AMAZON CELLPHONE REVIEW ANALYSIS

GROUP 74

INTRODUCTION

Mobile phones have revolutionized the way we purchase products online, making all the information available at our fingertips. Reviews and ratings submitted by consumers became an integral part of the customer's buying decision process. The review and rating platform provided by eCommerce players creates a transparent system for consumers to take decisions and feel confident about it.
However, it is difficult to read all the feedback for a particular item especially for the popular items with many comments. In this project, we will attempt to understand the factors that contribute to classifying reviews as positive or negative
We will be using Natural language processing to analyse the sentiment (positive or a negative) of the given review. A sample web application is integrated to the model built.

OBJECTIVES

1. To visualize the raw data and make something meaningful out of it.
2. Know how to pre-process/clean the data using different data preprocessing techniques.
3. Know fundamental concepts and techniques of natural language processing (NLP).
4. Know how to build a web application using Flask framework.

PROJECT FLOW

This project contains the following Activities

- Data Collection.
  - Collect the dataset or create the dataset
- Text Preprocessing.
  - Import the Libraries.
  - Importing the dataset.
  - Remove Punctuations
  - Convert each word into a lower case.
  - Stemming.
  - Splitting Data into Train and Test.
- Model Building
  - Import the model building Libraries
  - Initializing the model
  - Adding Input Layer
  - Adding Hidden Layer
  - Adding Output Layer
  - Configure the Learning Process

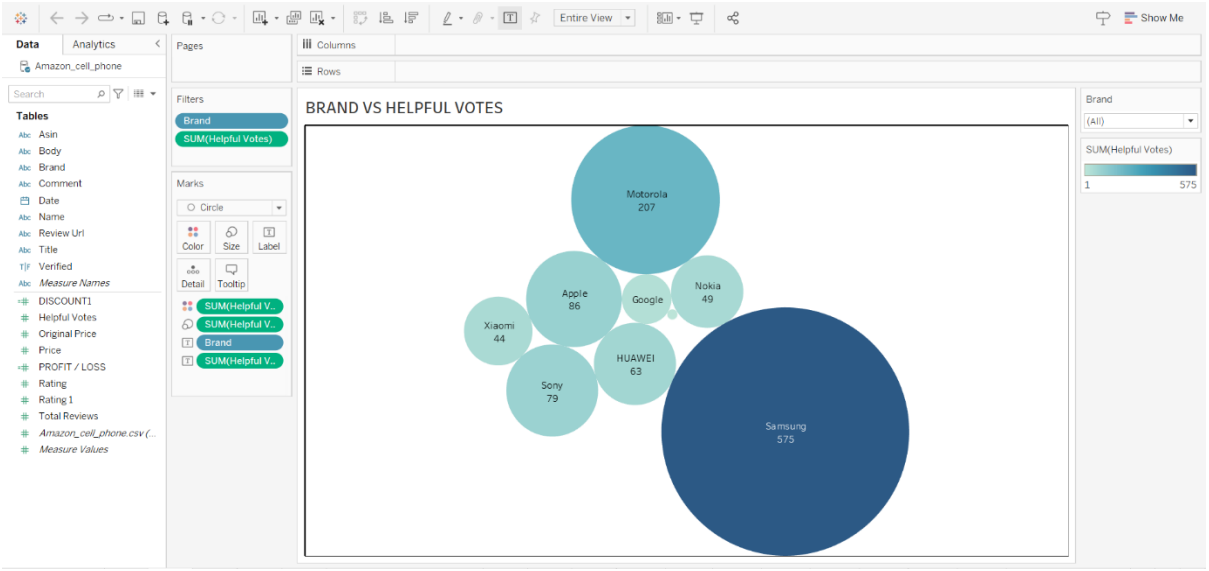
    - Training and testing the model
    - Optimize the Model
    - Save the Model
- Application Building
- Create an HTML file
- Build a Python Code

DATASET

https://drive.google.com/file/d/17One5FEv99DiKscqPxTFpVhbtvJmP4JH/view

VISUALIZATIONS

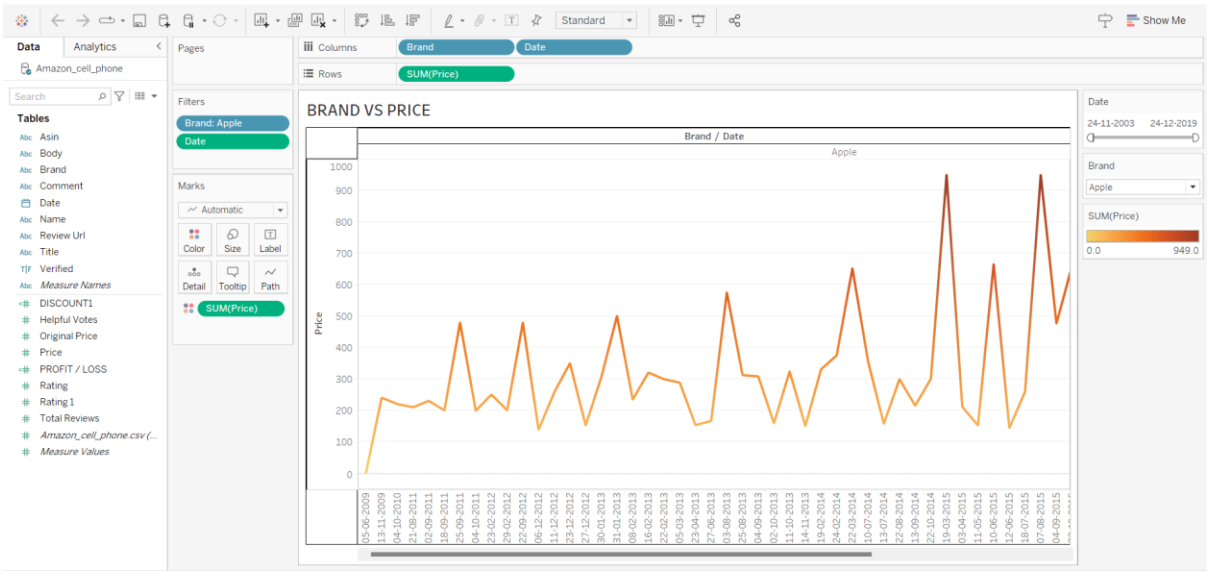
1. BRAND VS HELPFUL VOTES



This visualization helps us know the number of helpful votes each brand has. Bigger the size of the bubble more the helpful votes for that brand. And the number of helpful votes is also labelled. We still have one filter for the Brand where the user can choose the brand he wants to visualize and know the number of helpful votes.

https://drive.google.com/file/d/1-PzkLL0iW5FlKH1bydaBGncU7rXQXRjW/view?usp=drive_link

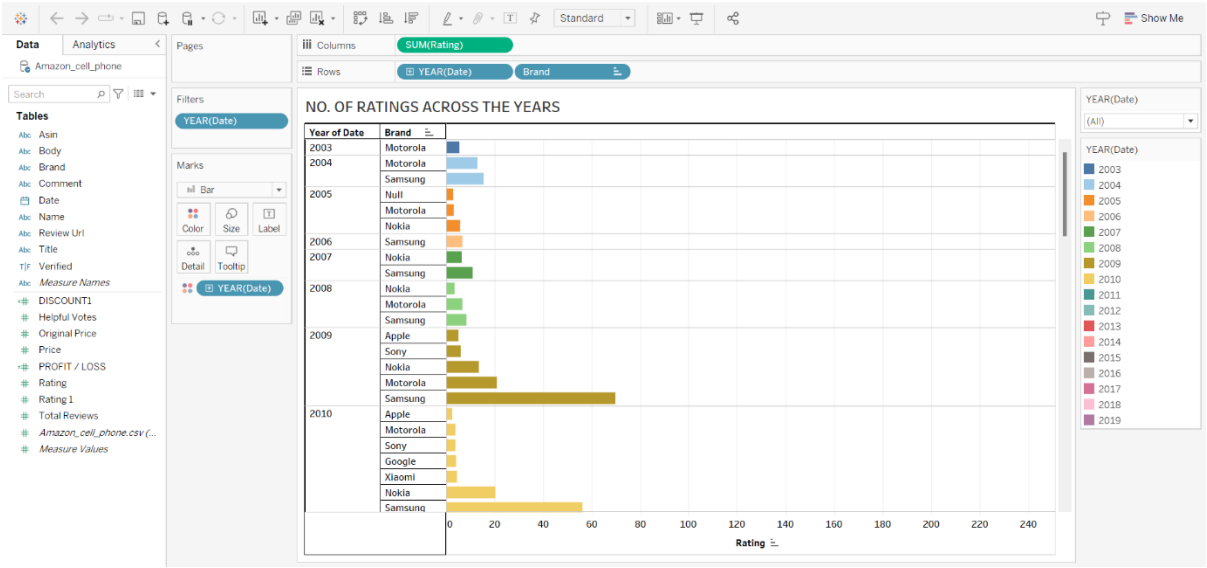
2. BRAND VS PRICE



The above visualization gives us the context of change in the price of the Brand across the years. As it is plotted with date and price. We have two filters one for the range of dates and the other one to select brands. So a user can select the brand he wants to visualize the price and still use the slider to know the price during that particular range.

https://drive.google.com/file/d/1CAnsGNY-jkzZZkkknLgivilYlyPXZ9O8/view?usp=drive_link

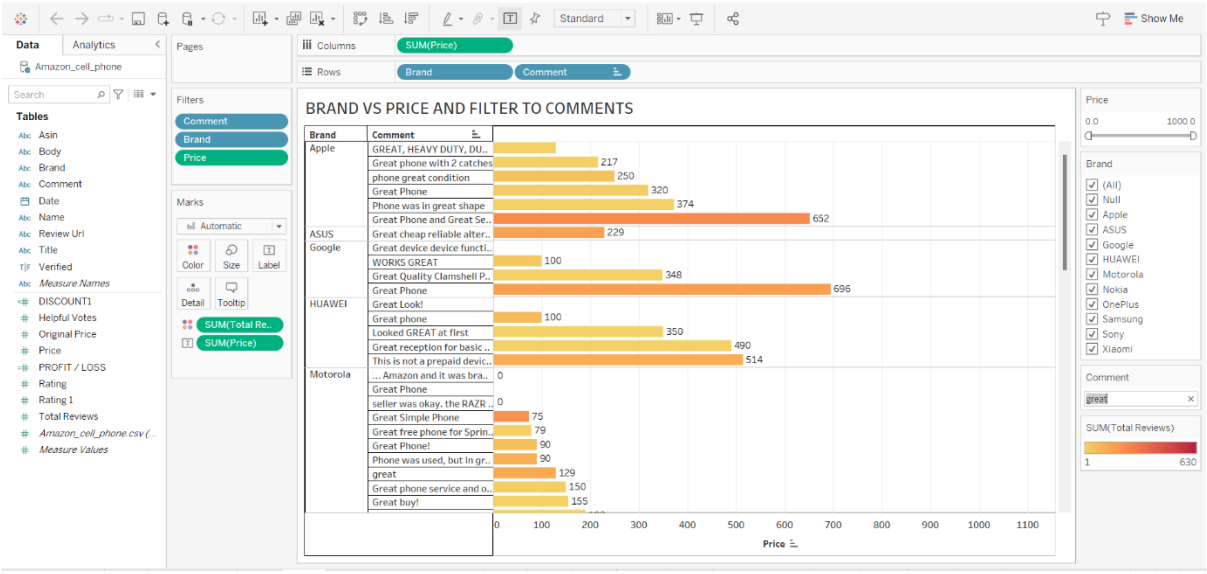
3. NUMBER OF RATINGS ACROSS THE YEARS



This visualization helps to find the number of ratings of different brands of phones over the years 2003 to 2019. Each year is represented by a specific colour and there is also a filter helpful to distinguish various years and solely focus on a particular year. It is useful to find out which brand got the most ratings across the years.

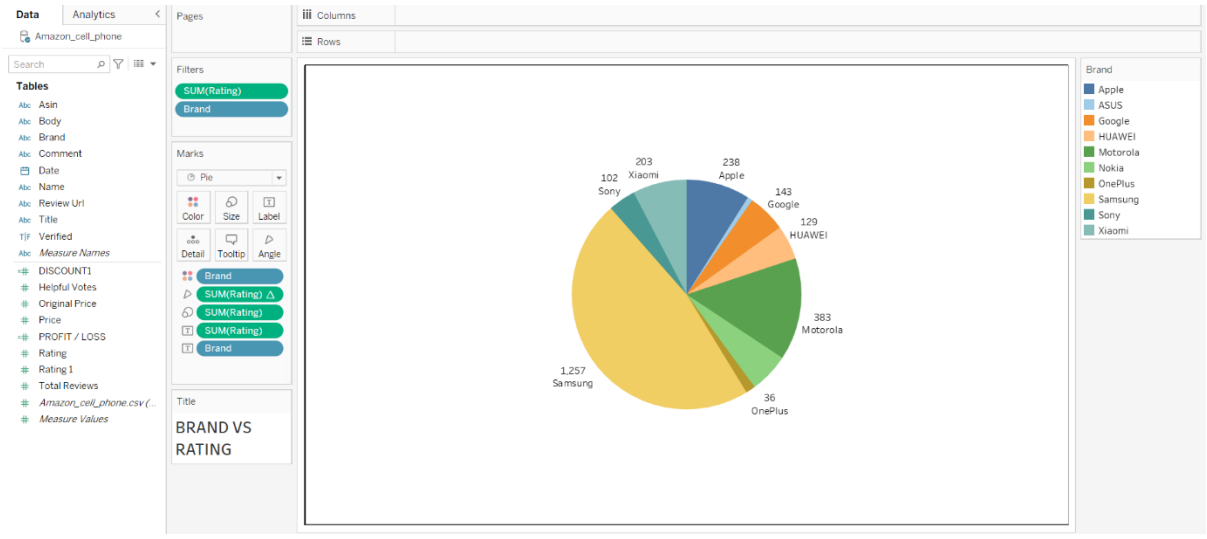
https://drive.google.com/file/d/1sip6_5DzEpemEzmYmCuO8f40sZrEQq27/view?usp=drive_link

4. BRAND VS PRICE AND FILTER TO COMMENTS



The above visualization has Brand and comment vs the price and has three filters applied, one to the comment, and the price amount is given as a slider with a range and a filter to the brand name. So, a user can type in a keyword and it shows all the comments with the keyword and the user can select the brands along with the price range he's looking for.

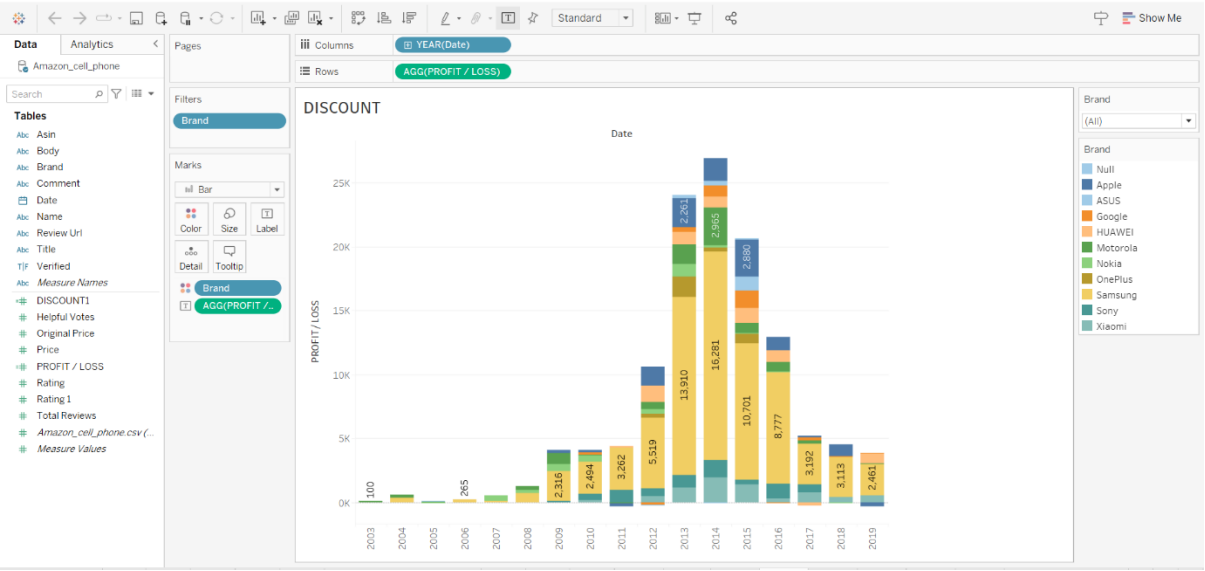
5. BRAND VS RATING



This visualization depicts a pie chart showcasing the total ratings given by the customers to a particular brand of a phone. This helps customers to make a purchase of the brand that is rated by most people. Each brand is identified with a specific colour.

https://drive.google.com/file/d/1xygqCiy4boTJgk8ezcjbqSUpGz_DmNs0/view?usp=drive_link

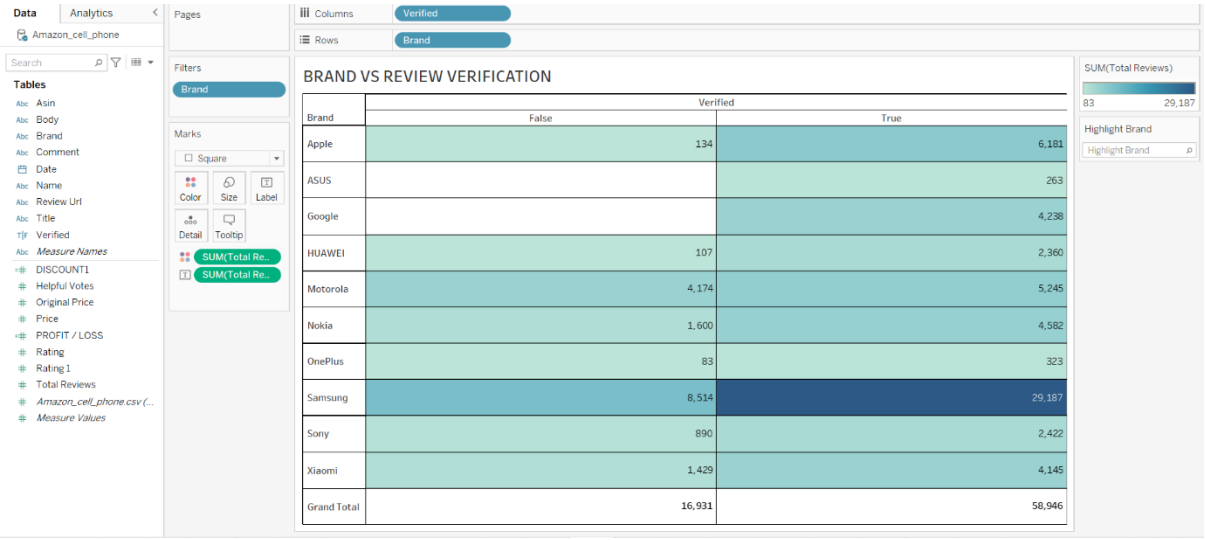
6. DISCOUNT



This visualization represents the various discounts given by different brands across the years. The discount is calculated using a calculated field by subtracting price from original price. The various colours indicate different brands across the cost and years.

https://drive.google.com/file/d/1pO9gweyzpKgLVpAJJxucXmh21bYov0ai/view?usp=drive_link

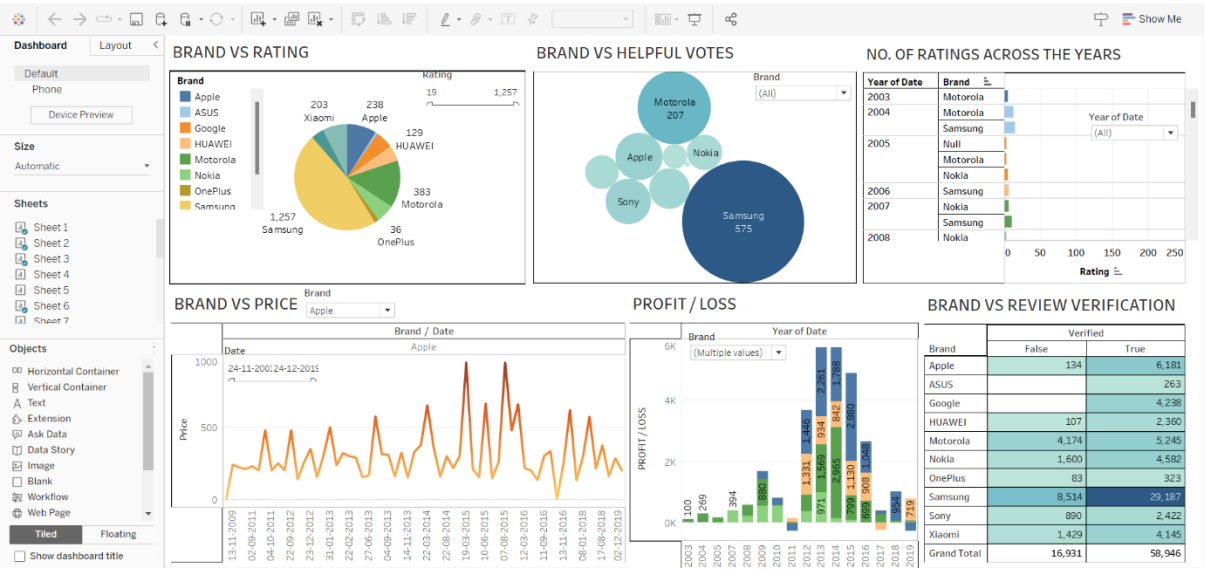
7. BRAND VS REVIEW VERIFICATION TABLE



This visualization helps customers to understand the authenticity of the reviews posted on Amazon. This will help them to choose widely. The darker gradient of colour represents more reviews. As we can see Samsung has not only highest number of reviews but also high number of verified reviews.

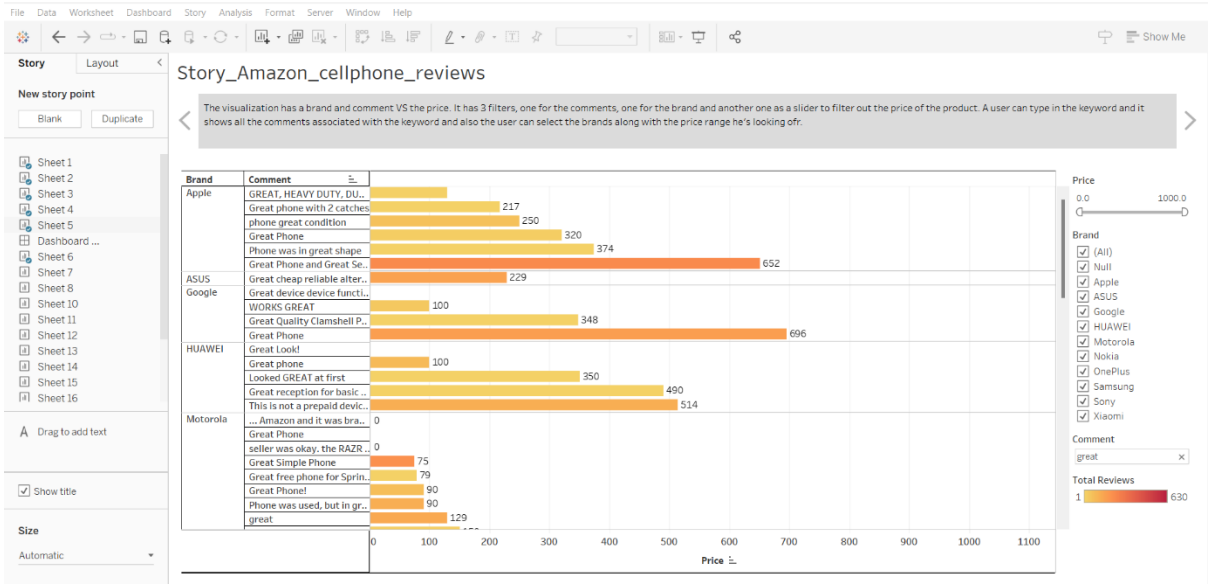
https://drive.google.com/file/d/16VxarBZsBcM0pXd-4EEAfQczIj3K_Rut/view?usp=drive_link

DASHBOARD



https://drive.google.com/file/d/1viohCnsV7bWSGBHooBlAvGiZFPG2vk4h/view?usp=drive_link

STORYBOARD

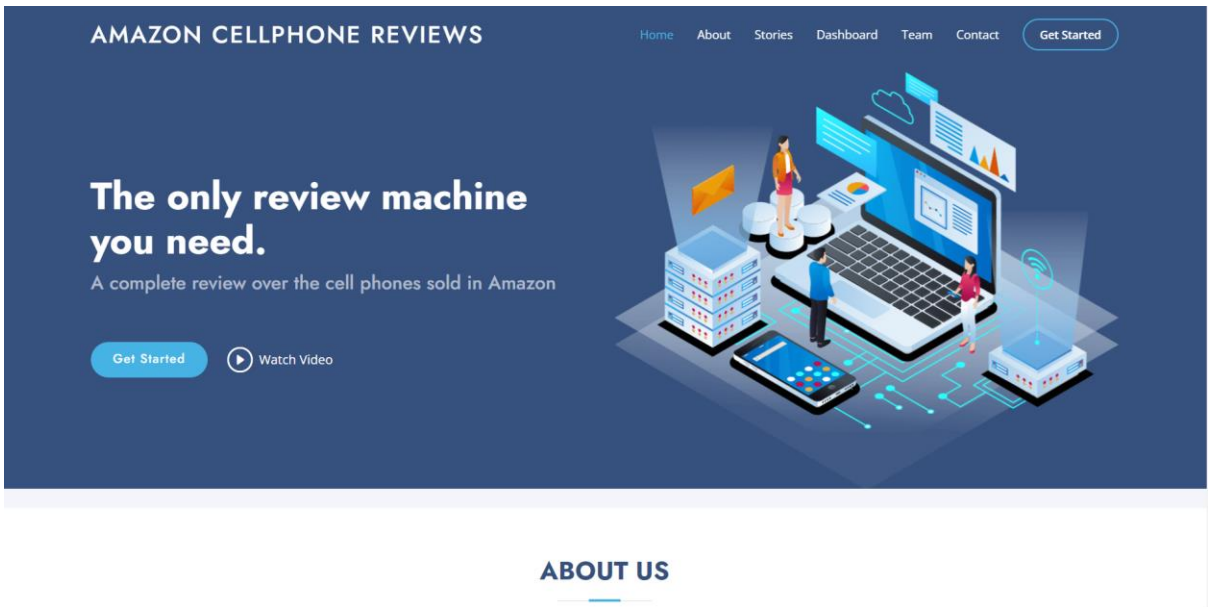


https://drive.google.com/file/d/1piRRgThPh9FlUYq03zuAnejtLSN7gnBx/view?usp=drive_link

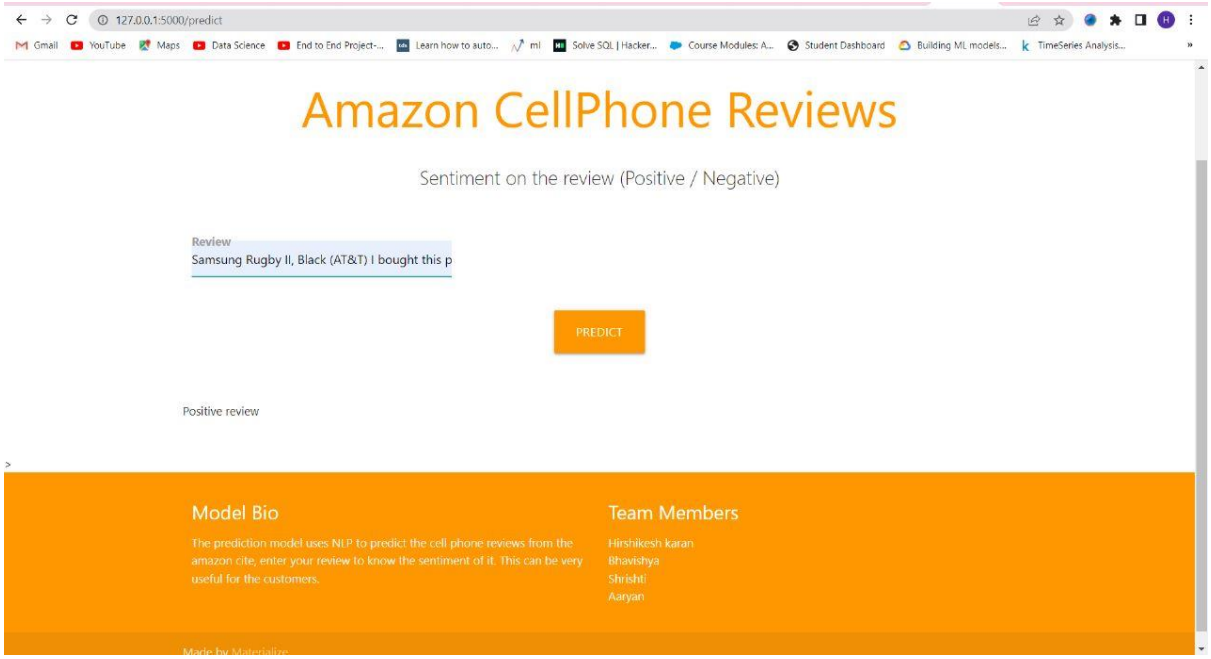
PUBLISHING ON TABLEAU PUBLIC

https://drive.google.com/file/d/10amVUfoilxLvg1oFrBS56Gy-EnUD-m8H/view?usp=drive_link

WEBSITE WITH DASHBOARD AND STORY INTEGRATION



SENTIMENT OF THE REVIEW ANALYSIS



NLP Sentimental Analysis is a branch of artificial intelligence that focuses on determining the sentiment or subjective opinion expressed in text data. It involves using machine learning and computational linguistics techniques to automatically analyse and classify the sentiment.

We have modified our model by adding different layers of neurons and dropout layers to the best test accuracy of around 93 percent.

Thus, helping in getting the work done of predicting if the review is positive or not.

ADVANTAGES

1. Customers often look at the pricing or the reviews of the mobile phones while making a purchase. So, it is important to have the reviews and pricing filtered accordingly.
2. Many reviews published on the internet might be true or might not. To provide the correct information to the customers, the authenticity of the reviews is tested and published.
3. It is also useful for the business because generating more authentic revies will help increase the sales.

APPLICATION

This project is intended to provide authentic information about the reviews posted on the Amazon to help customers figure out their purchases and make easy decisions. Whenever a product is to be bought, we often look up for the better products by analysing the reviews of the product, then compare the price, filter the price to our necessary range, check the reviews of the needed product and then make a purchase. This Tableau analysis is useful for the entire process as the dashboard and storyboard created have all these necessary filters for the pricing and the reviews.

A dashboard and storyboard are crucial to prevent the data from being misinterpreted. It contains various visualizations in a single page along with all the filters of the respective visualizations so as they make work easier without going to different sheets, looking at different visualisations and interpreting them.

Even businesses are at advantage with this analysis because the genuinely better products get sold after reviewing the reviews and the ones that are left out at the comparison can get a brief idea about what they lack and what they can provide the customers.

CONCLUSION

Analysing the data is very important so that we can have better business and consumer perspectives to make better sales and purchases. Reviews are one of the most important aspects that is looked at when it comes to buying a product right after the pricing. It is important to have good and genuine reviews. Amazon is one of the world’s best trusted websites for purchases as it provides ample reviews by the customers when compared to other websites. So, it should also make sure that those reviews are genuine to attract more customers and make better businesses. Thus, this analysis provides enough information for getting to know the review and testing the authenticity along with filtering the price ranges.

SUBMITTED BY -

1. AARYAN GUPTA – 20BCE10818
2. BHAVISHYA VARDHINENI – 20BCY10026
3. HIRSHIKESH KARAN – 20BEC0531
4. SHRISHTI SRIVASTAVA – 20BCY10133