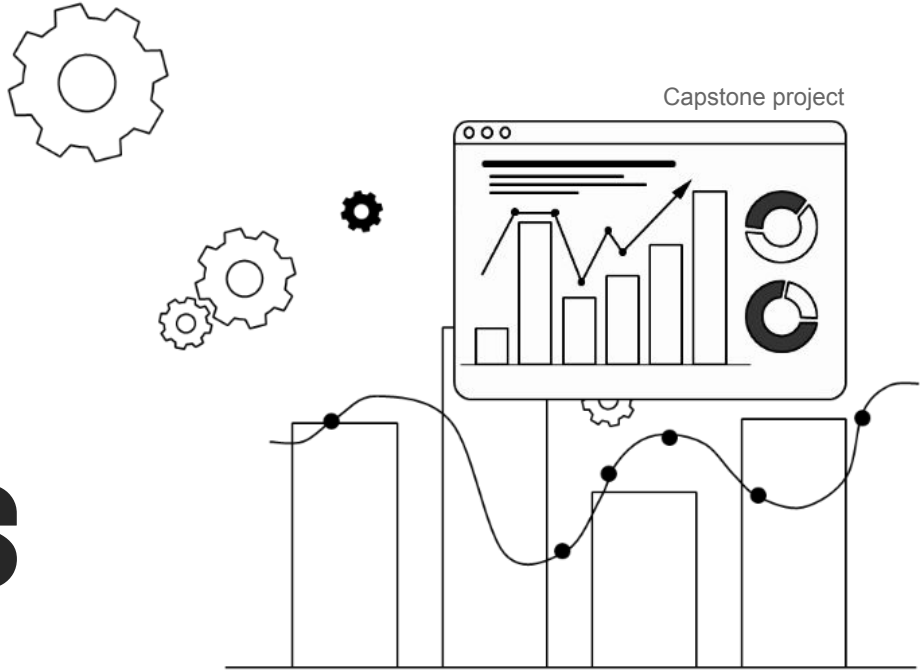


# AMAZON PRODUCT REVIEW ANALYSIS

TOYS & GAMES : INVENTORY OPTIMIZATION AND DEMAND FORECASTING



Guided by Dr. Amit Kumar

Kartik Sharma | Sagar | Tanish Namdev | Sivasankari Natarajan

# AGENDA OF THE PROJECT

1. Understand the business problem
2. Understand the data
  - 2.1. Derive insights from past data
  - 2.2. Identify the important constituents
  - 2.3. Identify the important characteristics
3. Plan for the next term
4. Recommendations to solve the business problem

# BUSINESS PROBLEM

Excess inventory takes up space and resources that could be used elsewhere. Business need balancing just enough inventory to meet the customer's demands. This ensures that logistics costs stay low by eliminating issues such as running out of stock, backorders and overstocking.

The business also needs to forecast demand to plan and strategize their products and activities accordingly for resourceful and efficient functioning.

# OBJECTIVES

- Optimize inventory management by identifying the product categories based on the product sentiments
- Predict categories that could be in demand based on the sentiments and sales

# DATA UNDERSTANDING

**What data is available?** Product details and Review details of the products

**How much data is available?** 1996 - 2018

## Challenges in the given data

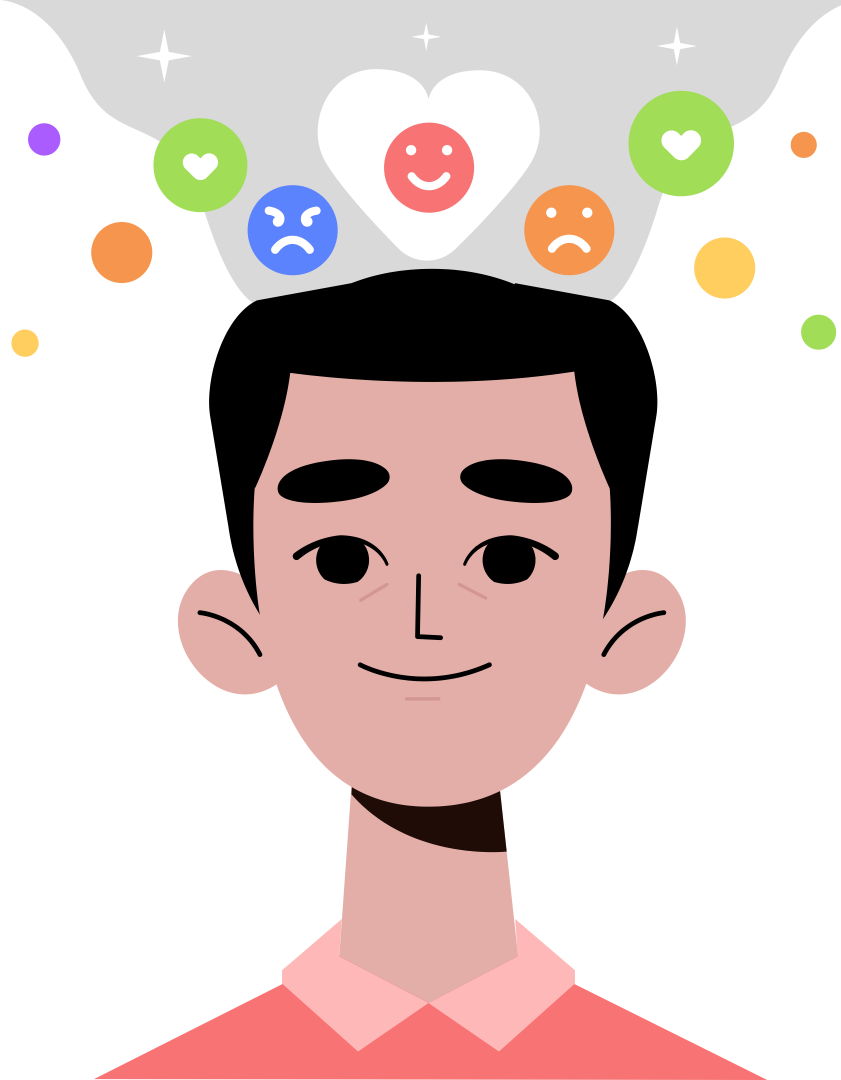
1. Few fields didn't have information- more than 70% of missing values - Drop fields
2. Records having missing values- more than 20% of missing values - Drop records
3. Irrelevant field data types - Type casting

```
Cleaned data
Shape: (1280016, 9)
COLUMN          DATATYPE          NULL VALUES
-----
overall          float64            0.00%
verified         bool               0.00%
reviewTime       datetime64[ns]    0.00%
asin            object             0.00%
reviewText       object             0.00%
summary          object             0.00%
category         object             0.00%
brand            object             0.00%
price           float64            0.00%
```

# EXPLORATORY DATA ANALYSIS

Inventory components: Category, Brand  
Measuring parameters: Price, Rating, Time

1. Which is the category & brand with best rating?
2. What are the contributing factors for good rating ?
3. Does price affect the rating of products? What is the relationship?
4. Is there a relationship existing between price and sales?
5. Are the top rated categories & brand consistent with time?
6. How are the poorly rated products different from well rated products?



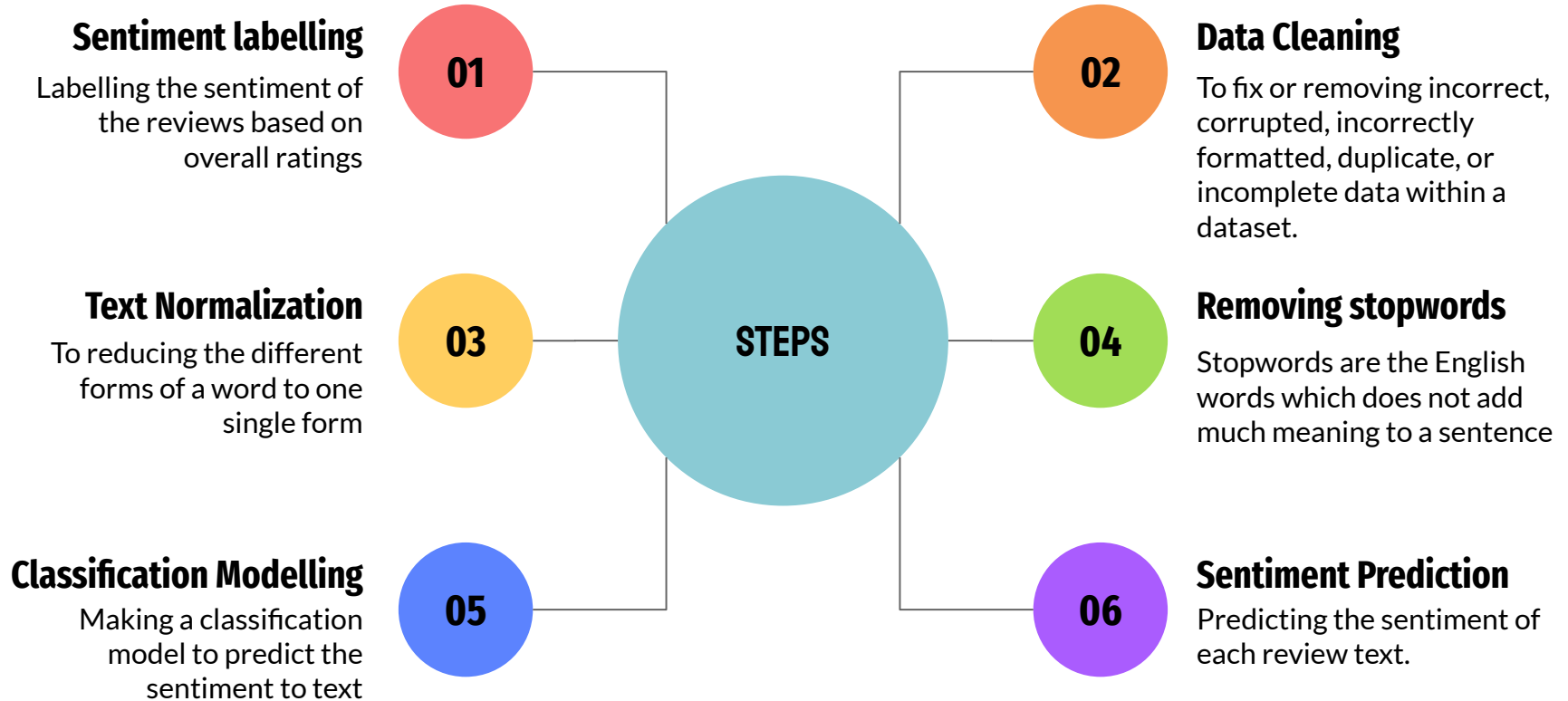
# SENTIMENT ANALYSIS

Sentiment analysis is used to determine if the data is positive, negative or neutral

## AIM

To analyse the sentiment of review text to determine the categories and brands having positive, negative or neutral sentiment

# STEPS INVOLVED IN SENTIMENT ANALYSIS



# CLASSIFICATION

## Aim

Automated system to analyze and monitor an enormous number of reviews

## Data

**Feature:** Review text

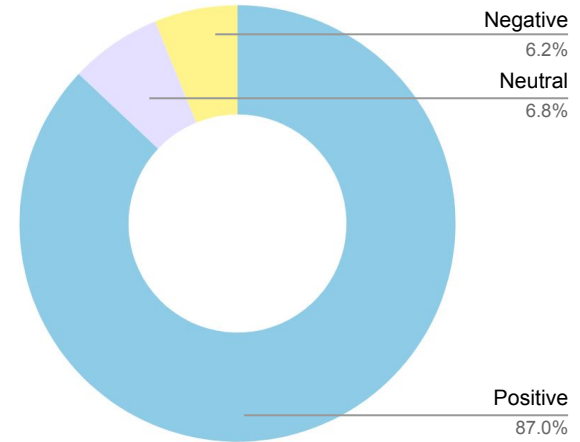
**Label:** Sentiment

**Challenges:** Large data size | Imbalance in data

**Solution:** Stratified random sampling  
Synthetic Minority Oversampling Technique

Data shape: 12,80,016 records

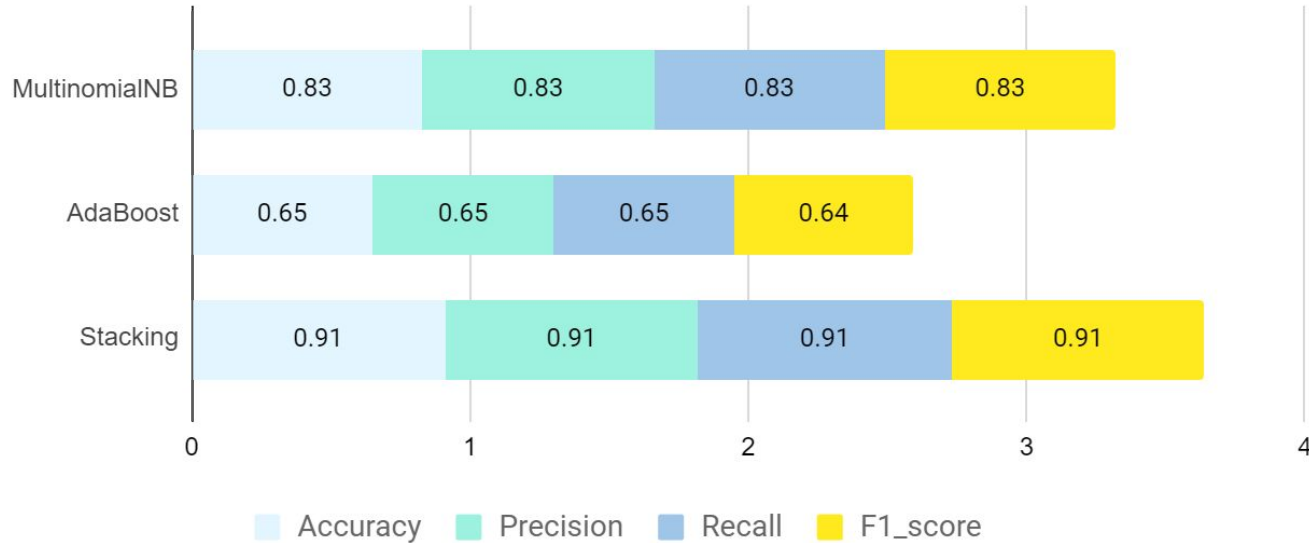
## Sentiment distribution





# EVALUATION OF CLASSIFICATION MODELS

Scores vs Classification model



**Stacking Classifier** is comparatively the best model

# INFERENCE

Based on reviewer sentiments

## TOP 5 CATEGORIES WITH POSITIVE SENTIMENT

Party Supplies  
Remote & App Controlled Vehicles & Parts  
Dress Up & Pretend Play Accessories Polyester  
Dress Up & Pretend Play Hats  
Dress Up & Pretend Play Accessories Plastic

## TOP 5 CATEGORIES WITH NEUTRAL SENTIMENT

Train Cars- Locomotives  
Trains & Accessories- Couplers & Trucks  
Dress Up & Pretend Play- Costumes  
Sports & Outdoor Play  
Trains & Accessories

## TOP 5 CATEGORIES WITH NEGATIVE SENTIMENT

Sports & Outdoor Play  
Remote & App Controlled Vehicles & Parts  
Hobby Building Tools & Hardware  
Arts & Crafts  
Baby & Toddler Toys

## TOP 5 BRANDS WITH POSITIVE SENTIMENT

Galison  
World of Harmony Music  
LEGO  
Bell AH-1S Cobra helicopter model  
Beanie Baby

## TOP 5 BRANDS WITH NEUTRAL SENTIMENT

Child's Play, Inc.  
Melissa & Doug  
Fisher-Price  
Ditty Bird  
U.S. Games Systems

## TOP 5 BRANDS WITH NEGATIVE SENTIMENT

Therapy Game HQ  
Brighter Child  
Gold and Bear  
DIdj  
Evil Hat Productions

# CLUSTERING

## Aim

Grouping similar categories into one cluster and based on cluster properties giving suggestions to improve the sales

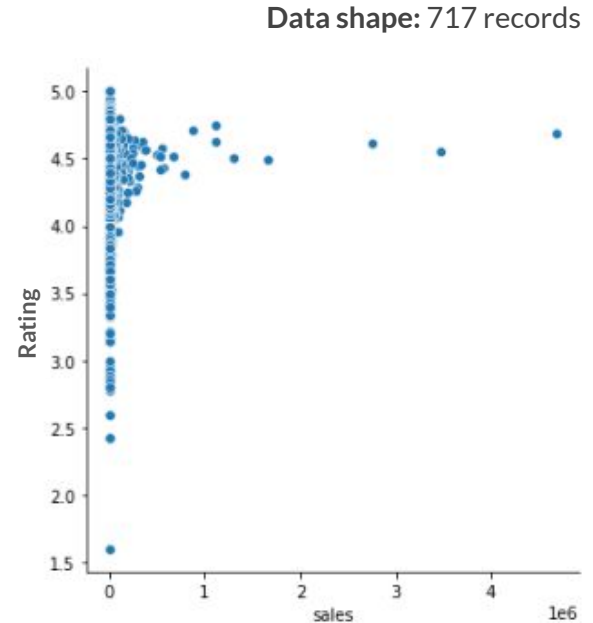
## Data

**Feature:** Sales, Rating

## Models

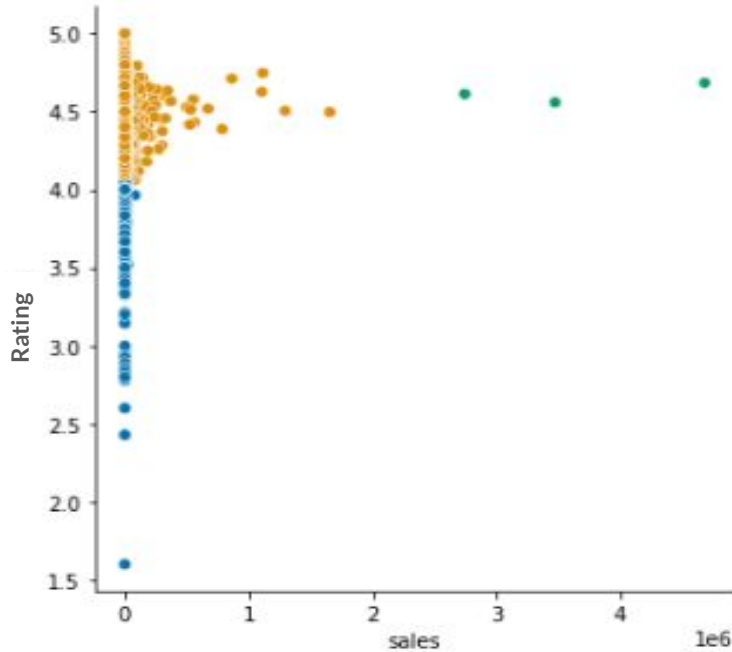
Based on our data we are performing distance based clustering algorithms:

- ☐ K Means
- ☐ Agglomerative

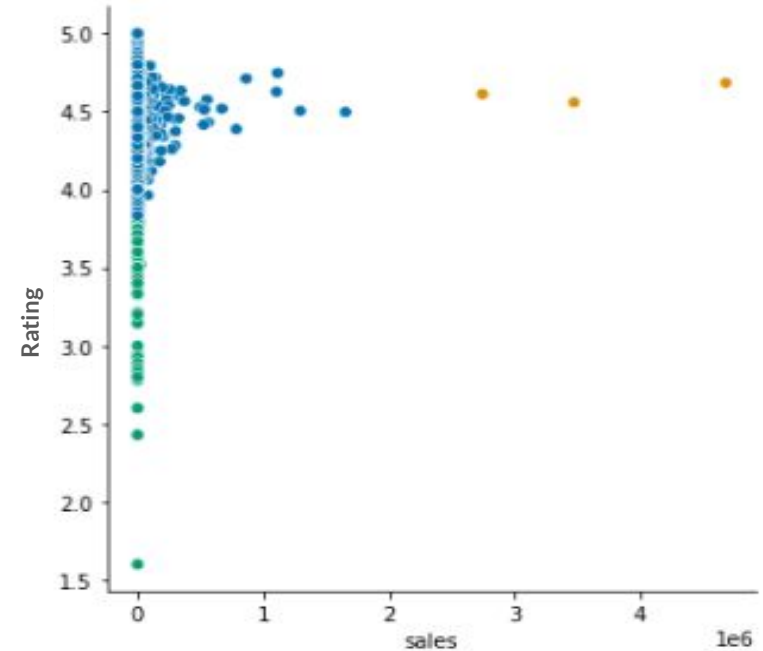


# MODEL COMPARISON

**K MEANS**



**AGGLOMERATIVE**



As we can observe both clustering algorithms provides almost similar results.  
We are selecting **K Means** model as it is least computational expensive as compare to Agglomerative clustering.

# INFERENCE

Best model: K Means Clusters

## CLUSTER 0: MODERATE PERFORMING

Cluster 0	Rating	Sales	Price
Mean	4.43	\$41,662	\$24
Max	5	\$1,653,801	\$523
Min	3.83	\$10	\$1

### CATEGORIES

Costumes  
Cases & Stands  
Party Favors

## CLUSTER 1: BEST PERFORMING

Cluster 1	Rating	Sales	Price
Mean	4.62	\$3,630,773	\$49
Max	4.68	\$4,678,258	\$64
Min	4.56	\$10	\$39

### CATEGORIES

Action Figures & Statues  
Dolls  
Building Toys & Building Sets

## CLUSTER 2: NEEDS IMPROVEMENT

Cluster 2	Rating	Sales	Price
Mean	3.38	\$1,132	\$13
Max	3.8	\$20,737	\$87
Min	1.6	\$10	\$1

### CATEGORIES

Hats  
Headbands  
Pretend Phones

# RECOMMENDATIONS

- First cluster (**Moderate performing**) is having good ratings but average sales. As observed the categories with less price is giving better sales and purchase frequency of those products is also high, so to improve the sales it is recommended to drop the price.
- Second cluster (**Best performing**) is having good ratings as well as sales. So, these categories should not be tampered.
- Third cluster (**Needs Improvement**) is having worst response in terms of rating as well as sales. So, complete overhauling is required to improve their ratings

# TIME SERIES ANALYSIS

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time.

## AIM

To get the scenario of the sales in future and optimize the inventory as per the demand.



# DEMAND FORECASTING OF PRODUCTS

with respect to their sentiments

## SPLITTING DATA INTO SENTIMENTS

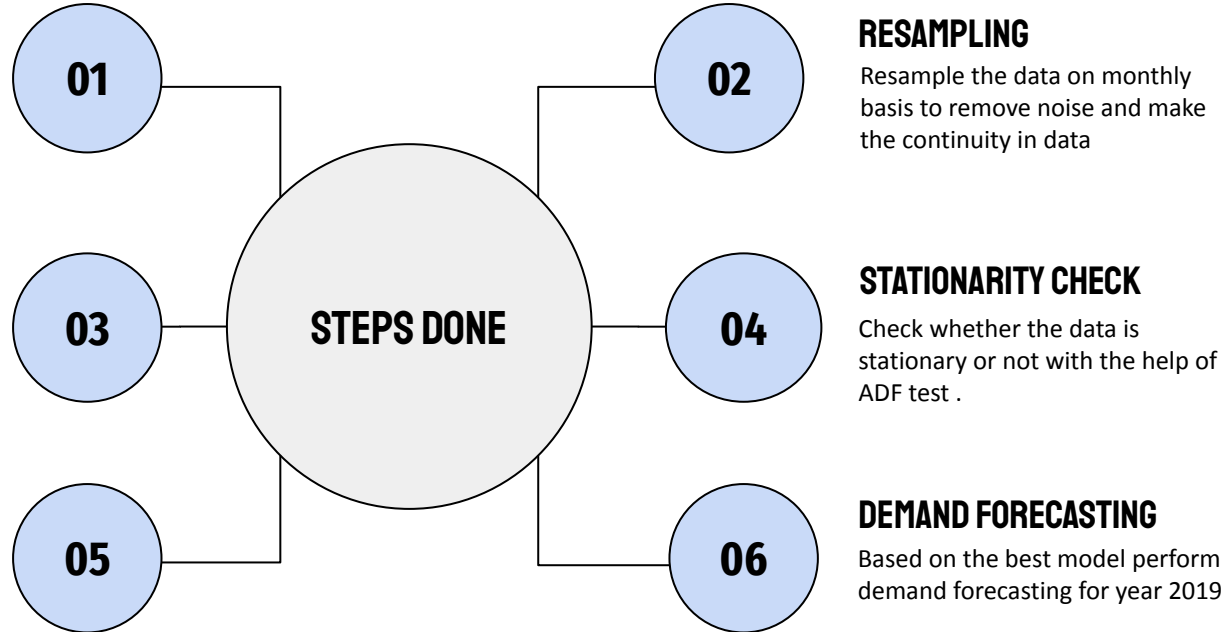
Using filtering split the data into positive, negative and neutral sentiments

## DECOMPOSITION OF DATA

Make the decomposition plot of each data to see the time series components in data .

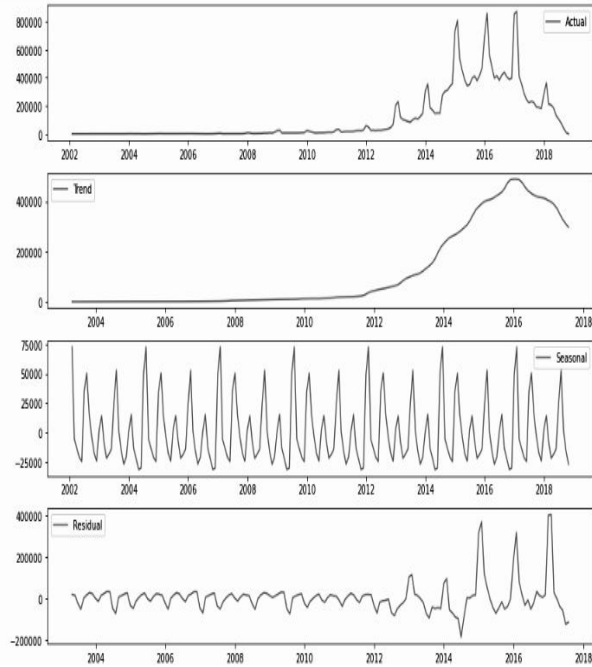
## TIME SERIES MODELING

SARIMA model on positive, negative, neutral sentiments products



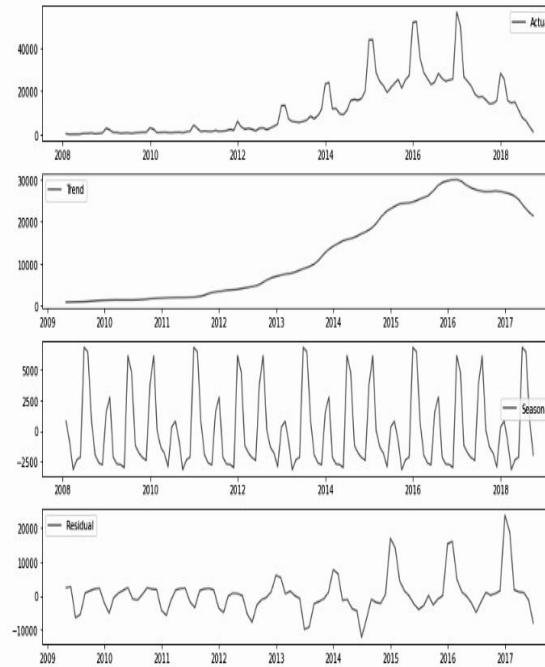


### DECOMPOSITION OF SALES OF PRODUCTS WITH POSITIVE SENTIMENTS



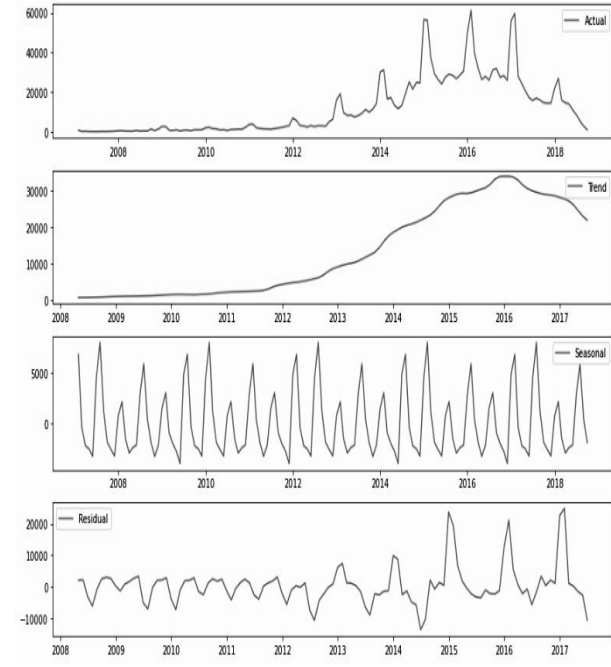
- **Trend:** Sales increased till 2016 and then started falling
- **Seasonality:** seasonality is observed after every 2 years and the fluctuations is between 2.5K to 75K

### DECOMPOSITION OF SALES OF PRODUCTS WITH NEGATIVE SENTIMENTS



- **Trend:** Sales increased till 2016 and then started falling
- **Seasonality:** seasonality is observed after every 2 years and the fluctuations is between 2.5K to 6K

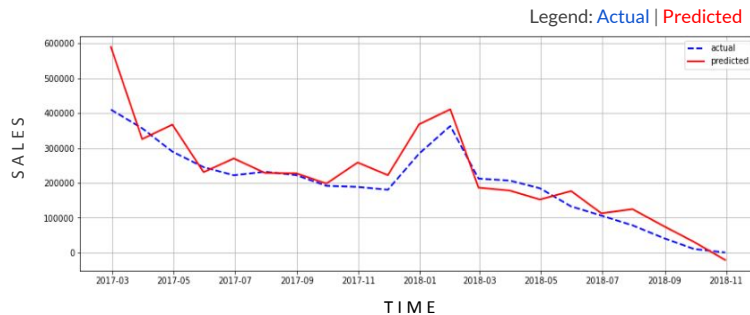
### DECOMPOSITION OF SALES OF PRODUCTS WITH NEUTRAL SENTIMENTS



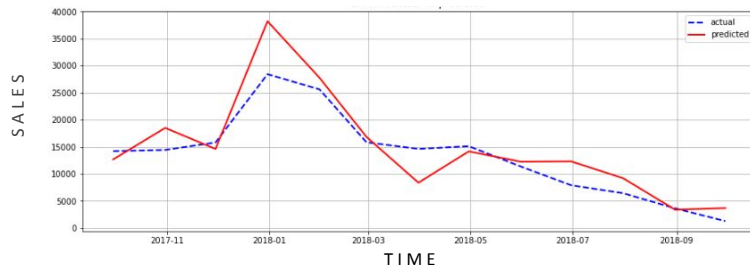
- **Trend:** Sales increased till 2016 and then started falling
- **Seasonality:** seasonality is observed after every 2 years and the fluctuations is between 1.5K to 6K

# ACTUAL AND PREDICTED SALES

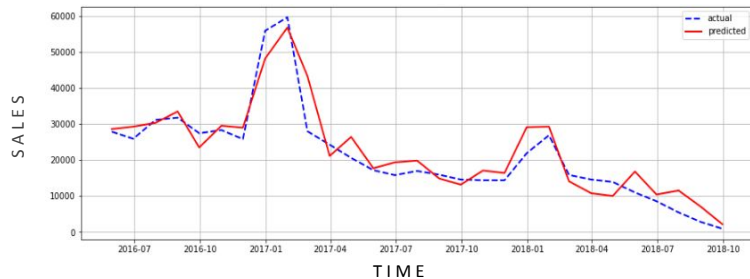
SALES OF PRODUCTS WITH POSITIVE SENTIMENTS



SALES OF PRODUCTS WITH NEGATIVE SENTIMENTS



SALES OF PRODUCTS WITH NEUTRAL SENTIMENTS



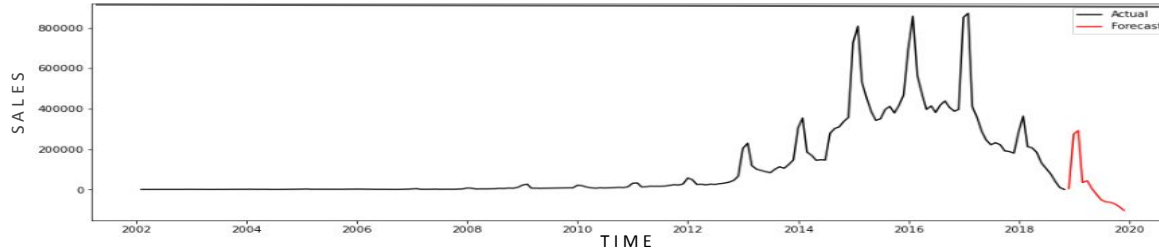
**SARIMA** model gives the low errors and the predicted values are closer to the actual values

# Forecasted total sales for 2019

for planning and strategizing inventory

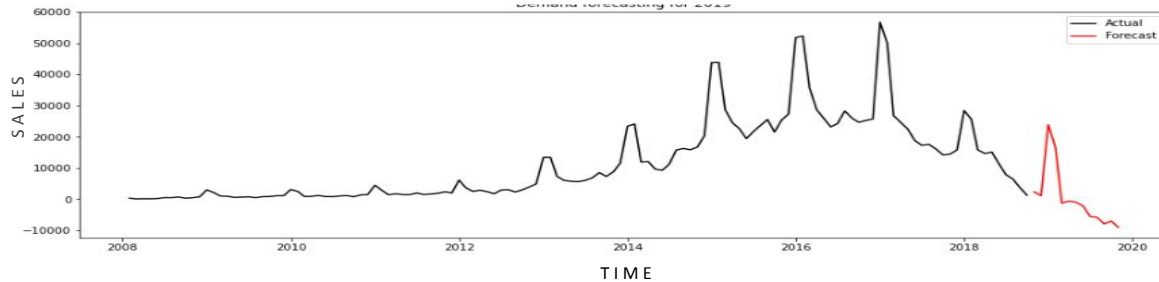
Legend: Actual | Forecast

Positive sentiment



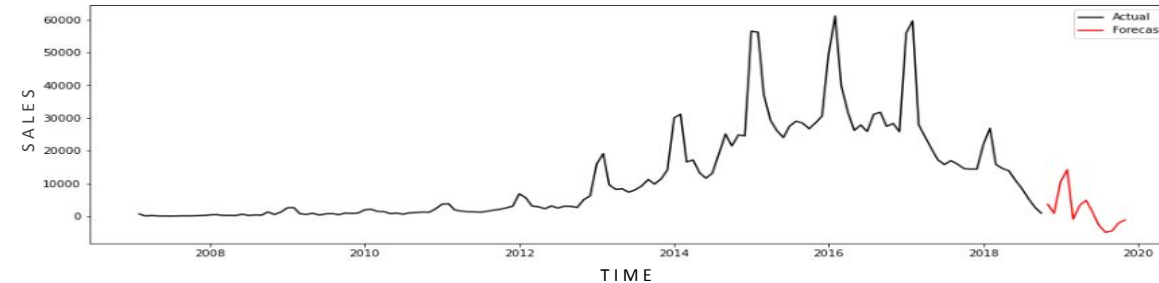
Sales for the products with positive reviews in 2019 will be highest in the end of first quarter around 2.5 lac and then decrease throughout the year .

Negative sentiment



Sales for negative reviewed products will hit peak at the end of 1st quarter around 20K and decrease as the year passes

Neutral sentiment

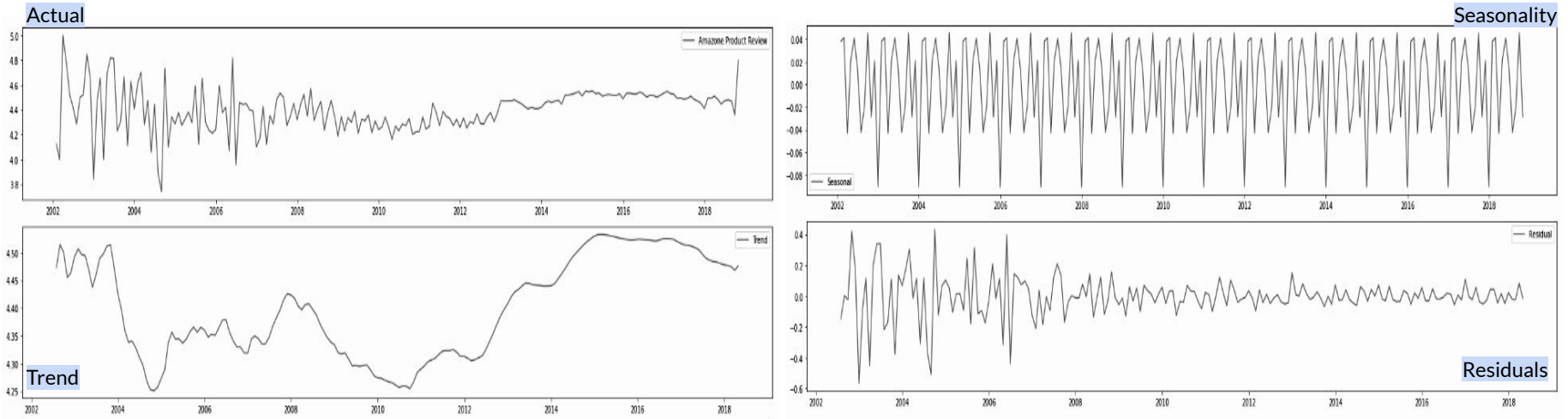


Sales for neural reviewed products will be highest around 10K at the end of first quarter and in 3rd quarter around 5K.

# SUGGESTIONS

- The total sales for the positively , negatively and neutral reviewed products will decrease in **2019** with comparison to **2018** can be the cause of default in products , high prices , not full fill the expectations of customer. So, we can optimize the inventory by managing the production because the demand of the products will be low throughout the year with comparison to the previous year.
- As we can observe that if the sales can decrease in 2019 we can **manage the production** and also we can give some **discounts & offers** and try to fulfill the **expectations** of the customers , and do more **advertisements** related to the products to attract the customers to increase the demand of the products .

# TIME SERIES ANALYSIS FOR OVERALL RATINGS



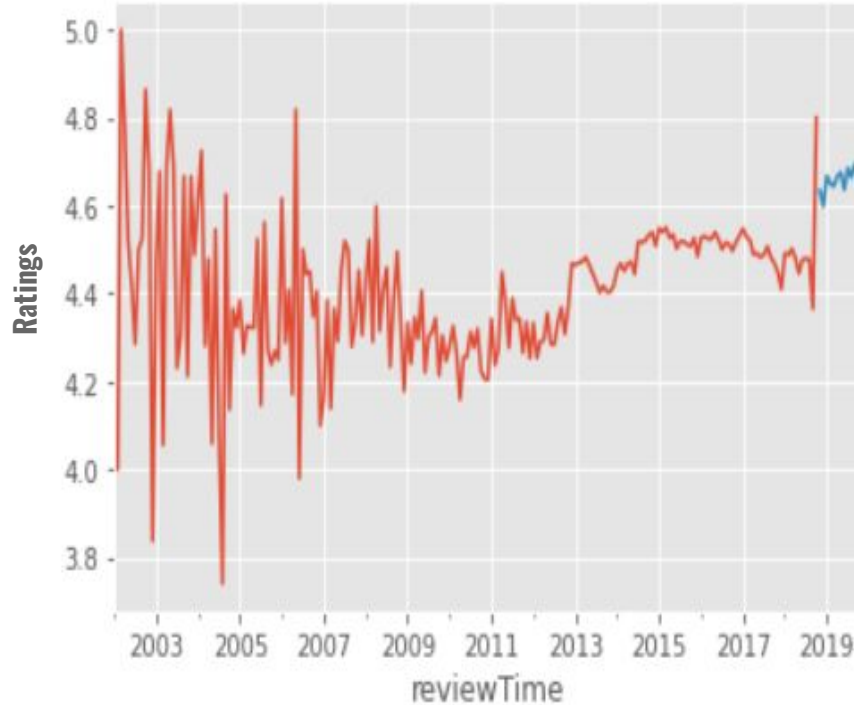
**Seasonality** - The pattern of data that is present over a specific period

**Trend** - The value of the data are increased or decreased over time, trend increased after 2012

**Residuals** - Irregular data falls under random type of chart

From visualization of the data we can find the trend and from exactly what year we can see a change in overall ratings

# FORECASTED RATINGS FOR YEAR 2019



## INFERENCE

- Overall **rating decreased after 2002** , after 2002 we can hardly see 5 point review for toys and games over the years.
- In 2007 where there is a steep high and low in overall reviews mainly because in 2007 Amazon released its **new logo** and the release of its first **kindle** in america.
- We are predicting the review Time , we can see a steep rise and stability after year 2013 , before 2013 we can see that there was no stability.
- Considering the data from 2002 to 2018 we get to a point that the customers are mostly **happy** with the **toys and games** section of amazon.

# RECOMMENDATIONS

- Moderate performing categories is having good ratings but average sales. As observed the categories with less price is giving better sales and purchase frequency of those products is also high, so to improve the sales it is recommended to drop the price.
- Categories which needs improvement is having worst response in terms of rating as well as sales. So, complete overhauling is required to improve their ratings.
- The total sales for the positively , negatively and neutral reviewed products will decrease in **2019** with comparison to **2018** can be the cause of default in products , high prices , not full fill the expectations of customer. So, we can optimize the inventory by managing the production because the demand of the products will be low throughout the year with comparison to the previous year.
- As we can observe that if the sales can decrease in 2019 we can **manage the production** and also we can give some **discounts & offers** and try to fulfill the **expectations** of the customers, and do more **advertisements** related to the products to attract the customers to increase the demand of the products .

**THANK YOU**