Report

I have chosen the product ‘Video Games’. Below is the number of rows containing the product.

Total number of rows for the product.

893

—----------

The Descriptive Statistics of the product is as follows:

Number of Reviews: 11057
Average Rating Score: 4.2253
Number of Unique Products: 228
Number of Good Ratings: 9705
Number of Bad Ratings: 1352
Number of Reviews corresponding to each Rating: {5.0: 6982, 4.0: 1785, 3.0: 938, 1.0: 849, 2.0: 503}
Sample of preprocessed text: i bought hoping play dreamcast without using rf connector came well cable even work

—----------

Extracting relevant statistics:

A.
Top 20 most reviewed brands:
1. blank " : 64
2. Gunnar Optiks : 22
3. Sony : 19
4. dreamGEAR : 17
5. Generic : 16
6. Masione : 16
7. Activision : 12
8. Mad Catz : 12
9. Microsoft : 11
10. KMD : 11
11. CTA Digital : 11
12. : 10
13. Nyko : 9
14. TekNmotion : 9
15. : 8

16. Canopy : 7
17. HDE : 7
18. Hyperkin : 7
19. INSTEN : 7
20. Importer520 : 7

B.
Top 20 least reviewed brands:
1. Beats by dre solo : 1
2. Razer : 1
3. Codemaster : 1
4. Bestipik : 1
5. G-Dreamer : 1
6. Tomee : 1
7. Oblanc : 1
8. Traveler's Choice : 1
9. Disney Infinity : 1
10. Performanced Designed Products LLC : 1
11. LASUS : 1
12. Assecure : 1
13. Accessory Genie : 1
14. Rapoo : 1
15. Lg : 1
16. Andoer : 1
17. Atlantic : 1
18. CWC-GROUP : 1
19. PWND Gear : 1
20. Xett Multimedia : 1

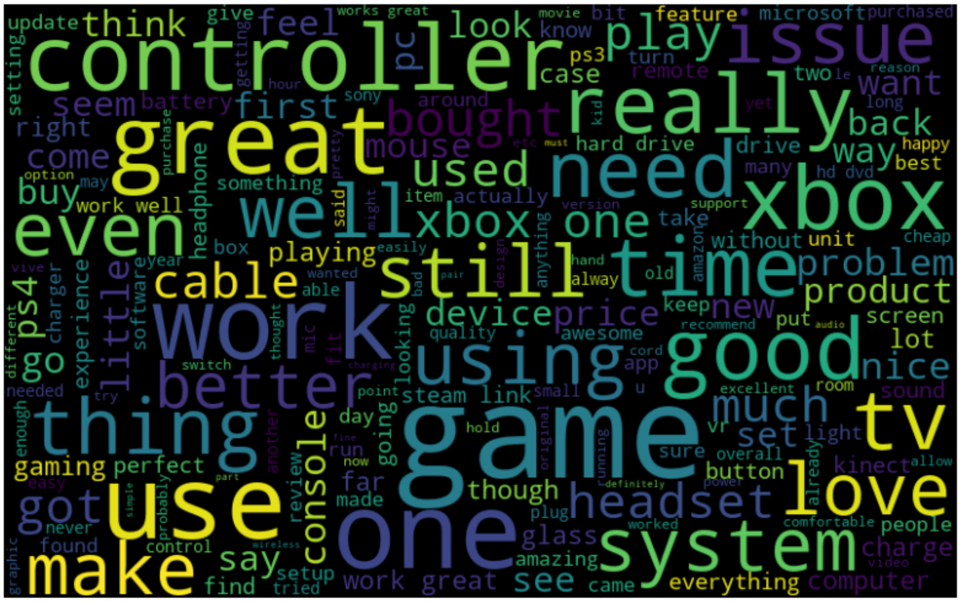
C.
Most positively reviewed Video Games: ('sunvalleytek', 3.2)

D.
Count of ratings over 5 consecutive years:
Year Count
2013 1703
2014 2074
2015 2332
2016 1981
2017 1373

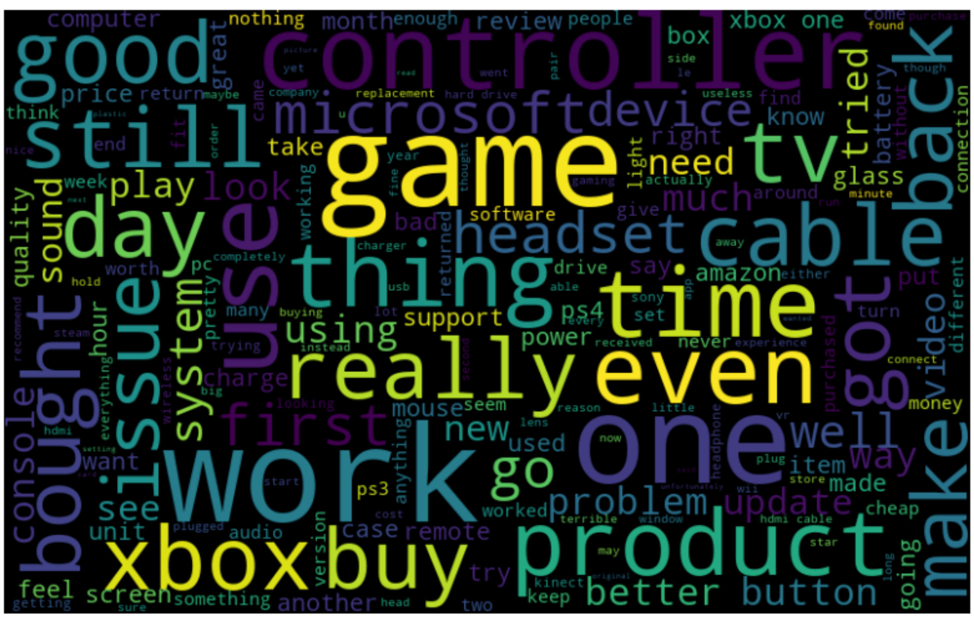
E.

```
By observing the word cloud we can say that the most common good and bad
word appears to be the same i.e. 'game' which can be seen largest in both
the word clouds although 'work' also appears equally large in bad words.
```

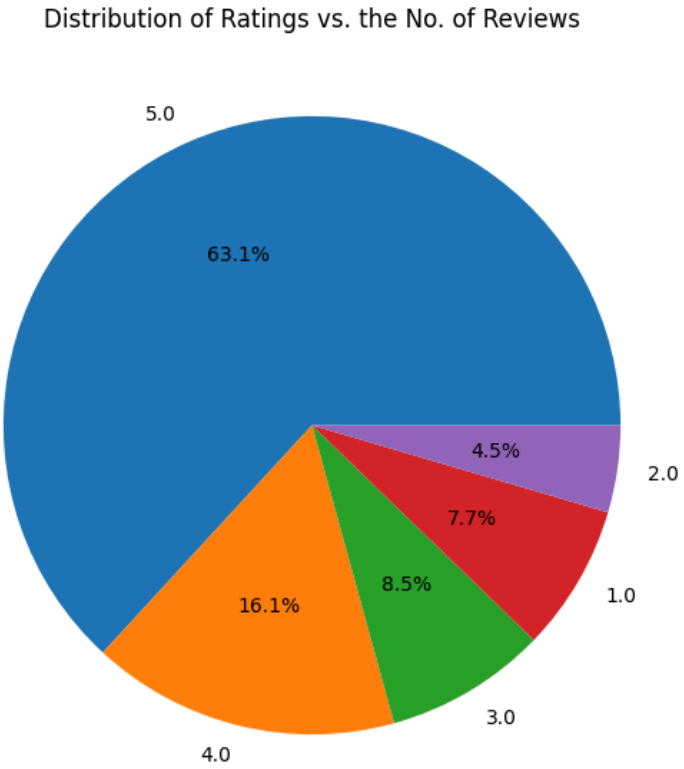
Word Cloud for Good Words:



Word Cloud for Bad Words:



F.



G.

Year the product got maximum reviews: 2015

H.

Year with the highest number of Customers: 2015

The Rating Class is divided into three categories:

- > 3 as Good
- =3 as Average
- <3 as Bad.

From the dataset, the Review Text was taken as input feature and Rating Class as target variable. The data was divided into Train and Test Data in the ratio of 75:25.

> Comparing the performance of 5 Machine Learning based models on the basis of Precision, Recall, F-1 Score and Support for each of the 3 target classes distinctly.

I chose the following 5 Machine Learning models:

1. Decision Tree
2. Logistic Regression
3. Random Forest
4. Multinomial Naive Bayes
5. K-Nearest Neighbors

Below are the results for each of the 5 models:

1. Decision Tree

	precision	***recall***	***f1-score***	***support***
Average	***0.15***	***0.15***	***0.15***	***231***
Bad	***0.44***	***0.42***	***0.43***	***336***
Good	***0.87***	***0.88***	***0.87***	***2198***
accuracy			***0.76***	***2765***
macro avg	***0.49***	***0.48***	***0.48***	***2765***
weighted avg	***0.76***	***0.76***	***0.76***	***2765***

2. Logistic Regression

	precision	***recall***	***f1-score***	***support***
Average	***0.28***	***0.17***	***0.21***	***231***
Bad	***0.61***	***0.50***	***0.55***	***336***
Good	***0.88***	***0.94***	***0.91***	***2198***
accuracy			***0.82***	***2765***
macro avg	***0.59***	***0.54***	***0.56***	***2765***
weighted avg	***0.80***	***0.82***	***0.81***	***2765***

3. Random Forest

	precision	*recall*	*f1-score*	*support*
Average	*0.27*	*0.01*	*0.02*	*231*
Bad	*0.70*	*0.26*	*0.38*	*336*
Good	*0.83*	*0.99*	*0.90*	*2198*
accuracy			*0.82*	*2765*
macro avg	*0.60*	*0.42*	*0.44*	*2765*
weighted avg	*0.76*	*0.82*	*0.76*	*2765*

4. Multinomial Naive Bayes

	precision	*recall*	*f1-score*	*support*
Average	*0.24*	*0.04*	*0.07*	*231*
Bad	*0.62*	*0.32*	*0.43*	*336*
Good	*0.84*	*0.97*	*0.90*	*2198*
accuracy			*0.81*	*2765*
macro avg	*0.56*	*0.44*	*0.46*	*2765*
weighted avg	*0.76*	*0.81*	*0.77*	*2765*

5. K-Nearest Neighbors

	precision	*recall*	*f1-score*	*support*
Average	*0.19*	*0.03*	*0.06*	*231*
Bad	*0.52*	*0.07*	*0.13*	*336*
Good	*0.81*	*0.98*	*0.89*	*2198*
accuracy			*0.79*	*2765*
macro avg	*0.50*	*0.36*	*0.36*	*2765*
weighted avg	*0.72*	*0.79*	*0.72*	*2765*

On observing all the values of the 5 chosen models the best performances are given by Random Forest and Logistic Regression.

The MAE (Mean Absolute Error) for taking K = 10, 20, 30, 40, 50 similar users.

1. For user-user recommender

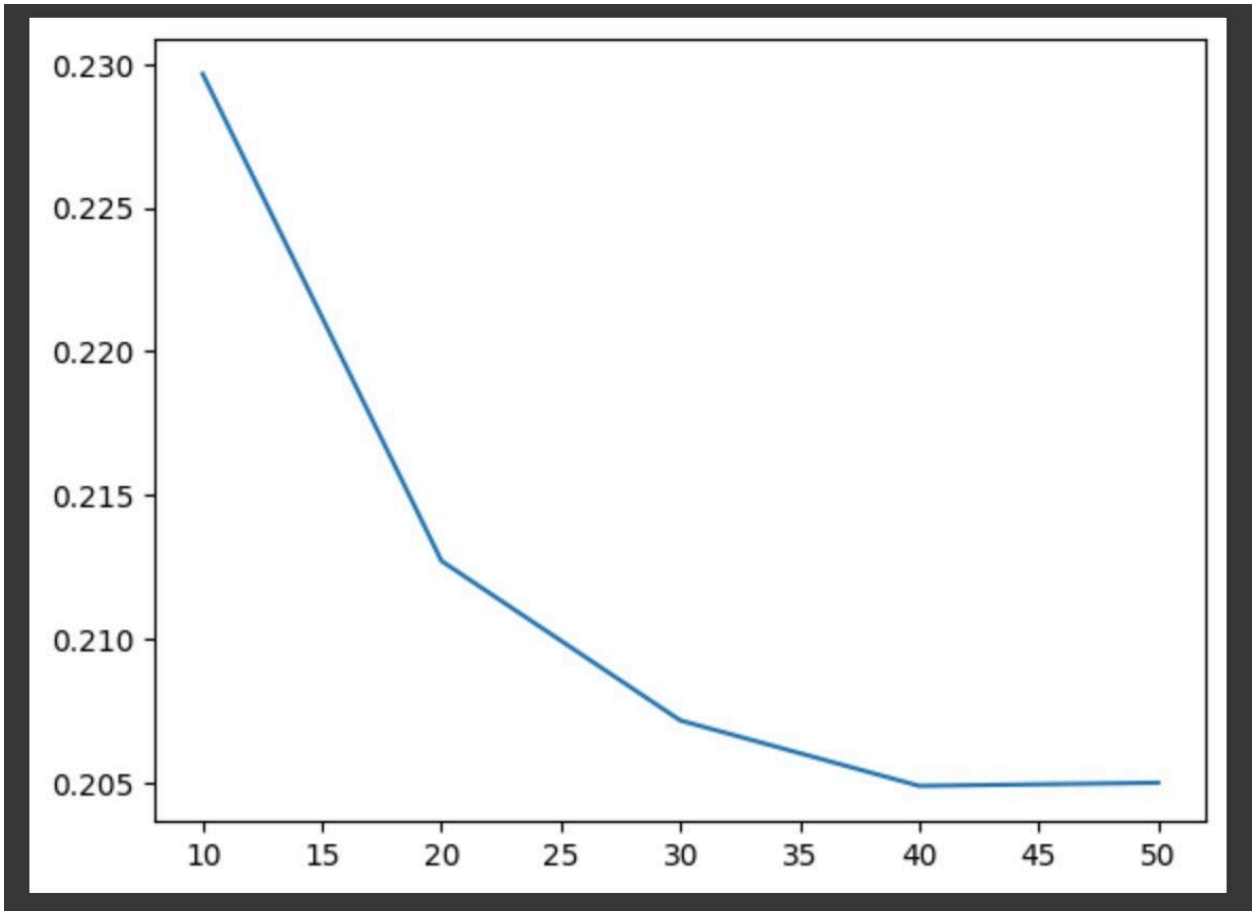
MAE for 50 = 0.2049
MAE for 40 = 0.2048
MAE for 30 = 0.2071
MAE for 20 = 0.2127
MAE for 10 = 0.2296

2. For item-item recommender

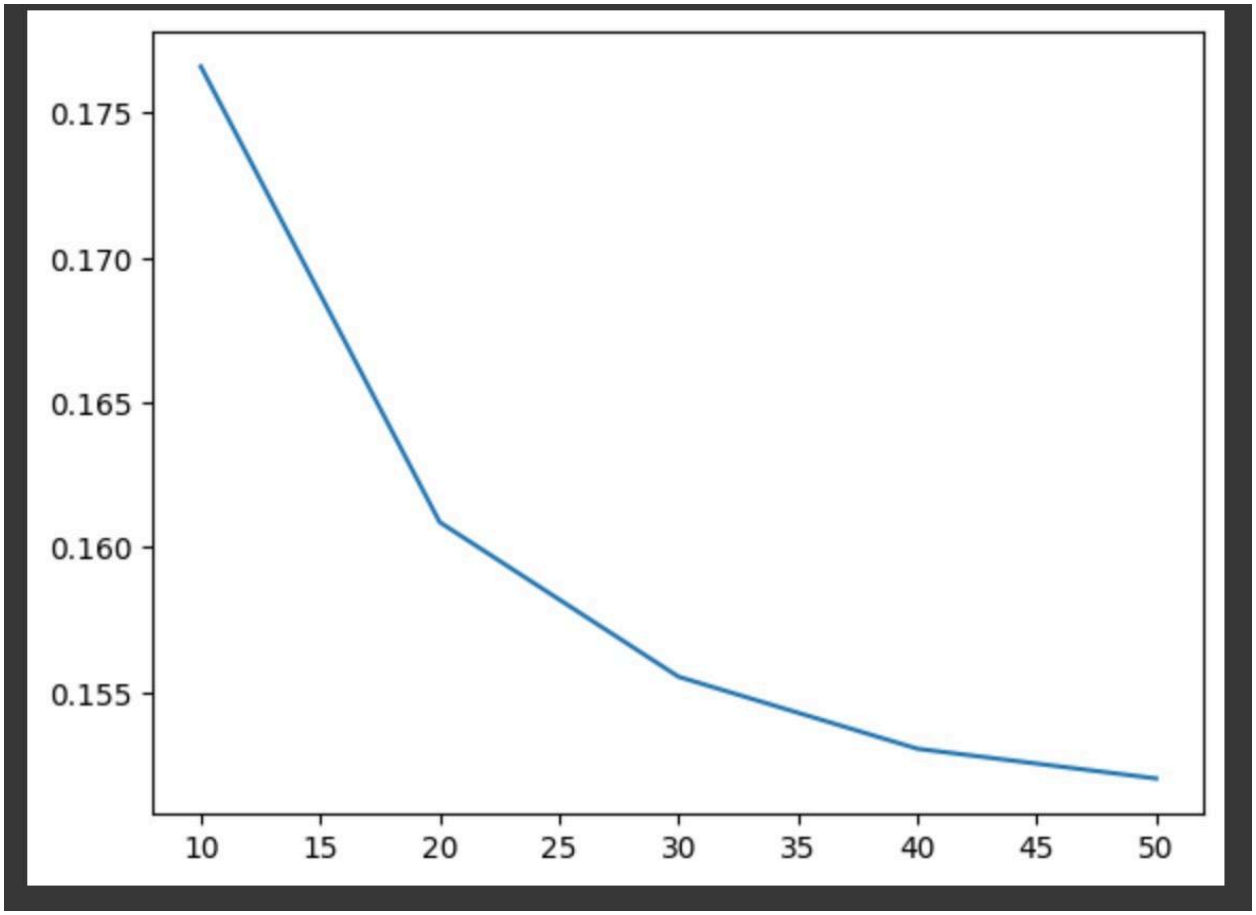
MAE for 50 = 0.1520
MAE for 40 = 0.1530
MAE for 30 = 0.1555
MAE for 20 = 0. 1608
MAE for 10 = 0.1766

MAE Graphs

1. For user-user recommender



2. For item-item recommender



The TOP 10 products by User Sum Ratings.

title

Cocoweb PortBlock Dual-Function Door Security Bar	***571.000000***
Country Heaven (Dare River)	***374.466667***
Mighty Bright XtraFlex Book Light, Silver	***280.000000***
Nightfall	***212.000000***
Sting/On A Winters Night (0602527171494)	***200.000000***
My MacBook	***174.000000***
Iclicker+ Student Remote	***164.333333***
Vintage Camera Photo Album	***159.000000***
Indelible : A Novel	***127.500000***
Nikon D90 Inbrief Laminated Reference Card	***109.000000***

Name: overall, dtype: float64