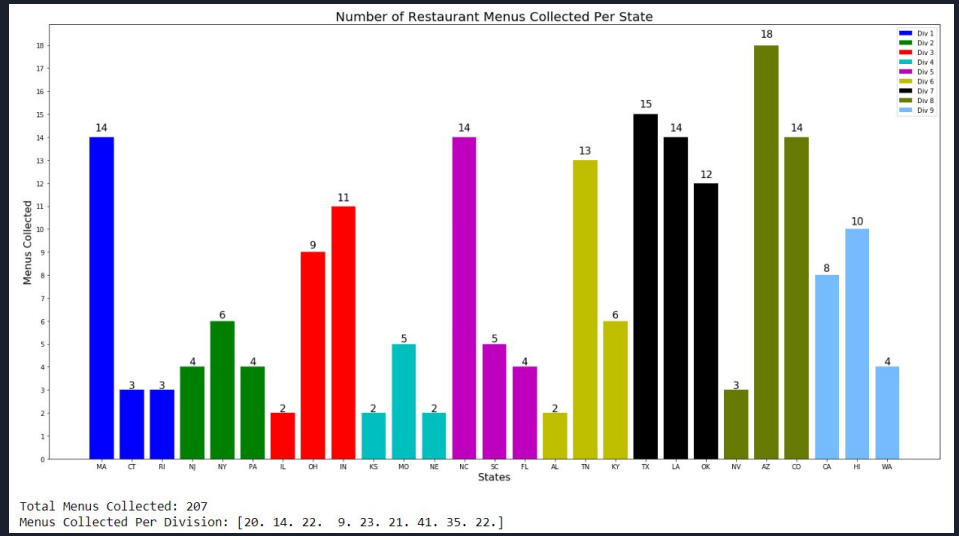


Data Obtaining

- Locations of restaurants were determined by first finding out major BBQ states.
- Afterwards, states were picked based on their population size as there would likely be more data for states with a higher population count.

Data Cleaning

- There were a number of restaurants that had missing attributes in them.

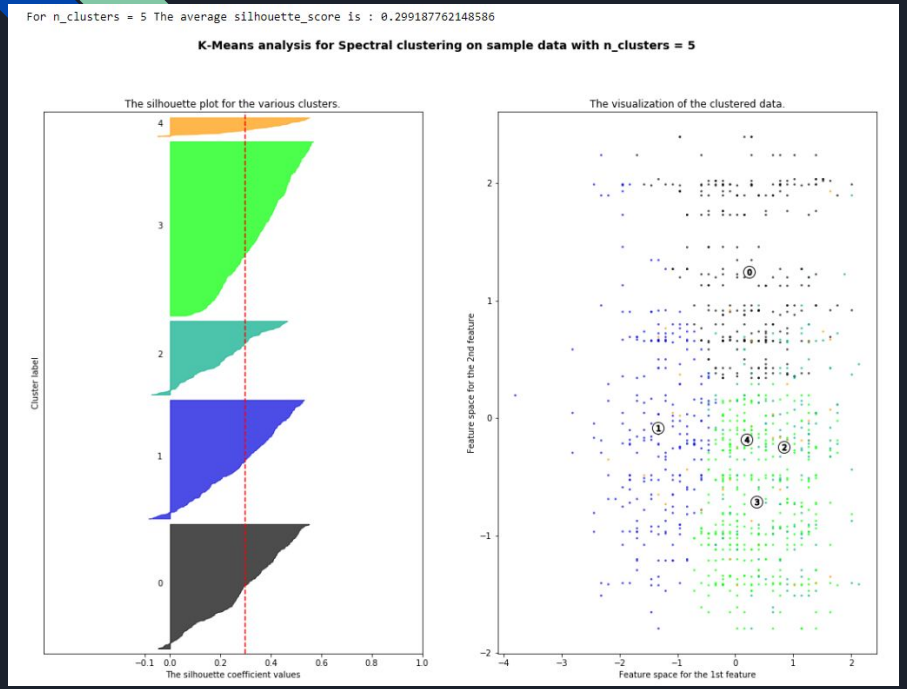


Data Cleaning

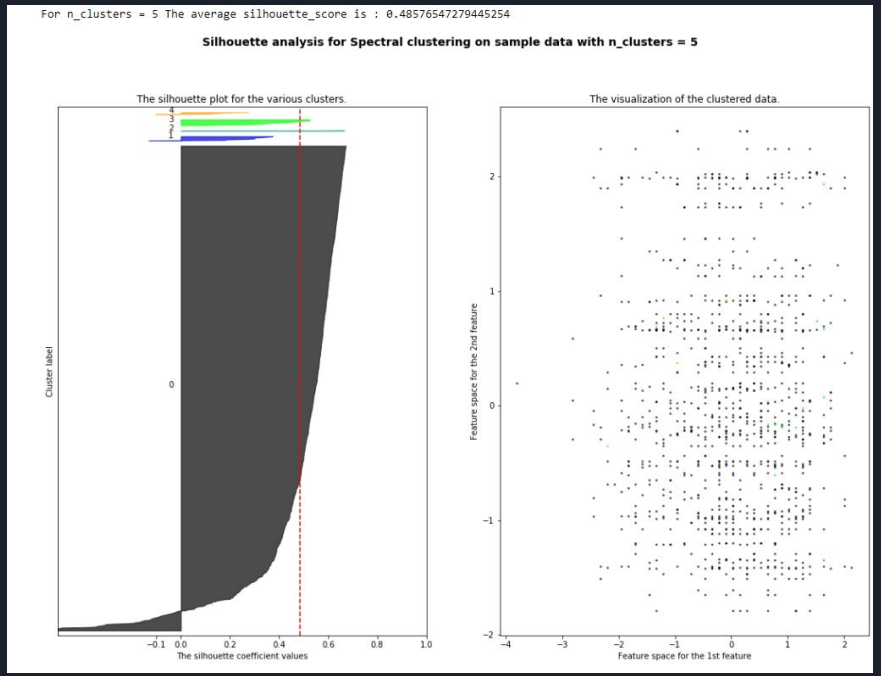
- Missing values were filled in by picking a random value from the list of data that was collected from that column.
- Missing price tier values were assigned to “Moderate” pricing due to many of the restaurants falling under that category.

Data Modeling

- Due to a significant number of unknown values, the analysis had to change from a supervised predicting model to a unsupervised clustering model.
- Main clustering models used were K-Means, Spectral, and DBSCAN.
- Main features that were kept were number of ratings, number of ratings, and average item length.

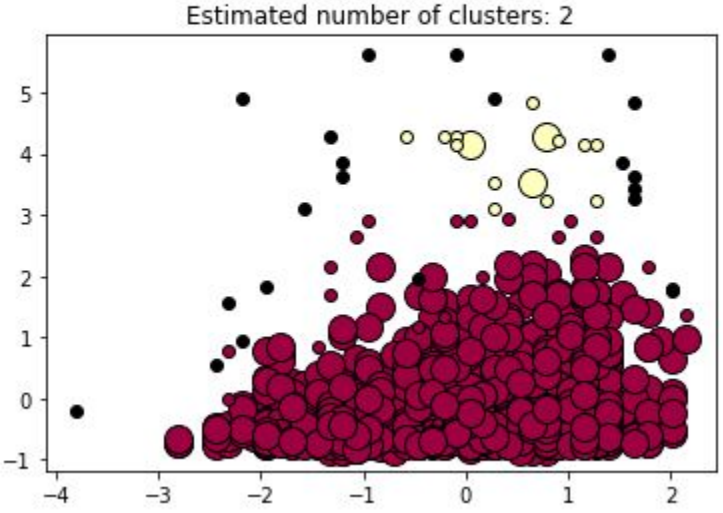


K-Means Result



Spectral Clustering Result

```
Estimated number of clusters: 2
Estimated number of noise points: 21
Silhouette Coefficient: 0.494
```



DBSCAN Result

Results

- K-Means results showed that each of the cluster were either overlapping or close with another.
- The silhouette score shows how well each point in the cluster are similar to each other.
- In this case, K-Means score showed that these clusters weren’t sufficient.

Results

- Spectral clustering ended up having one cluster take in a majority of points.
- This left no other points for the other cluster.
- Thus even though the silhouette score was high, this can’t be considered as a good cluster.

Results

- DBSCAN was able to create two cluster based on the given hyperparameters and features.
- With this model, there seems to be a good separation between the two cluster and silhouette score.
- Hence this model will be used for the final analysis.

Conclusion

- Common pattern that can be seen under this cluster are:
  - Good above average rating
  - Good number of ratings
  - Moderate Price Tier

	ID	Name	Latitude	Longitude	State	City	Category Name	Rating	Number of Rating	Price Tier	Average Menu Length	Average Item Price	Division
119	4cd573d7122ba143c4192ca1	Dinosaur Bar-B-Que	42.734600	-73.689244	NY	Troy	BBQ Joint	8.4	385.0	1.0	25.0	7.318182	2
127	5b8ad0a631fd14002c758e7c	Unihog	42.898373	-73.352850	NY	Hoosick Falls	BBQ Joint	7.6	388.0	1.0	25.0	6.476000	2
158	51b2306b498e43d578d7bd29	Bullbq's Burger Cafe	40.269864	-76.888302	PA	Harrisburg	BBQ Joint	8.3	311.0	1.0	10.0	6.272800	2
169	503923eee4b06348eb4d6260	Little Everett's BBQ	39.808870	-77.001290	PA	Hanover	BBQ Joint	7.7	379.0	1.0	10.0	7.168000	2
259	520e9ff911d2b12d923e69e1	Tamarack chicken and fish	40.076410	-82.966952	OH	Columbus	BBQ Joint	7.5	388.0	1.0	18.0	6.272800	3
271	4b64f537f964a52045dc2ae3	Weber Grill Restaurant	39.767612	-86.159902	IN	Indianapolis	American Restaurant	7.9	333.0	1.0	29.0	7.600000	3
442	5291360e11d20de90e42c30e	The Pit	36.003458	-78.899893	NC	Durham	BBQ Joint	7.9	301.0	1.0	46.0	5.610580	5
596	4c2b7217b34ad13ab4c3e9ce	Five Brothers BBQ	32.343436	-86.222641	AL	Montgomery	BBQ Joint	7.6	379.0	1.0	54.0	8.137143	6
612	515484d6e4b0bdae55d49c6d	Edley's East	36.175910	-86.756398	TN	Nashville	BBQ Joint	8.7	311.0	0.0	40.0	7.303571	6
667	4b747dfef964a520dddf2de3	Attus Apparel	38.000664	-84.524555	KY	Lexington	Clothing Store	7.2	388.0	1.0	5.0	5.843750	6
678	5676037b498e1b7d98c935c7	Cooper's Old Time Pit Bar-B-Que	30.264966	-97.743747	TX	Austin	BBQ Joint	8.6	379.0	1.0	10.0	7.303571	7
688	4a69f8d0f964a5204ecc1fe3	County Line on the Lake	30.357139	-97.785685	TX	Austin	BBQ Joint	8.3	388.0	1.0	25.0	7.360000	7
833	58fe9edf340a5840405cc618	Naked BBQ	33.579469	-111.887248	AZ	Scottsdale	BBQ Joint	8.2	430.0	1.0	49.0	7.855000	8
885	5dd705446292a10008685d70	AJ's Pit Bar-B-Q	39.677094	-104.992101	CO	Denver	BBQ Joint	8.7	379.0	1.0	48.0	7.716400	8
1027	4b36a724f964a520fe3925e3	The Butcher Boys Beef Outlet	47.119359	-122.293721	WA	Puyallup	Steakhouse	8.2	333.0	3.0	65.0	7.360000	9

Restaurants under Yellow Cluster

Conclusion

- With the results in mind, we can try making the following claim in that the number of ratings and price tier are the key important factors to consider.
- This could imply that people tend to leave a rating for a restaurant if they consistently have good service and food overall.
- The moderate price tier would also allow people to consistently visit restaurants rather than once.

Future Work

- Not enough data was collected for this analysis and thus more restaurants data would be needed.
- In order to avoid having many restaurants with missing values, non BBQ joint restaurants must be considered.
    - This might end up introducing different restaurant industry distribution however.