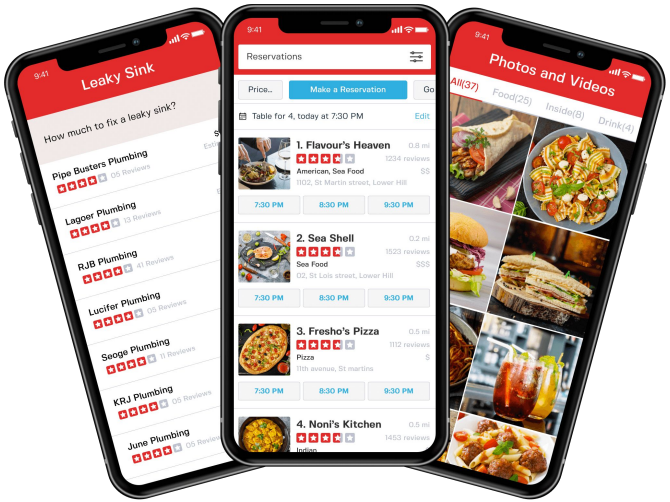
Data Science Applications for F&B Industry

Ruthvik Ravindra, Akshit Jain, Pranshu Diwan

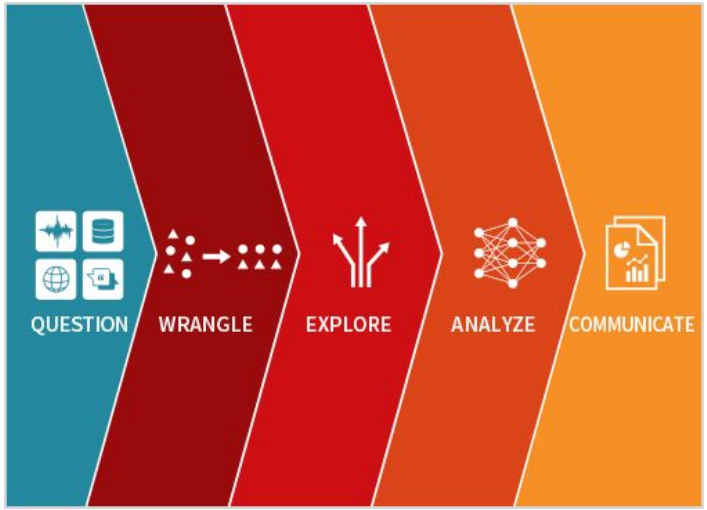
Introduction

- Yelp thrives on the numerous reviews that are left by users for local businesses and restaurants.
- Along with a ton of information about restaurants like operating hours, menus, pictures, addresses, etc.
- It is of considerable value to analyze this data and find out whether they help in directing the performance of a restaurant or whether restaurant performance is indeed dictated by other factors.

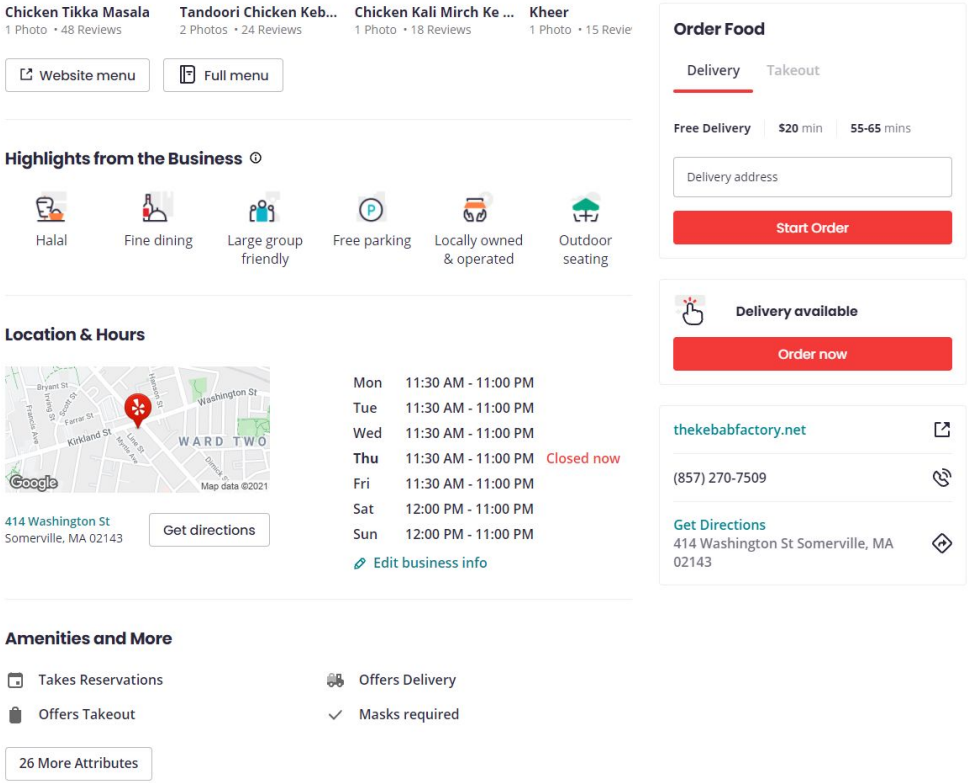


Objective

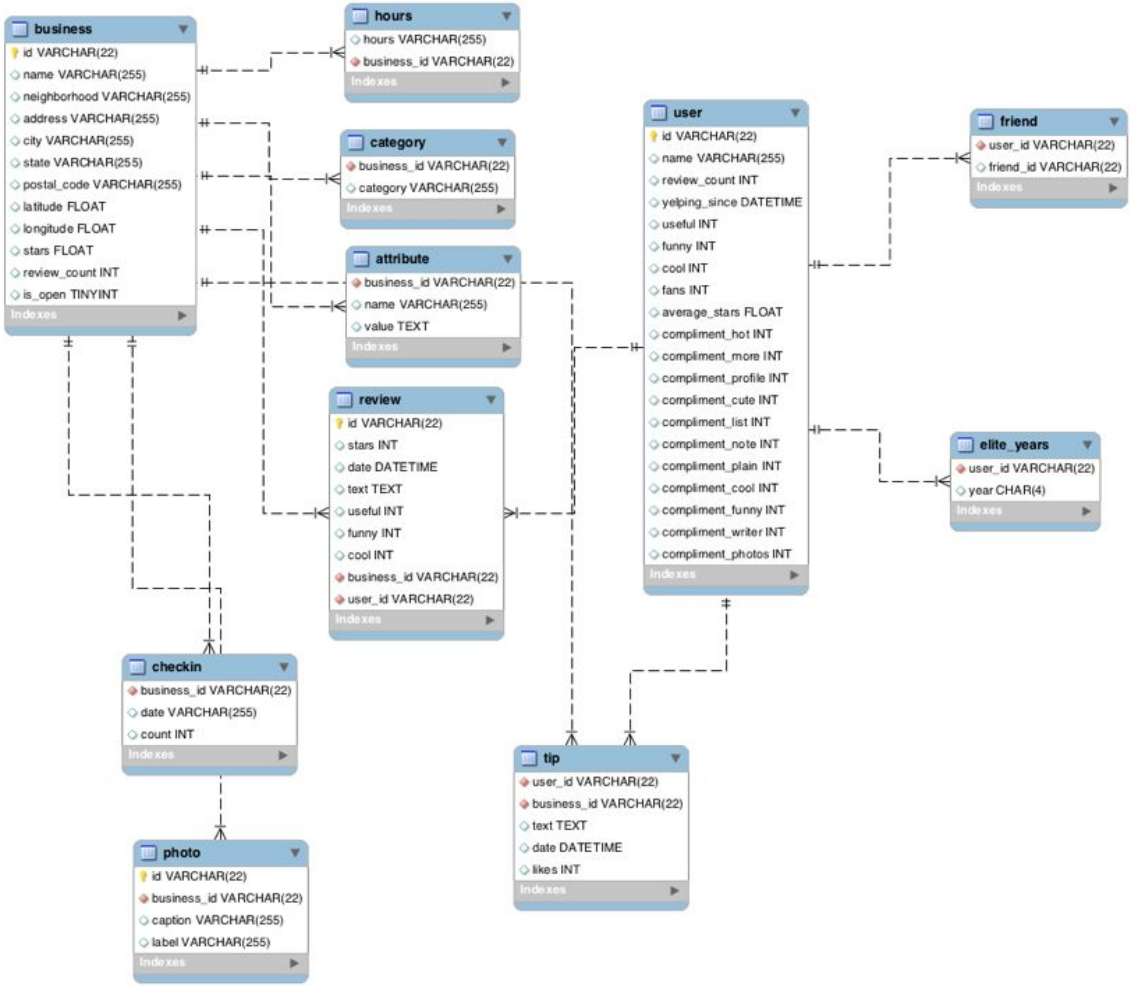
- Solve business problems using various data science techniques to help the F&B industry leverage their data.
- The project is divided into two phases:
  - Phase 1: Enable restaurants to gain more business and customers.
  - Phase 2: Enhance user experience while interacting with restaurants.



Dataset - Yelp Data



Dataset - UML


business
id VARCHAR(22)
name VARCHAR(255)
neighborhood VARCHAR(255)
address VARCHAR(255)
city VARCHAR(255)
state VARCHAR(255)
postal_code VARCHAR(255)
latitude FLOAT
longitude FLOAT
stars FLOAT
review_count INT
is_open TINYINT
Indexes
hours
hours VARCHAR(255)
business_id VARCHAR(22)
Indexes
category
business_id VARCHAR(22)
category VARCHAR(255)
Indexes
attribute
business_id VARCHAR(22)
name VARCHAR(255)
value TEXT
Indexes
review
id VARCHAR(22)
stars INT
date DATETIME
text TEXT
useful INT
funny INT
cool INT
business_id VARCHAR(22)
user_id VARCHAR(22)
Indexes
checkin
business_id VARCHAR(22)
date VARCHAR(255)
count INT
Indexes
photo
id VARCHAR(22)
business_id VARCHAR(22)
caption VARCHAR(255)
label VARCHAR(255)
Indexes
user
id VARCHAR(22)
name VARCHAR(255)
review_count INT
yelping_since DATETIME
useful INT
funny INT
cool INT
fans INT
average_stars FLOAT
compliment_hot INT
compliment_more INT
compliment_profile INT
compliment_cute INT
compliment_list INT
compliment_note INT
compliment_plain INT
compliment_cool INT
compliment_funny INT
compliment_writer INT
compliment_photos INT
Indexes
friend
user_id VARCHAR(22)
friend_id VARCHAR(22)
Indexes
elite_years
user_id VARCHAR(22)
year CHAR(4)
Indexes
tip
user_id VARCHAR(22)
business_id VARCHAR(22)
text TEXT
date DATETIME
likes INT
Indexes

Proposed Plan - Phase 1

Use Case 1

Predict the success of a restaurant based on features.

Models:

- Linear Regression
- Decision Tree
- Neural Networks

Impact:
Given business factors, Predicts the possible rating of the restaurant.

Use Case 2

Analyze reviews based on popular restaurant features.

Models:

- Topic Modeling
- Sentiment Analysis
- Fuzzy search

Impact:
Given a dish or service, restaurants can search how well it is doing.

Use Case 3

Forecasting the footfall for the next few days.

Models:

- Random Forests
- ARIMA
- LSTMs

Impact:
Restaurants can efficiently manage inventory or staff based on expected footfall.

Proposed Plan - Phase 2

In the second phase, we aim to deliver data science solutions to enhance user experience while interacting with restaurants

Proposed initiatives:

- Recommendation system to suggest users restaurants to visit and food items to try.
- Identify the food dish given a photo

Expected Timeline

Timeline for Phase 1	Business Case 1	Business Case 2	Business Case 3
	Predicting the success of a restaurant based on several factors.	Analyze reviews based on popular restaurant features.	Forecasting the footfall for the next few days.
Oct 5 (phase 1 pitches)	Data loading and analysis.	Data loading and analysis.	Data loading and analysis.
Oct 12 (phase 1 pitches)	Perform EDA, develop Regression models	Perform EDA, develop Topic Models, feature-based reviews using NLP.	Perform EDA, develop forecasting models like Random Forests, ARIMA, LSTMs, etc. Test the models.
Oct 19 (phase 1 updates)	Testing and optimizing regression models, visualizing and analyzing the results.	Annotating the results for accuracy of the model output. Productionizing the code for accepting any restaurant feature(s).	Optimize model results. Visualize the outputs and identify possible errors for certain time periods like festivals.
Oct 26 (phase 1 updates)	Writing phase 1 report and preparing for presentations.	Writing phase 1 report and preparing for presentations.	Re-train the model if needed. Writing phase 1 report and preparing for presentations.
Nov 2 (phase 1 presentations + phase 1 report due)	Finishing touches, reviewing the report and presentations.	Finishing touches, reviewing the report and presentations.	Finishing touches, reviewing the report and presentations.
Nov 9 (phase 1 presentations)	Presentations in class + report submission.	Presentations in class + report submission.	Presentations in class + report submission.

Q & A