Yelp Business Decision Analysis Project

By Chaz and Sunil

Summary / Overview

- This project analyzes the Yelp Data Set to properly educate and propose a new business idea to an investor
- Recommendations to be based on data analysis
- Based on our research, our findings conclude that opening a **High-End Indian Restaurant in the Washington D.C. and Los Angeles, CA** markets would be safe, stable and carries with it less risk.

Business Problem

We made certain assumptions w.r.t. both the business and utilization of effective data analysis techniques, which has carried over in our analysis

Sample Data Set has to be large

(> 100 businesses)

Implementation of quality and service of proposed business will be same as comparables

Proposed business will carry the least risk possible

Data / Yelp API

INPUTS

'category'

and

'location'

OUTPUTS

Business Information

- Name
- Location
- Category
- Price Level (1-4)
- Rating (0-5)
- # of Reviews

Review Information

- Text of review
- Review Rating
- Time Created

LIMITATIONS

- Only returning a maximum of 1,000 businesses per criteria
- Only providing business information within maximum radius of 25 miles
- Only providing three sample reviews per business.

SAMPLE DATA CAPTURED

Business Name	Business Categories	Address			Price Level	Reviews Count	Average Rating
Indika House	Indian, Asian Fusion	943 Broadway	Brooklyn	NY	2	74	5
Masala Grill	Indian	501B Atlantic Ave	Brooklyn	NY	2	215	3.5
Kinara II	Indian	368 Myrtle Ave	Brooklyn	NY	2	90	3.5
Dosa Hutt	Indian, Vegetarian	45-63 Bowne St	Flushing	NY	1	220	4.5
aRoqa	Indian, Bars, Tapas	206 9th Ave	New York	NY	2	300	4

Method Used Bottom's Up Analysis

Target Business Profile:

- **Indian** Restaurant (due to our interest)
- **High-End** Price (due to assumption that category offers most profitability)
- **Urban**

EVALUATE DIFFERENT MARKETS

Restaurants had best data

- ✓ Significant sample size
- ✓ Meaningful "Price Level" values

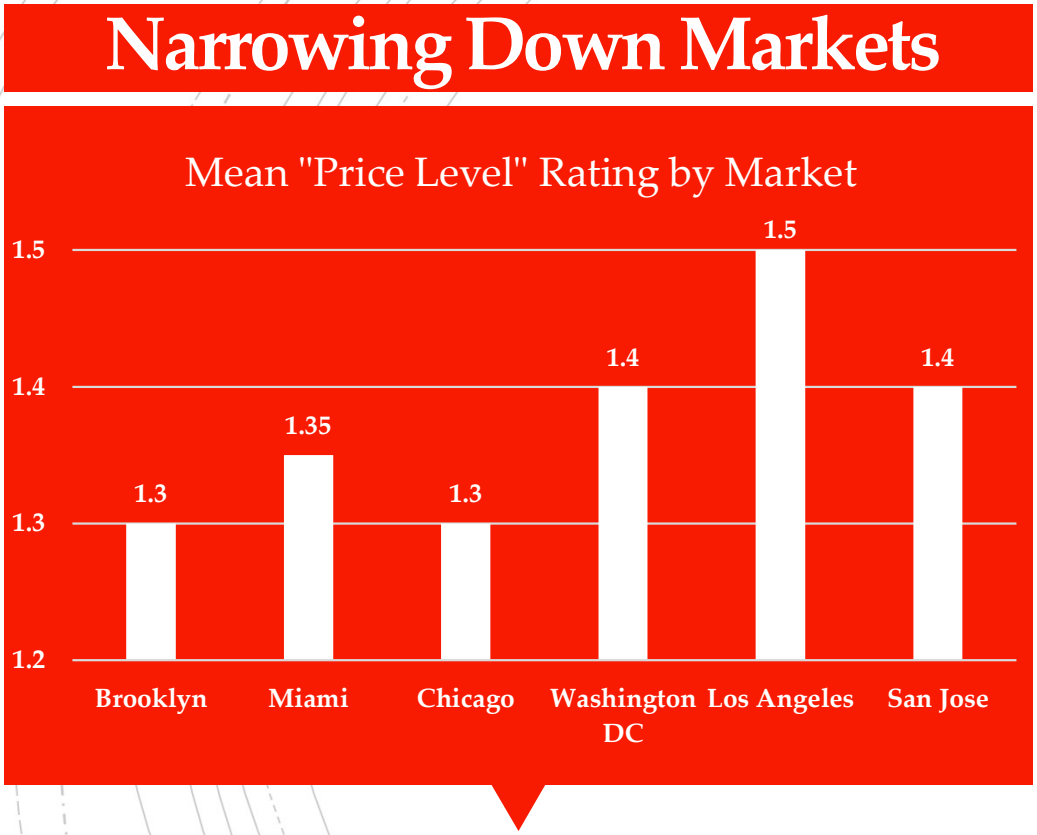
Analyzed data from each:

['HVAC Operator', 'Indian Restaurant', 'Estate/Divorce Attorney']

Identified Top 3 Yelp Categories:

[Home/Health Services, Restaurants, Professional Services]

Narrowing Down Markets



Starting with 6 Urban Markets…

Top 3 Markets with the highest average "Price Level"

Mean value: accurate representation of concentration of available high-quality restaurants

Markets with higher concentration of high-end restaurants

San Jose, CA

Washington D.C.

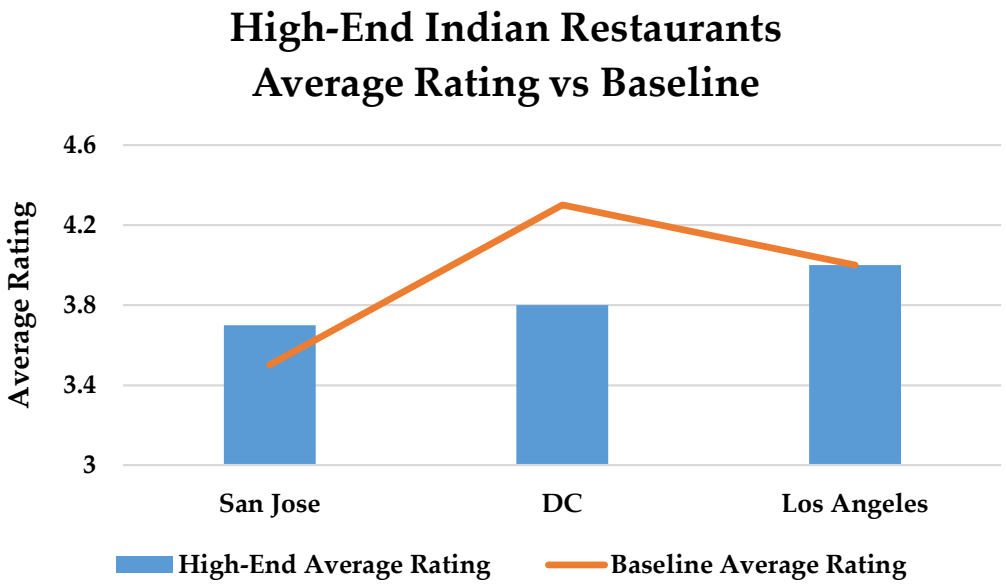
Los Angeles, CA

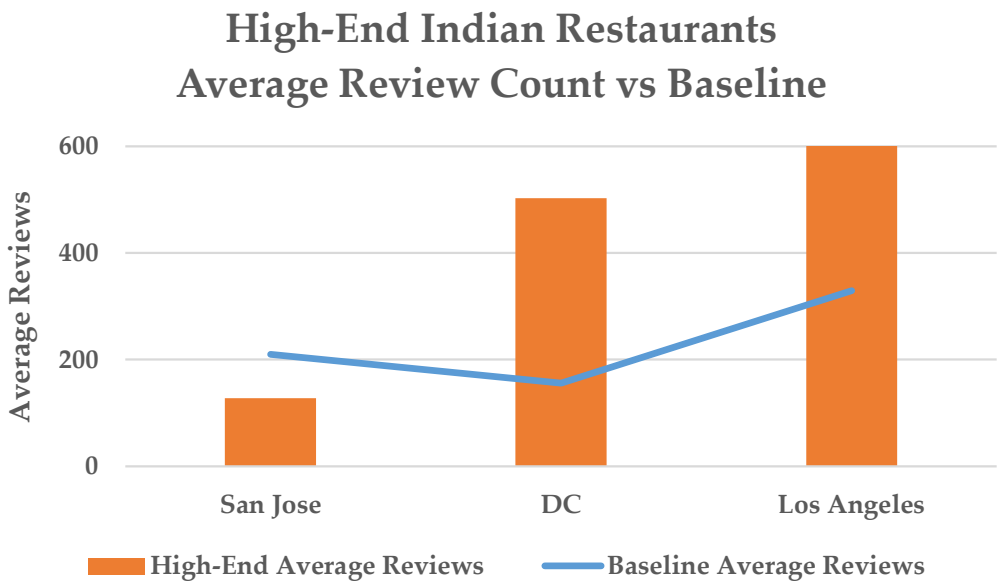
Creating Baselines / “Averages”

- **"Per Business Review Count" –**
  - Median value of all business review counts.
  - Reason: Accounts for outliers in the data
- **“Per Business Rating" –**
  - Mean value of all the business ratings.
  - Reason: Properly captures the "average" rating

	Baselines	
	Reviews	**Ratings**
San Jose	210	3.7
Washington D.C.	156	3.8
Los Angeles	329	4

GOAL IS TO COMPARE AGAINST PERFORMANCE OF EXISTING HIGH-END INDIAN RESTAURANTS IN RESPECTIVE MARKETS





High-End Indian Restaurant Performance

Assumptions made on "Averages"

Reviews: Just as popular, as much, if not more demand

Ratings: Consumers appreciative of quality and provide feedback

Better metrics mean they are healthy, less-risky markets to enter

Results & Conclusions

**** Open a High-End Indian Restaurant in the Washington D.C and Los Angeles, CA markets ****

There is a higher concentration of existing high-quality Indian restaurants in these markets, signifying demand and less-risk, and the ratings and reviews are equal to, if not better than, the baseline for the average Indian restaurant in their respective markets, signifying consumer appreciation and less-risk/stability.

Next Steps

Further analyses could yield additional insights to further improve our business process methods and our data analysis results

"Better prediction of business demand" This modeling could use already available data, such as Census demographics data and Consumer Surveys, to better determine the demand for the particular offering.

"Model need for business success indicators" This modeling use other business-financials-health relevant data to incorporate into our captured businesses to determine further correlations and better predictive outcomes

Thank You

By Chaz and Sunil