


# The Dining Experience

## San Francisco and Los Angeles

Ambar Mukherjea

5/31/2019



It is favorable to predict the appeal of a restaurant by it's cuisine and other factors

- San Francisco and Los Angeles are large, diverse cities in California
- Offer various cuisine types in the cities' restaurants and food venues
- Restaurant-goers can examine if the type of cuisine influences a venues popularity
- Entrepreneurs can examine where to open a restaurant based on the culinary type and other components

# Venue Data Acquisition and Cleaning

- Data for Venues with food type categories in the two cities are sourced from [www.foursquare.com](http://www.foursquare.com)
  - Popular cuisine types are derived from [www.yelp.com](http://www.yelp.com)
  - Samples containing data from cities outside of San Francisco and Los Angeles are removed
- Data components include the venues' type of cuisine, location price tier, rating, and # of likes
  - Samples with any missing data in these variables are removed
- The features selected for modeling and evaluation are the venues' type of cuisine, zip code, price tier, that influence their # of likes
  - Venues' rating is removed from consideration because it inherently contributes to # of likes



# Regression Models and Machine Learning Used

- Feature Correlation
- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Decision Tree Classification
- Grouped Aggregate Metrics

# These techniques produce negligible models!

- Correlation of Features:

	Venue Type Category	Zip Code	Price Tier
# of Likes (San Francisco venues)	0.01	-0.158	0.109
# of Likes (Los Angeles venues)	0.034	-0.101	0.197

- Model Evaluation:

Model	Predictor Variables	Response Variable	San Francisco Venues		Los Angeles Venues	
			R-square / Accuracy Score	MSE	R-square / Accuracy Score	MSE
Simple Linear Regression	Venue Type Category	# of Likes	0	28670.883	0.001	3832.542
Multiple Linear Regression	Venue Type Category, Zip Code, and Price Tier	# of Likes	0.041	27495.94	0.045	3663.814
Polynomial Regression	Venue Type Category, Zip Code, and Price Tier	# of Likes	0.089	26109.078	0.266	2818.226
Decision Tree Classifier	Venue Type Category, Zip Code, and Price Tier	# of Likes	0.022	-	0	-

..but aggregating features by Cuisine Type can serve as a basic guide

- Grouping of Features:

Cuisine Type	San Francisco Venues			Los Angeles Venues		
	# of Venues	Average # of Likes	Average Price Tier	# of Venues	Average # of Likes	Average Price Tier
Chinese	14	133.357	1.714	10	15	1.6
Cuban	1	218	2	6	81.833	1.833
Ethiopian	4	85.5	1.75	7	29	2
Indian	22	73.682	1.909	6	18.667	1.5
Indonesian	1	145	2	1	182	2
Italian	21	88.81	2.095	14	37.571	1.5
Japanese	19	190.789	1.947	20	53.95	1.85
Korean	14	73.357	2	15	85.2	2.067
Mediterranean	16	74.375	1.688	4	44	1.75
Mexican	29	148.724	1.31	17	39.705	1.235
Middle Eastern	2	150.5	1.5	3	5.333	2
Mongolian	0	0	0	2	7	1.5
Peruvian	4	161.25	1.75	6	32.667	1.833
Thai	46	96.891	1.826	31	48	1.71
Vietnamese	28	126.143	1.321	5	61.4	2

**Observation:** Thai restaurants are abundant in San Francisco, are moderately priced, and are highly liked by customers

# Conclusion and Future Directions

- # of Likes for a venue is not influenced by any other variable
- Diners and entrepreneurs should not use this data to enjoy or provide strong culinary experiences
- Foursquare venue data is limited in usage
- Additional data sources should be used to strengthen models that predict the appeal of a restaurant
  - Structured Data: Population Density, Median Age, Median Income, Cultural Diversity
  - Unstructured Data: Venues' Menus, Tips, Recommendations