

Mr. Beans Coffee



Title	Description
Organization	MR. BEANS COFFEE
Example Organization or Category	SMALL BUSINESS
Description of Organization	Small Coffee shop adapting ways to improve customer satisfaction and welcome return customers.
Two Specific Ethical Considerations in AI Adoption	Ensuring fair treatment of employees impacted by AI implementation. Maintaining customer privacy and data security in sentiment analysis.
Three Examples of Operational Decisions	Data scraping and cleaning of Google place reviews. Implementing sentiment analysis algorithms for classification. Developing a user-friendly interface for real-time feedback analysis.
Three Examples of Tactical Decisions	Data scraping and cleaning of Google place reviews. Implementing sentiment analysis algorithms for classification. Developing a user-friendly interface for real-time feedback analysis.
Three Examples of Strategic Decisions	Formulating a marketing strategy based on sentiment analysis insights. Collaborating with coffee shops to enhance customer experiences. Scaling the model for broader application in the food industry.

Why this is topic important?	Customer satisfaction drives business success in the food industry.
How could this project make an impact outside of class?	Enhancing customer experiences and business opportunities for coffee shops.
Why should class choose this topic?	Relevance to real-world applications and potential for practical impact.

DATASET:

<https://www.kaggle.com/datasets/ylchang/coffee-shop-sample-data-1113>

BUSINESS PROBLEM:

- As a busy coffee shop owner, It's difficult to read each and every review manually.
- The coffee shop wants to make its customers happier and improve its service.
- For this they need to go through the reviews but its impossible to manually go through each review for a busy coffee shop manager.

SOLUTION

- The Sentiment analysis model aims to provide valuable insights for coffee shop owners and managers to understand customer satisfaction and make informed business decisions.
- It will analyse the feedback and will show up the sentiment as positive/negative/neutral.
- Understanding customer sentiments can lead to improvements in service quality and overall customer experience.

IMPACT/RESULTS

- Understanding customer sentiments can lead to improvements in service quality and overall customer experience.

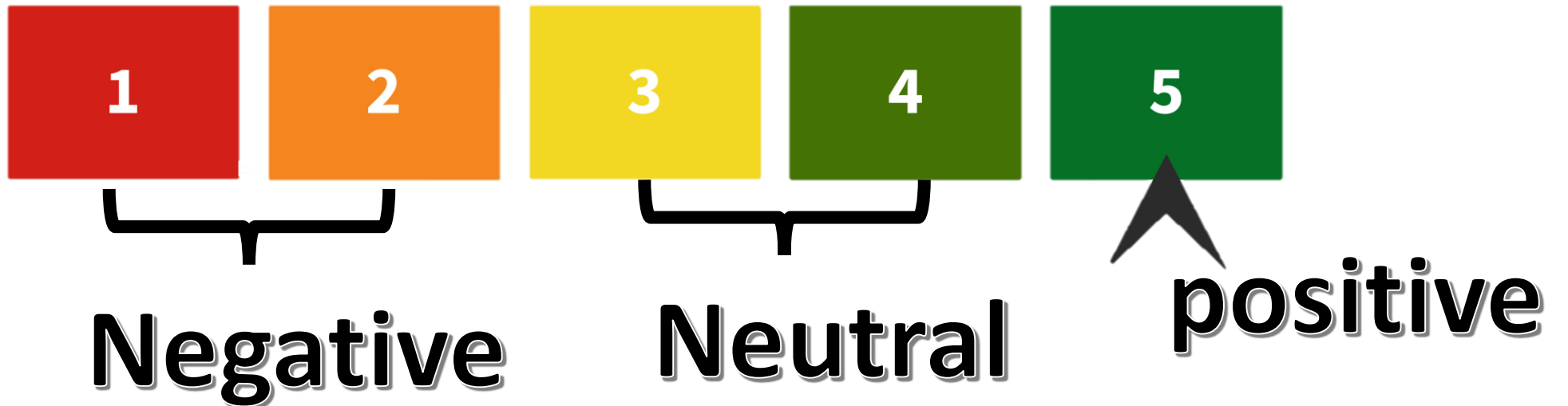
DATASET

- The data is scrapped from google place reviews and consists of 1165 reviews of Surry Hills coffee shops (20) in Sydney, Australia.

	page	name	link	thumbnail	rating	date	snippet	images	local_guide	CoffeeShop
0	1	J Wang	https://www.google.com/maps/contrib/1066405393...	https://lh3.googleusercontent.com/a/ALm5wu2QSW...	1.0	10 months ago	Ordered take away lunch and received this burn...	https://lh5.googleusercontent.com/p/AF1QipNc...	True	Paramount Coffee Project
1	1	Manish Rai	https://www.google.com/maps/contrib/1087306153...	https://lh3.googleusercontent.com/a/ALm5wu0hYM...	1.0	6 months ago	TLDR/Long story short: Wrong pricing online, o...	https://lh5.googleusercontent.com/p/AF1QipO-...	NaN	Paramount Coffee Project
2	1	Lisa	https://www.google.com/maps/contrib/1090187054...	https://lh3.googleusercontent.com/a-/ACNPEu-sy...	1.0	3 months ago	Drinks order was taken and drinks never arrive...	NaN	NaN	Paramount Coffee Project
3	1	Angel Manuel	https://www.google.com/maps/contrib/1127216205...	https://lh3.googleusercontent.com/a-/ACNPEu850...	1.0	2 months ago	I came here twice, poor service. I ordered gre...	NaN	NaN	Paramount Coffee Project
4	1	Jisoo	https://www.google.com/maps/contrib/1168975833...	https://lh3.googleusercontent.com/a/ALm5wu25sD...	1.0	a year ago	We came here because the coffee shop we were g...	https://lh5.googleusercontent.com/p/AF1QipOG...	NaN	Paramount Coffee Project

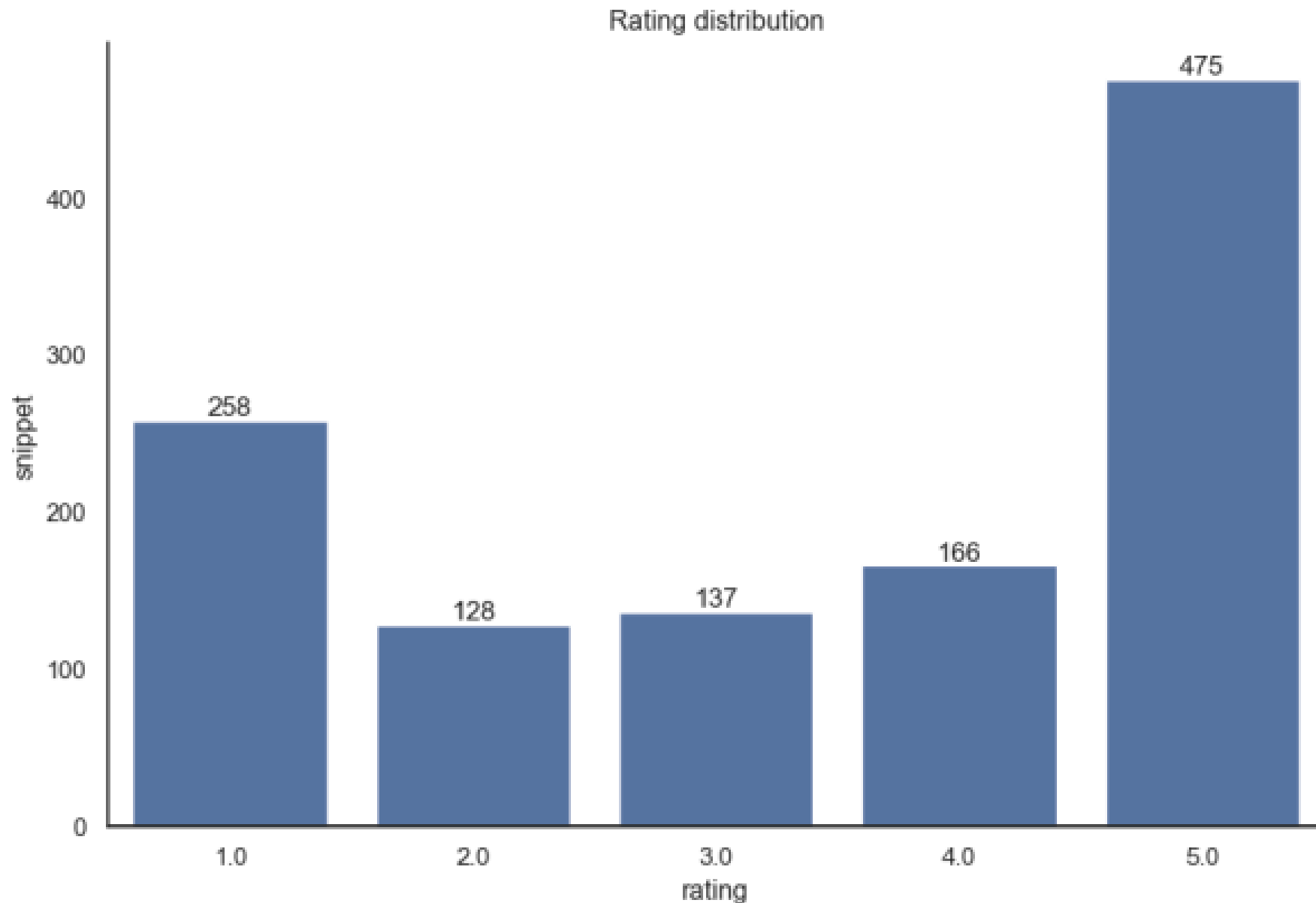
Rating

Rating Scale 1 - 5

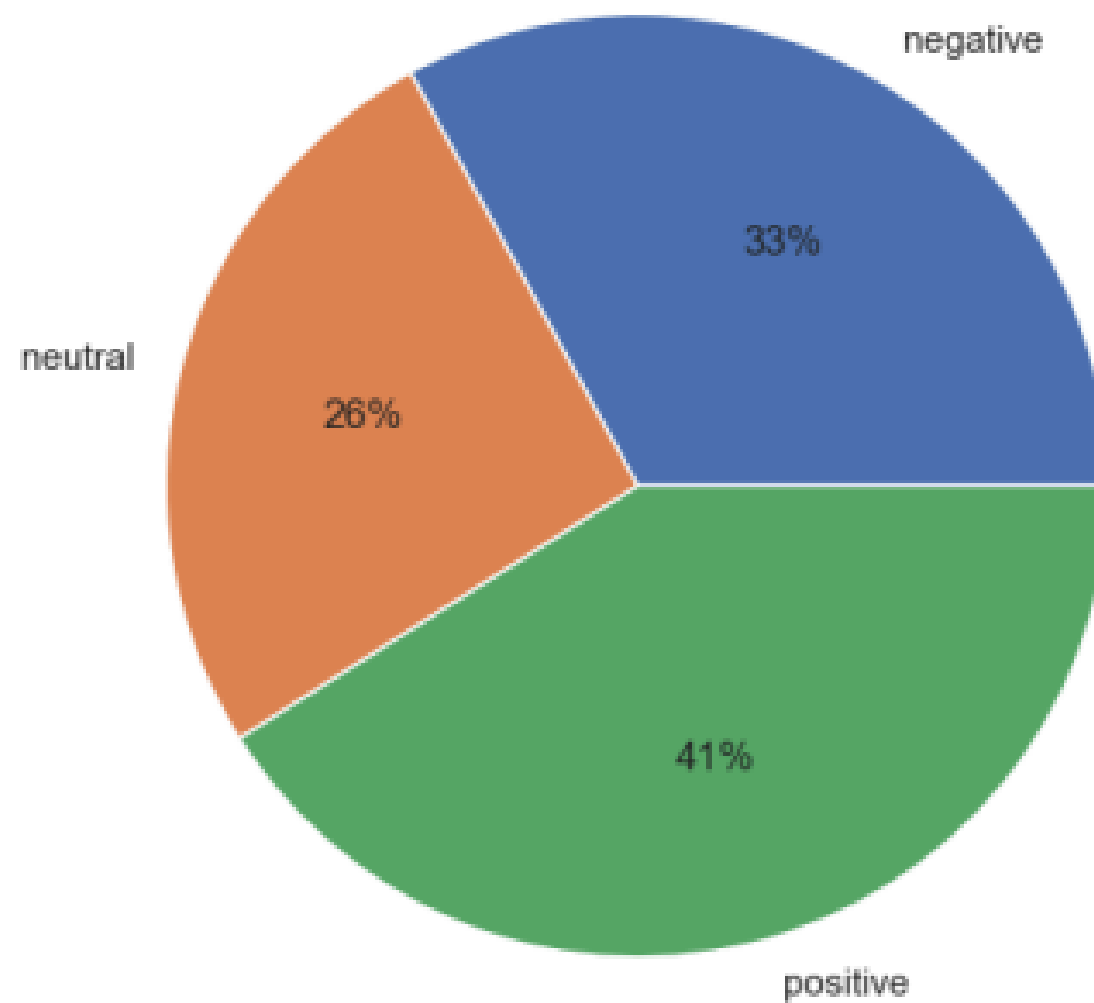


	rating	snippet
0	1.0	258
1	2.0	128
2	3.0	137
3	4.0	166
4	5.0	475

Review vs rating



This dataset is balanced: negative reviews: 0.33, neutral reviews: 0.26, positive reviews: 0.41



Cleaning Data:

```
: #Select only needed columns  
data = df[['snippet', 'sentiment']]
```

- Punctuation and unnecessary words removed
- Custom stop words defined and combined with NLTK stop words

```
# Define custom stop words  
my_stop_words = ['coffee', 'cafe', 'food', 'today', 'yesterday', 'place', 'staff', 'breakfast']
```

- Data prepared for stemming and lemmatization

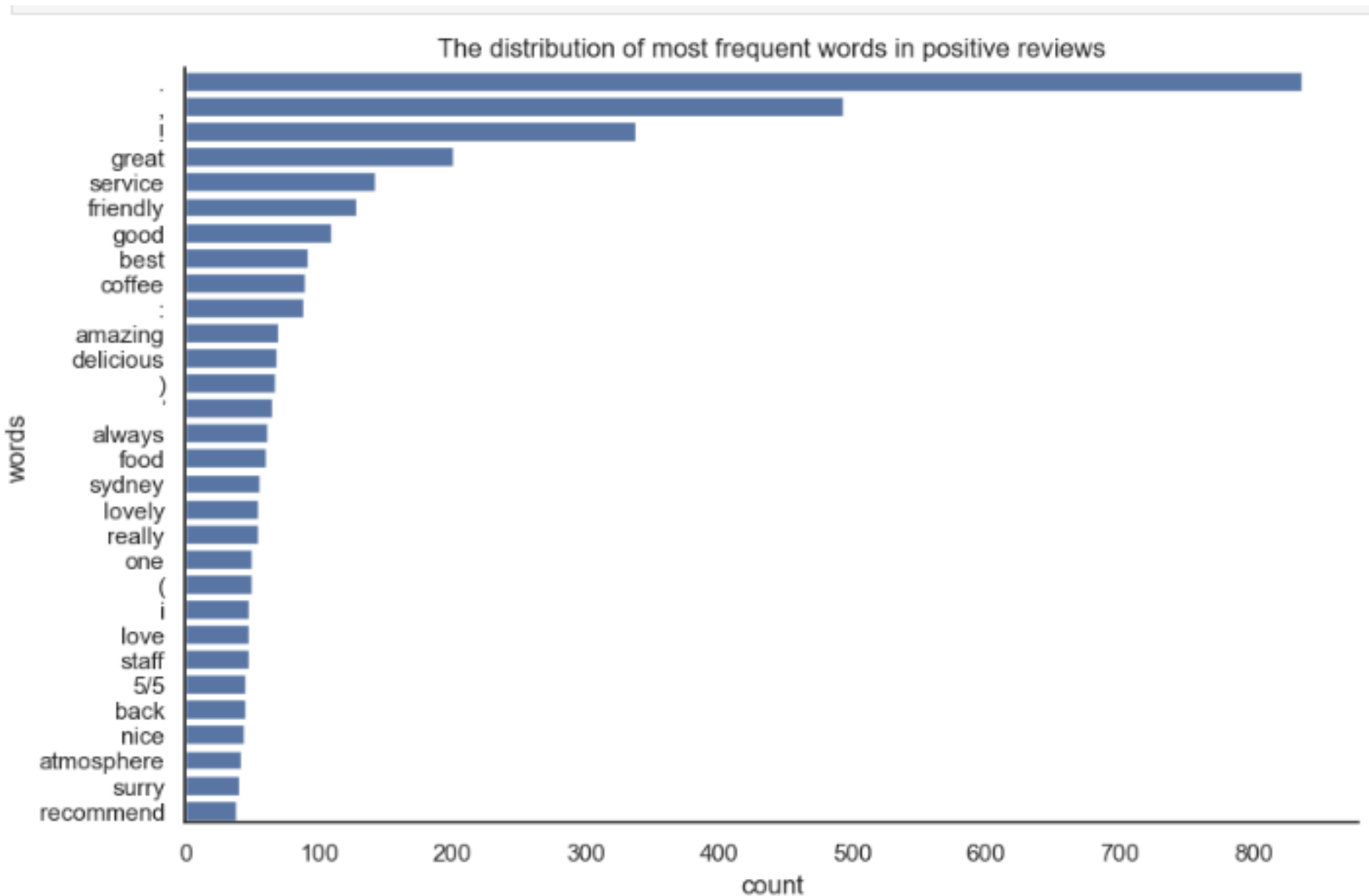
	snippet	sentiment
0	Ordered take away lunch and received this burn...	-1
1	TLDR/Long story short: Wrong pricing online, o...	-1
2	Drinks order was taken and drinks never arrive...	-1
3	I came here twice, poor service. I ordered gre...	-1
4	We came here because the coffee shop we were g...	-1

1. Stemming is a process that stems or removes last few characters from a word, often leading to incorrect meanings and spelling.
2. Lemmatization considers the context and converts the word to its meaningful base form, which is called Lemma.

	tidy_snippet	stem_snippet	lem_snippet
0	ordered take away lunch received burned toasti...	order take away lunch receiv burn toasti (hidd...	ordered take away lunch received burned toasti...
1	tldr/long story short: wrong pricing online, o...	tldr/long stori short: wrong price online, ove...	tldr/long story short: wrong pricing online, o...
2	drinks order taken drinks never arrived, chase...	drink order taken drink never arrived, chase u...	drink order taken drink never arrived, chase u...
3	came twice, poor service. ordered green tea, r...	came twice, poor service. order green tea, rec...	came twice, poor service. ordered green tea, r...
4	came shop gonna go closed. ordered housemade g...	came shop gonna go closed. order housemad gran...	came shop gonna go closed. ordered housemade g...

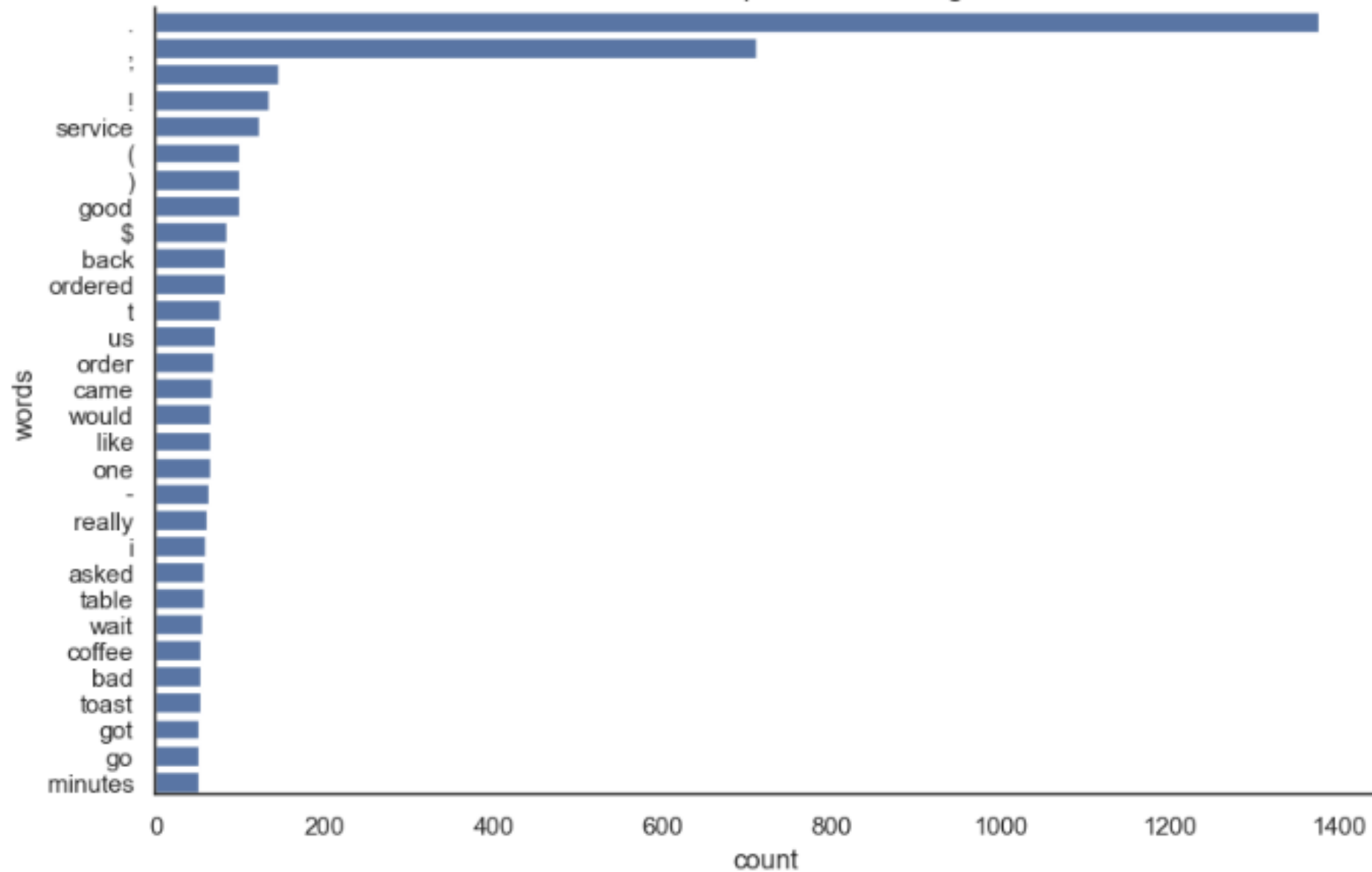
VECTORIZATION

- Vectorization is the process of converting text into numerical data
- Two options explored:
 - Count Vectorizer
 - TF-IDF
- Count Vectorizer: Converts strings into a frequency representation (Bag of Words)
- TF-IDF: Considers word frequency and importance



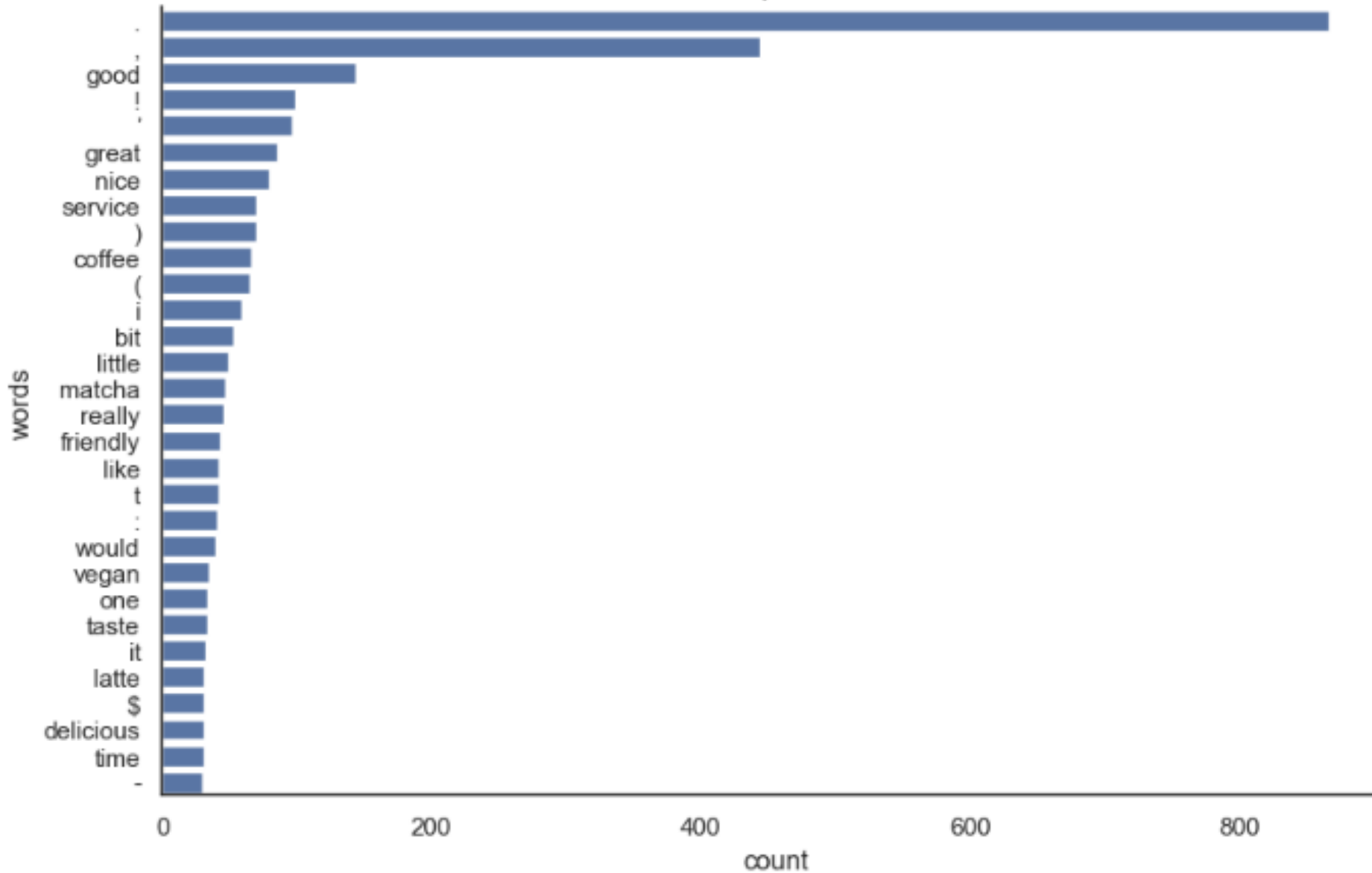
Insights
of
Dataset

The distribution of most frequent words in negative reviews



Insights
of
Dataset

The distribution of most frequent words in neutral reviews



Insights
of
Dataset

Models Examined

- Multinomial Naive Bayes
- Logistic Regression

Examining model in 4 ways

- Stemmed data prepared with Count Vectorizer
- Stemmed data prepared with TF-IDF
- Lemmatized data prepared with Count Vectorizer
- Lemmatized data prepared with TF-IDF

Multinomial NB

Accuracy Multinomial NB with Count Vectorizer and Data Stemming 0.79

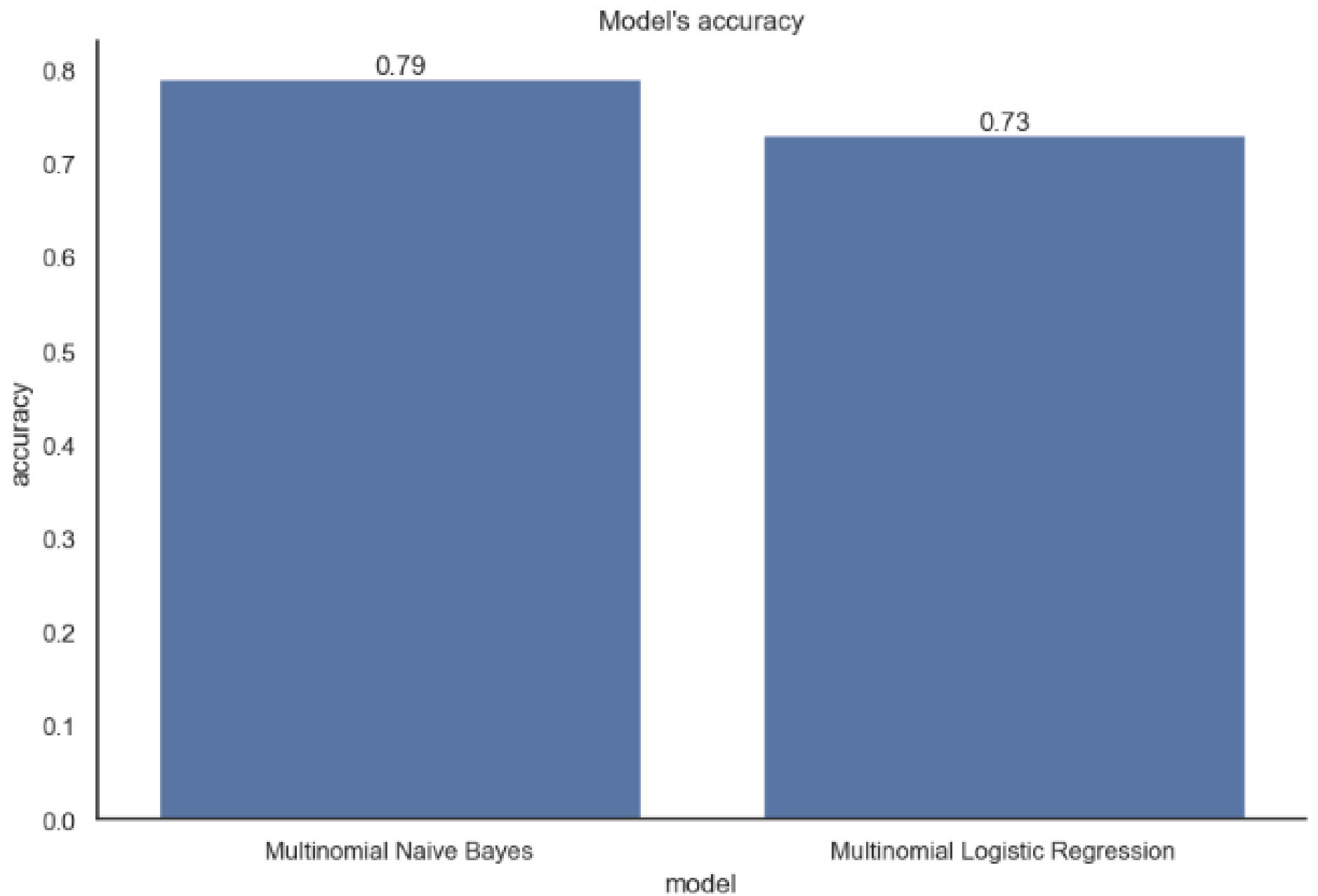
Accuracy Multinomial NB, TF-IDF, stemming 0.76

Accuracy Multinomial NB with Count Vectorizer and Data Lemmatization data 0.77

Accuracy Multinomial NB, TF-IDF, lemmatization data 0.75

Logistic Regression

Accuracy Logistic Regression 0.73



```
: predict_sentiment('very good place')  
# Correct!
```

Positive

```
: predict_sentiment('Charged me $8 for this. The highlight was the lemon. Making toast can be hard, but maybe you should practice a few times before you open a cafe. Good luck.')  
# Correct!
```

Negative

```
: predict_sentiment('2018 went in for a coffee, both my husband and I threw out the coffee - lucky we took our own takeaway cup. No longer a top coffee spot in sydney.')  
# Correct!
```

Negative