ANDERSON BALRAJ IBM DATA SCIENCE CAPSTONE: BATTLE OF THE NEIGHBORHOODS WEEK 2

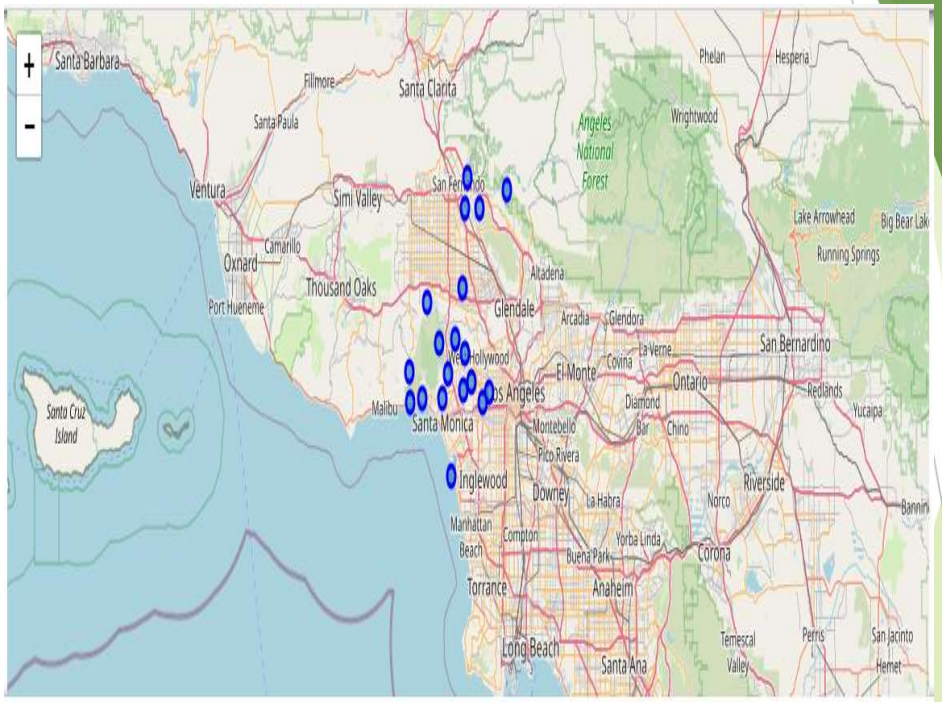
NEIGHBORHOOD RESTAURANT CUISINE ANALYSIS OF LOS ANGELES

Introduction

- Los Angeles (LA) is the largest city in California. LA has a population of approximately four (4) million people and ranks third in having the largest population in North America.
- Business investment in any venture poses high risk, utilizing data science it is possible to mitigate risk with an informed decision.
- The following are the queries that the report will answer to determine the best location and cuisine of a new restaurant in Los Angeles:
  - Determining neighborhood clusters according to the most popular type of restaurants cuisine e.g. Italian, Chinese.
  - Determining current rating of the different types of restaurant cuisine
  - Analyze the tips for a restaurant cuisine to determine popularity
  - Determining current trending types of restaurants cuisine

Obtaining, Validating and Cleaning the Data

- Data was obtained from the "SOCR Data - Los Angeles City Neighborhoods Data", this data was used to obtain the neighborhoods with the highest median income, the was validated on the Los Angeles Map
- These neighborhoods were then used to query the Foursquare API to obtain venue data
- The one hot encoding technique was used to reassign the Venue Categories as column headers and obtain the frequencies of visit for each Neighborhood
- The venue data was then filtered to obtain the restaurant data



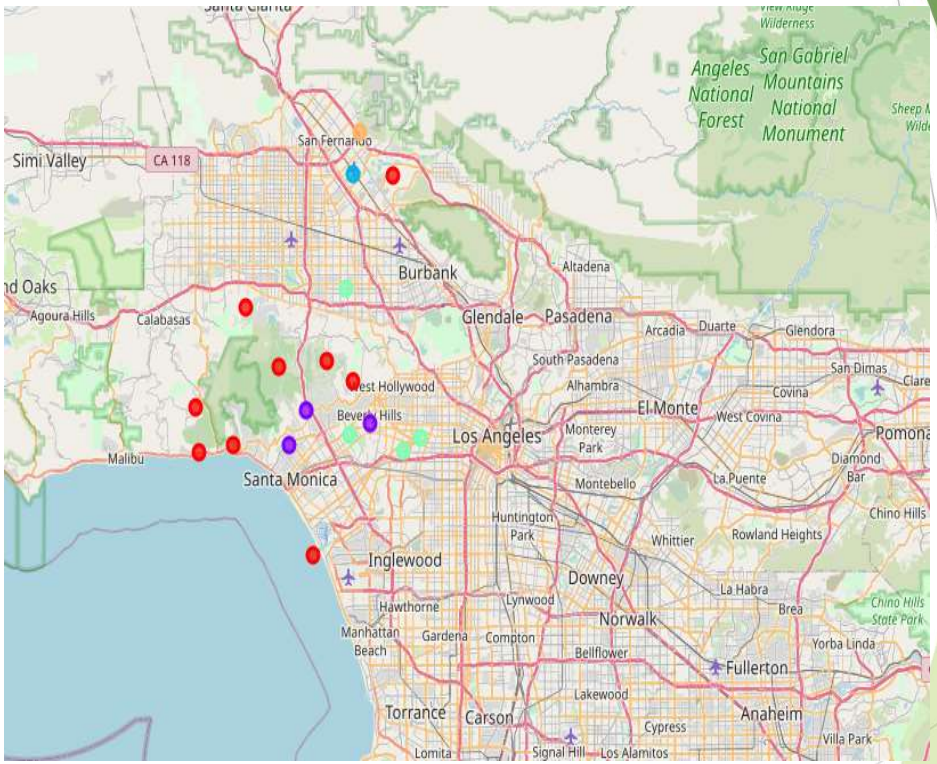
Data Analysis of Cleaned Dataframe

- Exploratory analysis of the clean dataframe was done to get an output of the top 5 restaurants for each neighborhood and the associated frequencies
- A new dataframe was completed that shows the top ten most common venues for each neighborhood, a snippet is shown below.

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	Bel-Air	Vegetarian / Vegan Restaurant	Japanese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Caribbean Restaurant	Chinese Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant
1	Beverly_Crest	Vegetarian / Vegan Restaurant	Japanese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Caribbean Restaurant	Chinese Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant
2	Beverlywood	American Restaurant	Fast Food Restaurant	Cajun / Creole Restaurant	Caribbean Restaurant	Restaurant	Mexican Restaurant	Latin American Restaurant	Italian Restaurant	Asian Restaurant	Chinese Restaurant
3	Brentwood	Mexican Restaurant	Chinese Restaurant	Latin American Restaurant	Vegetarian / Vegan Restaurant	Caribbean Restaurant	Kosher Restaurant	Korean Restaurant	American Restaurant	New American Restaurant	Mediterranean Restaurant
4	Century_City	Mexican Restaurant	Italian Restaurant	Chinese Restaurant	Korean Restaurant	Asian Restaurant	Fast Food Restaurant	Japanese Restaurant	American Restaurant	Restaurant	Southern / Soul Food Restaurant

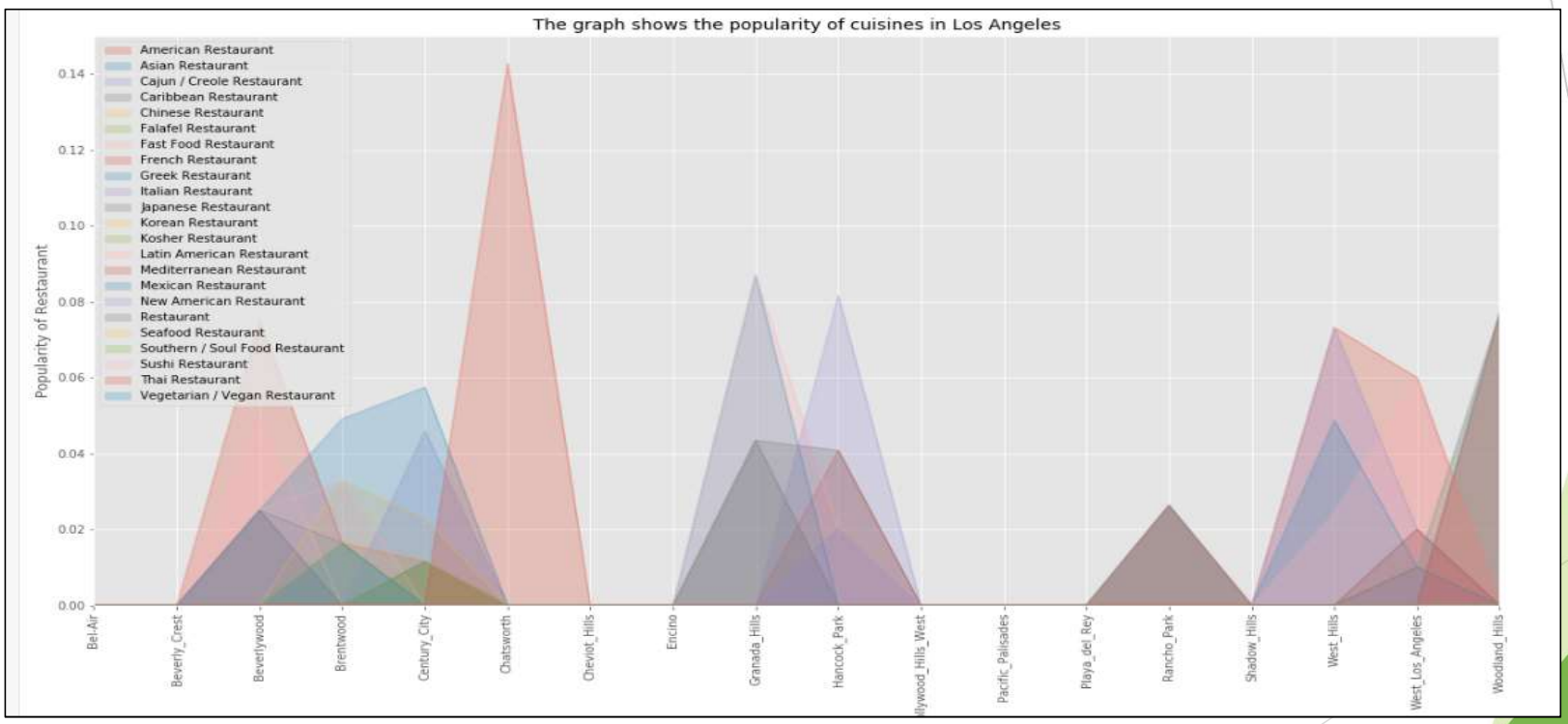
K-Means Clustering Analysis

- K-Means Clustering Analysis is used to group similar neighborhoods based on the most popular cuisines
- Five(5) clusters were used to classify the neighborhoods and then it was plotted on a map of Los Angeles
- The major groups were Cluster 1 and Cluster 5
- Cluster 1 neighborhoods preferred Vegetarian and Japanese cuisines and were located around the same area
- Cluster 5 neighborhoods were extremely popular in American Cusines



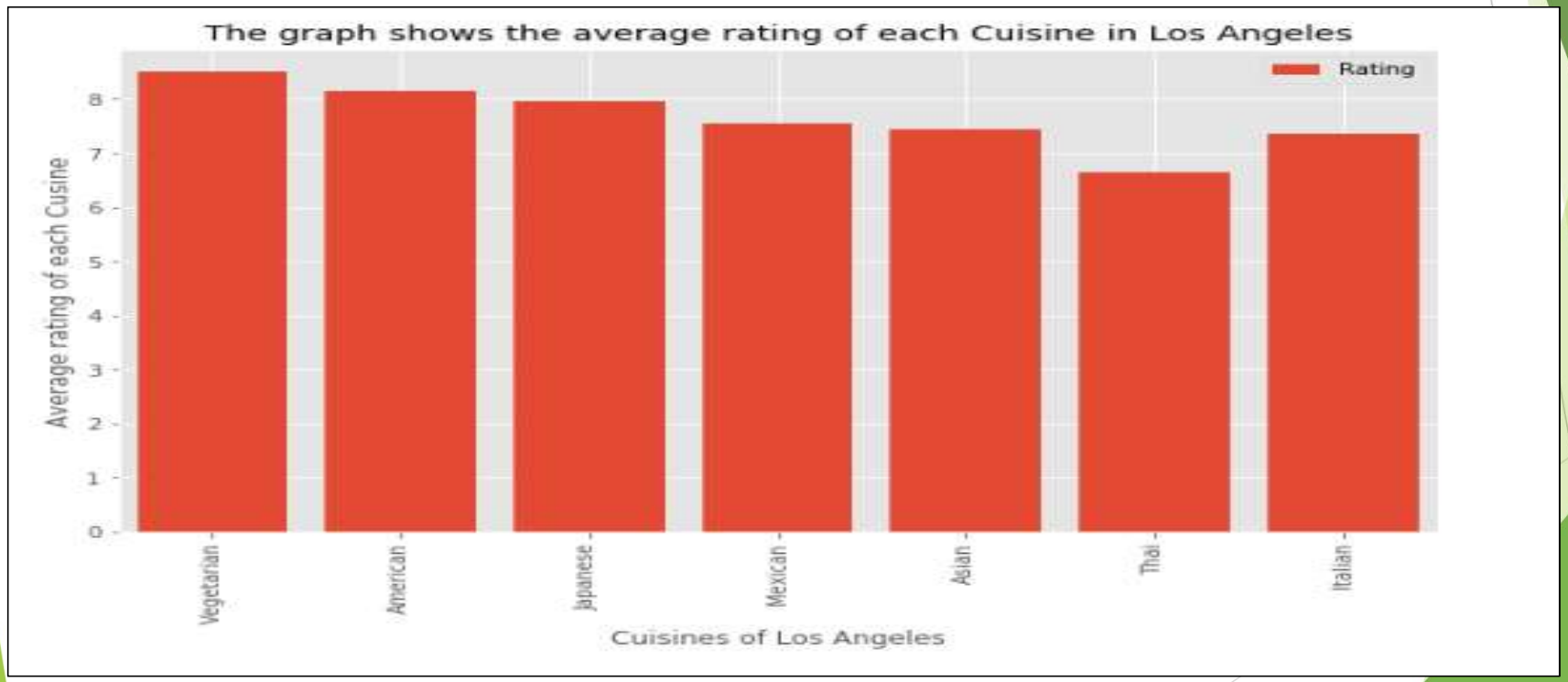
Most Frequented Restaurant Cuisine

- The most frequented restaurant cuisine was determined via an Area Plot shown below.
- The Area plot immediately showed American cuisine was extremely popular



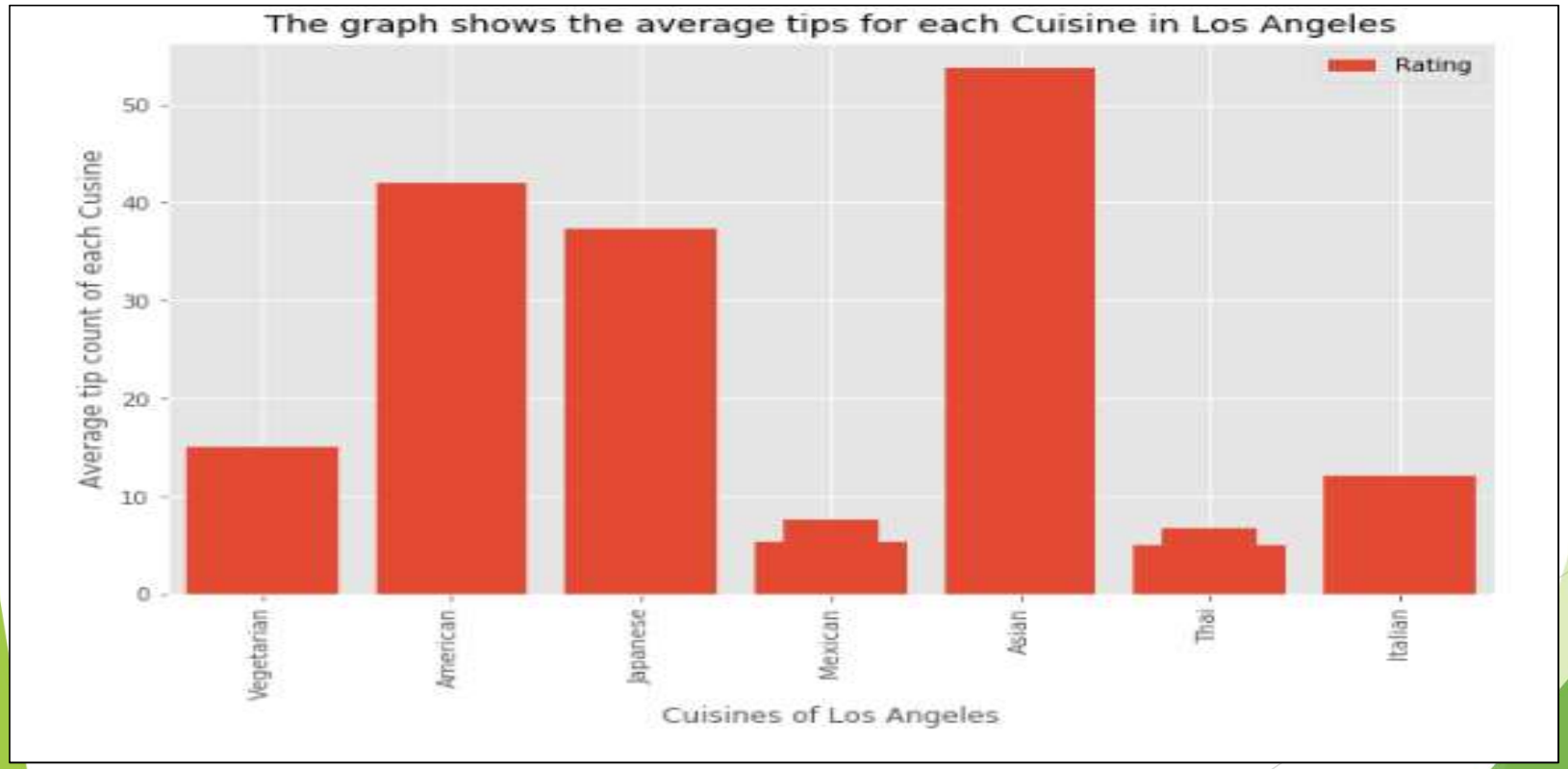
Cuisine Ratings

- The most popular cuisines was chosen from the previous analysis and a sample number of venues were checked for their ratings.
- The results were plotted on a bar graph below which shows Vegetarian, American and Japanese were the highest rated cuisines from the sample.



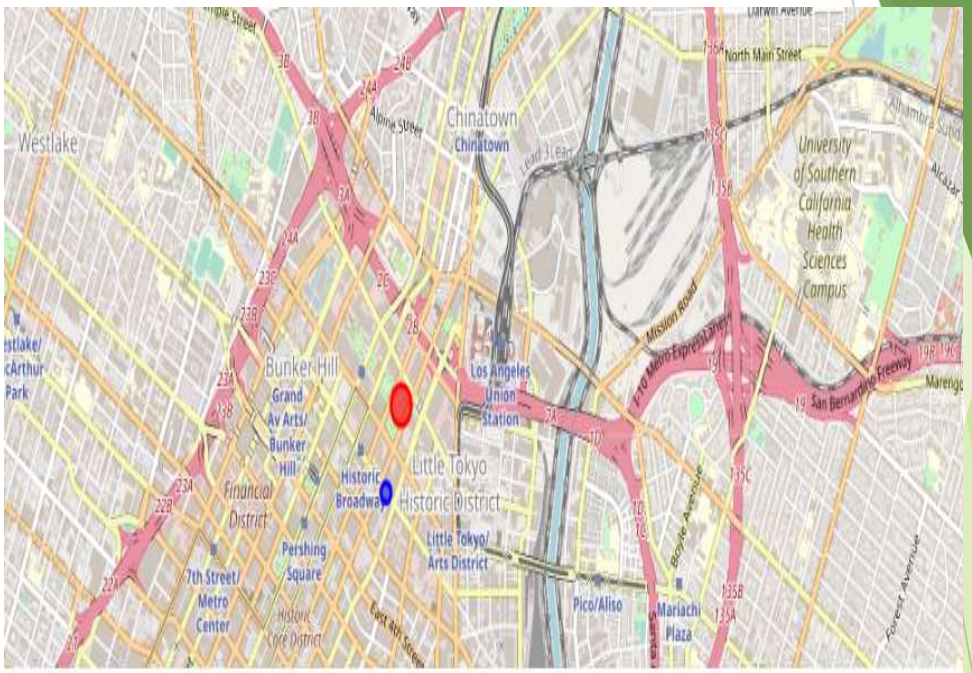
Tip Count Ratings

- The most popular cuisines was chosen from the previous analysis and a sample number of venues were checked for their tip count.
- The results were plotted on a bar graph below which shows Asian, American and Japanese got the highest tip counts.



Trending Venues

- The current trending restaurants were then determined by pulling the latest JSON file.
- The results showed only one American restaurant was trending (blue marker) called Redbird.
- Trending may have been affected by current worldwide issues.



Discussion

- Cluster 1 neighborhoods are close together, with the exception of neighborhoods Playa del Rey and Shadow Hills, so that there is a larger consumer base for a new restaurant that is opened in any neighborhood in Cluster 1.
- Cluster 1 neighborhoods shows that Vegetarian and Japanese are the most popular cuisines however there are limited venues thus it may be ideal to open restaurants with those cuisines.
- Clusters 2, 3 and 4 have one neighborhood per cluster and thus any new restaurant located in these neighborhoods will have a smaller consumer base.
- Cluster 5 neighborhoods are popular with American and Italian type cuisines which dominates the consumer base, thus opening a restaurant in these neighborhoods is not recommended.
- Vegetarian, American and Japanese cuisines have the highest ratings for cuisine
- Asian, American and Japanese have the highest tip counts

Conclusion

- In conclusion the data points to opening a Vegetarian or Japanese type restaurant in any of the following neighborhoods:
  - Bel Air
  - Beverley Crest
  - Pacific Palisades
  - Brentwood
  - Cheviot Hills
  - Hollywood Hills West
  - Encino
  - Rancho Park
- The reasons being for this include:
  - There are limited venues for both Vegetarian and Japanese cuisines meaning less competition amongst venues
  - These places are located in close proximity to each other resulting in a larger consumer base
  - Vegetarian and Japanese venues currently have good rating and good tip counts
  - Cluster 5 neighborhoods are not recommended due to the immense popularity and abundance of American cuisine
  - Cluster 2, 3 and 4 is not recommended due to the lower consumer base as these have single neighborhoods in each