



Preliminary analysis

How various features of East Asian Restaurants in CA
influence their review stars?

Data preprocessing

Extract Process Combine Polish

1. Extract **East Asian Restaurants** (Keyword: Chinese, Korean, etc.) in CA
2. Include all original features (e.g. **business name**, **stars** and **attributes** like **TakeOut option**, **Parking availability**, etc).
3. Data size: 23538 * 94

1. Remove permanently closed restaurant
2. Extract object-like data (e.g. hour, parking, etc)
3. Initialize new features (daily & weekly working hour)

1. Combine **business data** with **review data** using 'business_id';
2. Apply NLP strategies (e.g. remove stop words, punctuations, symbols, etc)
3. Initialize sentiment (based on customer's review star)

1. Unify type in mixed-type feature

Data preprocessing

Review id	Business id	Category	text	comment star	Hour per week
Id of each review comment	Id of each restaurant	Category of each restaurant (Chinese, Korean, Japanese)	Comment given by customer	Star given by customer (1,2,3,4,5)	Open hours of each restaurant per week

Data preprocessing

Bike Parking	HasTV	TakeOut	Good For Kids	Outdoor Seating	WiFi
Is there a parking space? (True/False)	Is there a TV in the restaurant? (True/False)	Does the restaurant provide takeout service? (True/False)		Is the food suitable for children? (True/False)	Is there any outdoor seating? (True/False)	Is the WiFi free for customers? (True/False)

Word Cloud from review text

Positive sentiment



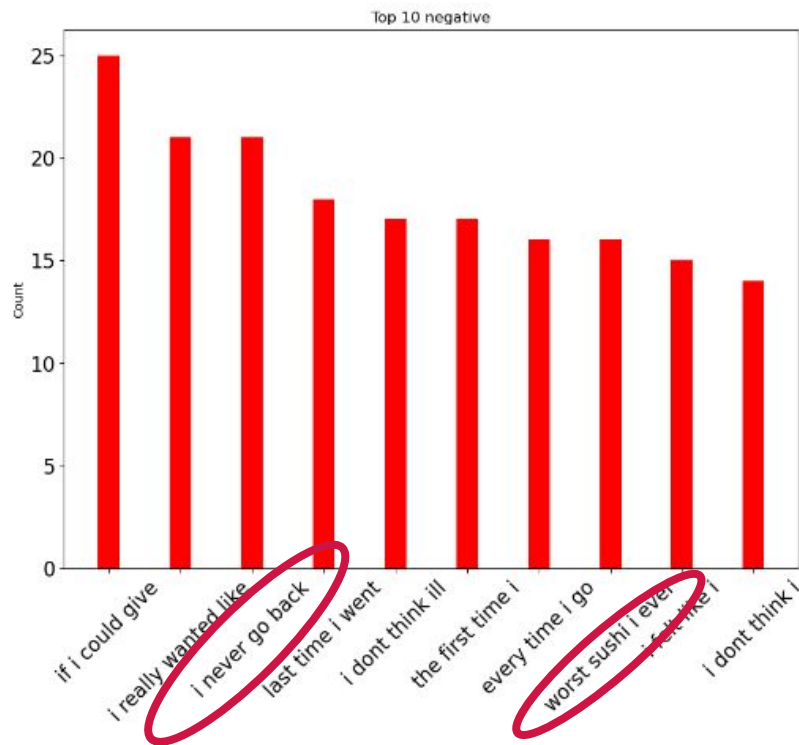
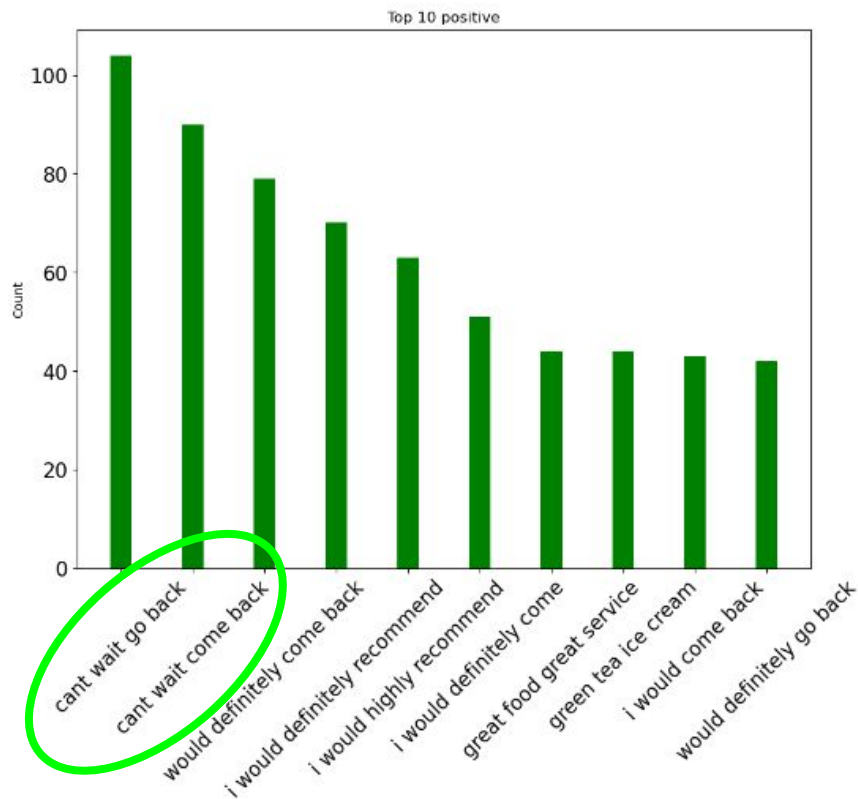
Neutral sentiment



Negative sentiment

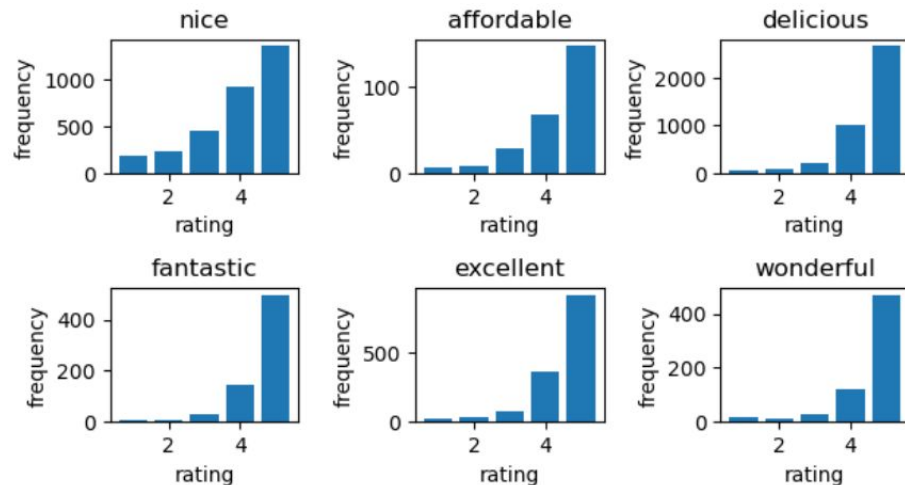


4-gram review text

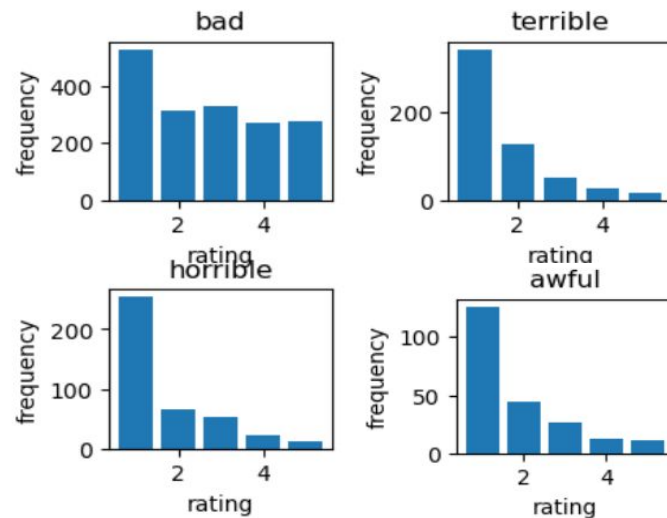


Information from review text

Positive words in review text

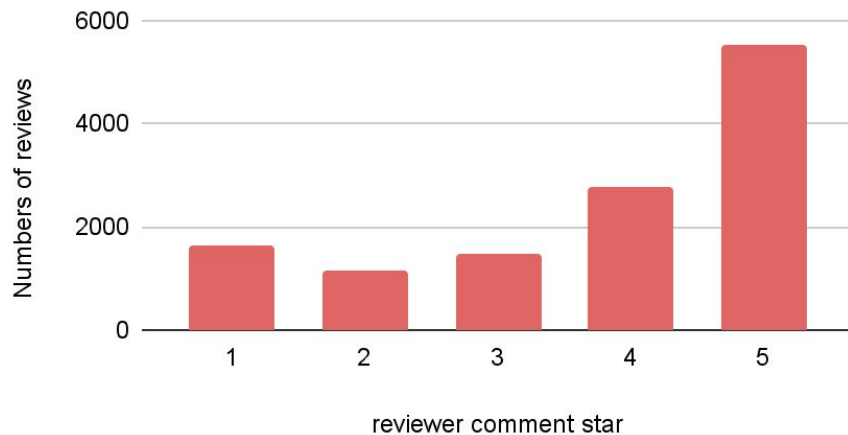


Negative words in review text

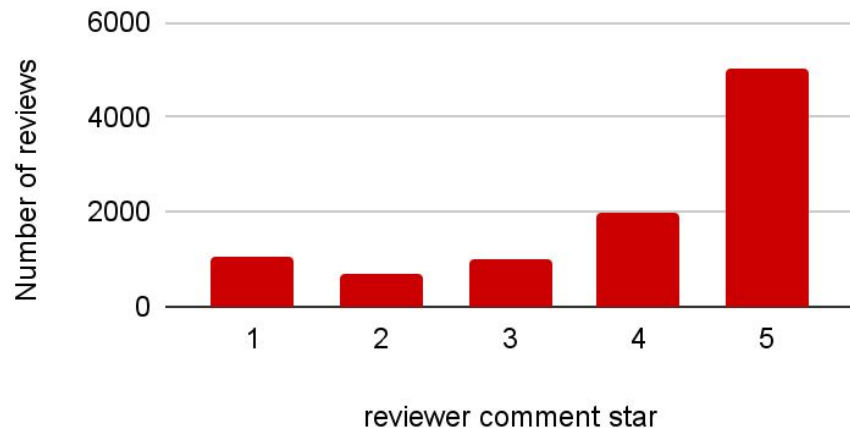


Information from attributes

TV available

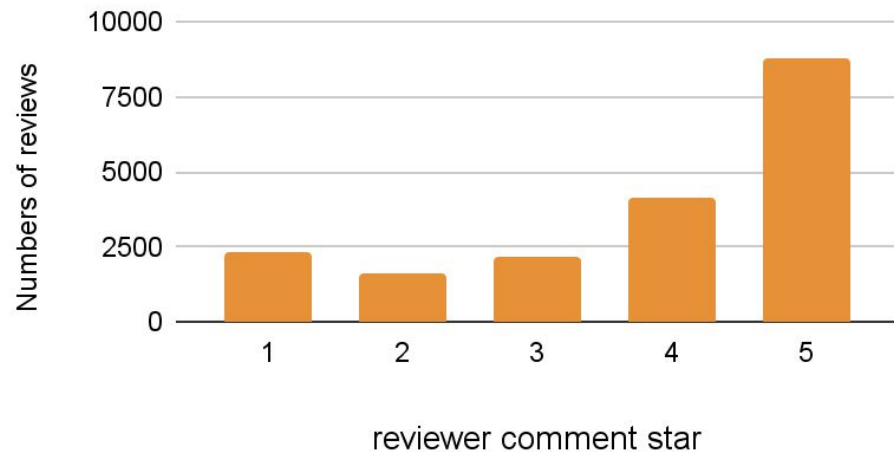


Free WiFi available

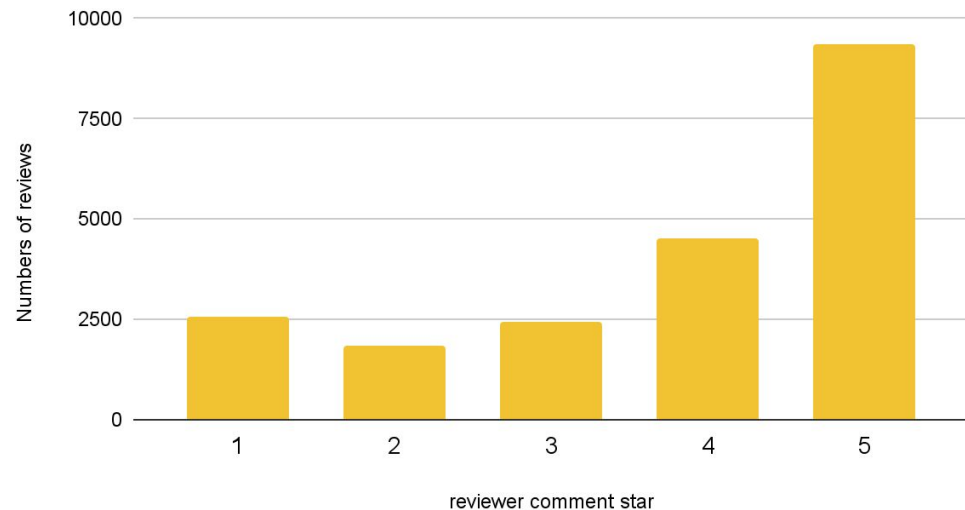


Information from attributes

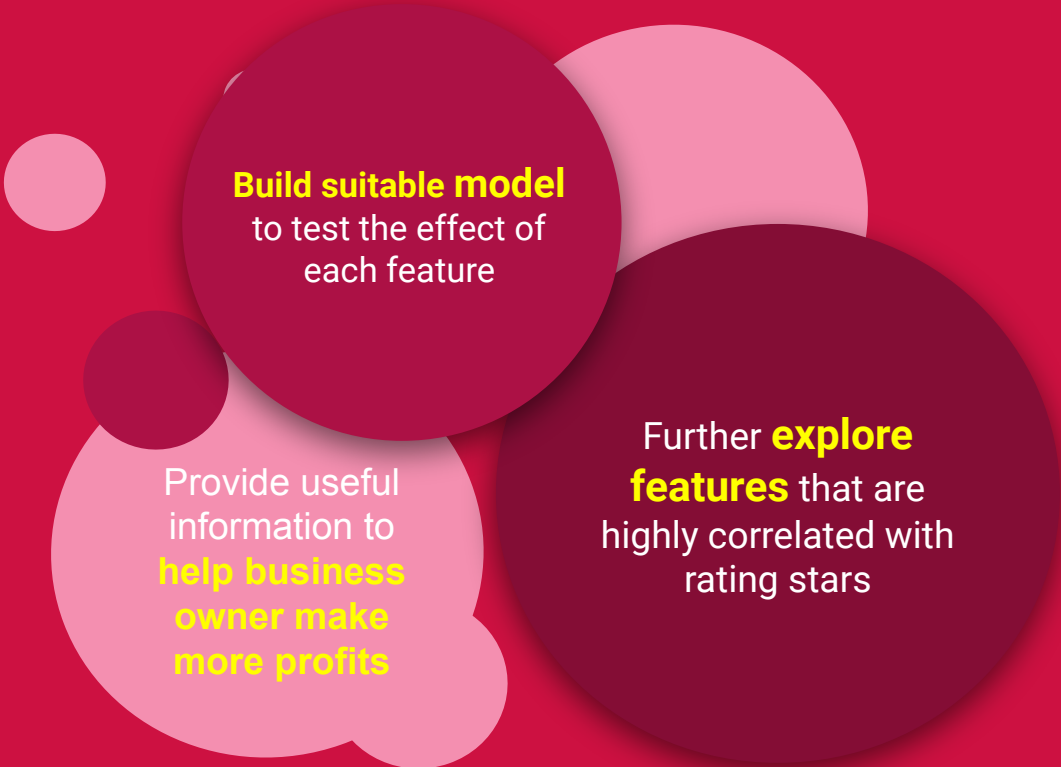
BikeParking Available



TakeOut Service Available



What we want to do next



Build suitable model
to test the effect of
each feature

Provide useful
information to
**help business
owner make
more profits**

Further **explore
features** that are
highly correlated with
rating stars



THANK YOU

