

AMAZON- PRODUCT REVIEW ANALYSIS


(OFFICE PRODUCTS AND CD's)

HIRAN S NAIR

VINAY RAWAT



Introduction



In this project we are using the amazon review data to examine the emotions and opinions expressed by the customers also discovering strategies to retain and satisfy existing customers, anticipating what future reviews must hold, projecting how customer sentiment may evolve in the coming years and their by grouping them based on their preference and feedback.



BUSINESS OBJECTIVE

- SENTIMENT ANALYSIS ON REVIEWS
- CUSTOMER RETENTION
- PREDICTION ON NEW REVIEWS
- SENTIMENT FORECASTING
- CUSTOMER SEGMENTATION

Why product Reviews are important ?

- Product reviews offer insights on quality, performance, and user experience.
- They are a trusted source for potential customers.
- Reviews reveal both positive and negative aspects of a product.
- They help shoppers assess risks and set realistic expectations.
- Reviews foster a sense of community and trust among consumers.
- They contribute to brand loyalty and influence purchasing decisions..



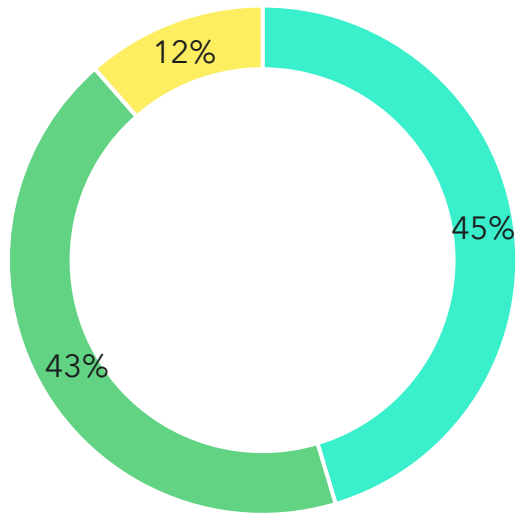


About the data

- We had taken 4 datasets, namely:
 - meta_CDs_and_Vinyl
 - reviews_CDs_and_Vinyl
 - reviews_Office_Products
 - meta_Office_Products
- Two product review file, and two product meta file, after importing all the data, the review files and meta file have been merged, respectively.

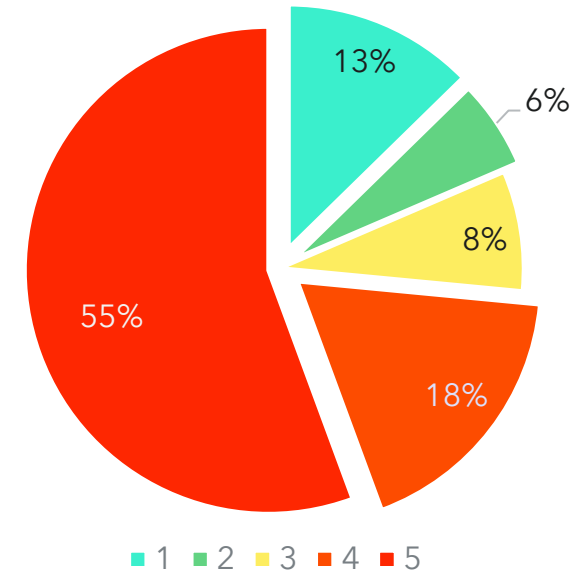
SENTIMENT ANALYSIS OF OFFICE PRODUCTS

Sentiment



■ Neutral ■ Positive ■ Negative

Rating

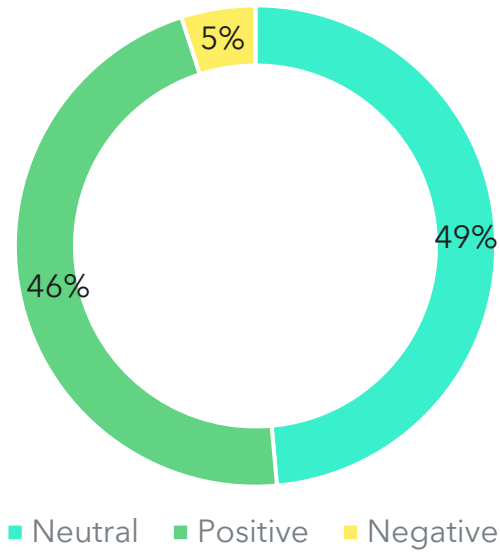


■ 1 ■ 2 ■ 3 ■ 4 ■ 5

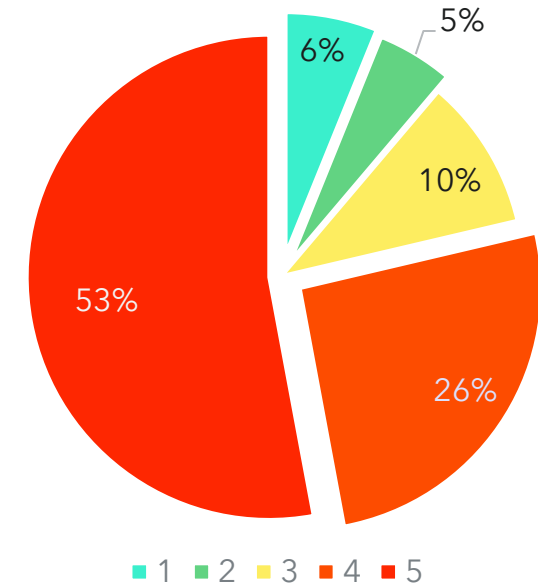
- 12% of office product's review is negative .
- 43% and 45% respectively of Positive and Negative Reviews.
- Similarly 55% had given 5 star rating and only 13% given 1 star rating.

SENTIMENT ANALYSIS OF CD AND VINYL

Sentiment

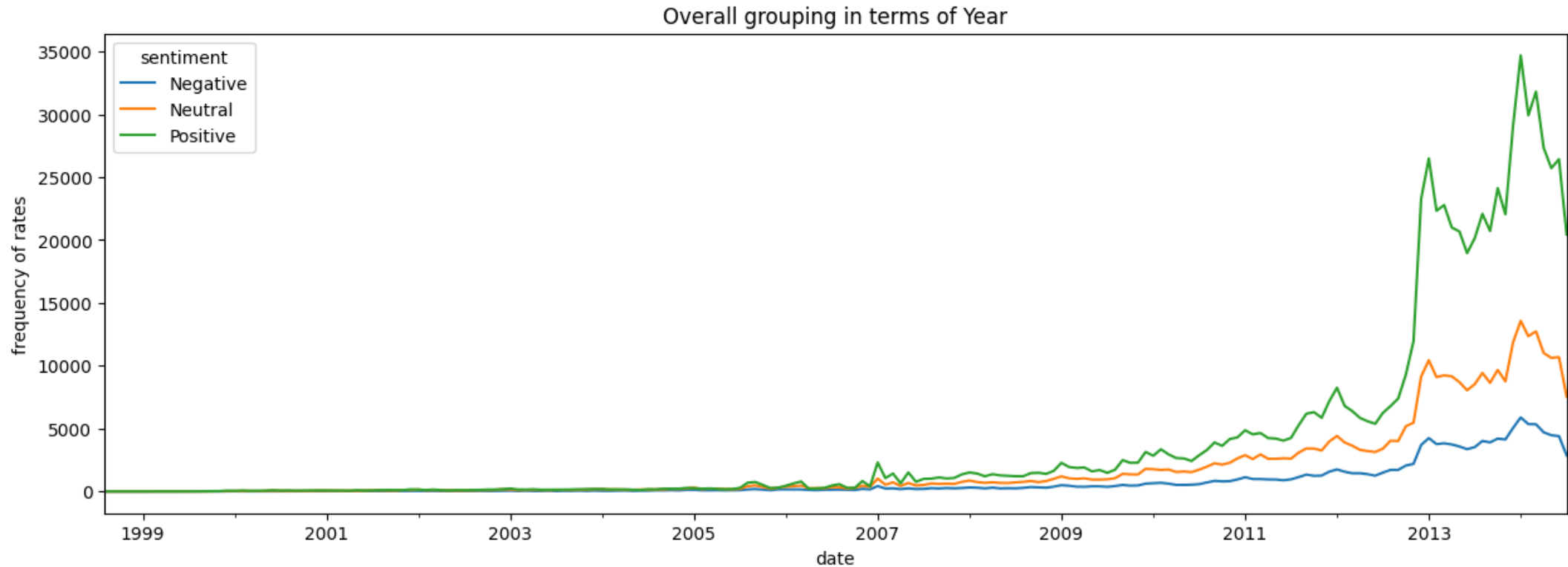


Rating



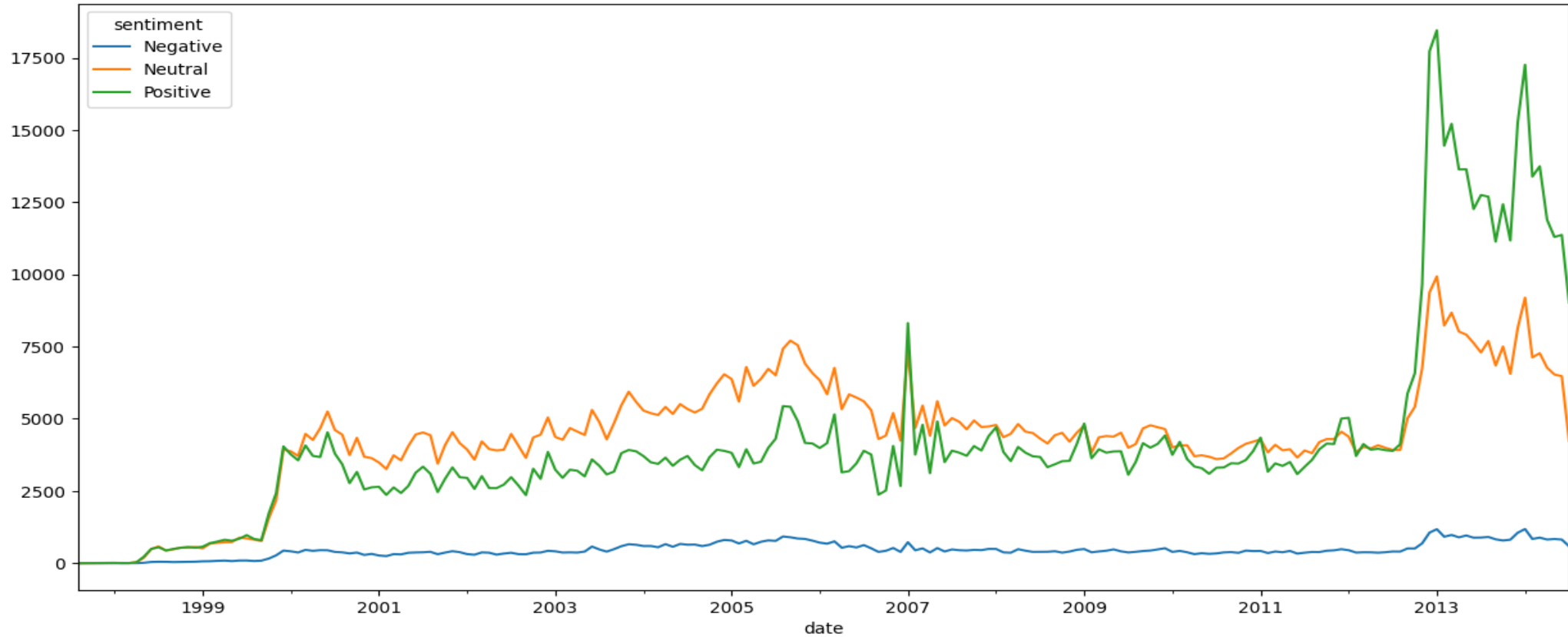
- 5% of cd and Vinyl's review is negative .
- 46% and 49% respectively of Positive and Negative Reviews.
- Similarly 53% had given 5 star rating and only 6% given 1 star rating.

Sentiment over the year-office products



- Customers started buying office products mainly after 2008.
- It have a gradual increase after 2013

Sentiment over the year- CD and vinyl

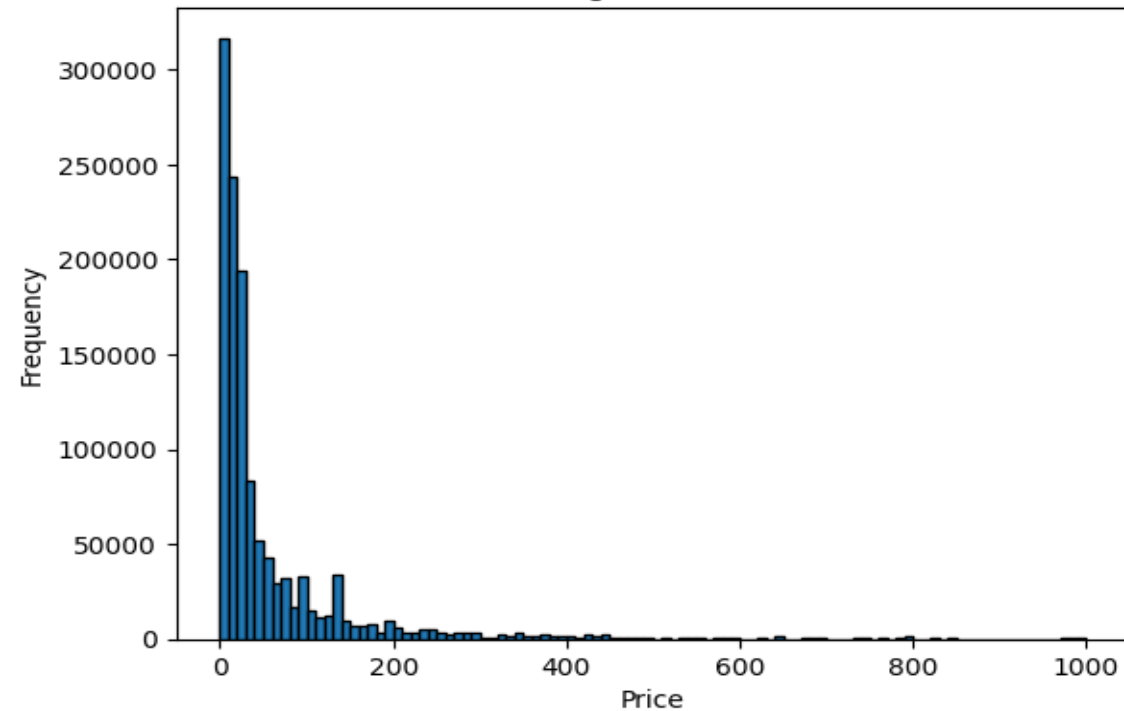


- Decent amount of sales for cd and vinyl from the initial years itself as it is like most demanded product.
- Negative sentiment is stable despite of huge spike in sales

How is price of products distributed?

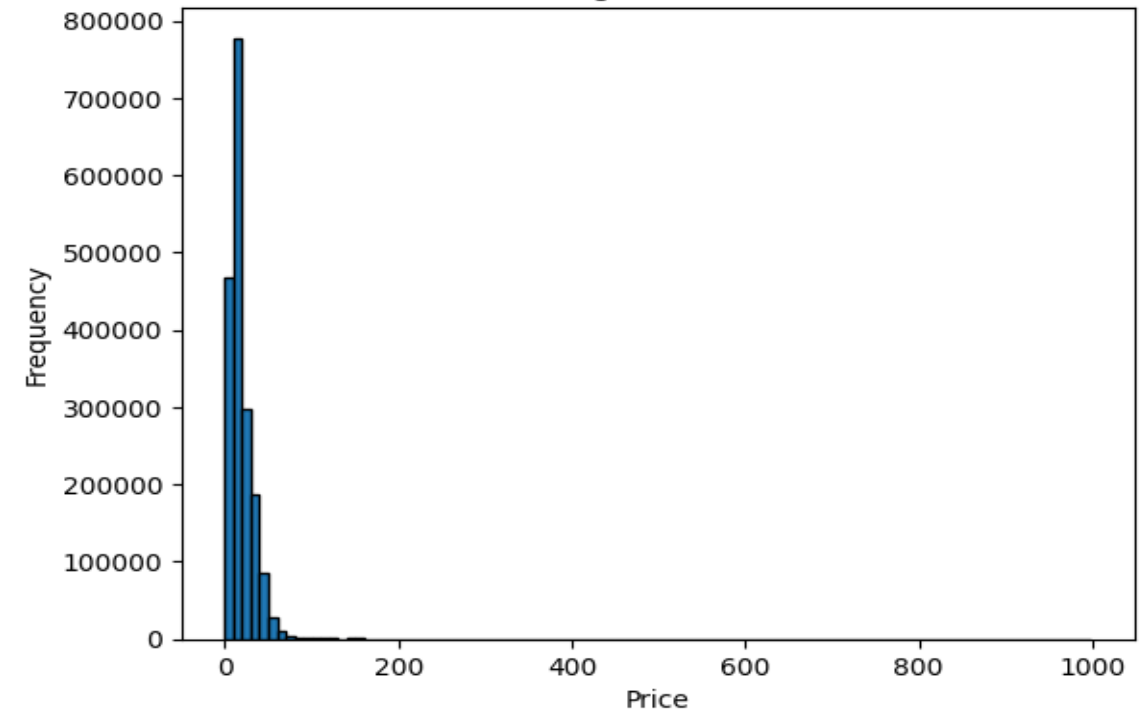
For office products

Histogram of Price



For CD and vinyl

Histogram of Price



- Most of the products came under the price of 100.
- Office products have more premium priced products than the cd and vinyl

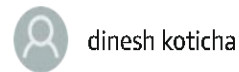
How fake reviews & promotional reviews affect customers?

Fake and promotional reviews can impact negatively to the business by creating;

- **Loss of Trust**
- **Misleading Decisions**
- **Reputation Damage**

How to churn out those ID's creating fake and promotional reviews ?

we can, by using the 'helpful' column in the data.



★★★★☆ **Digital safe standard**

Reviewed in India on 27 August 2023

Size: 41 liters | **Verified Purchase**

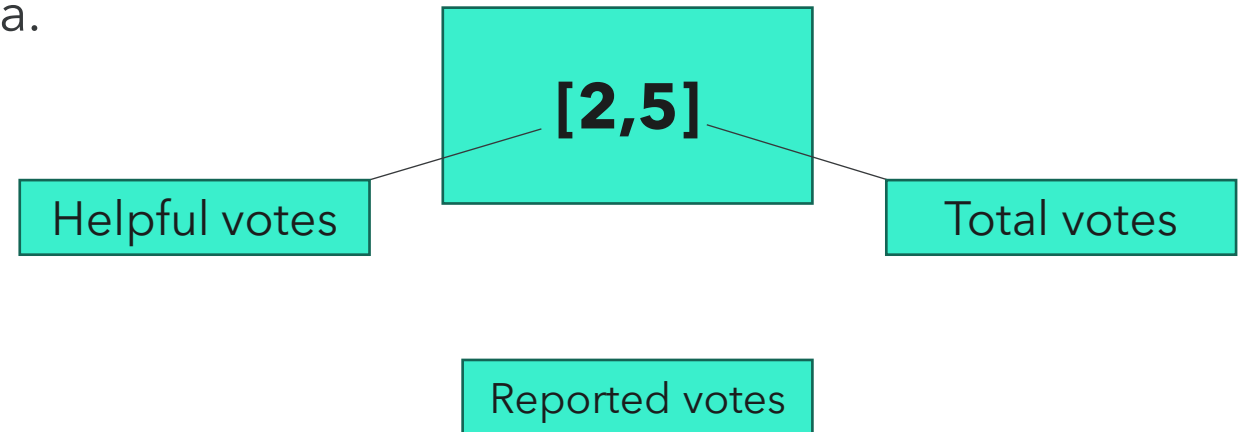
Good value for money. It's not fire proof and it's not biometric.

Instruction for use and installation is clear.

Installation bolts are very short need at least 3" long. So you will have to purchase them separately.

Helpful

Report



Customers that we need to churn out

s.No	ReviewerID	reviewer Name	unhelpful votes
1	A1M7VWY3SZ0Z6J	Intelligent Woman	395
2	A3GH4XY4VGE0TX	J. Day	375
3	A220XA9ZV9RRUY	da hammer	323
4	ACPCRJICJ024Z	Hay-Man	308
5	AA63JPV7NFRLI	J. Goodwin	297
6	A1SQENH064JLDC	Romneygirl	272
7	AVZKGXVR11V8	Momof4boys	261
8	A3I7LWJHMTVKJ9	Theresa Mayer "Bama Girl"	258
9	A30E269AJSBB3N	KS	235
10	A3PB5ZRSVRV550	Carla R Aannerud	232

- We had taken ReviewerID who are having more than 80% of total vote as unhelpful/reported.
- The table here shows top 10 Review-ID and there name.
- By churning out these Customers Amazon can regain the customer trust .
- Customers won't mislead into wrong purchasing decision thus generating happier customers.

*a separate csv for this is created

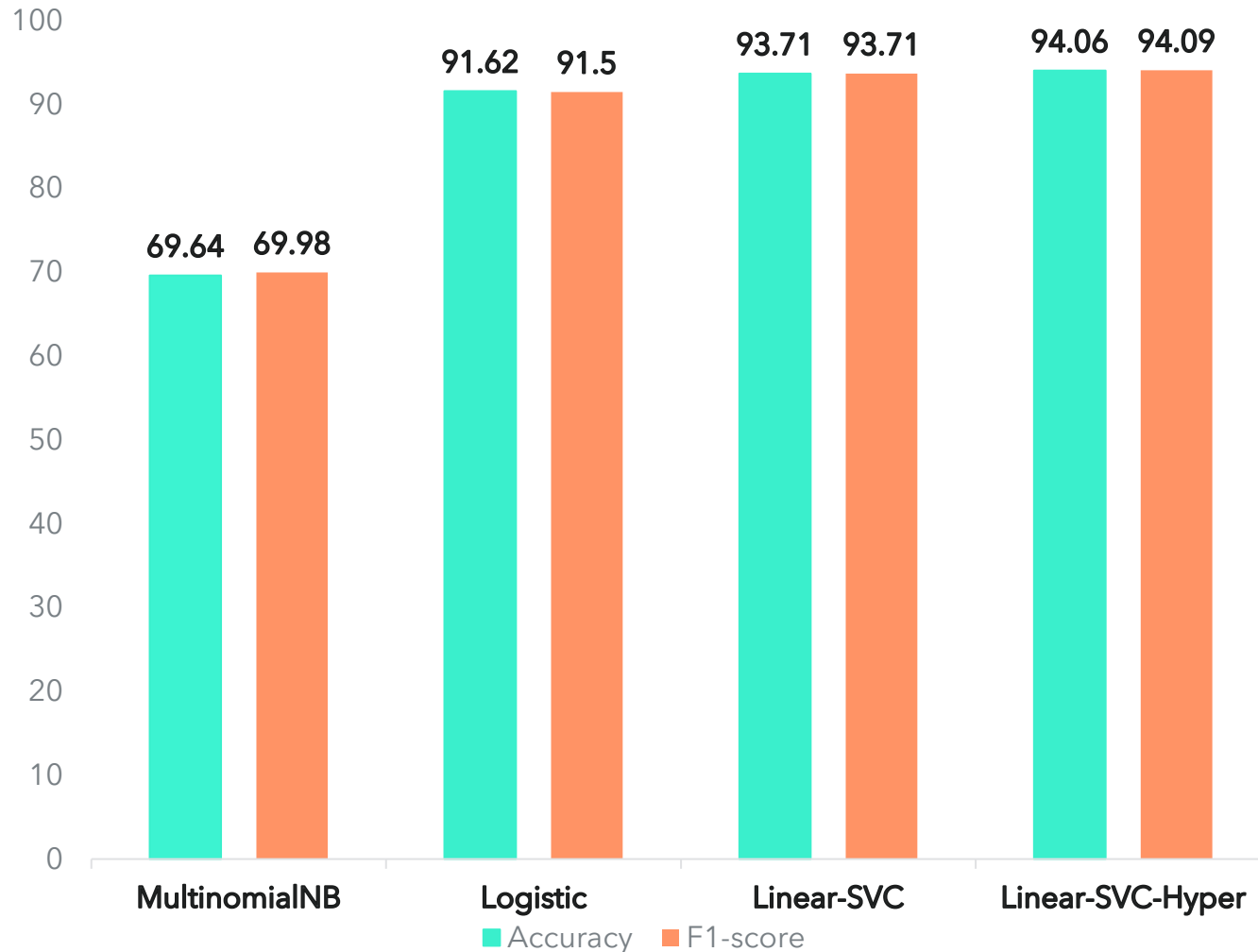
Creating a Classification Model.

- we've developed three distinct models leveraging Multinomial Naïve Bayes, Logistic Regression, and Linear Support Vector Classification (SVC) techniques.
- We harnessed the power of TF-IDF (Term Frequency-Inverse Document Frequency) for text vectorization.
- Bellow table shows how these models performed;

Metric	MultinomialNB-TFIDF	Logi-TFIDF	LinearSVC-TFIDF	LinearSVC-TFIDF Hyper-tuned
Confusion Matrix	[[128375 36563 2558] [33770 112219 21522] [6117 49554 103826]]	[[163012 4354 130] [15921 138114 13476] [79 7438 151980]]	[[165056 2370 70] [10501 145211 11799] [50 6294 153153]]	[[165373 2078 45] [8913 147021 11577] [45 6690 152762]]
Accuracy	0.6964958827431122	0.9162837914354586	0.9371410544707424	0.9406516428582984
F1-Score	0.6998616241696095	0.9150623070709138	0.9364367382792371	0.9400831329249217

Selecting The Best Classification Model.

Model performance

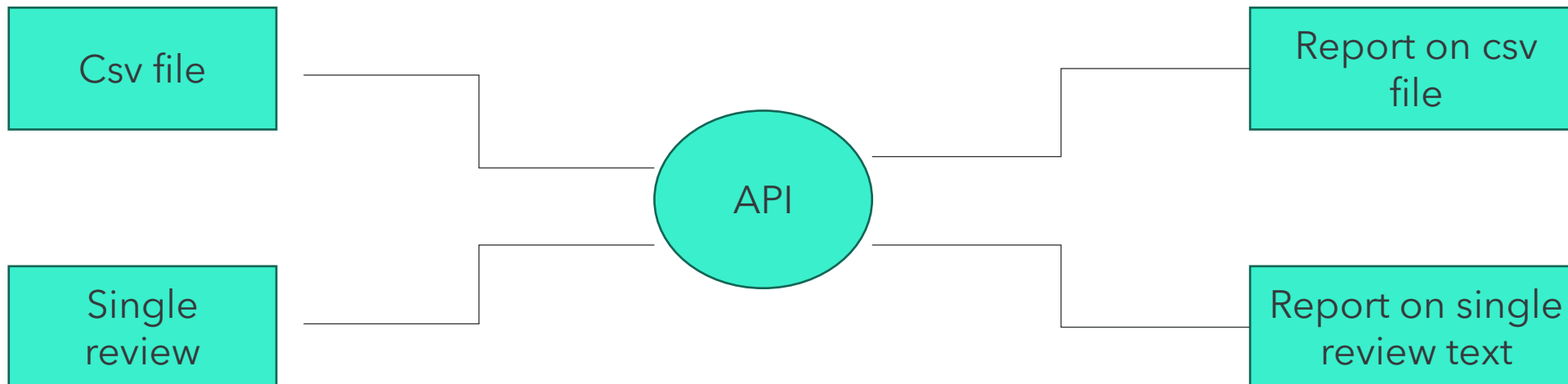


- Hyper-tuned Linear SVC model had the best results comparing all the others.
- MultinomialNB model takes less time for prediction.
- Hyper-tuned Linear SVC model takes less time than Logistic
- As Hyper-tuned Linear SVC is the best model ,selecting that as final model

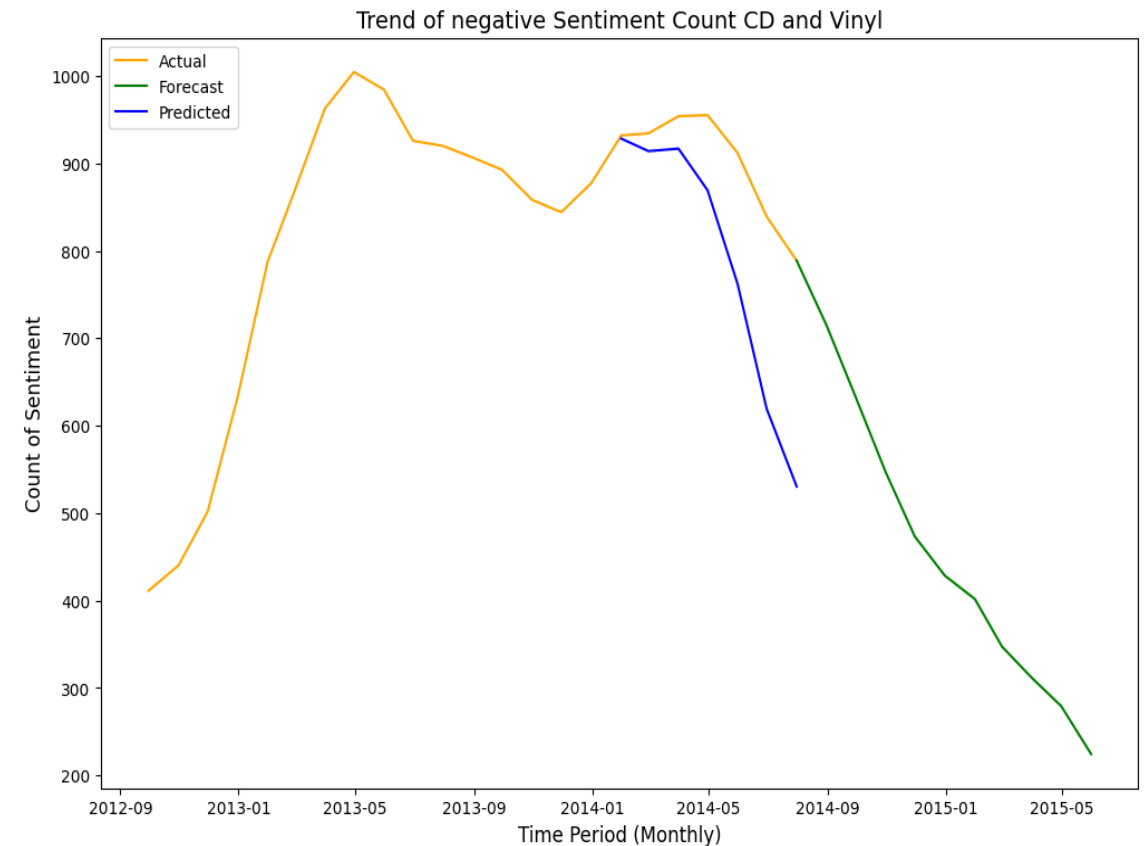
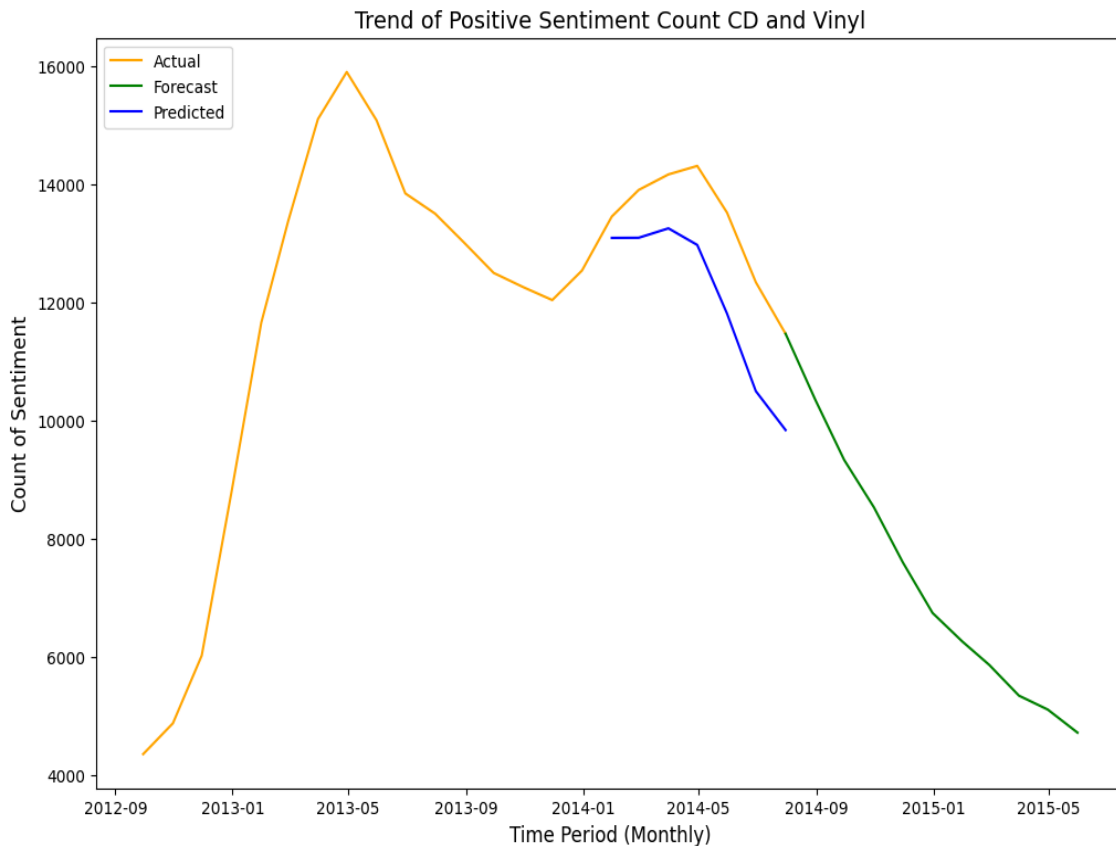
An API for sentiment analysis

This API work in 2 ways they are;

- By uploading a complete csv-file, it can return a summary of sentiment ,a pie chart and a bar graph of overall ratings.
- It can also use for individual sentiment analysis report.



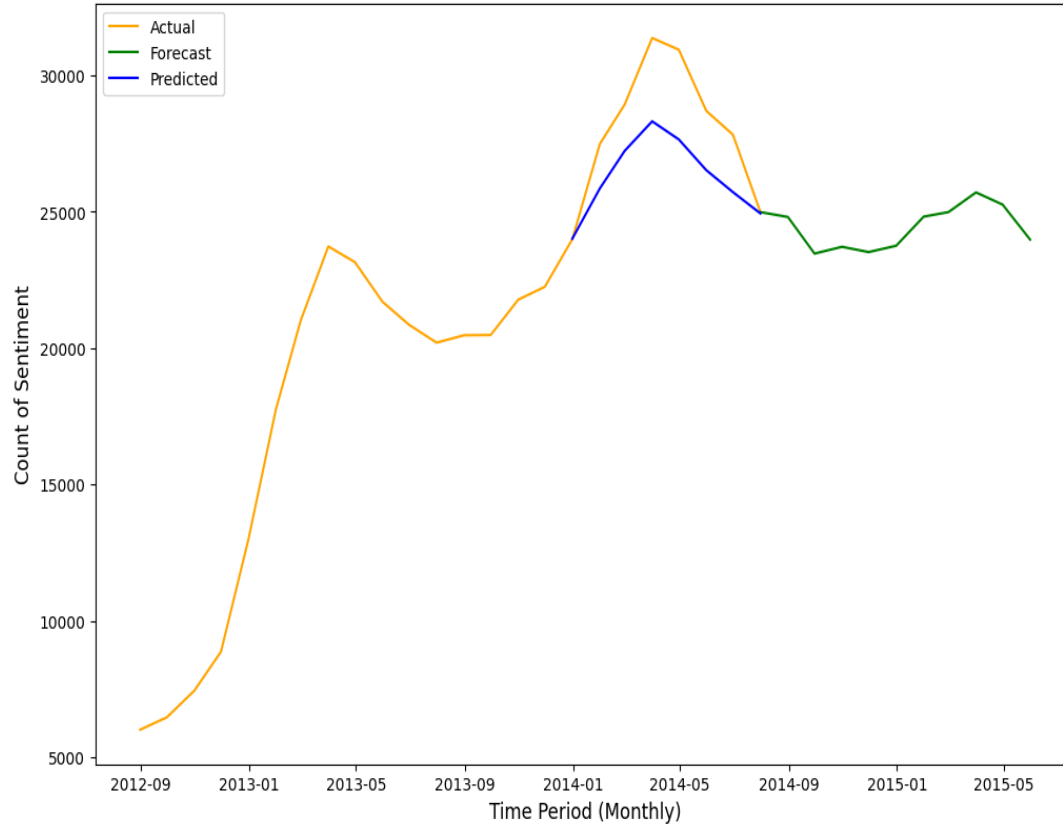
Time-series Forecast on CD and Vinyl



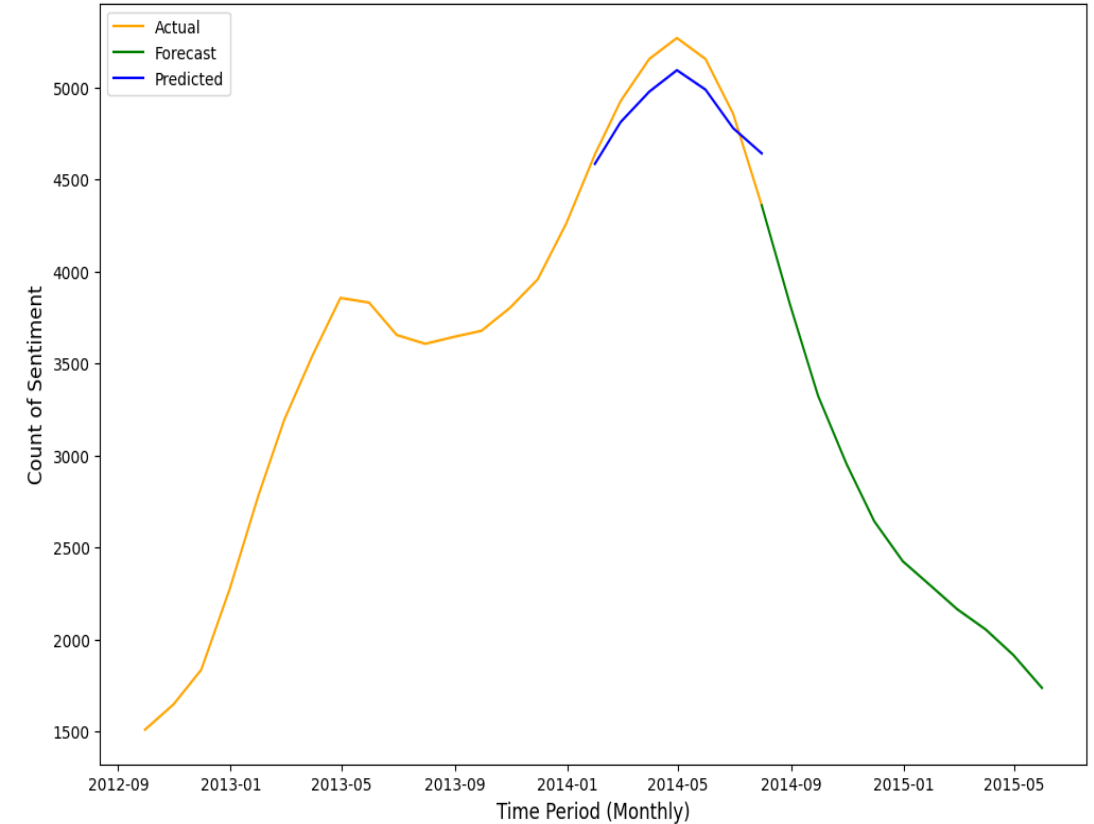
- It is predicting negative growth for both positive and negative reviews, that means the sales will go down drastically
- It is expected to go down as the CD trend is getting over and digital platform like Amazon Music is getting more popular.

Time-series Forecast on Office Products

Trend of Positive Sentiment Count office products



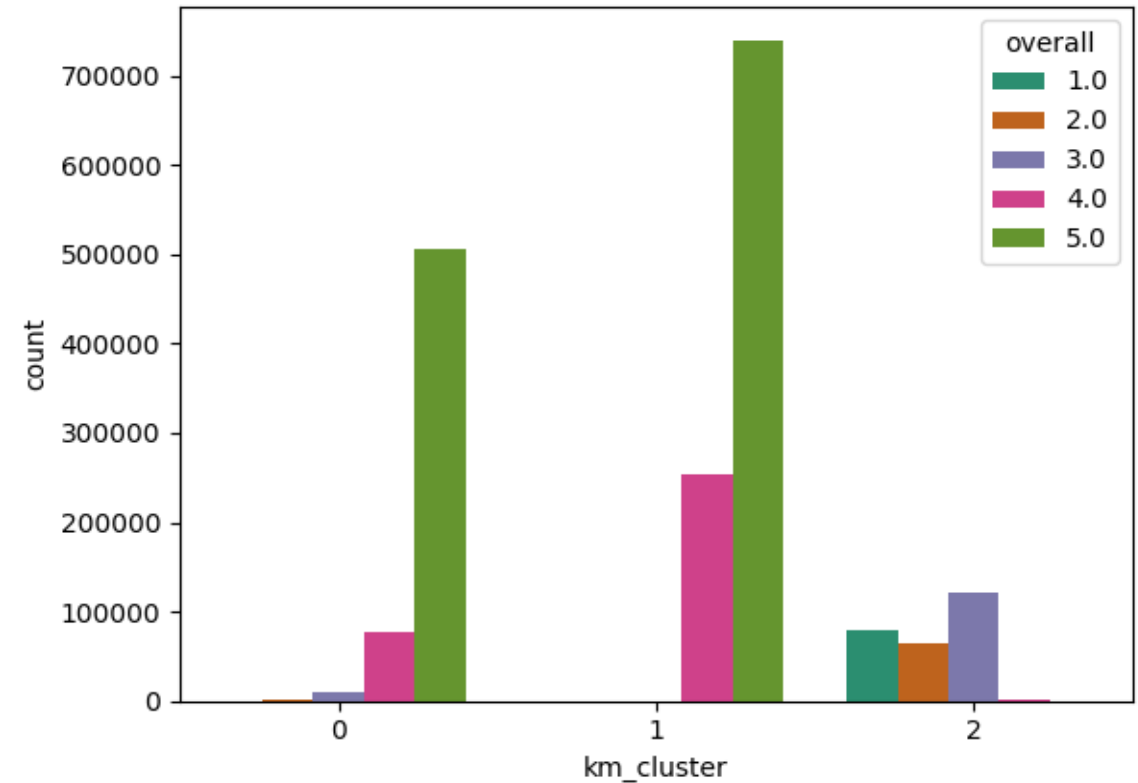
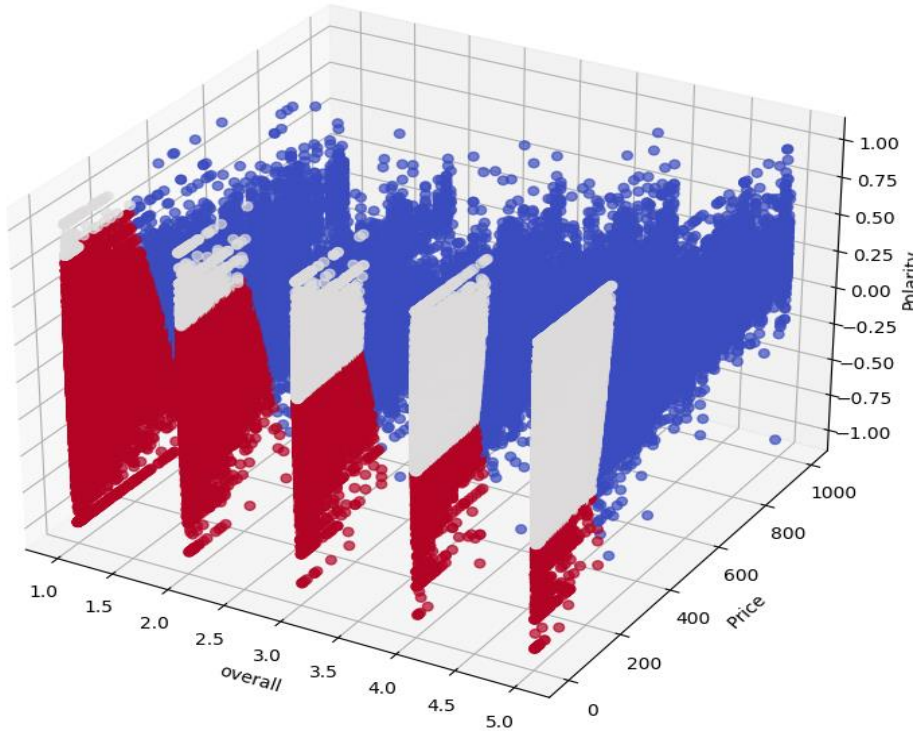
Trend of negative Sentiment office products



- Predicting positive growth for the positive sentiment, by that we can say that sales will definitely go up, and there will always be a demand for the office products in the coming future.
- Negative sentiment is predicted to go down.

Clustering on cd and vinyl

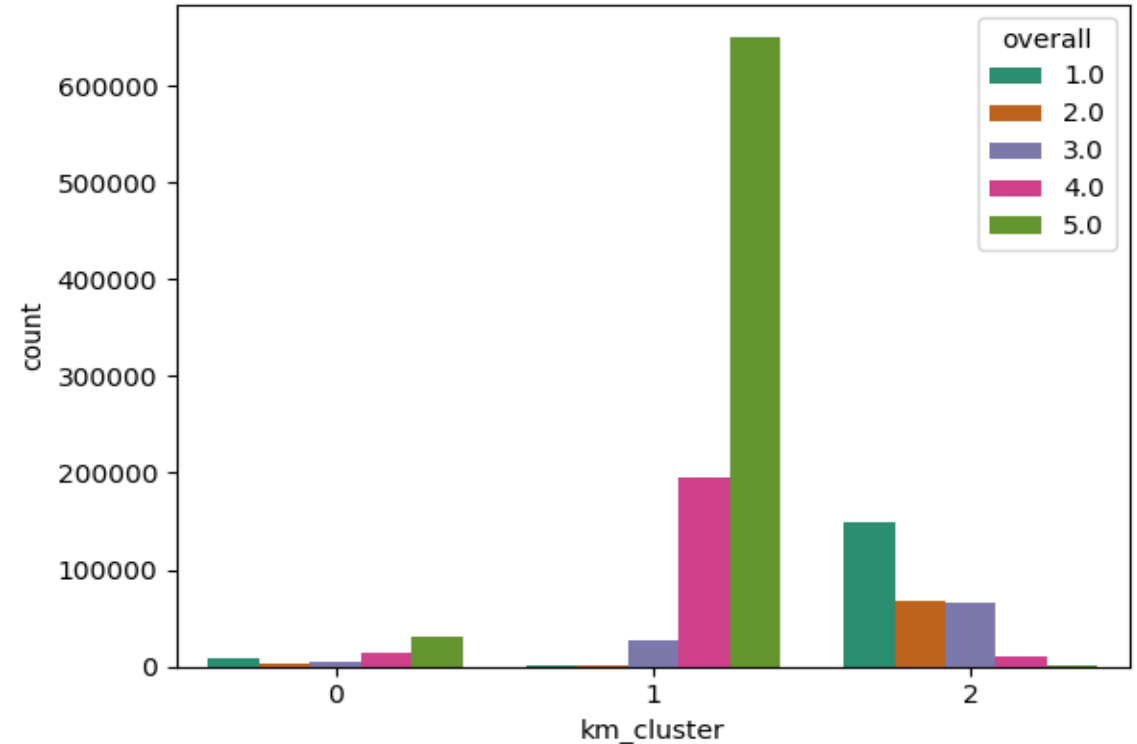
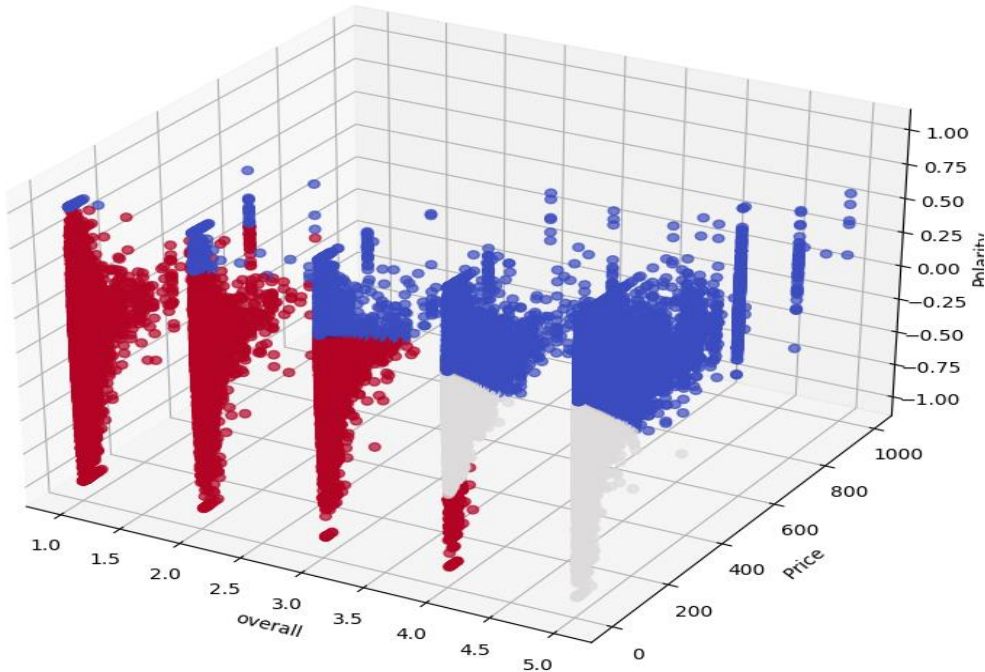
3D Scatter Plot of Clusters



- We use **k-means** clustering
- Utilized purchasing behavior[Rating ,Price, Polarity] data to group customers with similar buying patterns
- Cluster1 represent the customer who are happy with there purchase
- Cluster 2 represents products with negative reviews having lesser sales compare to other two clusters.

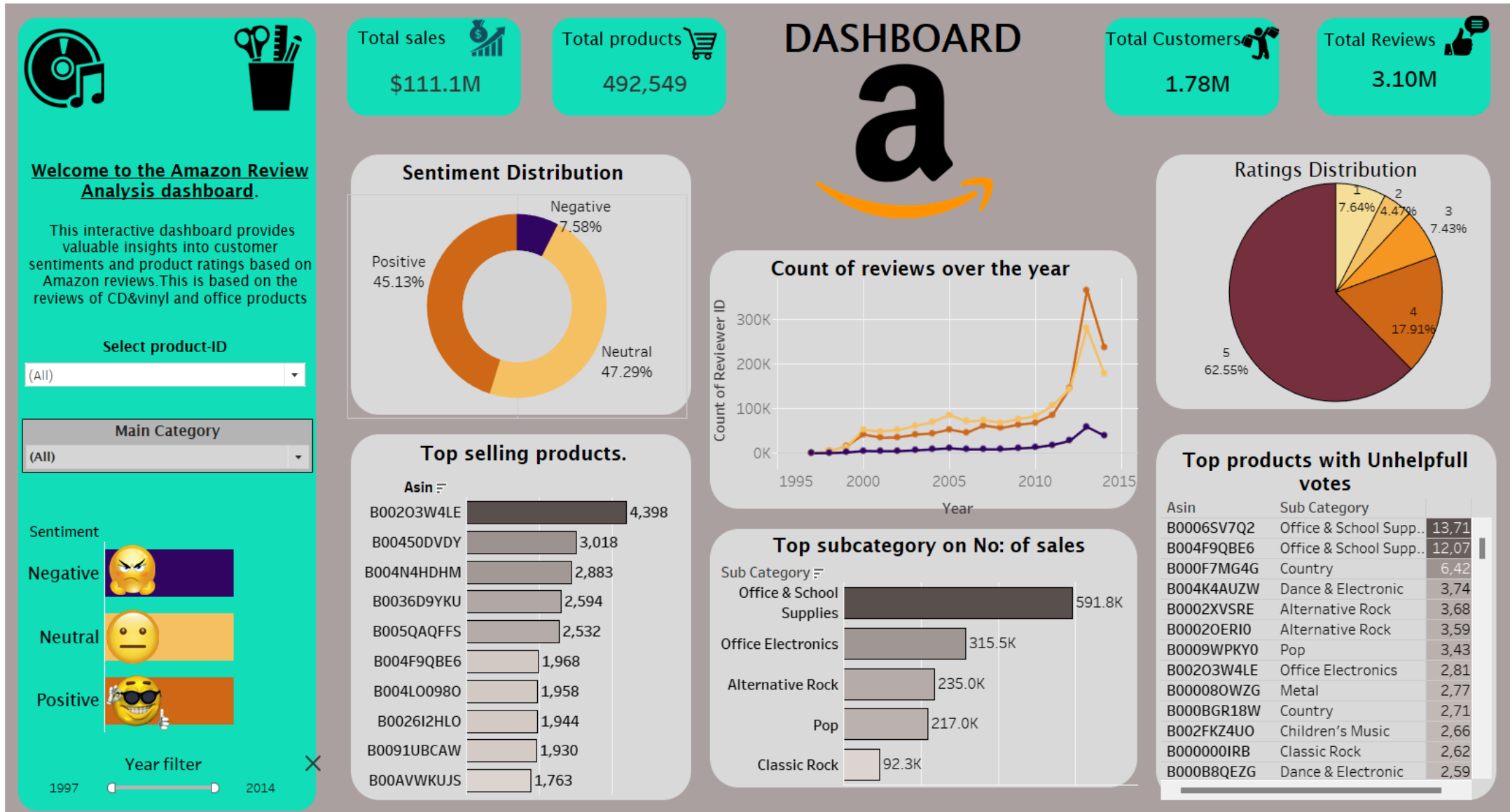
Clustering on office products

3D Scatter Plot of Clusters



- Cluster1 represent the customer who are happy with there purchase
- Cluster 2 represents products with negative reviews having lesser sales compare to other two clusters.
- So, we need to focus on customers of cluster 2 by taking measures to be them active in purchase and their by generating more income to the company.
- Also Amazon will understand the combinations in which people are more interested, and might start selling combo products which will bring in more profit for Amazon.

Tableau Dashboard




suggestion

- **Continuous Evaluation of customer sentiment is necessary**
- **By using our Api, sentiment evaluation on each segment will be much easier.**
- **There must be a false proof system to prevent the fake or promotional reviews.**
- **Continuous Monitoring on these reviews and respective seller is necessary.**
- **Amazon must initialise a campaign to retain unsatisfied costumers by giving them special discounts or credit card offers.**
- **Ensure no product is overly priced ,price to quality must maintain**



CONCLUSION

- 
- We estimated the sentiment of both datasets using the textBlob, and then analyzed results .
 - Explored the possibility of promotional and fake reviews and generated a list of ReviewID doing that malpractice.
 - Created a NLP model with good results for analyzing the future reviews.
 - Build an API incorporating the above NLP model. API have the dual functionality.
 - By using Time-Series Analysis we predicted the future of sentiment based on the historical data.
 - Using the Clustering techniques, we had grouped them based on their preference and feedback..
 - Build a dashboard on the entire dataset.



THANK YOU