



Capstone Project

Presentation by Hoda Shoghi

Predicting Restaurant Success

A DATA-DRIVEN APPROACH

- Restaurants are a vital part of the economy and cultural fabric.

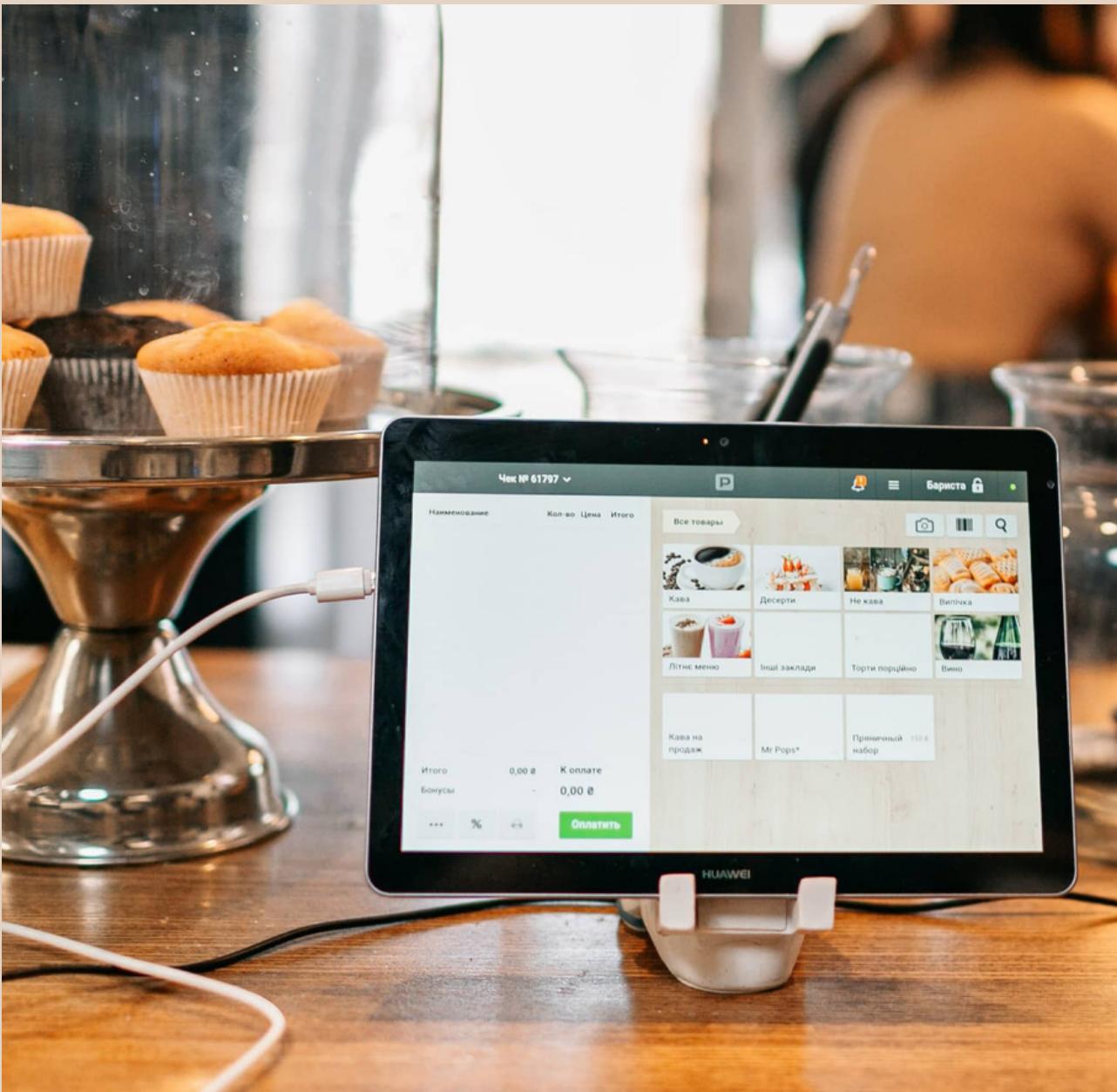




Problem

- Many restaurants fail due to **lack of market knowledge** and planning.
- With the ever-growing competition in the food industry, restaurant owners and investors face challenges in **understanding the factors** that contribute to the success of a restaurant.

Proposed Solution and Methodology



To build a **predictive model** that can accurately predict a restaurant's star rating by leveraging the vast amount of information in the dataset. This model can provide prospective Stakeholders with **insights into how certain factors might influence their new establishment's ratings.**



Yelp Open Dataset

An all-purpose dataset for learning



The Yelp dataset is a subset of our businesses, reviews, and user data for use in connection with academic research. Available as JSON files, use it to teach students about databases, to learn NLP, or for sample production data while you learn how to make mobile apps.

The Dataset



6,990,280 reviews



150,346 businesses



200,100 pictures



11 metropolitan areas

Dataset

I Yelp's business data as our primary dataset, specifically focusing on restaurants.

Potential Impact of the Solution

■ market dynamics

The insights gained from this project can be useful for investors, culinary entrepreneurs, and city planners to understand the restaurant market dynamics better.

■ Failure Reduction

The reduction of the failure rate among restaurants could be a significant outcome, promoting a more robust and vibrant food industry.

■ strategic decisions

Leveraging this data-driven model, restaurant owners and stakeholders can make strategic decisions backed by concrete insights, potentially increasing their chances of success.

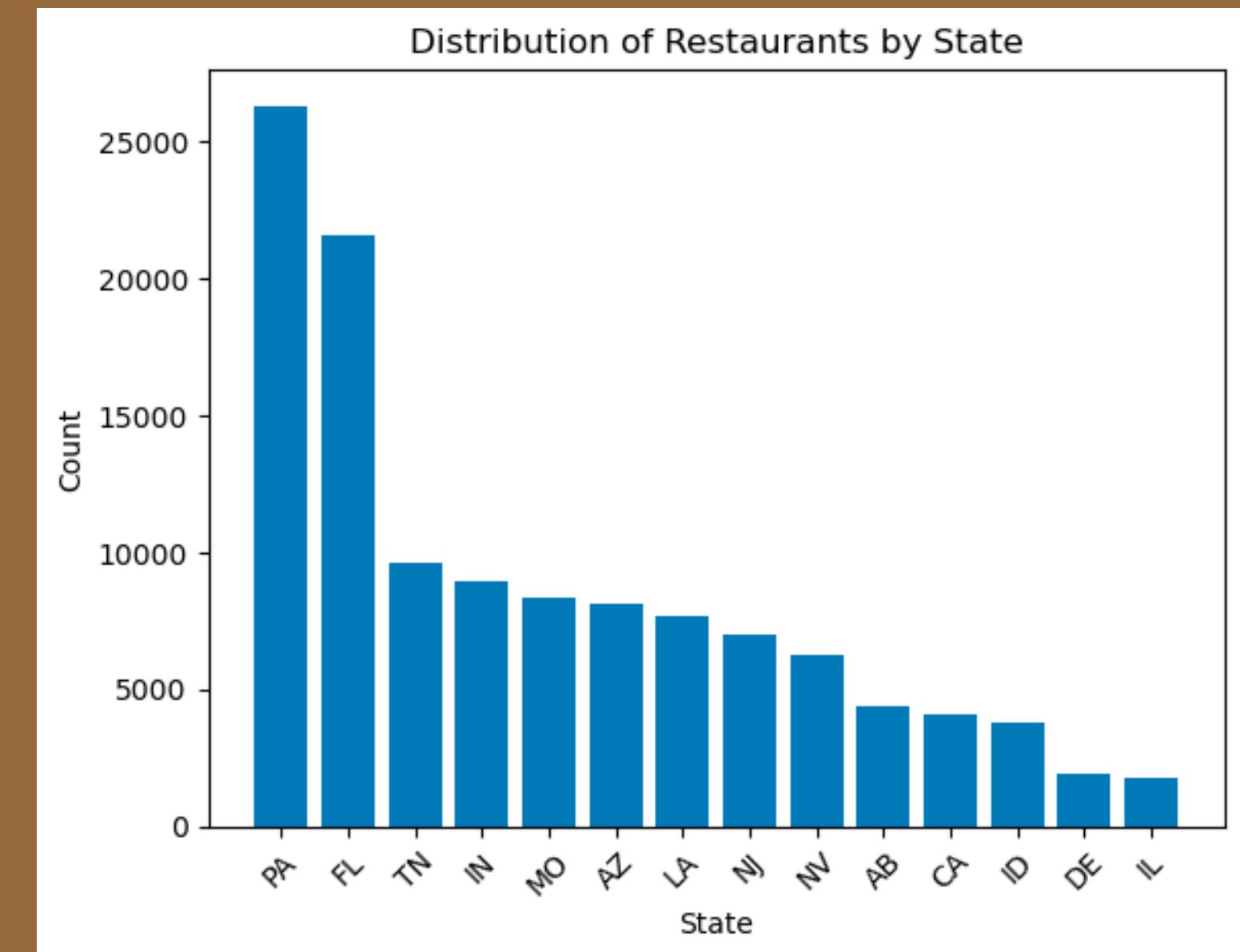
Initial Findings

COLUMN NAME	DESCRIPTION	DATA TYPE
business_id	Unique identifier for the business	string
name	Name of the business	string
address	Address where the business is located	string
city	City where the business is located	string
state	State where the business is located	string
postal_code	postal_code where the business is located	string
latitude	Geographical latitude of the business	float64
longitude	Geographical longitude of the business	float64
stars	Star rating of the business	float64
review_count	Number of reviews the business has received	int64
is_open	0 is closed and 1 is open	int64
attributes	different attribute like payment method, delivery options, etc	string
categories	Categories the business falls under	string
hours	Hours of operation	string

What kind of data is stored in each column?

- 1 **IS OPEN** This dataset has 150346 rows
is open
1 119698. 0 30648
- 2 **STATE**
- 3 **CATEGORIES**
- 4 **ATTRIBUTES**

States

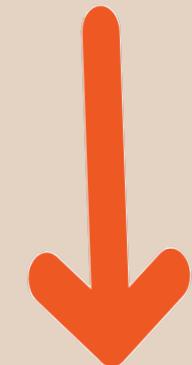


CLEANING CATEGORIES

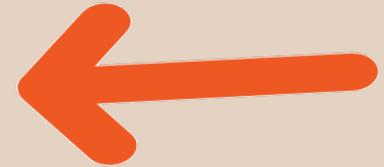
Beauty & Spas, Nail Salons	900
Nail Salons, Beauty & Spas	849
Restaurants, Pizza	668
Pizza, Restaurants	575
Restaurants, Chinese	537
sandwiches	6075
nightlife	5779
bars	5567
american (traditional)	5531
fast food	5516
pizza	5090
breakfast & brunch	4415
burgers	4275
american (new)	3629
mexican	3315
coffee & tea	2977



restaurants, pizza	668
pizza, restaurants	575
restaurants, chinese	537
chinese, restaurants	513
restaurants, mexican	447
mexican, restaurants	417
restaurants, fast food, burgers	147
italian, restaurants	132
fast food, restaurants, burgers	127



restaurants	34987
food	10859
sandwiches	6075
nightlife	5779
bars	5567
american (traditional)	5531
fast food	5516
pizza	5090
breakfast & brunch	4415
burgers	4275
american (new)	3629



Inside attribute Column

Dictionary

```
3  {'RestaurantsDelivery': 'False', 'OutdoorSeati...  
4  {'BusinessAcceptsCreditCards': 'True', 'Wheelc...  
15  {'RestaurantsReservations': 'True', 'Restauran...  
19  {'NoiseLevel': 'u'quiet", 'GoodForMeal': '{'d...
```

25

None

Name: attributes, dtype: object



Keys

```
{'BestNights', 'DogsAllowed', 'GoodForDancing', 'ByAppointmentOnly',  
'HappyHour', 'RestaurantsTableService', 'CoatCheck', 'BYOBCorkage',  
'RestaurantsCounterService', 'Alcohol', 'Open24Hours', 'OutdoorSeating', 'WiFi',  
'HairSpecializesIn', 'BikeParking', 'RestaurantsPriceRange2', 'RestaurantsTakeOut',  
'AcceptsInsurance', 'BusinessAcceptsBitcoin', 'BusinessParking',  
'DietaryRestrictions', 'RestaurantsAttire', 'Caters', 'RestaurantsDelivery', 'HasTV',  
'BusinessAcceptsCreditCards', 'WheelchairAccessible', 'BYOB', 'NoiseLevel',  
'DriveThru', 'RestaurantsReservations', 'Music', 'Smoking', 'GoodForKids',  
'GoodForMeal', 'RestaurantsGoodForGroups', 'Ambience', 'AgesAllowed',  
'Corkage'}
```

Nested Dictionaries in Attribute Columns !?

'BusinessParking'.

Keys. >>>>> 'validated', 'garage', 'lot', 'street', 'valet'



'business_id', 'name', 'address', 'city', 'state', 'postal_code',
'latitude', 'longitude', 'stars', 'review_count', 'is_open',
'attributes', 'categories', 'hours', 'HappyHour',
'RestaurantsReservations', 'DogsAllowed', 'OutdoorSeating',
'WheelchairAccessible', 'RestaurantsGoodForGroups', 'Open24Hours',
'BusinessParking', 'validated', 'garage', 'lot', 'street', 'valet'

Forging Ahead

- 1** Clean and preprocess the data further by handling any missing or anomalous values.
- 2** Conduct a more in-depth exploratory data analysis to identify patterns and relationships.
- 3** Develop different baseline models using suitable machine learning algorithms.
- 4** Evaluate and refine the model for better accuracy.



Thank
You