

# Amazon Analysis



By:

**Mohammed Affan Sabir Khan**  
[sabirkhan.affan8499@gmail.com](mailto:sabirkhan.affan8499@gmail.com)

**Neha J Kotwal**  
[nehajbk@gmail.com](mailto:nehajbk@gmail.com)

**Prashant Yadav**  
[3.prash@gmail.com](mailto:3.prash@gmail.com)

**Batch : DSFT 12 NIIT STACK ROUTE**

# Contents

## PROJECT OVERVIEW

To make data inspired decision by turning review history into knowledge, monitoring trends that would help in increasing the success rate of existing and new products.

**UNDERSTANDING**  
Data Breakdown & EDA

**ANALYSIS & STATISTICS**  
Providing Insights

**SOLUTIONS**  
Insights to Conclusions



=

# Why is the analysis required ?



## Product Reviews and their influence.

Online product reviews help influence the buying decisions of potential customers. Reviews left by customers can help businesses develop more effective marketing plans.

## Category Division

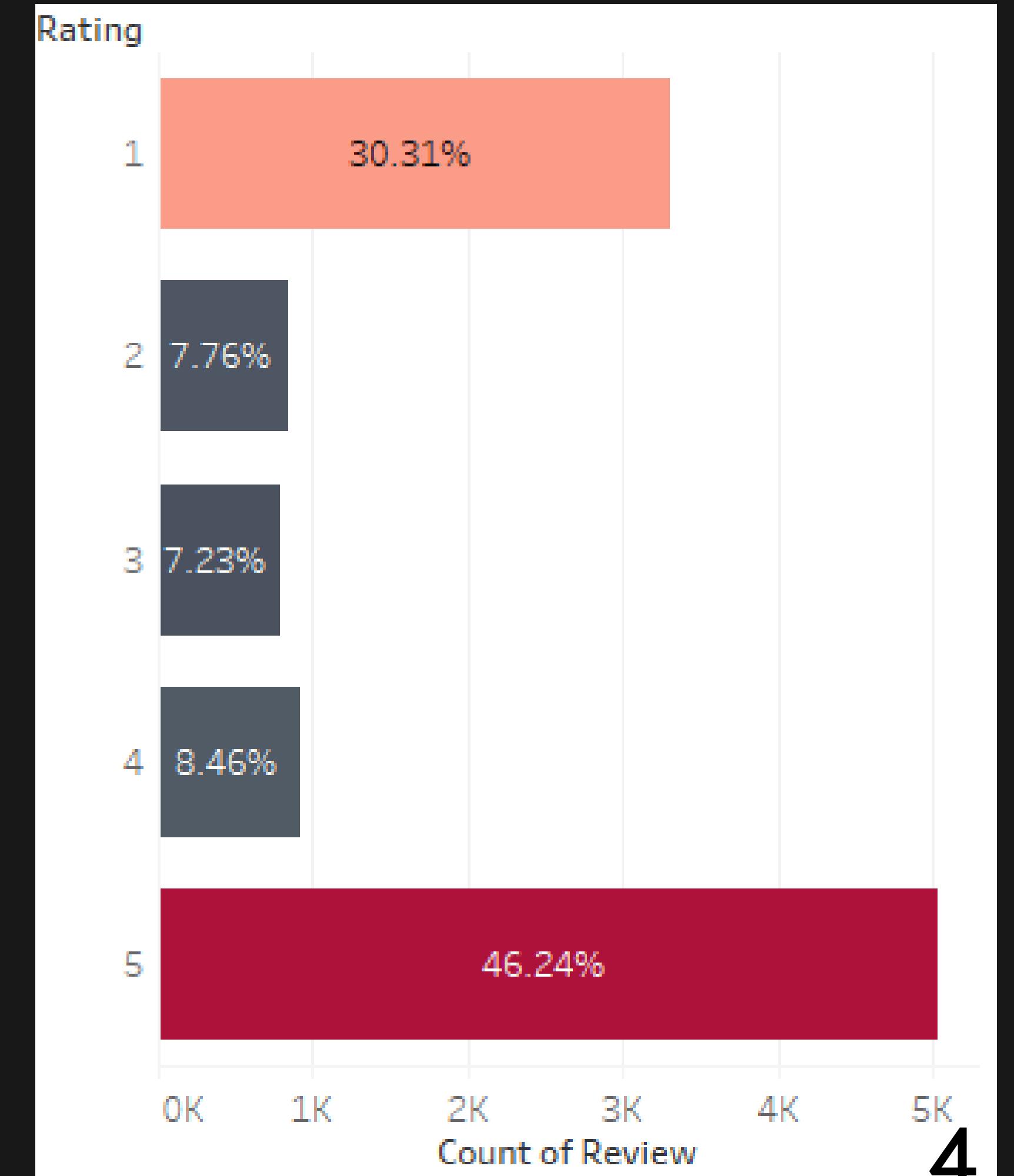
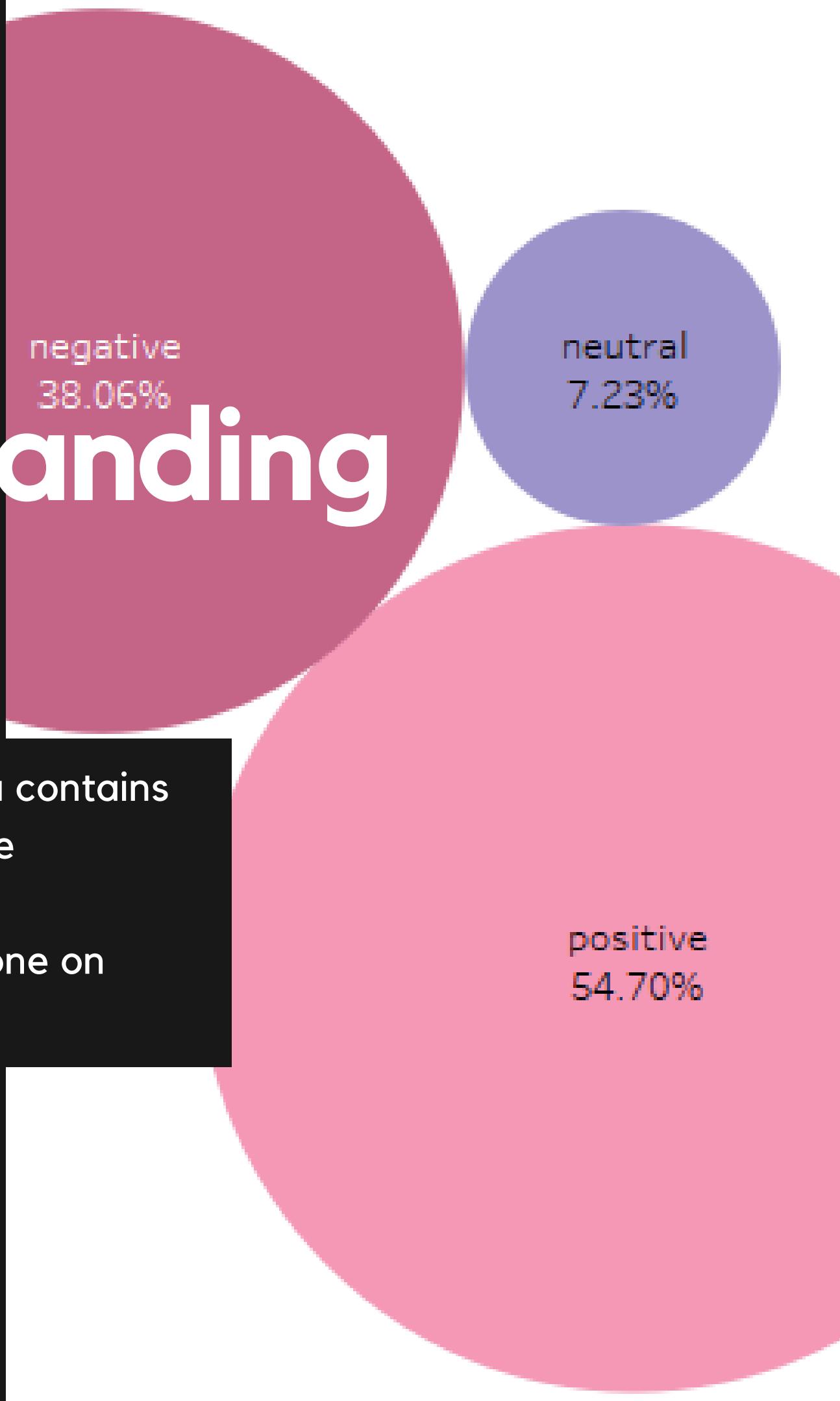
1. Grocery & Gourmet Food
2. Amazon Pantry

Both the datasets taken here are based on similar topics here.

# Data Understanding

More than half of the data contains reviews containing positive sentiment.

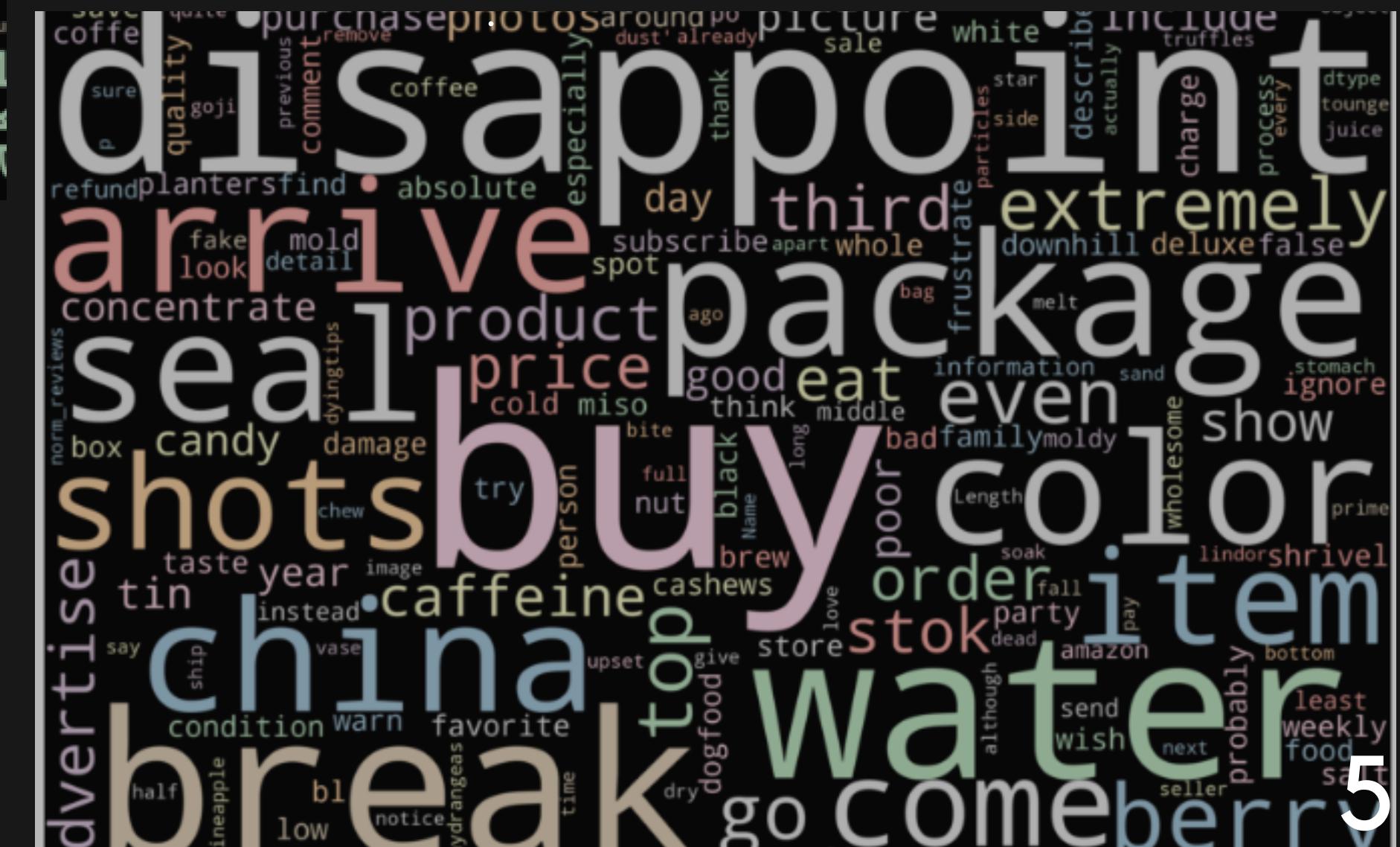
Sentiment classification done on ratings provided.

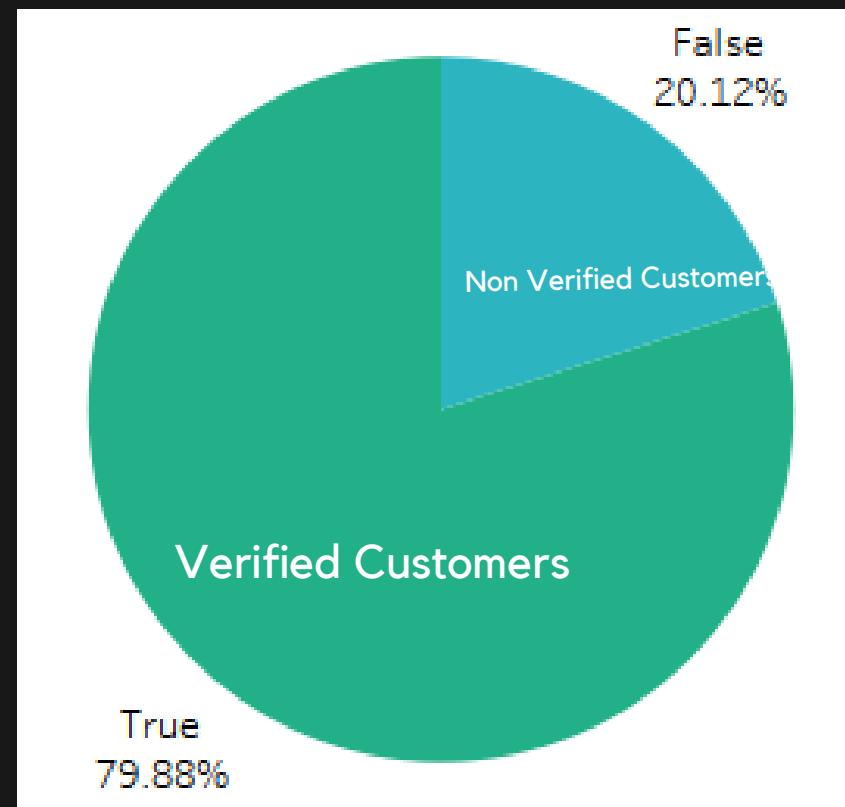


# Popular words in Positive Reviews.

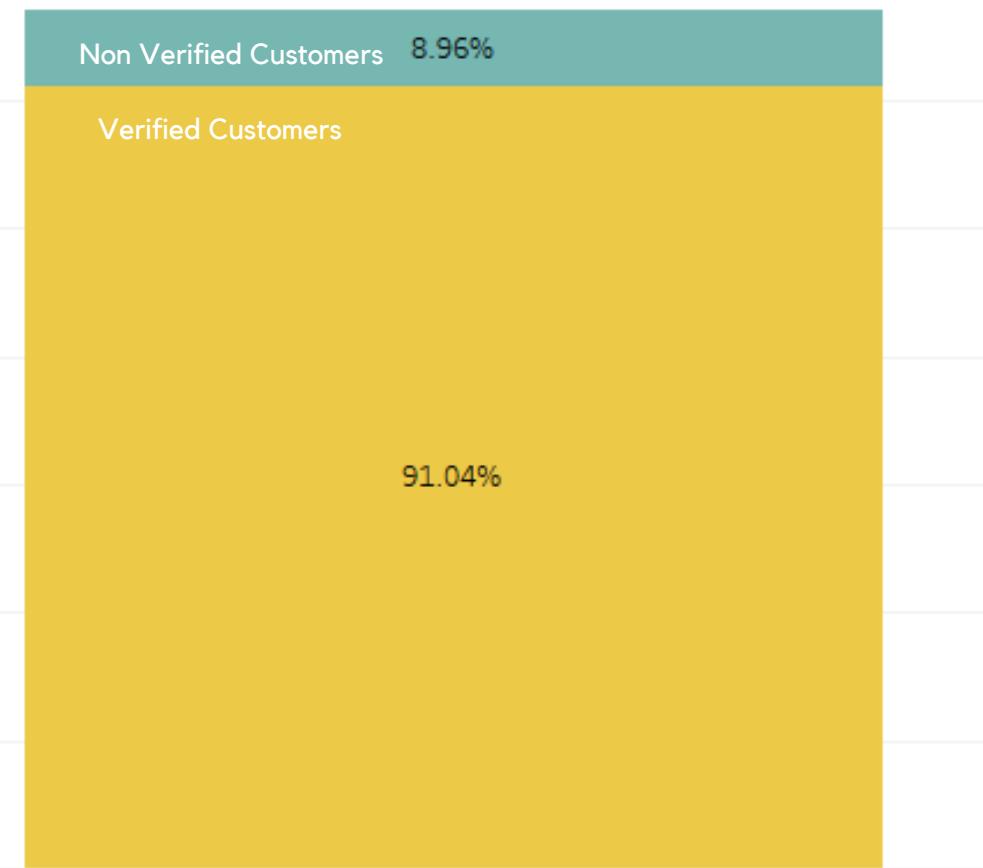
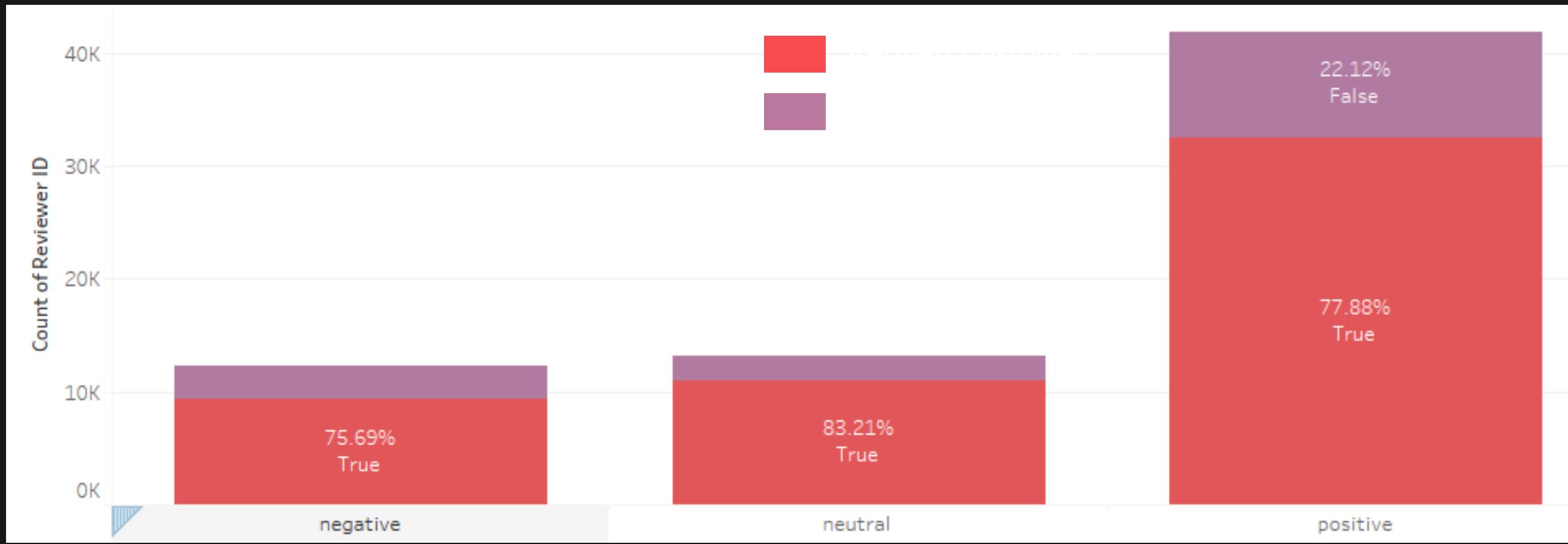


# Popular words in Negative Reviews.





# FRAUD VISUALISATION

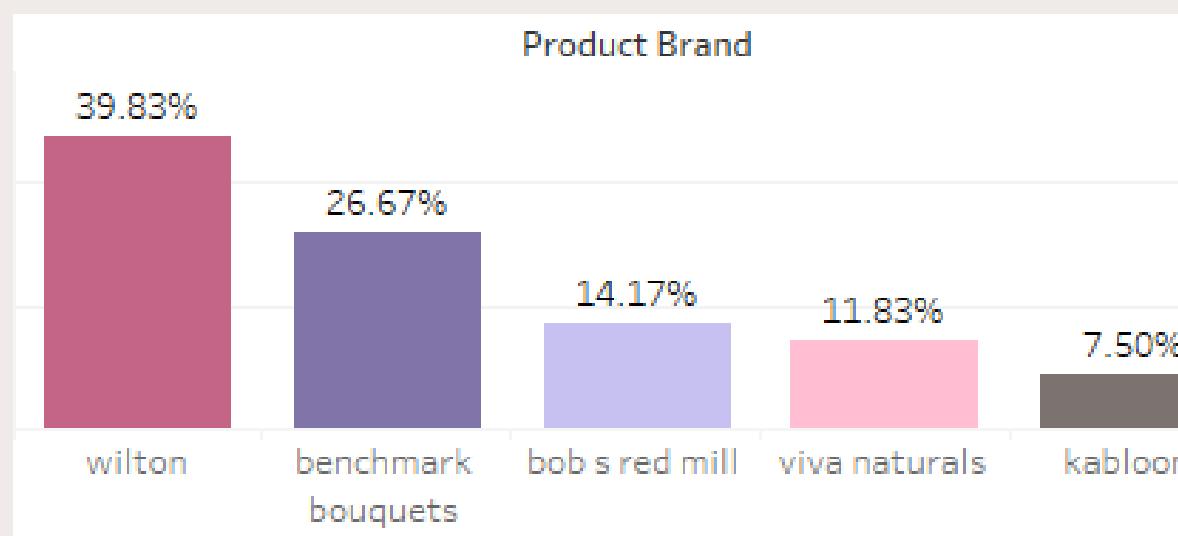


Comparing with overall dataset

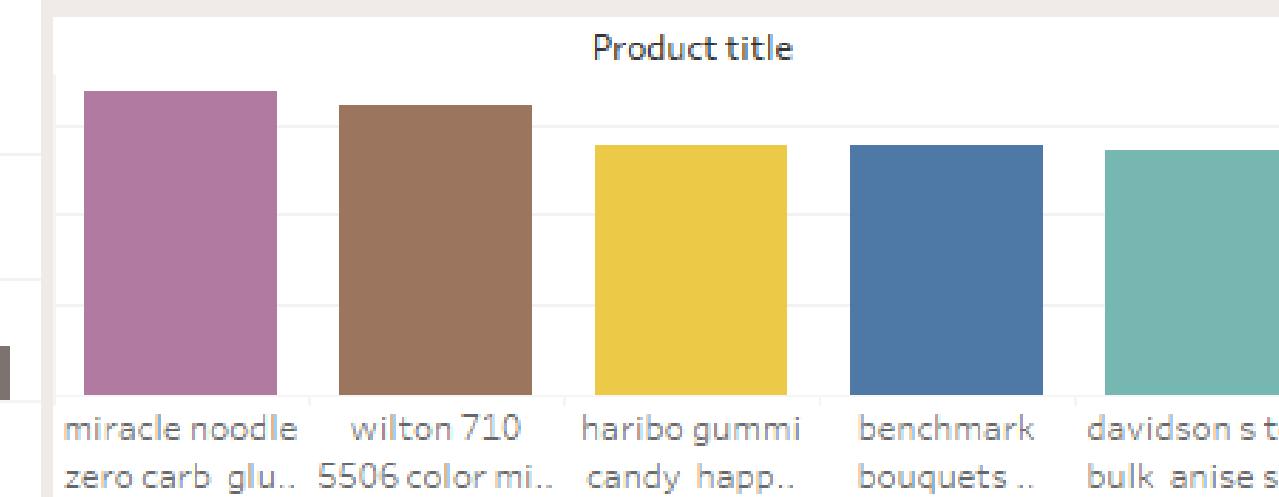
Spotting fake amazon reviews! When 5 star reviews aren't what they seem like :(

# GROCERY PRODUCT REVIEW BASIC ANALYSIS

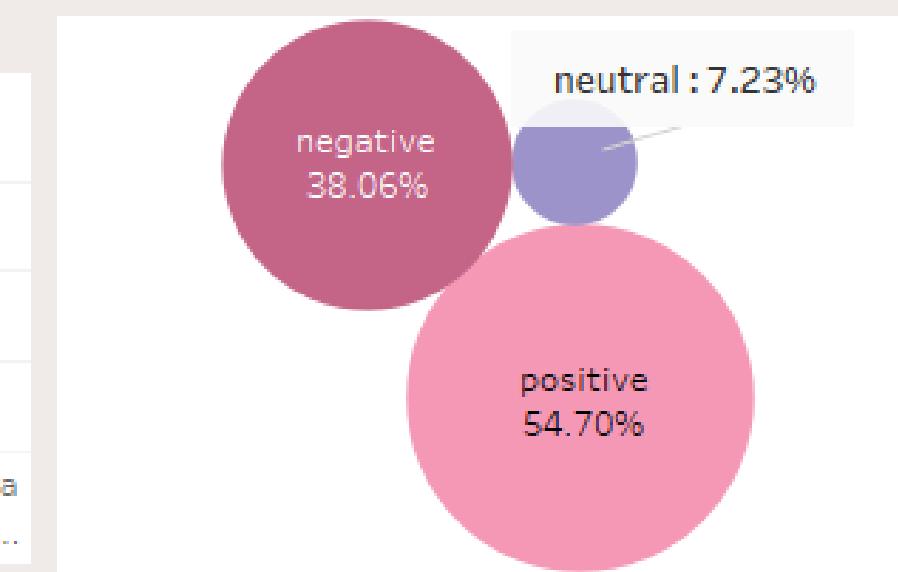
TOP 5 BRANDS WITH MAX POSITIVE REVIEWS



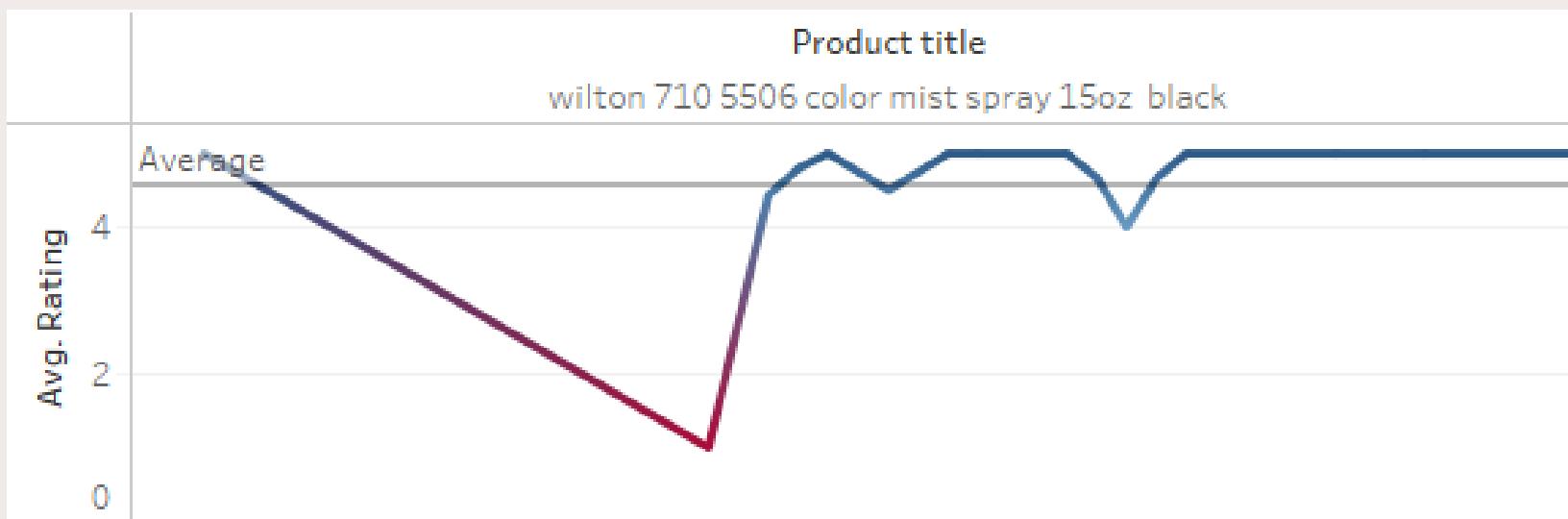
TOP 5 MOST REVIEWED PRODUCTS



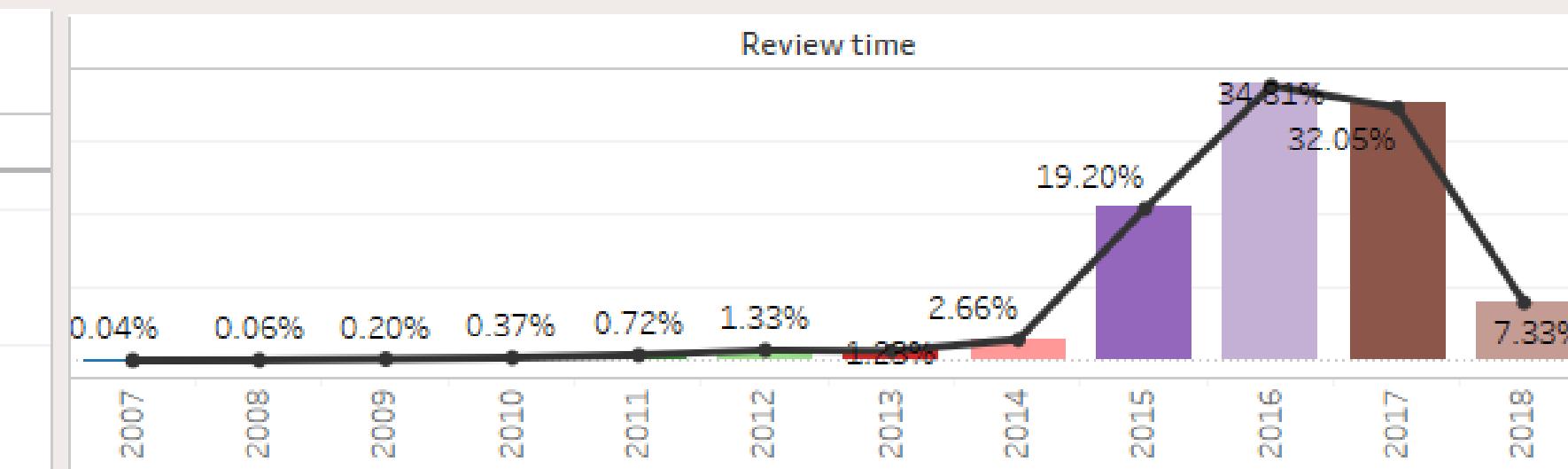
DATA DISTRIBUTION SENTIMENT WISE



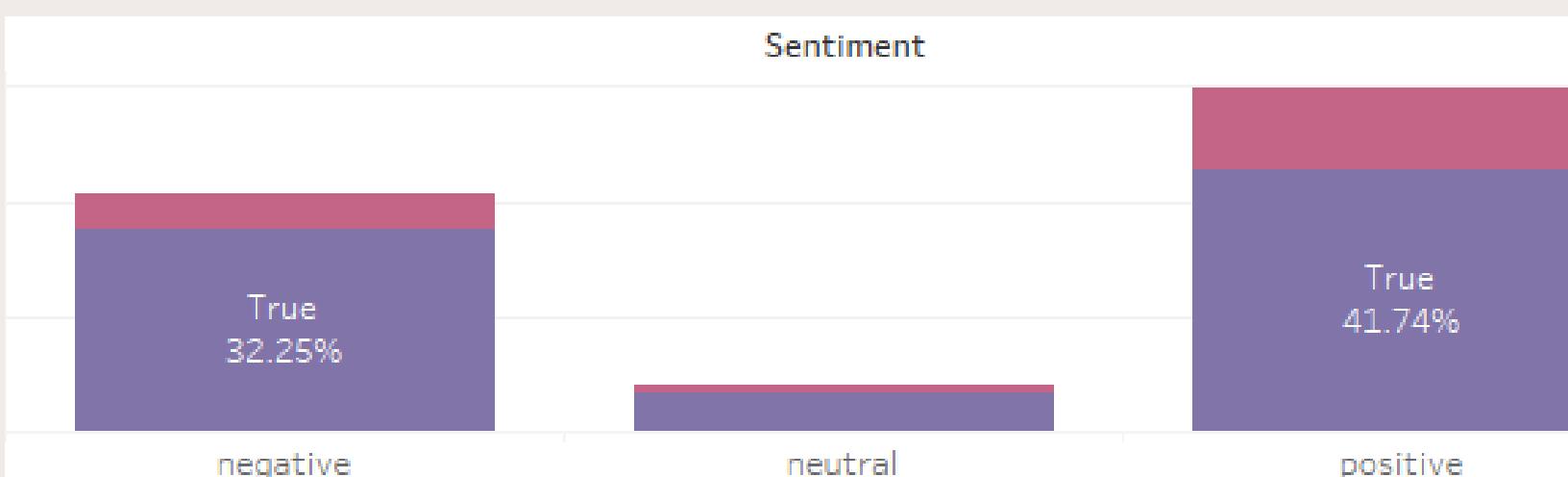
RATING TRENDS OF MOST REVIEWED PRODUCT



WHICH YEAR HAS THE HIGHEST NUMBER OF CUSTOMERS?



HOW MANY SENTIMENT CATEGORIES ARE ACTUALLY VERIFIED PURCHASES?

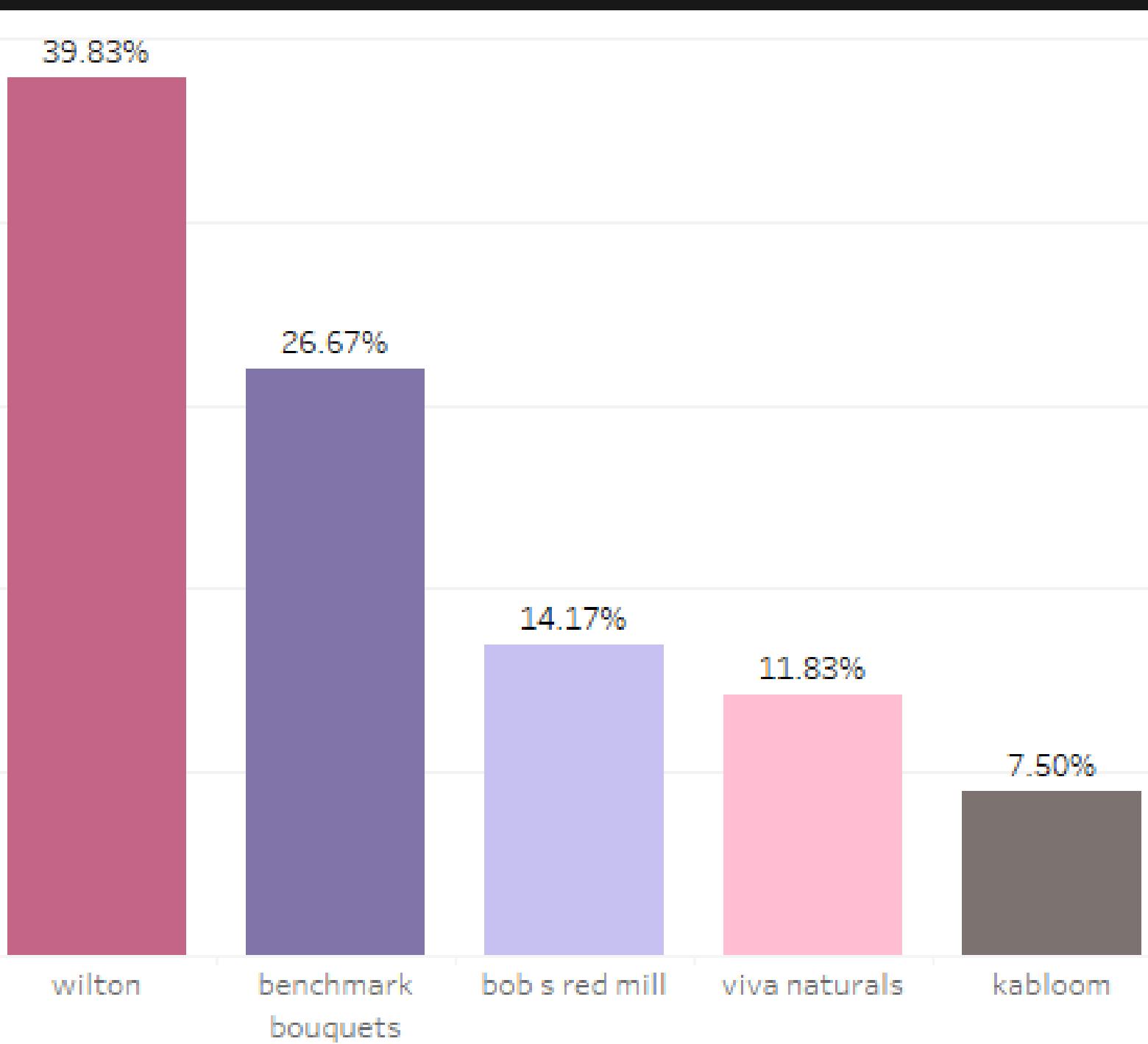


## INTERPRETATION:

1. 54.70% of the data has positive (rating range 4-5) reviews.
2. Wilton and Benchmark Bouquets are top most emerging brands review wise.
3. Miracle Noodle Shirataki Pasta product has the highest demand with respect to the amount of reviews flowing in.
4. Wilton's most in demand product has an average rating of 4.6
5. 2016 has seen great inflow of customers accounting to around 35% of the total customers from 2007 to 2018 year range.
6. 42% of the positive reviews are verified customers with 13% being in the non verified category.

=

# Emerging Brands for Grocery Data



## Wilton Industries

Leading food crafting company in the industry, dealing in cake decorating, bakeware etc.



## Benchmark Bouquets

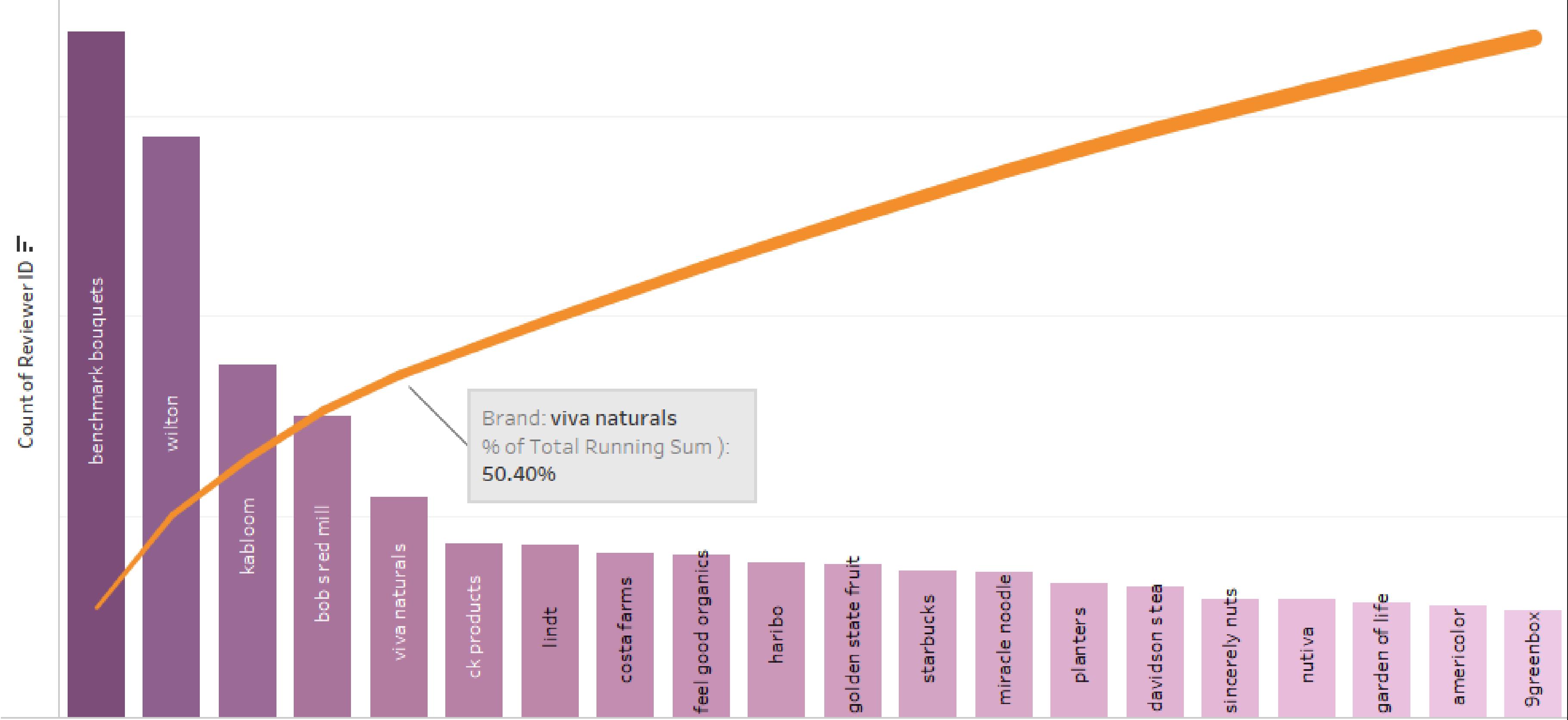
Farm fresh flowers delivered fast, throughout the year, through amazon delivery.

## Most Positively Reviewd Brands

A list of brands ordered by their positive review count. The brands are displayed in a staggered, overlapping style. Some brand names are in green, while others are in red.

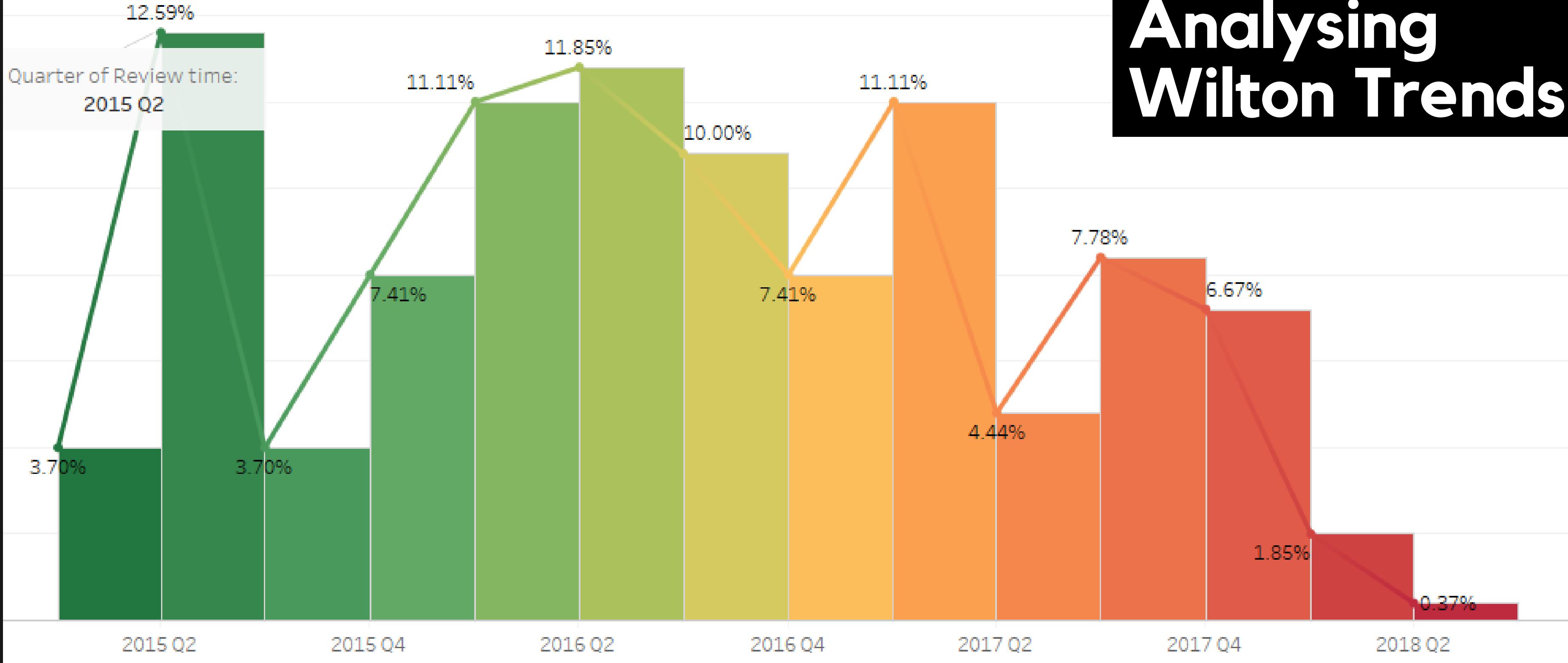
- wilton
- davidson s tea
- kabloom
- lindt
- benchmark bouquets
- haribo
- ck products
- bob s red mill
- viva naturals

# Insights on Grocery Data



50% of the total revenue comes from the initial five brands.

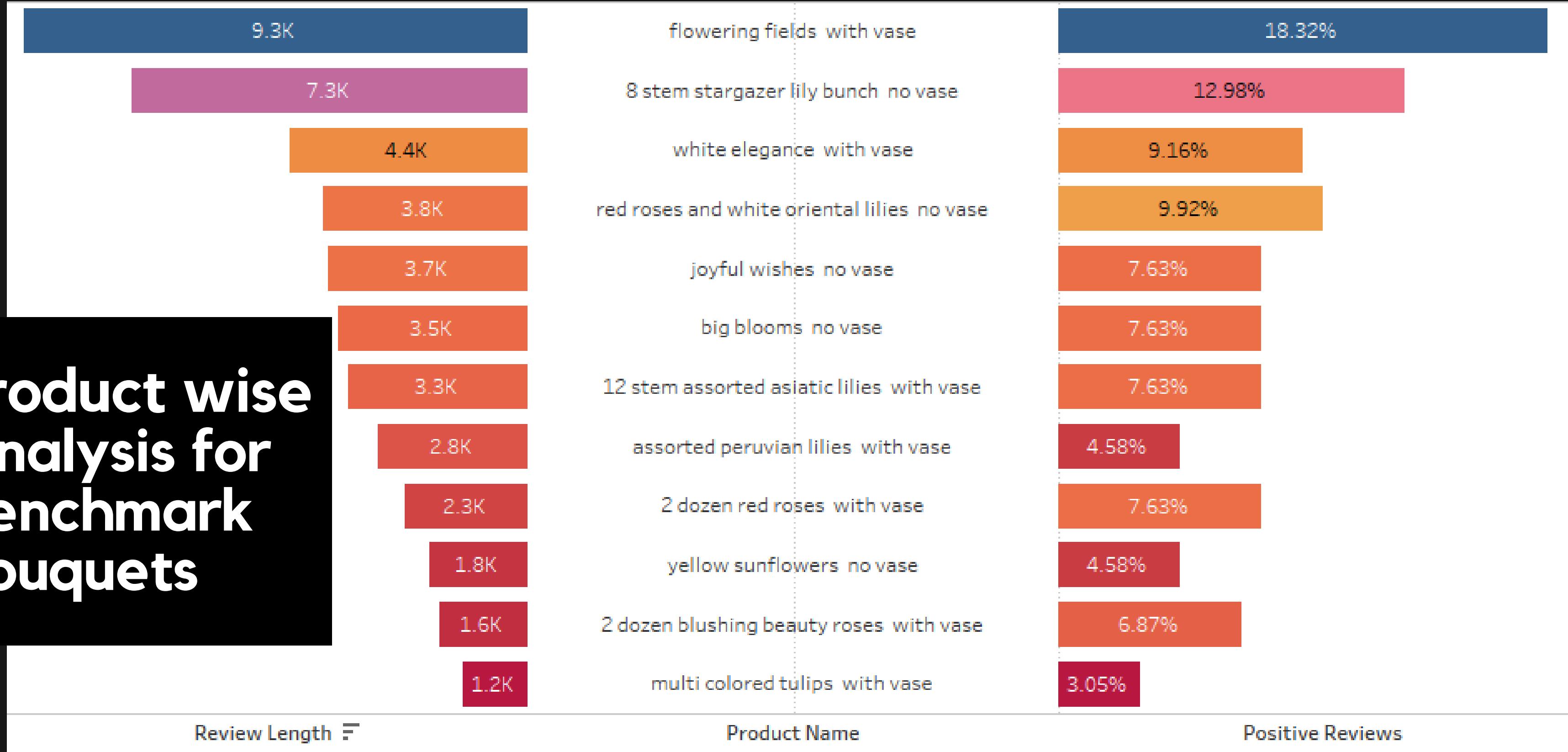
# Analysing Wilton Trends



The year 2015 to 2017 have been very stable for Wilton, with 2016 being the most profitable year and the sales was max in the Q2 of 2015.

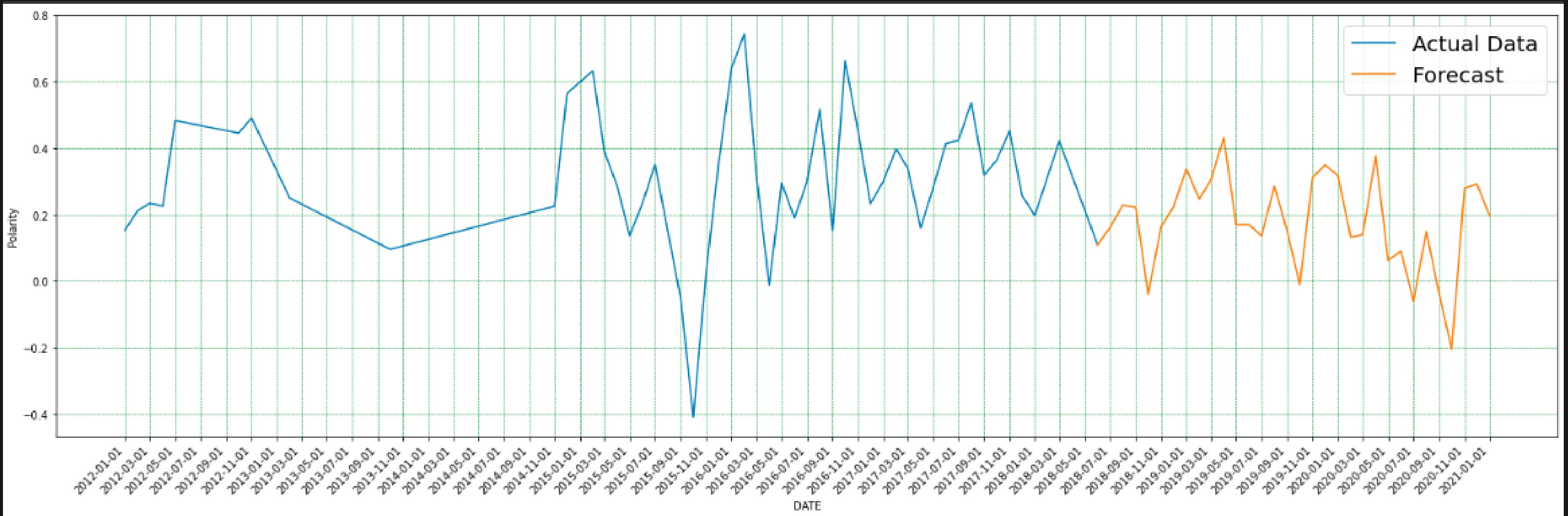
=

# Product wise Analysis for Benchmark Bouquets



More excited and satisfied customers are with the product, more they write on their reviews :).  
No doubt flowering fields is their number 1 seller.

# Forecasting for Benchmark Bouquets



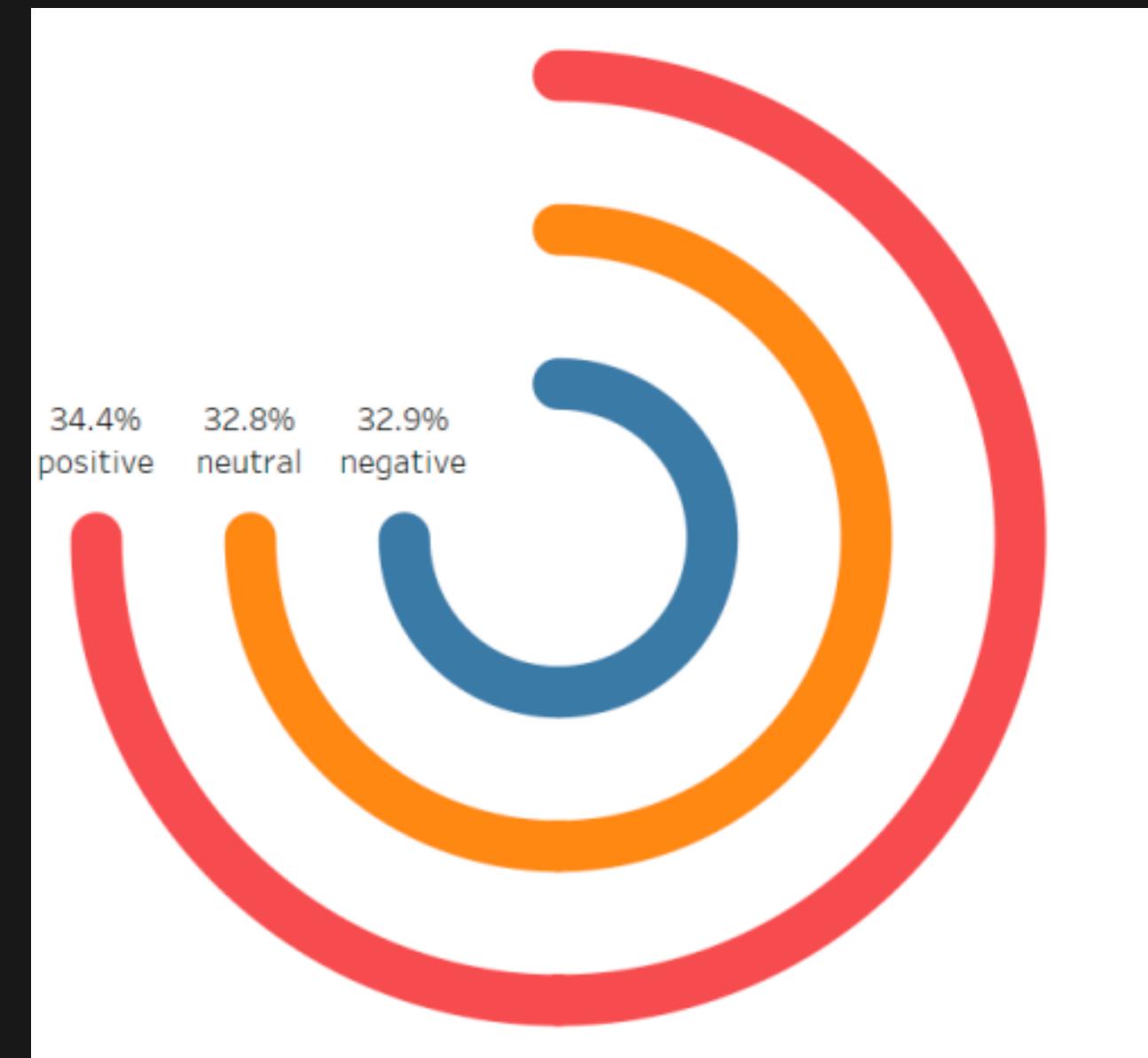
Dips are seen in the month of November in almost every year from 2015 - 2018 and the patterns are observed in the forecasted data as well.

=

# Emerging Brands for Pantry Data



Twinings has the most number of verified reviews.  
Data is distributed almost equally sentiment wise.

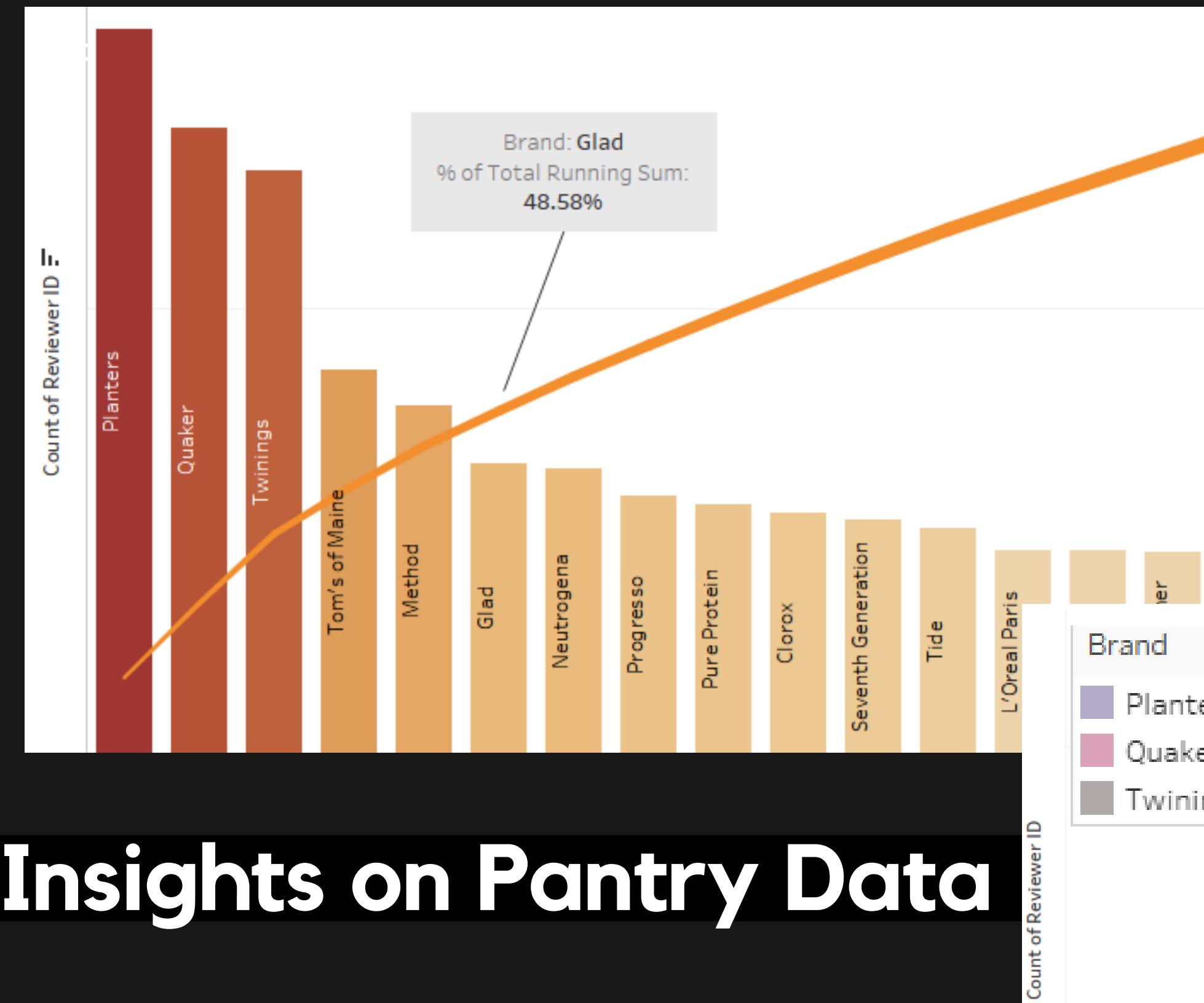


## Sentiment Distribution for Pantry Data

Pantry Data Tableau Link

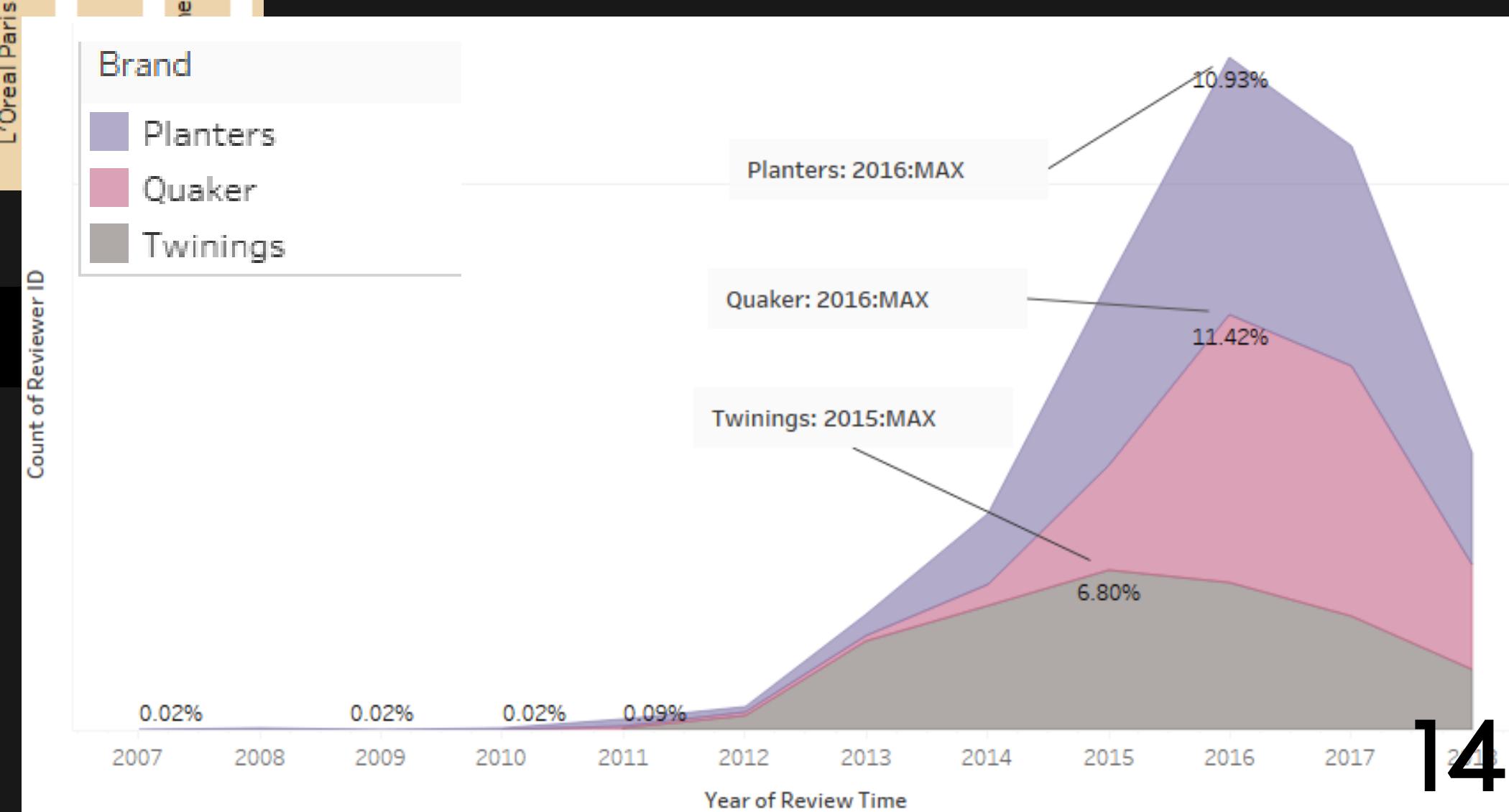
## Insights on Pantry Data

50% of the total revenue comes from the initial six brands.



## Analysing Trends

In 2016 Planters, Quaker brand offered most of the profit in the Pantry Category, while in 2015 Twining pitched in the most.



# Business Improvement Strategies

## 1. HOT SELLING PRODUCTS TO BE IN STOCK

miracle noodle zero carb gluten free shirataki pasta spinach angel hair 7 ounce

35.34%

twinings of london pure peppermint herbal tea k cups for keurig 12 count

wilton 710 5506 color mist spray 15oz black

22.64%

planters nutrition cashews almonds macadamias 9 75 oz

haribo gummi candy happy cola 5 pound bag

17.33%

pure protein 23 grams vanilla creme 4 count 12 fl oz each

benchmark bouquets flowering fields with vase

12.37%

tide original scent he turbo clean liquid laundry detergent 64 loads 100 fl oz packaging may vary

davidson s tea bulk anise seed 16 ounce bag

## 2. KEY BENEFITS OF INVENTORY MANAGEMENT

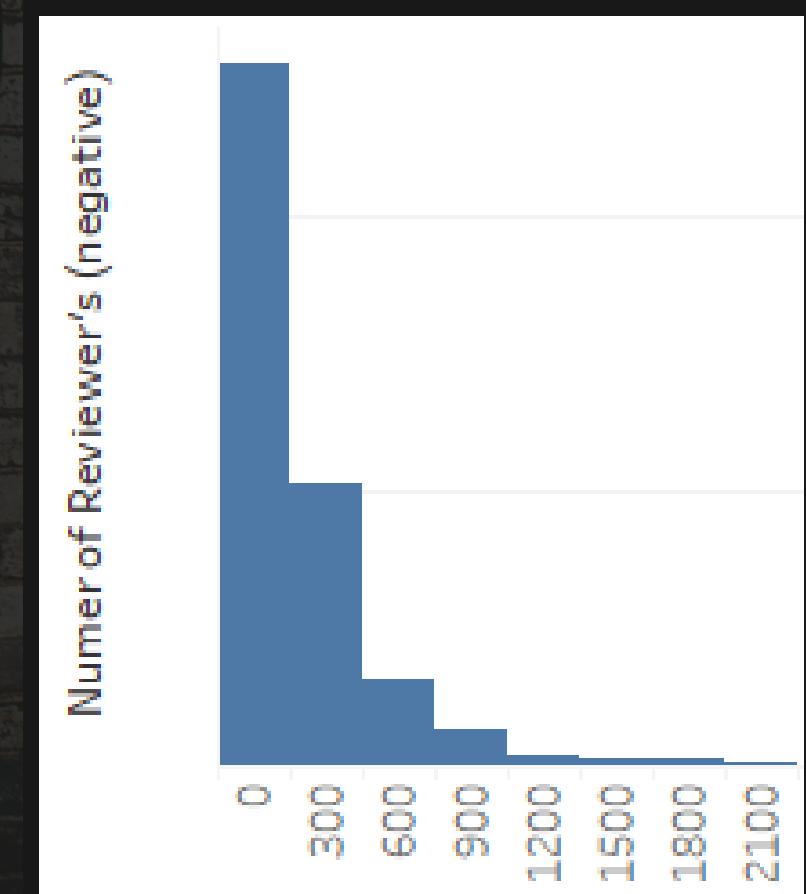
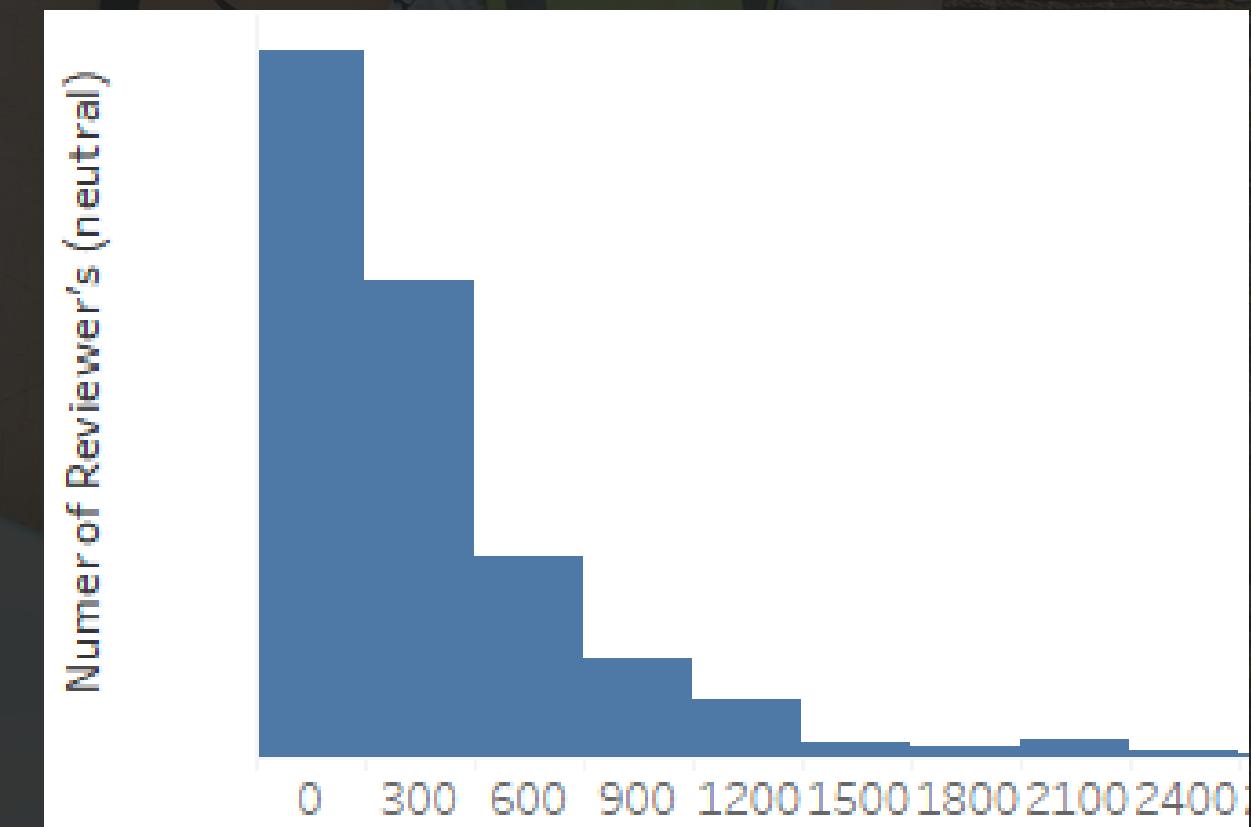
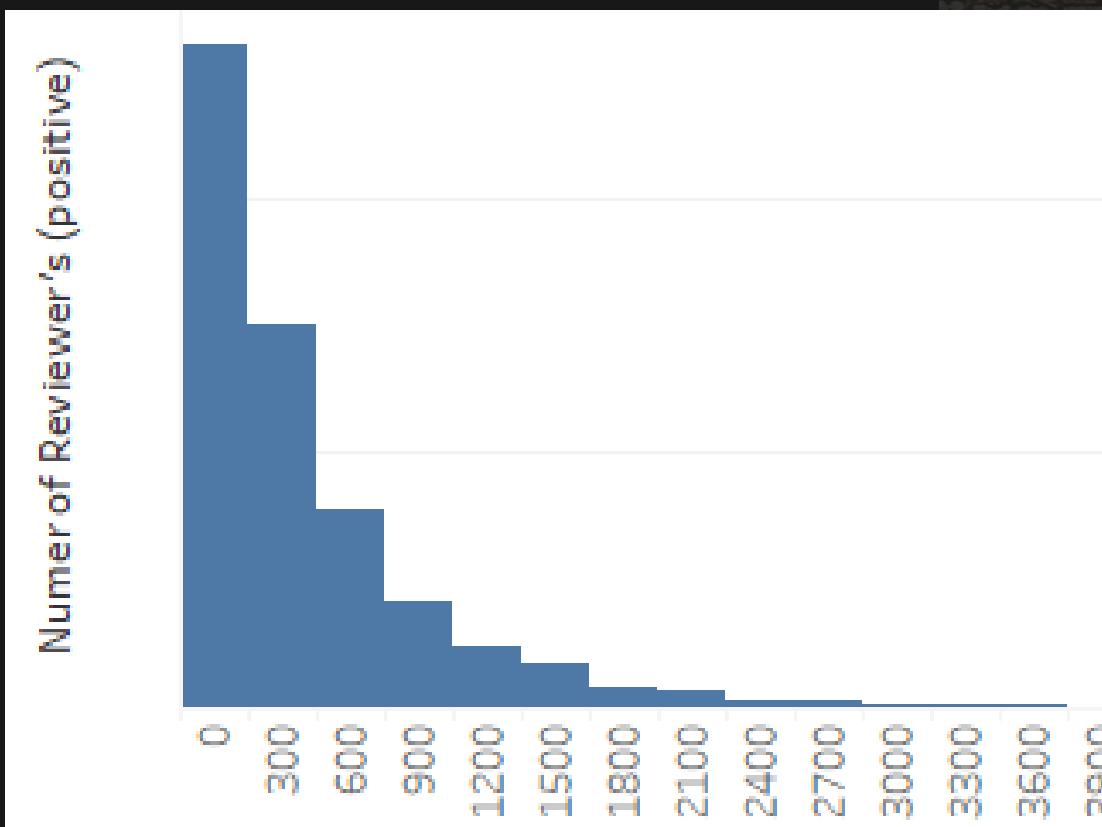
- Keeping track of recommended stocks - Managing stock-outs - Meeting customer expectations - Improved delivery.
- Forecasting, Safety Stock and Batch Tracking best strategies which can be followed.
- Brands to be prioritised according to our analysis in
  - Kitchen Pantry Planters, Quackers, Twinings / - Home Pantry Glad, Clorox / - Personal Grooming category we have Method, Neutrogena.

### 3. MANAGING PLANNING USING FORECASTING

- Stock in management should be doubly checked for all months, especially around festivities in the US market.
- Taking Benchmark Bouquet's trends, dips are seen in the month of November in almost every year due to the rainy season, stock storage should be handled and prioritised carefully.

### 4. AMAZON'S DEPENDENCY ON RECOMMENDATION ENGINE

- By behavioral analytics, customers tend to purchase products with enormous reviews and best star ratings.
- From our analysis since amazon leverages its data via its recommendation engine length of the reviews text with the star rating taken into consideration is double effective.



# How we turned data into knowledge?

Model Performances & More.

**Sentiment Analysis**  
Anaysing tones,  
launguage, keyowrds.

**Time Series Approach**  
Building descriptive and  
predictive models.

**Machine Learning Approach**  
Applying Supervised ML  
algorithms.

# Libraries Used

## Sentiment Analysis

Vader, Afinn, Spacy, Gensim, Text Blob.

## Supervised Learning

Sklearn Library, PyCaret Library.

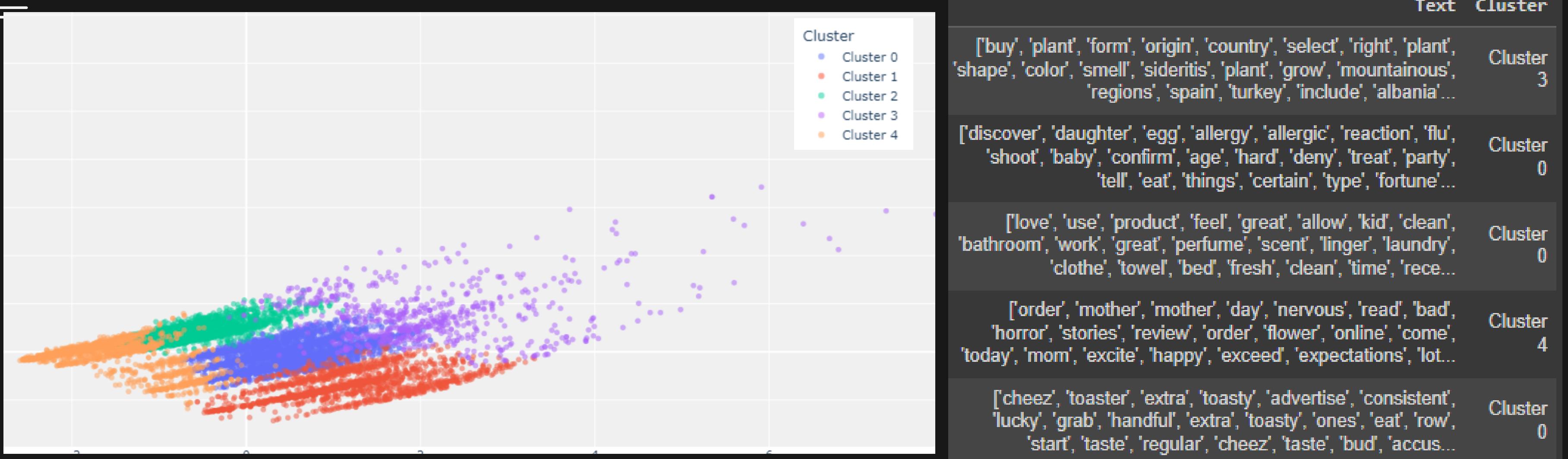
## Time Series Analysis

SARIMAX, ARIMA Models. AutoTS Library.

Glimse of some technical detailing

MODEL	F1_score	Accuracy	recall	precision
Text_blob	0.422071	0.454543	0.454543	0.461814
Vader_analysis	0.422940	0.459211	0.459211	0.547035
Affin_Analysis	0.444613	0.474956	0.474956	0.510949

MODEL	F1_score	Accuracy	recall	precision
Logistic_regression	0.926153	0.927350	0.927350	0.926353
RANDOM FOREST	0.768004	0.793944	0.793944	0.830501
SVC	0.941821	0.942514	0.942514	0.941654



Dominant_Topic	Topic_Keywords	Num_Documents	Perc_Documents
0.0	flour, cookies, bread, gluten, cookie, yeast, free, pizza, dough, bake	56	0.0051
1.0	coffee, cup, bean, roast, flavor, pod, brew, keurig, dark, taste	475	0.0436
2.0	box, order, package, product, open, receive, bag, arrive, look, picture	3765	0.3457
3.0	mix, use, salt, add, seed, bake, egg, powder, cook, recipe	333	0.0306
4.0	taste, bottle, buy, good, time, water, use, try, know, drink	1500	0.1377
5.0	good, taste, love, great, buy, flavor, bag, price, eat, review	2205	0.2024
6.0	tea, flavor, green, drink, taste, bag, teas, matcha, good, water	706	0.0648
7.0	oil, cook, add, rice, sauce, noodles, water, chicken, use, spice	422	0.0387
8.0	cake, candy, color, great, look, perfect, love, work, use, birthday	757	0.0695

# Technical Innovation

## PyCaret Library

### LDA using GENSIM

# Thanks!

**Questions? Comments? Concerns?**

Contact us to discuss anything you'd like. We love (virtual) coffees and always welcome the opportunity.

## Meet the team

**Mohammed Affan Sabir Khan**  
[sabirkhan.affan8499@gmail.com](mailto:sabirkhan.affan8499@gmail.com)

**Neha J Kotwal**  
[nehajbk@gmail.com](mailto:nehajbk@gmail.com)

**Prashant Yadav**  
[3.prash@gmail.com](mailto:3.prash@gmail.com)

**Batch : DSFT 12 NIIT STACK ROUTE**