Lauren\_Calso\_David\_Colonia\_Project1\_Yelp\_Presentation

1. Introducing data set (yelp\_business.csv)

* Use 7 files from Yelp Open Data Set
* Why: Lots of resources, previous code to build on, seems comprehensive

1. Original question:

Project was to uncover which features from Yelp from the “More Business Info” section (such as take reservations, parking accepts credit card, and more) has the most impact on a Yelp’s star rating for restaurants in San Francisco, CA.

Result:

Dataset lacked variety because lots of data were from some cities and displayed less data than other cities.

Solution:

New question of food types (Pizza vs. Mexican, Pizza in CA vs. Pizza in NY, Mexican food in CA vs. Mexican food in NY)

Result:

Same outcome: Missing data (ex. No Dominoes, Pizza Hut in NY)

1. Analysis:

Good business data for Las Vegas: variety, details, reviews, most businesses were restaurants

1. Argument:

Possible bias persists.

1. Null Hypothesis: Business data sets from csv files and from the API key will have the similar results such as tables, graphs, ratings, number of reviews.
2. Alternate Hypothesis: Business data sets from csv files and from the API key will not have the similar results such as tables, graphs, ratings, number of reviews.
3. Explain data for API key and results.
4. Explain code/graphs of the restaurants in Las Vegas (top business). Explain user’s choice of review works for a restaurant in Las Vegas with a high star rating compared to a restaurant in Las Vegas with a low star rating.
5. Final conclusions between datasets from CSV and from API key.