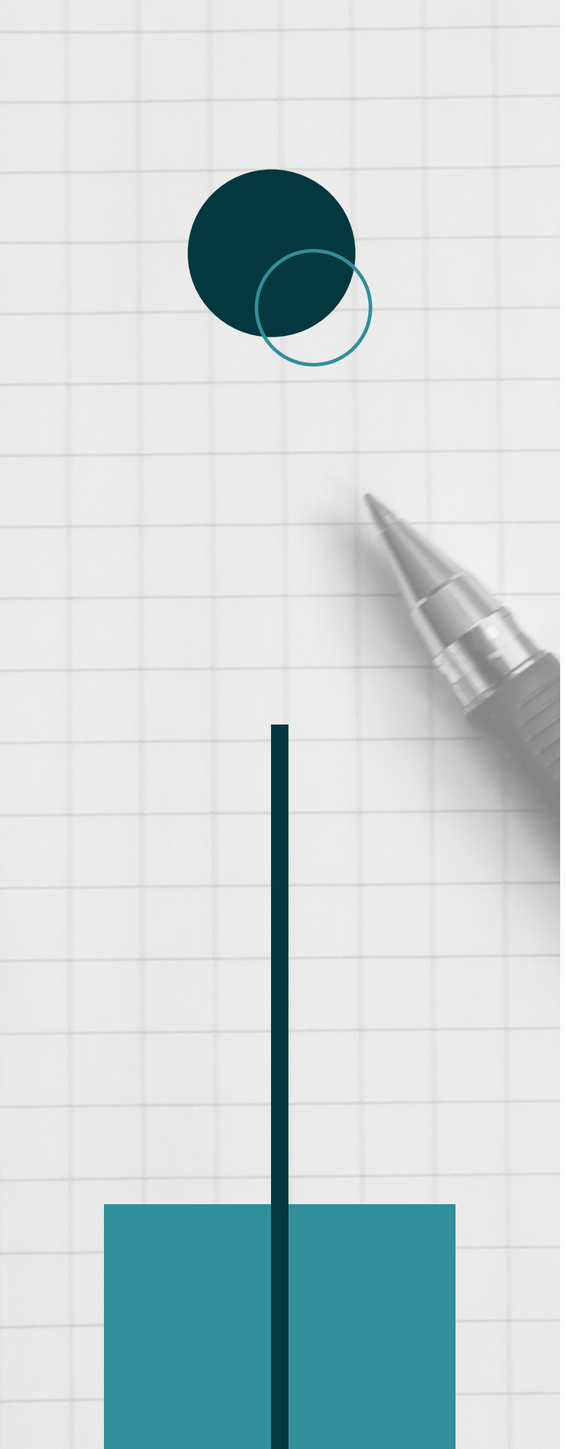
TWITTER SENTIMENT ANALYSIS

TABLE OF CONTENT



BUSINESS UNDERSTANDING

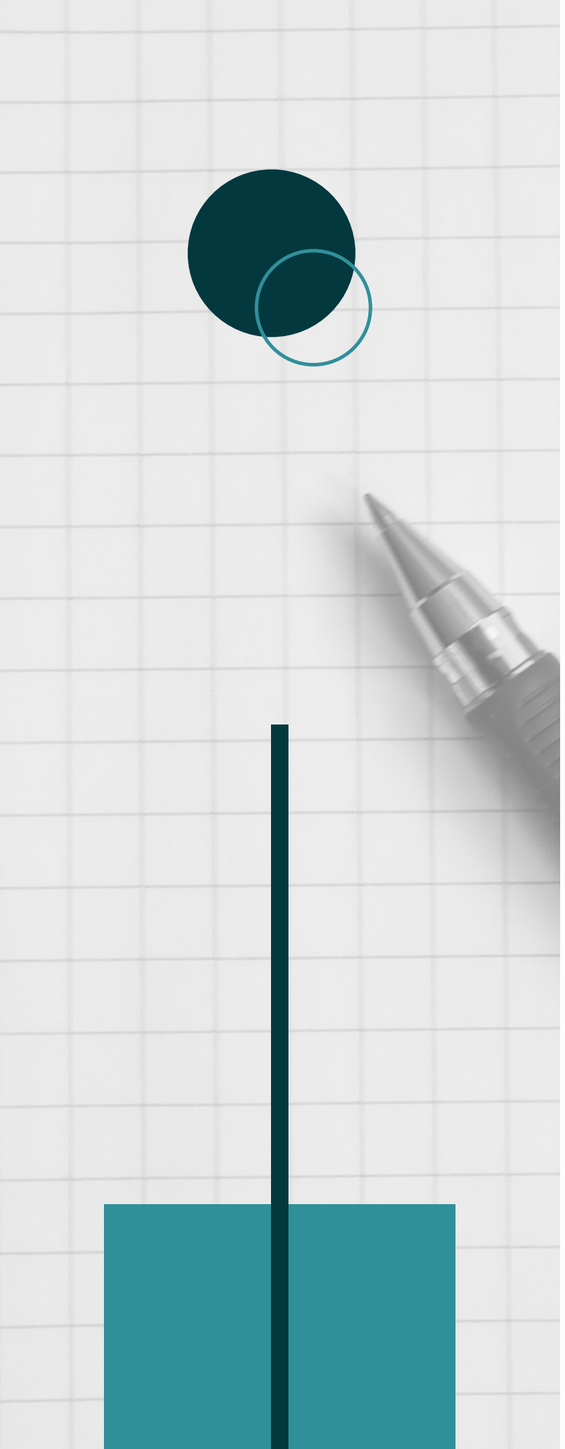
Consumer perception of a brand can provide valuable information about their purchasing behavior and, in turn, the financial performance of the business that produces them.

In order to determine which brands to research further for potential investment, Longview tech ventures want a generalizable model to measure sentiment across various brands.

They have tasked us with developing a predictive model that keeps track of recent tweets about tech products so they can make smart investment choices

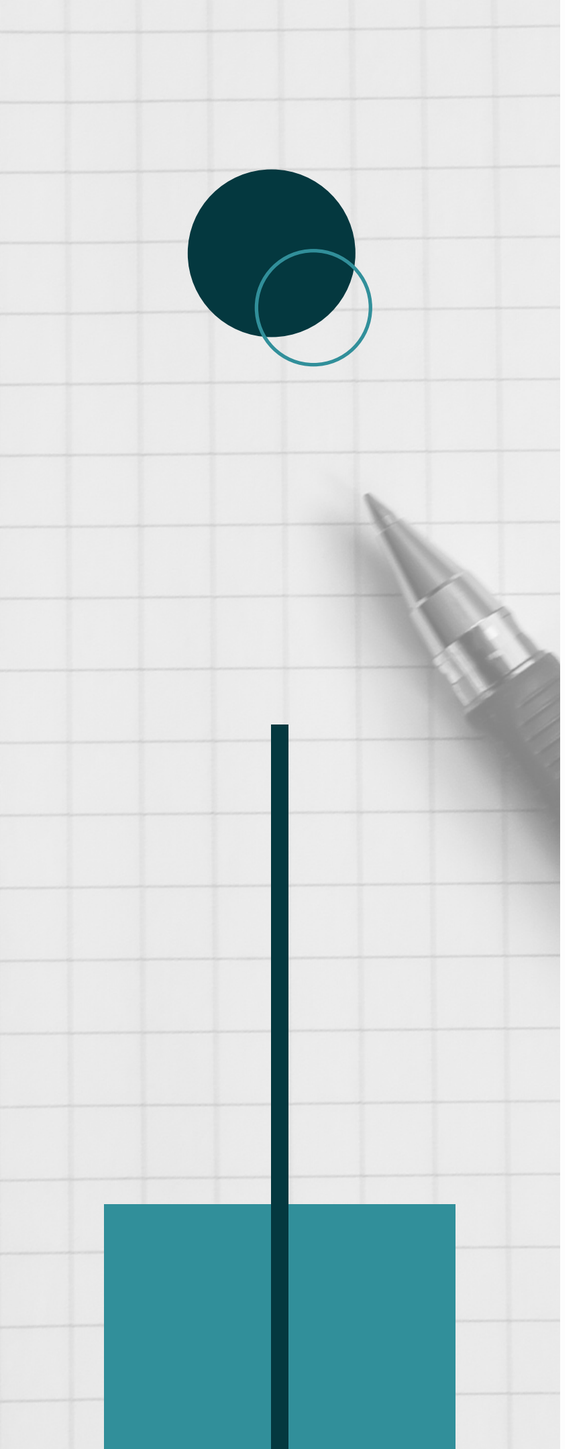
PROBLEM STATEMENT

- Due to rise in popularity among product brands, Longview-tech ventures would like to know which brand is doing best, in order to make an investment.



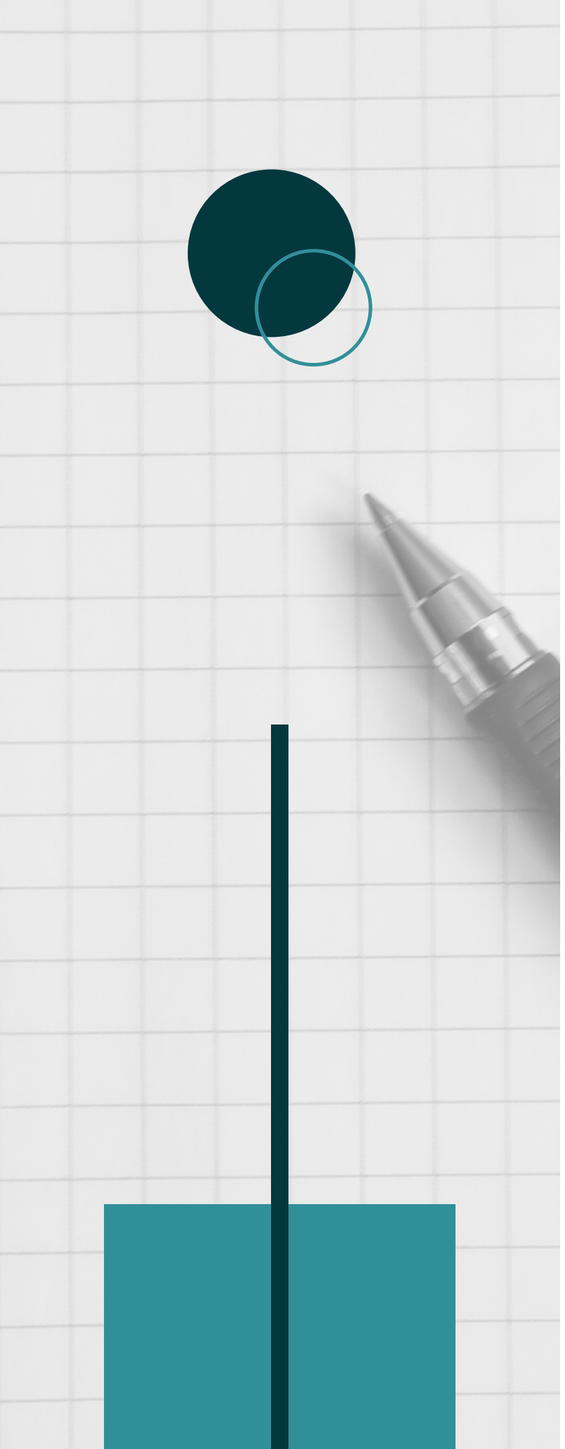
PROPOSED SOLUTION

- We create a model that analyzes tweets to ascertain how customers feel about a brand.



OBJECTIVES

- To keep track of tweets about tech brands.
- To determine the best brand to make an investment in.



DATA PREPARATION

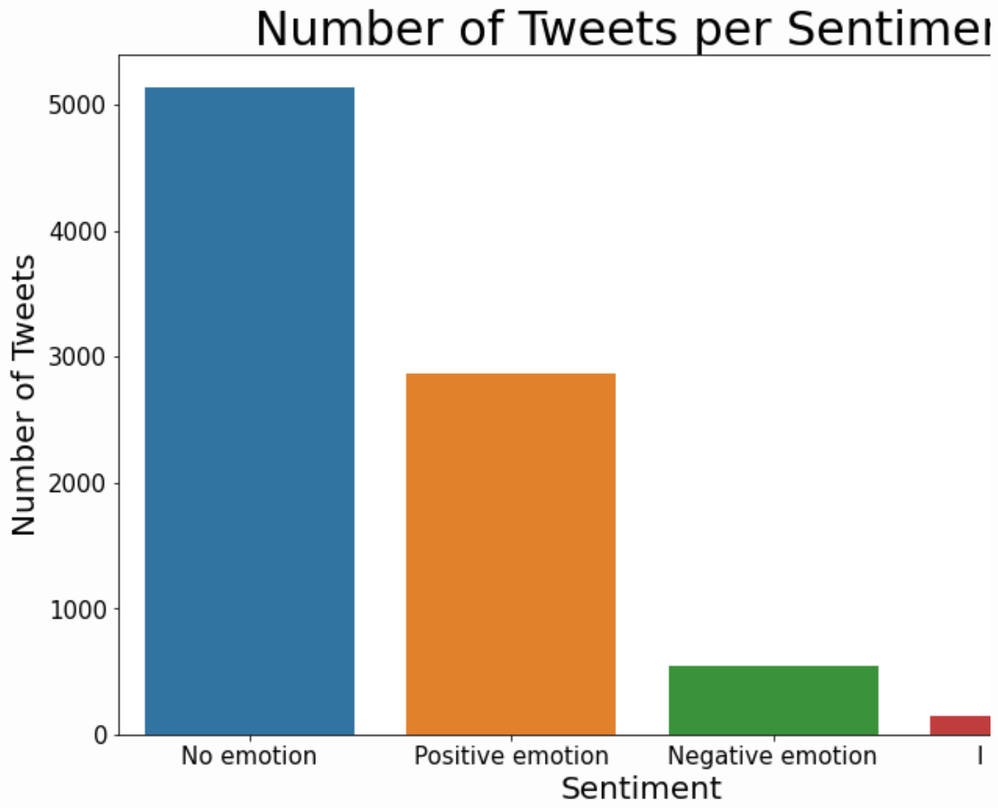
Data Cleaning

Getting the Data Into a neat format

Data Preprocessing

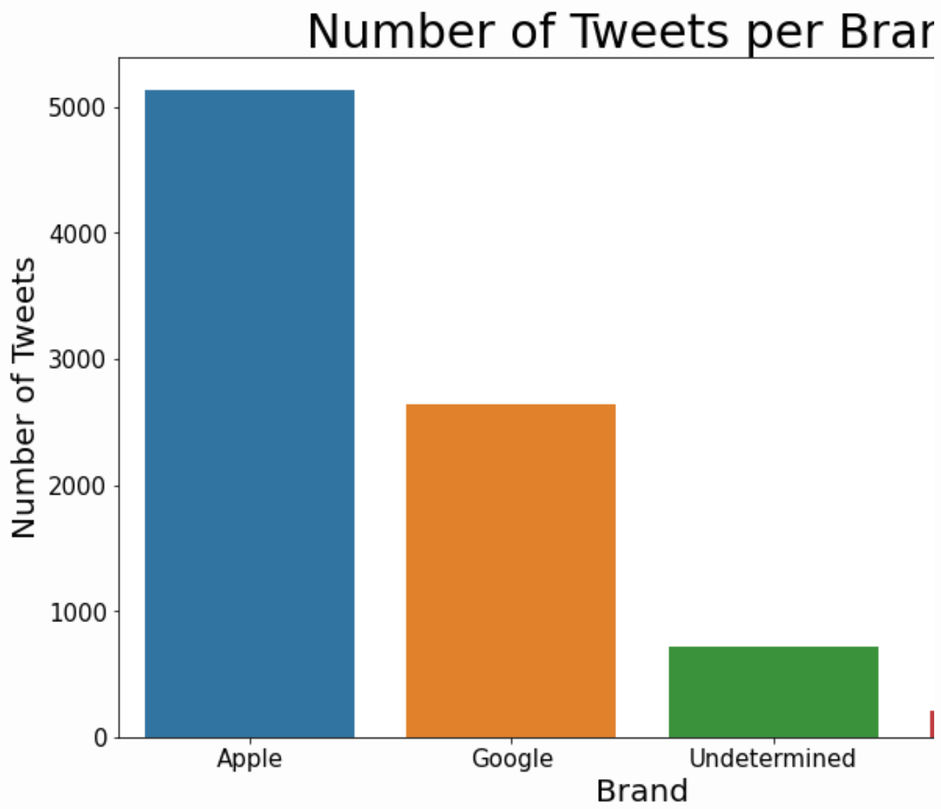
Getting the data into a format that can be fed into the model

DATA EXPLORATION



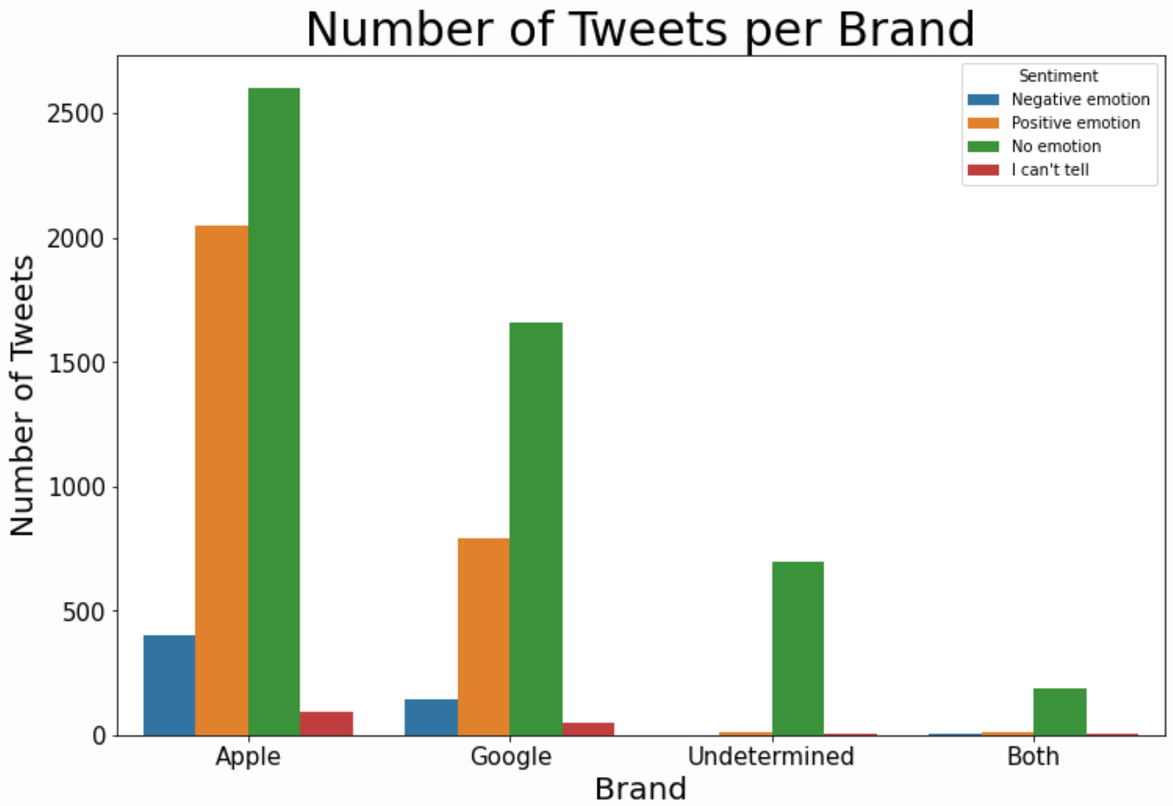
- The plot shows the number of tweets per sentiment. Based on the plot 'No Emotion" has the highest tweet count 0f about 60 % of all tweets followed by positive emotion.

DATA EXPLORATION



- The plot shows the number of tweets per brand. It's clear that the highest number of tweets are on Apple of about 60% followed by Google which composed of 30% of all tweets. With tweets on both brands being the least.

DATA EXPLORATION



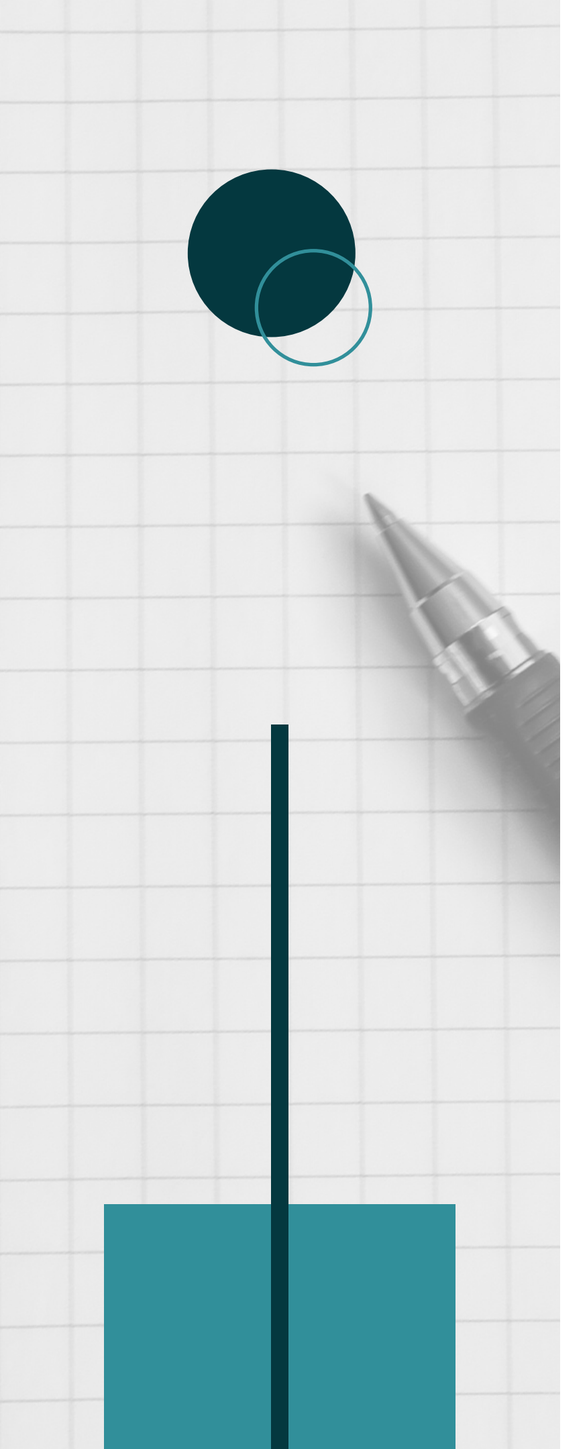
- The plot shows each brand and provides the count of sentiments per brand. From the plot, we see that for every brand 'no emotion' has the highest count then followed by positive emotion with Apple receiving the highest number of tweets with positive emotions.

DATA EXPLORATION

Word Cloud

MODELING AND EVALUATION

Multinomial Naive Bayes Classifier model had an f1 score of 74% on train data and 63% on test data this is our best model because which means that F1-score takes both precision and recall into account, which also means it accounts for both FPs and FNs.
Created a Random Forest Classifier where the training accuracy is at 0.95 against 0.65 for the test data hence not the best due to overfitting
LSTM Neurals model had a training accuracy is at 0.88 and 0.65 for the test data ,the model was overfitted



RECOMMENDATION

Using data from other tech companies, particularly small startups and PR companies which are often overlooked

We recommend our long view-tech investor to go with apple products since it had a higher positive rate as compared to google

CONCLUSION

- From the visuals we plotted, it is evident that most of the tweets were targeted toward apple products with the apple brand receiving a high number of both positive and negative sentiments then followed by Google.
- There were also clear signs of class imbalances in our target variable with 'no emotion' having the highest number of tweets this might have affected our model's accuracy in one way or another.
- Building a model on real-time data to detect consumer sentiments can be more useful and applicable to most brands instead of using data from years back.

THANK YOU!
QUESTIONS??