

AMAZON PRODUCT REVIEW ANALYSIS

PRESENTED BY -

SHRITESH JAGTAP
JITENDRA GIRISH PATIL
PUNEETHKRISHNA B.
SAMRADDHI SHRIVASTAVA



Introduction

- In the retail e-commerce world of the online marketplace, where experiencing products is not feasible. Therefore, customers need to rely largely on product reviews to make up their minds for better decision-making on purchases.
- However, searching and comparing text reviews can be frustrating for users. Hence we need a better numerical rating system based on the reviews which will make customers' purchase decisions with ease.
- The amazon review dataset for Musical Instrument, Digital Music, and Amazon Instant Video Category was considered. And hence based on different criteria obtained from this project we help the consumer and the marketing industry to get the best out of the review system.

COLUMNS	DESCRIPTION
ASIN	PRODUCT ID
REVIEWTEXT	REVIEWS
REVIEWERNAME	REVIEWER NAME
OVERALL	RATING
UNIXREVIEWTIME	REVIEW TIME
REVIEWTIME	REVIEW DATE
USERID	REVIEWER ID
TITLE	PRODUCT NAME
CATEGORIES	SUB CATEGORIES
PRICE	PRICE OF PRODUCT
BRAND	BRAND OF PRODUCT



THE OBJECTIVE OF THE PROJECT

- To develop an automated system to analyze and monitor an enormous number of reviews.
- To Increase the success rate of existing and new products and market campaigns.

TABLE OF CONTENT

- PREPROCESSING AND CLEANING
- EXPLORATORY DATA ANALYSIS
- SENTIMENT ANALYSIS
- TEXT CLASSIFICATION
- TIME SERIES ANALYSIS
- CLUSTERING
- PRODUCT RECOMMENDATION
- CUSTOMER SEGMENTATION
- SUMMARY TABLE
- BUSINESS SOLUTIONS

PREPROCESSING AND CLEANING

Data Importation



1

Data Exploration



2

Missing Value Treatment



3

Feature Selection



4

Feature Scaling



5

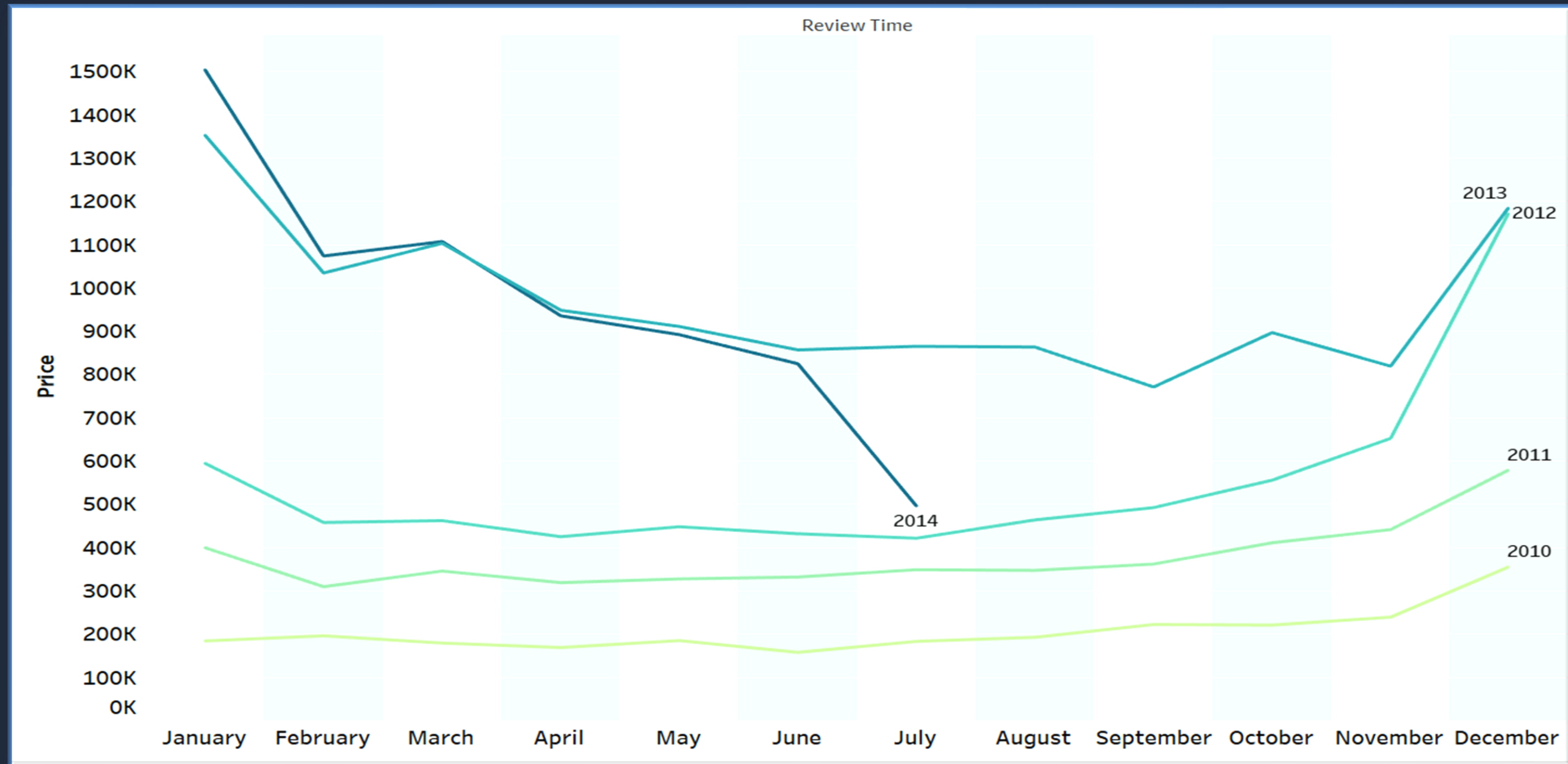
EXPLORATORY DATA ANALYSIS

UNIVARIATE ANALYSIS

BIVARIATE ANALYSIS

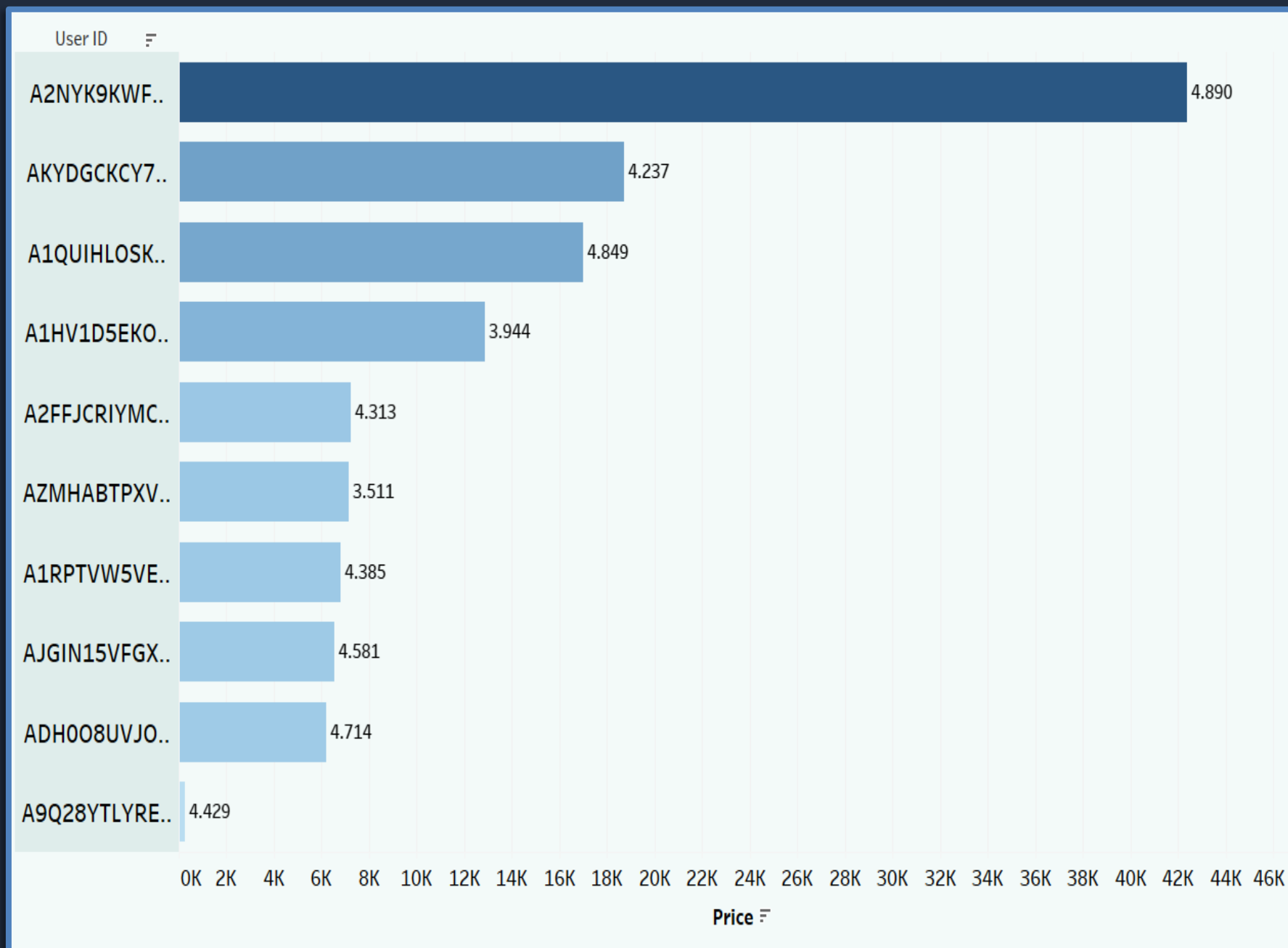
MULTIVARIATE ANALYSIS

MONTH WISE REVENUE

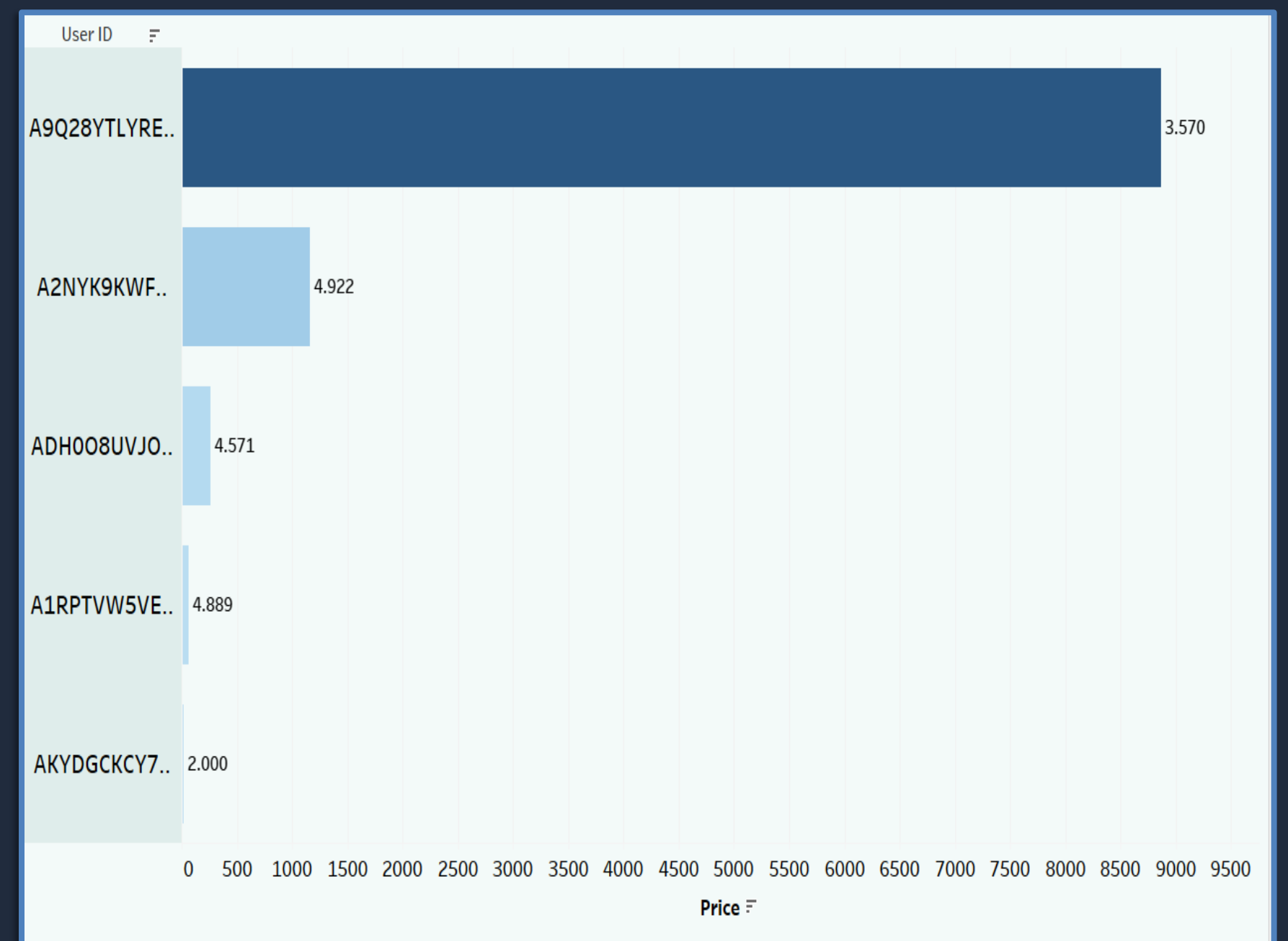


REVENUE WISE TOP CUSTOMERS

Musical Instrument Category

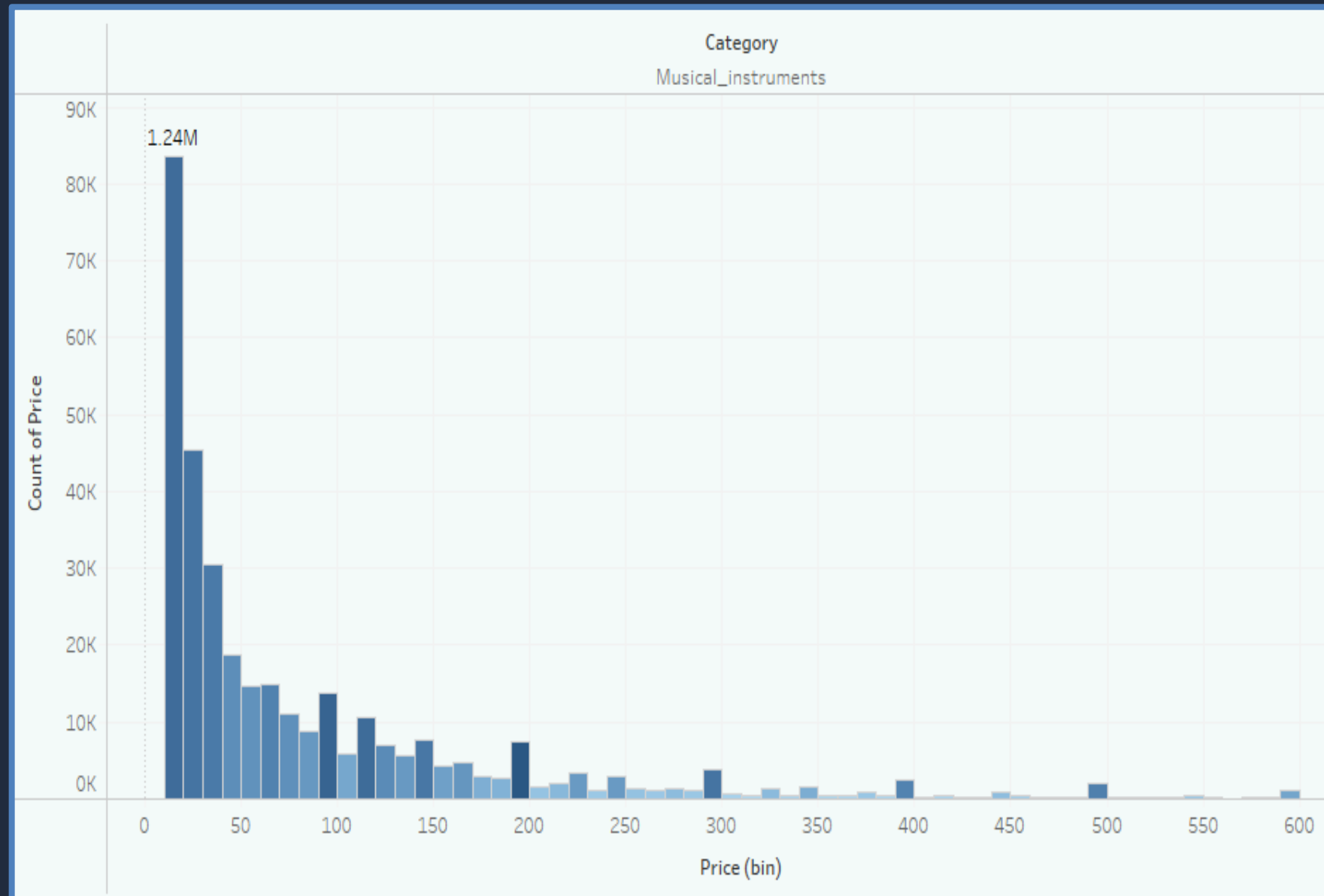


Digital Music Category

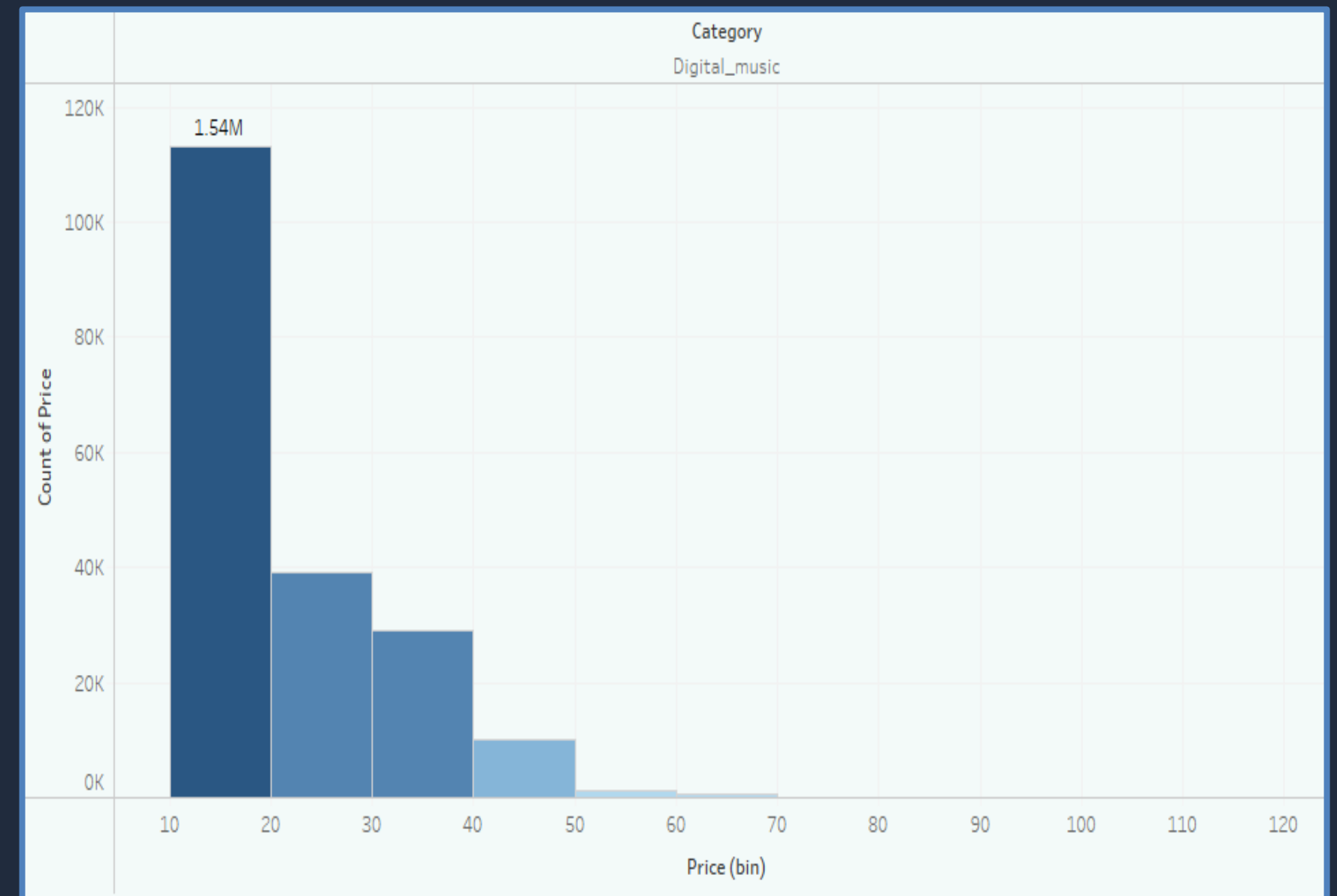


PRICE DISTRIBUTION OF PRODUCTS

Musical Instruments Category



Digital Music Category



For further analysis of data, we will head towards [Dashboard](#)

SENTIMENT ANALYSIS

Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative, or neutral.

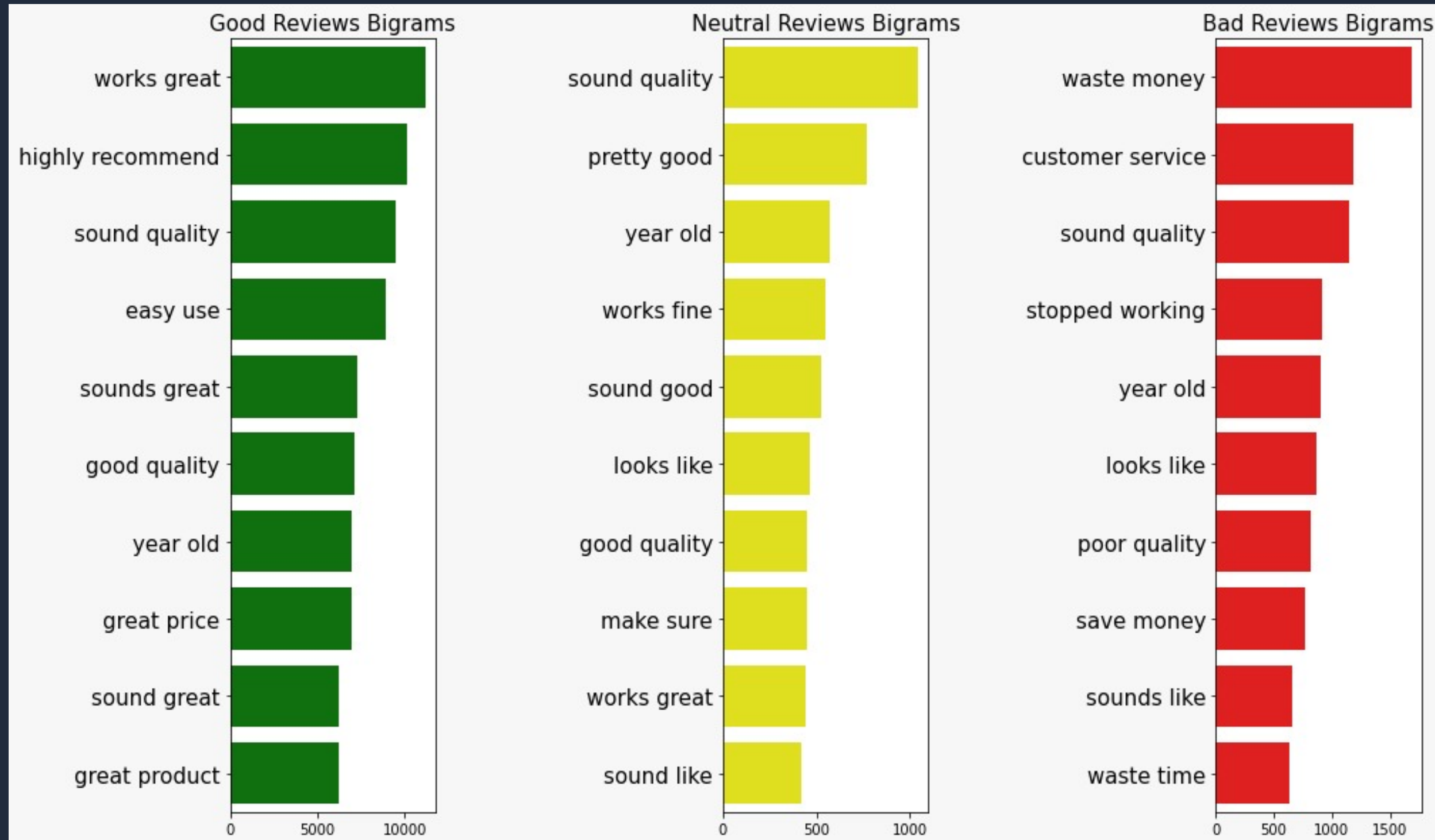
- To make the data more understandable the reviews of the customers are converted into Sentiments using VADER (Valence Aware Dictionary for Sentiment Reasoning).
- Language Detection
- Language Translation

TEXT CLASSIFICATION

Text Classification is the process of categorizing text into organized groups.

- Using TFIDF (Term Frequency Inverse Document Frequency) for converting text into numbers.
- N-Gram
- Logistic Regression Model is used to predict sentiments from review text.

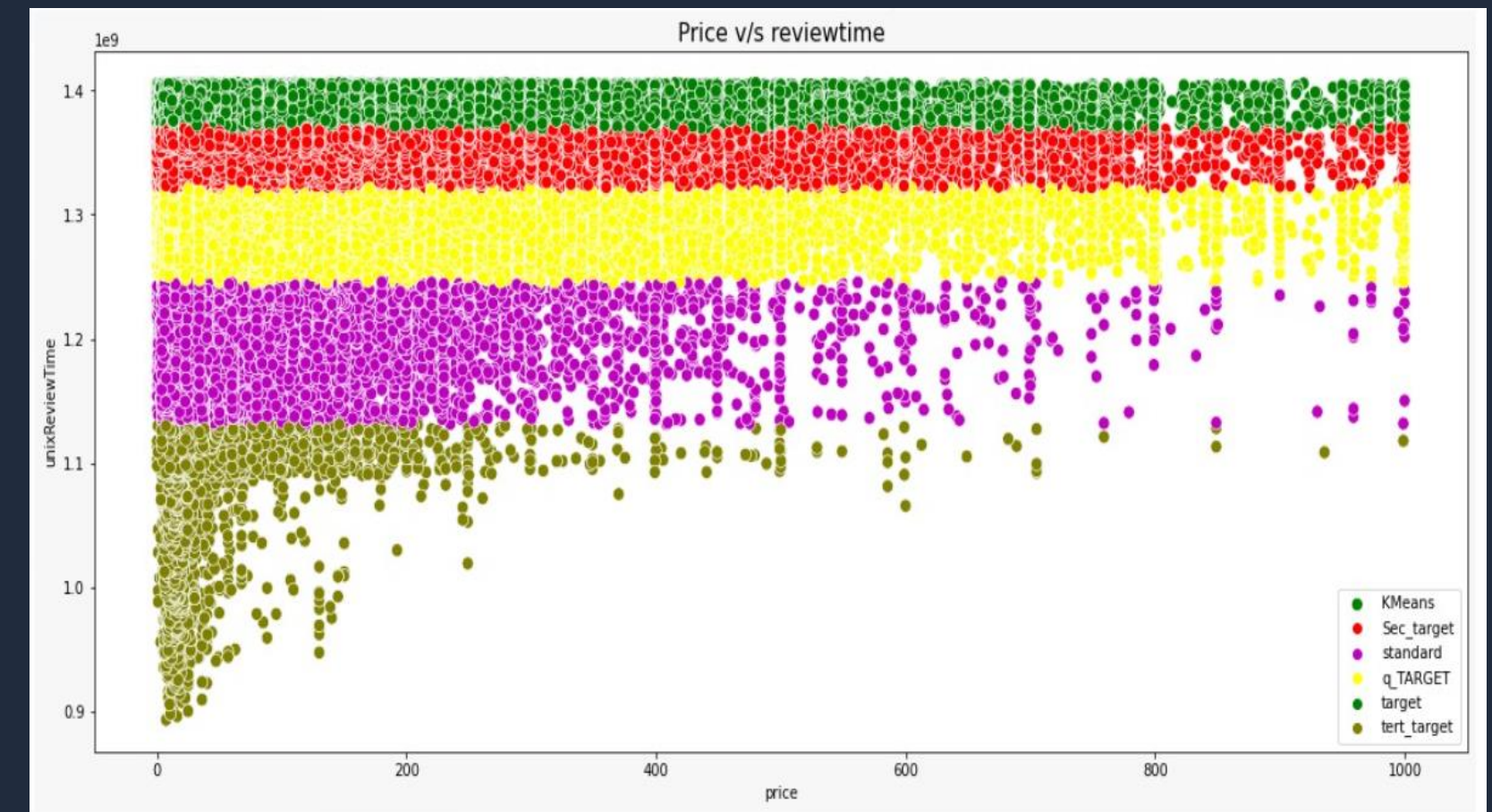
BIGRAM VISUALISATION



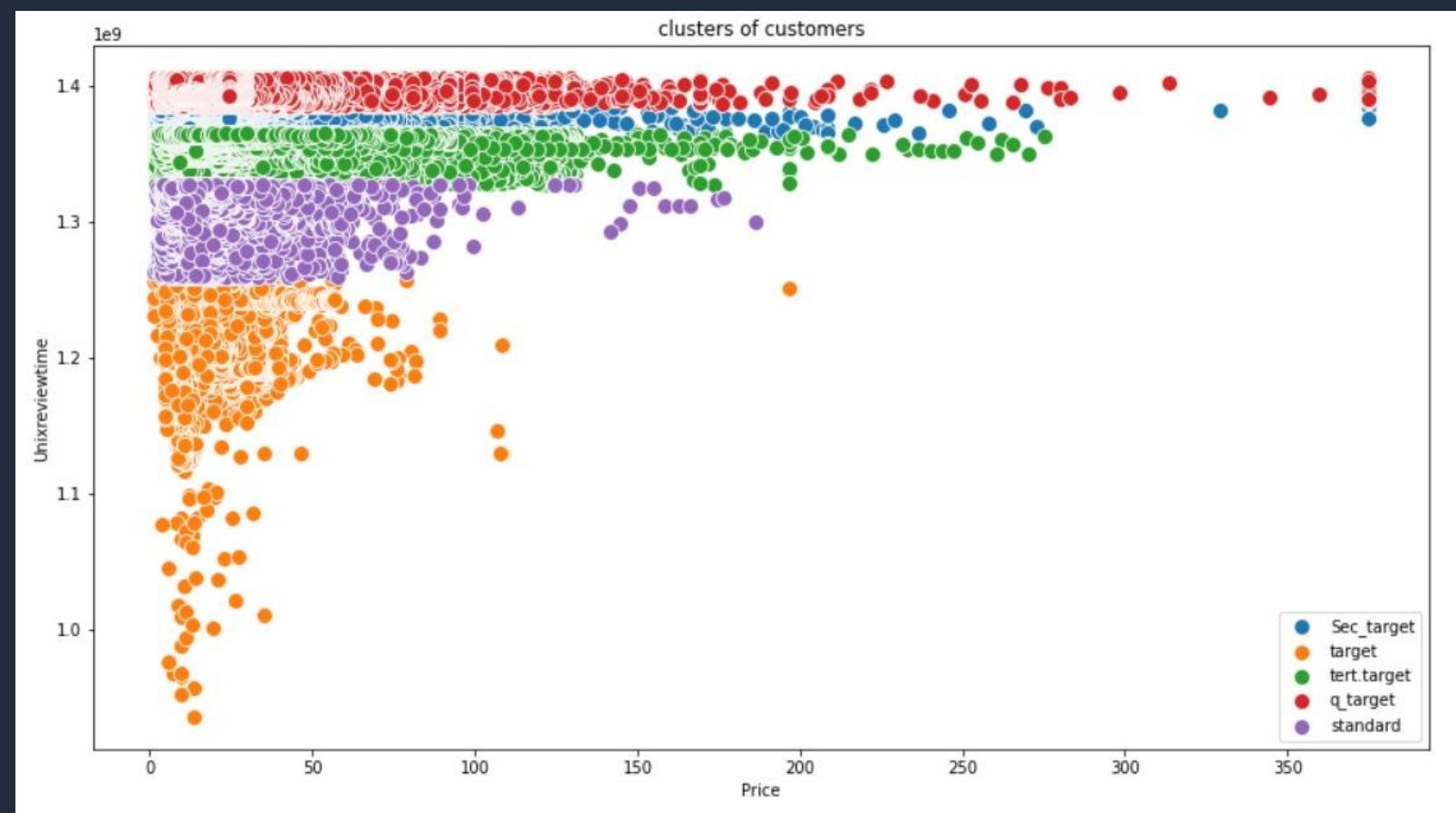
Frequently used words in Reviews

CLUSTERING

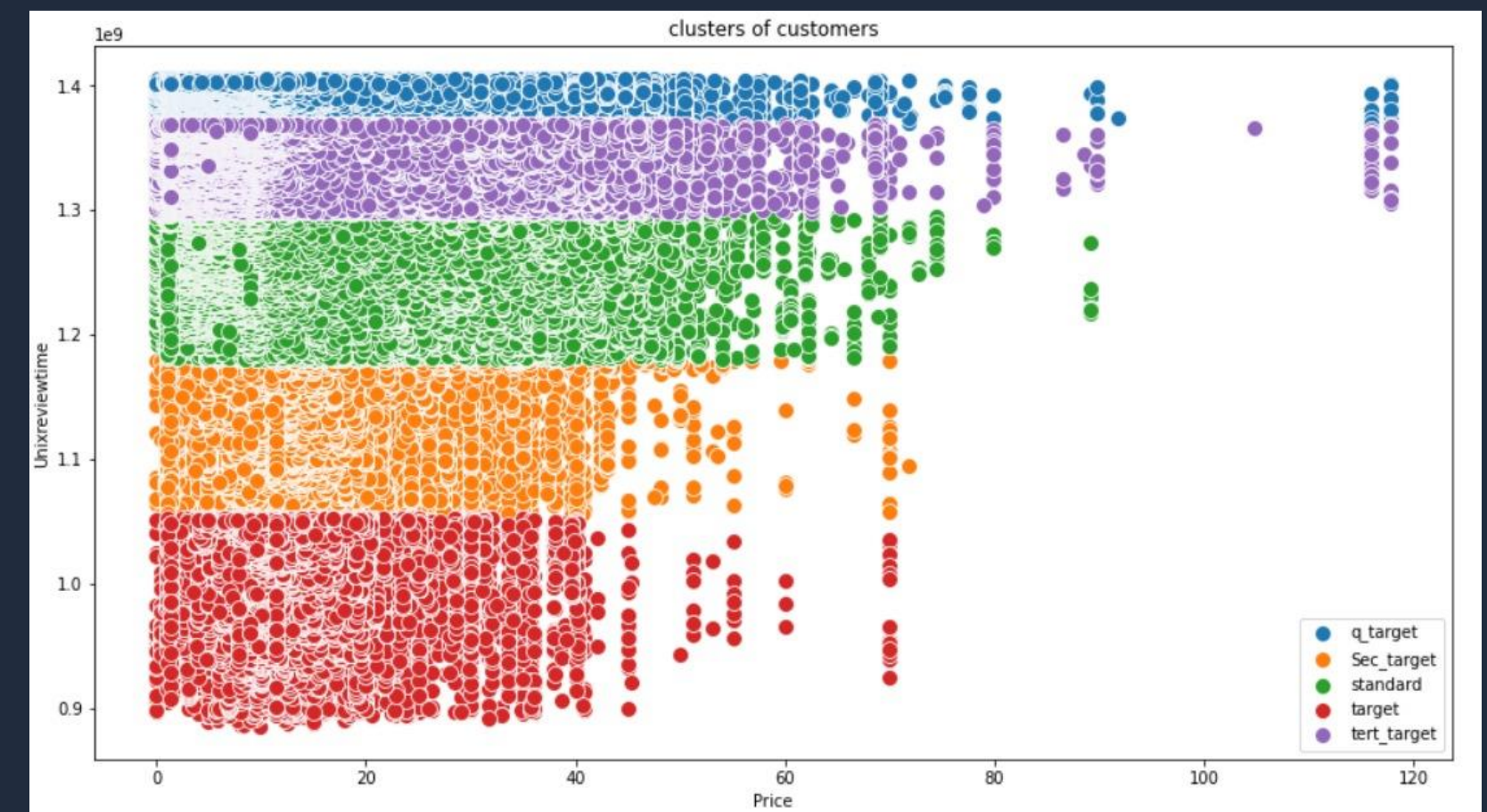
Clustering is basically a grouping of data points on the basis of similarity and dissimilarity between them.



FOR MUSICAL INSTRUMENT CATEGORY



FOR AMAZON INSTANT VIDEO CATEGORY



FOR DIGITAL MUSIC CATEGORY

TOP AND BOTTOM PRODUCTS OF VARIOUS CATEGORIES

Musical Instruments

B00004XQ83	289
B00005EBJG	79
0767851013	59
B00005QZH2	33
B000002LB7	25
B000098XFM	18
B0000255CL	17
B00003IE4L	17
B00009MDEO	16
B00000I038	15

**T
O
P**

**B
O
T
T
O
M**

B00006343B	9
B00004XQ83	5
B0002ISEIU	5
B000VSKPZG	5
0767851013	5
B00004YK00	4
B00009ETK4	4
B0000AHS AU	3
B00001QE HM	3
B000PTF0E2	3

Digital Music

B004D1GZ2E	1187
B004K4AUZW	827
B0075MFS9Y	495
B00970FKQ8	432
B008K9SG9K	407
B006M4RN3U	382
B007DIQAXI	380
B00900PBTC	354
B009KUA0H0	332
B005QJZ5FA	303

**T
O
P**

**B
O
T
T
O
M**

B004K4AUZW	220
B002HP8EKE	128
B004R0MF0W	47
B008UXALQ2	44
B004D1GZ2E	43
B004NSULHM	42
B00B5JY86W	31
B004S5JBZ8	26
B008RYN44U	25
B00970FKQ8	23

Amazon Instant Video

B002962RKE	231
B001F6ZIXC	145
B0012QRPU4	141
B000VZUWZW	137
B002L4BQ42	124
B002KGEORM	122
B002TU02SU	114
B000J09JHW	105
B004X2M3N8	105
B00337ZGIS	88

**T
O
P**

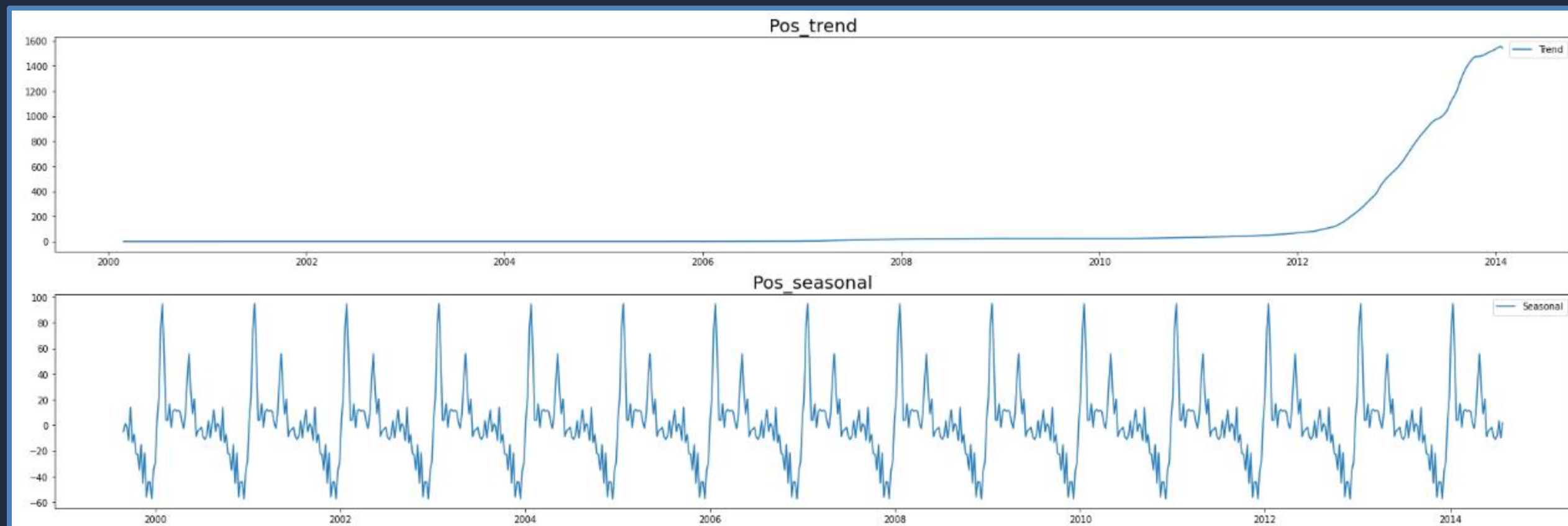
**B
O
T
T
O
M**

B004GEYL9M	53
B0044BY98K	52
B004EW7AWG	33
B0064JGWZS	28
B002962RKE	28
B003075T38	25
B004I90IW4	19
B003CWHG8M	19
B002QS50Q4	18
B002IVNLKA	15

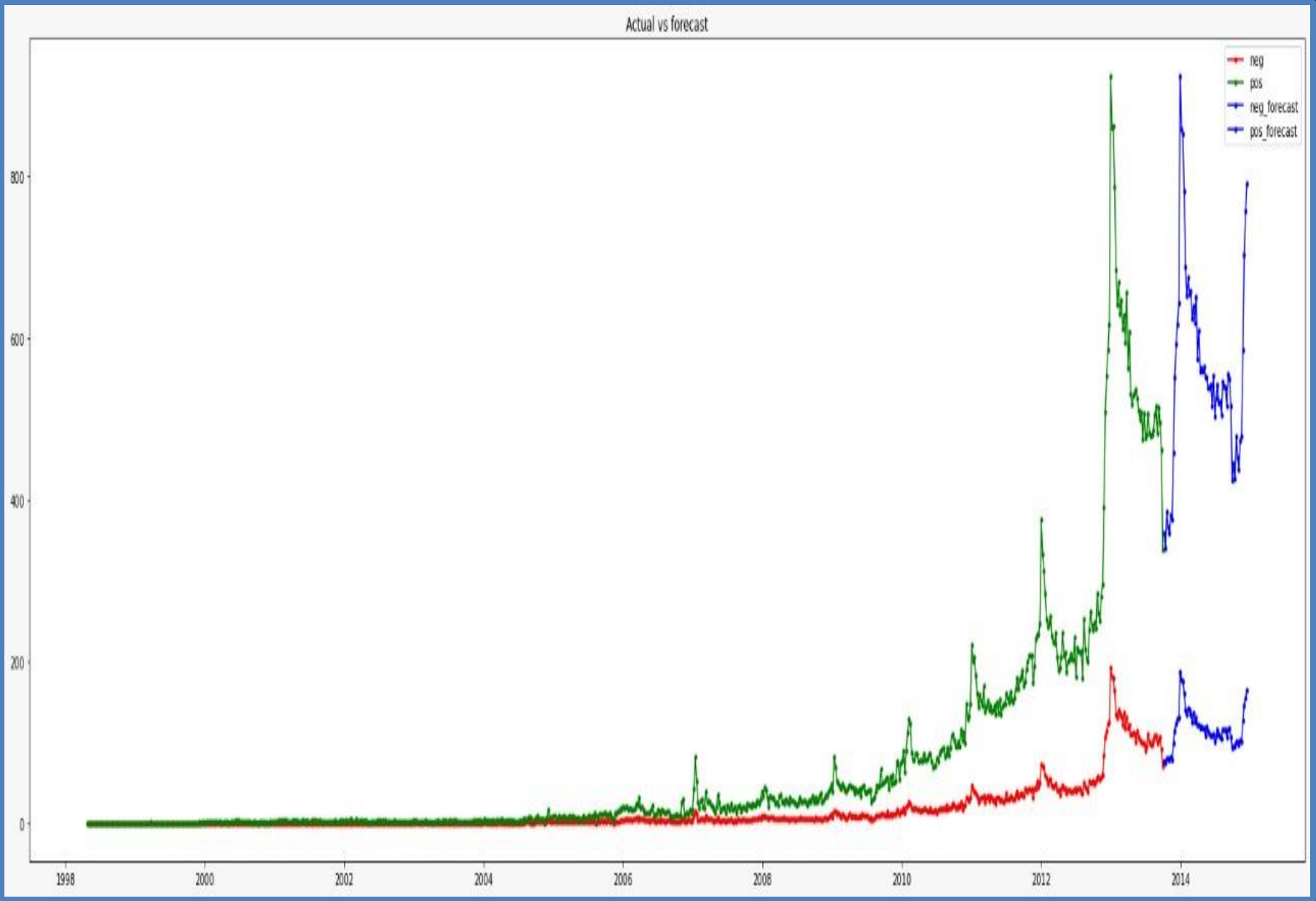
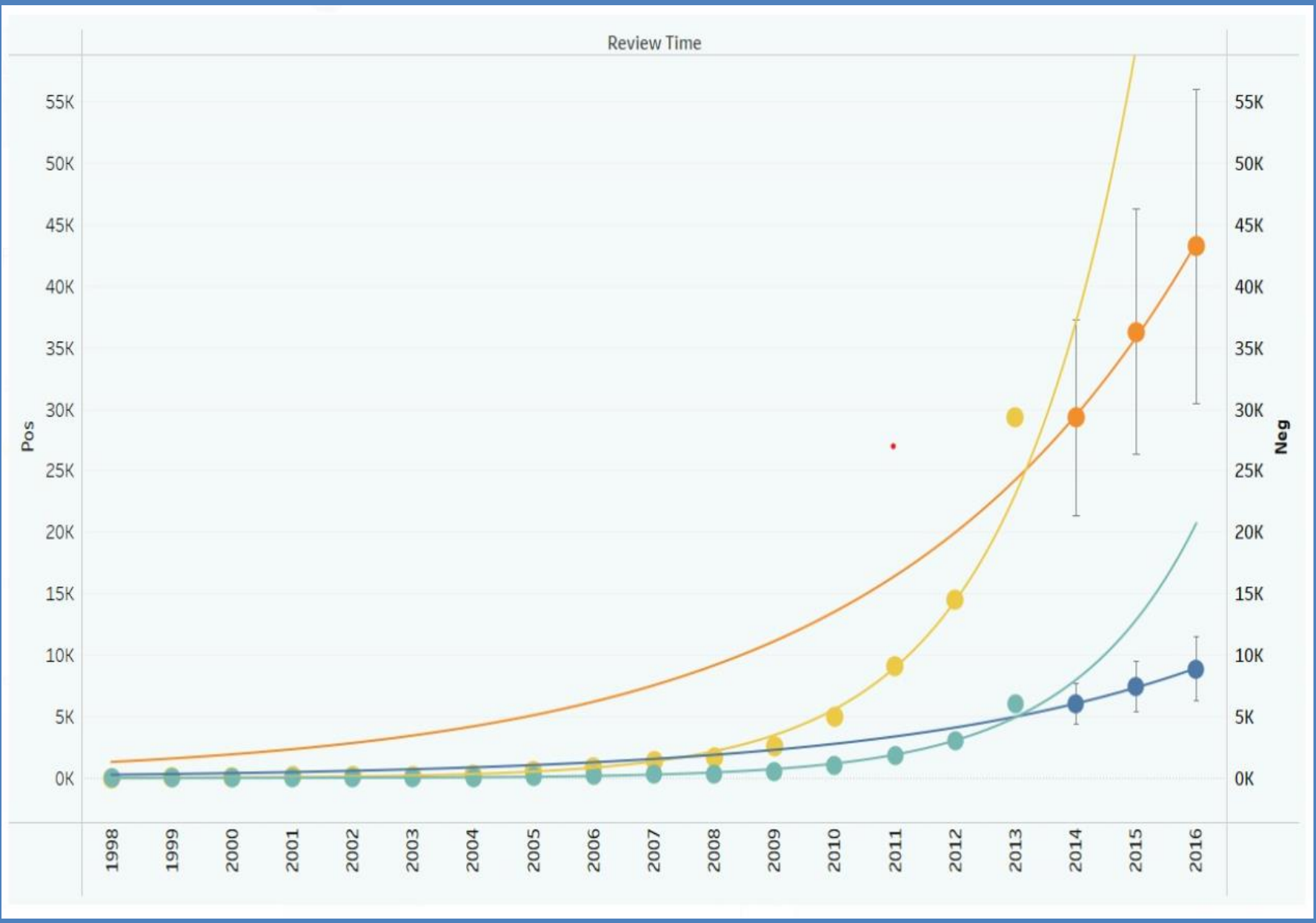
TIME SERIES ANALYSIS

Time Series Analysis is the way of studying the characteristics of the variables with respect to time.

- The Overall data have some seasonality and trend.
- The data has been converted into Weekly data.
- With the help of Time Series Analysis we can forecast various continuous entities which can help us understand the trend better.

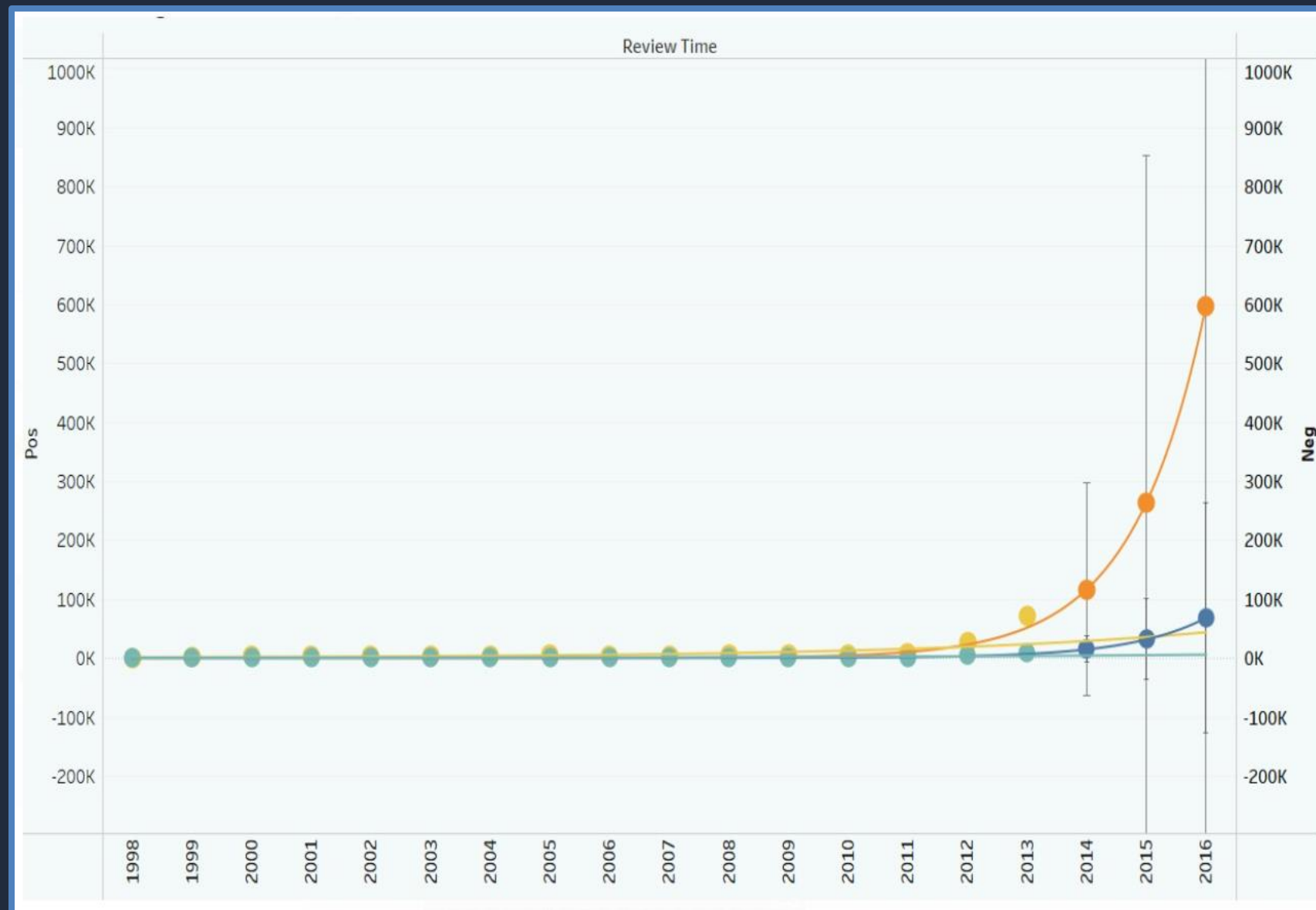


FORECASTING OF POSITIVE AND NEGATIVE SENTIMENTS

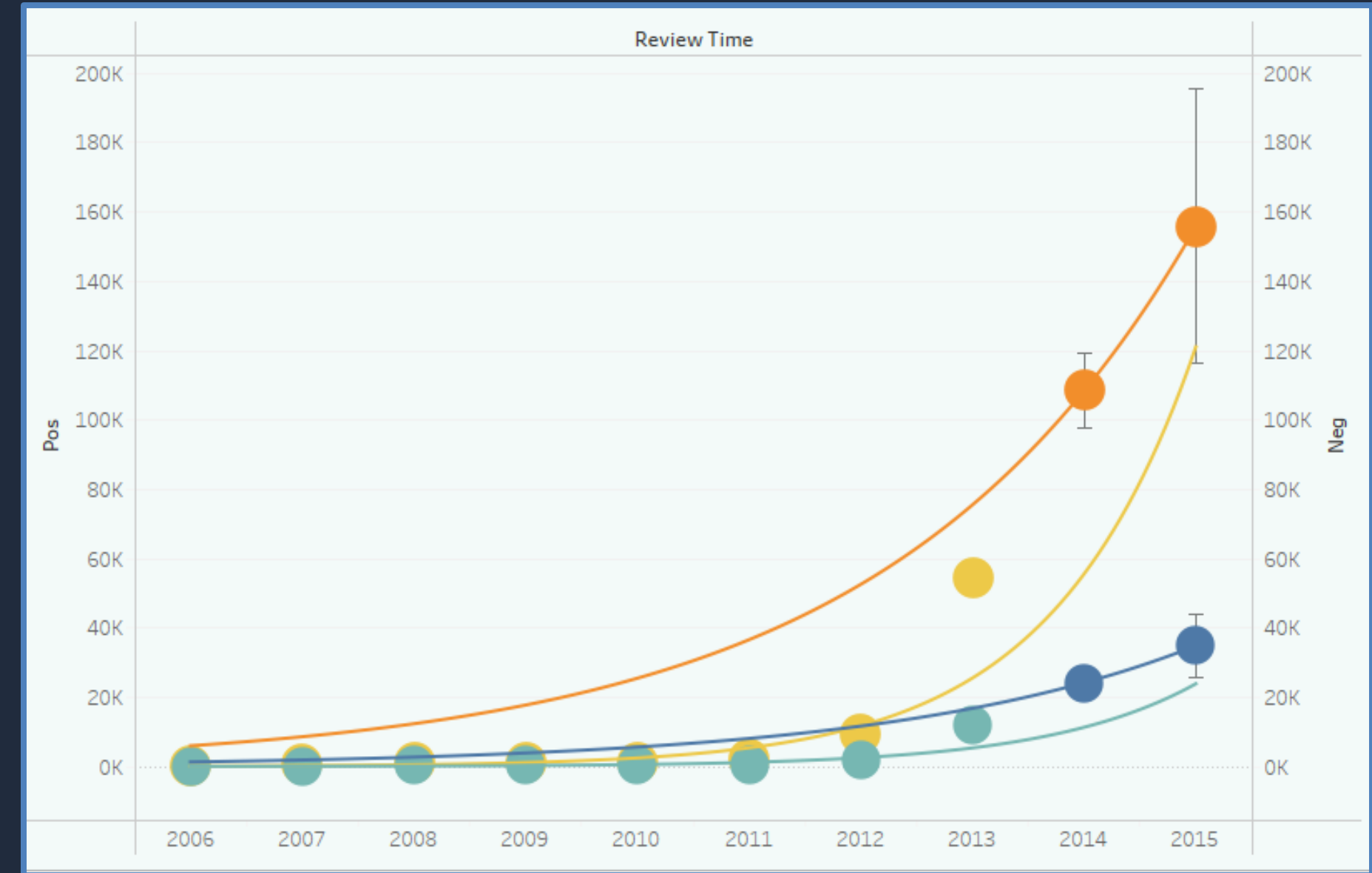


FOR MUSICAL INSTRUMENT CATEGORY

FORECASTING OF POSITIVE AND NEGATIVE SENTIMENTS



FOR DIGITAL MUSIC CATEGORY



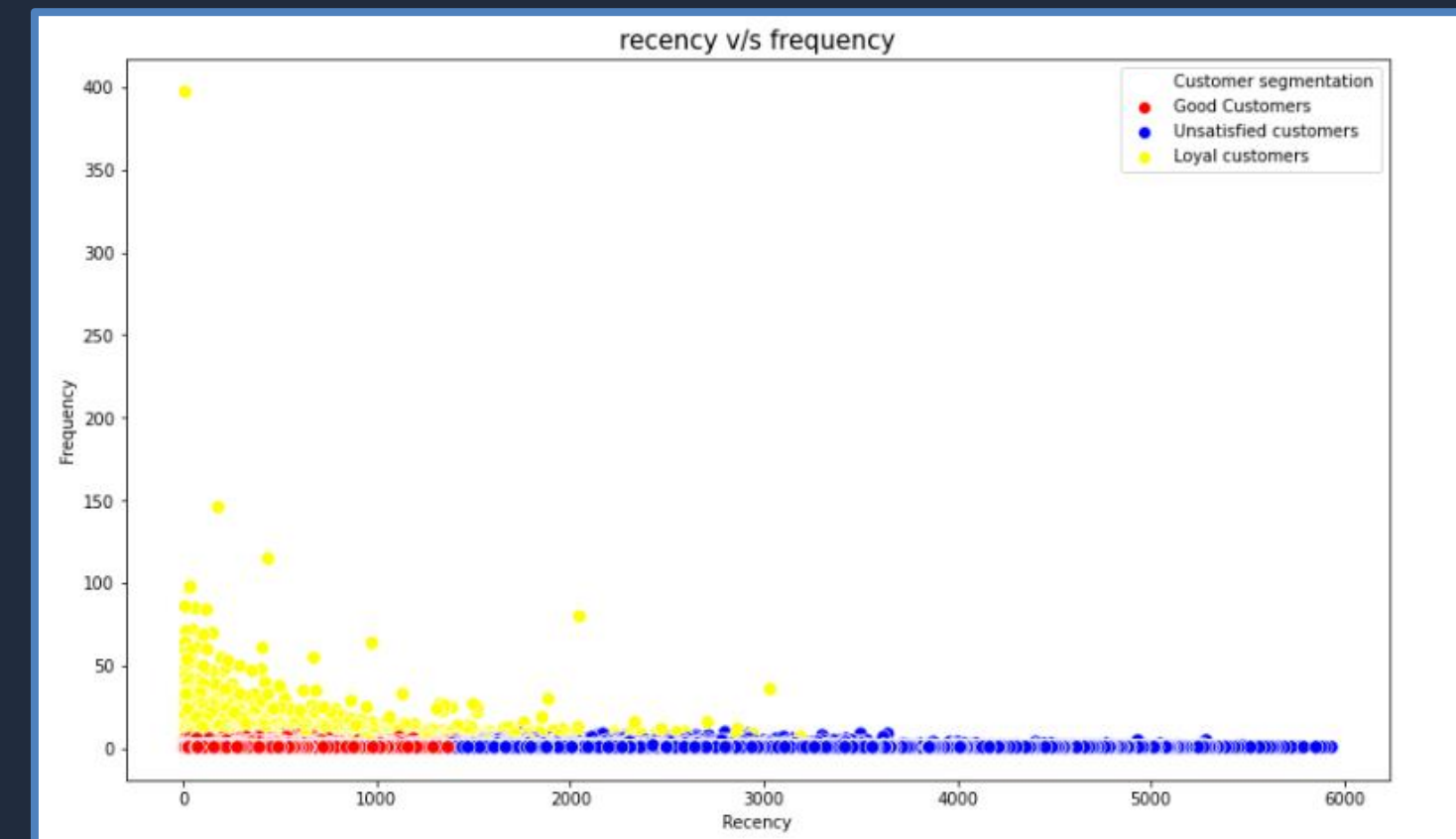
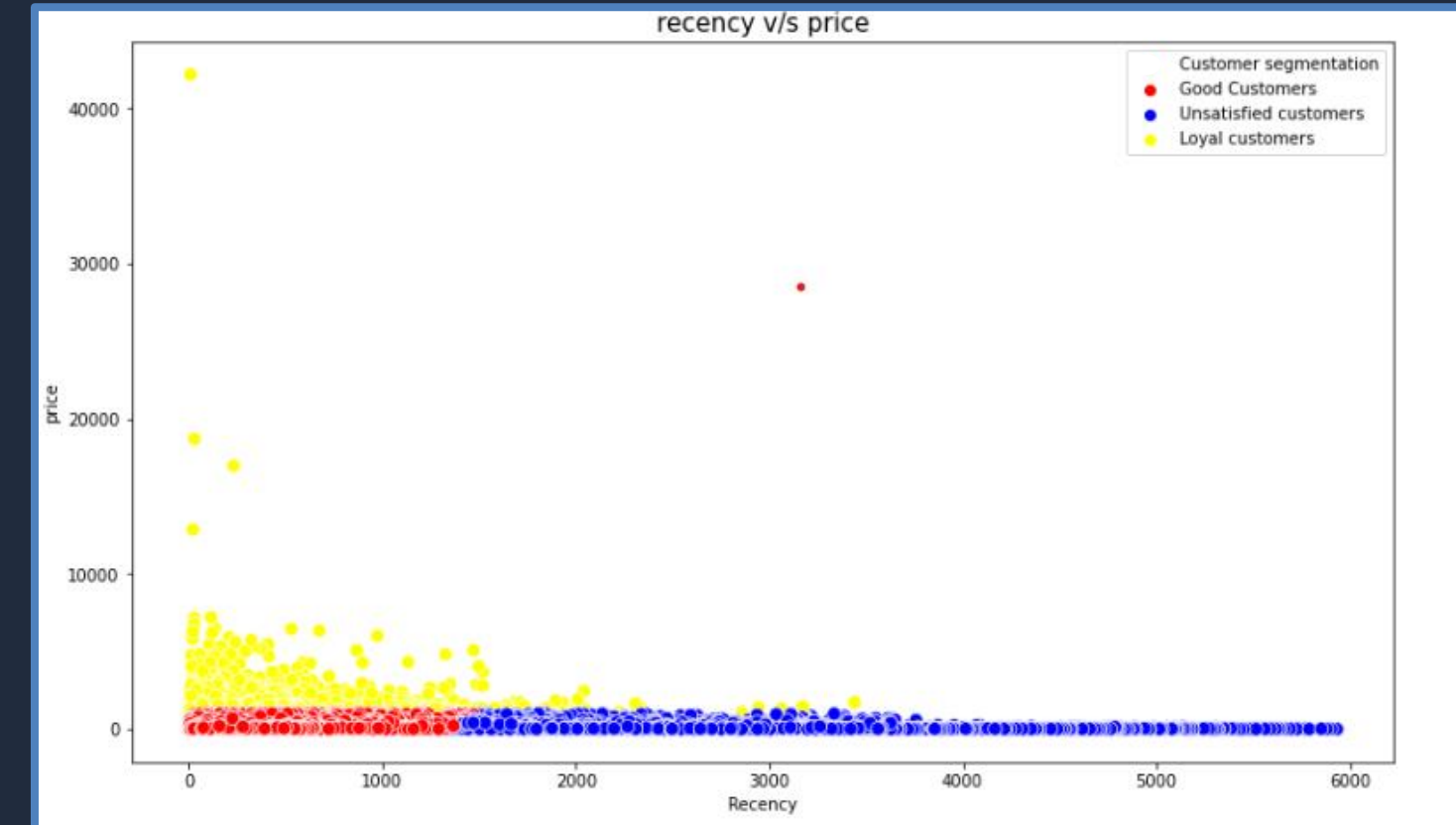
FOR AMAZON INSTANT VIDEO CATEGORY

CUSTOMER SEGMENTATION

Customer segmentation is playing an important role in terms of retaining the customers

➤ RFM Model

- Recency
- Frequency
- Monetary value



PRODUCT RECOMMENDATION

Recommendation systems collect customer data and auto-analyze this data to generate customized recommendations for your customers

Amazon Recommendation System

- Collaborative filtering(mode-based)
- Correlation Matrix
- Truncated SVD

Recommended Products for Musical Instruments

PRODUCT 1 – B0002BG2DQ

GATOR NOTE KEYBOARD COVER GKC

0739079883 – ALFRED'S KIDS GUITAR COURSE
COMPLETE STARTER PACK

B00001W0DT – ONE FOR THE KIDS

B000068IDU – CASIO GSDX DELUX KEYBOARD
STAND

B000068NTU – FOOT LONG MIDI TO MIDI
CONNECT CABLE

B000068NVI – HOSA CABLE GTR210 GUITAR
INSTRUMENT CABLE FOOT

PRODUCT 2 – B0002ABOFS

NUMARK PT01 PORTABLE DJ TURNTABLE

0739079883 – ALFRED'S KIDS GUITAR
COURSE COMPLETE STARTER PACK

1417030321 – BROKEBACK MOUNTAIN

B000002LB7 - ELECTRIC

B00004UFRO – SONY WCS999 WIRELESS
CAMCORDER MICROPHONE FOR DCR-DVD101

B00004XQ83 – O BROTHER WHERE ART
THOU

SUMMARY TABLE

Title	Musical Instrument	Digital Music Data	Amazon Instant Video
1. Pre-processing	Null Value Dropped	Null Values Imputation	Null Values Imputation
2. Sentiment Analysis	TFIDF, VADER	TFIDF, VADER	TFIDF, VADER
3. Models	Logistic Regression	Logistic Regression	Logistic Regression
4. Evaluation - Accuracy	86%	92%	88%
5. Evaluation - F1	84%	90%	86%
6. Clustering	Kmeans Clustering	Kmeans Clustering	Kmeans Clustering
7. TSA	SARIMA	SARIMA	SARIMA
8. Forecasted Pos, Neg Sentiment	Sharp upward, mild upward	Sharp upward, mild upward	Sharp upward, mild upward
9. Reviewer Name (Max)	John	Michael	Amazon Customer
10. Top and Bottom Product	B00004XQ83,B00006343B	B004D1GZ2E,B0004K4AUZ W	B002962RKE,B004GEYL9M
11. Recommended Product	B00004UFRO	B00000016T	B000HZEHL6

BUSINESS SOLUTIONS



EDA

Count of reviews increasing over the period of time

Revenue is increasing over the period of time



SENTIMENTAL ANALYSIS

Model is able to detect and translate all the languages to English

Model is able to Automate Sentiment Predictions



CLUSTERING

Model is able to segregate top and bottom products

Model is able to create segments based on customer perceptions



PRODUCT RECOMMENDATION

Model is able to recommend related products based on customer purchase

Model is able to forecast future trend of the sentiments

THANK YOU