

# INGREDIENTS FOR A SUCCESSFUL RESTAURANT



Predicting restaurant success and identifying good business practices

---

Mai Ueno

# Can machine learning predict restaurant success?

## Clients

I want to open a new restaurant

I want to find what makes restaurants successful



## Challenges In The Industry



1. **50+%** of restaurants close before their three year anniversary



2. Starting a restaurant is **very expensive**



3. Owners and investors often have to heavily rely on their anecdotal experience, **not on data**, for business improvement.

## Request



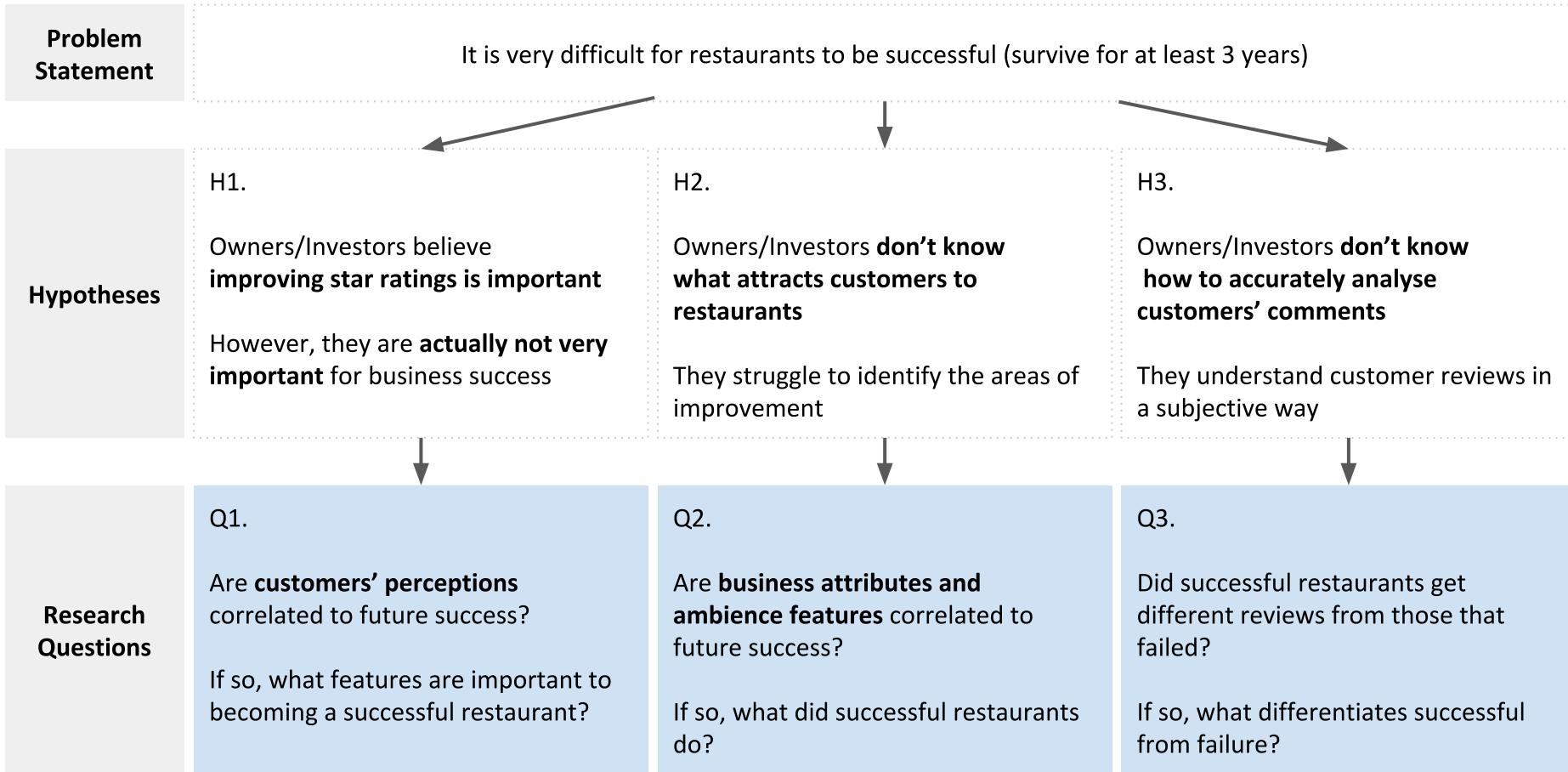
"Can machine learning predict restaurant's success?"

"Can it determine key elements of success?"

"Can it identify good business practices?"

"Can it help restaurants understand what customers think about their business?"

# Determine the ingredients of a successful restaurant



# Let's find the answers

---

## 1. Create a dataset to answer the research questions

This research focused on understanding **the high level competitive landscape**, then dug into the local trends per US state

## 2. Clean up the dataset - keep relevant data and remove the unnecessary

Left with 60K business information, 4M reviews and other relevant information after data cleaning

## 3. Put the data into a model

Selected **classification models** to predict business success

## 4. Check model performance scores

Model performance was evaluated based on **train, test, mean cross validation, and precision score**

## 5. If the model performed fairly, analyse results and find insights

Analysed data by coefficient values obtained **from the best performing classification models** and **by NLP analysis**

# The secret of success might be hidden in Yelp.com

## Yelp Dataset

- Yelp offers data from 2005 to 2018
- Downloaded 5 JSON files which contain:
  - Information on **1.6M** users
  - Information on **200K** businesses
    - Reviews (**7+M**)
    - Check-ins and tips (**3+M**)
    - 200+ business attributes

The screenshot shows the Yelp search interface with a red box highlighting the 'Top 10 Restaurants in New York, NY, United States' section. Below it, a red arrow points to a detailed view of a restaurant, specifically '1. Amélie', showing its address, reviews, and a map.

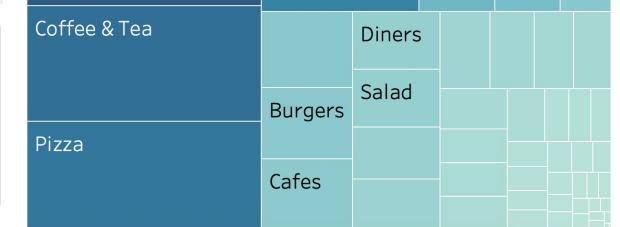
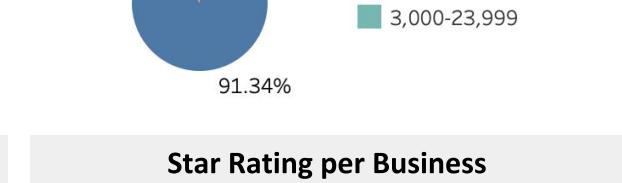
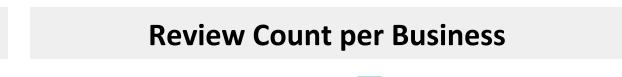
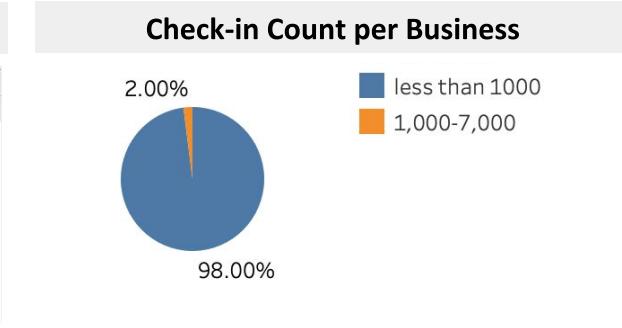
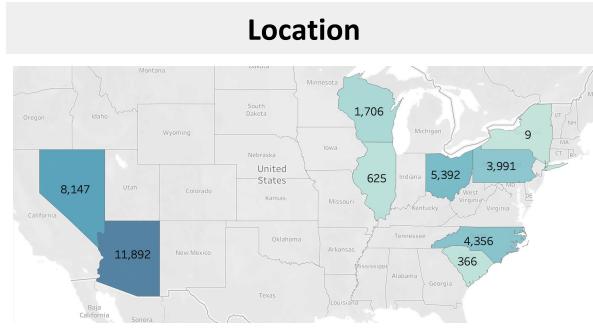
## Datasets created for building models

1. **Customer perception dataset (4M rows x 10+ columns)**
  - Star ratings per review
  - Sentiment scores per review
  - Reviews per restaurant
2. **Business attribute and ambience data (60K rows x 250+ columns)**
  - Restaurant and food store information only
  - Extracted from websites

The screenshot displays a 'Top 10 Restaurants in New York, NY, United States' page. A red box highlights the main table, and a red arrow points to a specific row for '1. Amélie'. The right side of the screen shows a sidebar with 'More Features' for filtering by various attributes like General Features, Alcohol, and Meals Served.

- Generated features
  - **Estimated operating year**
  - Total check-in count per restaurant
  - Total review count per restaurant
  - Category and food type

# Discover what insights can be distilled from the data



The background image shows a modern restaurant or cafe interior. It features a long wooden bar counter on the left where a staff member is working. The seating area consists of several round tables with wooden chairs. Large, black, industrial-style pendant lights hang from the ceiling. The ceiling itself is made of a perforated metal grid. In the background, there's a staircase and a sign that says "TOILET".

Question 1.

# Are customers' perceptions correlated to future success?

If so, what features are important to becoming a successful restaurant?

# Text features failed to predict future restaurant success

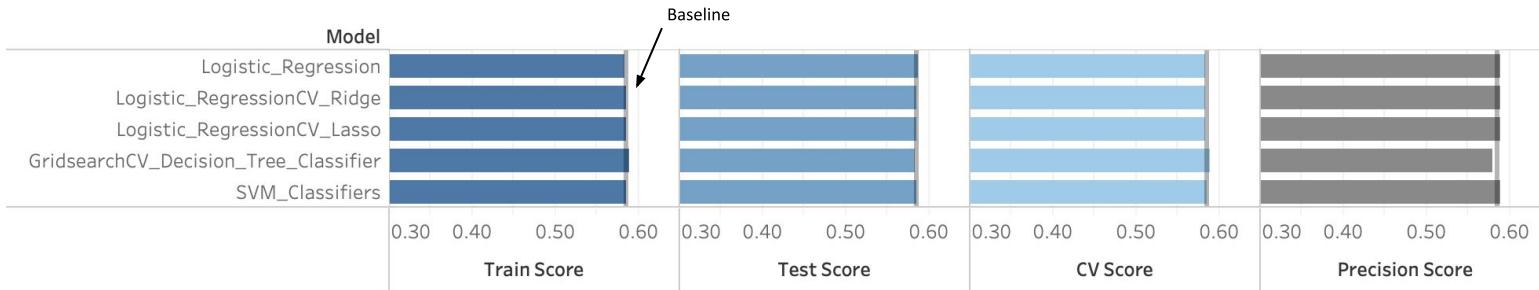
## Data

- Target variables: Successful (1), Failed (0)
- Predictors:
  - **Star rating**
  - **Positive/negative/compound sentiment score**

Star Category	Vader Result	Text
positive		Damn spicy. Damn greasy. Damn cheap. Damn tasty.
	negative	Big ass burger is no joke. Sometimes service is slowwww.
		I love having 100 channels. I hate not being able to see them because of a bad signal. I also hate dodging taxis..
		You expect horrible, overpriced food and awful service at any airport restaurant. Here, the food and service is ..
		So Good , OMG very fresh.. Was a little worried about the place had a friend say it was horrible but not sure wh..
		Double double animal style with a strawberry shake. Life doesn't get much better than this.

- Customers have different definitions per star rating
- Customers wrote negative reviews but gave 4 stars on the restaurant

- Customer perception data **did not** contribute to business prediction\*



## Models & Results

### Note:

- Baseline: "If I did not have any models, this is my best guess."
- Train score: "What it is vs what it thought it was"
- Test score: "How will my model perform on the data that it has not seen?"
- Cross-validation score: "How consistent does my models perform?"
- Precision score: "How precise is the classifier when predicting positive instances?"

\*Several previous research confirm the same results

The background image shows a modern restaurant or cafe interior. The ceiling is made of dark metal beams with several large, black, industrial-style pendant lights hanging down. The floor is made of light-colored concrete tiles. There are several wooden tables and chairs. In the foreground, a person is sitting at a table, looking at a laptop. In the background, there are more tables, some with people sitting at them, and a bar area with a chalkboard menu. The overall atmosphere is casual and modern.

Question 2.

## Are business attributes and ambience features correlated to future success?

If so, what did successful restaurants do?

# Business features are strongly correlated to restaurant performance

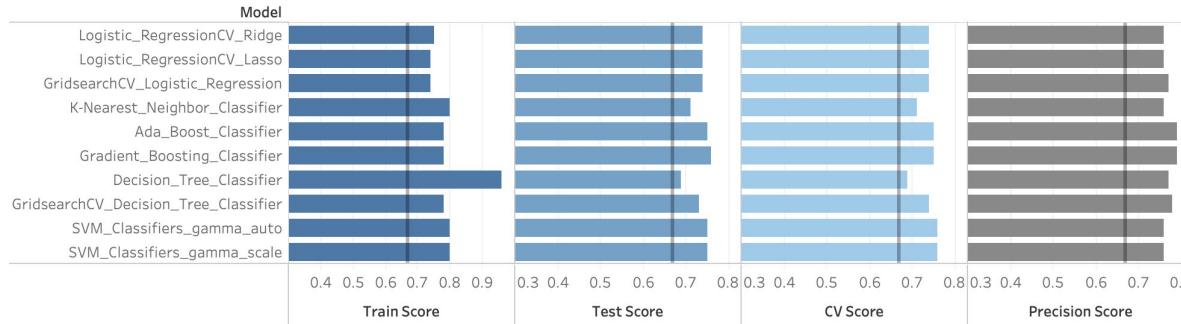
## Data

- Target variables: Successful (1), Failed (0)
- Predictors: **200+ columns of business attributes and ambience features**

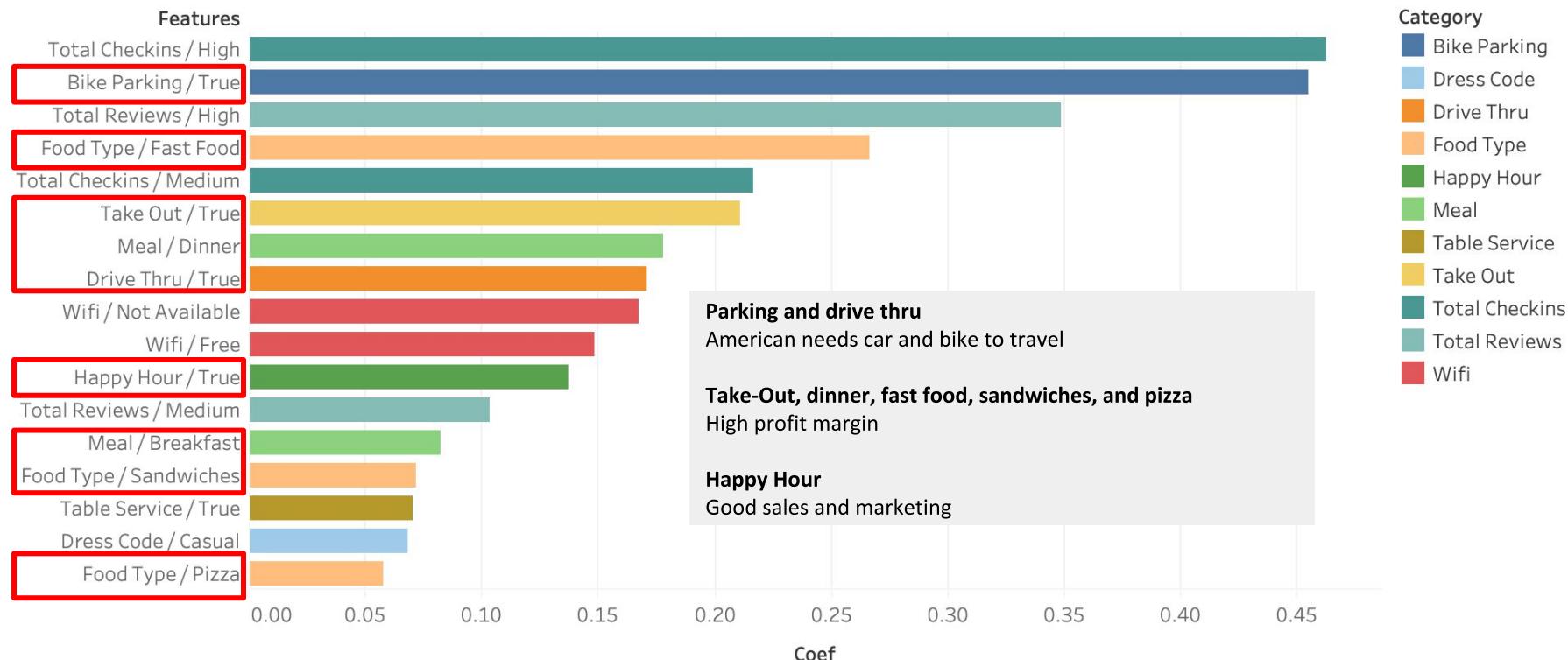


## Models & Results

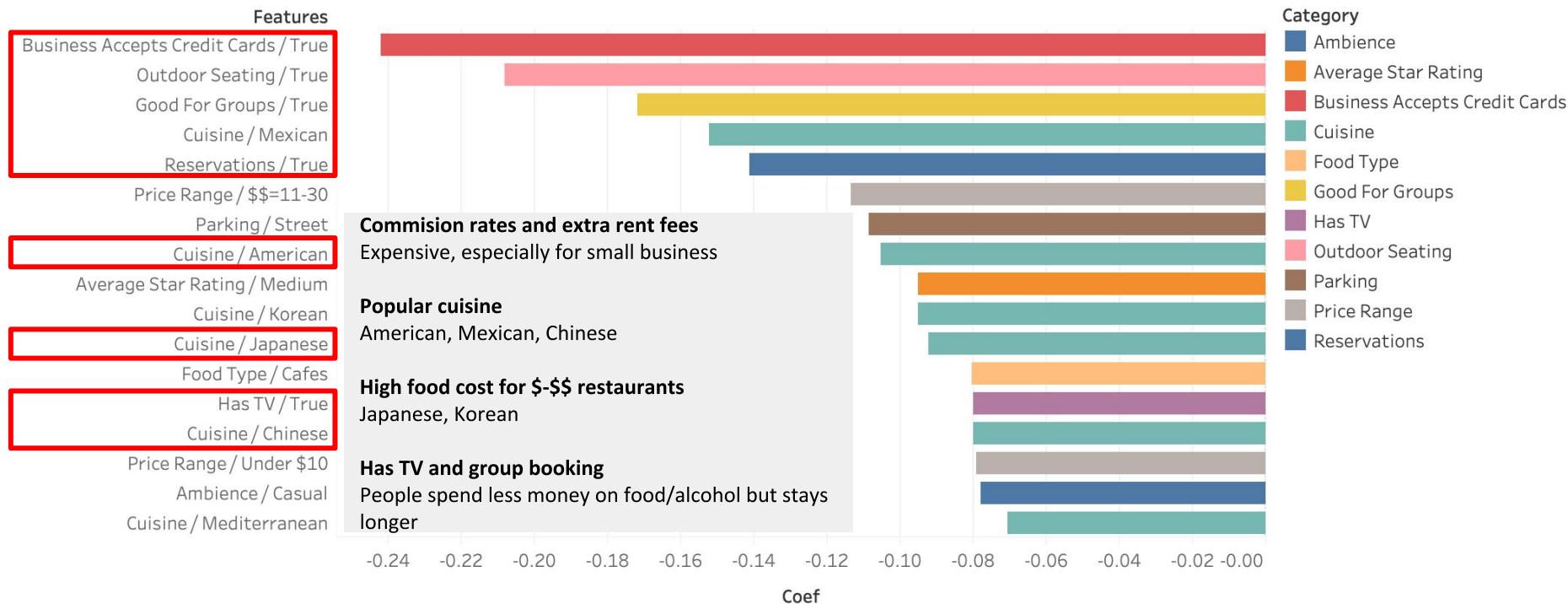
- Classification models can predict business success
- Results of Logistic Regression with a Lasso penalty were used for further analysis



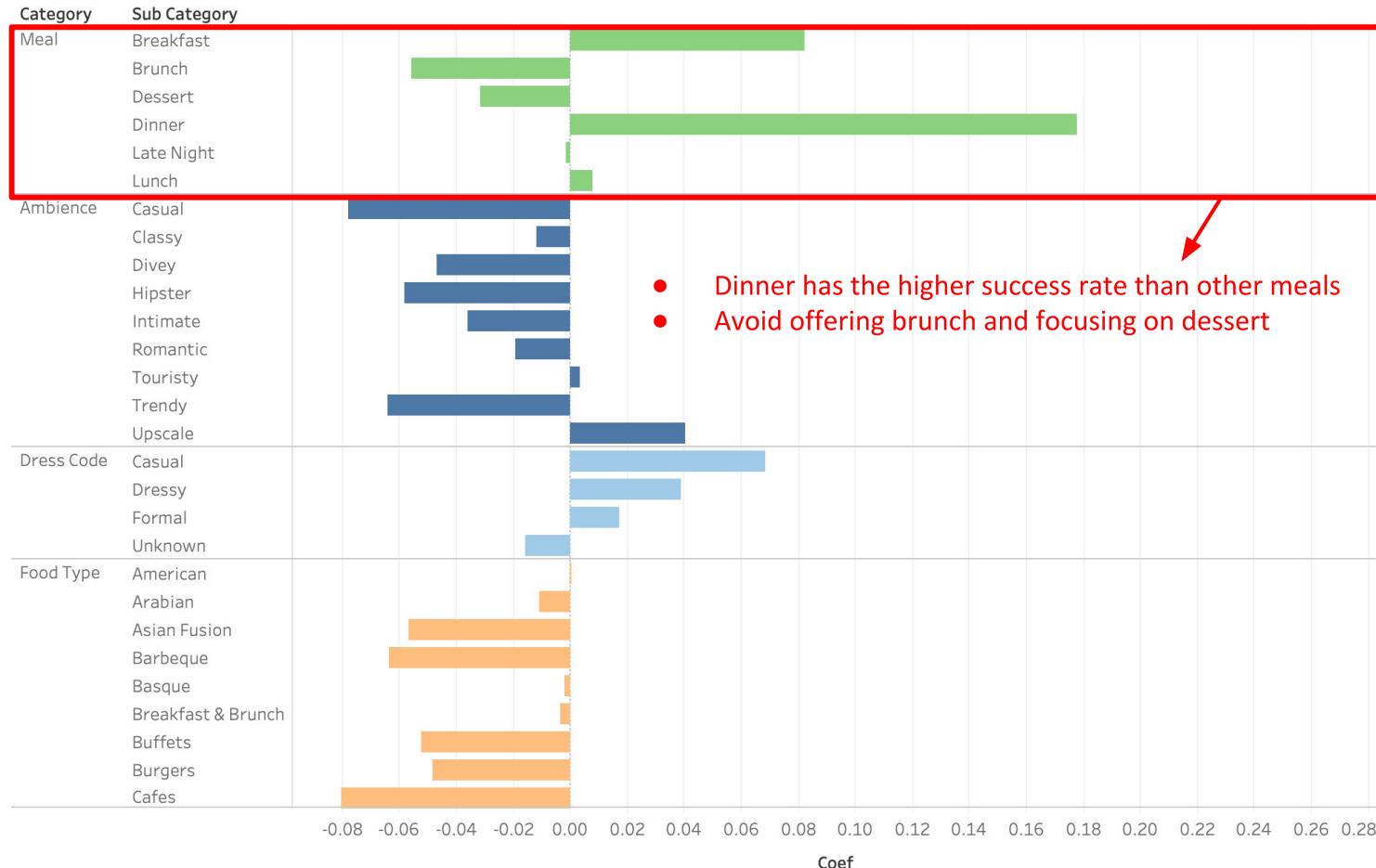
# Physical attributes have more of an impact than atmospheric variables



# Lower profit margins could be the key determinant of business failure



# Discovering areas of improvement at a glance



The background image shows a modern restaurant or cafe interior. The ceiling is made of dark wood with several large, black, industrial-style pendant lights hanging down. The floor is made of light-colored concrete tiles. There are wooden tables and chairs arranged throughout the space. In the background, there's a bar area with a chalkboard menu and some plants. The overall atmosphere is casual and contemporary.

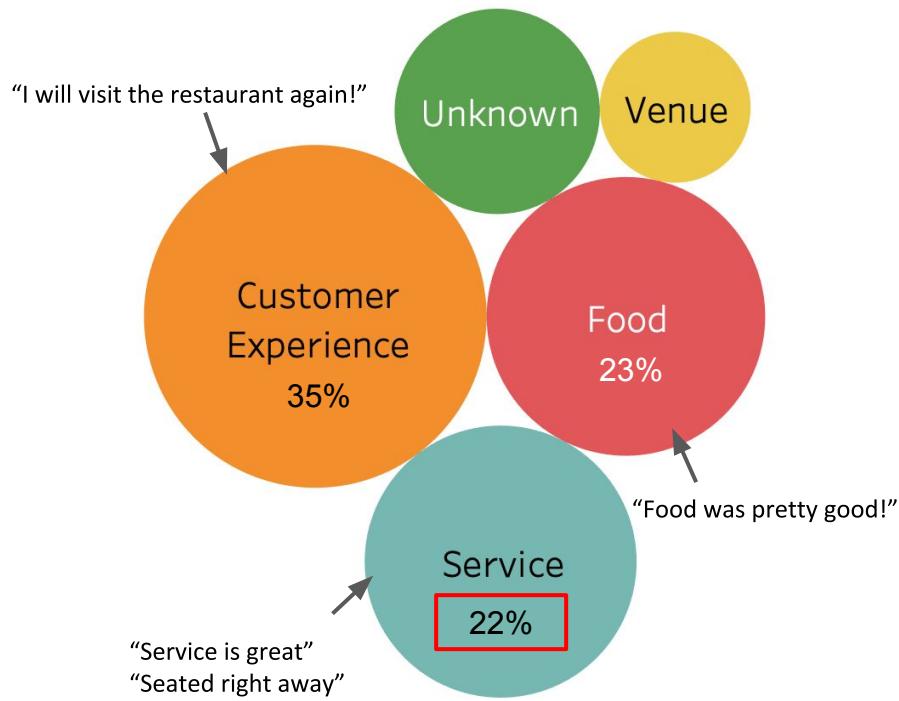
Question 3.

# Did successful restaurants get different reviews from those that failed?

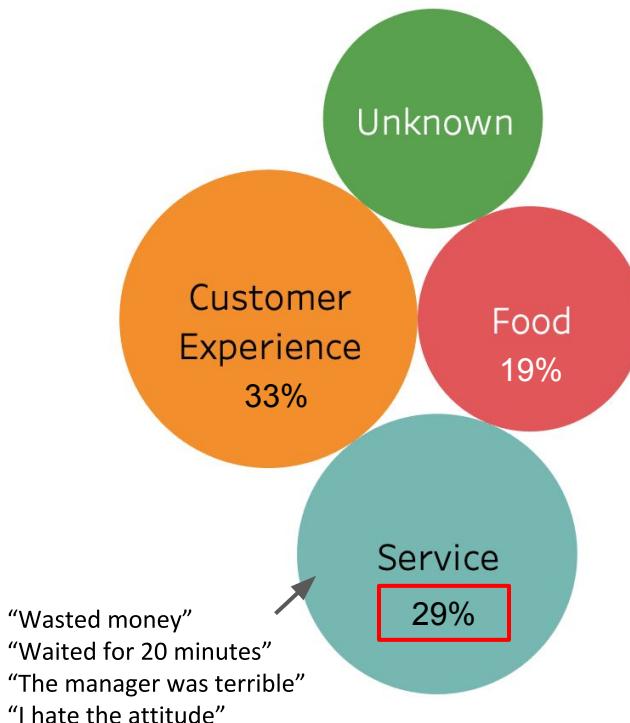
If so, what differentiates successful from failure?

# Happy customers talk about food. Unhappy customers complain about service quality

Positive reviews of successful restaurants



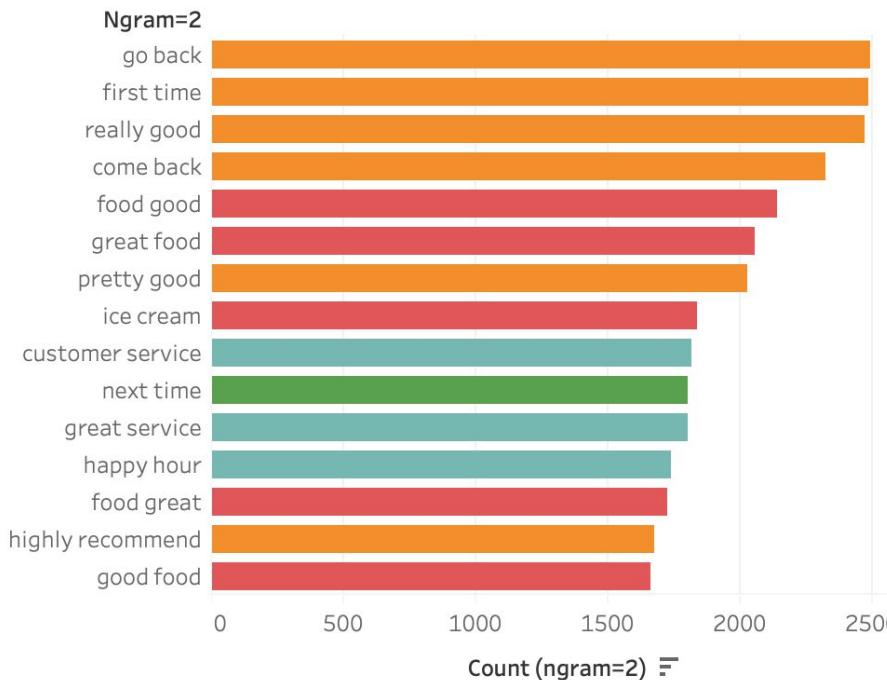
Negative Reviews of failed restaurants



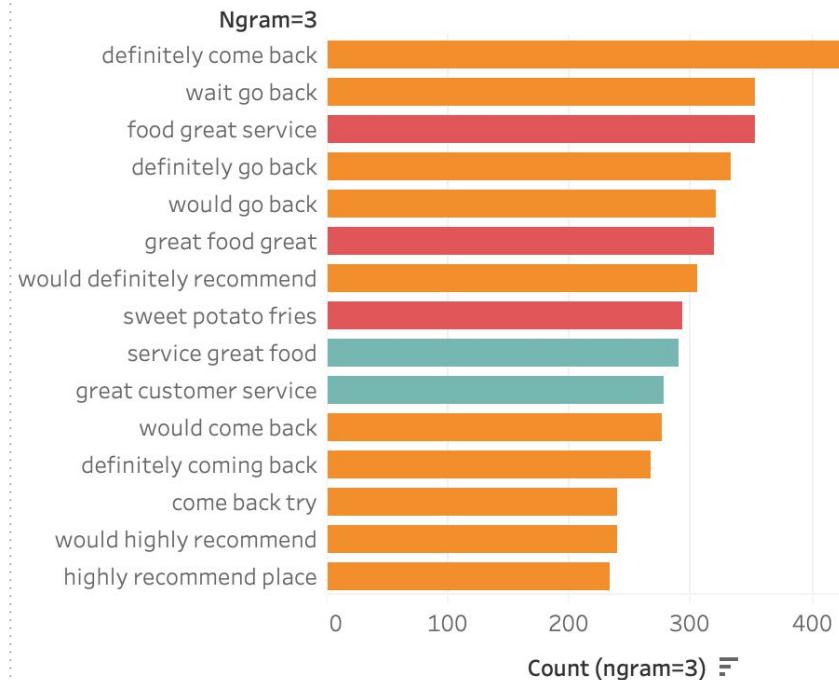
# Positive reviews of successful restaurants: “I WILL COME BACK!”

Customer Experience      Service  
Food                          Unknown

## Positive Reviews (ngram=2)



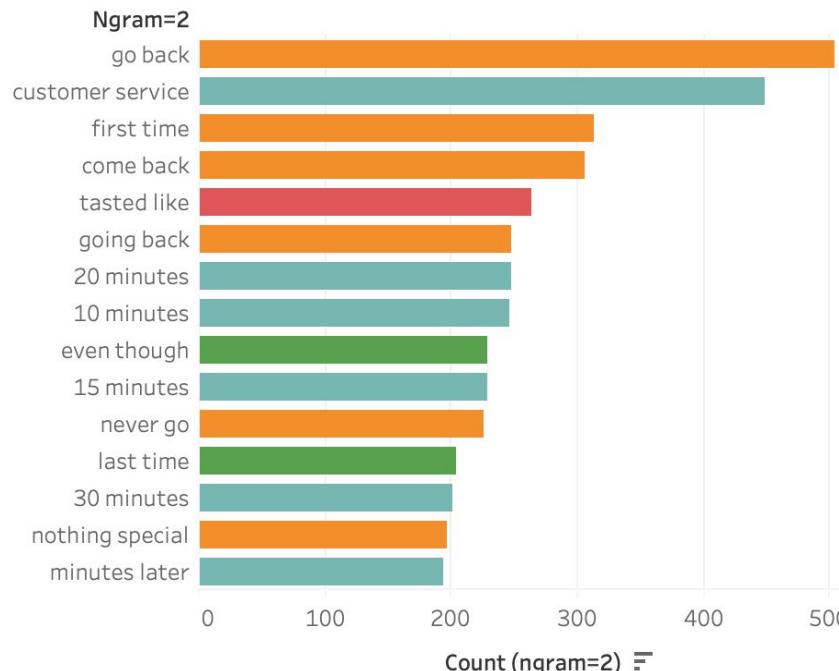
## Positive Reviews (ngram=3)



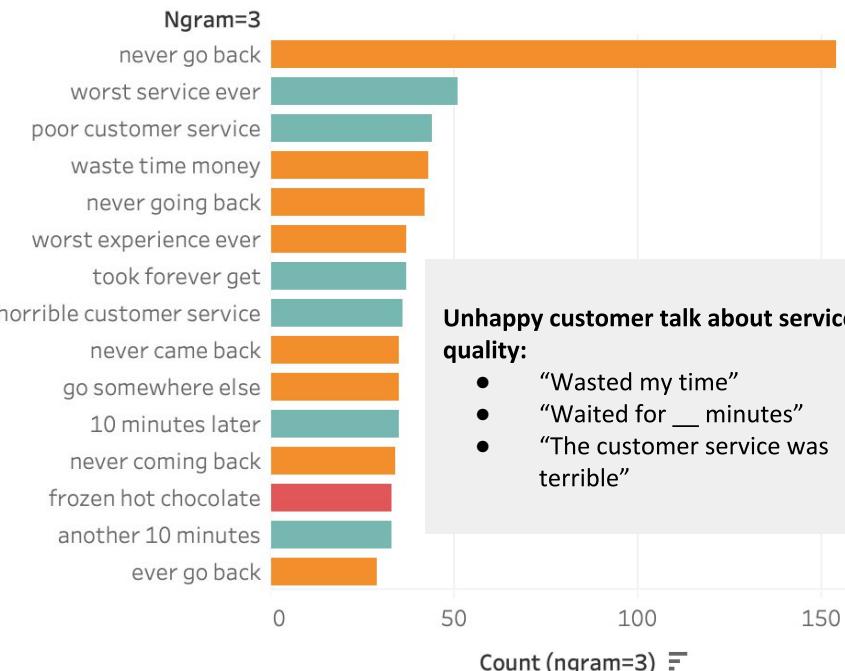
# Negative reviews of failed restaurants: “TERRIBLE SERVICE”



## Negative Reviews (ngram=2)



## Negative Reviews (ngram=3)



# Conclusion



# Suggestions for future business success



- Classification models managed to:
  - Predict future success with **75% accuracy** (vs. baseline 67%)
  - Identify:
    - Good business practices
    - Risks for business failure
- Based on the results of sentiment analysis, restaurants should focus on **more physical attributes than customer perception**
- Based on nationwide analysis, restaurants would do well by:
  - **Offering parking spaces**
  - **Focusing on high margin food**
  - **Avoiding commissions and unnecessary rent fees**
- Restaurants should regularly analyse customer comments by:
  - Checking **frequently used words in reviews**
  - **Avoiding subjective statements**

# Research limitations and future recommendation



- **Limitations**
  - Datasets were **curated by Yelp team**
    - The dataset only covers specific states in the US
    - There were many missing data for some states
  - **Models would be more accurate by using web scraped data from yelp.com** in the same column format as this project
  - Models would perform differently for restaurants outside the US
- **Future recommendation**
  - Obtain information from yelp.com by **web scraping**
  - **Include more columns** such as chain restaurants
  - Analyse the information per state or city to **see local trends**
  - Improve the feature generations