yelp Ratings & Restaurant Attributes in Orange County

Team 5B: Chris Gu, Jason Jia, Srita Kothuri, Boram Gaudet

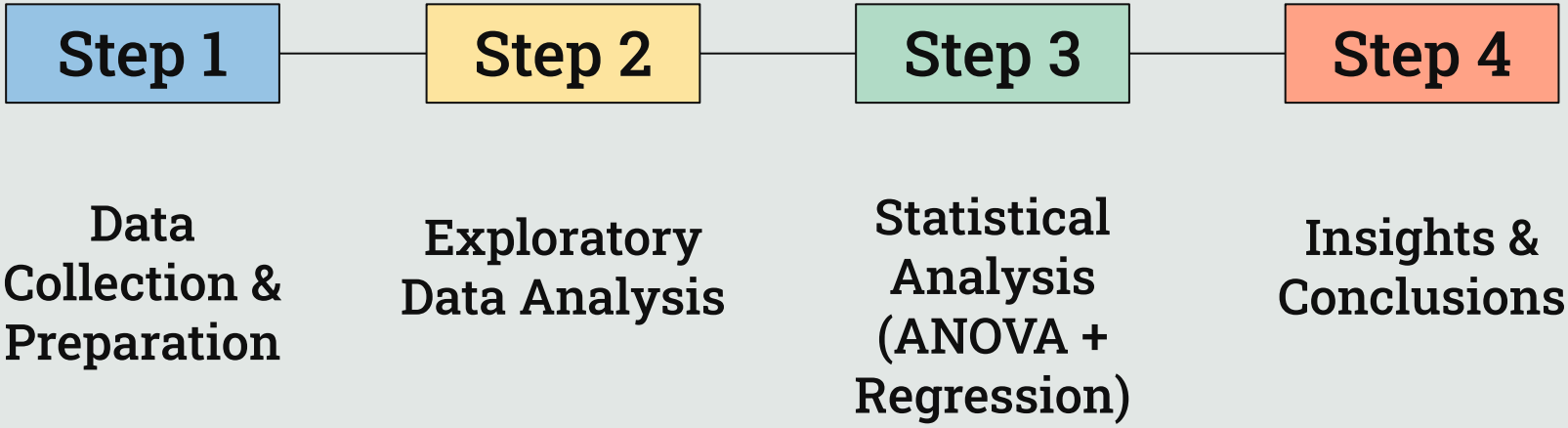
Introduction

- Orange County, CA → diverse, high-quality food scene
- Yelp is a major platform for ratings & review insights
- Key attributes: food quality, ambience, service
- Goal: understand what drives Yelp ratings across OC restaurants

Research Questions

- Which restaurant attributes influence Yelp ratings the most?
- How do these factors differ across cuisines types?
- Do customers prioritize different attributes across cuisines?

Workflow Overview

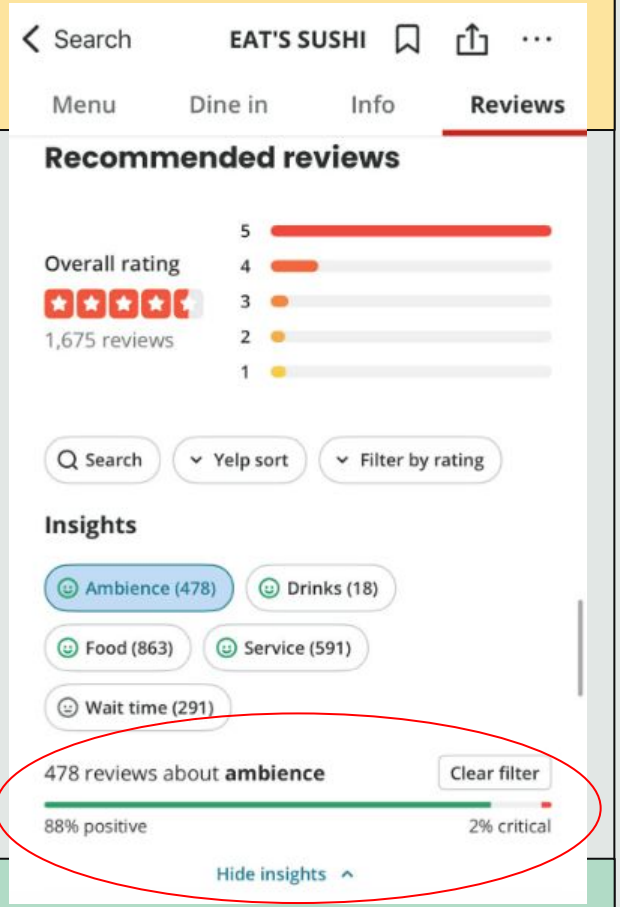


01

Data Collection & Preparation

Data Collection & Preparation

- Selected 10 Yelp restaurants per cuisine (Vietnamese, Korean, Japanese, Mexican, Italian)
  - 50 restaurants total
- Collected basic Yelp details: city, price range, review count, overall rating
- Pulled attribute data from Yelp's Review Insights (mentions within positive reviews):
  - Ambience
  - Food quality
  - Service
- Compiled the dataset into Excel, then loaded into Python for analysis



Dataset Snapshot

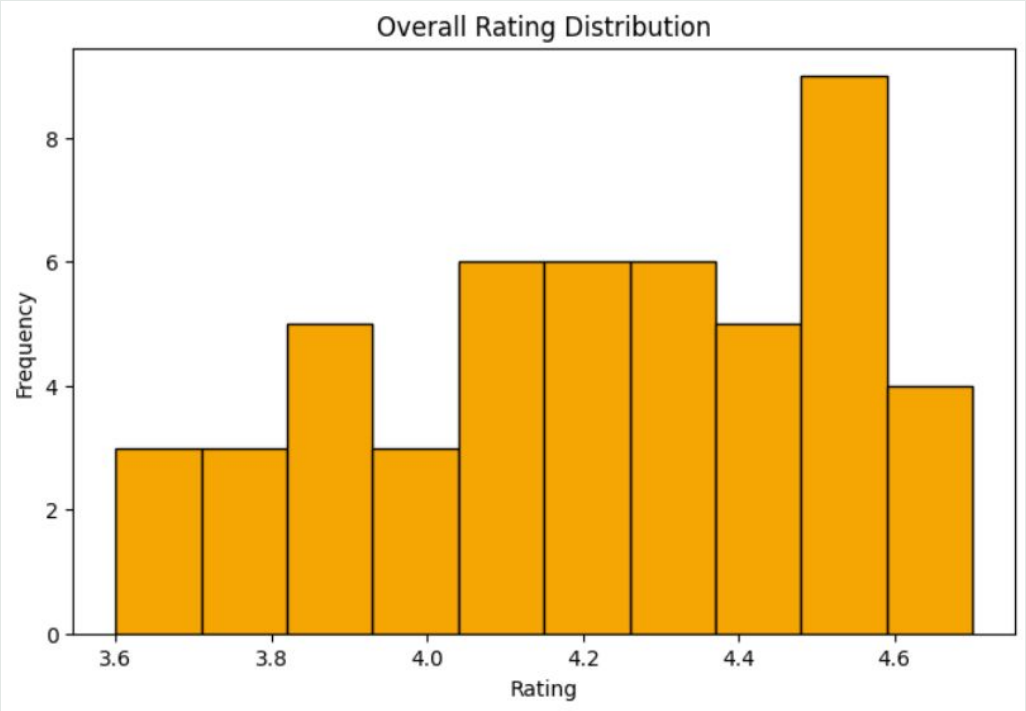
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	index	name	cuisine	location	price	overall_num_review	overall_rating	ambience_num_review	ambience_rating	food_num_review	food_rating	service_num_review	service_rating
2	0	VOX Kitchen	Vietnamese	Fountain Valley	$$	7964	4.5	846	0.81	1677	0.84	1194	0.83
3	1	NEP Café	Vietnamese	Fountain Valley	$$	4537	4.5	1035	0.86	1870	0.83	1314	0.83
4	2	Brodard Restaurant	Vietnamese	Fountain Valley	$$	3924	3.8	423	0.68	908	0.78	583	0.63
5	3	Pho 79 Restaurant	Vietnamese	Garden Grove	$$	3769	4.2	199	0.56	581	0.79	381	0.48
6	4	Oc & Lau Restaurant	Vietnamese	Graden Grove	$$	3273	4	267	0.57	560	0.8	363	0.71
7	5	Sup Noodle Bar	Vietnamese	Irvine	$$	3074	4.3	350	0.7	947	0.76	662	0.68
8	6	Nguyen's Kitchen	Vietnamese	Orange	$$	2714	3.9	99	0.52	242	0.67	132	0.63
9	7	Nep Café	Vietnamese	Irvine	$$	1759	4.5	949	0.82	1531	0.8	1111	0.81
10	8	Pho Ba Co	Vietnamese	Irvine	$$	1752	3.9	51	0.83	123	0.74	98	0.82
11	9	Au Lac	Vietnamese	Fountain Valley	$$	1743	4.4	51	0.82	111	0.8	84	0.72
12	0	Bakjeong	Korean	Irvine	$$$	5472	4.5	462	0.8	1248	0.86	1360	0.91
13	1	BCD Tofu House	Korean	Irvine	$$	3519	3.9	167	0.67	428	0.83	319	0.65
14	2	All That Barbecue	Korean	Irvine	$$$	2933	3.8	118	0.44	219	0.71	210	0.51
15	3	Mo Ran Gak Restaurant	Korean	Garden Grove	$$	2636	4.5	165	0.71	497	0.89	337	0.8

- ambience_rating = Proportion of positive reviews mentioning restaurant's ambience
- food_rating = Proportion of positive reviews mentioning restaurant's food quality
- service_rating = Proportion of positive reviews mentioning restaurant's service

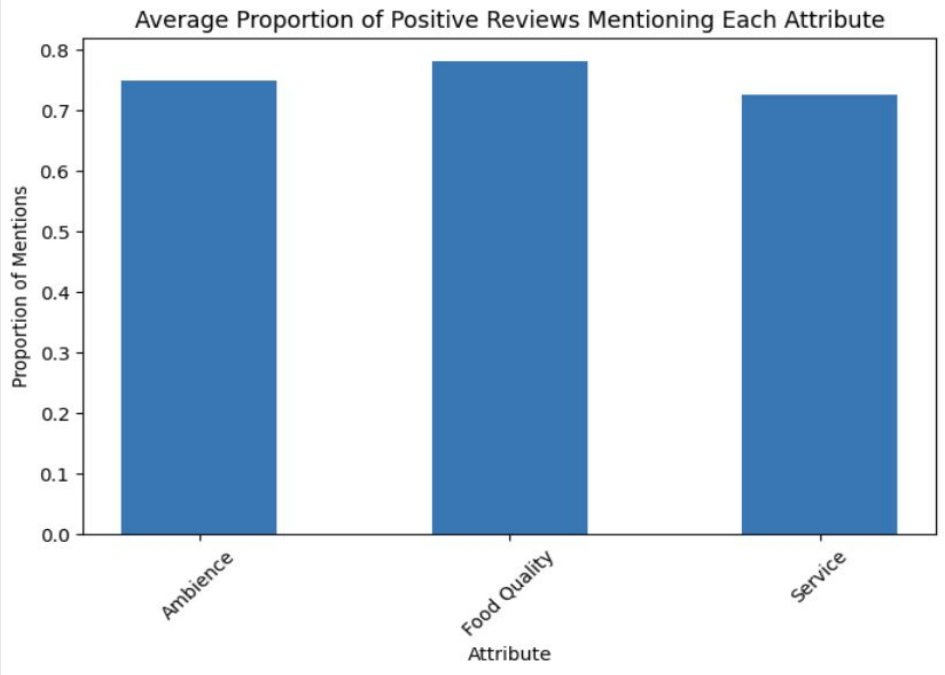
02

Exploratory Data Analysis

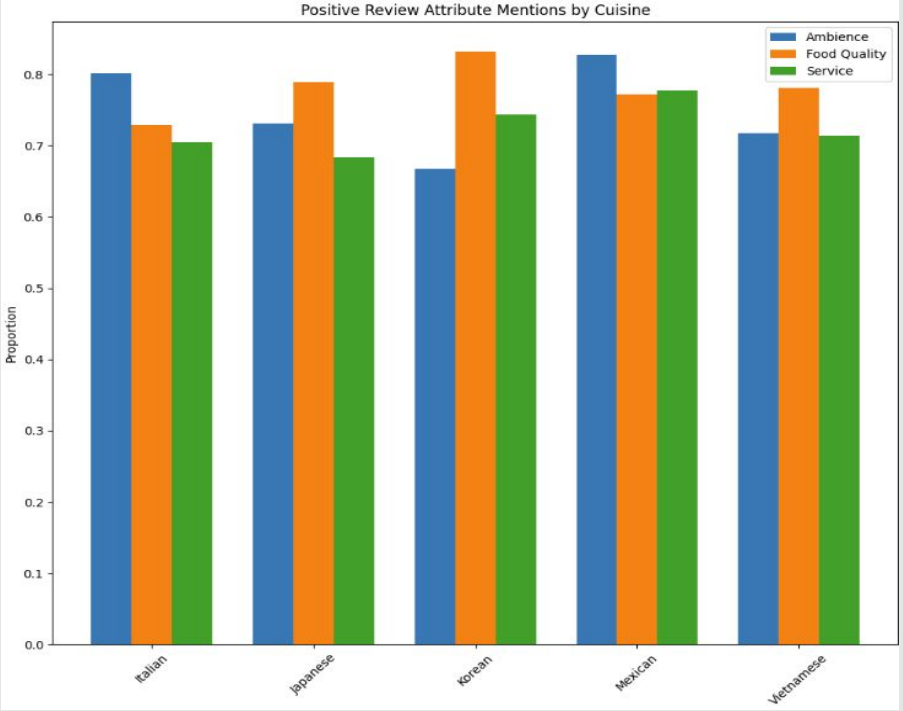
Overall Restaurant Rating Distribution



Attribute Mention Proportions (Overall)



Positive Review Attribute Mention by Cuisine



We will break each of these attributes down in the following slides

03

Statistical Analysis

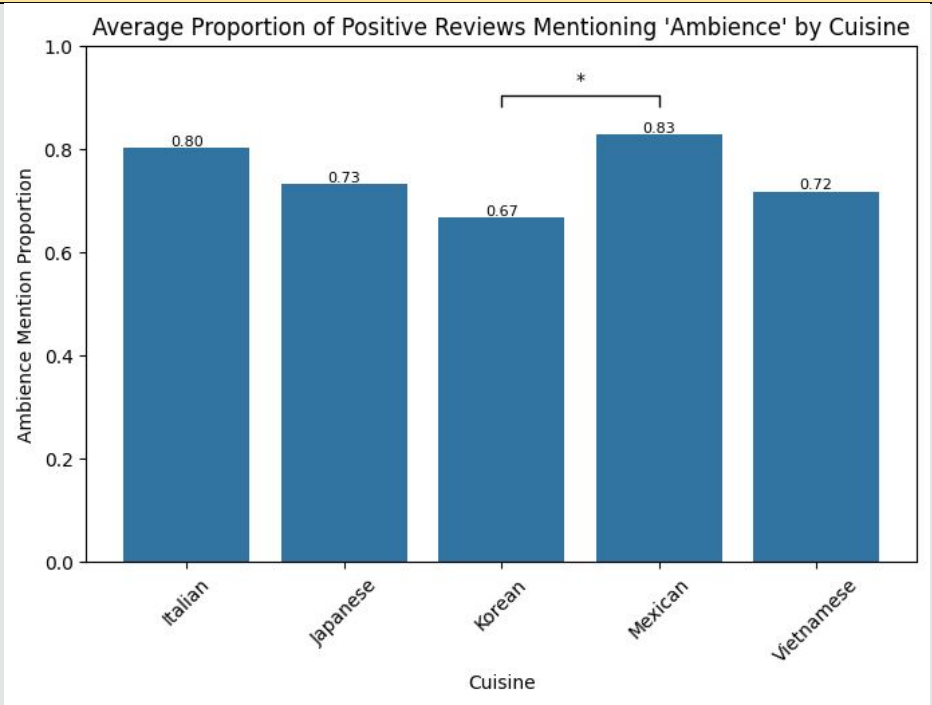
Statistical Analysis: ANOVA

- ANOVA approach
  - ANOVA run on each attribute
  - Two-tailed test
  - Purpose: Assess if average proportion of attribute mentions differed **significantly** between cuisines (pairwise)

- Assumptions & Prep
  - Normality checked with Shapiro Wilks Test
  - Ambience attribute was not normally distributed
    - Applied arcsine square root transformation

- Why ANOVA?
  - Identify attribute differences across cuisines
  - Highlights what customers notice most per cuisine

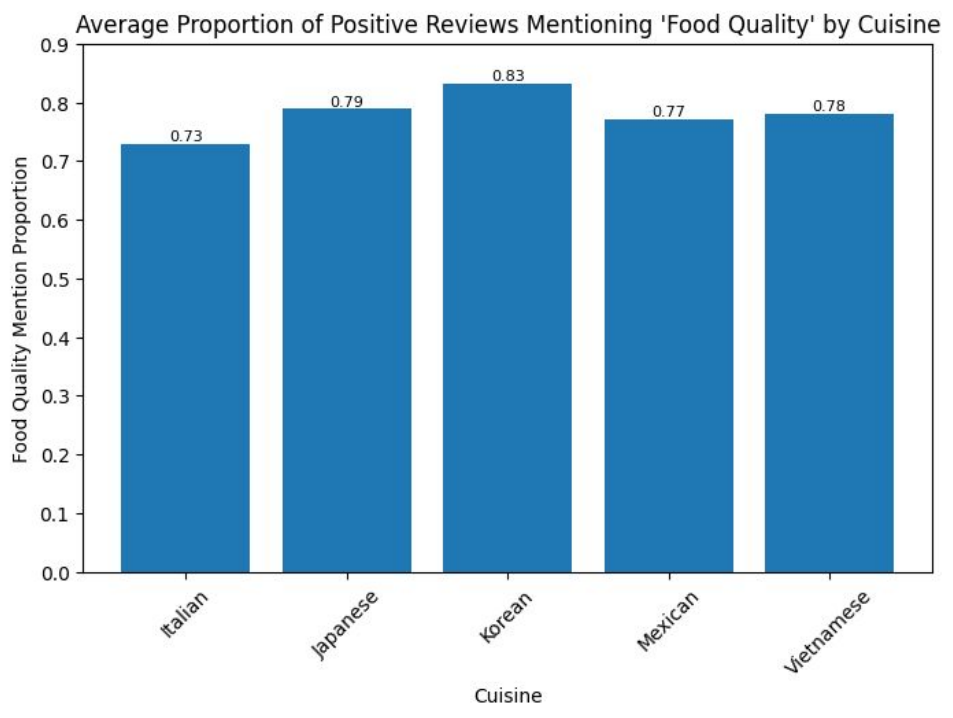
ANOVA: Ambience Rating



- **Significant difference** in proportion of reviews mentioning ambience by cuisine
  - $F(2,50) = 3.429$, $p = 0.015$
- Post Hoc (Tukey): Significant difference only between Korean vs. Mexican
  - $p = 0.021$

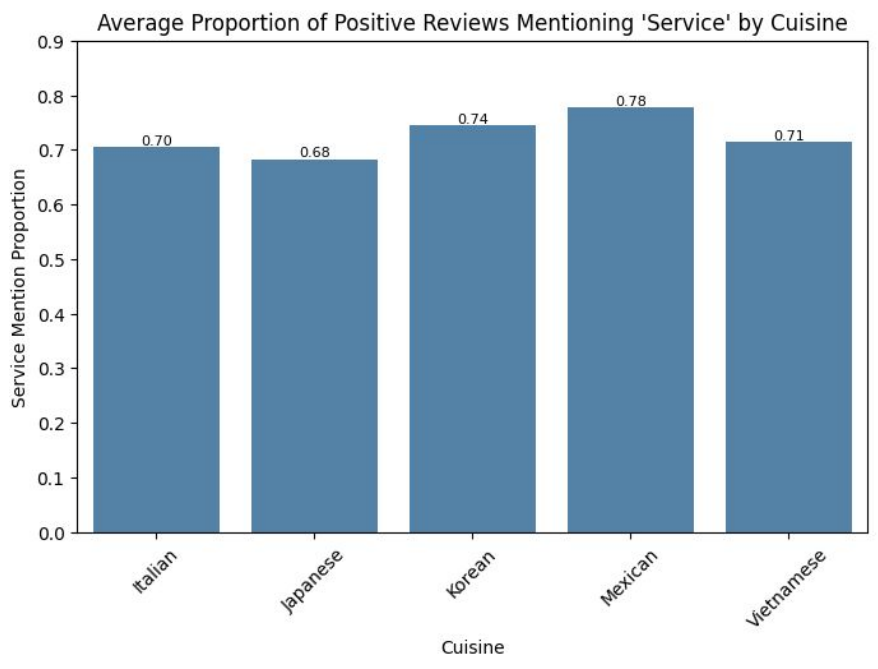
Note: this is the transformed data

ANOVA: Food Quality



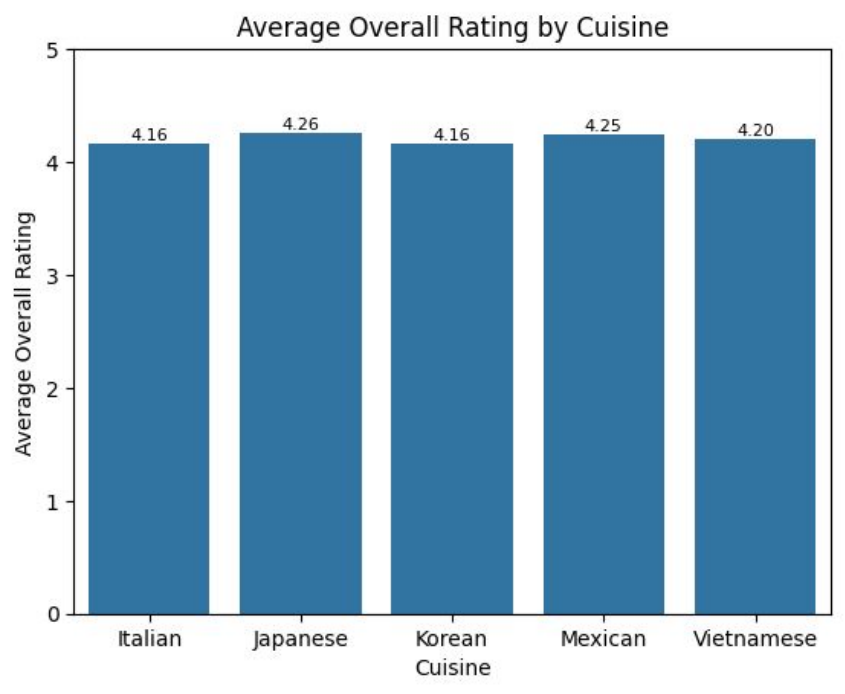
- **No significant difference** in proportion of reviews mentioning food quality by cuisine
  - $F(2,50) = 2.27$, $p = 0.076$

ANOVA: Service Rating



- **No significant difference** in proportion of reviews mentioning service by cuisine
  - $F(2,50) = 0.74$, $p = 0.569$

ANOVA: Overall Rating By Cuisine



- **No significant difference** in overall ratings between each cuisine
  - $F(2,50) = 0.26$, $p = 0.90$

Statistical Analysis: Regression

- Goal: Understand which attribute explains the most variance in overall rating of a cuisine

Regression: Each Cuisine

Italian

Ambience:
$r^2 = 0.45$*

Food Quality:
$r^2 = 0.76$**

Service:
$r^2 = 0.48$*

Japanese

Ambience:
$r^2 = 0.18$

Food Quality:
$r^2 = 0.83$***

Service:
$r^2 = 0.44$*

Korean

Ambience:
$r^2 = 0.50$*

Food Quality:
$r^2 = 0.26$

Service:
$r^2 = 0.21$

Vietnamese

Ambience:
$r^2 = 0.35$

Food Quality:
$r^2 = 0.45$*

Service:
$r^2 = 0.18$

Mexican

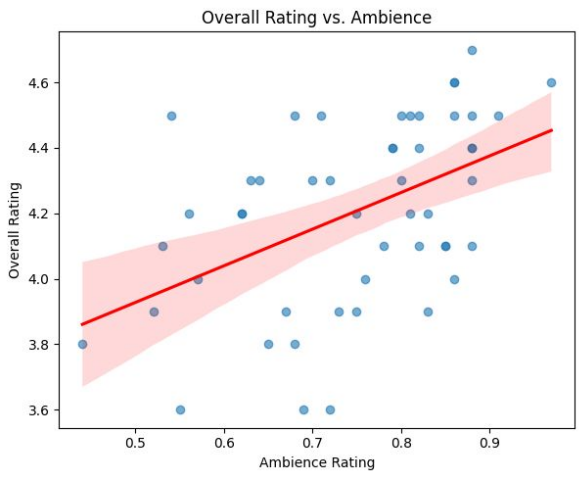
Ambience:
$r^2 = 0.14$

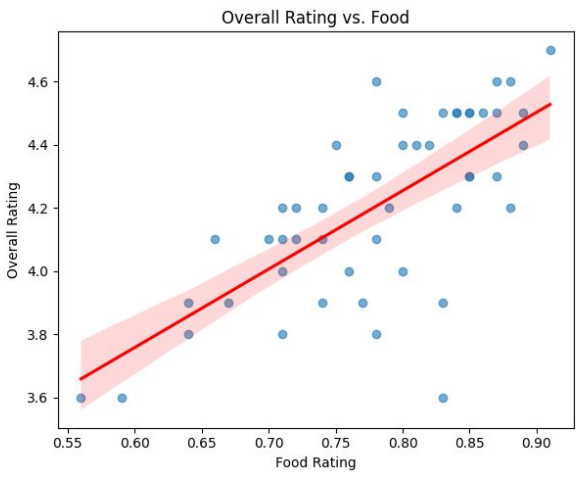
Food Quality:
$r^2 = 0.85$***

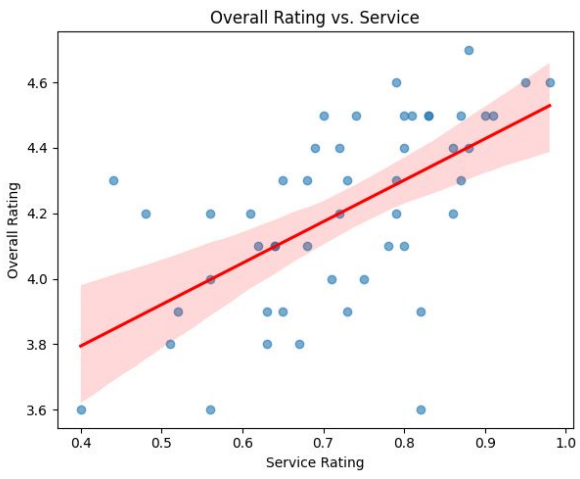
Service:
$r^2 = 0.72$**

*indicates p-value < 0.05 ; ** indicates p-value < 0.01; *** indicates p-value < 0.001

Regression: Overall







$r^2 = 0.23^{***}$

$r^2 = 0.50^{***}$

$r^2 = 0.35^{***}$

*indicates p-value < 0.05 ; ** indicates p-value < 0.01; *** indicates p-value < 0.001

04

Insights & Conclusions

Insights & Conclusions

1) No difference in average overall ratings across cuisines

2) Significant higher proportion of ambience mentions in **Mexican** restaurants vs. **Korean** restaurants
    a) Suggests that **ambience** is sought after in Mexican restaurants

3) Food Quality explains most variation in overall rating (4/5 cuisines)

Limitations

1) Limited sample size and reliance on summary-level Yelp data

2) Sample restricted to Orange County restaurants
   a) Potential self-selection bias

3) Focused on only three main attributes (food, service, ambience)
   a) Many other factors may influence restaurant experience

Thank You