



'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



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 Results

6 Business Use

7 Future Scope

1.1 Product Recommendation

amazon prime

sneakers

Deliver to Kanak - Brooklyn 11209

Over 30,000 results

Filter

Sponsored by KEEZMZ

KEEZMS Mens Womens Quick Drying Aqua Water Shoes

Shop now >

Price and other details may vary based on size and color

Sponsored TIOSEBON Women's Athletic Walking Shoes Casual Mesh-...

from \$19⁹⁹

Sponsored PepStep Canvas Sneakers for Women/Light...

from \$24³⁰

amazon prime

my go-to shoe with jeans

These Skechers are quite comfortable, and they are my go-to shoe with jeans for casual day at the office.

See full review

See all recommendations

Customers who bought this item also bought

TIOSEBON

from \$19⁹⁹

Daily Ritual

from \$24³⁰

konh

from

amazon prime

sneakers

Deliver to Devarshi - Brooklyn 11220

Over 30,000 results

Filter

Price and other details may vary based on size and color

Sponsored TIOSEBON Women's Athletic Walking Shoes Casual...

from \$29⁹⁹

Sponsored PepStep Canvas Sneakers for Women/Light...

from \$19⁷⁷

amazon prime

my go-to shoe with jeans

These Skechers are quite comfortable, and they are my go-to shoe with jeans for casual day at the office.

See full review

See all recommendations

Customers who bought this item also bought

TIOSEBON

from \$19⁹⁹

Daily Ritual

from \$24³⁰

konh

from

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

3

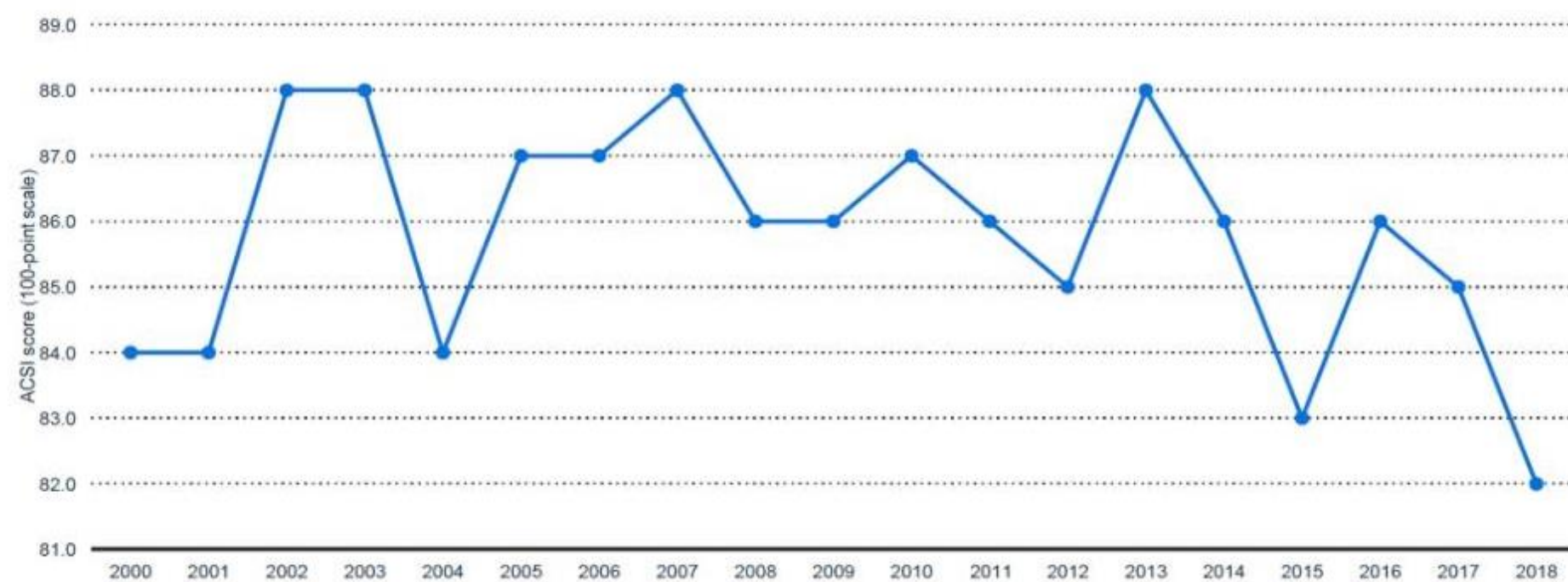
1.2 Help Improve personalization based on user-reviews

1.1 Problem

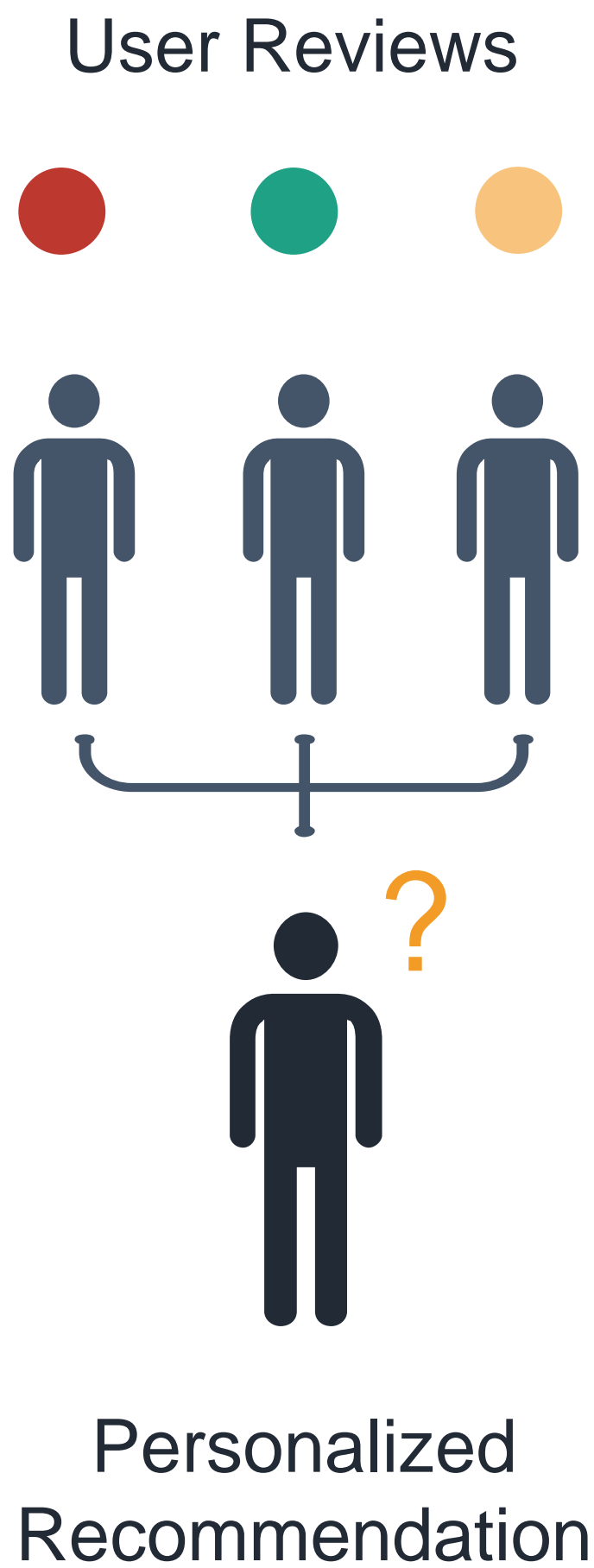
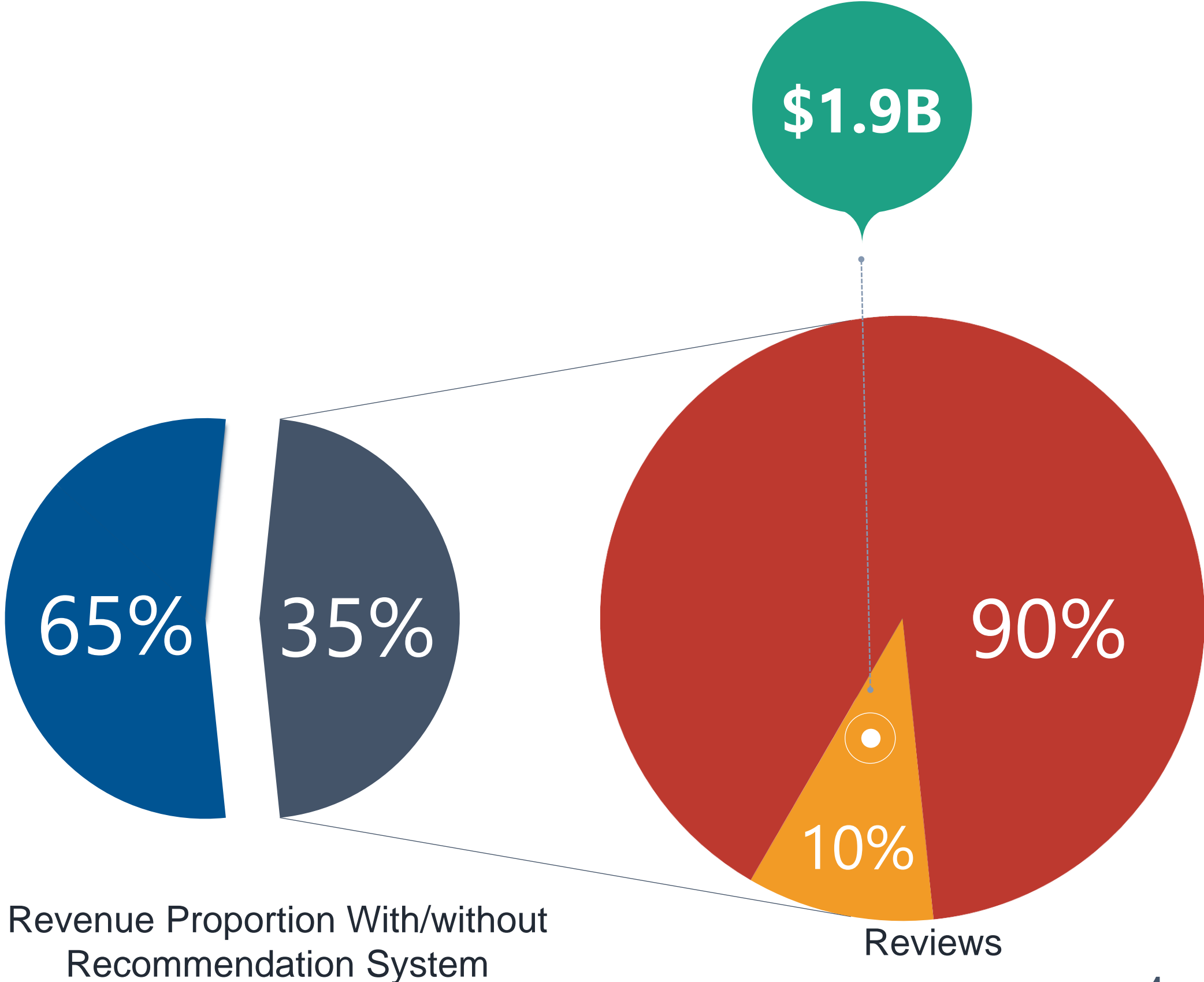
- Lack of personalization in recommendation systems

1.2 Market Study

U.S. customer satisfaction with Amazon.com from 2000 to 2018 (index score)
ACSI - U.S. customer satisfaction with Amazon.com as of 2018



57% of the purchasing decision driven by product reviews.



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

ASIN:
Unique Identification Number

Product details
Product Dimensions: 9.2 x 6.5 x 1 inches
Shipping Weight: 10.4 ounces (View shipping rates and policies)
ASIN: B00KWR6G4A
UPC: 628586674028

Username/ User ID

Ratings

Reviews

People who find
it helpful:
Helpfulness Num
Helpfulness Den

1

2

3

4

5

6

9

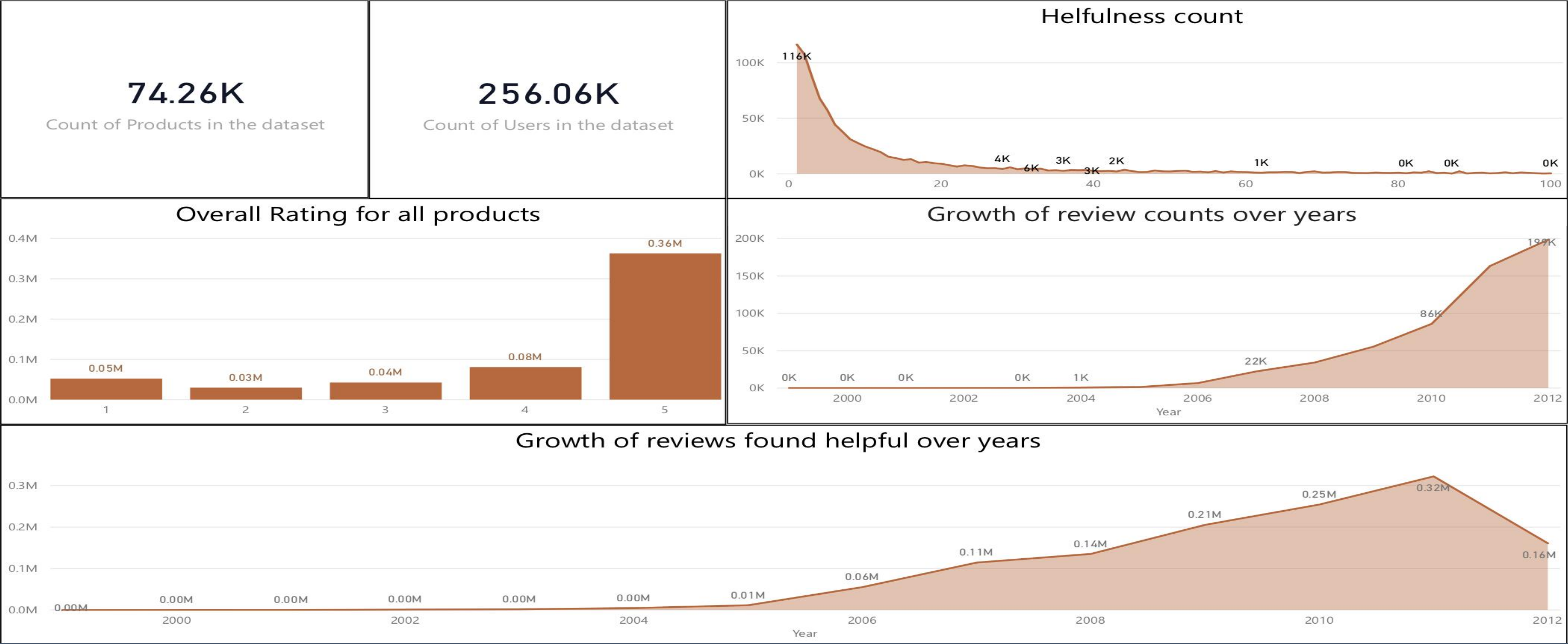
8

Unix Review
Time

7

Summary

3.1 Exploratory Analysis



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

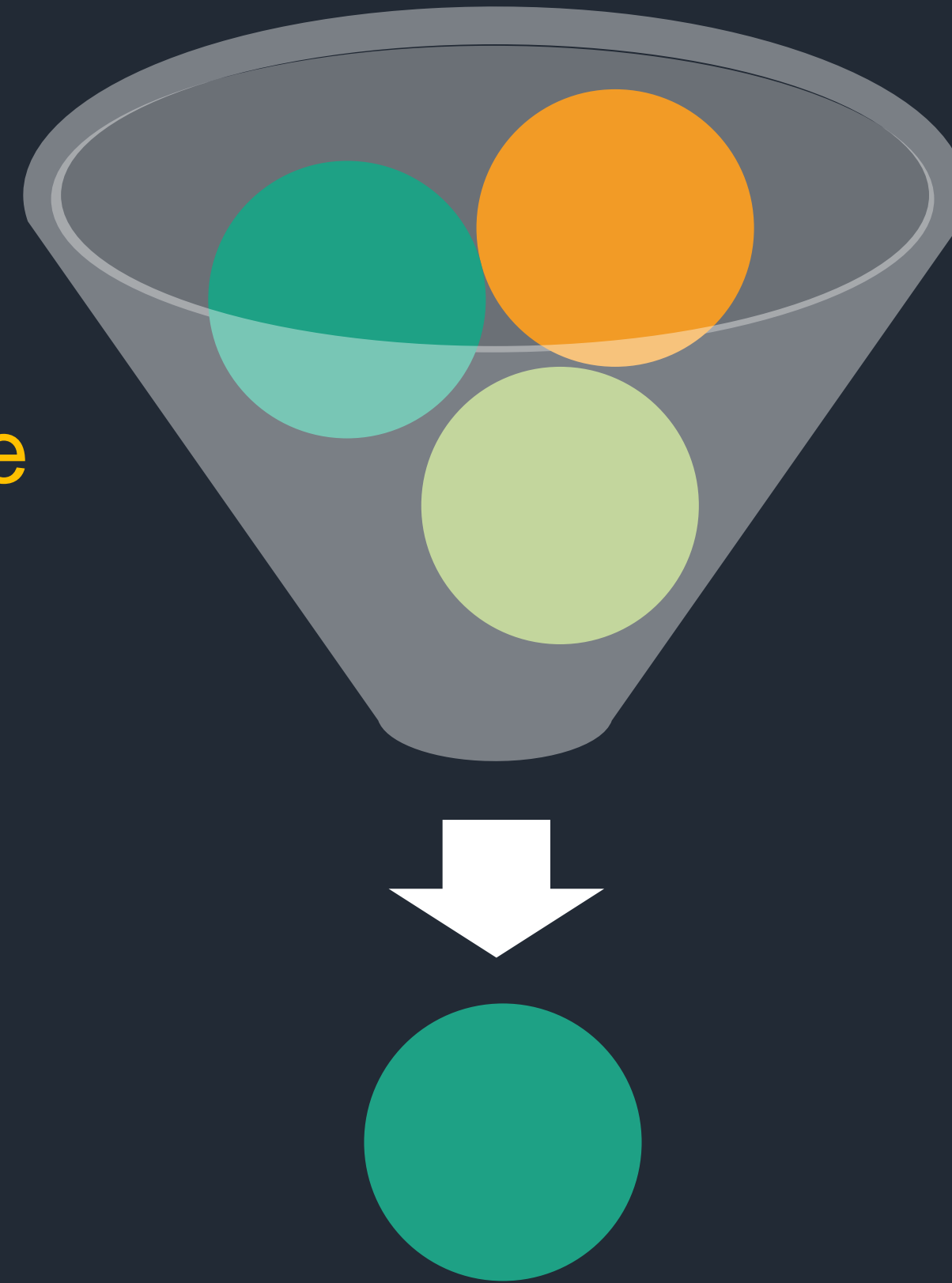
| Asin | Reviewer ID | Reviews |
|----------|-------------|------------------|
| B001E4KF | A3SGXH | I had bought ... |
| F101E4EF | A5PGDF | Nice Product |
| B001E4KF | A3SGXH | I had bought ... |



| Asin | Reviewer ID | Reviews |
|----------|-------------|------------------|
| B001E4KF | A3SGXH | I had bought ... |
| F101E4EF | A5PGDF | Nice Product |

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



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



"We were Disappointed with 1, cookies".

3

"Disappointed 1, cookies".

4

Disappointed cookies

5

disappoint cookies

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

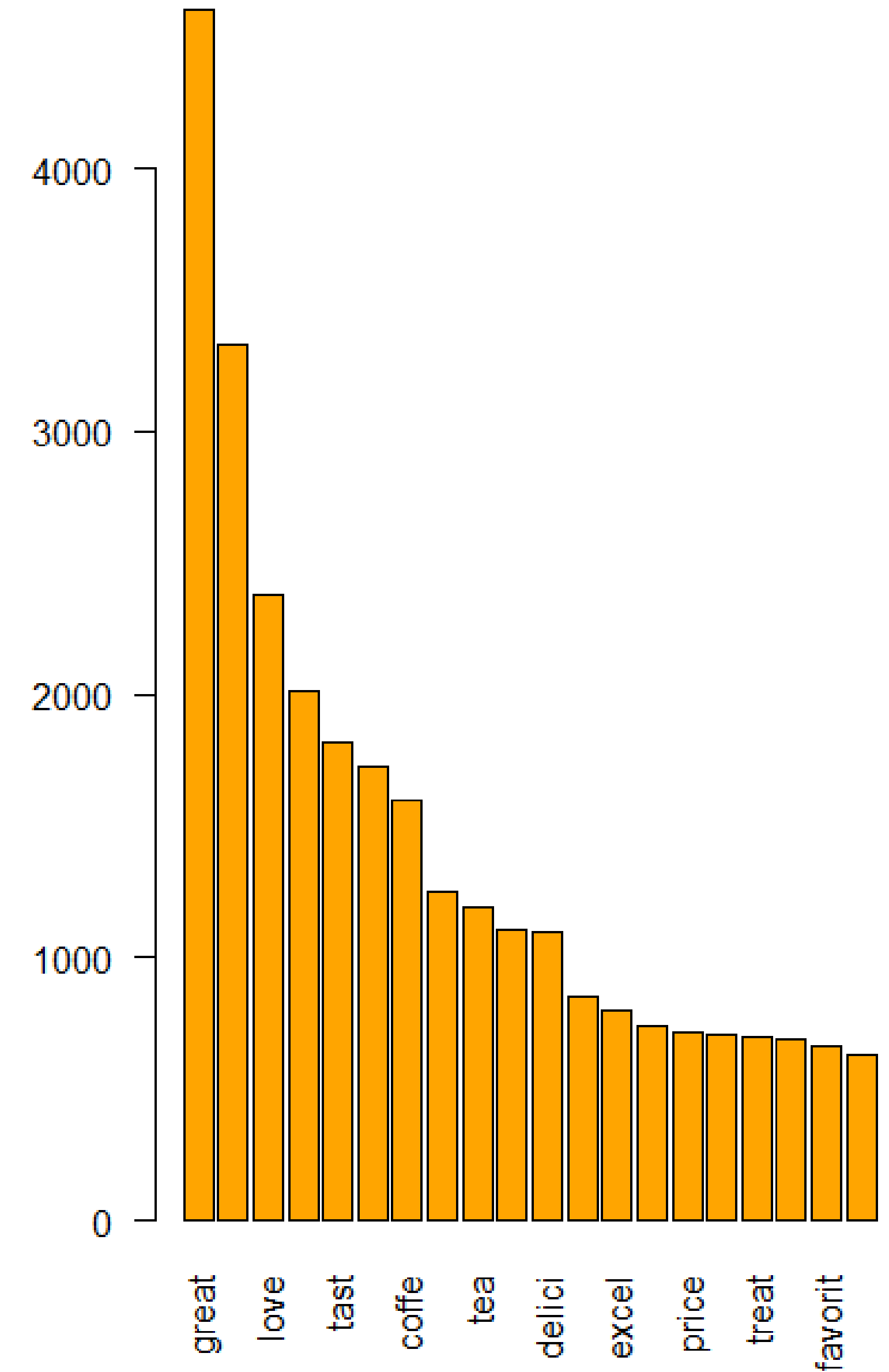
| | Awesome | Best | Features | Nice | Product | |
|-----------|---------|------|----------|------|---------|--|
| B1001E4KF | 0 | 1 | 0 | 1 | 2 | |
| F101E4EF | 1 | 0 | 1 | 0 | 0 | |

Top 300 High Frequency Words

4.1 Sentiment Analysis

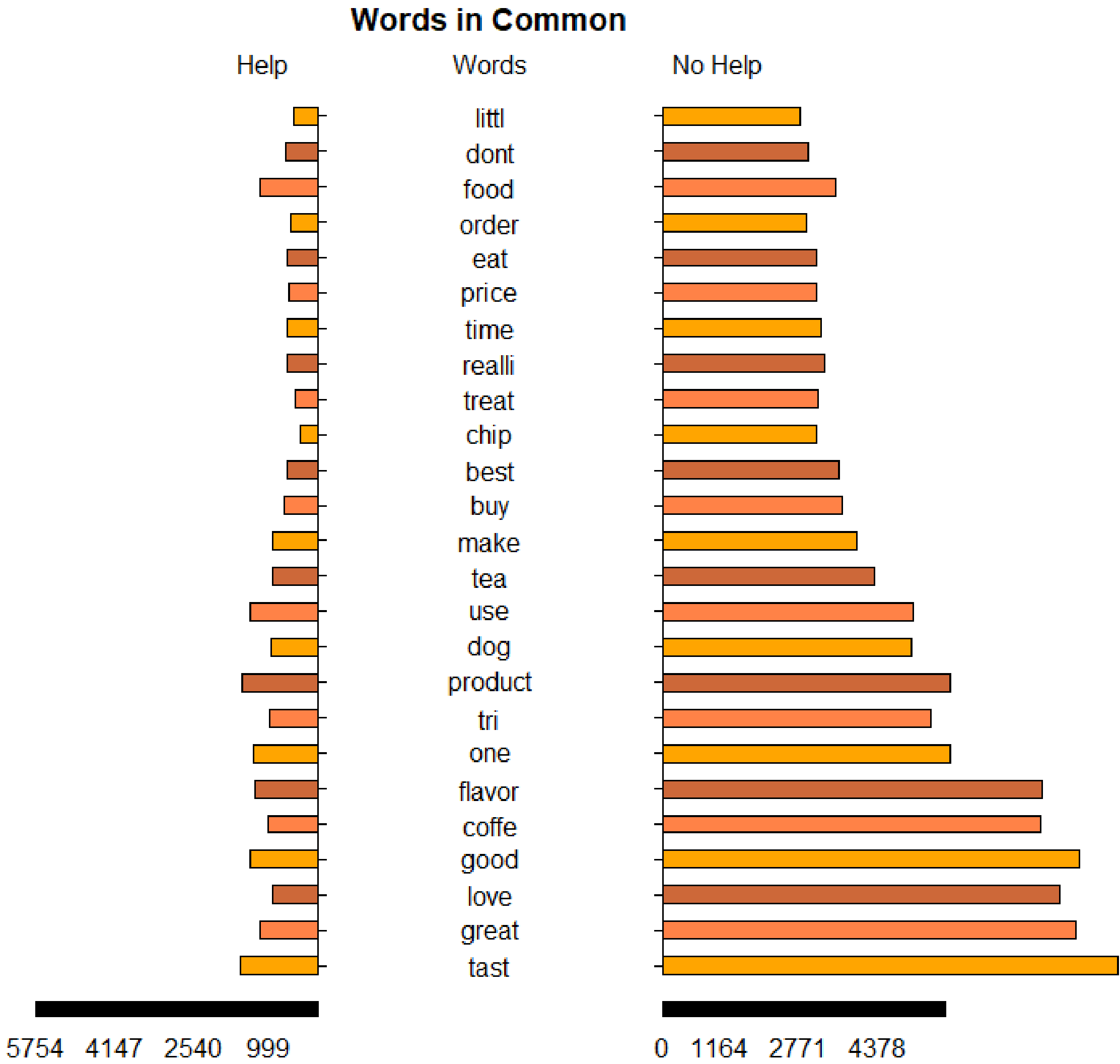
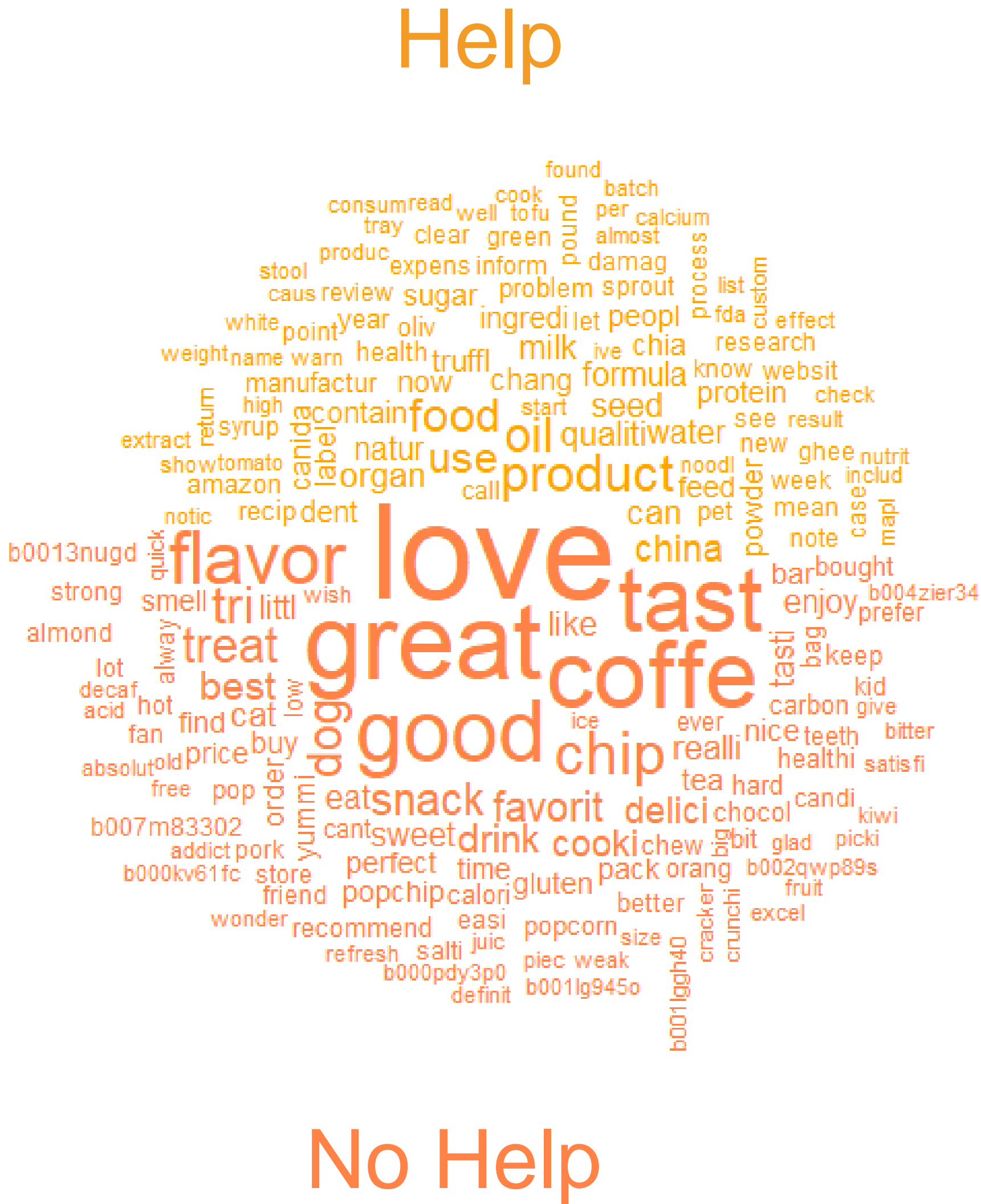


Text Analysis

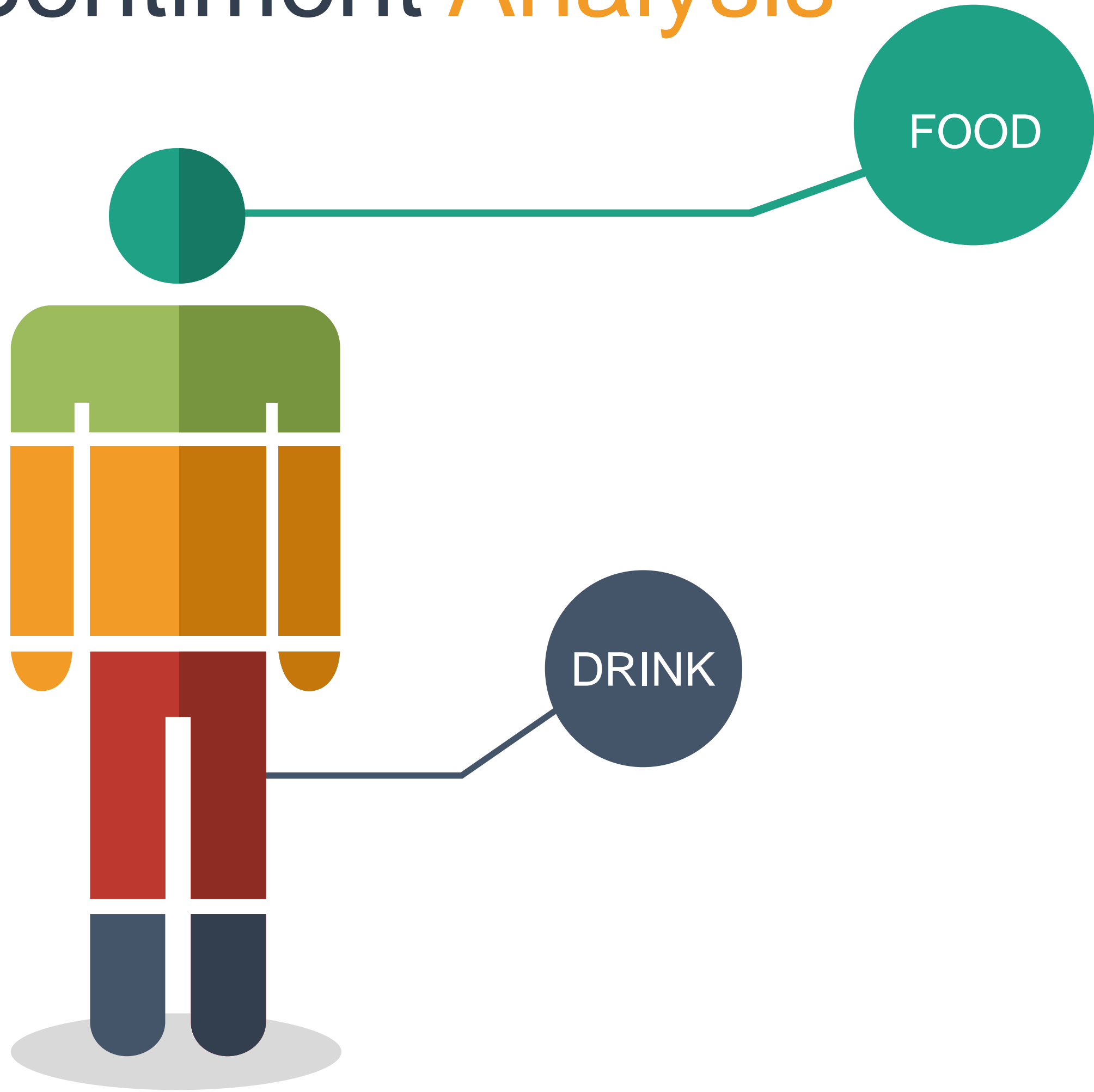


4.1 Sentiment Analysis

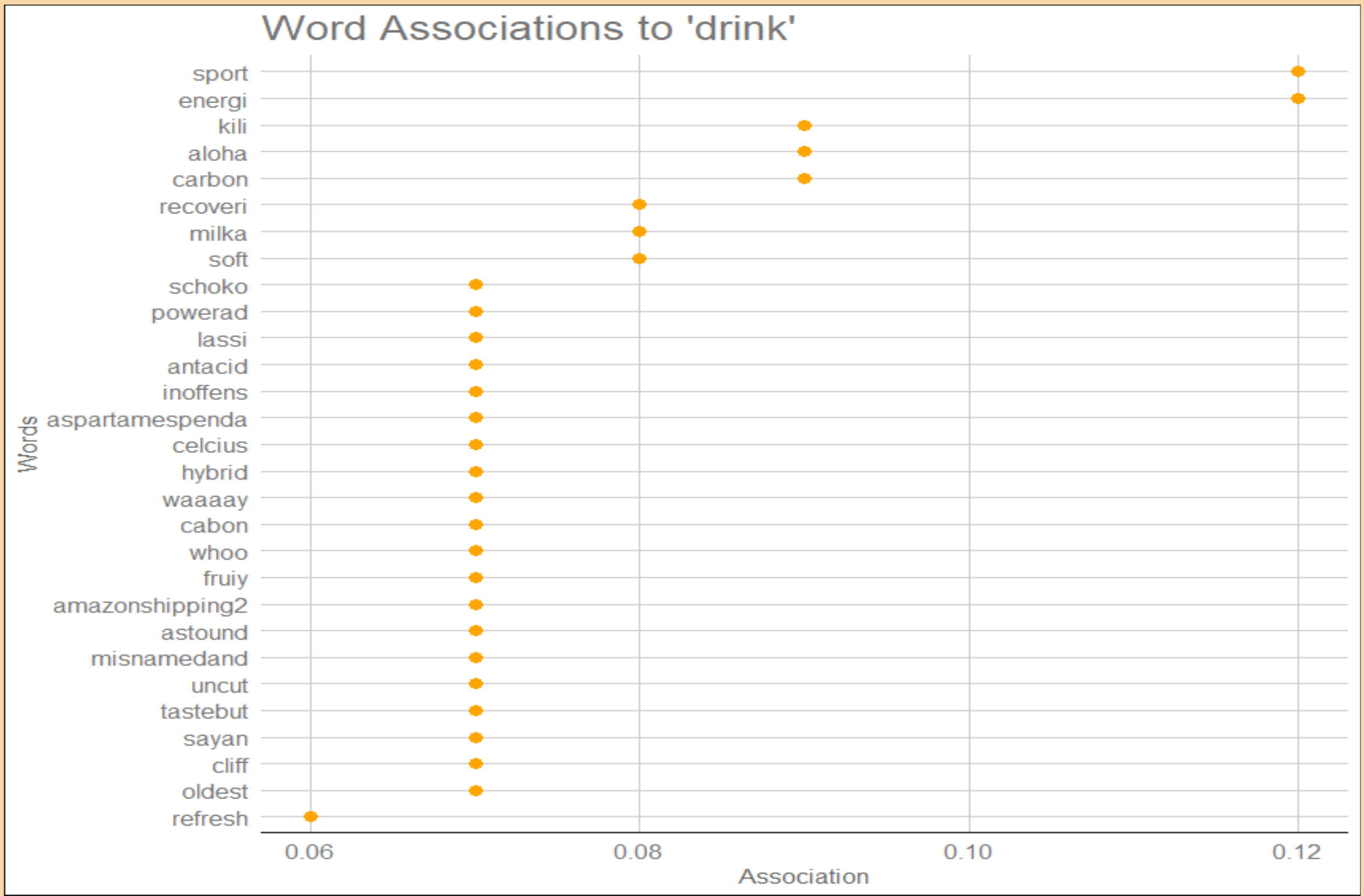
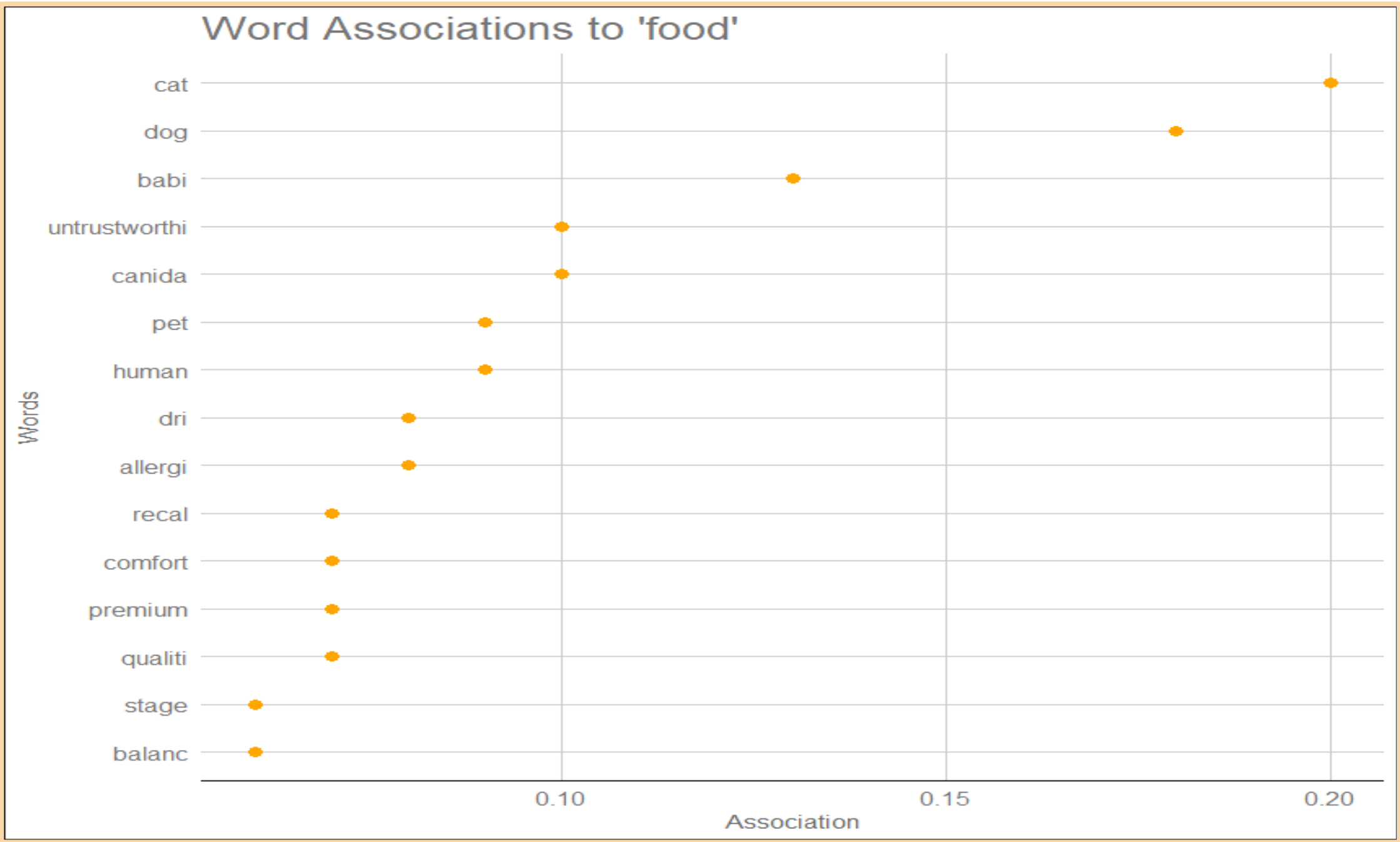
Helpfulness



4.1 Sentiment Analysis

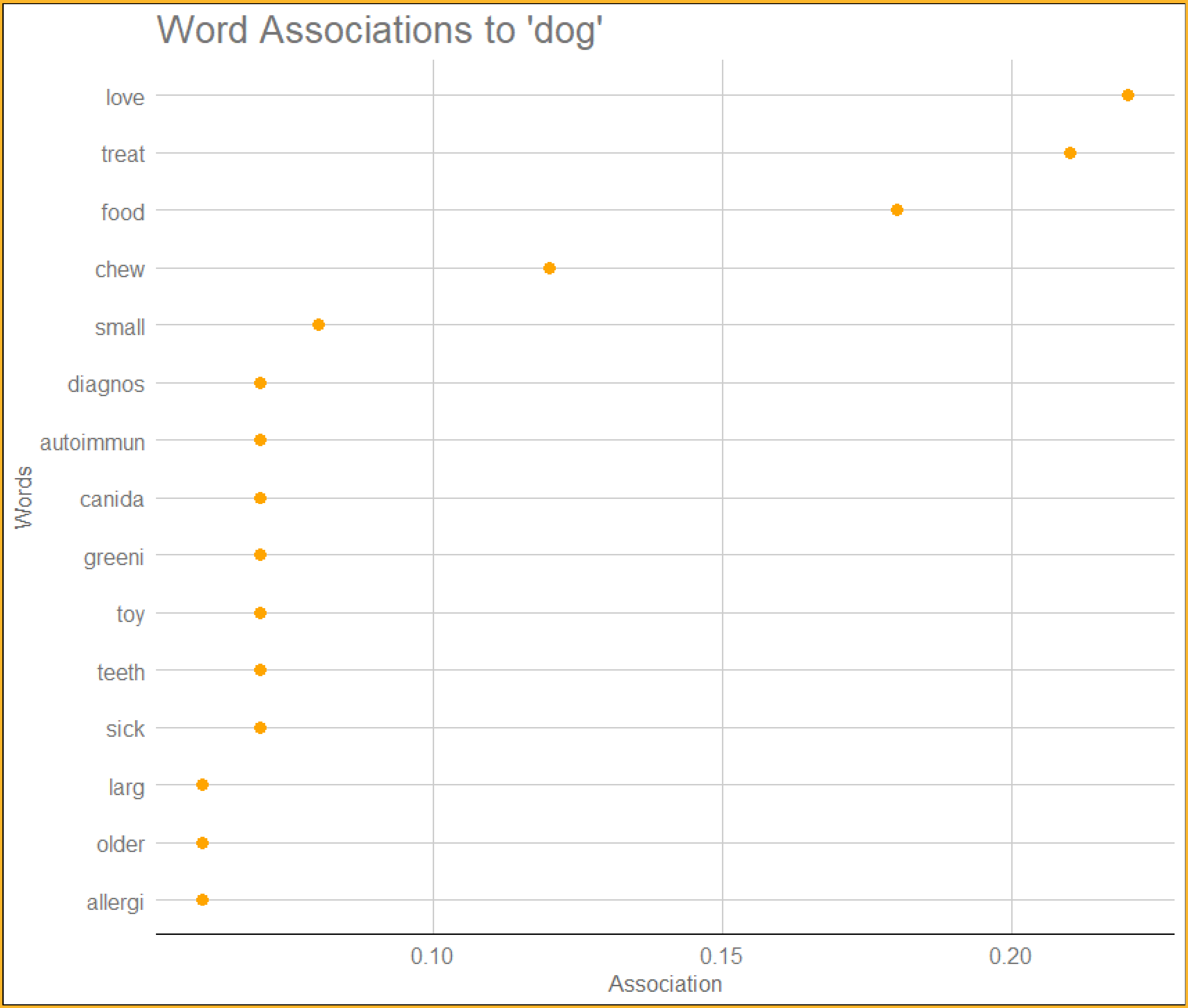


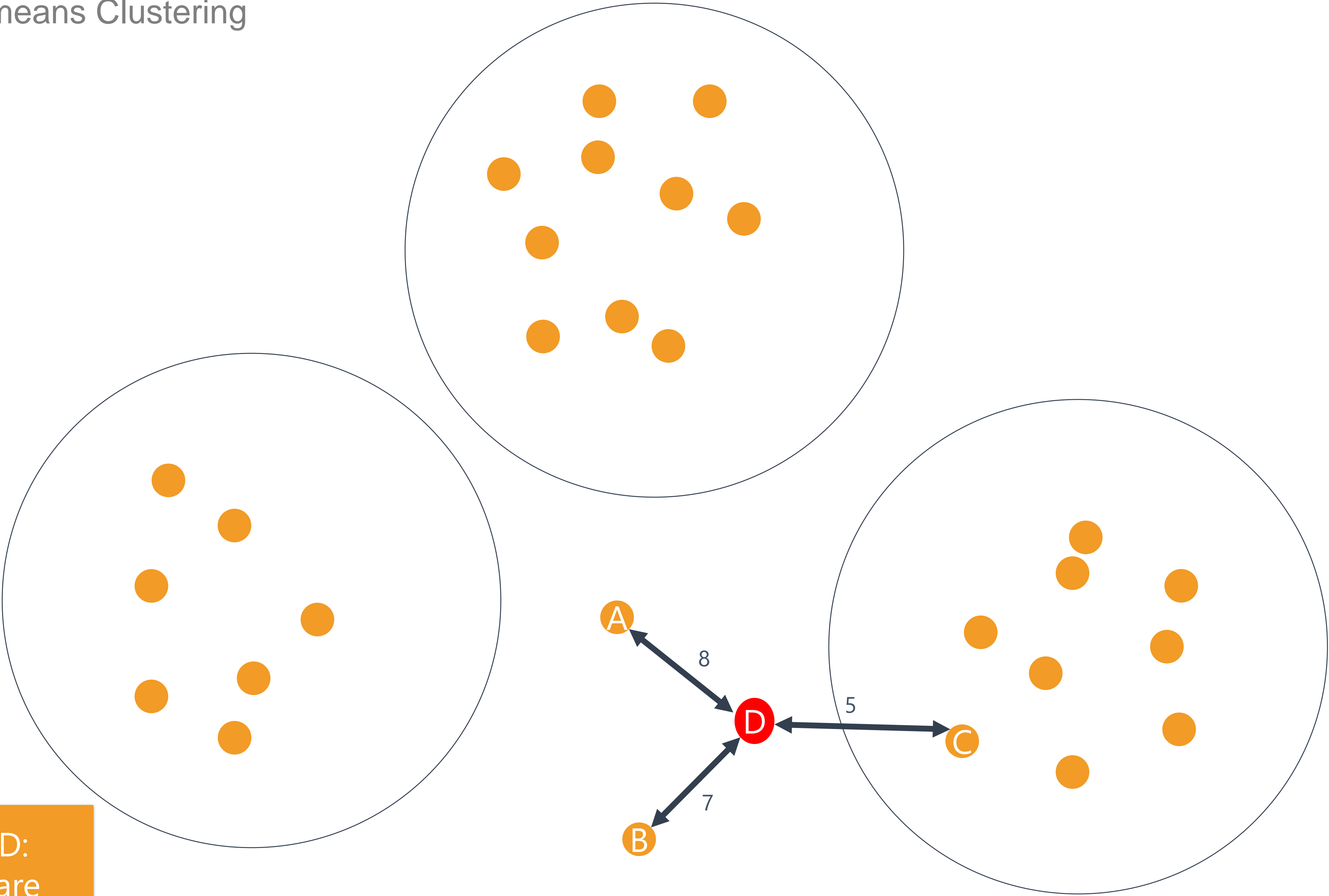
Word Association: Food & Drink



4.1 Sentiment Analysis

Word Association: Dog



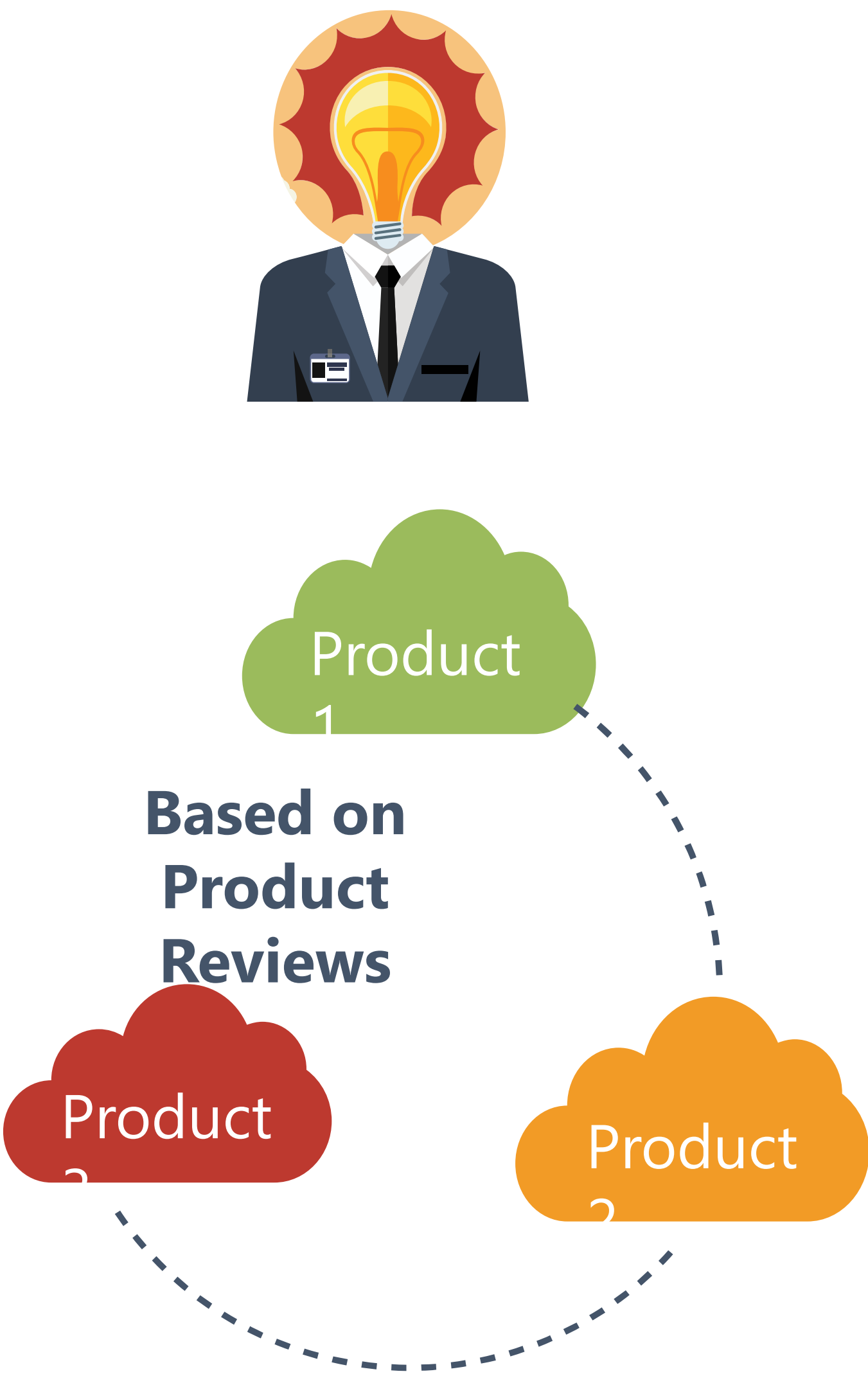


For Product D:
Suggestions are
Product C & B



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 | B002AQ00L2 | 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 | B002AQ00L2 | average rating is | 4.058035714285714 |



6. Business Use

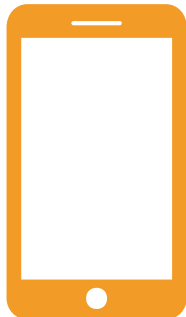
New Personalized Recommendation
(based on reviews)



**PERSONALIZED
EXPERIENCE**



**CROSS SELLING
COMPLIMENTARY PRODUCTS**



**ENGAGE
SHOPPERS**



Reach



Act

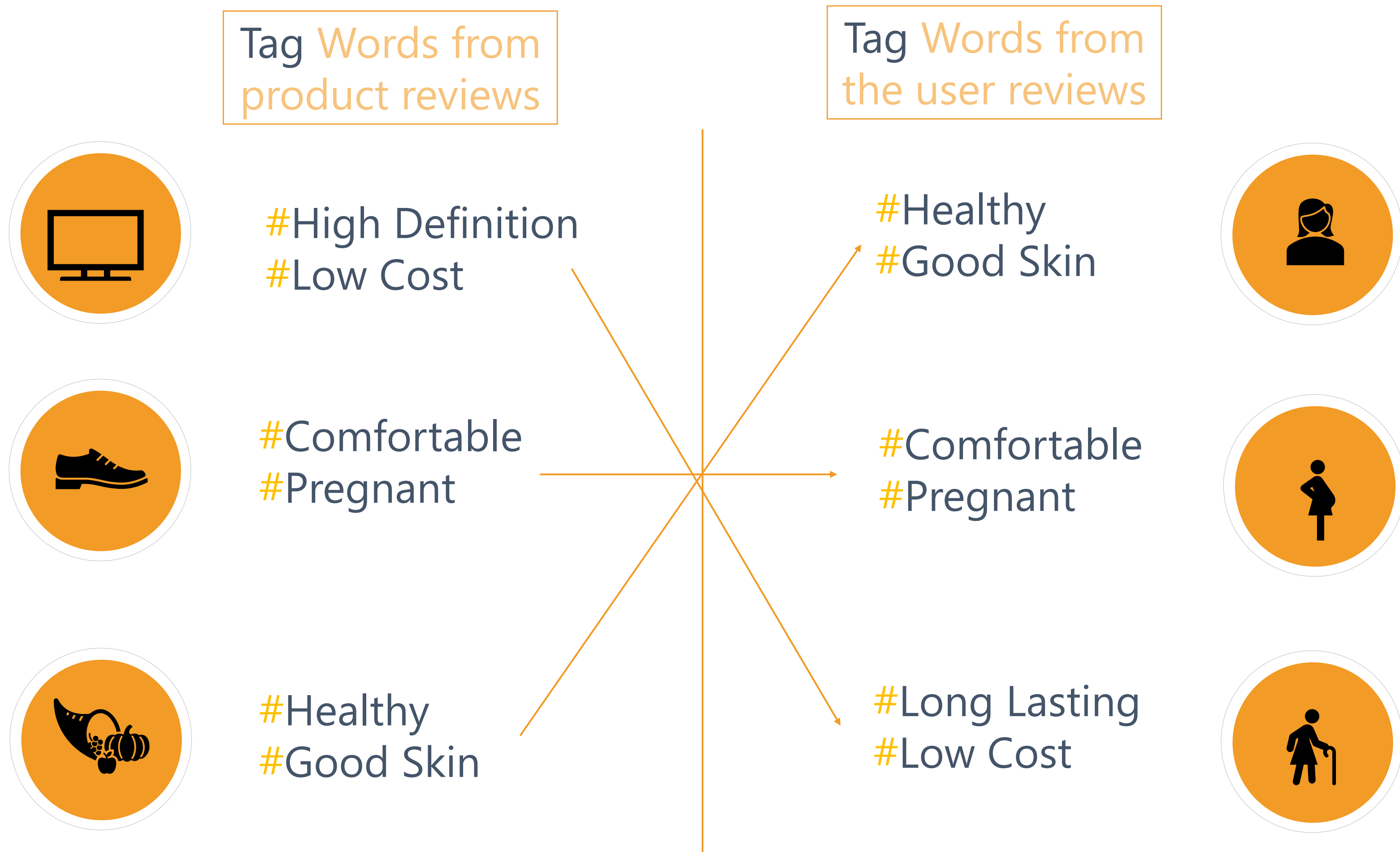


Convert



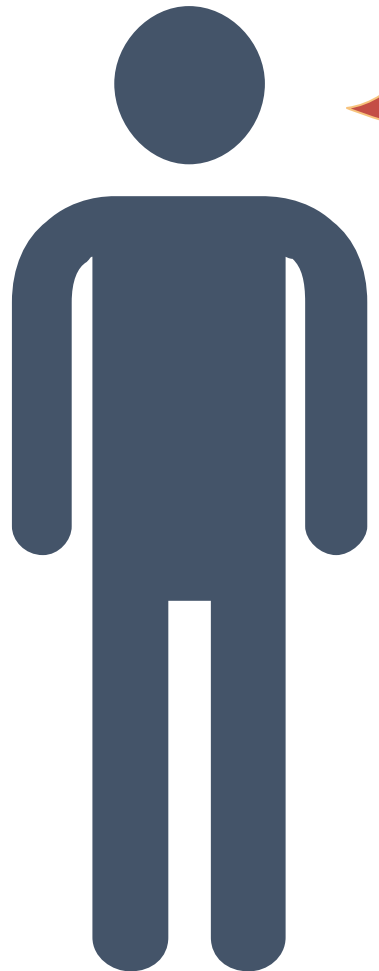
Engage

Future Scope : User Profiling by Tag words



Questions ?

Thank You



#Amazon Prime #Toys #Black Friday #Games #Movies #Personalized
#Ecommerce #Recommendation #United States #Year 2020
#AWS #Clothing #Office Products #Digital Music
#Gadgets #Brie #Customer Centric #Champions #Digital