

Yelp Review Analysis

By Sasha Prokhorova



01

**project
overview**

We will talk about
the dataset and
the goals of this
analysis

02

methodology

Tools and
scientific
approach

03

analysis

Research process

04

**takeaways and
future work**


We will discuss
what we learned



01

project overview

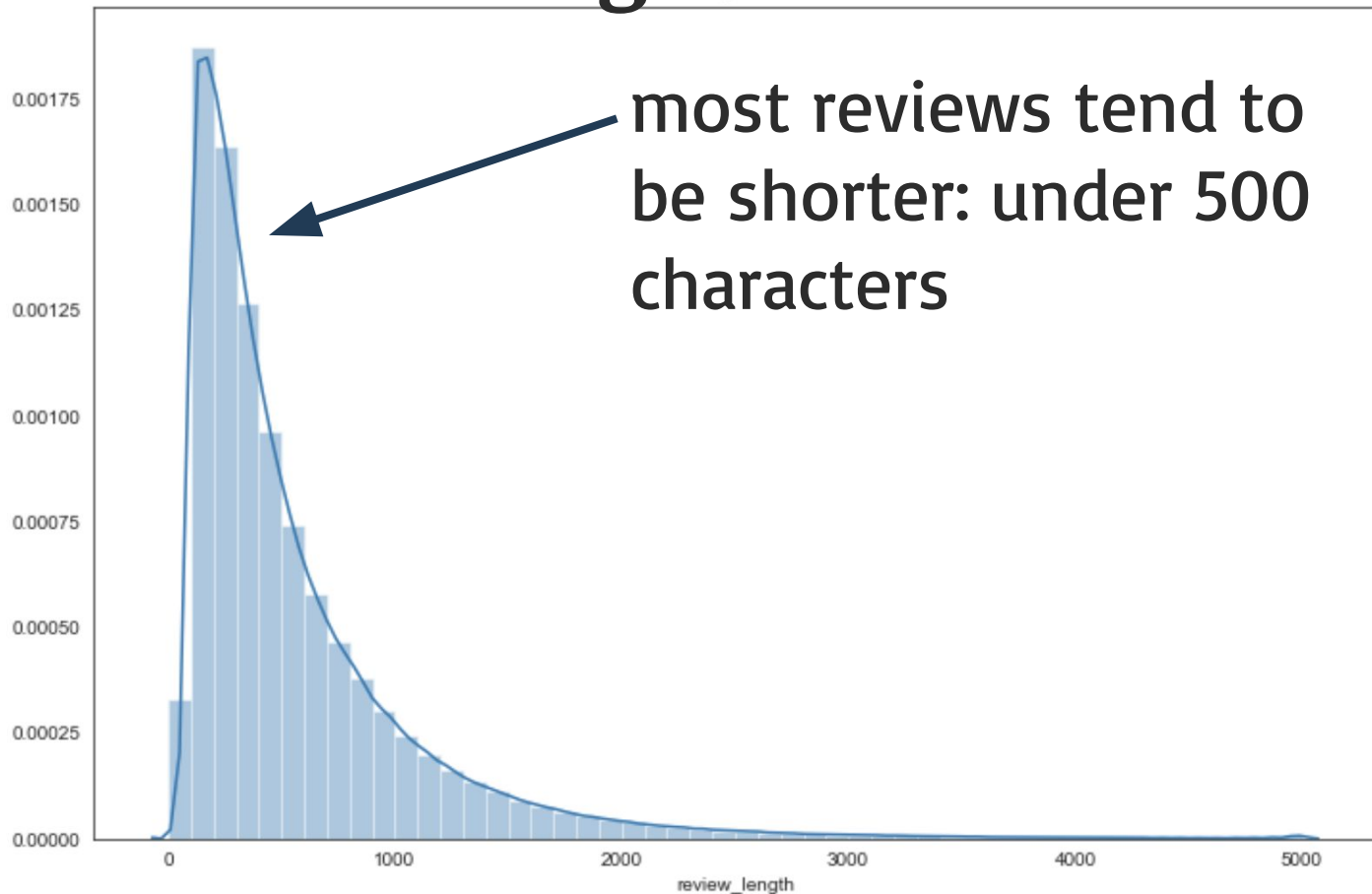
What are the users talking about?

The background is a solid yellow color. It is decorated with various abstract geometric shapes and patterns in blue, red, black, and white. These include circles, squares, triangles, and rectangles, some of which are divided into smaller sections or feature stripes. The patterns are scattered around the central text, with a higher concentration on the left and right sides.

**“I’ve learned that people
will forget what you said,
people will forget what you
did, but people will never
forget how you made them
feel.”**

—Maya Angelou

review length distribution





02

methodology

Analysis roadmap

design process

step 1



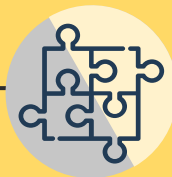
Yelp review data
collection

step 2



Data
pre-processing

step 3



Topic modeling

step 4



Business insights

Libraries and tools



NLTK

The pillar and the
pioneer of NLP



SpaCy

Helps with
effective stemming
and lemmatization



gensim

Works well for
topic modeling and
visualization



sklearn

A trusted machine
learning library

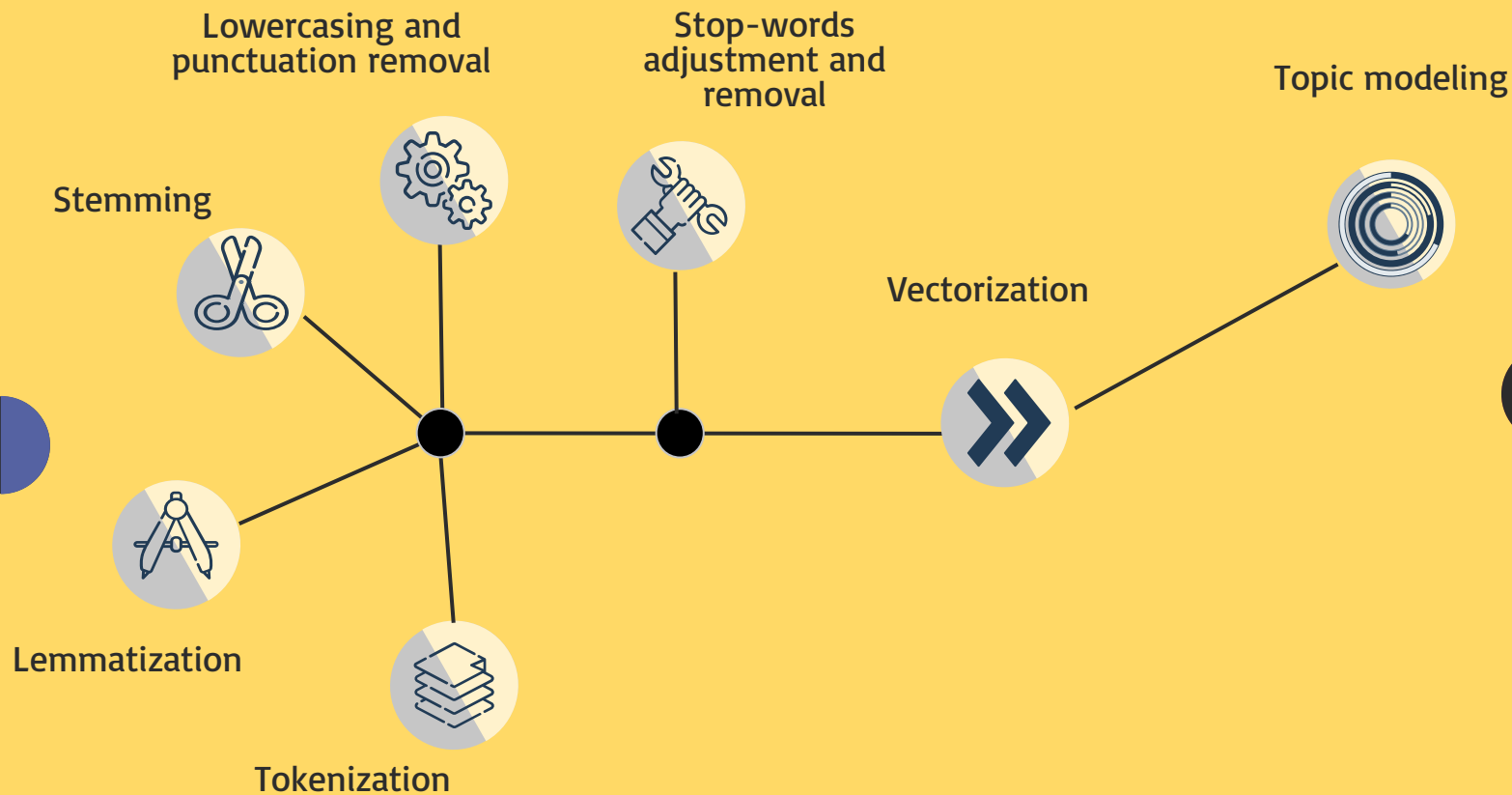


03

analysis

Research process

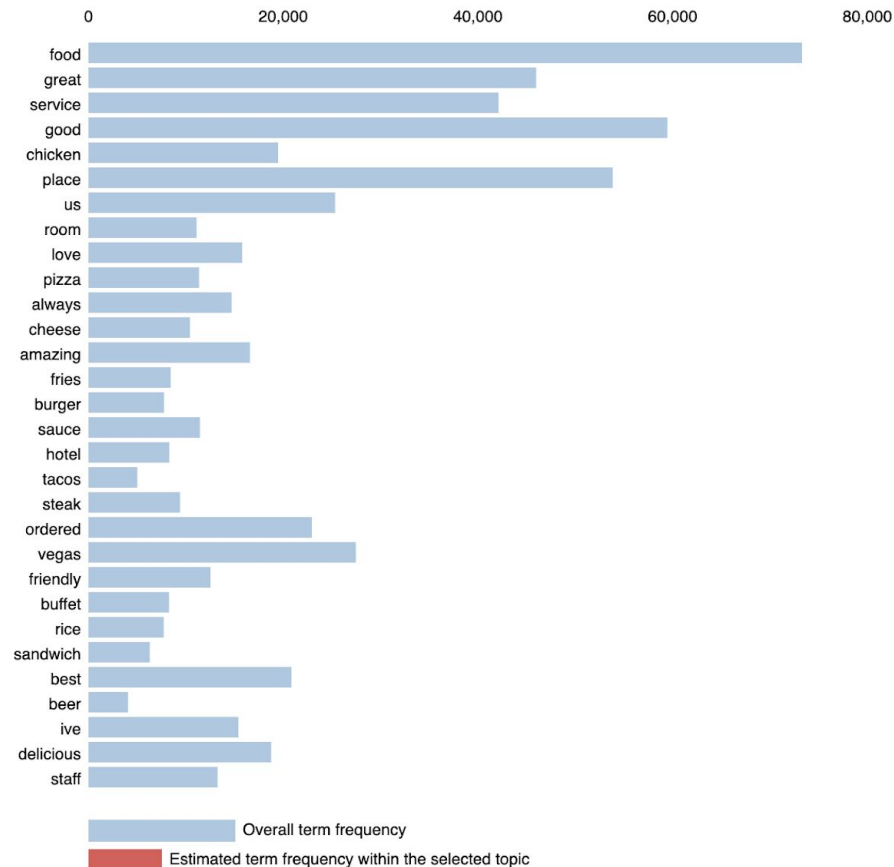
NLP Pipeline



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Salient Terms¹



1. $\text{saliency}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w) / p(t))]$ for topics t ; see Chuang et. al (2012)

2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t) / p(w)$; see Sievert & Shirley (2014)





04

takeaways and future work

Conclusions and ideas



100,000

Las Vegas restaurant reviews analyzed

6

Prevalent topics discovered

11

Customers' favorite dishes found

Review topics discovered



Hotel

Vegas strip experiences: bars, casinos and rooms

Promptness

Mentioning food and service taking long time

Brunch

Breakfast buffet and morning rituals

Surf & Turf

High-end dinner, steak, lobster, shrimp, dessert

Casual dining

Pizza, burgers, sandwiches and tacos

Positivity

Service and food complimented

what do users want?

People want
experiences.



future work



Recommendation system

Providing recommendations based on users' check-ins and likes



Flask app

Web app to help users pick a place to eat



Larger dataset

All the U.S. restaurants with fully automated NLP pipeline



THANKS!

Questions?

CREDITS: This presentation template was created by Slidesgo,
including icons by Flaticon, and infographics & images by Freepik.

please keep this slide for attribution :)





05

appendix

The background of the image is a vibrant yellow. It is decorated with several abstract geometric patterns in the corners. These patterns consist of various shapes including circles, squares, triangles, and rectangles, some of which are filled with solid colors like red, blue, or black, while others feature horizontal or vertical stripes. The overall aesthetic is modern and graphic.

184,000,000

Reviews worldwide



178,000,000
Unique visitors monthly

9%
Increase in revenue for every star
earned
44th
Most visited website in the U.S.

Company timeline

