

Yelp Business Insights

Shan He
Oct 2018

Agenda

- Data Context
- Data Overview
- Business Questions
- Data Transformation
- Key Insights

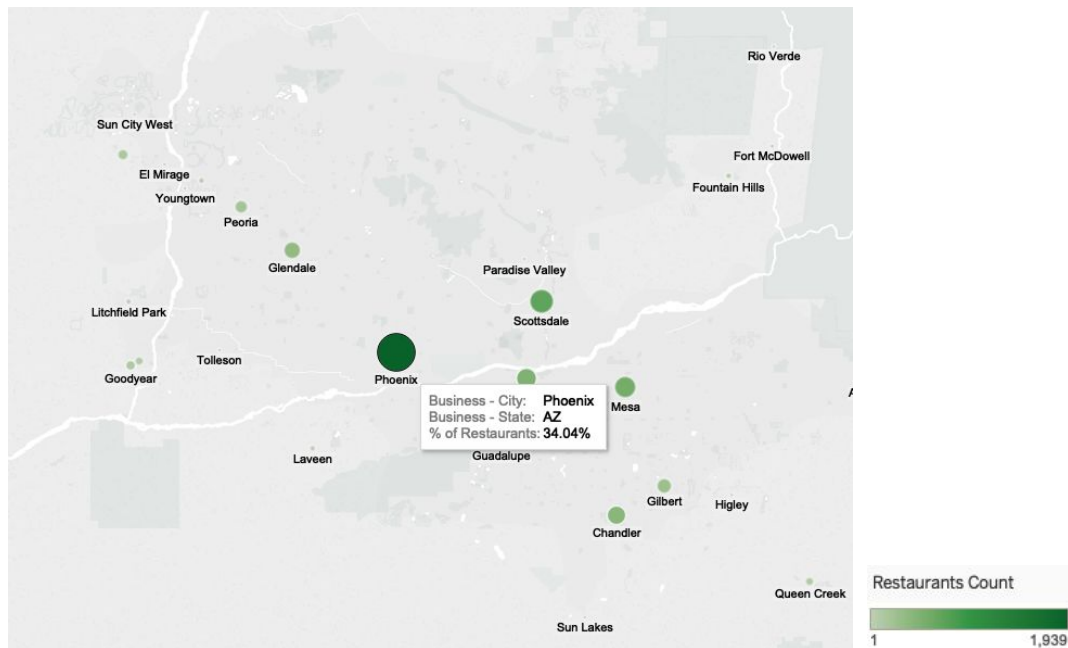
Data Context

Source data includes Yelp user reviews for restaurants with a scope of

- 285,764 reviews
- 71,089 unique users
- 5,696 unique restaurants
- Reviews dated from 02/01/2015 - 07/16/2014
- Restaurants from 52 cities in AZ

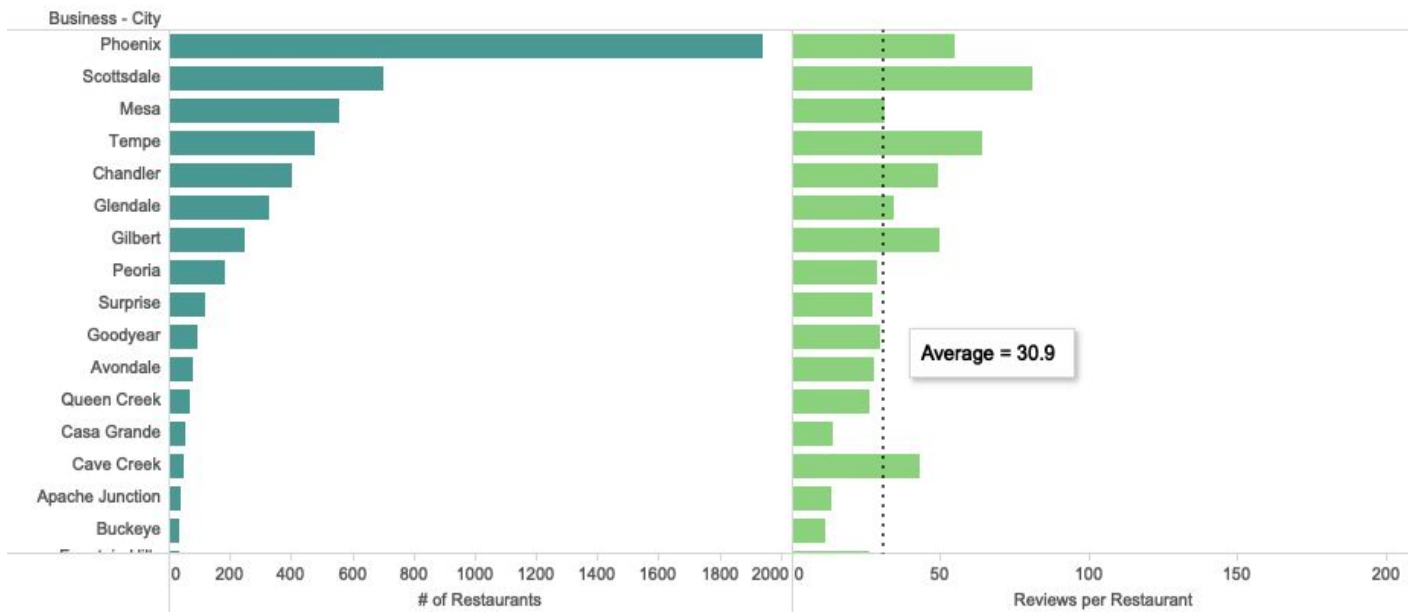
Data Overview - Restaurant Location

1. Phoenix alone accounts for 34% of the reviewed restaurants



Data Overview - Restaurant Review

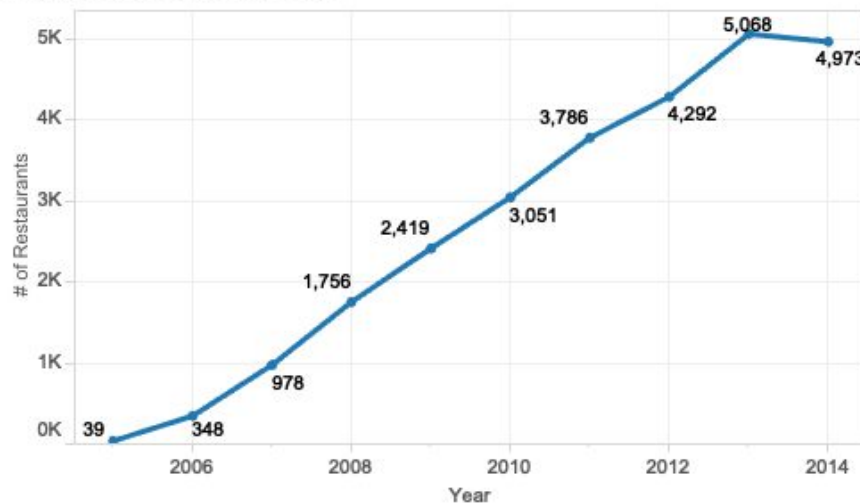
2. The average number of reviews for each restaurant is 30.9, with Scottsdale having the most active reviewers among the top cities (in terms of # of restaurants)



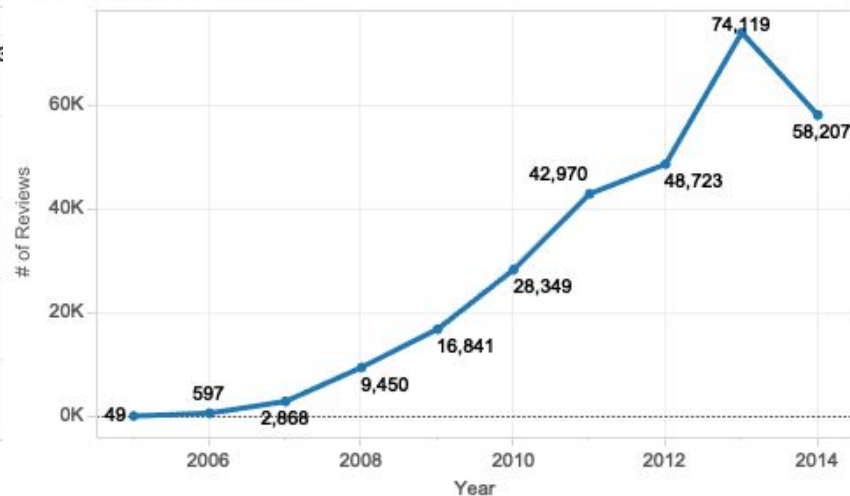
Data Overview - Restaurant Review

3. Although cutoff in mid 2014, data shows high growth rate in terms of # of reviews and # of restaurants reviewed on Yelp

of Restaurants over time



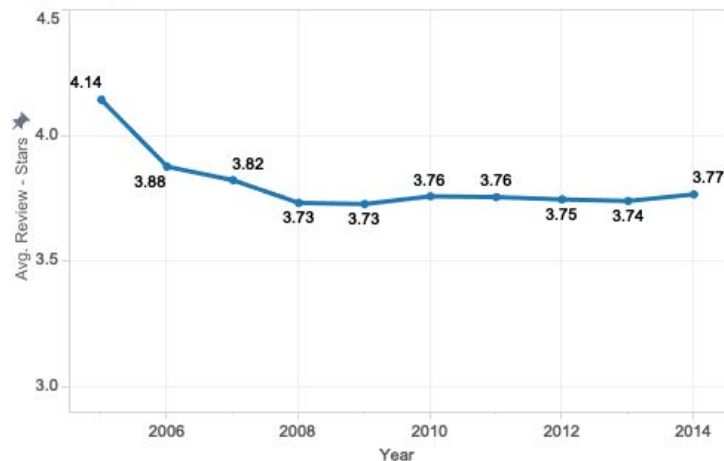
of Reviews over time



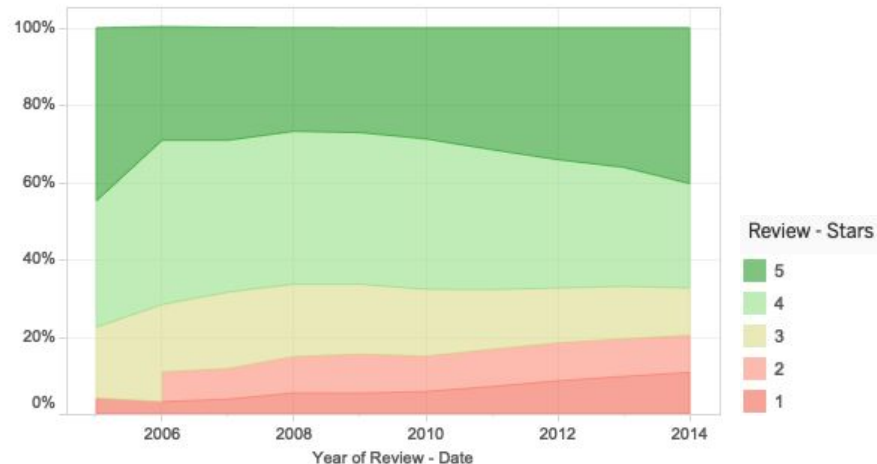
Data Overview - Rating

4. Besides a skewed sample in 2005-2006, the average ratings have been stable. Most reviews are 4 or 5 stars and recently the proportion of 5 star reviews has been rising

Avg Rating over time



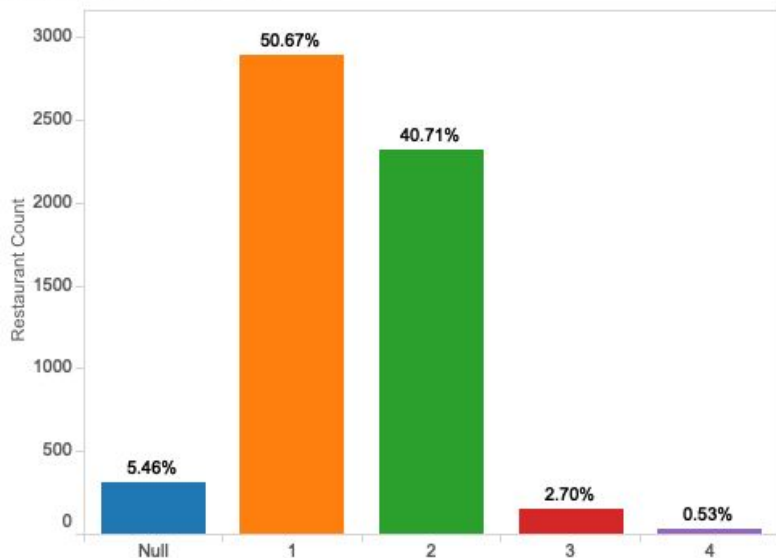
Review Distribution



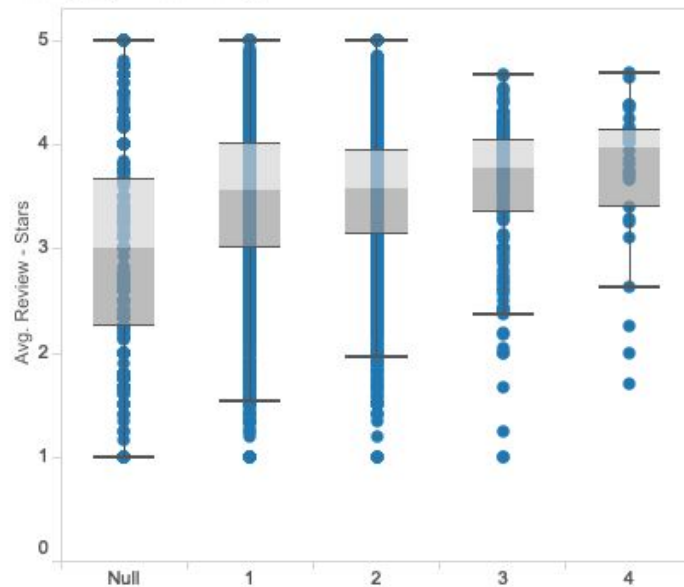
Data Overview - Price Range

5. Half of the reviewed restaurants are in the '\$' price range. The median and 25th quantiles for average ratings seem positively correlated to price range

Price Range Distribution



Rating by Price Range



Key Business Insight - Business Questions

We have a good overview of how the restaurants are reviewed as a whole. But can we identify any category-specific trend?

- What are the most dominant/popular restaurant categories?
- What restaurant categories are getting more interest over time?
- What restaurant categories are growing the fastest?

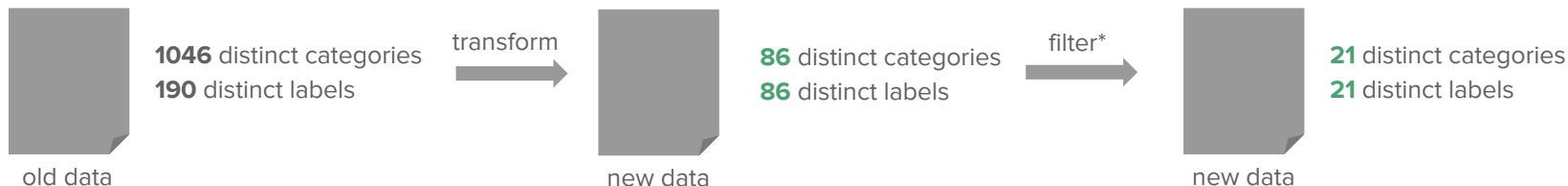
However, Yelp allows restaurants to choose multiple categories for their business and it's challenging to compare across categories meaningfully

Key Business Insight - Data Transformation

Data transformation is then performed to **extract** the **general categories**. We take the “**popular vote**” approach to decide the most descriptive label for a restaurant with multiple labels. For example:

[‘Sushi Bar’, ‘Asian’, ‘Restaurant’] --> [‘Sushi Bar’]

since ‘Sushi Bar’ occurs most frequently out of all unique combinations. As a result -

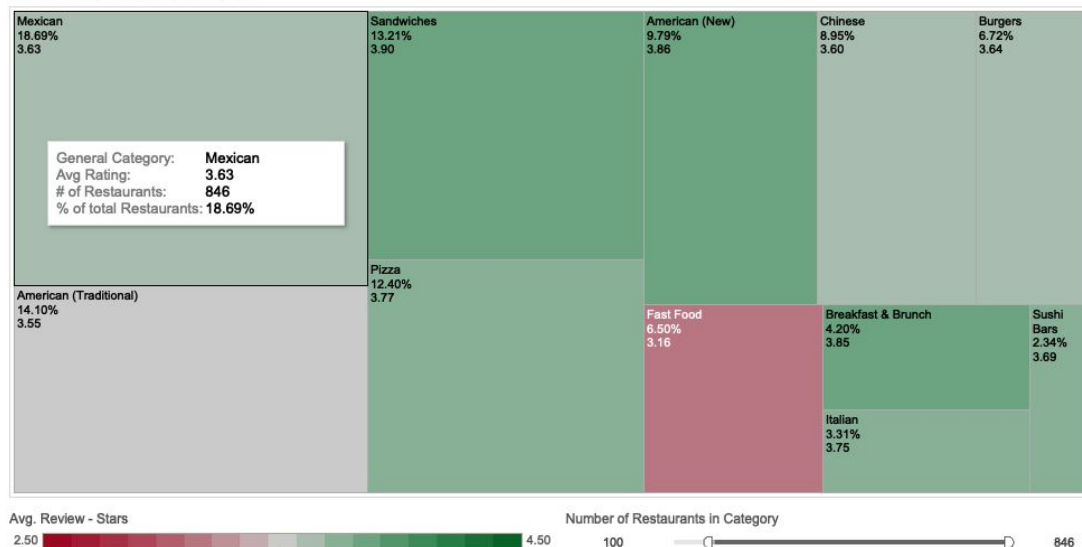


* Filtering involves looking at the non-generic categories with over 20 restaurants

Key Business Insight - Popular Restaurants

The average ratings of the most popular categories vary a lot, with Sandwiches shops being the most highly reviewed and fast food restaurants being the lowest.

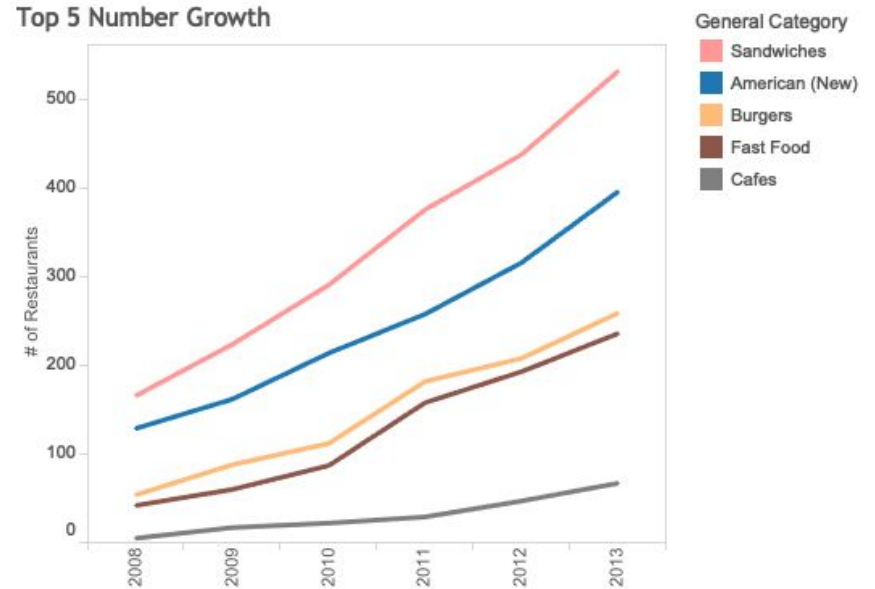
Overview by General Category



Key Business Insight - Establishment Growth Rate

Looking at the 2010 - 2013, these are the 5 categories that have the fastest growth rate in terms of their restaurant establishment

Rk.	Category	Avg. Annual Growth (Number)
1	Cafes	45%
2	Fast Food	42%
3	Burgers	34%
4	American(New)	23%
5	Sandwiches	22%

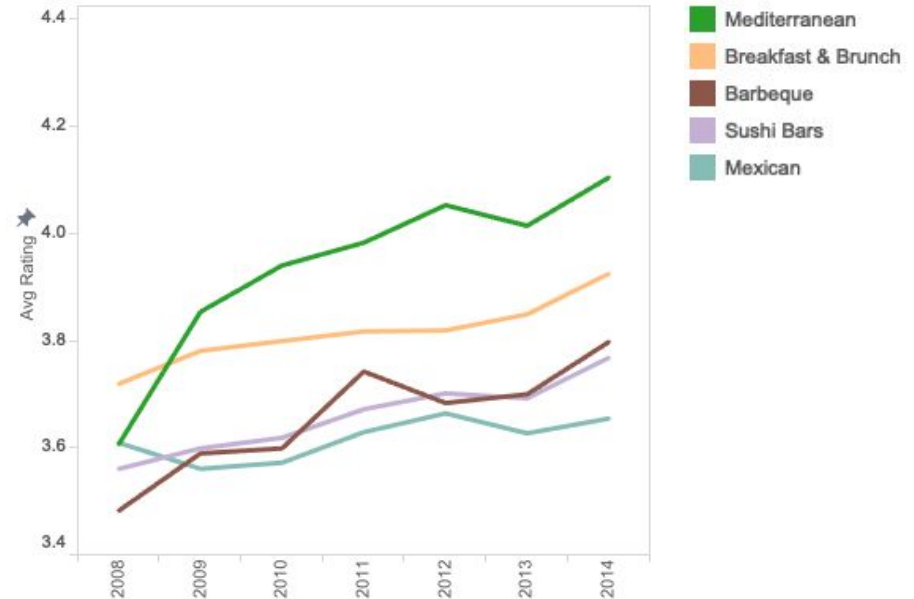


Key Business Insight - Rating Growth Rate

Looking at the 2010 - 2013, these are the 5 categories that have the fastest growth rate in terms of their average ratings

Rk.	Category	Avg. Annual Growth (Rating)
1	Barbeque	1.0%
2	Sushi Bars	0.7%
3	Mediterranean	0.6%
4	Mexican	0.5%
5	Breakfast & Brunch	0.4%

Top 5 Rating Growth



Thank you!

Click [here](#) for iPython Notebook and here for [Tableau Dashboard](#)