

# Sentiment Analysis on Electronics Customer Reviews on Amazon

## 1. Problem:

At Amazon, customer feedback is a vital component of the business model, guiding prospective buyers and shaping the reputation of products on the platform. Within the expansive Amazon Electronics section, reviews serve as a cornerstone for customer decision-making, influencing both perception and sales. However, sifting through the vast quantity of text-based reviews to discern overall sentiment is a daunting task for both customers and sellers. Our proposed project aims to address this challenge by developing a machine learning model capable of accurately predicting the sentiment of electronics reviews—classifying them as positive or negative.

The questions we seek to answer through this investigation are many: Can we automate the process of sentiment analysis to assist in the rapid and accurate categorization of review sentiment? How can sentiment trends be quantified and tracked over time to identify shifts in customer satisfaction? Furthermore, what correlations exist between review sentiments and sales performance, and can we leverage this information to predict future sales trends or to identify electronics that might benefit from targeted marketing strategies?

The insights derived from this model will have a direct business impact by enhancing the customer experience through improved search and recommendation algorithms, thus driving sales. It will provide valuable, actionable insights into customer satisfaction and areas for improvement. The implementation of this model will streamline operations, bolster customer engagement, and reinforce Amazon's commitment to customer-centric innovation.

## 2. Data set: <https://www.kaggle.com/code/sunil199/amazon-sentiment-analysis-model-comparison-2019/input>

## 3. Dataset Description:

### 4. Number of Rows:

- a. The dataset includes 34,660 rows, each representing a customer feedback entry for a specific product.

### 5. Number of Columns:

- a. There are 21 columns available in the dataset.

### 6. Types of Variables:

#### a. Integer Variables:

There are two columns with integer data, which are "reviews.numHelpful" and "reviews.rating". Both use the "Int64" data type to accommodate integer values.

#### b. Object Variables:

The remaining columns are of the object type, typically containing text or mixed data.

## 5. Anticipated Results

For this dataset, firstly, we will be performing **sentiment analysis** by using subjective reviews which express personal opinions while the objective reviews tend to be more factual. We will then aggregate the sentiments of the multiple reviews for a product to get an overall sentiment score for that particular product. Sentiment analysis will be performed here using the following columns: **reviews.text** and **reviews.title**. This sentiment analysis will highlight areas for product improvement or customer service enhancement.

Secondly, we will be performing **regression analysis** where we will predict product reviews or ratings based on other features such as **reviews.numHelpful** and **reviews.date**. This will predict how customers rate products and this is worth undertaking because the product ratings will allow companies to address issues and enhance customer experience. Thirdly, we will be performing **anomaly detection** which will be used to identify and catch unusual patterns or fraud in the customer reviews and interactions. This will allow us to reveal suspicious and fraudulent reviews. It is worth undertaking because it will maintain data integrity on the amazon website and people will gain in online reviews.

In summary, this project is worth taking because this reveals a lot of hidden insights about the amazon website and how reviews really impact the product sales.

## 6. Implication of our result :

Humans are inherently risk-averse creatures, wired by evolution to seek comfort and avoid danger. This inclination persists in our daily lives, where we encounter frequent risks. In this age of information abundance, we tend to overanalyze even the simplest of purchases.

Consider a scenario where a person spends at least 10 minutes reading product reviews before making a purchase. Negative reviews can significantly influence consumer behavior. Therefore, sellers must pay close attention to customer feedback to understand what works well and what needs improvement in their products.

By dedicating time and effort to assess product reviews, companies can gain valuable business insights that might otherwise remain hidden. If acted upon effectively, this can lead to the development of superior products and increased business success. Tesla, for instance, has thrived by leveraging customer feedback to promptly release software updates. Their reliance on word-of-mouth marketing, rather than paid advertising, illustrates the power of customer reviews. In my view, harnessing the potential of customer feedback and combining it with analytics can unlock untapped business insights for any organization.

## Appendix

## Loading data in python and summary statistics

Summary Statistics:

```
#Providing some Summary Statistics
df.describe()
```

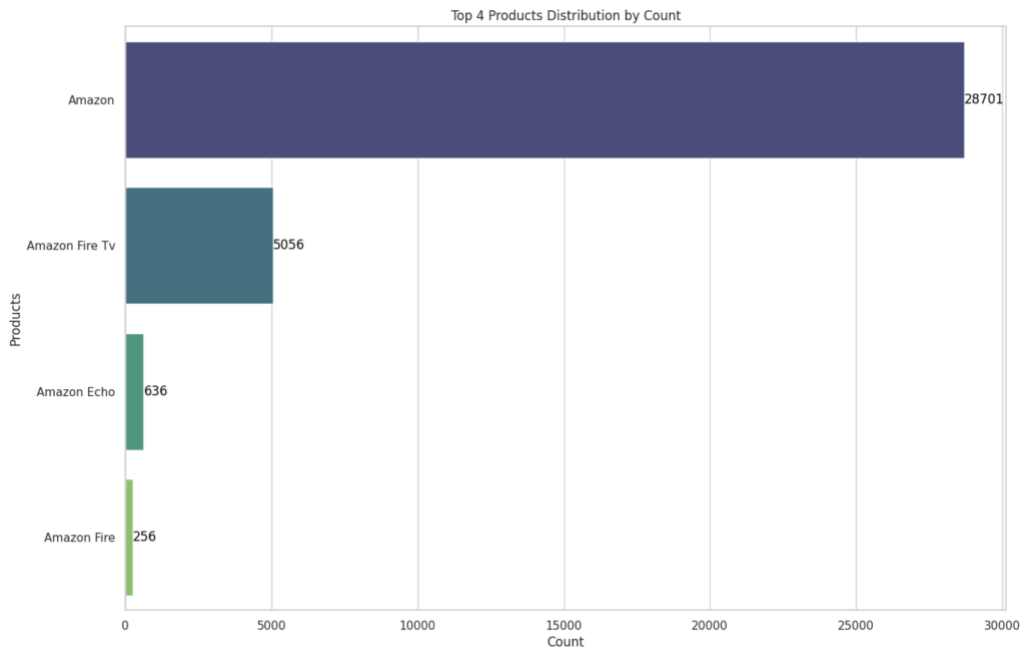
	reviews.id	reviews.numHelpful	reviews.rating	reviews.userCity	reviews.userProvince
count	1.0	34131.000000	34627.000000	0.0	0.0
mean	111372787.0	0.630248	4.584573	NaN	NaN
std	NaN	13.215775	0.735653	NaN	NaN
min	111372787.0	0.000000	1.000000	NaN	NaN
25%	111372787.0	0.000000	4.000000	NaN	NaN
50%	111372787.0	0.000000	5.000000	NaN	NaN
75%	111372787.0	0.000000	5.000000	NaN	NaN
max	111372787.0	814.000000	5.000000	NaN	NaN

Data Type of Columns:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34660 entries, 0 to 34659
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    34660 non-null  object
1   name                  27900 non-null  object
2   asins                 34658 non-null  object
3   brand                 34660 non-null  object
4   categories            34660 non-null  object
5   keys                  34660 non-null  object
6   manufacturer          34660 non-null  object
7   reviews.date          34621 non-null  object
8   reviews.dateAdded     24039 non-null  object
9   reviews.dateSeen      34660 non-null  object
10  reviews.didPurchase   1 non-null      object
11  reviews.doRecommend   34066 non-null  object
12  reviews.id            1 non-null      float64
13  reviews.numHelpful    34131 non-null  float64
14  reviews.rating        34627 non-null  float64
15  reviews.sourceURLs    34660 non-null  object
16  reviews.text          34660 non-null  object
17  reviews.title         34655 non-null  object
18  reviews.userCity      0 non-null      float64
19  reviews.userProvince  0 non-null      float64
20  reviews.username      34658 non-null  object
dtypes: float64(5), object(16)
memory usage: 5.6+ MB
```

Distribution of top 4 Products by Count:



Average number of Ratings by Brands:



Word Cloud of Reviews Column:

