'To be earth most customer centric company and build a place where people can discover anything' -Amazon

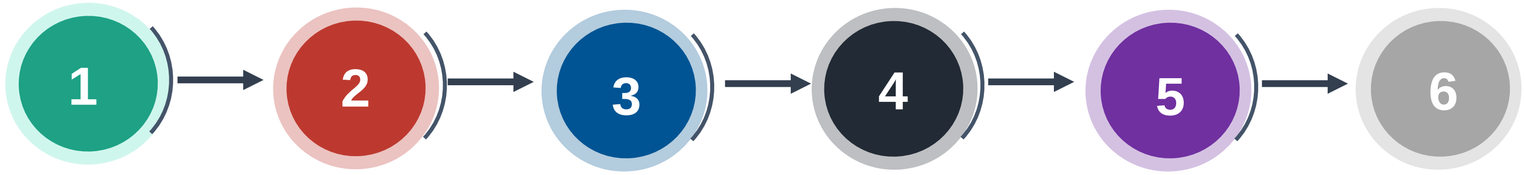
Recommendation System

Based on Amazon Reviews | Analyzed & Designed by Team BRIE

amazon

Summary Report

Overview



1 Business Analysis

1.1 Business Case

- Improve personalization

1.2 Market Study

2 Data Understanding

2.1 Data Extraction

- GZ zip format -> Json file -> .csv data frame

2.2 Data Understanding

3 Data Preparation

3.1 Exploratory Analysis

9
Columns

568,454
Instances

3.2 Pre-processing

- Text Analysis
- Tokenization
- Word Count
- Final Dataset

(Dataset: 365 x 360)

- Test Train Split

4 Modeling

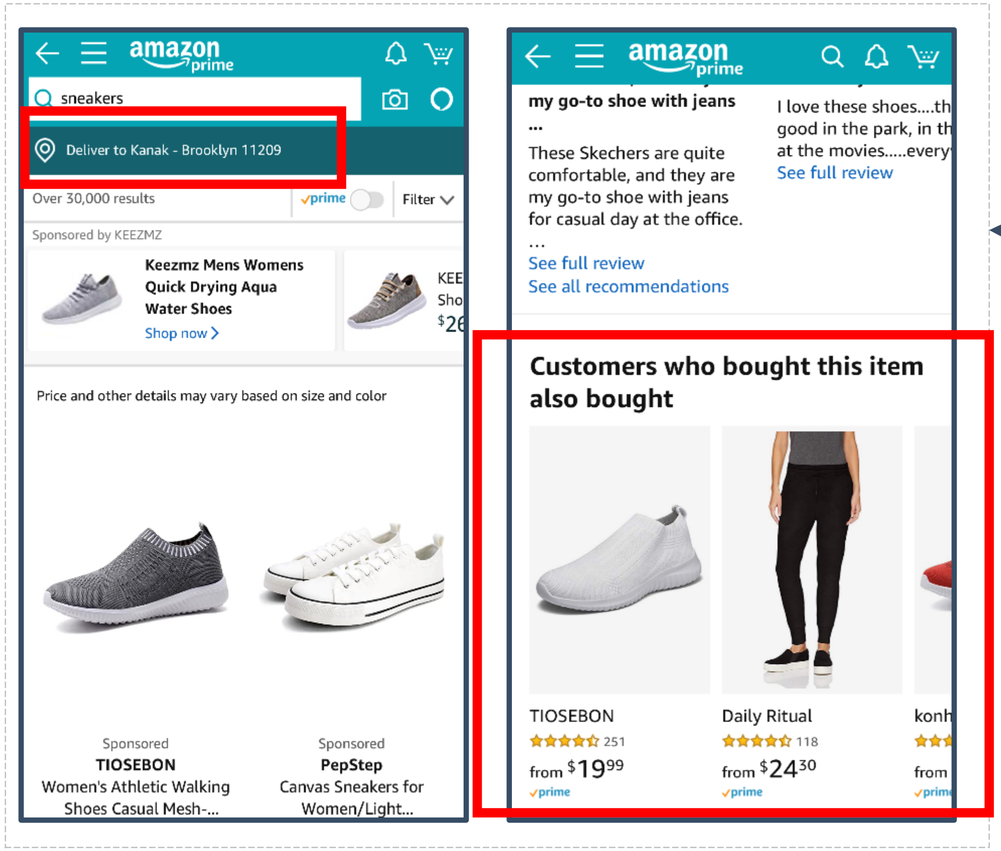
4.1 Sentiment Analysis-Text Analysis

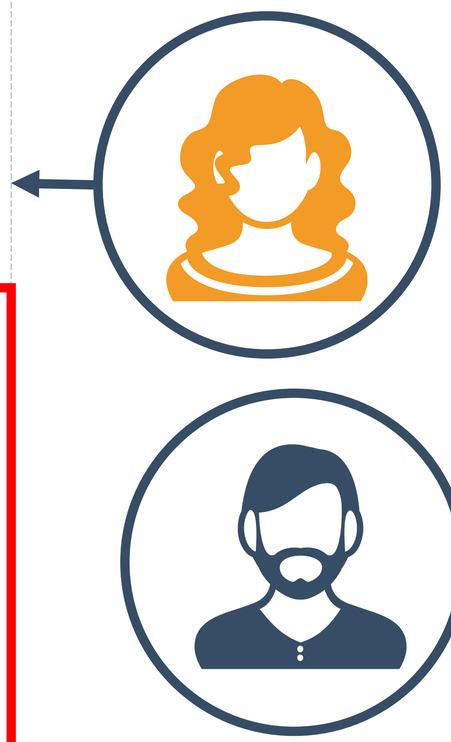
4.2 K-means Clustering

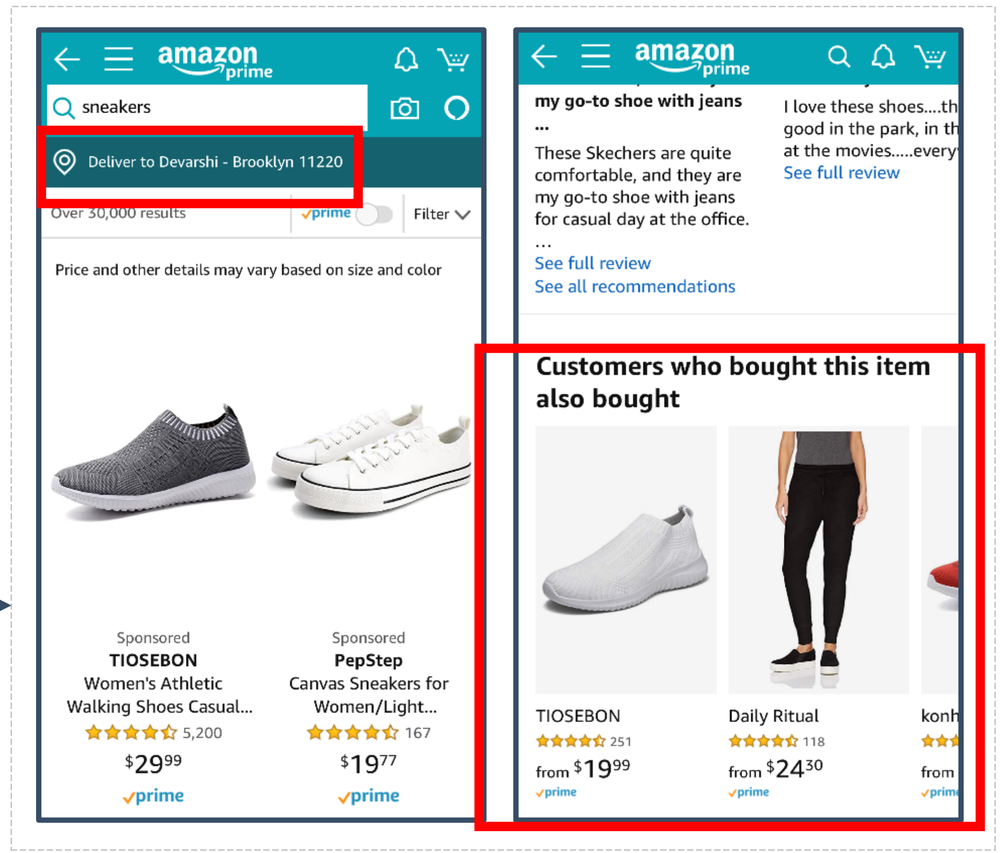
5 Implications

6 Business Insights

Amazon > 1.1 Product Recommendation





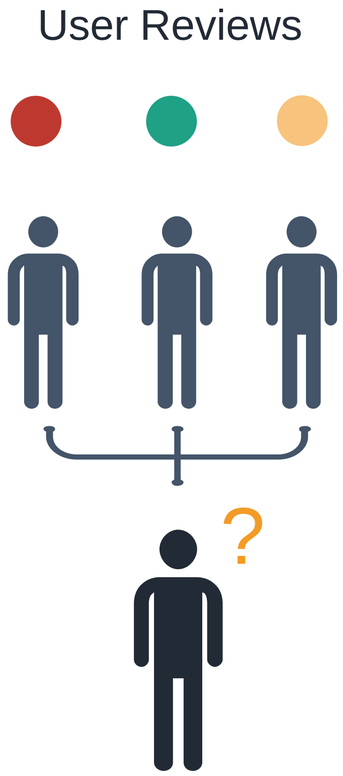


Above images are screenshots from two different Amazon user accounts providing similar recommendation

Amazon strives to be the pioneer of personalization to improve customer experience

TechCrunch

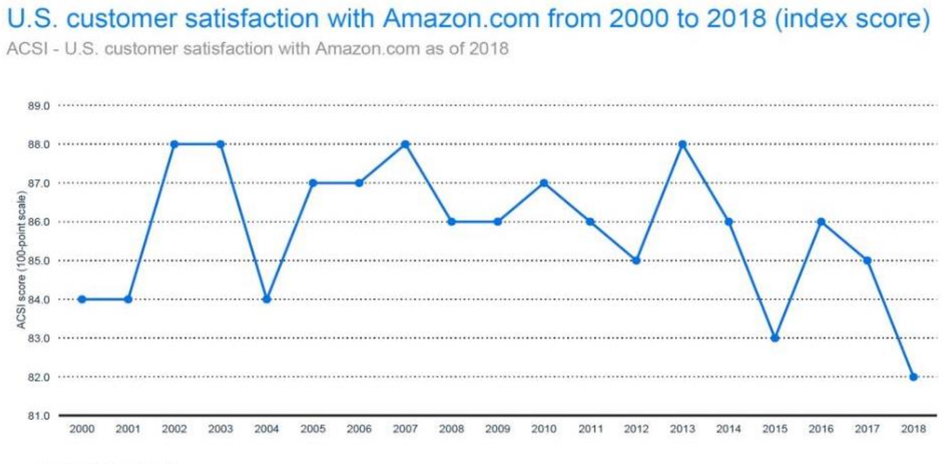
1.2 Help Improve personalization based on user-reviews



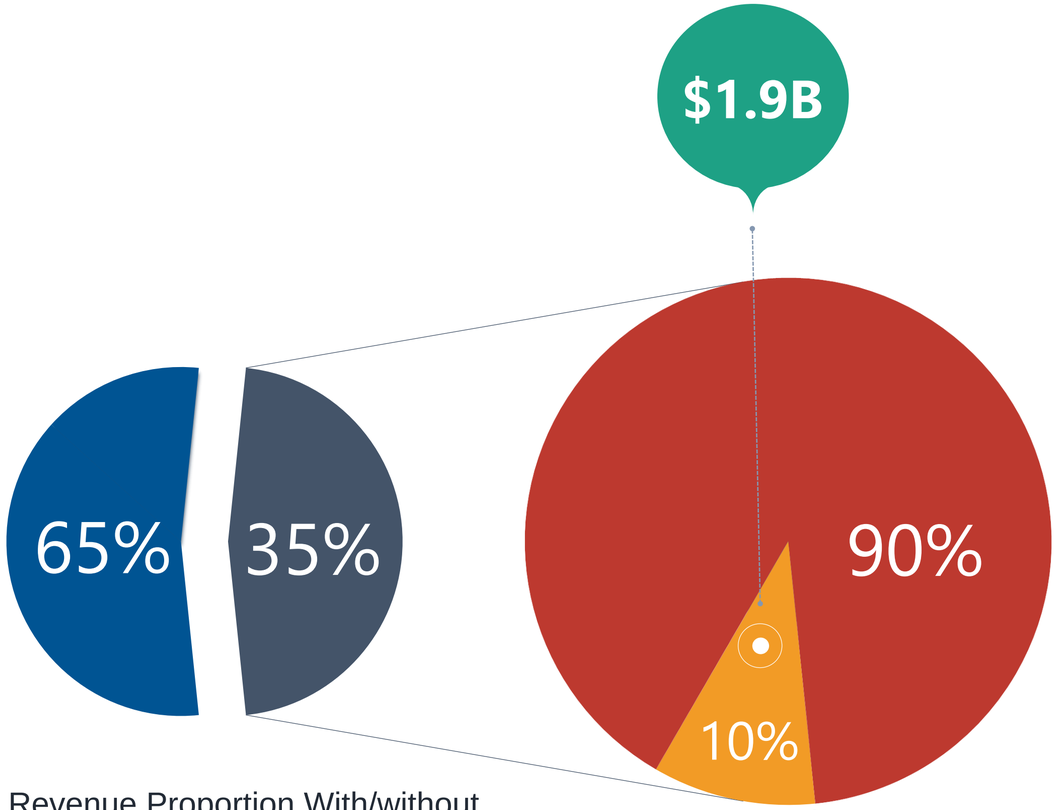
1.1 Problem

- Lack of personalization in recommendation systems

1.2 Market Study



57% of the purchasing decision driven by product reviews.



Revenue Proportion With/without Recommendation System

2. Extracting and understanding dataset

Source: Amazon Fine foods
Stanford.edu, J. Leskovec

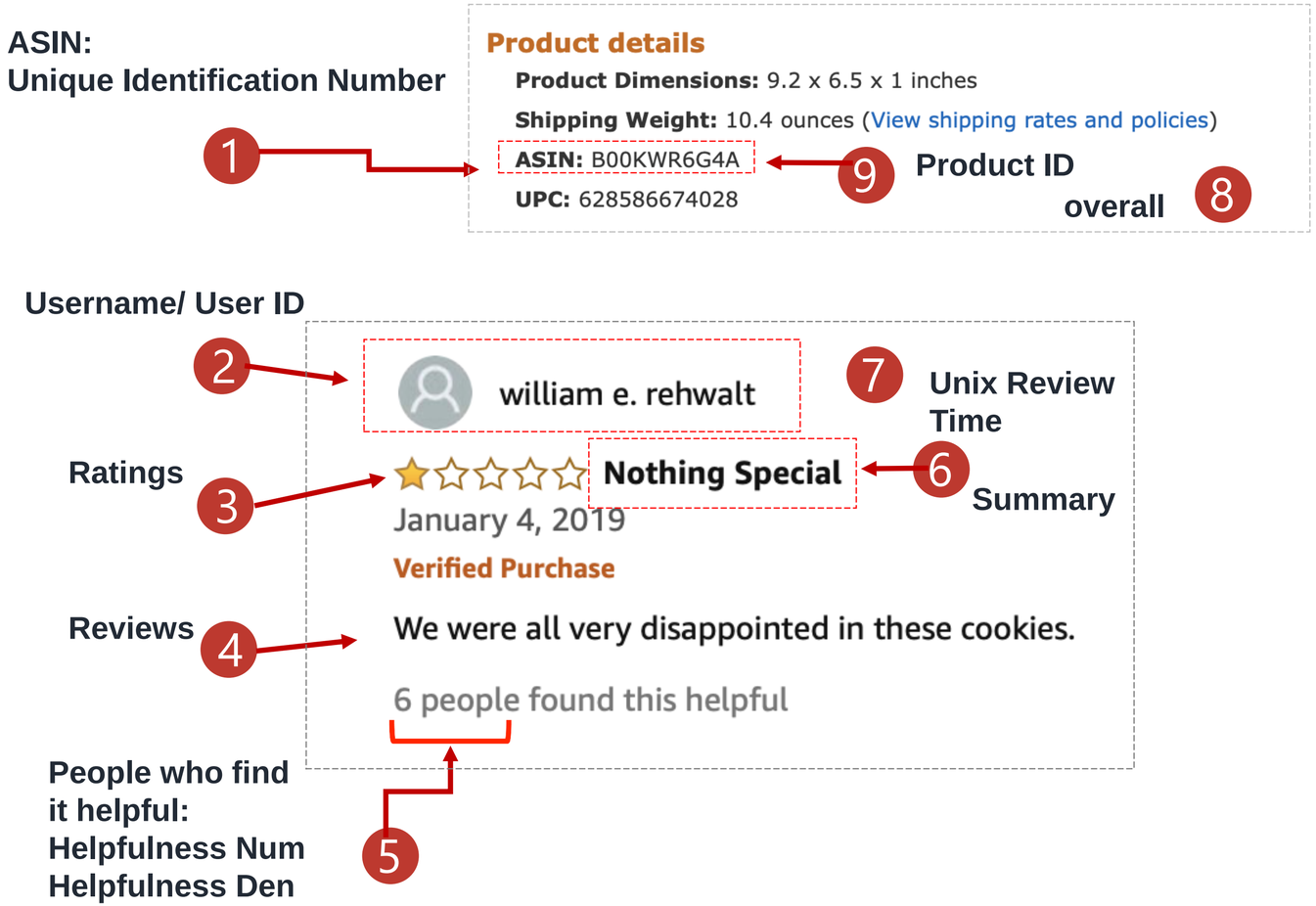
GZ Zip Format

Json file

csv file

9
Columns

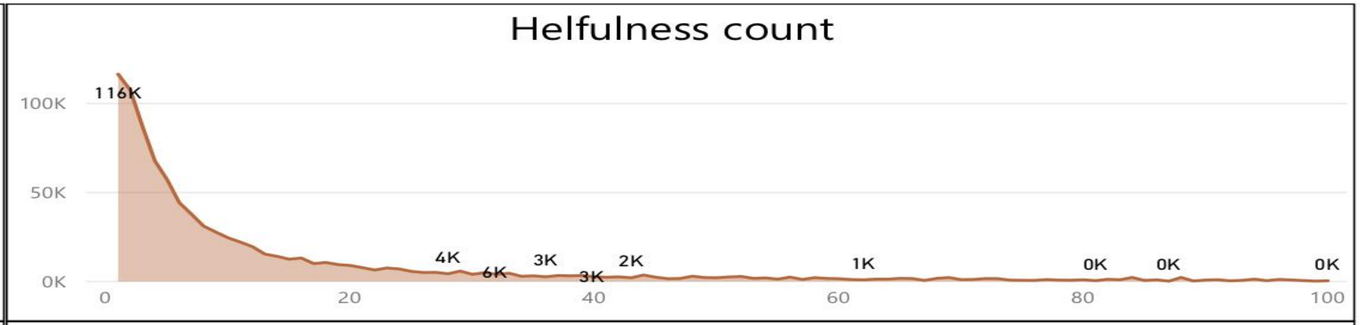
568,454
Instances

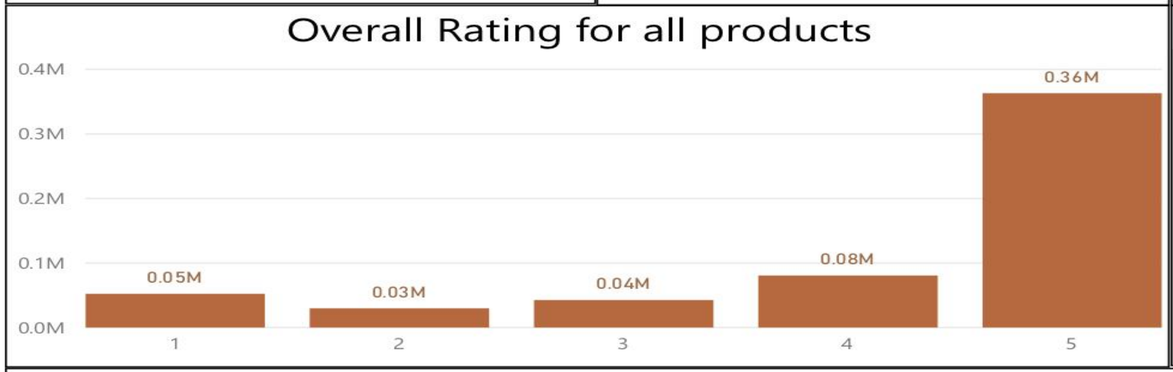


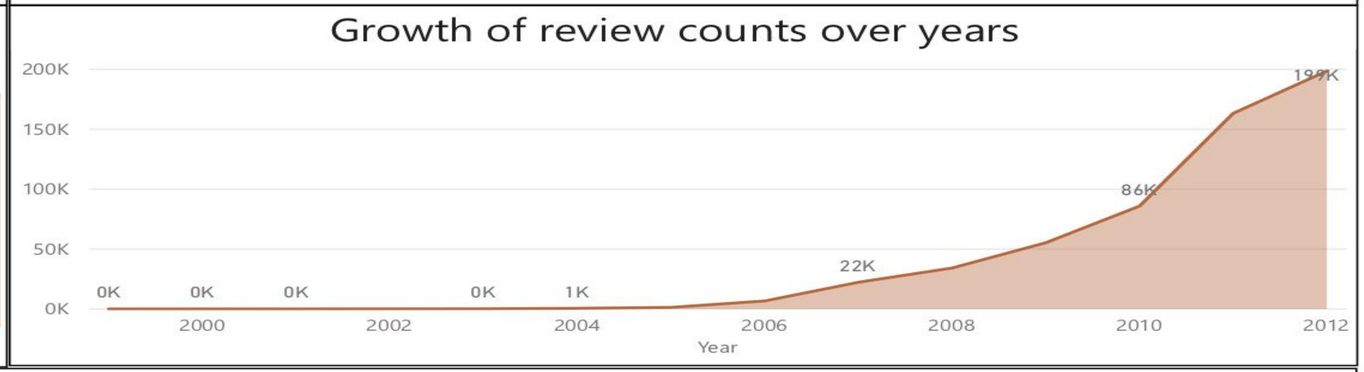
3.1 Exploratory Analysis

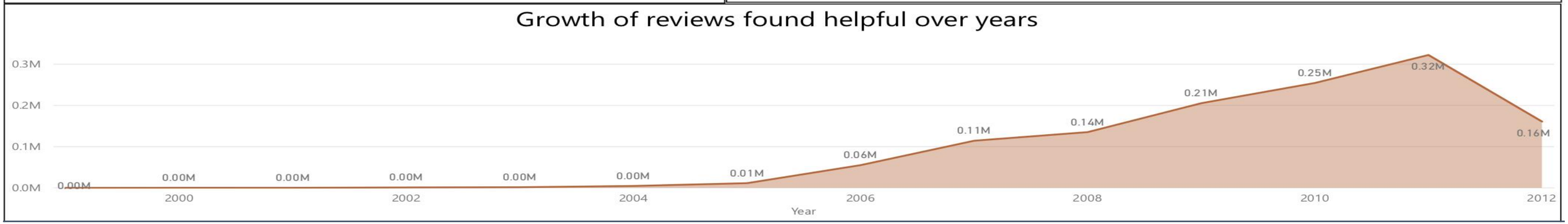
74.26K
Count of Products in the dataset

256.06K
Count of Users in the dataset









3.2 Data Preprocessing

- ✓ 1. Removed Duplicates
- ✓ 2. Data Filtering
- ✓ 3. Removed Stop Words, White Space
- ✓ 4. Removed Punctuations, Numbers
- ✓ 5. Normalize: Lower Case, Stemming
- ✓ 6. Word Count
- ✓ 7. Processed Dataset

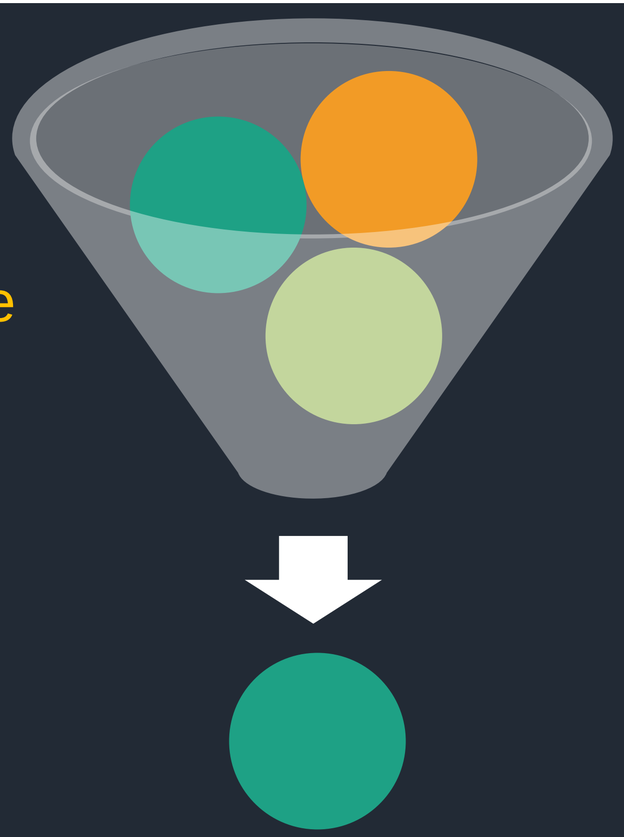
Asin	Reviewer ID	Reviews
B001E4KF	A3SGXH	I had bought …
F101E4EF	A5PGDF	Nice Product
B001E4KF	A3SGXH	I had bought …

Asin	Reviewer ID	Reviews
B001E4KF	A3SGXH	I had bought …
F101E4EF	A5PGDF	Nice Product

Processed Dataset

3.2 Data Preprocessing

- 1. Removed Duplicates
- 2. Data Filtering
- 3. Removed Stop Words, White Space
- 4. Removed Punctuations, Numbers
- 5. Normalize: Lower Case, Stemming
- 6. Word Count
- 7. Processed Dataset



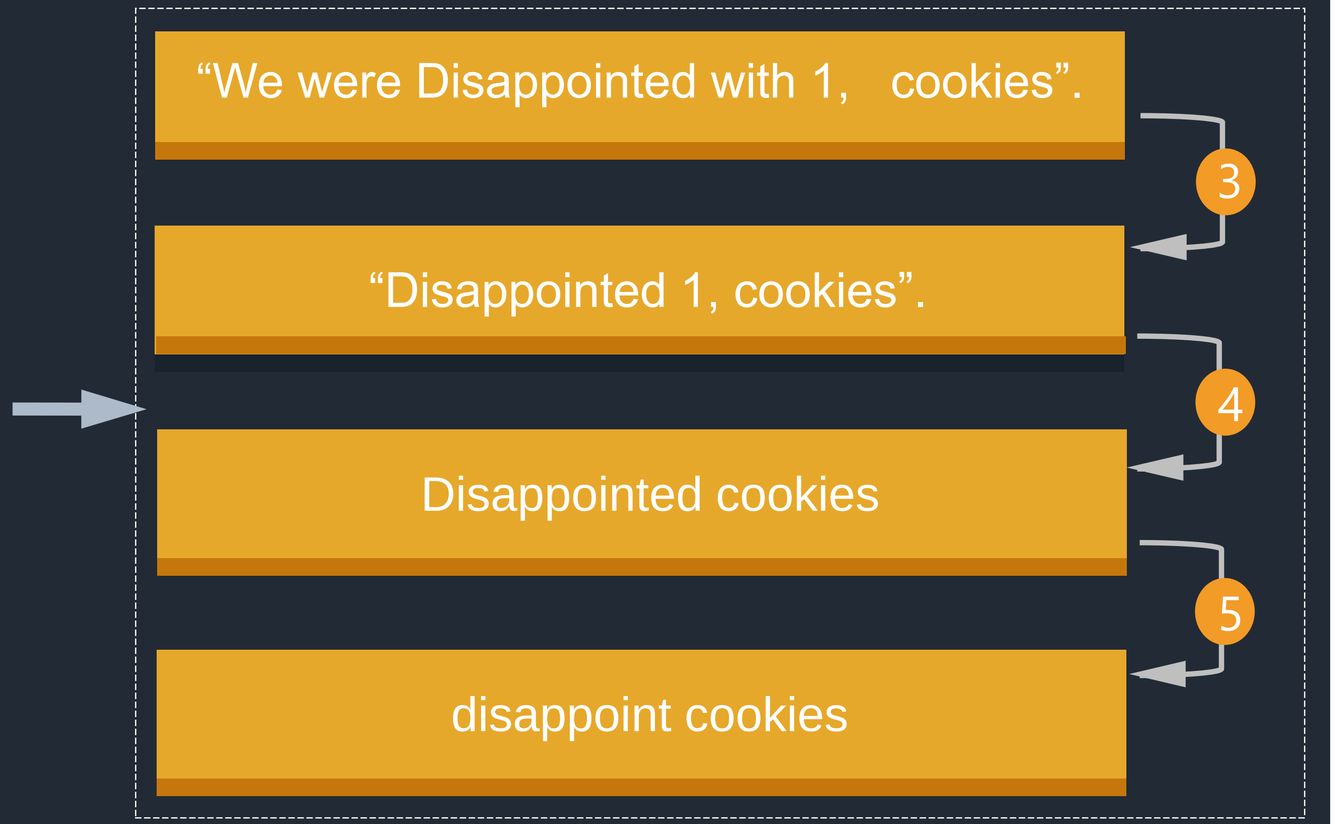
Filtered by Product Reviews

Processed Dataset

3.2 Data Preprocessing

#Amazon Review

- 1. Removed Duplicates
- 2. Data Filtering
- 3. Removed Stop Words, White Space
- 4. Removed Punctuations, Numbers
- 5. Normalize: Lower Case, Stemming
- 6. Word Count
- 7. Processed Dataset



3.2 Data Preprocessing

- 1. Removed Duplicates
- 2. Data Filtering
- 3. Removed Stop Words, White Space
- 4. Removed Punctuations, Numbers
- 5. Normalize: Lower Case, Stemming
- 6. Word Count
- 7. Processed Dataset

Asin	Reviews
B001E4KF	Nice Product
F101E4EF	Awesome features
B001E4KF	Best Product

Asin	Reviews
B001E4KF	Nice Product Best Product
F101E4EF	Awesome features

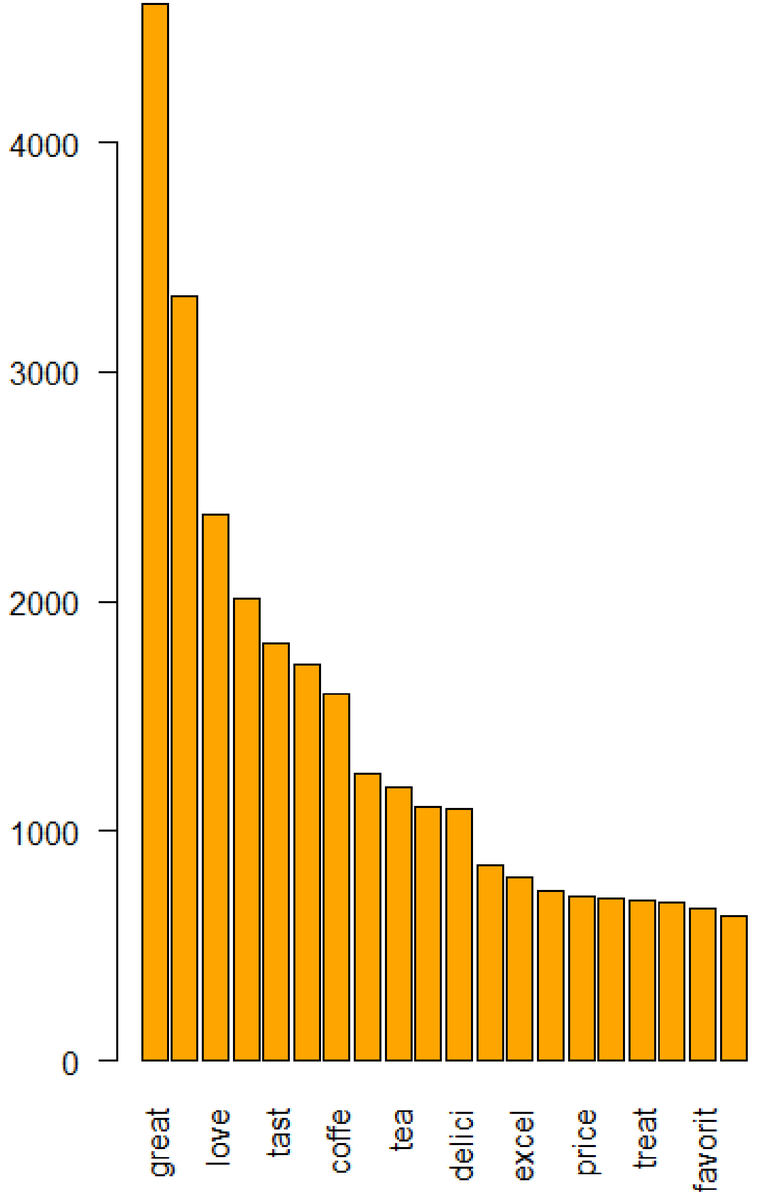
Processed Dataset

	Awesome	Best	Features	Nice	Product
B1001E4KF	0	1	0	1	2
F101E4EF	1	0	1	0	0

Top 300 High Frequency Words

4.1 Sentiment Analysis

Text Analysis

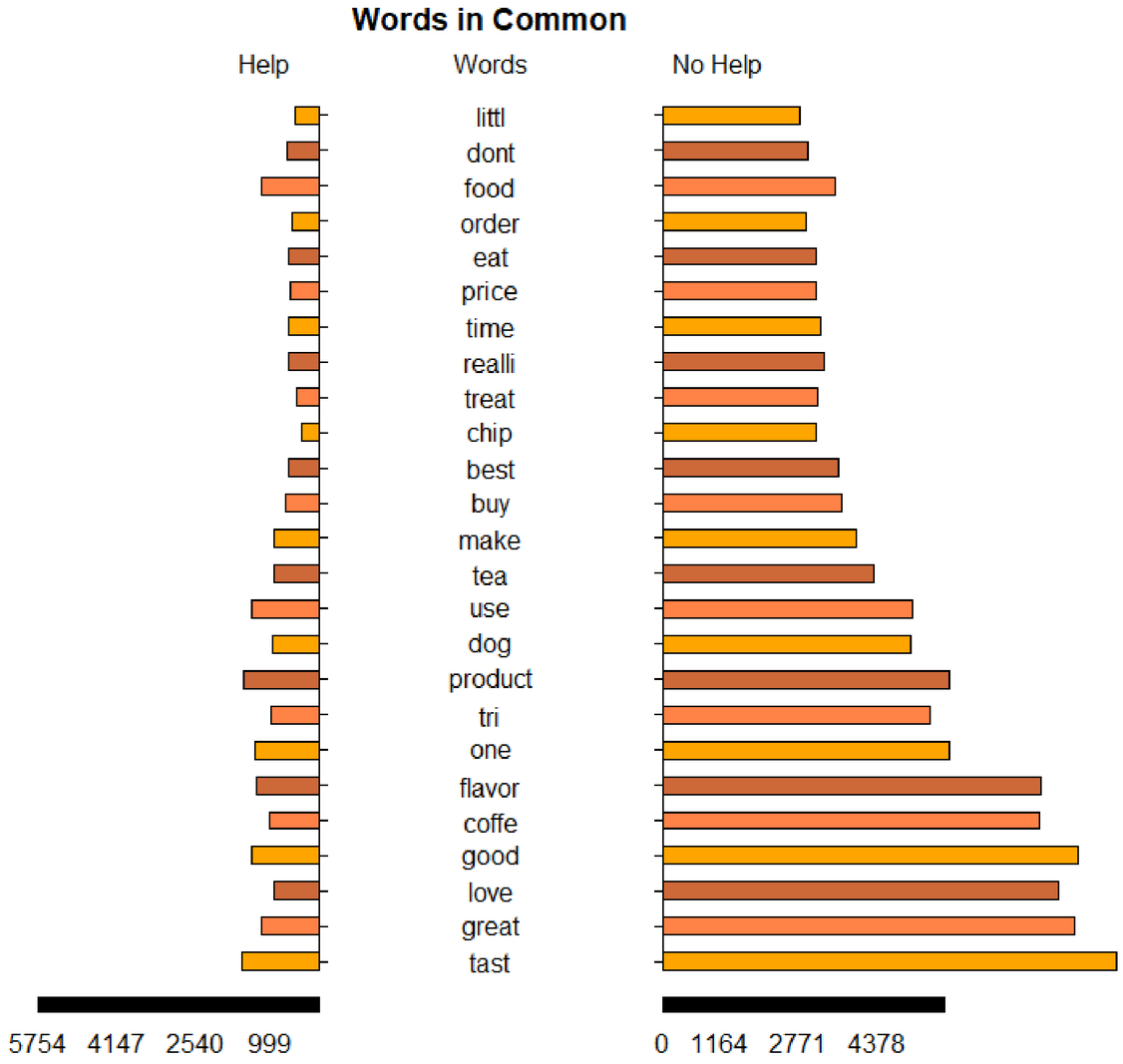


4.1 Sentiment Analysis

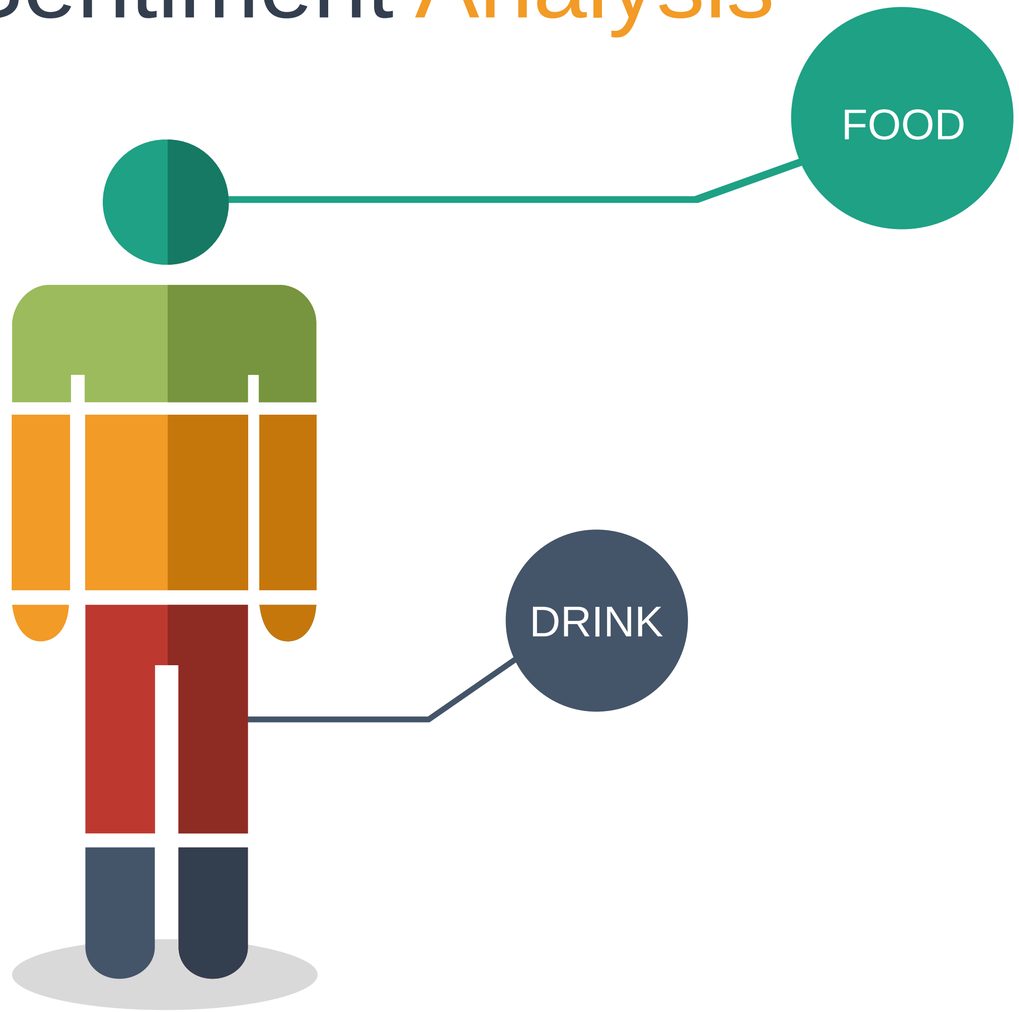
Helpfulness

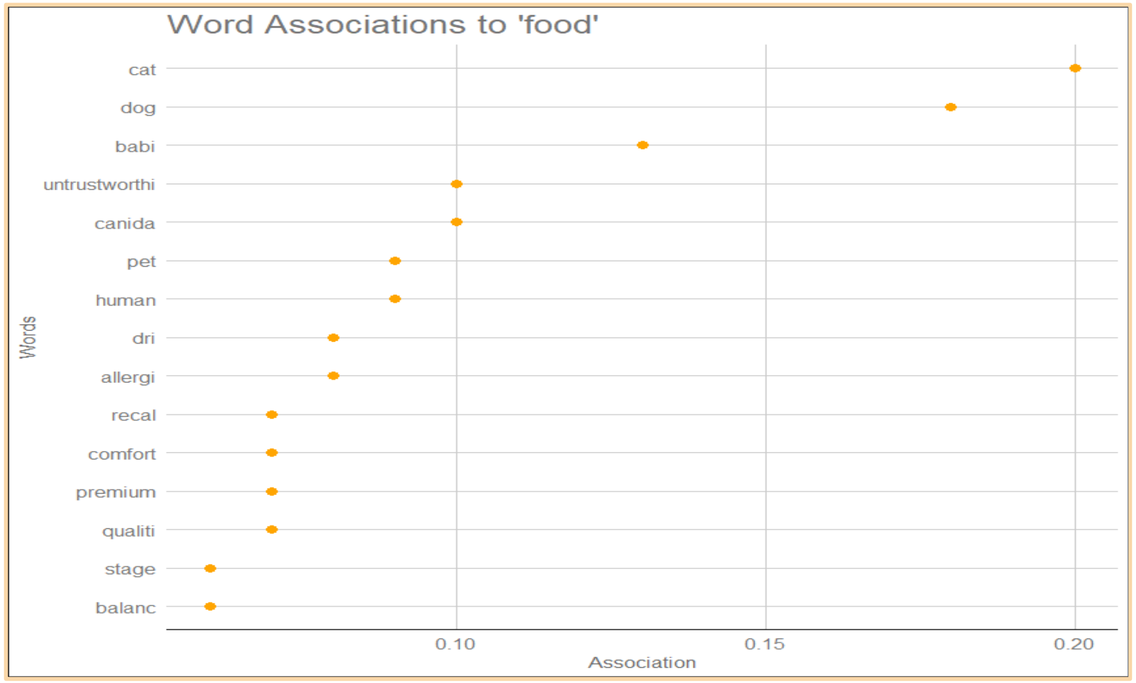
Help

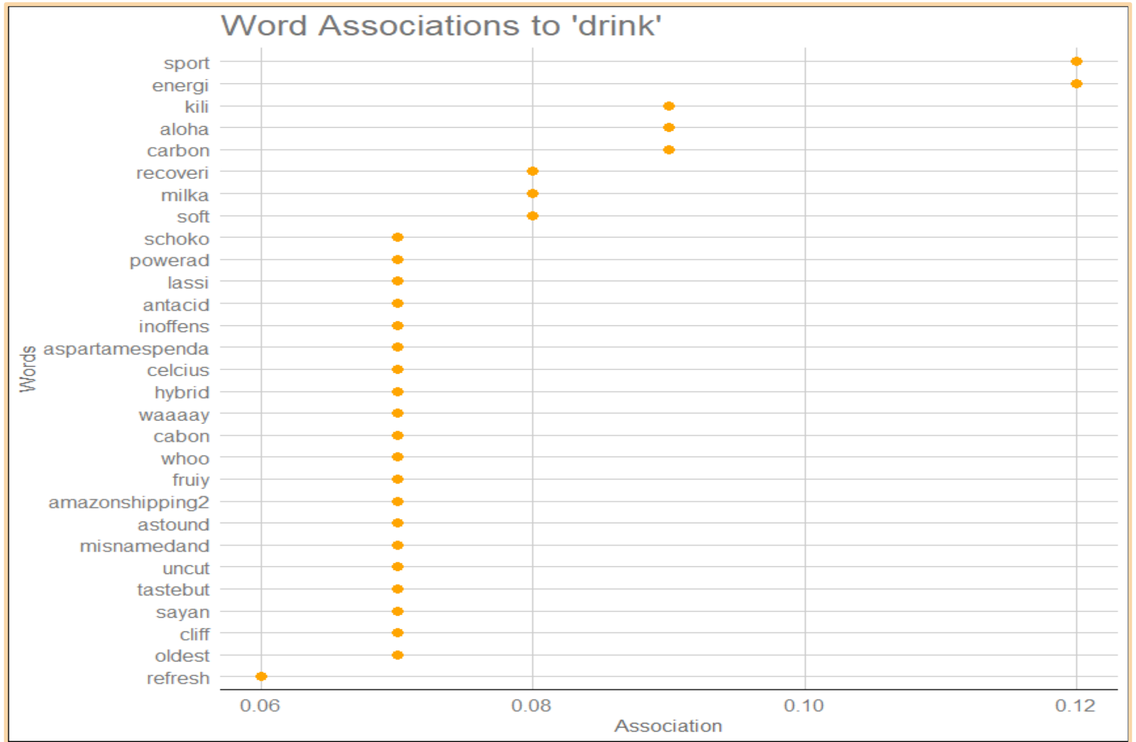
No Help



4.1 Sentiment Analysis



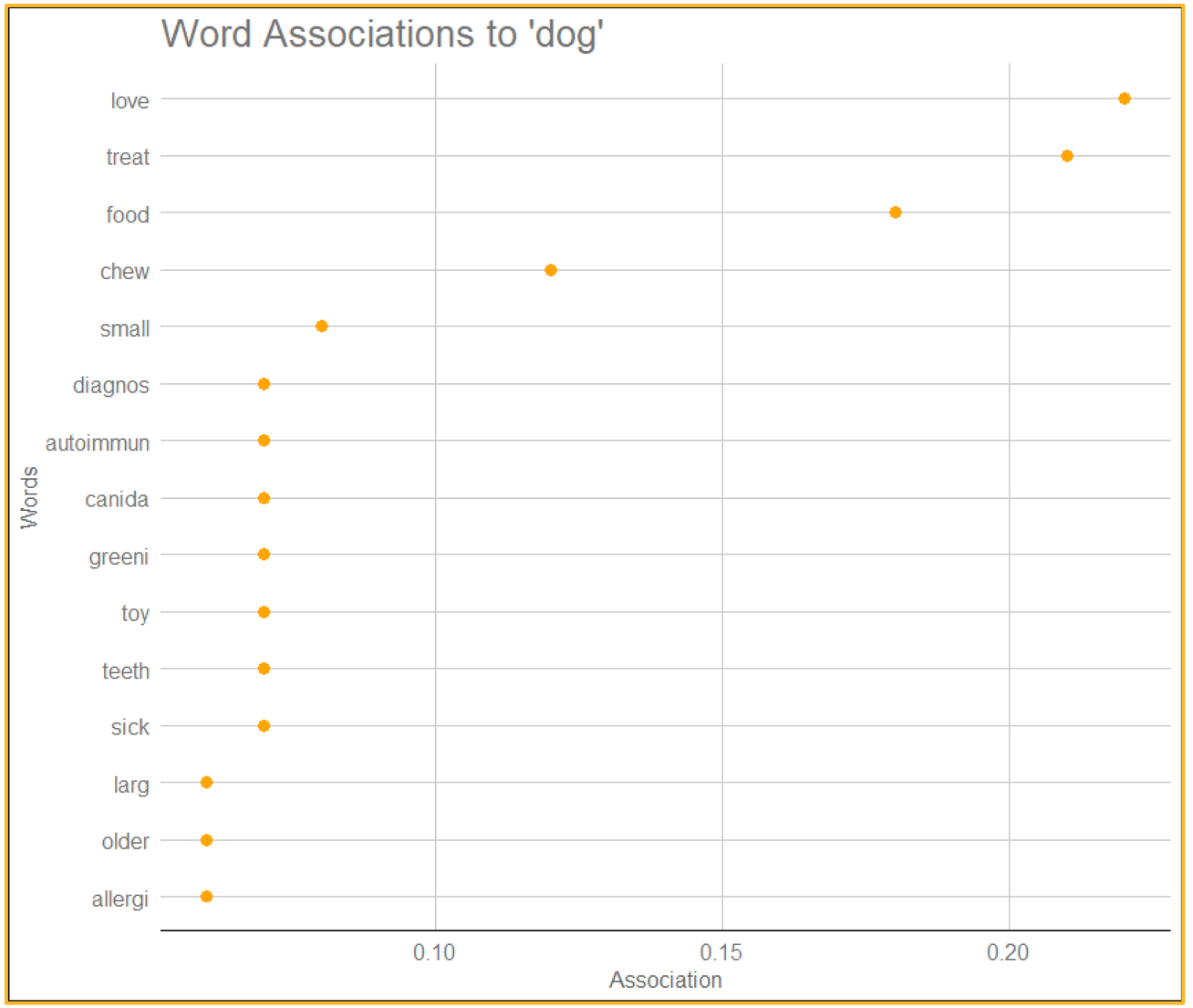


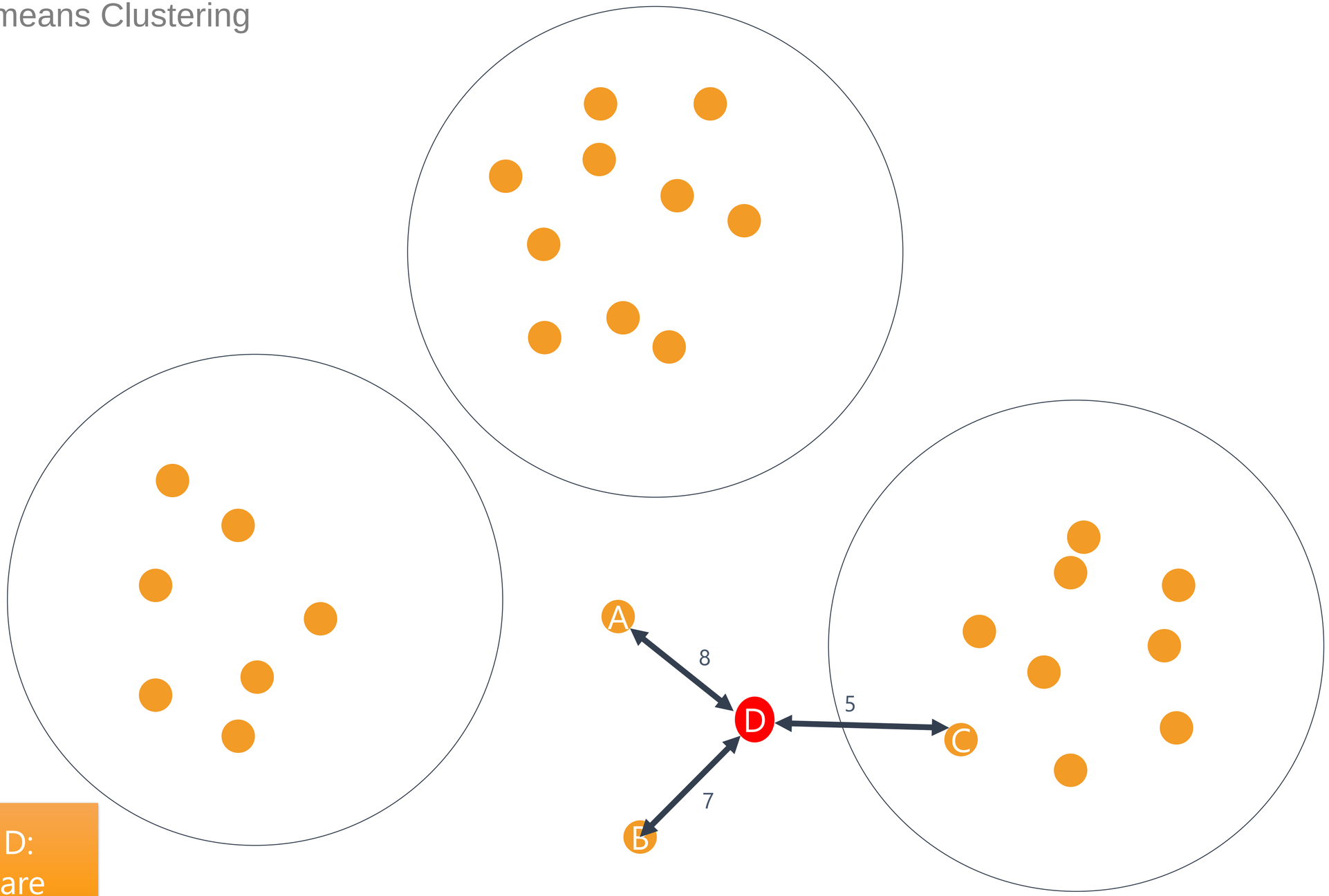


Word Association: Food &Drink

4.1 Sentiment Analysis

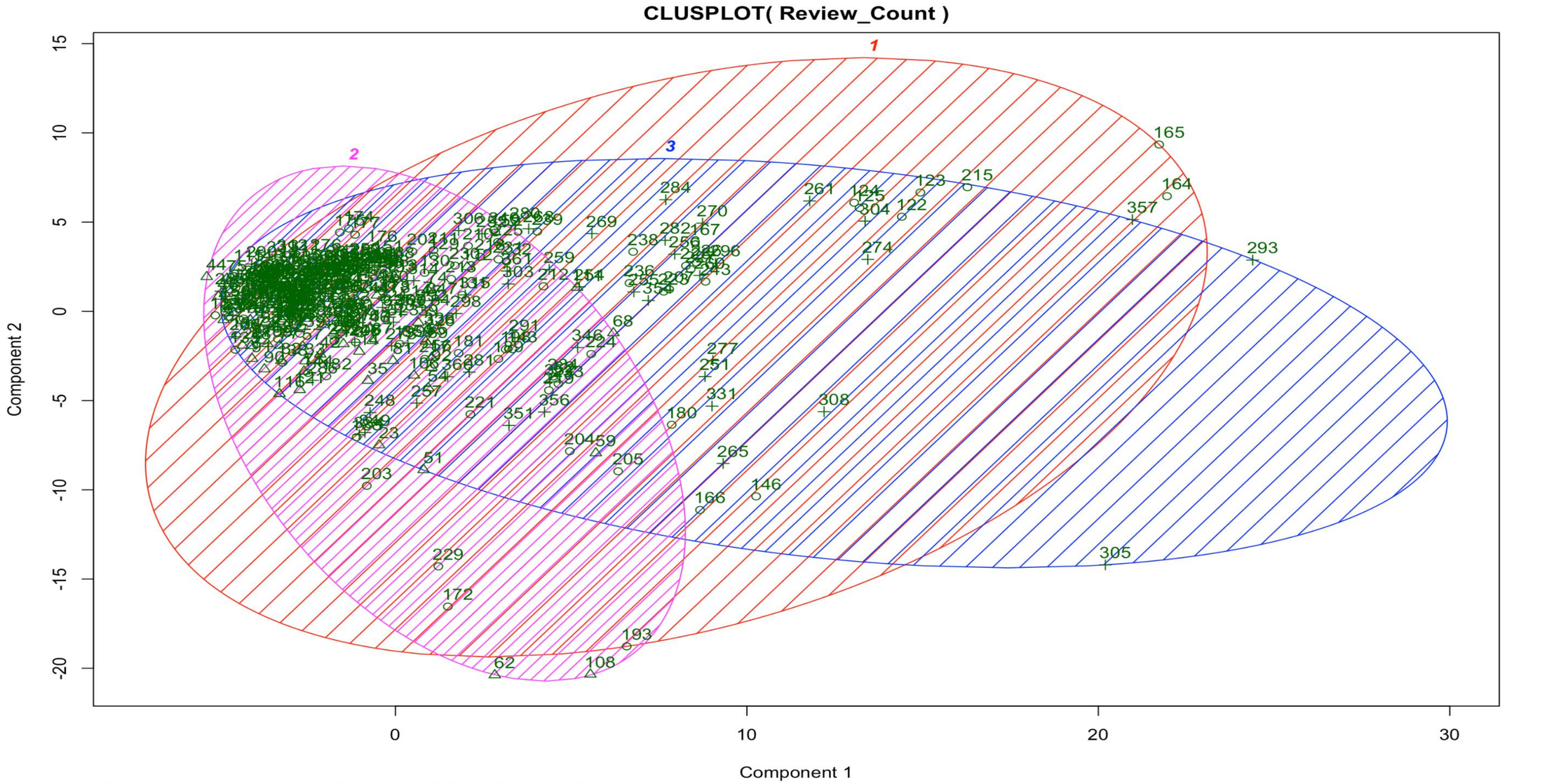
Word Association: Dog





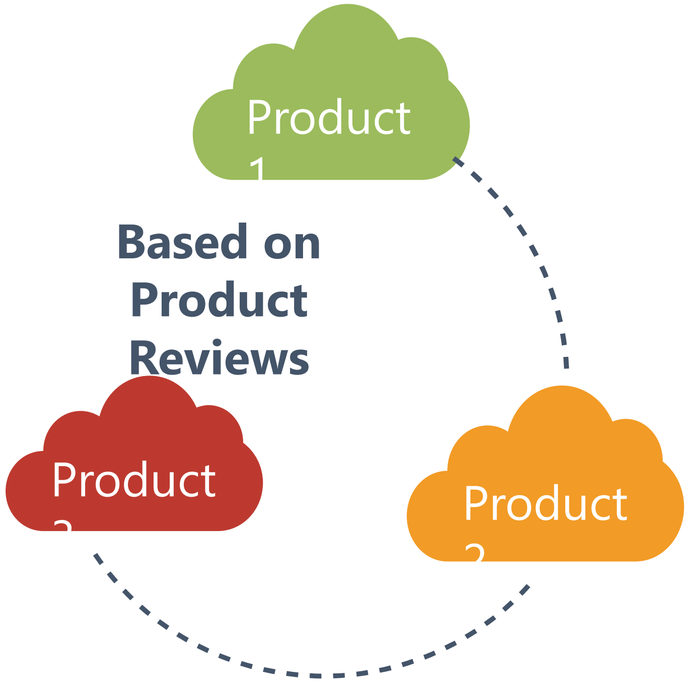
For Product D:
Suggestions are
Product C & B

4.2 K-means Clustering



5. Results

```
Based on product reviews, for  B007TGDXMU  average rating is  4.517006802721088
The first similar product is  B001E50THY  average rating is  4.044642857142857
The second similar product is  B007TGDXMK  average rating is  4.143540669856459
-----------------------------------------------------------
Based on product reviews, for  B007TGDXNO  average rating is  4.3478260869565215
The first similar product is  B007TGDXMK  average rating is  4.143540669856459
The second similar product is  B002AQ0OL2  average rating is  4.058035714285714
-----------------------------------------------------------
Based on product reviews, for  B007TJGY5K  average rating is  4.119496855345912
The first similar product is  B005ZBZLPI  average rating is  4.0
The second similar product is  B005HUVI40  average rating is  3.9917355371900825
-----------------------------------------------------------
Based on product reviews, for  B007TJGZ0Y  average rating is  4.384615384615385
The first similar product is  B001CHJ01A  average rating is  4.364485981308412
The second similar product is  B002QGK2V8  average rating is  3.4484848484848483
-----------------------------------------------------------
Based on product reviews, for  B007TJGZ54  average rating is  4.243801652892562
The first similar product is  B006N3I29E  average rating is  4.245689655172414
The second similar product is  B002AQ0OL2  average rating is  4.058035714285714
```



6. Business Usage

PERSONALIZED EXPERIENCE

CROSS SELLING COMPLIMENTARY PRODUCTS

ENGAGE SHOPPERS

Reach | Act | Convert | Engage

Amazon's general RACE strategy

6. Future Scope : User Profiling by Tag words

Tag Words from product reviews

Tag Words from the user reviews

Questions ?

Thank You

#Amazon Prime #Toys #Black Friday #Games #Movies #Personalized
#Ecommerce #Recommendation #United States #Year 2020
#AWS #Clothing #Office Products #Digital Music
#Gadgets #No Words #Brie #Customer Centric #Champions #Digital