# Title

Amazon Product Review Analysis

Md. Abdul Qadir Khan Soofi

Prakash J.M.

NIIT:DSFT15

Ms. Komilla Bhatia

June-July ,2023

# CERTIFICATE OF ORIGINALITY

This is to certify, that the research paper submitted by us is an outcome of my independent and original work. We have duly acknowledged all the sources from which the ideas and extracts have been taken. The project is free from any plagiarism and has not been submitted elsewhere for publication.

Name of Author (s):

Md. Abdul Qadir Khan Soofi

Prakash J.M.

Designation: Students

Affiliated Institution: NIIT

Title of the project: Amazon Product Review Analysis

E-mail: [qadirsoofi129@gmail.com](mailto:qadirsoofi129@gmail.com)

[saiprakash51999@gmail.com](mailto:saiprakash51999@gmail.com)

**ACKNOWLEDGEMENT**

­­ I would like to acknowledge that my assignment has been completed and I am ensuring that this was done by me and not copied.

In this accomplishment, I would like to express my special gratitude to all my teachers and most importantly our mentor Ms. Komilla Bhatia of NIIT, without their guidance and feedback it is not possible to complete this assignment.

Finally, as one of the team members, I would like to appreciate all my group member Mr. Prakash J.M. for their support and coordination, I hope we will achieve more in our future endeavors.

Md. Abdul Qadir Khan Soofi

Prakash J.M.

Contents

[Title 1](#_Toc140157989)

[CERTIFICATE OF ORIGINALITY 2](#_Toc140157990)

[INTRODUCTION 5](#_Toc140157991)

[ABOUT DATA 6](#_Toc140157992)

[PROCEDURE 8](#_Toc140157993)

[Imported data: 8](#_Toc140157994)

[Cleaning the data: 8](#_Toc140157995)

[Sentiment Analysis: 8](#_Toc140157996)

[Merging the two datasets: 9](#_Toc140157997)

[EDA in PYTHON / TABLEAU: 9](#_Toc140157998)

[6.Time Series Forecasting: 9](#_Toc140157999)

[ANALYSIS AND CALCULATIONS 10](#_Toc140158000)

[Scope of Improvement: - 11](#_Toc140158001)

[Bibliography and References:- 12](#_Toc140158002)

# INTRODUCTION

Reviews tell which products are trending, what’s in demand, what’s no longer relevant, how products and competitors are doing and much more. As product reviews are often the deciding factor for customers in buying a product, it’s important to have an automated system for monitoring them.

Our main agenda in this review analysis is to find the reasons for negative reviews from customers, the reasons for customer attrition, brands having negative responses, trends in negative response over years, classification of responses based on verified and non-verified ids, Ids giving negative responses from non-verified accounts, Sales response over years, Brands having high sales as well as positive response

Identifying customer reaction and sentiment in order to better understand and improve the business, exploring the data on the basis of sentiment and ratings, forecasting customer sentiments to provide valuable insights into understanding future trends.

trends.

# ABOUT DATA

DataSource:<https://jmcauley.ucsd.edu/data/amazon/>

We have two datasets

1. Toys and Games
2. Office Products

Toys and Games:

There are 1 lakhs observations

There are 11 columns

Office Products:

There are 1 lakhs observations

There are 19 columns

The products reviewed between years 1996 and 2018.

# PROCEDURE

## Imported data:

We imported both Toys&Games and Office products datasets using the Pandas library where the data was in JSON format (pd.read\_json)

## Cleaning the data:

After seeing the data we could there columns which are not useful for analysis so we dropped those specific columns ,In cleaning the data we removed punctuation , hashtags ,all non-letters and non-spaces , underscore , URLs , HTML tags , all hyphens except between two words , multiple spaces and lowercase and also to improve the overall quality we remove stop words and convert words into their base word using lemmatization on the review text column with the meta data of both datasets we found that there are irregular and null observations found and Defined a function to filter the observations using regex match function and removed unnecessary columns.

## Sentiment Analysis:

For doing sentiment analysis in the review text column, we found the new library named TextBlob, which gives us the polarity score from the review-text column and from that polarity score we can define the sentiment of the review. To attain this, first, we created a function for finding a polarity score and, with that score giving a range for it, we got the sentiment of review and stored the observations of both polarity and sentiment in separate column

## Merging the two datasets:

After finding the sentiment analysis, we found that the ASIN column was present in both datasets. We merged both cleaned data into one and converted it into CSV format for doing EDA in tableau

## EDA in PYTHON / TABLEAU:

By using python for visualization, we did a bar-plot, count-plot and pie-chart in the EDA of Python, whereas in tableau we did a Donut chart, lollipop chart for forecasting, line chart and bar-chart

## 6.Time Series Forecasting:

As the first step, we imported the libraries required for doing modelling, then to do the modelling for positive sentiment through time, we created a dataset with polarity and review time from the main data where sentiment is positive. Then , converting the date feature into a standard date-time feature and set the date-time feature as index of data, and decomposed the data into components of time series to analyze the data, then we did a stationary check on the data for plotting the pacf and acf graph so that we get values of p and q from the graph after this process we build the model based on the conditions then fitted the model and by Using Ljung box method to find the goodness of the model and predict the output from unseen input and plot the graph of forecast, then we did this same process for negative sentiment.

# ANALYSIS AND CALCULATIONS

There were 3 main analyses that were done. They are: -

1.Sentiment Analysis.

2.Time Series Analysis.

3.Time Series Forecasting.

1.Sentiment Analysis: -

* This was done using the TextBlob Library.
* We created a new column named polarity which was used to give polarity scores to the reviews depending on which they were classified into positive, negative and neutral.

2.Time Series Forecasting: -

* It was used to analyze seasonal trends and reasons for their occurrences.
* Decomposition graph was plotted and seasonality and trend were analyzed.

# Scope of Improvement: -

* The complaints of the customer’s needs to be categorized such as complain related to packing, product quality, shipping delivery executive behavior etc. So that it would be easier to understand customers issues in a better way.
* Details such as exact product name should be provided to analyze what are the leading products from each brand and what are the products that are generating negative response w.r.t. the brand.
* Also, details regarding the courier company should be provided to understand issues related to supply.
* Information regarding the city ,state/province of both the customer and buyer should be provided as it would be useful in a lot of ways such as understanding the issues related to delivery, effect of climate (if any) on the products etc.

# Bibliography and References:-

TextBlob:-

* Official documentation: <https://textblob.readthedocs.io/en/dev/>
* Uses of TextBlob:

1. <https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>
2. <https://www.analyticsvidhya.com/blog/2021/10/making-natural-language-processing-easy-with-textblob/>

* How to add polarity and how is it useful: <https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>

Wordcloud:-

* Official Documentation: <https://pypi.org/project/wordcloud/>
* About Wordcloud: <https://www.geeksforgeeks.org/generating-word-cloud-python/>
* How to build a Wordcloud:-

1. <https://www.analyticsvidhya.com/blog/2021/05/how-to-build-word-cloud-in-python/>

<https://www.projectpro.io/recipes/create-word-cloud-python>

Auto-EDA:-

1. Sweet-viz:-

* Official Documentation: <https://pypi.org/project/sweetviz/>
* About SweetViz: <https://www.analyticsvidhya.com/blog/2021/05/sweetviz-library-eda-in-seconds/>

1. Pandas-Profiling:-

* Official Documentation: [The Python Profilers — Python 3.11.4 documentation](https://docs.python.org/3/library/profile.html)
* About SweetViz: <https://www.geeksforgeeks.org/pandas-profiling-in-python/>