



Content



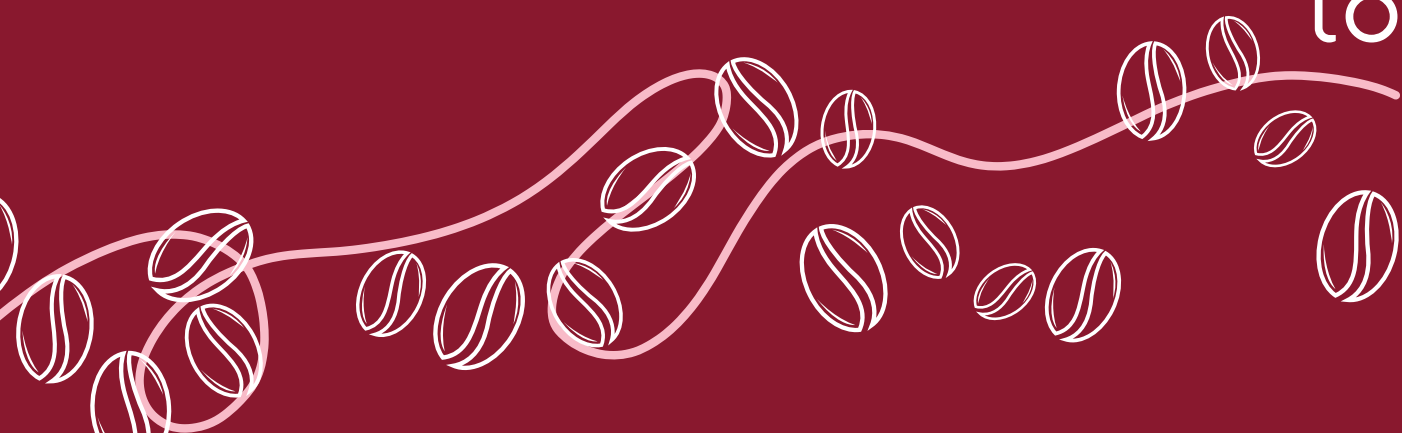
- 1 Introduction
- 2 Data exploration
- 3 Data preprocessing
- 4 Model planning and building
- 5 Data visualization
- 6 Conclusion

Introduction



costa coffee

Costa Coffee is a British multinational coffeehouse company founded in London in 1971. With over 3,800 stores across 32 countries, Costa Coffee is the second-largest coffeehouse chain in the world and the largest in the United Kingdom. The company serves a wide range of coffee-based drinks, as well as tea, hot chocolate, and a variety of food items, such as sandwiches, pastries, and cakes. Over the years, the company has garnered a loyal customer base, who appreciate its high-quality coffee and friendly atmosphere. Whether you're looking for a quick coffee on the go or a relaxing spot to catch up with friends, Costa Coffee has something to offer for everyone.



project goals



USE DATA ANALYSIS
SKILLS TO ANALYZE
TWEETS INVOLVING
COSTA COFFEE.



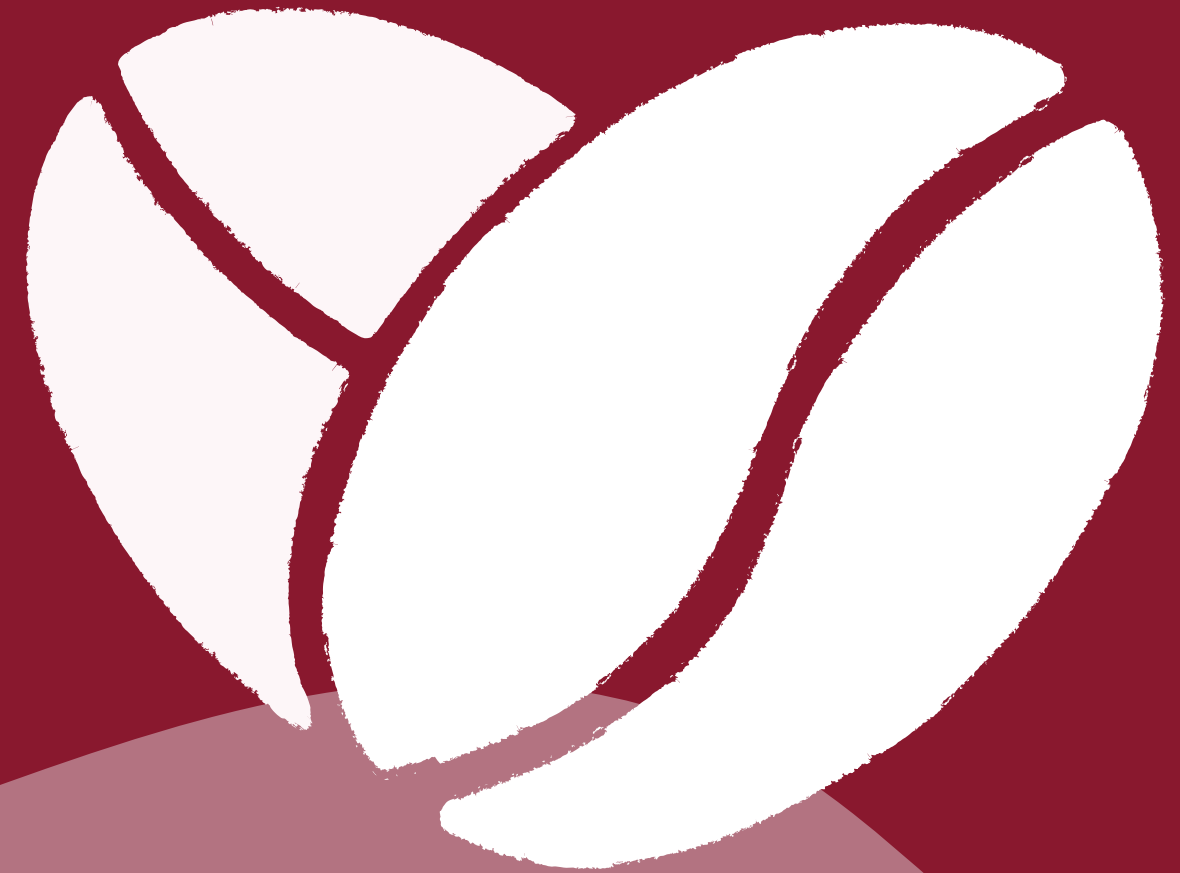
LEARN WHAT
CONSUMERS FEEL
ABOUT COSTA COFFEE
AND THE ISSUES THEY
EXPERIENCE.



ASSIST COSTA COFFEE
LEARN ABOUT
OPINIONS AND TAKE
CUSTOMER INPUT
INTO CONSIDERATION

Initial Hypothesis

AT THE END OF OUR ANALYSIS, WE SHALL EITHER CONFIRM OR DENY THE CLAIM THAT THE MOST OF THE CONSUMERS TEND TO LOVE TRADITIONAL COFFEE, THEY LOVE THE COFFEE ENVIRONMENT, AND THEY THINK THAT COSTA COFFEE PRICES ARE EXPENSIVE.



The background is a solid dark red color. In the top right corner, there is a cluster of white-outlined coffee beans arranged in a loose, upward-curving pattern. On the left side, there is a lighter red, wavy shape that overlaps the main background. At the bottom right, there is a thin, white, wavy line that curves upwards and then downwards.

Data Exploration

At the beginning of phase 2, we explored the data we collected in phase 1. We explored a total of 1,049 data collected about costa coffee.

The following is
what we found
about our data.

COLUMNS NAME		DATA TYPE		INFO OF DATASET	
Total dataset		COUNT OF NON-NULL VALUES			

The background is a solid dark red color. In the top right corner, there is a cluster of white-outlined coffee beans arranged in a diagonal line. On the left side, there is a large, light red wavy shape that partially overlaps the dark red background. In the bottom right corner, there is a white wavy line that curves upwards and to the left.

Data preprocessing

After exploring the dataset collected, we looking through the data, we identified some issues that need to be taken into account.

issues that we addressed

CleanedTweets	remove_Emojis
We should launch a kindness c...	We should launch a kindness c...
use my link to treat us both xx\n\nUse the Co...	use my link to treat us both xx Use the Cost...
After the release of a new video of costa coff...	After the release of a new video of costa coff...
مع هشام باشا ميسوط أوي 🤗🤗 (Costa Coffee Dr...	Costa Coffee Dr... مع هشام باشا ميسوط أوي
Never use Costa or any of the other coffee sh...	Never use Costa or any of the other coffee sh...

REMOVE
DUPLICTE
RECORDS

REMOVE
EMOJIS

Tweets	CleanedTweets
@chrisdysonHT @SafeSENCOSaeed @BrightLeadChris...	We should launch a kindness c...
#Costa use my link to treat us both xx\n\nUse ...	use my link to treat us both xx\n\nUse the Co...

REMOVE
LINKS, MENTIONS, HASHTAGS

issues that we addressed


REMOVE
STOP WORDS

REMOVE
SPECIAL
CHARACTERS


REMOVE
FOREIGN WORDS

remove_Emojis	remove_nonEnglish
We should launch a kindness c...	We should launch a kindness c...
use my link to treat us both xx Use the Cost...	use my link to treat us both xx Use the Cost...
After the release of a new video of costa coff...	After the release of a new video of costa coff...
Costa Coffee Drive Thru in Mecc... مع هشام باشا ميسوط أوي	Costa Coffee Drive Thru in Mecc...

MODEL PLANNING AND BUILDING



DESCRIPTIVE
ANALYSIS



PREDICTIVE
ANALYSIS

DESCRIPTIVE ANALYSIS

Attempts to characterize or summarize past and present data, helping to create accessible data insights.

IT ALLOWS US TO
DETECT COMMONALITIES
BETWEEN VARIABLES,
PREPARING US FOR FURTHER
STATISTICAL ANALYSIS

IDENTIFY THE SHAPE,
COLUMN AND SUMMARY
OF THE DATA FRAME

WHAT WE USED

- summary of data using `info()`
- `describe()` function to calculate mean, std, and IQR values for numeric columns
- `count()` function



PREDICTIVE ANALYSIS

A BRANCH OF ADVANCED ANALYTICS THAT MAKES
PREDICTIONS ABOUT FUTURE OUTCOMES USING
HISTORICAL DATA COMBINED WITH STATISTICAL
MODELING, DATA MINING TECHNIQUES, AND
MACHINE LEARNING



PREDICTIVE ANALYSIS

- Model Comparison
- Training Pipeline
- Identifying The Training & testing set
- Remove Neutral
- Convert Textual To Numerical

PREDICTIVE ANALYSIS HAS TWO MODELS



Naïve Bayes

The diagram consists of two large circles side-by-side. The left circle is white with a dark red border and contains the text 'Naïve Bayes'. The right circle is a solid dark red color and contains the text 'Logistic Regression' in white. The background is a solid dark red color.

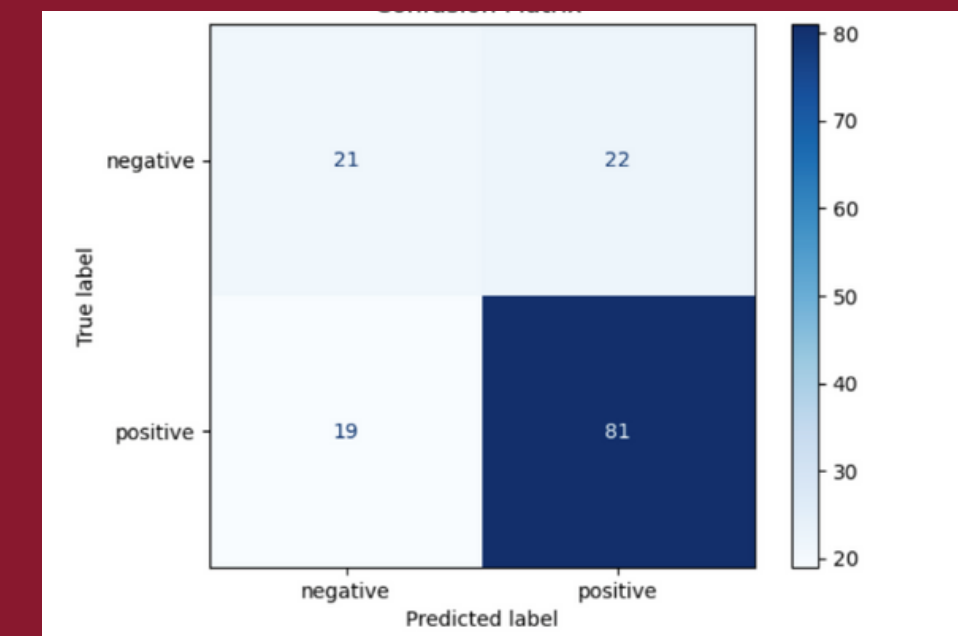
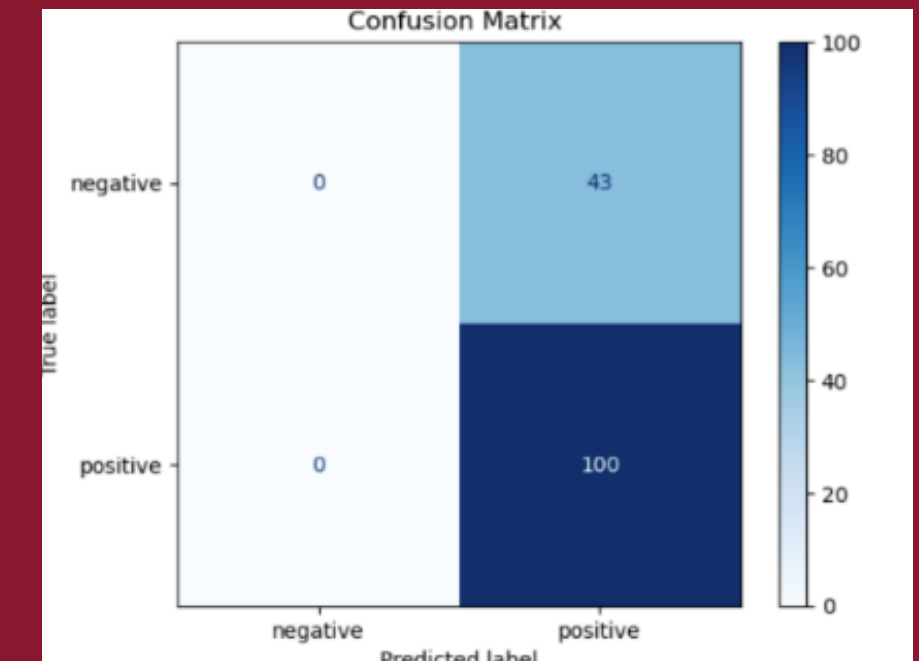
Logistic Regression

Naïve Bayes

A SUPERVISED LEARNING ALGORITHM,
WHICH IS BASED ON BAYES THEOREM AND
USED FOR SOLVING CLASSIFICATION
PROBLEMS

- SEVERAL LIBRARIES WERE USED,
INCLUDING PANDAS, NUMPY, SKLEARN,
AND MATPLOTLIB
- BALANCED AND UNBALANCED DATASET

NAÏVE BAYES ACCURACY IS 0.8 OR
80% REPORT OF UNBALANCED.



NAÏVE BAYES ACCURACY IS
1.00 OR 100% REPORT OF
BALANCED

Logistic Regression

- A CLASSIFICATION ALGORITHM USED TO ASSIGN OBSERVATIONS TO A DISCRETE SET OF CLASSES
- THE OBJECTIVE IS TO BUILD A BINARY CLASSIFICATION MODEL WITH TWO CLASSE POSITIVE AND NEGATIVE.

logistic accuracy: 0.75

10-fold cross-validation:

[0.72916667 0.72916667 0.72916667 0.72916667 0.72916667 0.72916667
0.72916667 0.74468085 0.74468085 0.74468085]

the average accuracy of the logistic classifier is : 00

nConfusion matrix of the logistic classifier:

[[108 0]
[36 0]]

classification_report of logistic classifier :
precision recall f1-score support

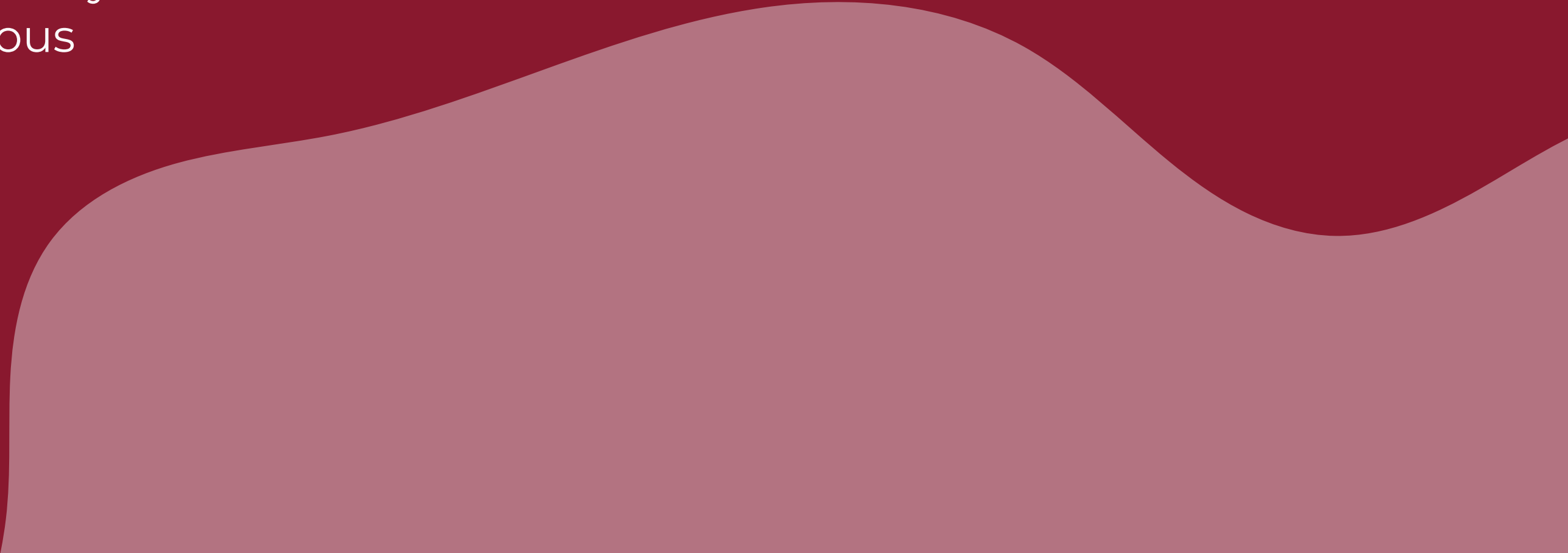
0.0	0.75	1.00	0.86	108
1.0	1.00	0.00	0.00	36
accuracy			0.75	144
macro avg	0.88	0.50	0.43	144
weighted avg	0.81	0.75	0.64	144

THE LOGISTIC REGRESSION
ACCURACY IS 0.75 OR
75%

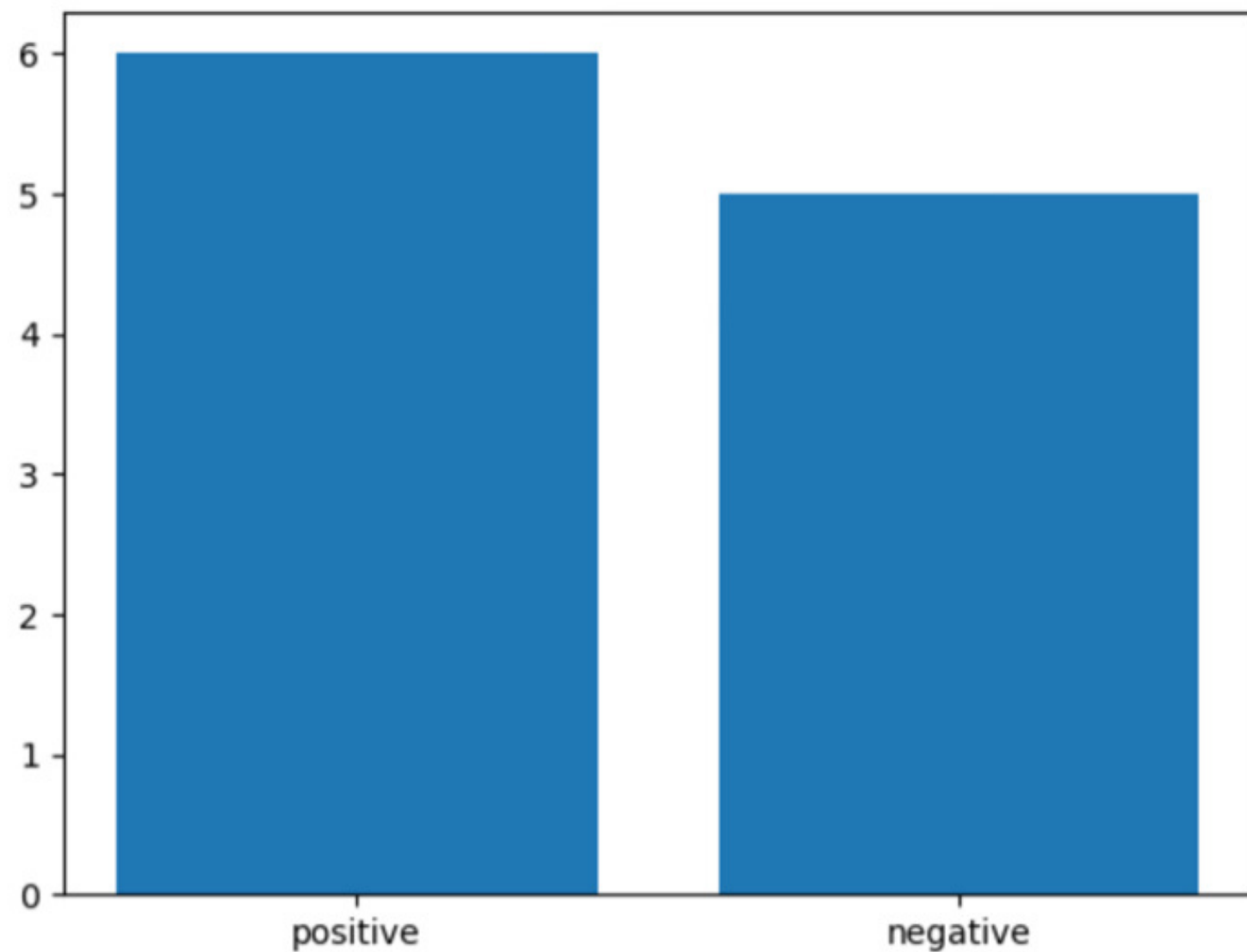


Communicate Results

One of the most important abilities for data scientists to have is the ability to clearly convey results to various stakeholders.



first hypothesis “Customers tend to love traditional coffee.”



We got insight from the first diagram that people seem happy about costa coffee taste. Based on our data analysis, the number of positive tweets about costa coffee taste are higher than the negative

Second hypothesis “people opinion on costa coffee environment”

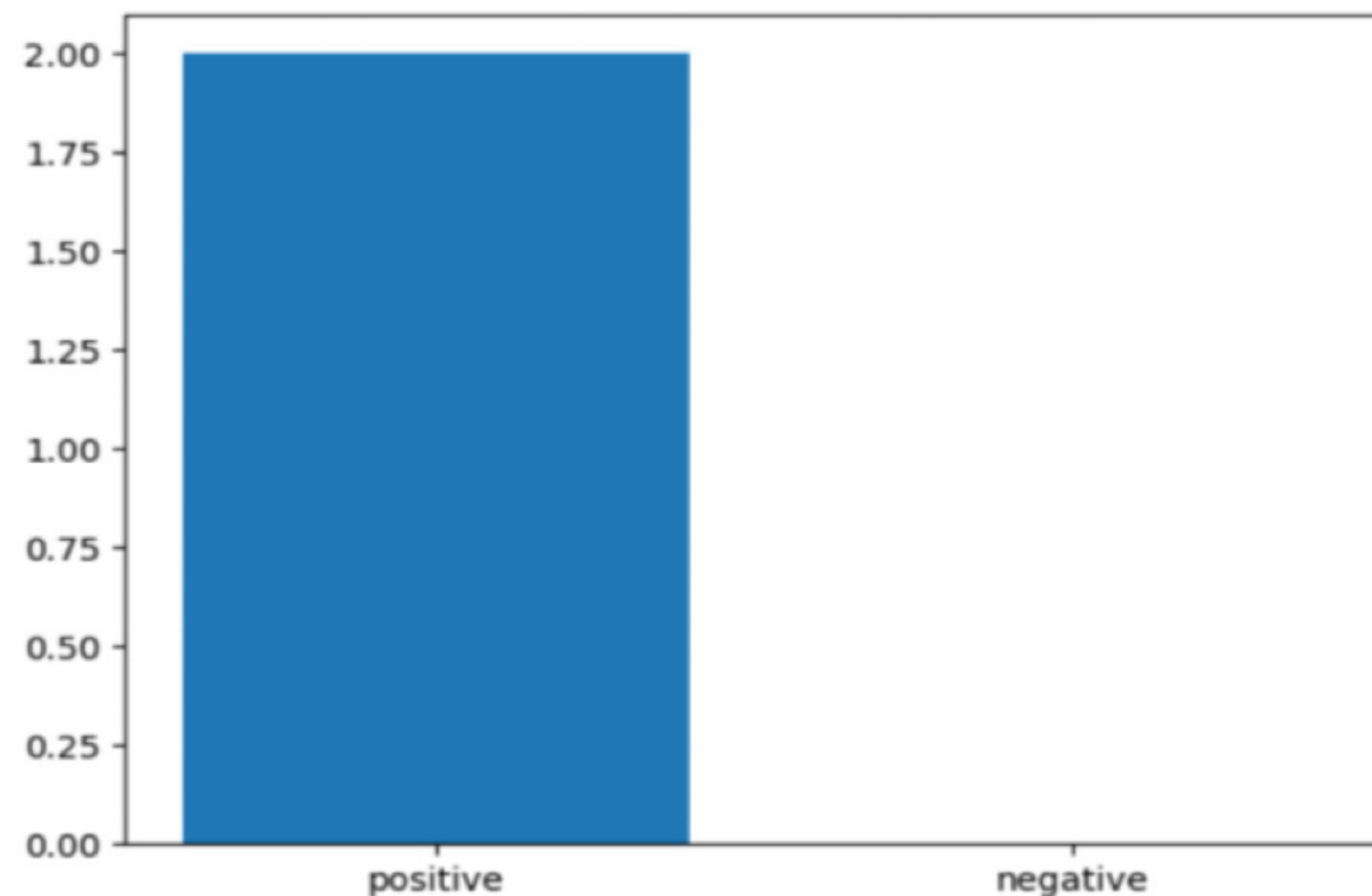


Figure 53 : bar plot chart

We got insight from this diagram that people seem very happy about costa coffee environment. Based on our data analysis, the number of positive tweets about costa coffee environment are the highest

Third hypothesis : ' People's Opinions on costa prices



Based on the diagram, it appears that people are even with Costa's prices. , we found that negative tweets about Costa prices are equal with positive tweets. Our hypotheses are disproved by customers who think Costa prices are good.

Conclusion

- Data analysis will always help us get better results and know what is going on and then choose the right decisions. Unclean data makes it difficult to choose the right path, so we need to take care of our data to make our decisions better.



Thank you for listening