



Web Scraping Project: Yelp vs Michelin Ratings in the Bay Area

August 14, 2016



Agenda



- Context and Overview
- Analysis and Findings
- Next Steps

Agenda

- **Context and Overview**
- Analysis and Findings
- Next Steps

Michelin stars are coveted by restaurants across the globe, yet we have little insight into what drives ratings

Yelp vs Michelin Rating Systems

Factor	Yelp 	Michelin 
Scale	<ul style="list-style-type: none"> 1 to 5 stars given by reviewer Yelp provides aggregate rating 	<ul style="list-style-type: none"> 1, 2, or 3 stars Very good, excellent cooking, exceptional cuisine respectively
Reviewer	<ul style="list-style-type: none"> Anyone with an internet connection Provides range of data points / opinions 	<ul style="list-style-type: none"> Highly trained professional Secret identity "Snob" factor
Scope	<ul style="list-style-type: none"> In theory, all restaurants in the world 	<ul style="list-style-type: none"> < 300 restaurants globally with star In US, primarily NYC, SF, and Chicago are reviewed
Prestige	<ul style="list-style-type: none"> Less prestigious than Michelin rating, but still has impact on bottom line Varies by city 	<ul style="list-style-type: none"> Highest honor "Oscars" equivalent
Criteria	<ul style="list-style-type: none"> Public: reviewers often write up detailed reports 	<ul style="list-style-type: none"> Secret: Michelin does not disclose rationale
Timing	<ul style="list-style-type: none"> Reviews published each day at will 	<ul style="list-style-type: none"> Annual guidebook published in October with many surprise changes in ratings

On the other hand, Yelp makes its critiques public: what if we used this data to evaluate Michelin restaurants?

Questions to Evaluate

- How do Michelin ratings in the Bay Area compare to Yelp? Are they consistent?
- What is the typical Yelp profile of a Michelin restaurant?
- During the year in which restaurants gain or lose Michelin stars, what happens to their Yelp reviews?
- Can anything in Yelp help us predict upcoming Michelin ratings?



Potential Impact

On Restaurant Owners

- Dramatic effect on prices charged → restaurants would love to better understand how to gain a Michelin star

On Consumers

- Foodies love to visit restaurants “before they’re cool”
- Could potentially dine at trendy new place before it becomes too expensive

By scraping Yelp data we can establish a baseline for Michelin restaurants and see which up-and-comers might be due for a star

Yelp vs Michelin Web Scraping Approach

Extract and Transform

- Scraped 50,000+ reviews from every SF area restaurant with Michelin star since 2010
- Captured key fields such as review rating, whether reviewer is “elite”, etc.

Baseline Analysis

- Created typical profile of each Michelin star level
- Used two sample t-test to check whether variable is significant for restaurants that gain star vs status quo

Prediction

- Sampled 38 non-Michelin high end restaurants
- Applied against baseline profile and significant variables
- Ranked candidates to predict which likely to gain Michelin stars

One SF restaurant is outperforming based on Yelp metrics: I predict it will gain a star in either the 2017 or 2018 Michelin Guide

Agenda

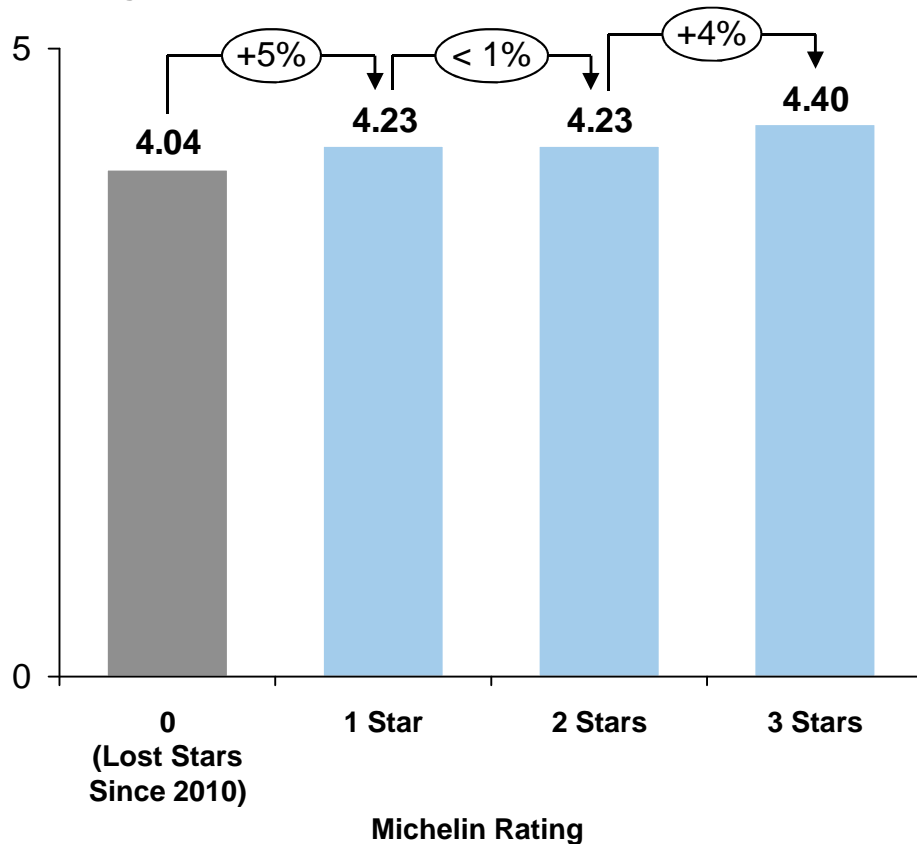
- Context and Overview
- **Analysis and Findings**
- Next Steps

As expected, Yelp and Michelin ratings are correlated; at first glance these restaurants seem to get a lot of elite reviews

San Francisco Bay Area Yelp vs Michelin Ratings

(As of 8/2016, includes 62 restaurants w/ at least 1 star since 2010)

Yelp Rating

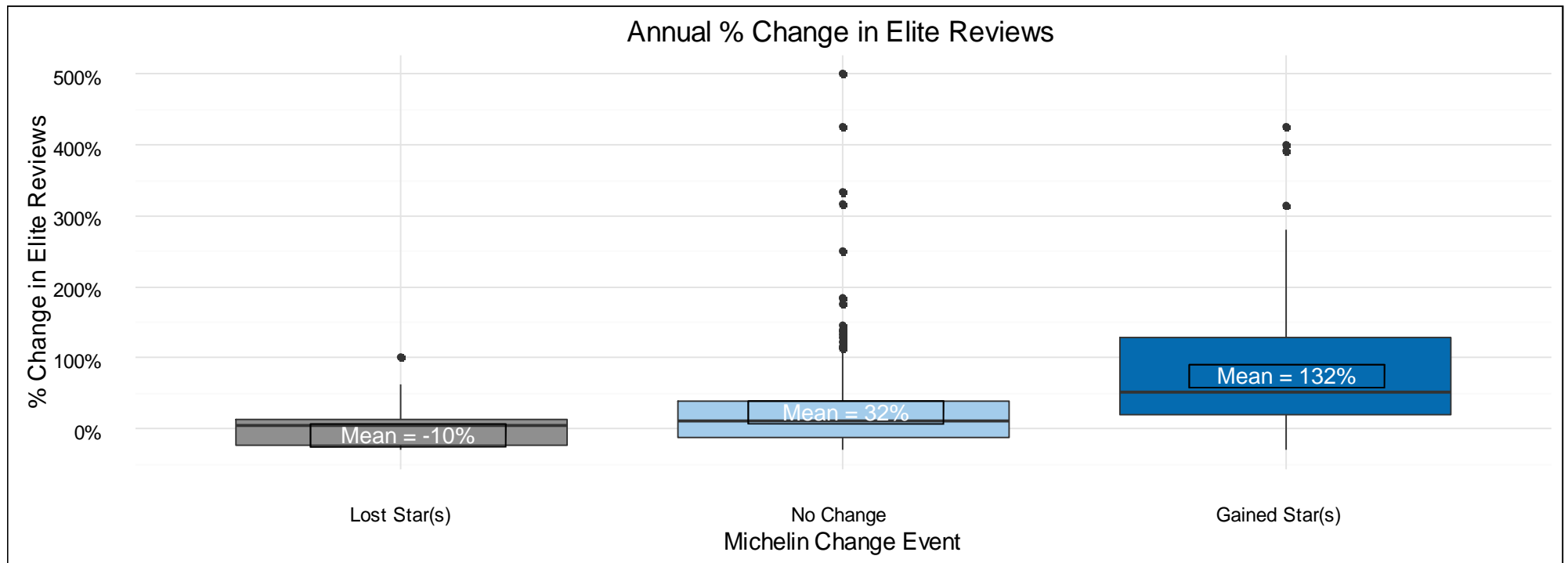


Typical Yelp Profile for SF Michelin Restaurant

(As of 8/2016, includes 62 restaurants w/ at least 1 star since 2010)

Characteristic	# of Michelin Stars (2016)			
	0 (Lost Stars)	1	2	3
Restaurant Count	15 (24%)	35 (56%)	7 (11%)	5 (8%)
Average Reviews	827	623	563	714
Average 5 Star Reviews	364 (44%)	310 (50%)	298 (53%)	437 (61%)
Average Elite Reviews	112 (14%)	109 (17%)	85 (15%)	123 (17%)
Min Total Reviews	194 (Santé)	54 (All Spice)	299 (Baumé)	272 (Saison)
Max Total Reviews	1,581 (Boulevard)	3,109 (Gary Danko)	839 (Quince)	1,511 (French Laundry)

In the year leading up to a restaurant gaining Michelin stars, we see a (potentially) statistically significant change in elite reviews



Two sample T-test confirms: % change in elite reviews matters

Two Sample T-Test: YOY % Change in Elite Reviews

(Sample 1: Restaurants Gaining Stars, Sample 2: No Change in # Stars)

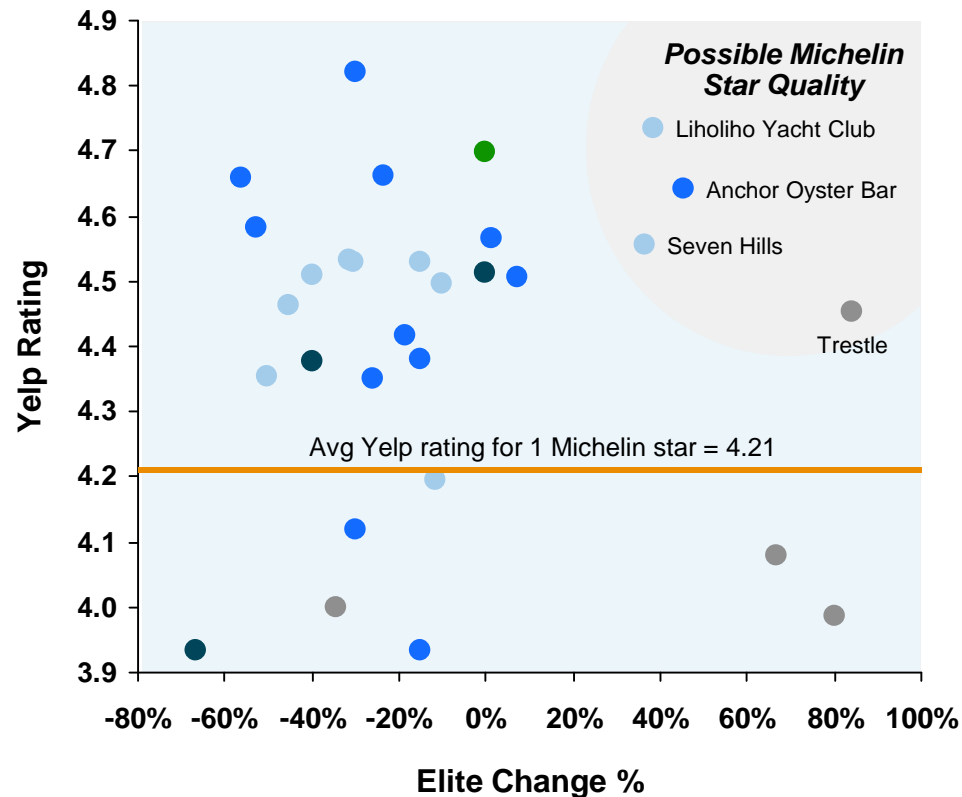
```
1 t.test(t.gain.elite_ann_chg, t.none.elite_ann_chg, alternative = "greater")
2
3 Welch Two Sample t-test
4
5 data: t.gain.elite_ann_chg and t.none.elite_ann_chg
6 t = 2.6138, df = 48.33, p-value = 0.005952
7 alternative hypothesis: true difference in means is greater than 0
8 95 percent confidence interval:
9  0.3561711      Inf
10 sample estimates:
11 mean of x mean of y
12 1.3172140 0.3234178
```

Prior to annual Michelin announcement we see an increase in elite reviews of 132% on average for restaurants gaining a star vs 32% for those maintaining stars

Plotting % change in elite reviews vs Yelp ratings highlights a few strong contenders for a Michelin star

Non-Michelin Star Restaurants Ranked

(YOY % Change in Elite Reviews vs Yelp Rating)



Discussion

- Best contender for a 2017 Michelin star appears to be **Trestle** based on its ramp in elite reviews
- Also meets the typical profile with a rating over 4.4 and 300+ total reviews
- 3 other contenders might qualify, though their percent change in elite reviews is not as convincing

Restaurants Gaining Michelin Star Word Cloud



Trestle Word Cloud



Agenda

- Context and Overview
- Analysis and Findings
- **Next Steps**

It would be interesting to scrape Michelin data from other cities and understand how Yelp profile and metrics vary

Next Steps

- With more time, additional areas of interest include:
 - NYC and Chicago Michelin restaurants (do same patterns apply?)
 - Years prior to 2010
 - Building more formal model (e.g., logistic regression, random forest)
- Questions?