Orange Food Scene

Spencer Au, Dan Boudagian, Miles Allen

Variables

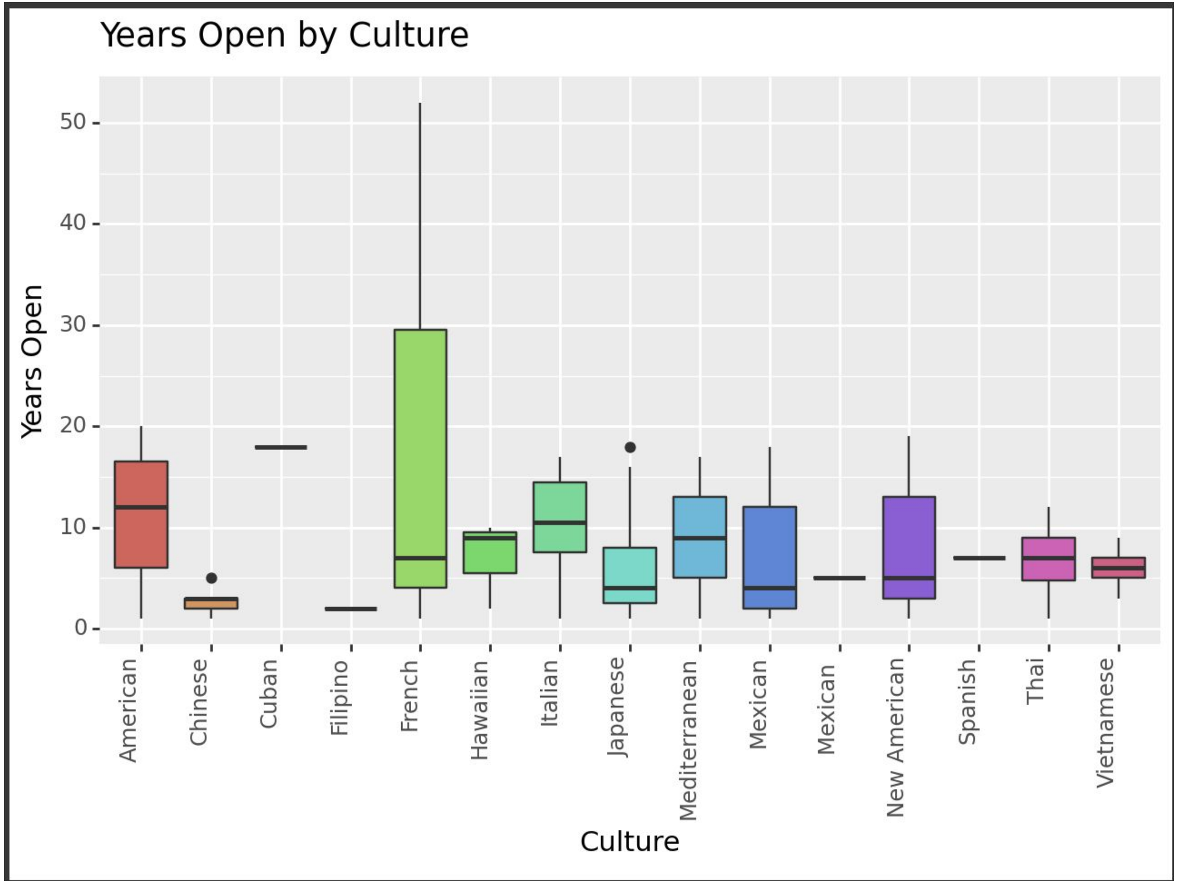
- Restaurant Type: Single restaurant, national chain, or local chain.
- Years in Operation: How long the restaurant has been open.
- Culture: Cultural origin of the restaurant's cuisine.
- Specialty: Main type of food offered by the restaurant.
- Average Meal Price (in USD): Price range of the restaurant.
- Distance from Chapman University: Proximity to the university.
- Seating Capacity: Number of customers the restaurant can serve at a time.
- Ratings and Reviews: Aggregated from platforms like Yelp, Google, etc.
- Competitor Density: Number of similar restaurants in the vicinity.
- Health Inspection Rating: Cleanliness ratings from California health inspectors.
- Alcohol Availability: Categories include Beer, Wine, Liquor, or No Alcohol.

Some Factors at a Glance

- Culture (Categorical)
- Specialty (Categorical)
- Alcohol Availability (Categorical)
- Price (Continuous)
- Years in Operation (Continuous)

Cleaning: Z Score Continuous and One Hot Encoding

Modeling/Computation: Utilize a Supervised linear regression model to predict the years of operation

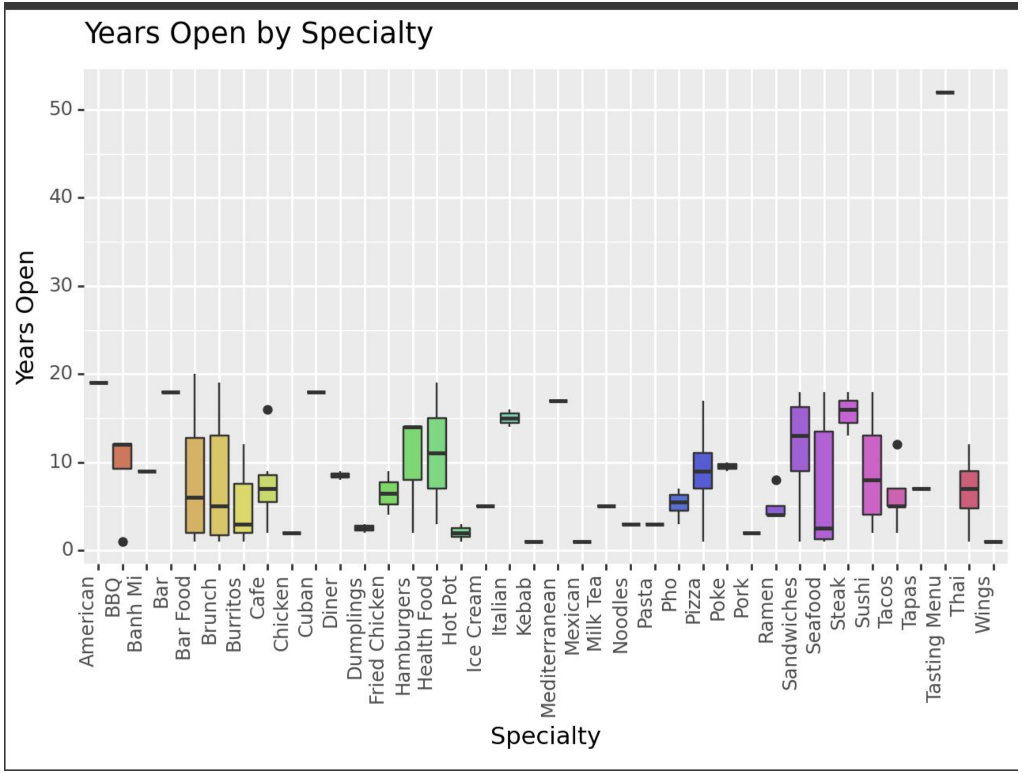


Some Factors at a Glance

- Culture (Categorical)
- Specialty (Categorical)
- Alcohol Availability (Categorical)
- Price (Continuous)
- Years in Operation (Continuous)

Cleaning: Z Score Continuous and One Hot Encoding

Modeling/Computation: Utilize a Supervised linear regression model to predict the years of operation

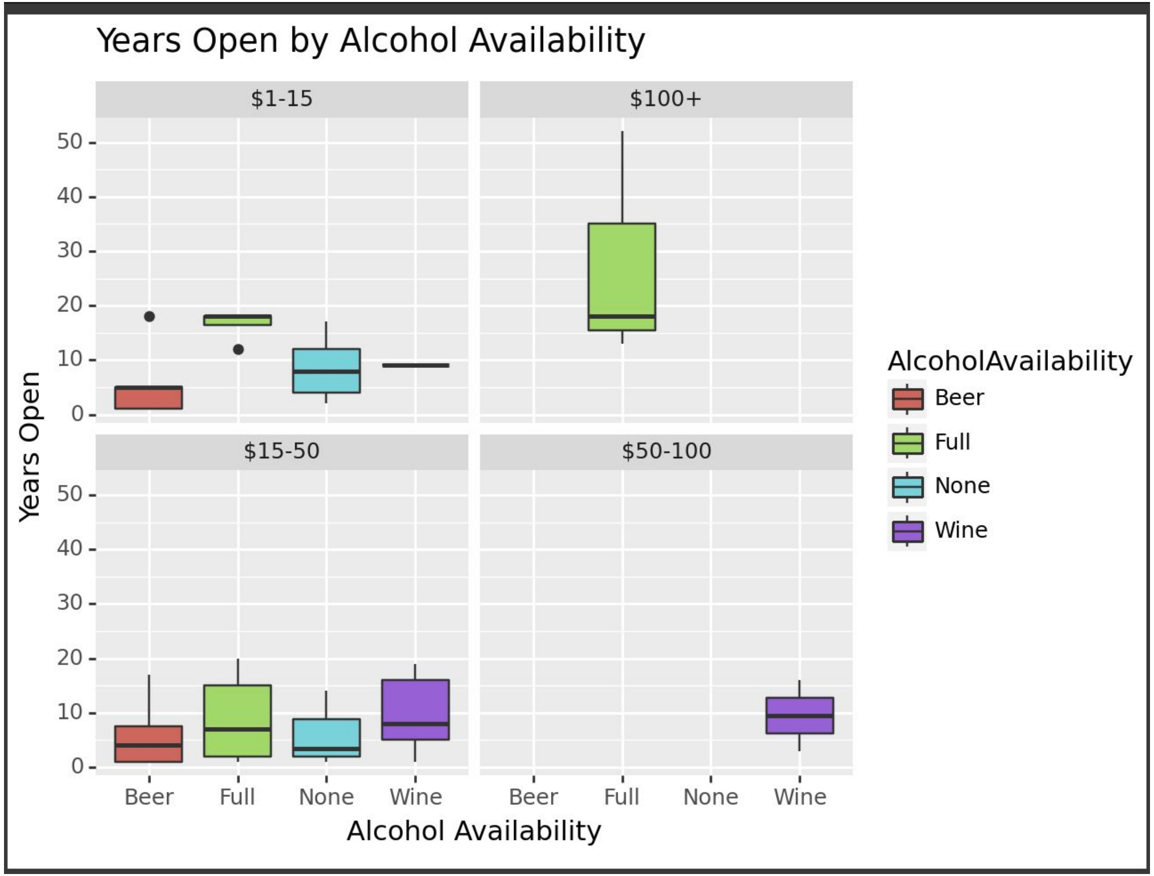


Some Factors at a Glance

- Culture (Categorical)
- Specialty (Categorical)
- Alcohol Availability (Categorical)
- Price (Continuous)
- Years in Operation (Continuous)

Cleaning: Z Score Continuous and One Hot Encoding

Modeling/Computation: Utilize a Supervised linear regression model to predict the years of operation

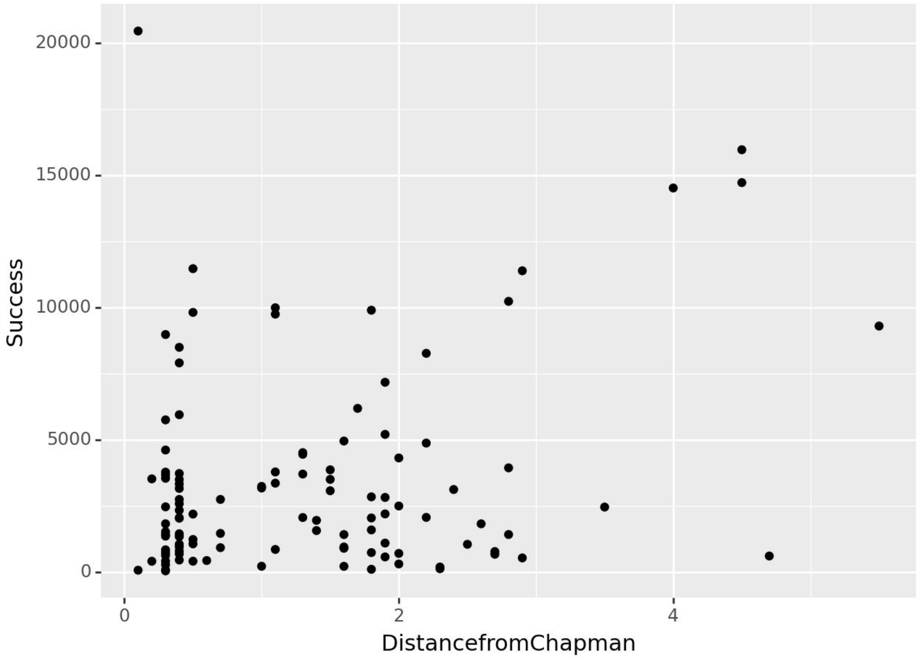


How Important is Being Near Chapman?

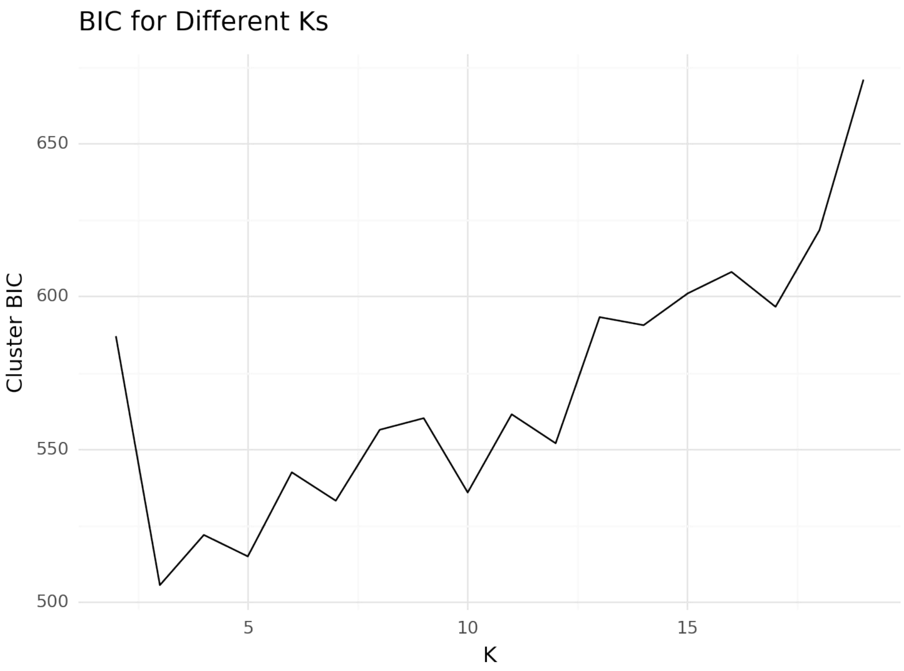
- AverageMealPrice (Categorical Numerical)
- DistanceFromChapman (Continuous)
- Ratings (Continuous)

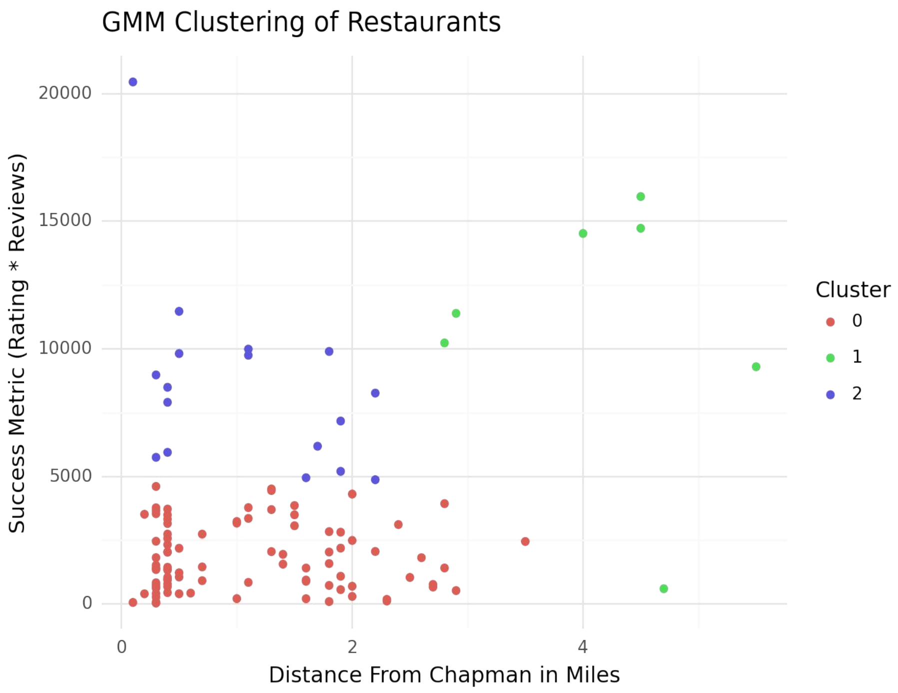
Business Use Case:

- Getting your restaurant closer to Chapman statistically would mean better success rates



Results - Success Clustered





Identifying distinct clusters near Chapman University (Q2)

Variables Involved:

- avg_meal_price (categorical)
- rating(continuous),
- distance_chapman(continuous)

Cleaning:

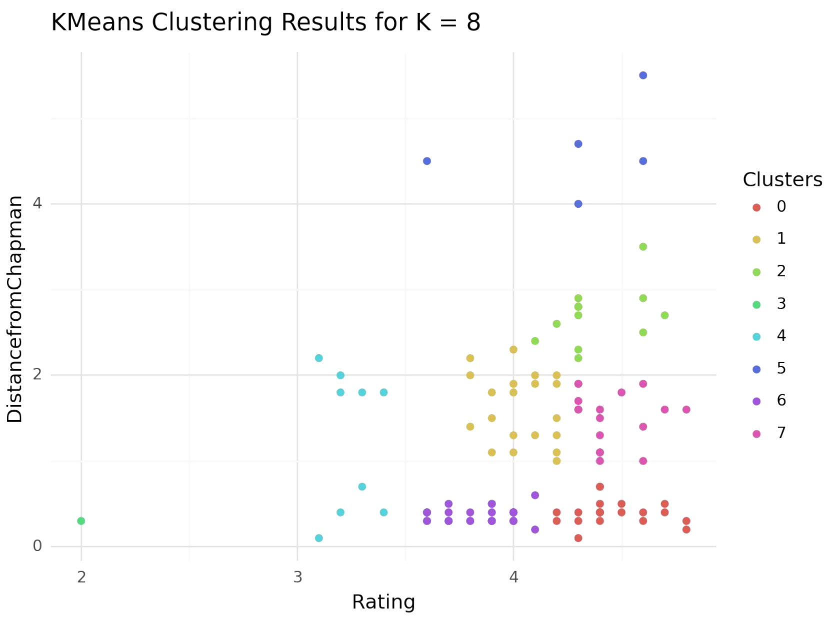
- Missing values will be dropped, and reindexed. Standard Scale continuous Variables and make Avg Meal Price into numerical categories

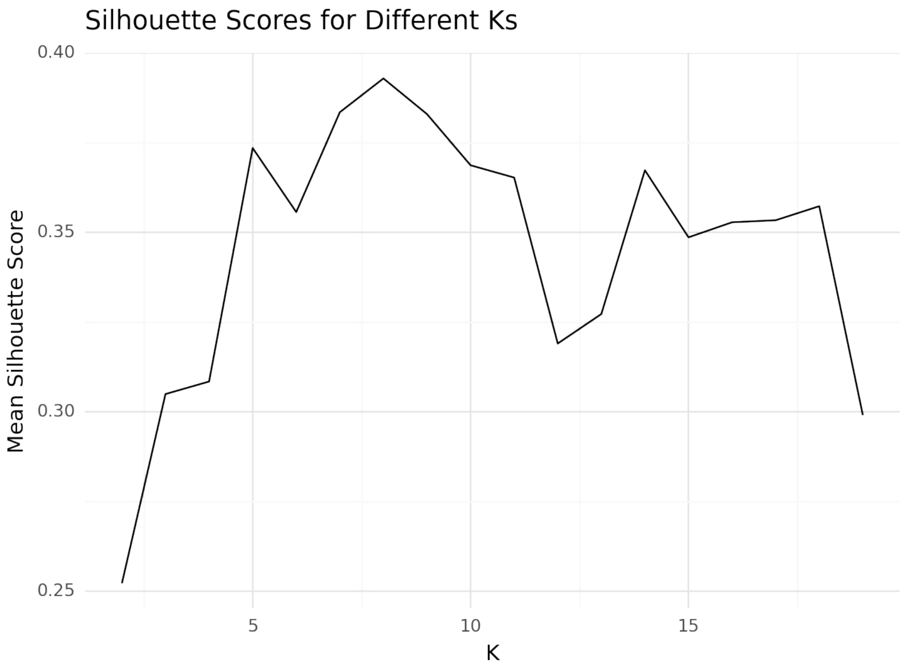
Modeling/Computation:

Make two clustering models K-Means, and Gaussian Mixtures:

- Use the elbow method for K-Means to find the optimal number of clusters.
- For Gaussian Mixtures I use Bayesian Information Criterion (BIC) to choose the optimal number of components (clusters) based on the trade-off between model fit and complexity

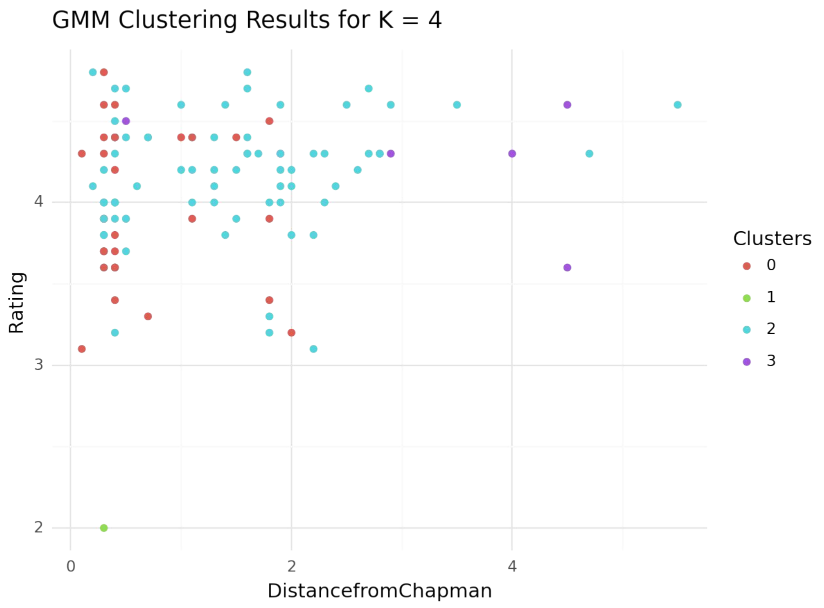
Results - K-means (Q2)

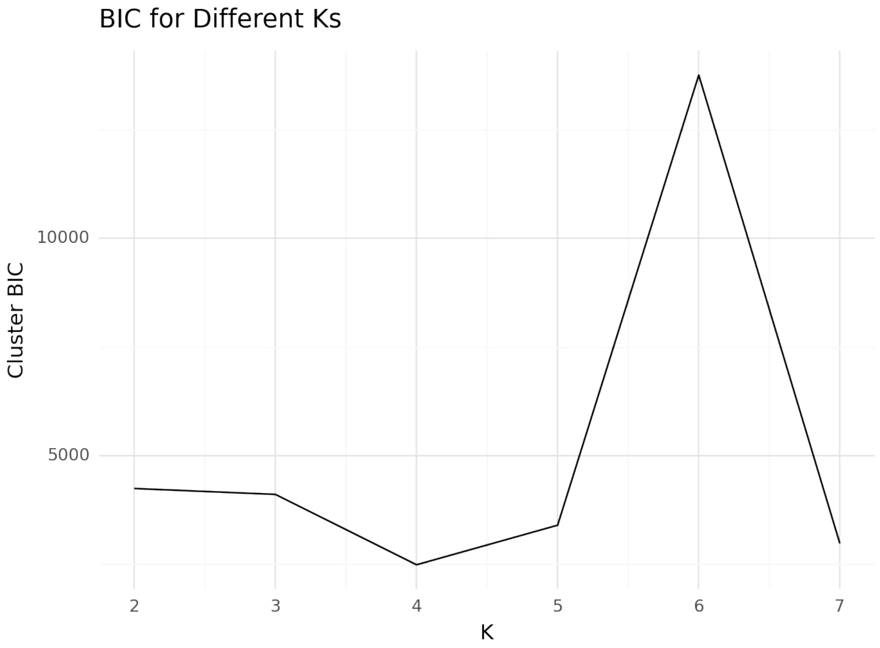




Silhouette Score for KMeans: 0.3929737670502199 7

Results - GMM (Q2)





The Silhouette score is: 0.2531488461571392

Business Use Case:

Market Segmentation:

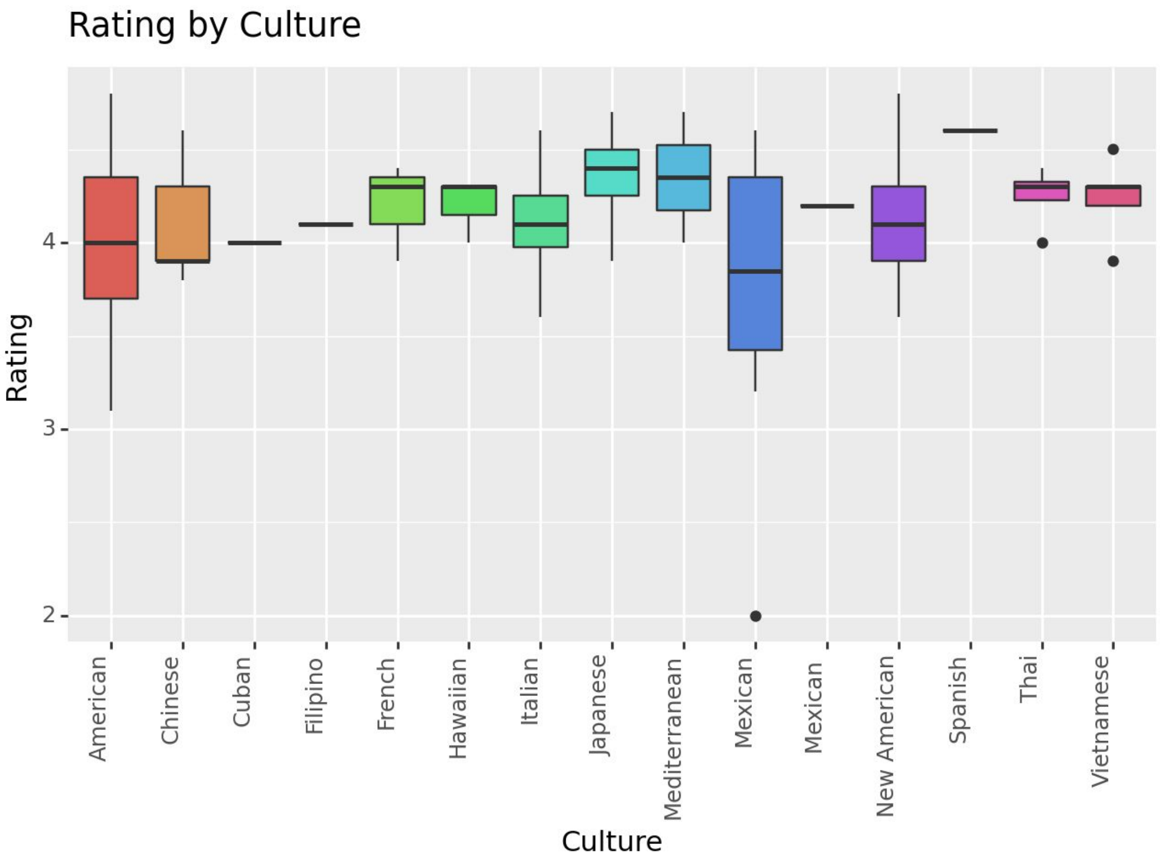
- **Targeted Marketing:**
    - Understanding distinct categories of restaurants can enable targeted marketing efforts. For example, restaurants in a certain cluster with high ratings and short distance from Chapman University might be marketed differently than those with high ratings but a farther distance from Chapman University

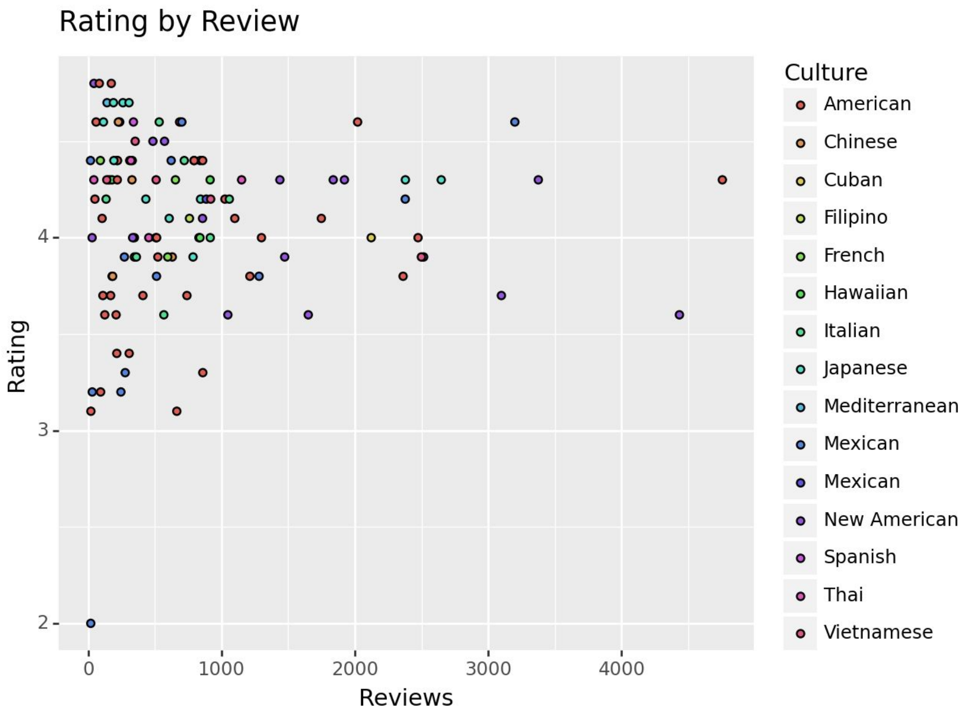
Site Selection:

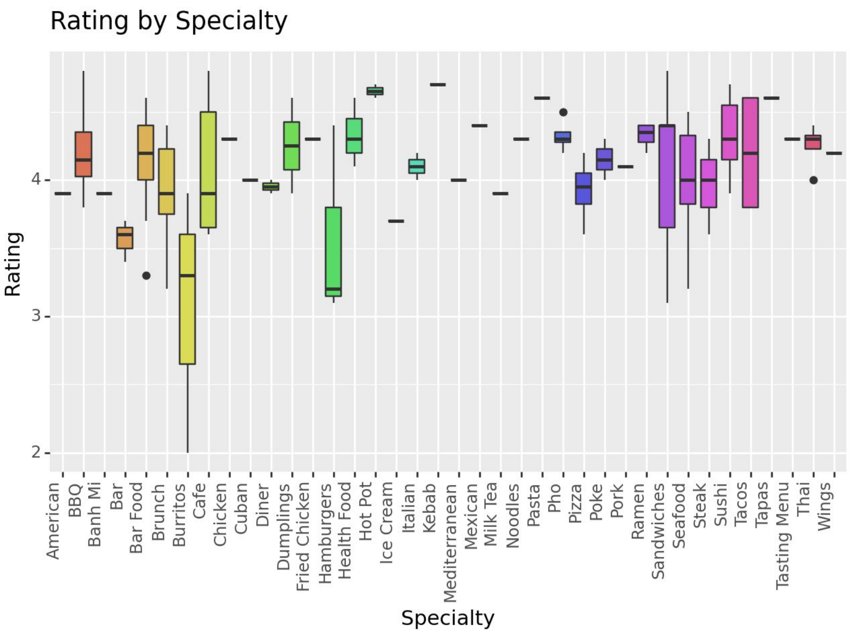
- **Expansion Planning:**
    - Clustering analysis can aid in expansion planning by identifying areas where certain types of restaurants are lacking. It can guide decisions on where to open new establishments based on the existing market landscape

Impact of Culture and Specialty on Ratings (Q3)

- Culture is not a good indicator, however....







Business Use Case:

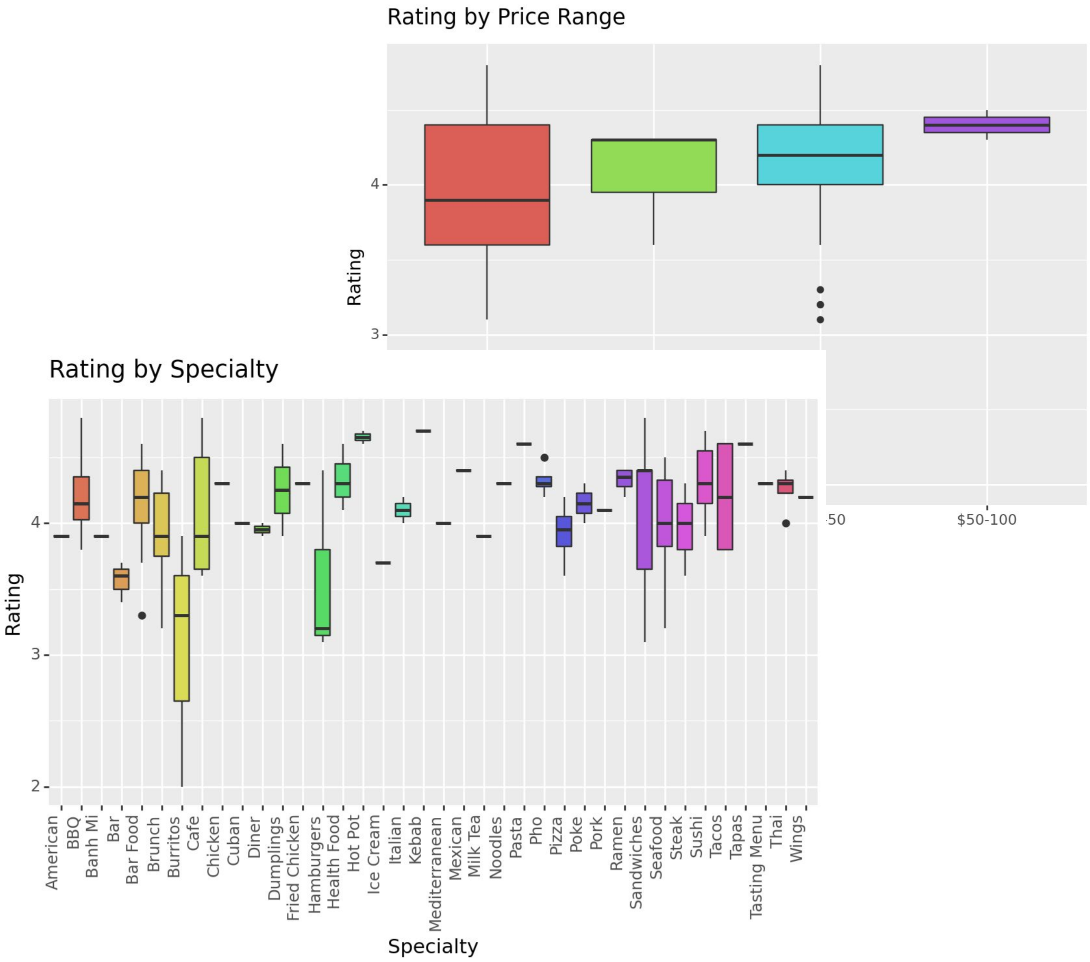
- Cultures with a more casual, finger food approach are more likely to have success among the current demographics.

Influence of Meal Price and Specialty on Ratings (Q5)

- Average Meal Price (Categorical)
- Specialty (Categorical)
- Ratings (Interval))

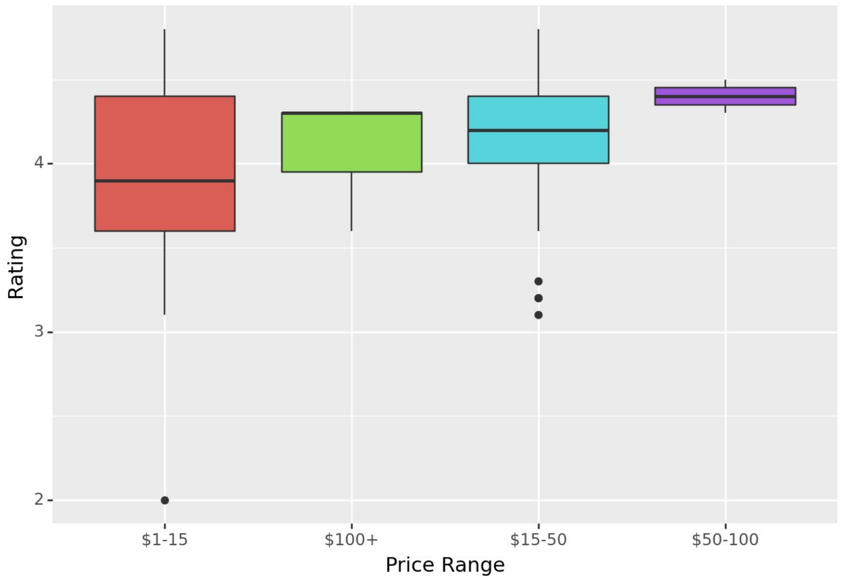
Cleaning: Z Score Continuous and One Hot Encoding

Modeling/Computation: Utilize a Gradient Boosting Tree to predict linear and nonlinear reltationships

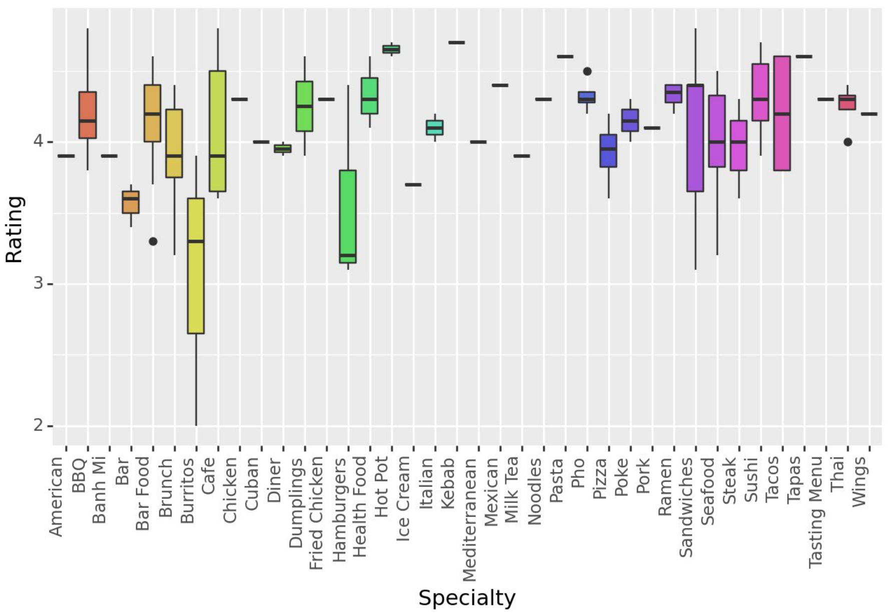


Results (Q5)

Rating by Price Range



Rating by Specialty



Q6: Feature Importance on Average Meal Price

Variables:

- Average Meal Price (numerical categories)
- RestaurantType (Categorical)
- Alcohol Availability(Categorical)
- Rating (Continuous)
- CompetitorDensity (Continuous)
- YearsSinceOpen (Continuous)
- Reviews (Continuous)
- DistancefromChapman (Continuous)

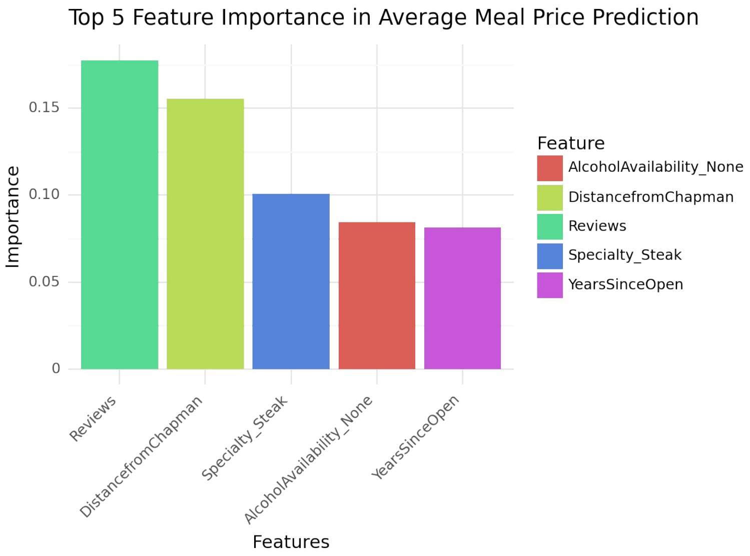
Cleaning and Preprocessing:

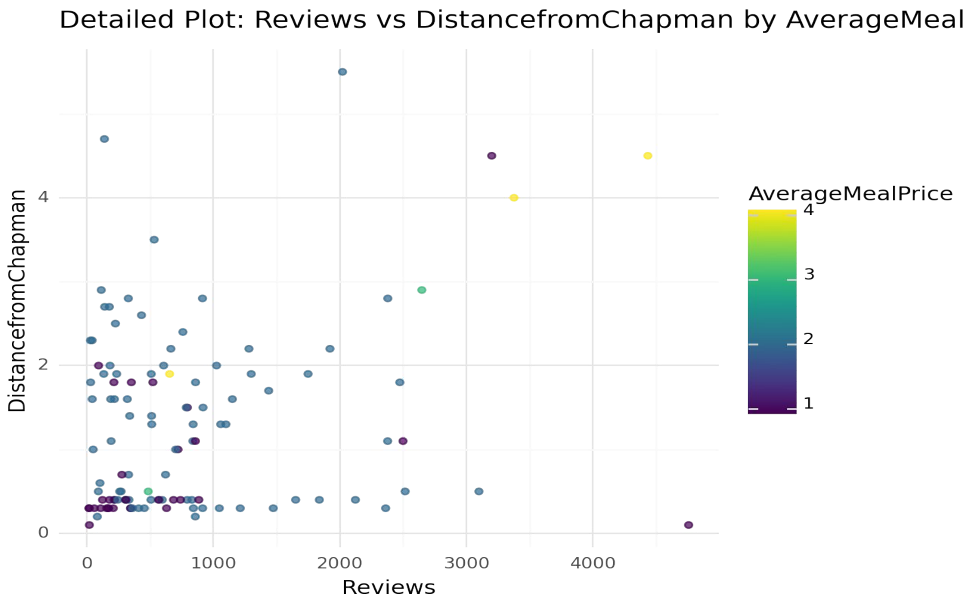
- Handle missing values.
- OneHotEncoder() for categorical variables.
- Z Score Continuous

Modeling/Computation:

- Utilize a Random Forest Model to find most important features
- Use model to predict what are the most important features for predicting average meal price (Categorical)

Results (Q6)





Training Set:
Mean Squared Error: 0.04163133309378368
R-squared: 0.8915772151438O2

Testing Set:
Mean Squared Error: 0.14997484728710025
R-squared: 0.3388608815426998

Business Use Case:

Dynamic Pricing Strategy:

- **Quality-Based Pricing**:
    - Using ratings and reviews to influence pricing aligns with a quality-based dynamic pricing strategy. Higher-rated establishments may ask more premium prices
- **Location-Based Pricing**:
    - Adjusting prices based on the restaurant's proximity to Chapman University reflects a location-specific strategy. Closer proximity might justify higher prices due to convenience or increased demand.