Amazon Product Recommender System

Jin Hui Xu

UMBC Data Science Capstone

Contents

Introduction

Introduction

Why recommender system is important?

- It can drive traffic through personalized email messages to the store site and increase average order value.
- It also enhances the shopping experience by delivering relevant content based on personalized preferences.
- It can reduce workload for inventory management and boost work effectiveness.
- It can create comprehensive reports to support making the right decision for business direction.
- Overall, product recommender systems not only boost the companies' revenue but also increase customer satisfaction and loyalty.

Research Questions

Research Questions

- What characteristics are useful to generate personalized recommendations?
- Which recommender systems algorithms/methods are most successful and practical?
- Can textual data improve recommender systems' performance?

Data Sources

Data Sources

The data for this project is the Amazon Review Data (2018) which is collected by the University of California San Diego (https://nijianmo.github.io/amazon/index.html).

The dataset includes reviews data and product metadata.

It contains a total number of 233.1 million real reviews with the size of 34 gigabytes from Amazon.

Due to the computing resource limitation, a subset in Appliances category will be used for this project.

Review Data

There are a total of 602,777 review records in the Appliances category, and the dataset has 12 different features.

The interested feature are overall, reviewTime, reviewerID, asin, reviewText, summary.

Feature	Data Type	Description
reviewerID	String	ID of the reviewer
asin	String	ID of the product
reviewerName	String	name of the reviewer
vote	Integer	helpful votes of the review
style	String	a dictionary of the product metadata, e.g., "Format" is "Hardcover"
reviewText	String	text of the review
overall	float	rating of the product
summary	String	summary of the review
unixReviewTime	Integer	time of the review (unix time)
reviewTime	Datetime	time of the review (raw)
verified	Boolean	verified review
image	Object	images that users post after they have received the product

Product Data

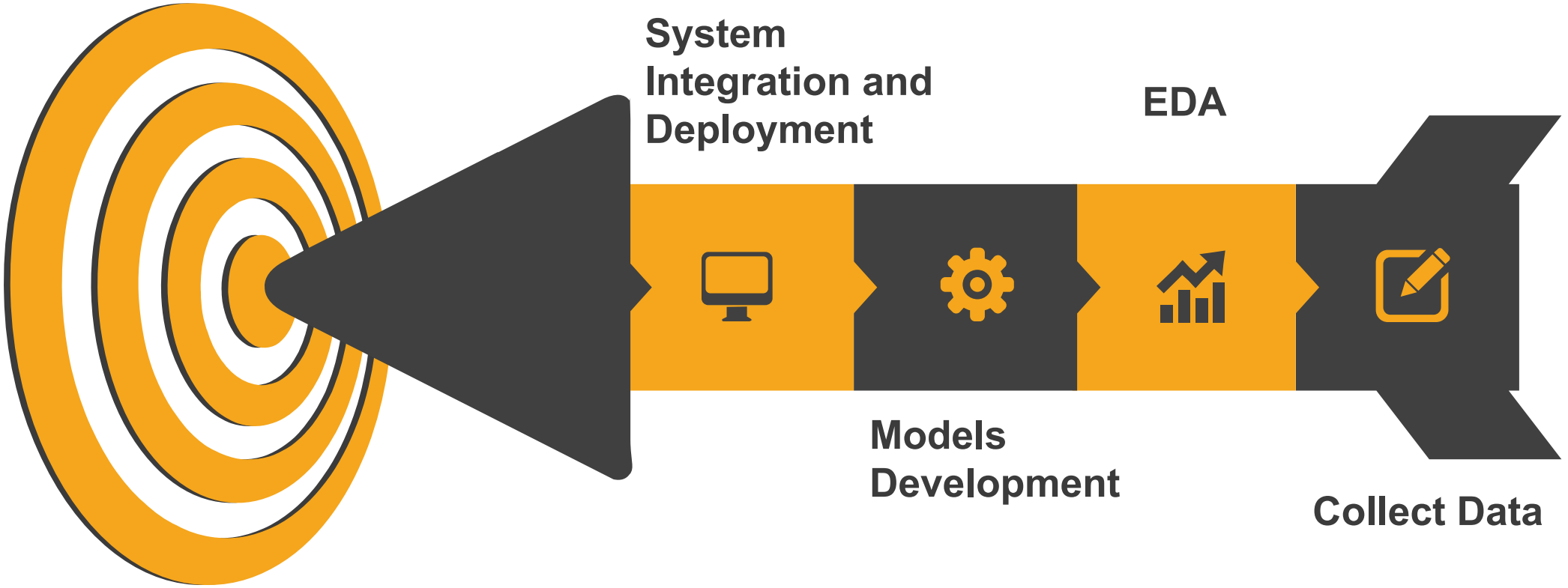
There are a total of 30,239 product records in this category, and the dataset has 19 different features.

The interested feature are category, description, title, brand, feature, main_cat, date, price, asin, imageURLHighRes.

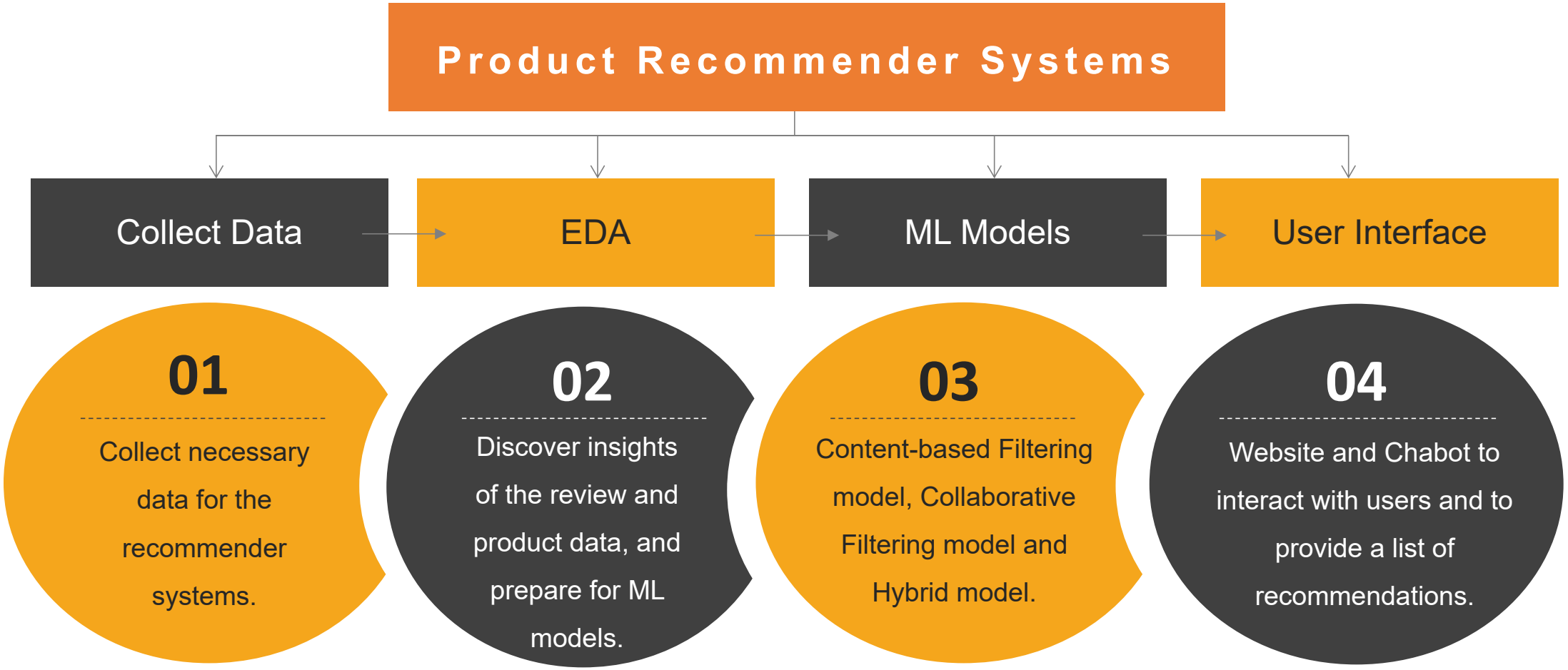
Feature	Data Type	Description
asin	String	ID of the product, e.g. 0000031852
title	String	name of the product
feature	String	bullet-point format features of the product
description	String	description of the product
price	String	price in US dollars (at time of crawl)
imageURL	Object	url of the product image
imageURLHighRes	Object	url of the high resolution product image
salesRank	String	sales rank information
brand	String	brand name
categories	String	list of categories the product belongs to
tech1	String	the first technical detail table of the product
tech2	String	the second technical detail table of the product
similar	String	similar product table
fit	String	size description of the product
also_buy	String	related products
also_view	String	related products
details	String	related product details
main_cat	String	main category the product belongs to
date	String	product release date

Research Process

Research Process

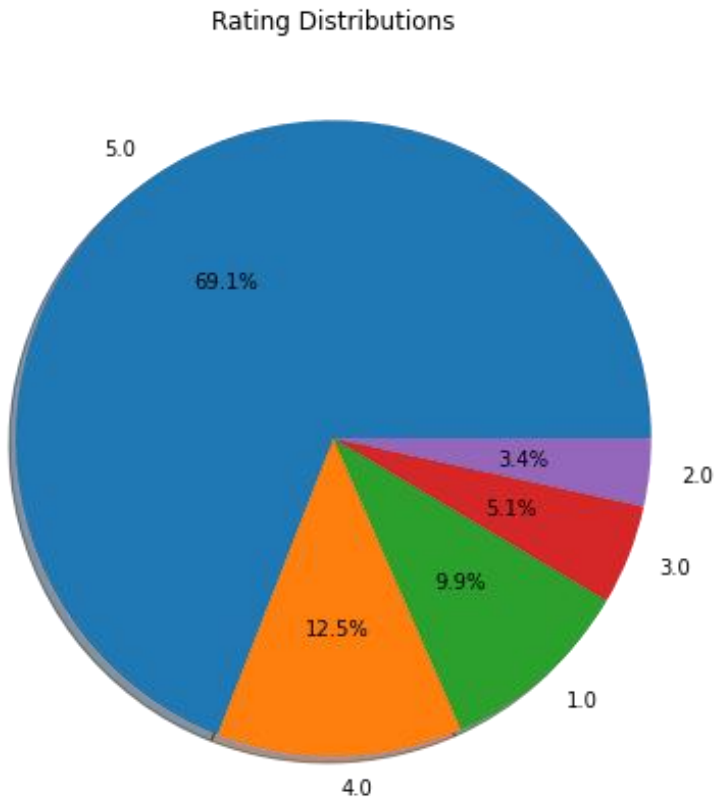


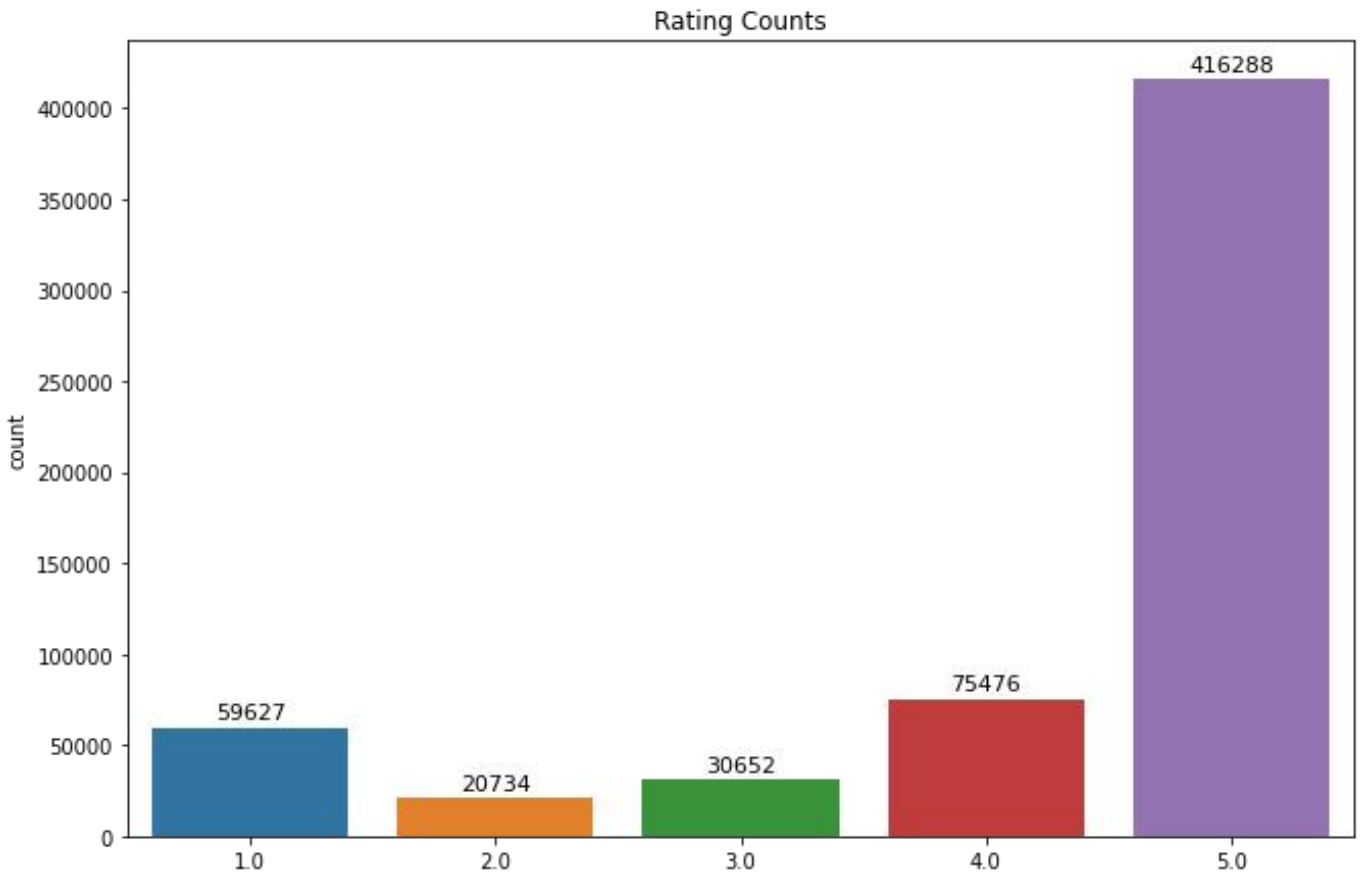
Research Process



Exploratory Data Analysis

EDA – Review Data





The rating distribution graphs show that the overall ratings in this review data set are highly imbalanced, which contains more than 69% of 5 stars rating.

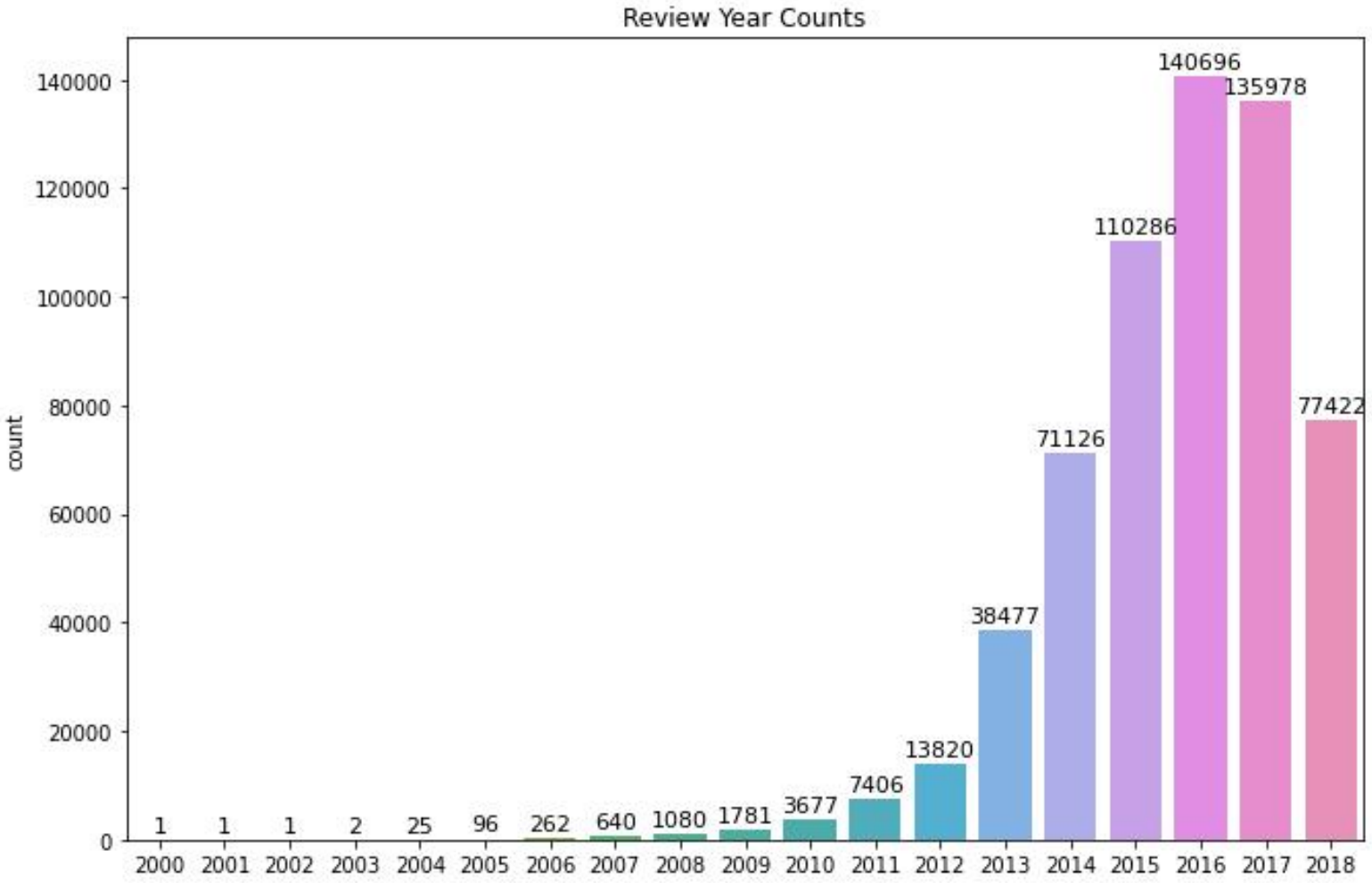
EDA – Review Data

	reviewerID	counts	rating_mean
412749	A8WEXFRWX1ZHH	208	4.980769
71295	A1IT56MV1C09VS	207	4.995169
142776	A21TPY9BVC9IKZ	206	5.000000
156061	A25C30G90PKSQA	206	3.000000
384058	A3TMNU7VNK5JJE	206	3.000000
...	...	...	...
183327	A2CH9B6K2QJS6Z	1	5.000000
183326	A2CH8ZFJWN1R60	1	5.000000
183325	A2CH7FPVP7H0XX	1	5.000000
183323	A2CH75VNS4T0GV	1	4.000000
515649	AZZZY1W55XHZR	1	3.000000

515650 rows × 3 columns

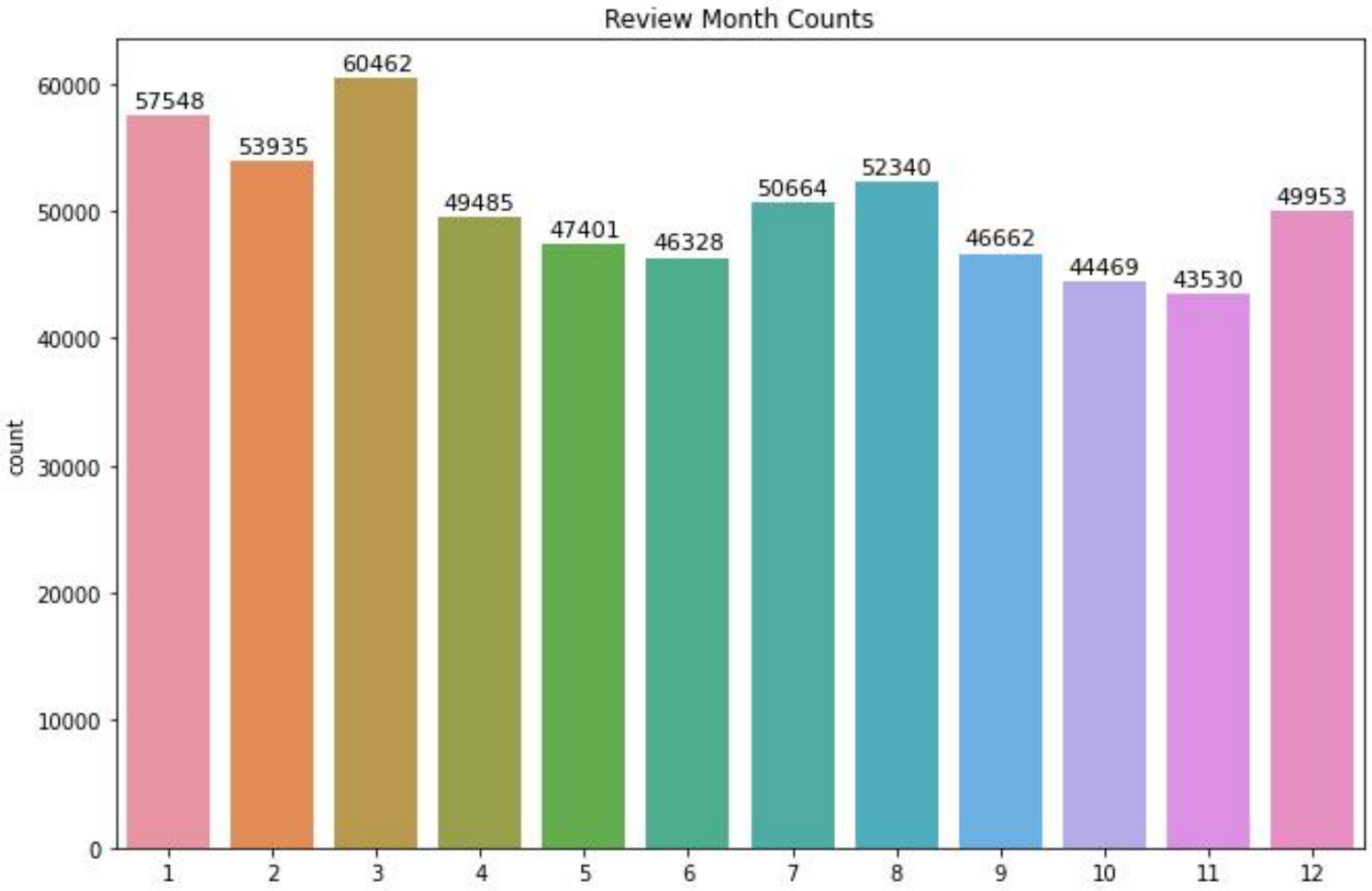
There are a total of 515,650 distinct reviewers in this dataset, and the most active reviewer had reviewed 208 products with an average 4.98 rating score.

EDA – Review Data



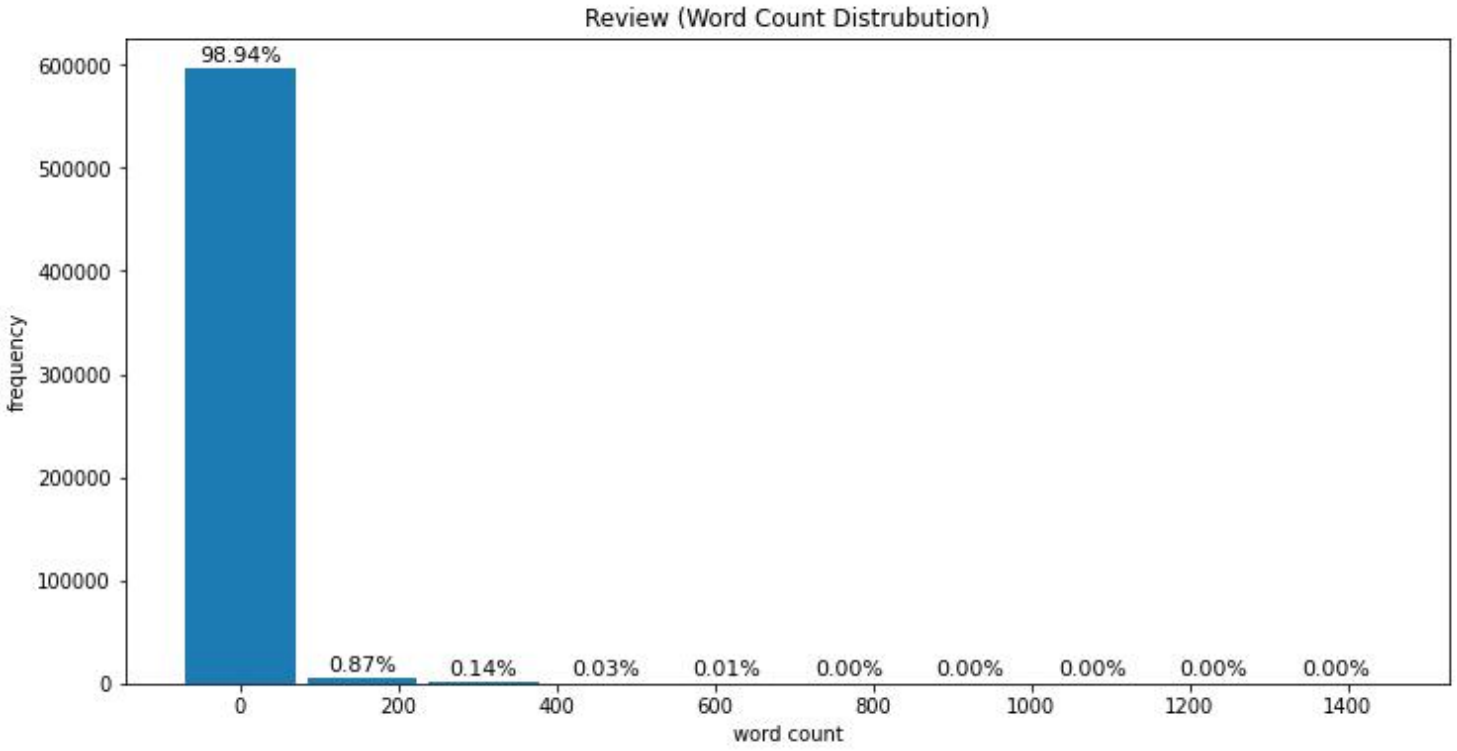
The review year distribution graphs show that the reviews in this dataset are heavily collected after the year 2013, which can quite well represent the current generation customers' preferences.

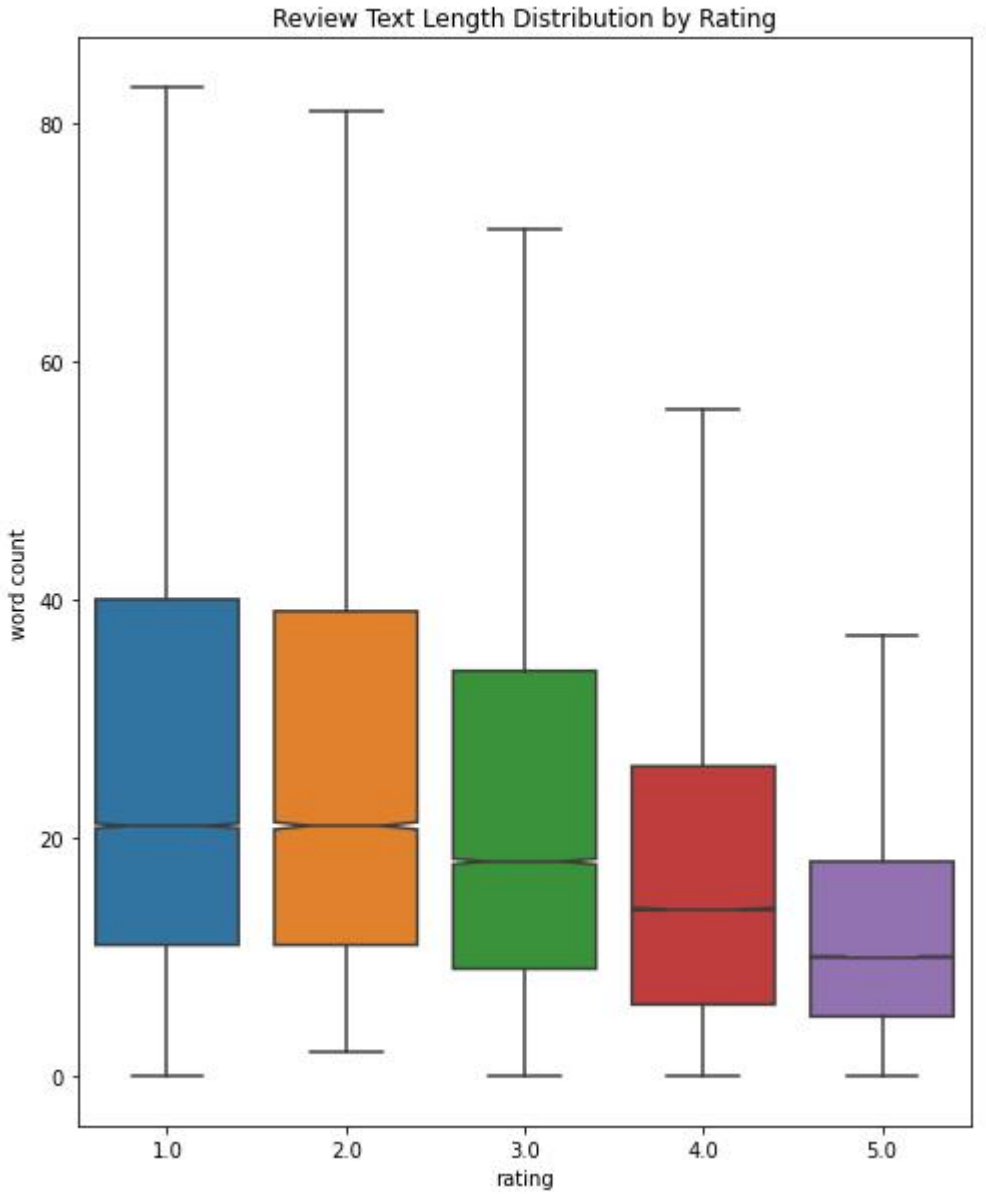
EDA – Review Data



The review month distribution graphs show that the months are quite evenly distributed in the dataset, which we can conclude that the season doesn't play a significant role in the influence of the purchase of the appliances.

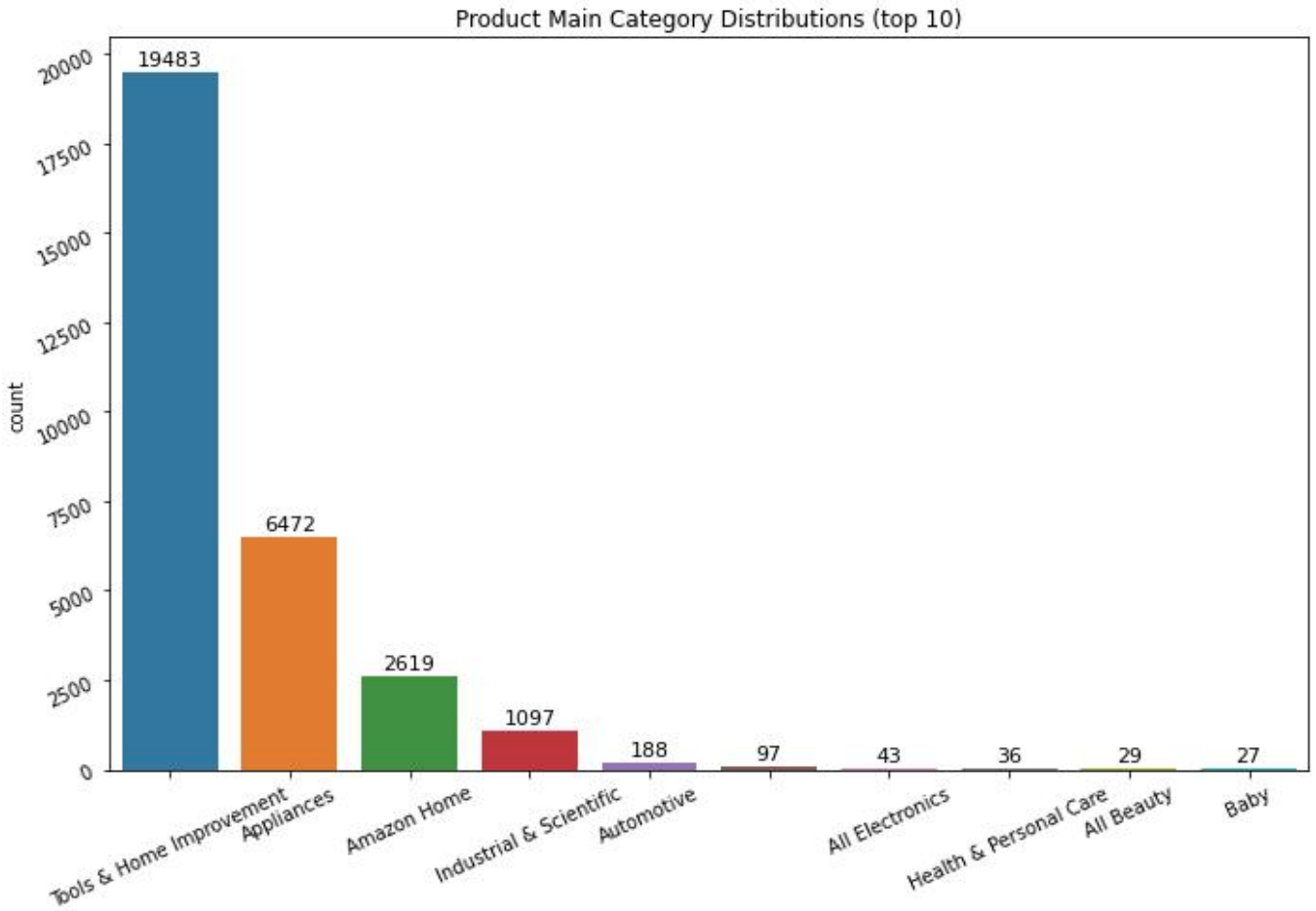
EDA – Review Data

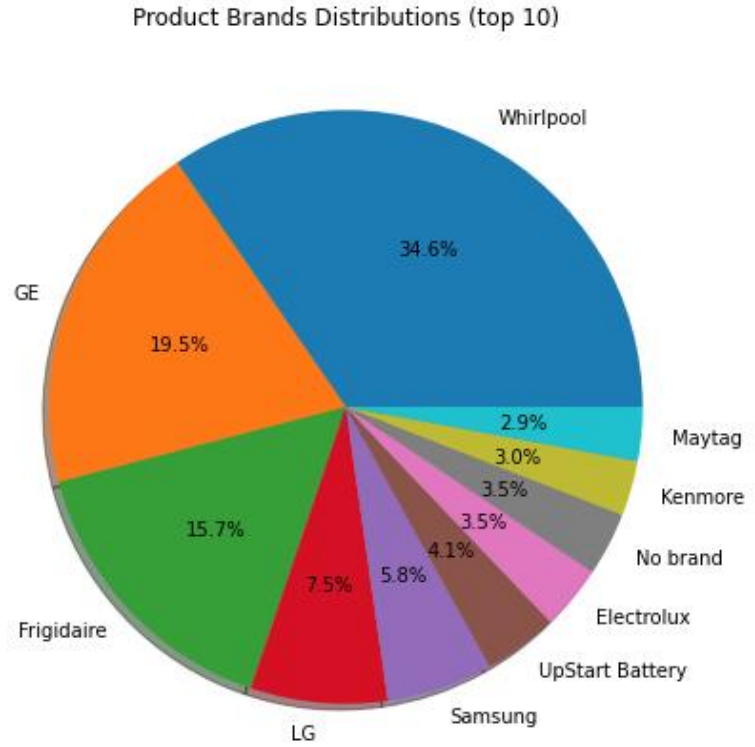




- Most of the reviews contain less than 100 words.
- The word counts distributions for each star rating review are similar. The box plot shows that the 5 stars rating reviews have the lowest interquartile range (IQR) compared to the other 4 ratings, which implies that it has average the shortest review text.

EDA – Product Data





- The majority of the products (64.7%) are in the Tools & Home Improvement category, and the Appliances category also holds 21.5%.
- There are a total of 2,762 brands, and Whirlpool is at the rank 1 position of amount of products.

EDA – Product Data

	asin	counts	rating_mean
421	B000AST3AK	6510	4.422427
5891	B004UB1O9Q	5702	4.341810
1634	B0014CN8Y8	4048	4.676383
17054	B00KJ07SEM	3200	4.409063
5289	B0045LLC7K	2936	4.403270
...	...	...	...
15012	B00GMJ0QCU	1	5.000000
15013	B00GMJ1IDQ	1	5.000000
15014	B00GMJ1XYU	1	5.000000
15018	B00GMJ5SGY	1	4.000000
30251	B01HJHHQM6	1	5.000000

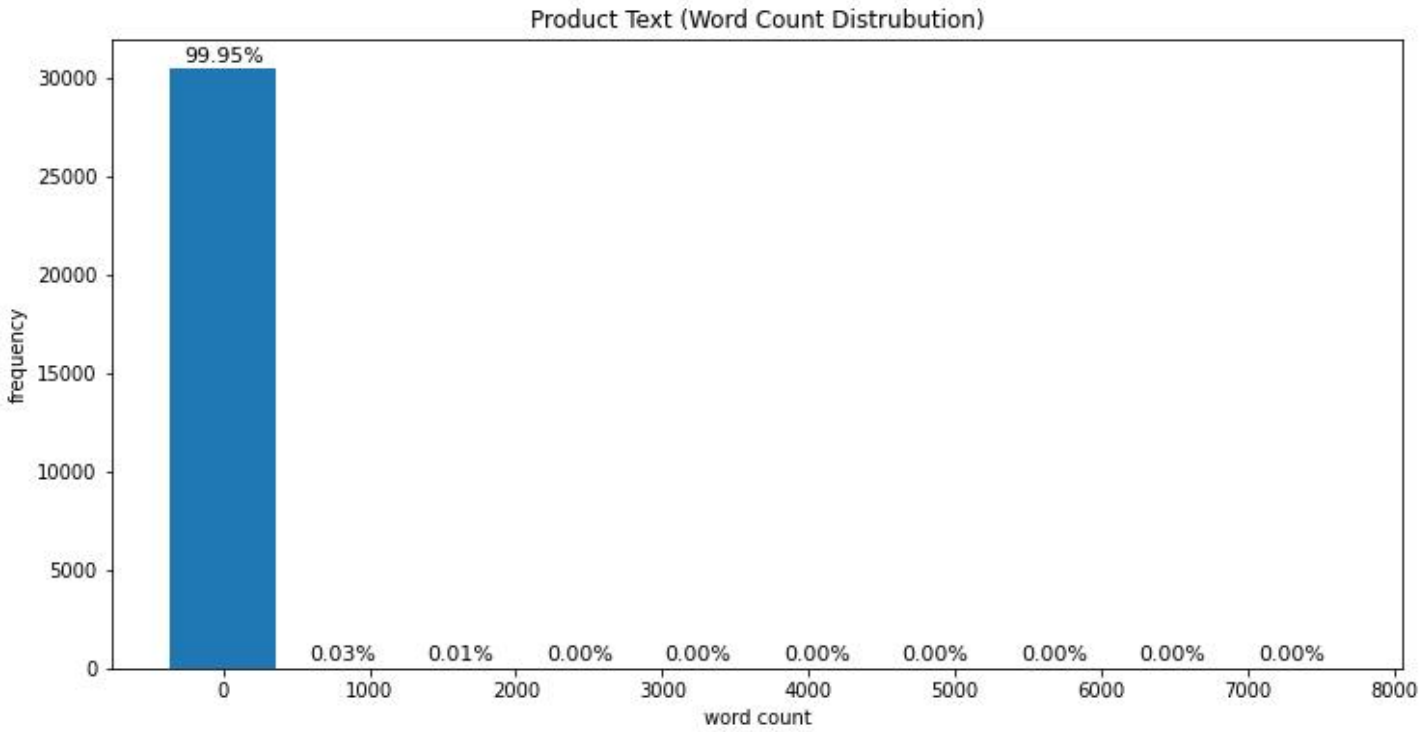
30252 rows × 3 columns

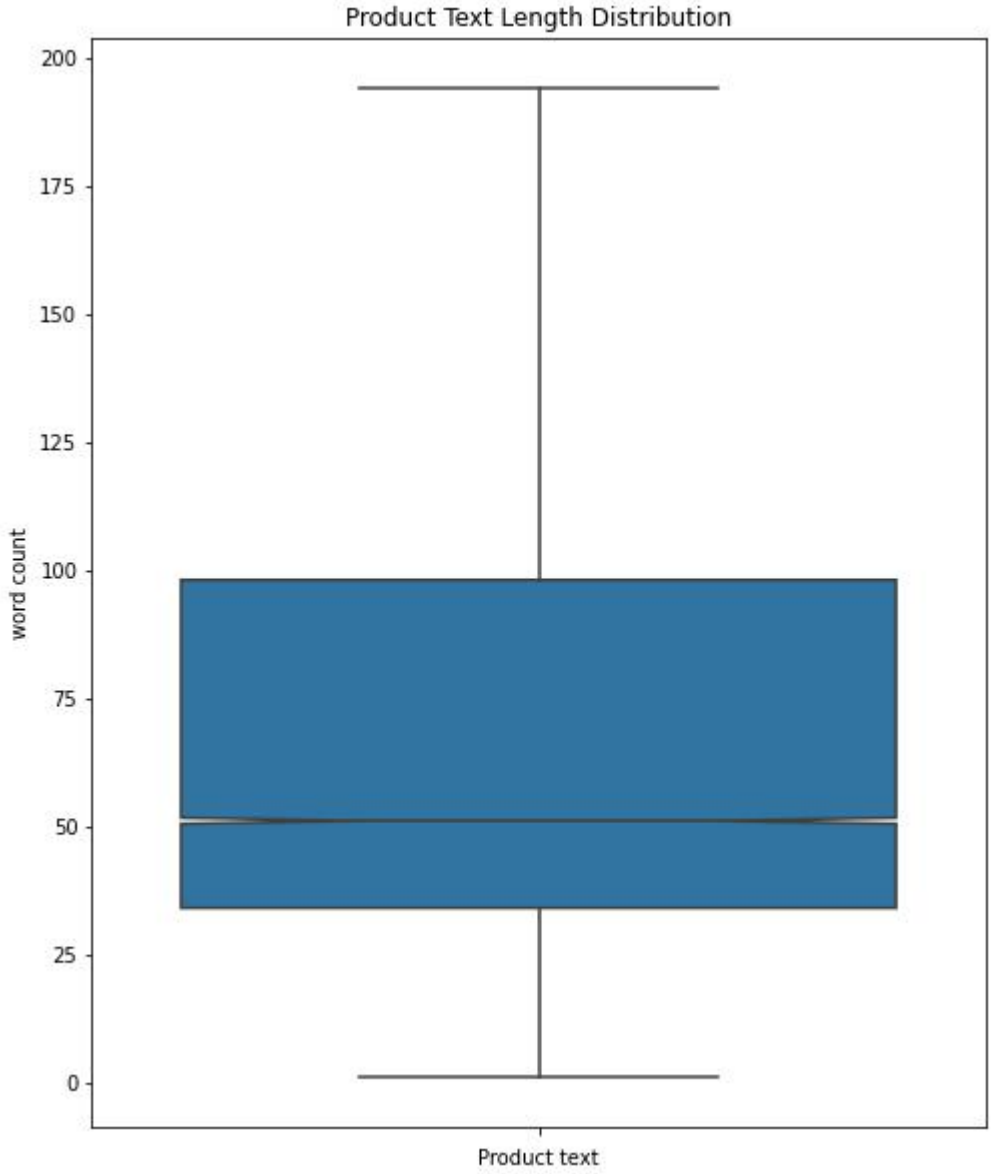
This is the list of the ranking of most reviewed products and their average ratings. Among 30,239 Appliances products, there are 30,252 products were reviewed. So there are some products are not included in the product dataset.

Within this list, the most reviewed product is General Electric MWF Refrigerator Water Filter, and the second most reviewed product is Samsung Genuine DA29-00020B Refrigerator Water Filter, 3 Pack. Both of them are Refrigerator Water Filters.

Most reviewed product: General Electric MWF Refrigerator Water Filter Second most reviewed product: Samsung Genuine DA29-00020B Refrigerator Water Filter, 3 Pack

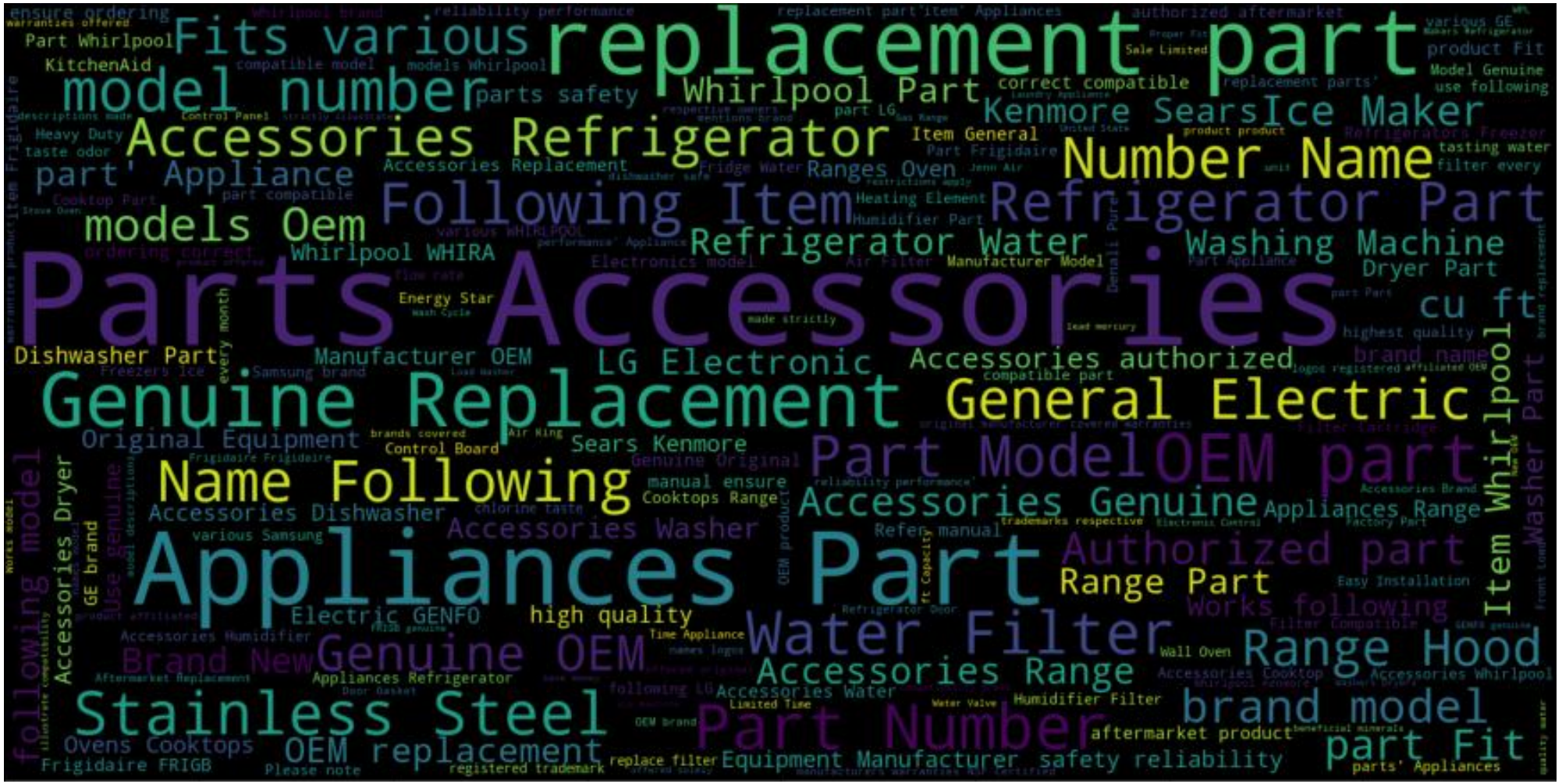
EDA – Product Data





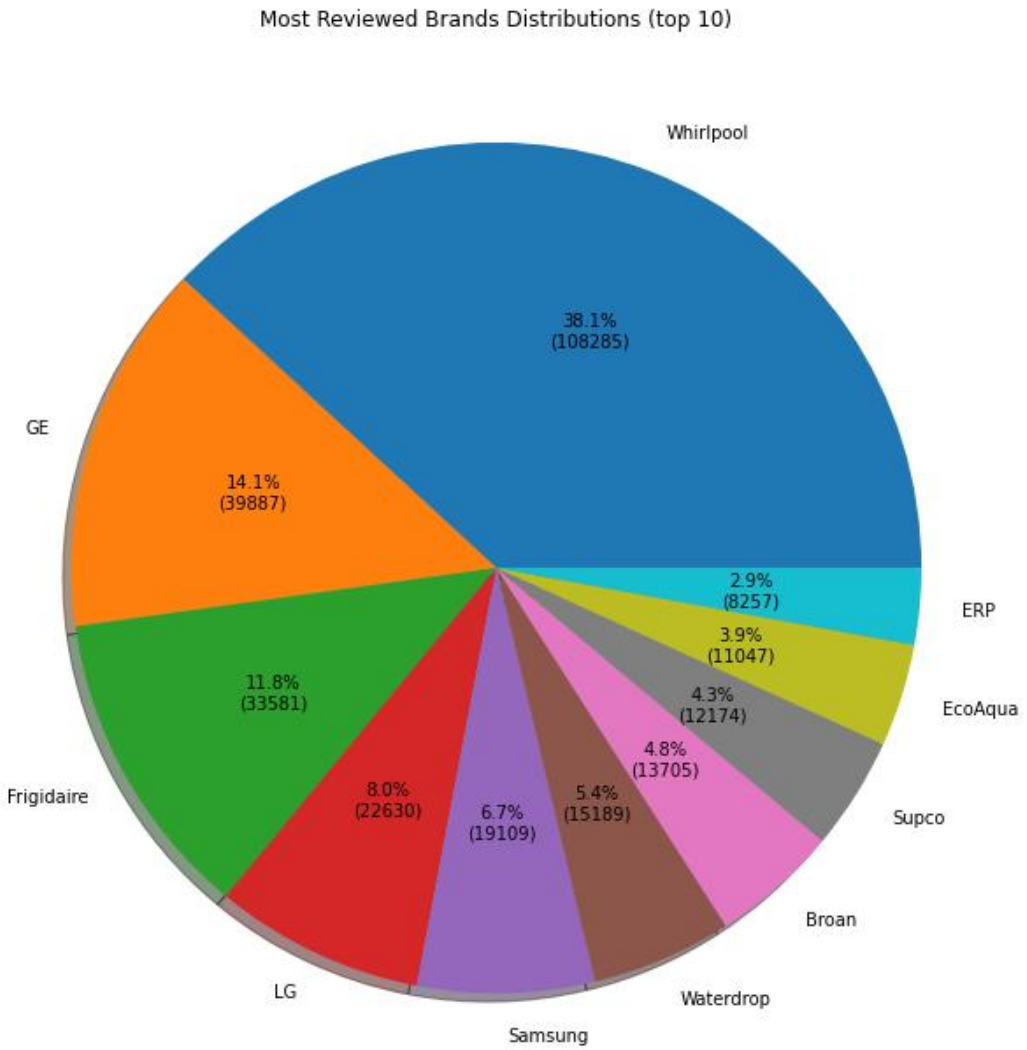
The product text distribution histogram and box plot show that majority of the product text is less than 1000 words. There are only a few outliers that are greater than 2000 words, so for future NLP model development, in order to reduce the padding size, we can consider a smaller number instead.

EDA – Product Data



The word cloud shows that the most frequently used words for Appliances products are related to replacement, part, and model number.

EDA – Merged Data



- Most Reviewed Brands Distributions (top 10) graphs show that Whirlpool products have the rank 1 position of amount of reviews.
- There are some other brands in the list that are not in the list of top 10 product numbers, which means offering more products doesn't imply more sales and revenue.

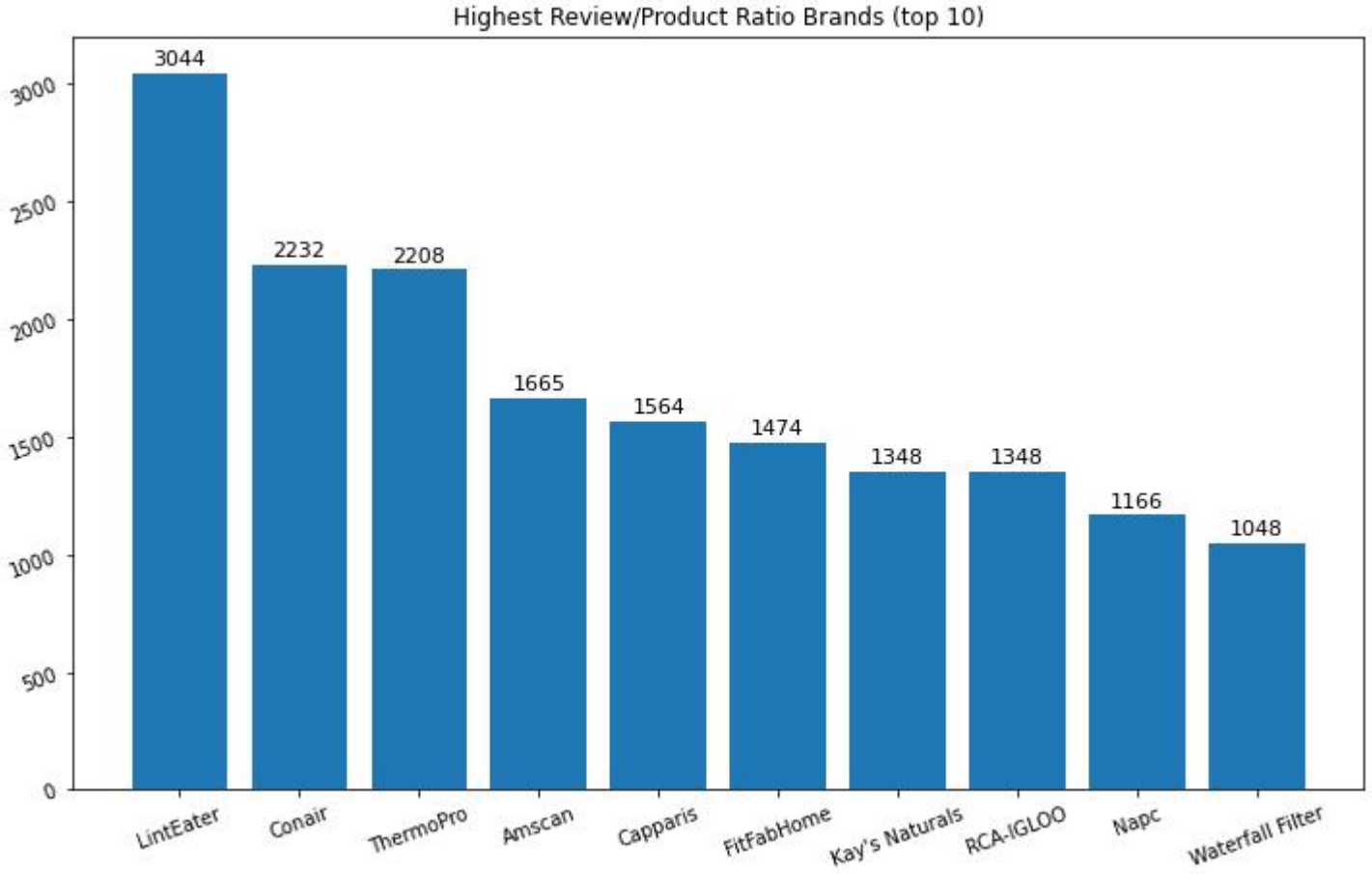
EDA – Merged Data

	brand	review_counts	rating_mean
1536	LintEater	6088	4.617280
2570	Waterdrop	15189	4.525183
2294	Supco	12231	4.494808
2591	Whirlpool	108295	4.477280
1419	Kenmore	5490	4.389253
1491	LG	22630	4.382501
769	ERP	8257	4.378225
793	EcoAqua	11047	4.374038
967	Frigidaire	33581	4.367738
991	GE	40213	4.357049

This table shows the top 10 average rating brand (reviews > 5000) in the dataset.

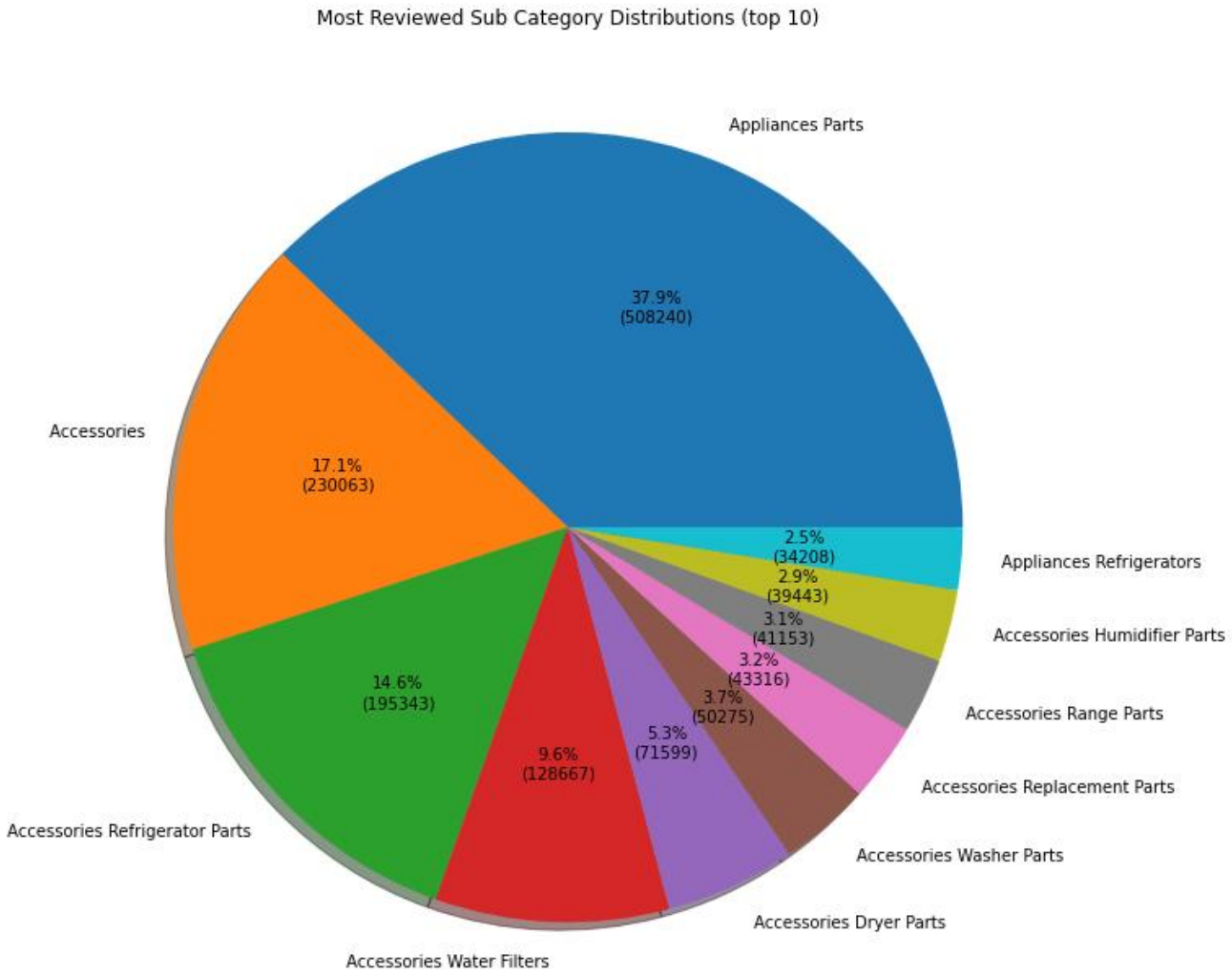
The result show that brand LintEater has the highest average rating 4.62 with over 6,000 reviews. Whereas Whirlpool has the rank 4 in this list.

EDA – Merged Data



This graph shows that LintEater has the best review per product ratio in the dataset. And most of the brands are not in the top ranking of the number of products, which again proves that offering more products doesn't imply more sales and revenue.

EDA – Merged Data



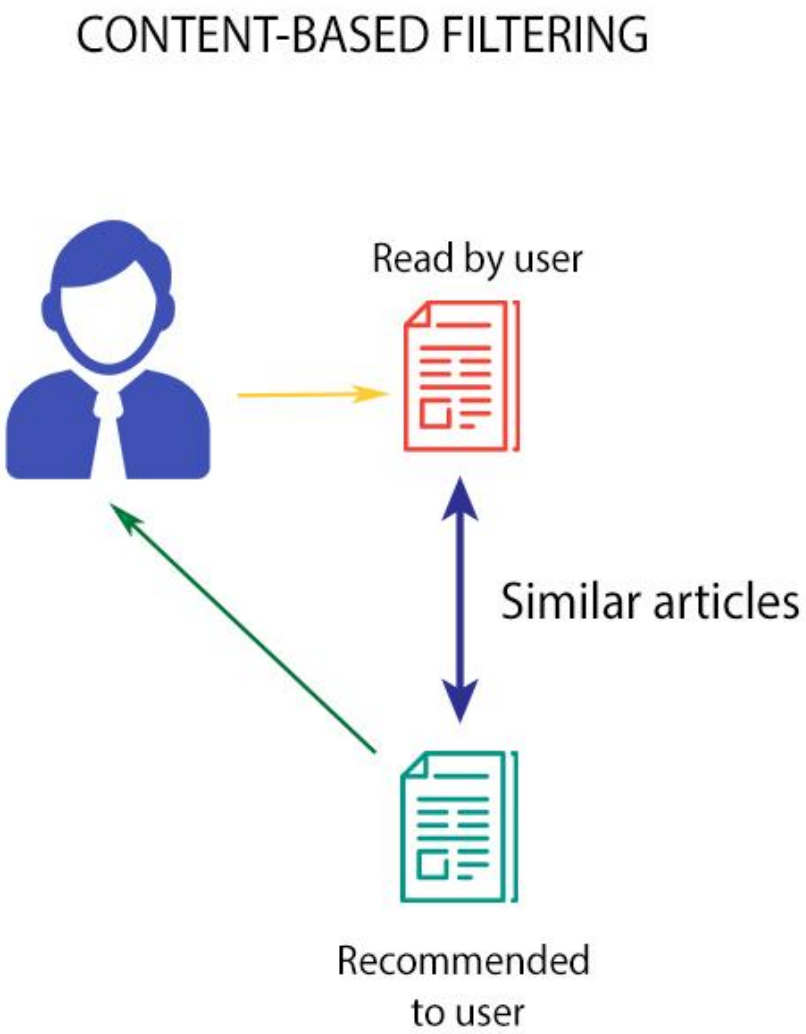
- Most Reviewed Sub Category Distributions (top 10) graphs show that 37.9% of the reviews are in the Appliances Parts sub category, and the Accessories sub category also holds 17.1% in the dataset.

Machine Learning Models

Base Model

A base model is a simple knowledge-based recommender that takes user inputs such as product category, brand, release year, and targeted price to search for matching products. It usually doesn't leverage machine learning to provide recommendations.

Content-Based Filtering



The idea of content-based filtering is to find the similarity products based on either metadata or product description. The most feasible approach is to apply the cosine similarity method against the textual data to find the most similar products.

Content-Based Filtering

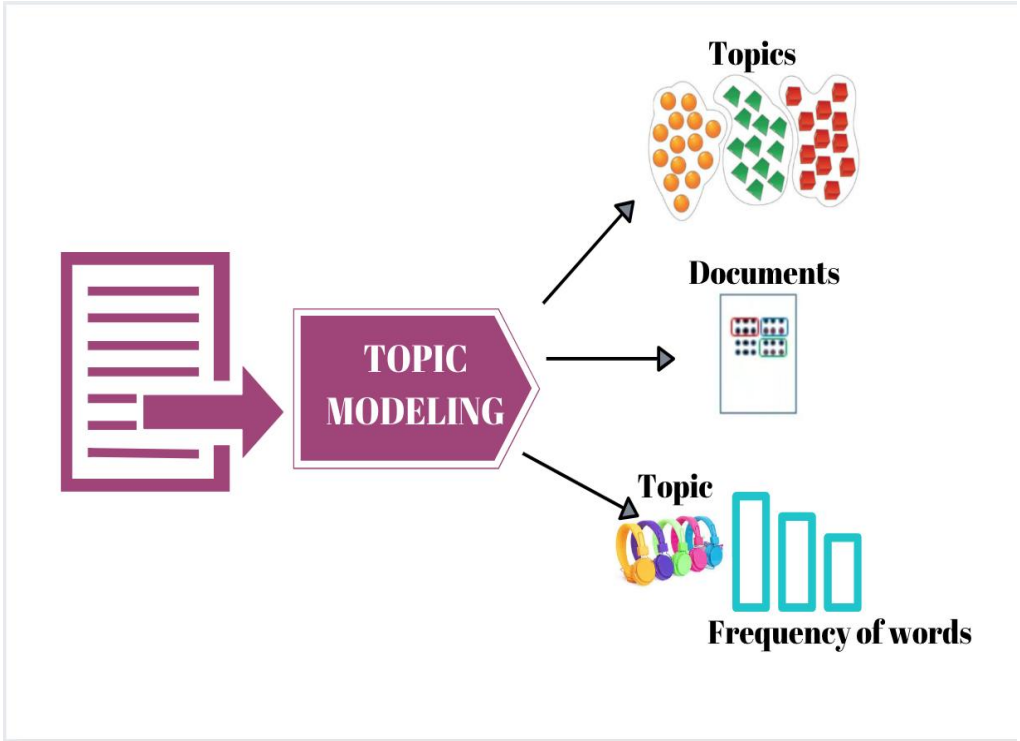
Product Description Cosine Similarity Model

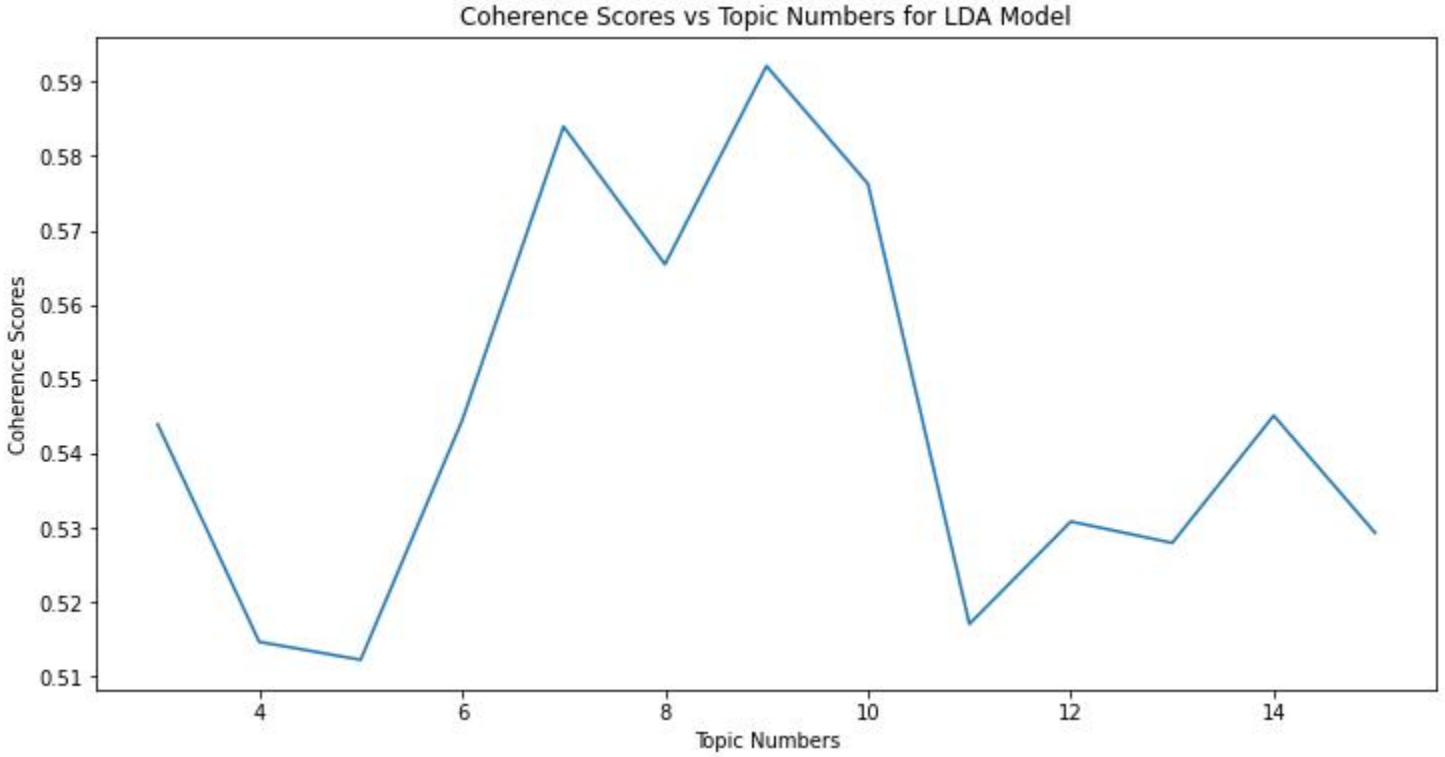
```
Description Based Recommender Result for B0001YH10C, coldmate mr 128 mini cooler warmer deluxe mini refrigerator:
1. B00ID8CLMG, Avanti FF45006W 4.3 CF Frost Free Refrigerator Freezer, White
2. B00RNAH5OY, goFridge Mini Fridge Portable Electric Cooler
3. B001H80RN4, Frigidaire 241505301 Refrigerator Door Bin Genuine Original Equipment Manufacturer (OEM) Part
4. B004NEYPYQ, Frost-Free 4.3 Cu. Ft. Refrigerator/Freezer White
5. B000JLL3BK, Pek Vino Vault Wine Preserving Refrigerator, Silver
6. B01F79MKME, Amana AMA43BK Compact Single Door Refrigerator, 4.3 cu. ft, Black
7. B00OVI6HHW, Avanti AR4456SS Counterhigh Refrigerator, 4.5 cu. ft, Black/Stainless Steel
8. B001TIYPI0, Whirlpool Part Number 2179374: Wine Rack
9. B001F7H4RY, PORTABLE COOLER WARMER MINI FRIDGE WINE BEER
10. B001775T4C, Nostalgia Electrics CRF170RETRORED Retro Series Mini Fridge, 1.7 Cubic Feet
```

Product Metadata Cosine Similarity Model

```
Description Based Recommender Result for B0001YH10C, coldmate mr 128 mini cooler warmer deluxe mini refrigerator:
1. B001F7H4RY, PORTABLE COOLER WARMER MINI FRIDGE WINE BEER
2. B00YNNEC8Q, Mini Wine Cooler
3. B00ND5CWAA, Phoenix USB 5v Portable One Zip-top Can Cooler-mini Car Compact Refrigerator and Warmer
4. B00YNMUYV6, Mini Wine Cooler Refrigerator with Lock
5. B00RNAH5OY, goFridge Mini Fridge Portable Electric Cooler
6. B016K4J3U2, Honeykoko Mini USB PC Refrigerator Fridge Beverage Drink Can Cooler Warmer Heater Gadget One Can in Home Office
7. B016KQ7X8E, ThreeH New Mini Red USB Fridge Cooler Beverage Drink Cans Cooler/Warmer Refrigerator for Laptop PC Computer Red H-UF05Red
8. B0187KYRQC, Coca-Cola Mini Can Cooler
9. B00KE7FM3O, Mini USB Desktop Fridge Cooler Refrigerator
10. B005JAVC94, Mini Desktop Fridge Cooler Personal Fridge(Black)
```

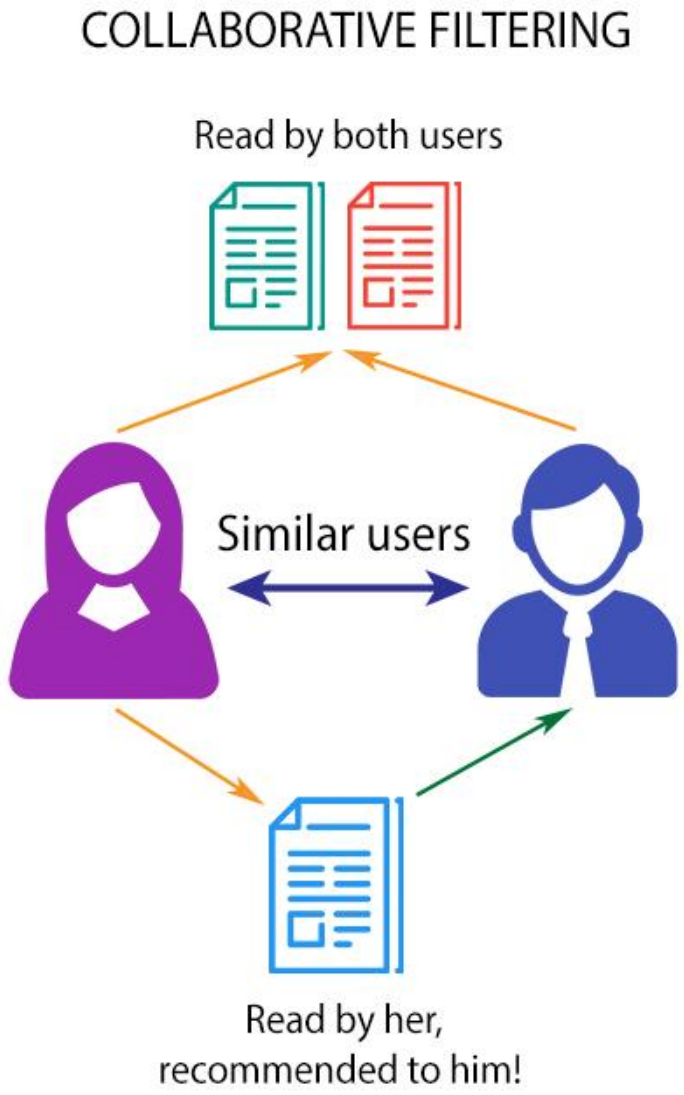
Content-Based Filtering





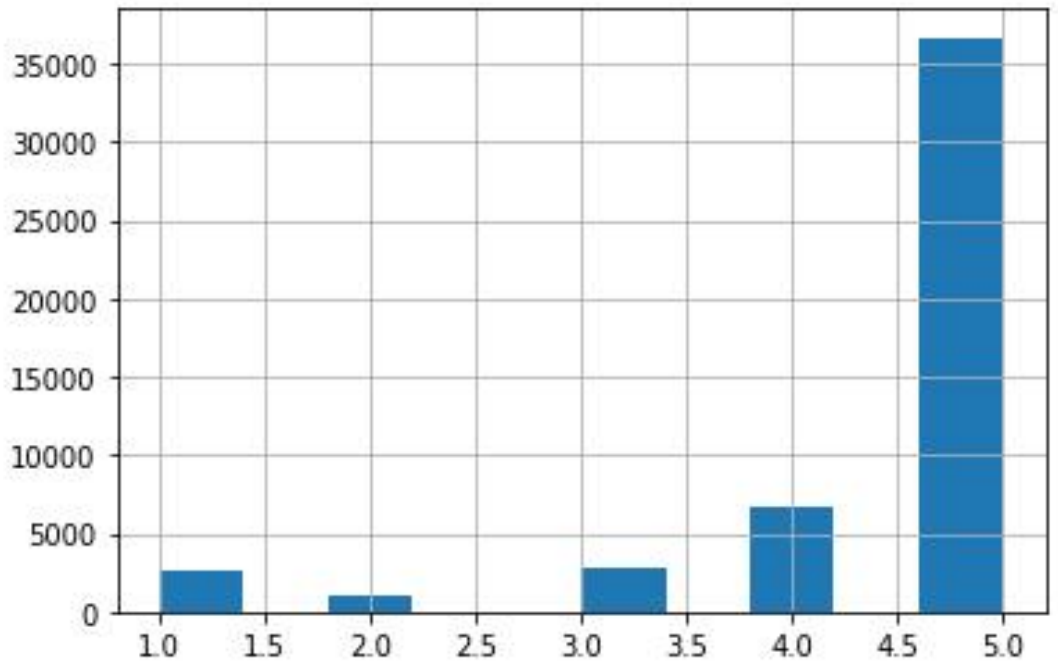

```
Topic Modeling Recommender Result for B0001YH10C, coldmate mr 128 mini cooler warmer deluxe mini refrigerator:
1. B013PRRB4W, Power Pair Special-LG Turbo Series Ultra-Capacity Laundry System with Steam*PURE WHITE COLOR*(WM4270HWA_DLEX4270W)
2. B013PSOBNA, Power Pair Special-LG Turbo Series Ultra-Capacity Laundry System with Steam and Matching Storage Pedestals *GRAPHITE STEEL*(WM4270HVA_DLEX4270V_WDP4V X 2)
3. B013PT1PFQ, Power Pair Special-LG Turbo Series Ultra-Capacity Laundry System with Steam*GRAPHITE STEEL*(WM4270HVA_DLEX4270V)
4. B00HX3ZJKS, PAIR SPECIAL- LG Turbo Series Ultra Capacity Laundry System With Steam Technology (WM3470HVA,DLEX3470V,WDP4V x2)
5. B0049OSUD2, Maytag MFI2665XEM Ice2O 25.5 Cu. Ft. Stainless Steel French Door Refrigerator - Energy Star
6. B00MG225MQ, Power Pair Special- LG Turbo Series Ultra Capacity Laundry System with Steam Technology(WM3570HWA_DLEX3570W)*PURE WHITE IN COLOR*
7. B00MG17WBQ, POWER PAIR SPECIAL-LG TURBO SERIES ULTRA CAPACITY LAUNDRY SYSTEM WITH STEAM TECHNOLOGY, AND STAINLESS DRUMS (WM3570HVA_DLEX3570V) *GRAPHITE STEEL COLOR*
8. B00MG1FT5M, POWER PAIR SPECIAL-LG Turbo Series ultra Large Capacity Laundry System With Steam Technology (WM3570HVA_DLEX3570V_WDP4V X 2) *GRAPHITE STEEL COLOR*
9. B00P9VMNPK, LG H/E Ultra Large Capacity Top Load Laundry System with Turbo Wash Technology (WT5680HWA_DLEX5680W) ELECTRIC DRYER
10. B00UNTG9XK, Electrolux Frigidaire Professional FPBS2777RF 27.8 Cu.Ft. Stainless French Door Refrigerator
```

Collaborative Filtering

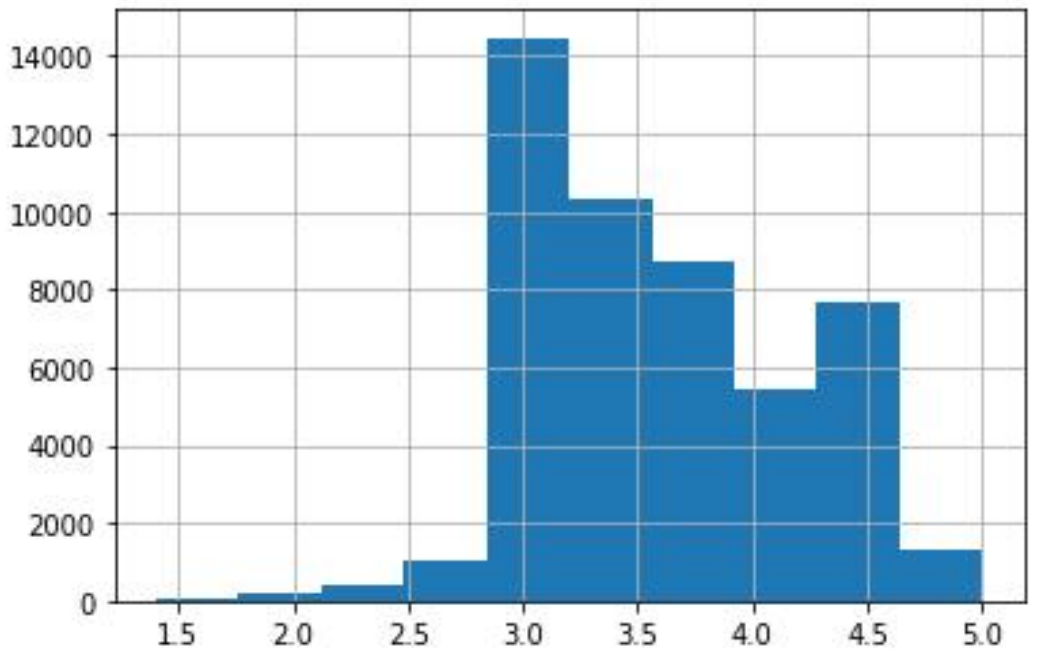


The idea of collaborative filtering for recommender systems are methods based on past interactions recorded between users and items to generate new recommendations. The past user-item interactions represent the bases to detect similar users and/or similar items and to make predictions based on estimated proximities

Collaborative Filtering



Review Rating Distribution



Review Sentiment Score Distribution

Collaborative Filtering

Algorithm	test_rmse	fit_time	test_time
SVDpp	0.570962	7.235222	0.225179
SVD	0.577325	2.171684	0.069904
BaselineOnly	0.581488	0.129146	0.046562
KNNBaseline	0.591287	5.741593	0.804107
KNNBasic	0.615071	5.544654	0.778922
KNNWithMeans	0.637690	5.649759	0.762275
Neural Network	0.639841	NaN	NaN
KNNWithZScore	0.640280	5.923294	0.777118
SlopeOne	0.660248	2.148028	0.123483
NMF	0.743142	3.774009	0.104871
CoClustering	0.752756	2.216395	0.048579
NormalPredictor	0.850930	0.065574	0.066035

Collaborative Filtering

Evaluating RMSE of algorithm SVD on 10 split(s).

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Mean	Std
RMSE (testset)	0.5565	0.5578	0.5681	0.5585	0.5612	0.5643	0.5577	0.5623	0.5570	0.5593	0.5603	0.0035
Fit time	20.40	20.28	20.39	20.48	20.58	20.71	21.01	21.09	20.50	21.30	20.67	0.33
Test time	0.13	0.13	0.13	0.13	0.13	0.15	0.20	0.13	0.17	0.12	0.14	0.02

Collaborative Recommender Result for customer A1CY6CQC5HPQGL:

1. B00LPDJSV8, Maytag 22002315 Washer Drive Pulley Genuine Original Equipment Manufacturer (OEM) part for Maytag & Amana
2. B002YTOCP4, Whirlpool Part Number 8181673: Gasket, Tub
3. B00IMO2Y1Y, Whirlpool 8181887 Handle for Washing Machine
4. B00L4JD7S8, LG Electronics 5005JJ2014A Refrigerator Door Shelf/Bin, White with Clear Trim
5. B00AFSMIP2, Supco MP21MA 2100W Cooktop Stove Surface Element Replacement for 0E00801799, Y04100166, AP4283501
6. B0053F8Y0A, 2206671B Whirlpool Refrigerator Grille Overflow (black)
7. B0156NCY8G, General Electric WR17X11264 Funnel Ice Display
8. B001DHNT7K, General Electric WR72X209 Drawer Slide Rail
9. B0156NDG4M, Whirlpool W10074200 Gasket Door
10. B00DZUA3DQ, Whirlpool 2206670W Grille

Hybrid Model

Content-based Filtering Model

Collaborative Filtering Model

Hybrid Recommenders Model

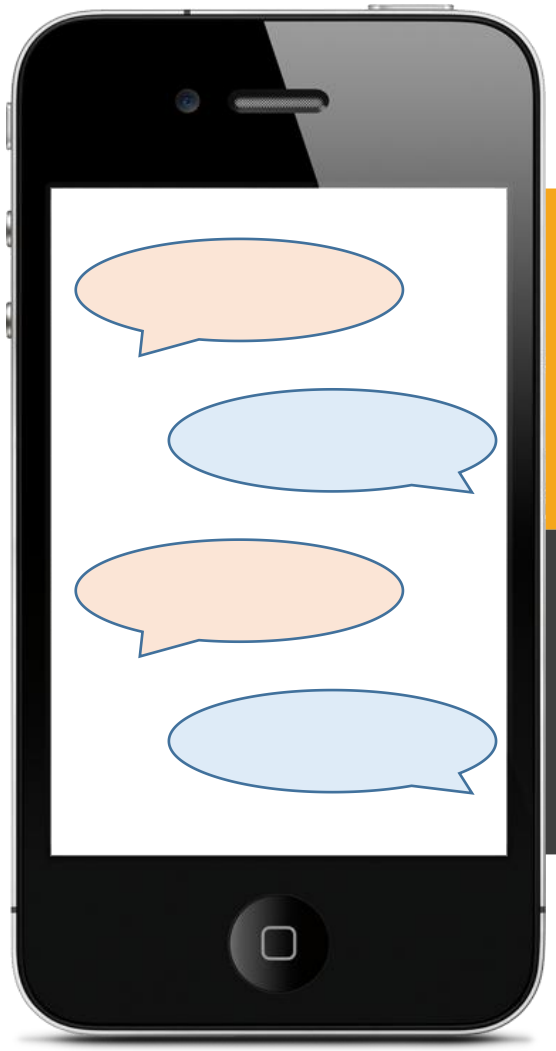
Hybrid Model

The recommendation result shows that the hybrid model is suggesting more products that are similar to the product ID B0001YH10C for customer ID A1CY6CQC5HPQGL

```
Collaborative Recommender Result for customer A1CY6CQC5HPQGL:
1. B00LPDJSV8, Maytag 22002315 Washer Drive Pulley Genuine Original Equipment Manufacturer (OEM) part for Maytag &amp;amp; Amana
2. B002YTOCP4, Whirlpool Part Number 8181673: Gasket, Tub
3. B00IMO2Y1Y, Whirlpool 8181887 Handle for Washing Machine
4. B00L4JD7S8, LG Electronics 5005JJ2014A Refrigerator Door Shelf/Bin, White with Clear Trim
5. B00AFSMIP2, Supco MP21MA 2100W Cooktop Stove Surface Element Replacement for 0E00801799, Y04100166, AP4283501
6. B0053F8Y0A, 2206671B Whirlpool Refrigerator Grille Overflow (black)
7. B0156NCY8G, General Electric WR17X11264 Funnel Ice Display
8. B001DHNT7K, General Electric WR72X209 Drawer Slide Rail
9. B0156NDG4M, Whirlpool W10074200 Gasket Door
10. B00DZUA3DQ, Whirlpool 2206670W Grille
Hybrid Recommender Result for customer A1CY6CQC5HPQGL and product B0001YH10C:
1. B00JV8FUTI, LG LRBP1031T10.0 Cu. Ft. Titanium Counter Depth Bottom Freezer Refrigerator
2. B013PRRB4W, Electrolux EWMED70JIWWave-Touch 8.0 Cu. Ft. White Stackable With Steam Cycle Electric Dryer
3. B0050OZRLS, Maytag MFI2569YEW
4. B01B3Q6U6W, Maytag MDC4809PAB JetClean Plus 24&amp;quot; Black Portable Full Console Dishwasher - Energy Star
5. B00EE89JLU, LG WM3050CW4.0 Cu. Ft. White Stackable Front Load Washer - Energy Star
6. B00SZAH9Y2, LG PAIR SPECIAL- Turbo Series With Steam Technology WM3470HWA+DLEX3470W
7. B00HX3RS3O, LG DLEX3570W 7.4 Cu. Ft. Electric SteamDryer with NFC Tag On - White
8. B000UVWS0Y, Speed Queen ADEE8RGS 27&amp;quot; ADA Compliant Button Control Front Load Electric Dryer with 7.0 Cu. Ft. Capacity Reversible Door
9. B003S6HC9A, Power Pair Special-LG Turbo Series Ultra-Capacity Laundry System with Steam*PURE WHITE COLOR*(WM4270HWA_DLEX4270W)
10. B007RJZLX8, LG LSXS26386D 26.0 Cu. Ft. Black Stainless Steel Side-By-Side Refrigerator - Energy Star
```

System Integration

System Integration



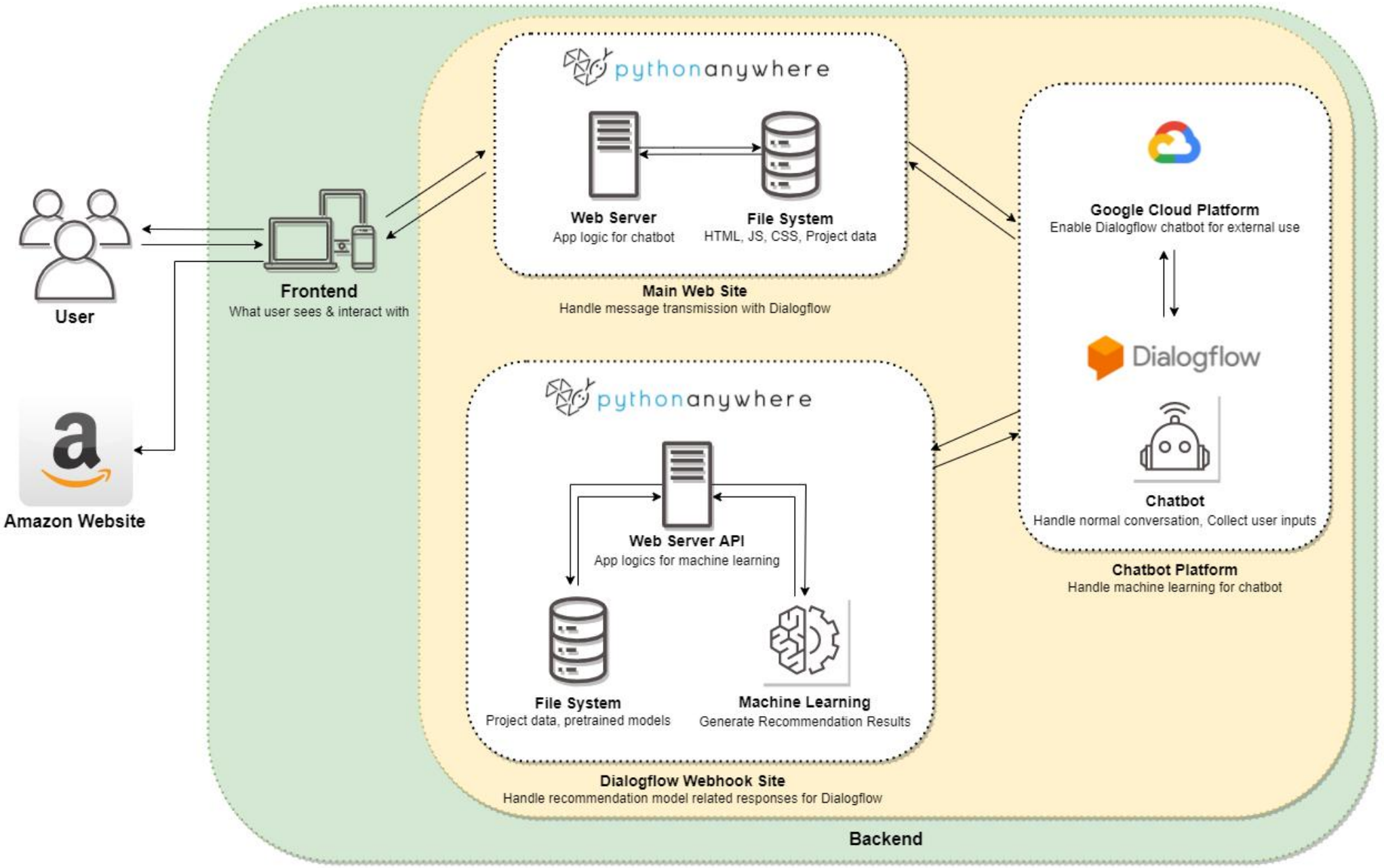
pythonanywhere

Dialogflow

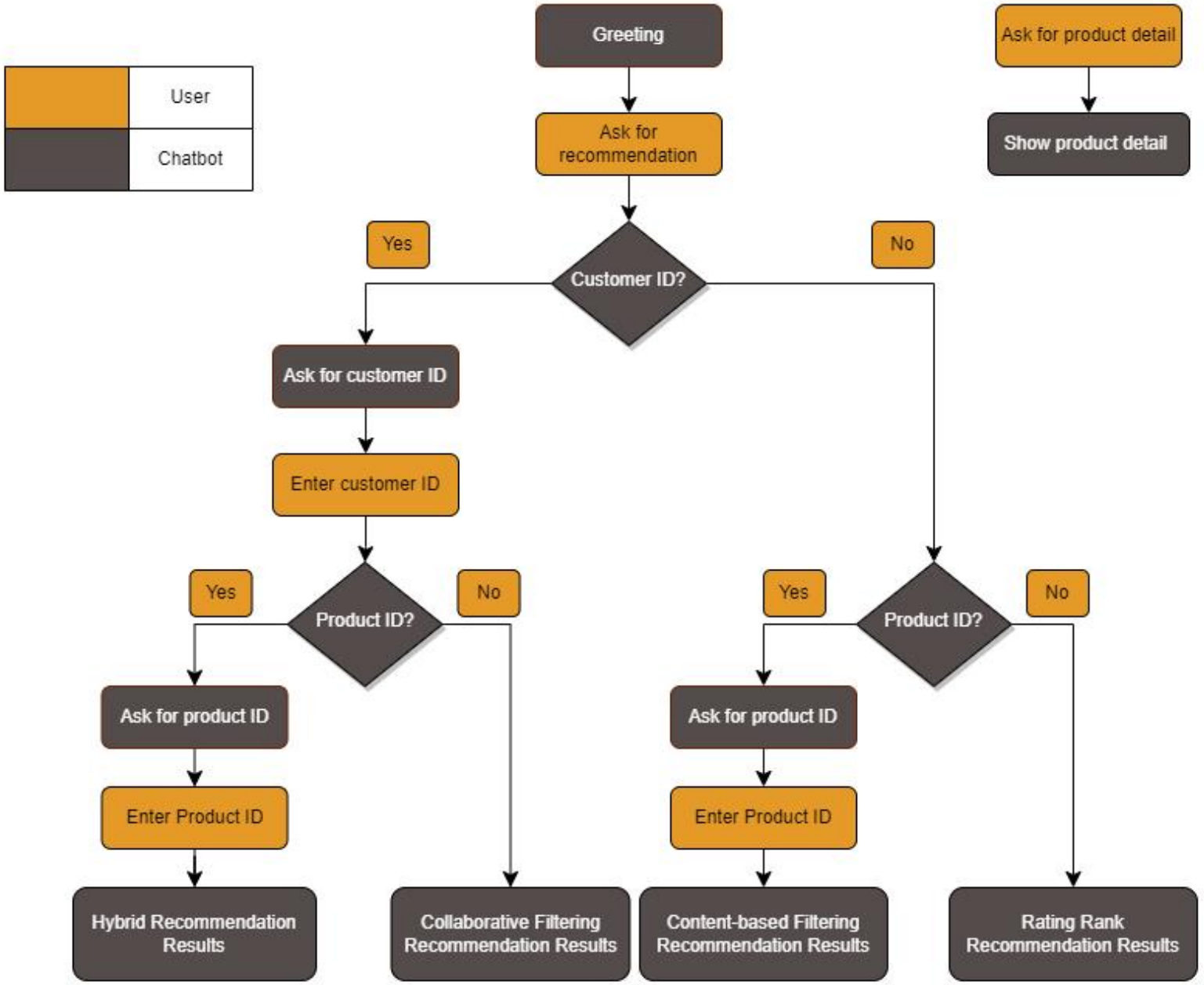
Develop a Flask website and use PythonAnywhere web hosting service to host our recommender system.

Develop a Chabot using DialogFlow platform to assist users for product recommendation and integrate it with the Flask website.

Integrated System Architecture

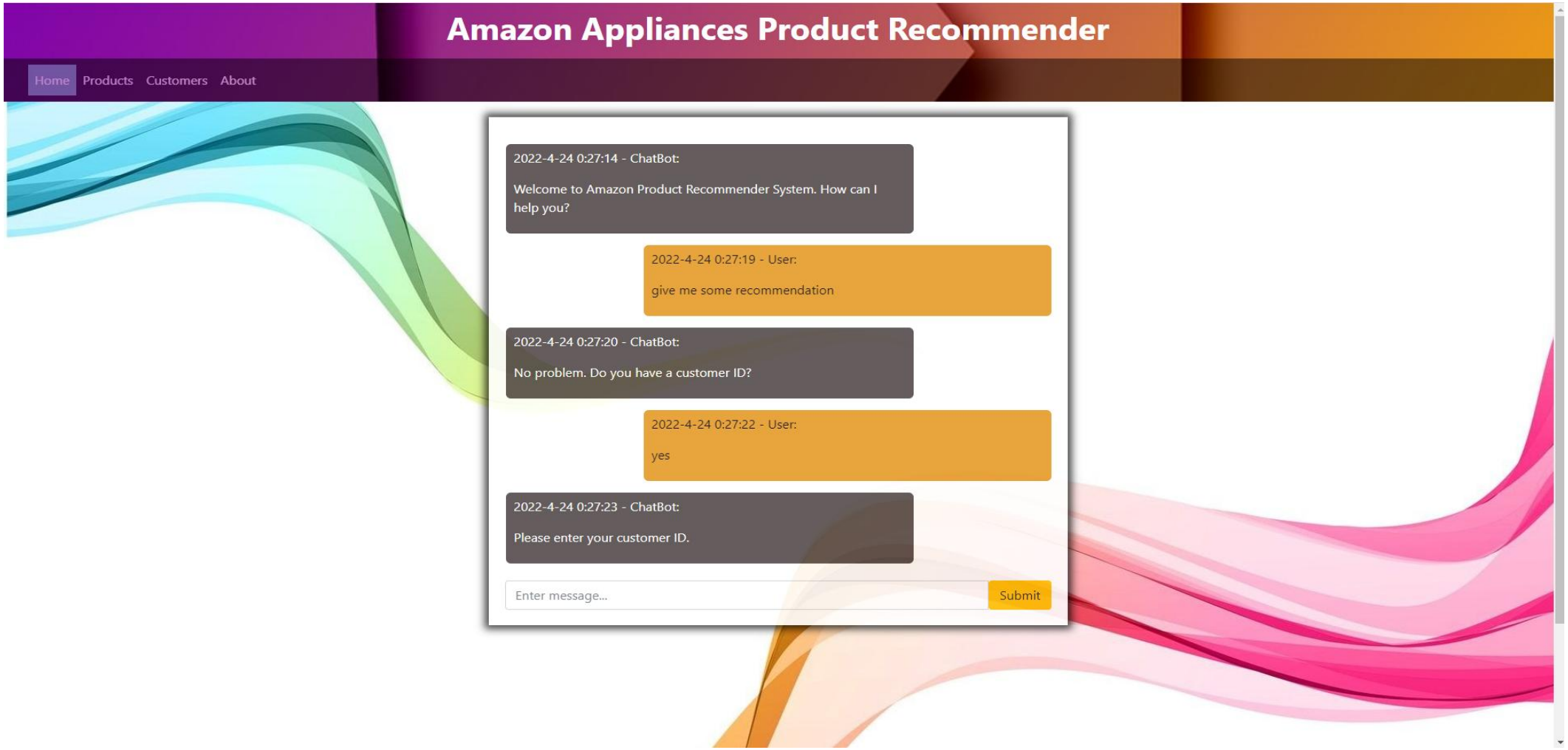


Dialogflow Chatbot Conversation Flow Chart



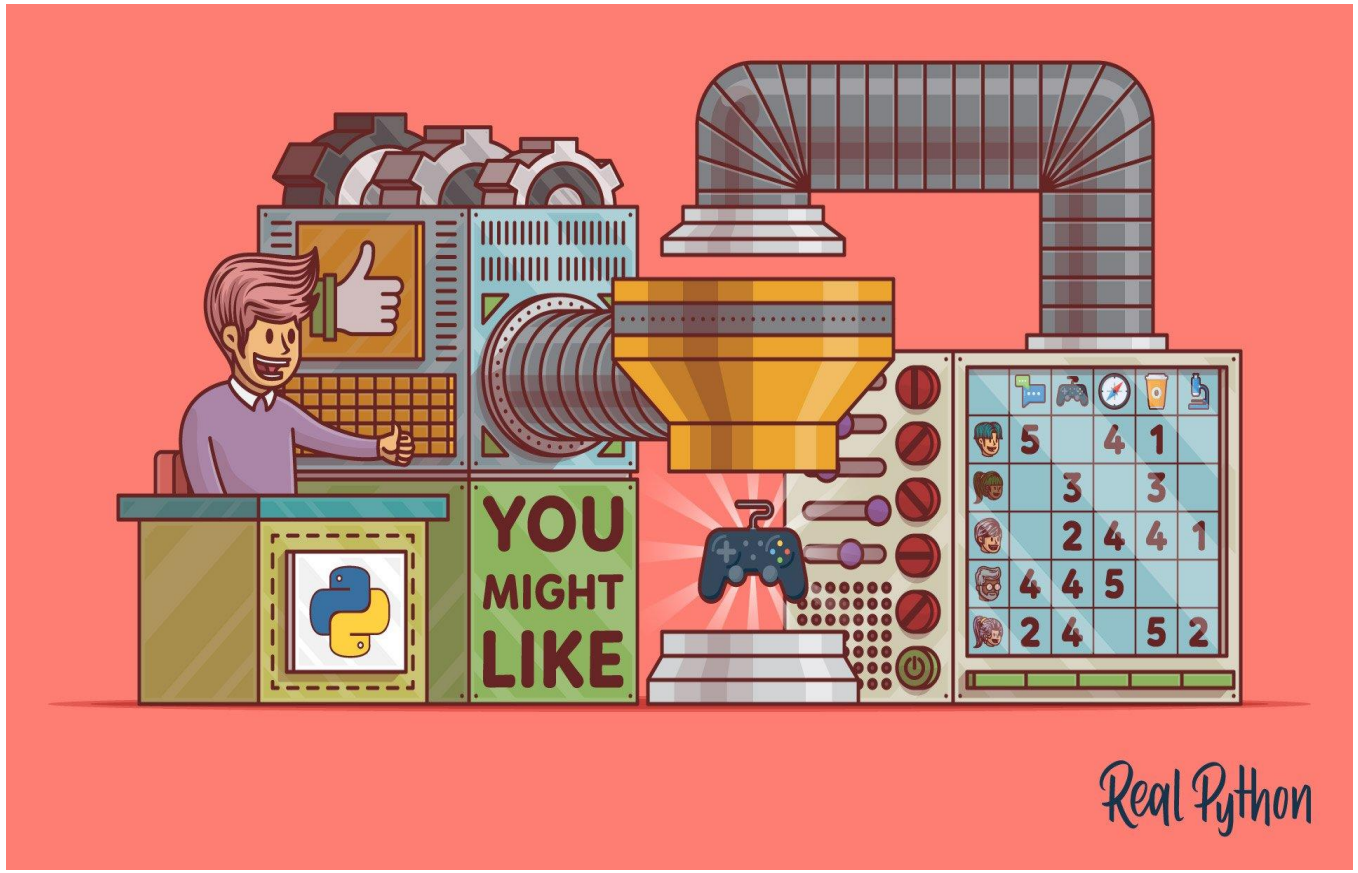
Live Recommender System

The recommendation chatbot website is hosted at: https://data606project.pythonanywhere.com



Conclusion

Outcomes



- Developed a product recommender system that can accurately predict customers' preferences
- There is no optimal recommendation algorithm/method
- The most useful characteristics to promote certain products to customers are product description and review rating/text
- Textual data plays a significant role in recommender systems
- Provide a website and Chatbot to assist amazon users to make purchase decisions.

Limitation

The data used for this project is a subset of the original data

The deployed models are not optimal

Offline Recommender System

Future Research

- Cross-domain recommender system
- Online Recommender Methods

References

- Jianmo Ni, Jiacheng Li, Julian McAuley Empirical Methods in Natural Language Processing (EMNLP), 2019 http://cseweb.ucsd.edu/~jmcauley/pdfs/emnlp19a.pdf
- Doshi, S. (2019, February 20). Brief on Recommender Systems. Medium. Retrieved February 13, 2022, from https://towardsdatascience.com/brief-on-recommender-systems-b86a1068a4dd
- Engineering@ZenOfAI. (2019, August 7). Creating chatbot with Webhooks using python (FLASK) and dialogflow. Medium. Retrieved March 5, 2022, from https://medium.com/zenofai/creating-chatbot-using-python-flask-d6947d8ef805
- BANIK, R. O. U. N. A. K. (2018). Hands-on recommendation systems with Python: Start building powerful and personalized, ... recommendation engines with python. PACKT Publishing Limited.
- Kapadia, S. (2020, December 29). Topic modeling in Python: Latent dirichlet allocation (LDA). Medium. Retrieved April 24, 2022, from https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0
- Tanner, G. (n.d.). Building a book recommendation system using Keras. Gilbert Tanner. Retrieved April 24, 2022, from https://gilberttanner.com/blog/building-a-book-recommendation-system-usingkeras


Thanks