**INTRODUCTION**

For this project, I will be evaluating the overall response of cell phone reviews of previous models of the iPhone from the dataset sourced from the Kaggle website. Source: <https://www.kaggle.com/datasets/grikomsn/amazon-cell-phones-reviews>

To be specific, this will be about iphone X/XE/XS/XR models. The objective is to solve the business problem of how to improve the quality of product further considering the customer reviews and help achieve better profits for the future products.

How do we improve the iPhone (X,XE,XS,XR models considered here) product?

Assuming data scientist role at Apple and to point out all concerns and crucial points of opportunity regarding previous iphones from Amazon review's datasets. The objective is that I will isolate and categorize the topic that is being mentioned through this dataset by picking up on certain key aspects/variables. There are currently eight different aspects that are being examined.

Variables/aspects: Appearance, Battery, Screen, Software, Music, Storage, Camera and Speed.

Sentiment analysis will be conducted and examined to truly understand the weak points of the phones. The idea is that once weak points are isolated, different departments can find out how to most effectively pool their resources to improve customer satisfaction leading to future sales and better profits.

**MILESTONE1:**

I will be analyzing what features of the phones work well and which ones don't. There are different variables and columns, which make up this project. Some of them include the ASIN number, the review title, review body, the date, the helpful review counter and prices paid. My goal is to perform thorough investigation and learn more about the phones. This is a type of practice that saves money, increases customer service satisfaction, and protects brand loyalty for a company. I believe that this is important for any company that cares about providing a responsible customer service experience, including Apple. From the graphs, I learned that I can trust the data because there is approximately 140 verified reviews when compared to about 5 unverified reviews. Also, it seems like there are generally positive reviews. However, in another graph, I found that the graphs do change from one another. This is helpful because it tells me which areas may need extra attention. Finally, I found a number of reviews by the different Item #.

Some of the charts generated from the data are given below:

**Chart

Description automatically generated**

**Chart

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**MILESTONE 2:**

There are a quite a number of unwanted columns from the original dataset. The columns being dropped are: image url, the url to the review, the price purchased, the name of the reviewer and other review from a duplicate rating, because they are not really required for sentiment analysis.

In addition, I have removed the brand and reviews category because the brand is the same for all rows, and the reviews doesn't give credence to the credibility of the review.

Post the creation of categorical columns through get\_dummies function, any values that are strings or text-based have been deleted because their usefulness have been extracted and dropping these columns would give clarity.

More visualizations have been created such as time series analysis, histograms, correlation maps, and linear regression. Coefficients have also been determined and calculated.

Some of the graphs generated as part of this milestone are given below which are around certain key aspects.

Battery:

Chart, bar chart

Description automatically generated

Camera:

Chart

Description automatically generated

Speed:

Chart, treemap chart

Description automatically generated

Storage:

Chart, bar chart

Description automatically generated

Software:

Chart, bar chart, treemap chart

Description automatically generated

Music:

Chart, bar chart

Description automatically generated

Score:

Chart, bar chart

Description automatically generated

**MILESTONE 3:**

Here, I have normalized the Price and Rating columns, feature reduction and showcased plots of linear regression on different aspects of the product reviews as part of build and evaluate the model process. Some of the key plots are shown below.

Rectangle

Description automatically generated with low confidence

A picture containing text

Description automatically generated

Rectangle

Description automatically generated with low confidence

**CONCLUSION**

The analysis on the topic of the Amazon review of the Apple iPhones was good. We found that overall, there is a modest increase over time from most topics as they increase, but an area that deserves the most amount of attention moving forward is that of the camera. However, it should be noted that within the last year of reviews, the biggest drop of review positivity scores come from both music and storage.

One fact that was consistent throughout all analysis is that almost always, an increase of subjectivity correlated with high positivity rankings. This means that if we can excite all customers about the iPhones that it would translate to higher customer satisfaction.This contrasts from the least popular aspect, the camera, which had a very lower average helpful rating. The scores appear to be steady as of recently, although the recent subjectivity scores from the last year appear to be positive, which could indicate that the positivity scores will increase should be steady with subjectivity. This prediction is based off a prediction rating of roughly .4, which is generally considered moderate in nature, especially given the p value of less than .001.