

IR

Assignment 3

Report

- Aditya Nangia
- 2020168

Part 4

```
Number of reviews: 264868
Average Rating Score: 4.07106936285244
Number of Unique Products: 5393
Number of Good Ratings: 225823
Number of Bad Ratings: 39045
Number of Reviews corresponding to each Rating
overall
5.0    145081
4.0    52937
3.0    27805
1.0    20362
2.0    18683
Name: count, dtype: int64
```

Part 6

Top 20 brands

```
Top 20 Most Reviewed Brands:
brand
Sony                23063
Sennheiser          14497
Bose                 9879
Audio-Technica      6521
Bluedio              5301
Panasonic            4961
Mpow                 4907
Philips              4738
Etre Jeune           4682
Symphonized          4284
MEE audio            4087
TaoTronics           4082
JVC                  4008
Koss                 3864
Plantronics          3602
Toysdone             3506
JLAB                 3164
SoundPEATS           3052
EldHus               2724
Kidz Gear            2511
Name: count, dtype: int64
```

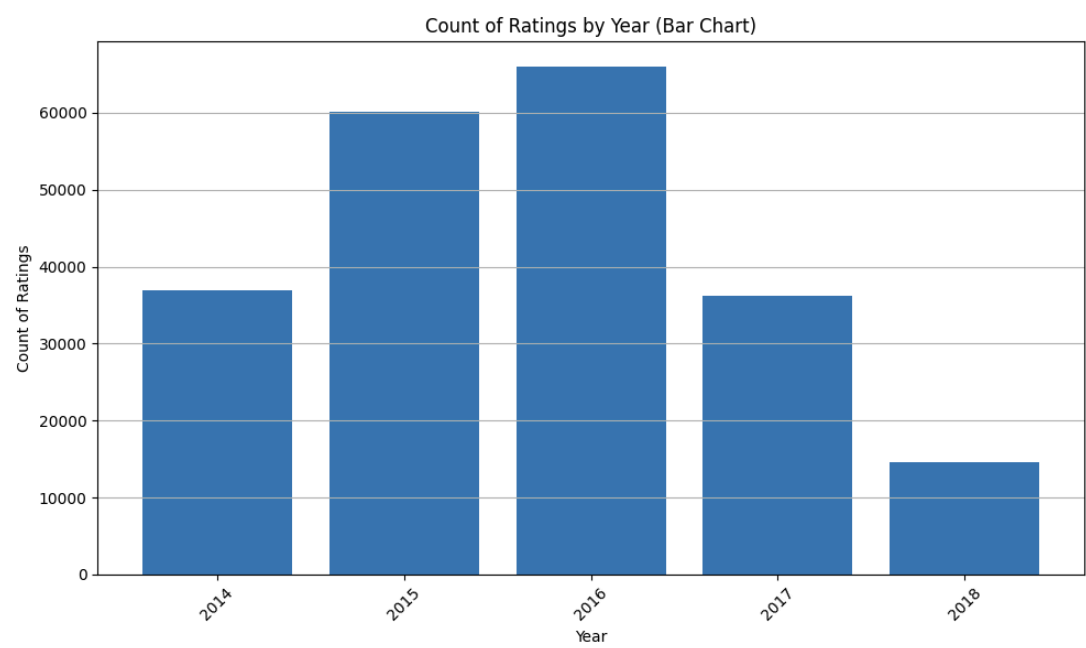
Least 20 Brands

Top 20 Least Reviewed Brands:	
brand	
Honda	2
AIRDRIVES	3
NOIZY Brands	3
PopClik	4
California Cable Market	4
EUBUY	5
Fuji	5
Alitutumao	5
Blue House	5
Kaisi	5
lexastech	5
meda one	5
Innovage	5
Jiaohong Quan	5
Daioolor	5
Rockford Fosgate	5
Central.Tech.Gear	5
Audiology	5
SpongeBob SquarePants	5
Giga Goods	5
Name: count, dtype: int64	

Most Positively Reviewed Brand

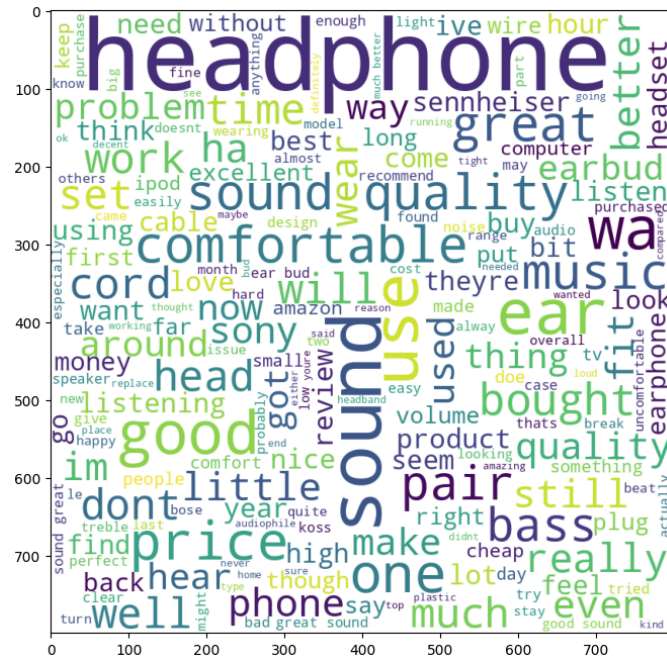
'ARCTIC'

Count of Ratings



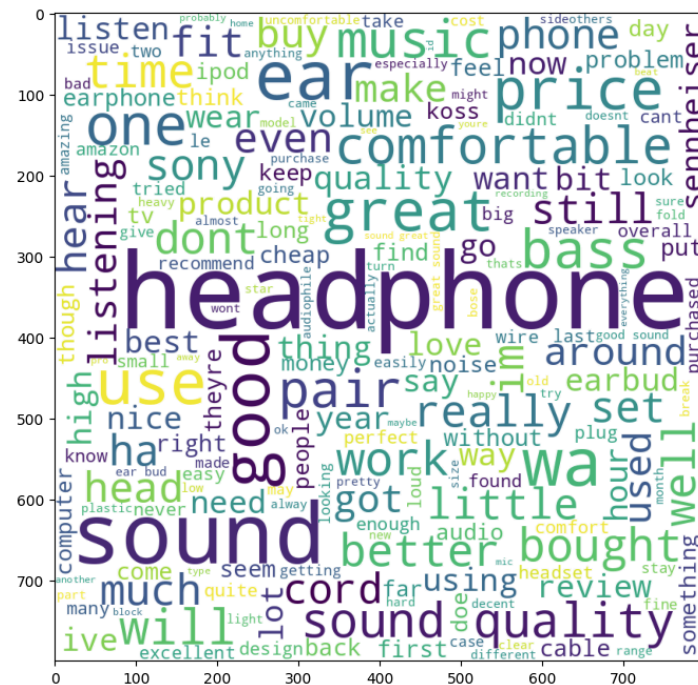
Word Cloud

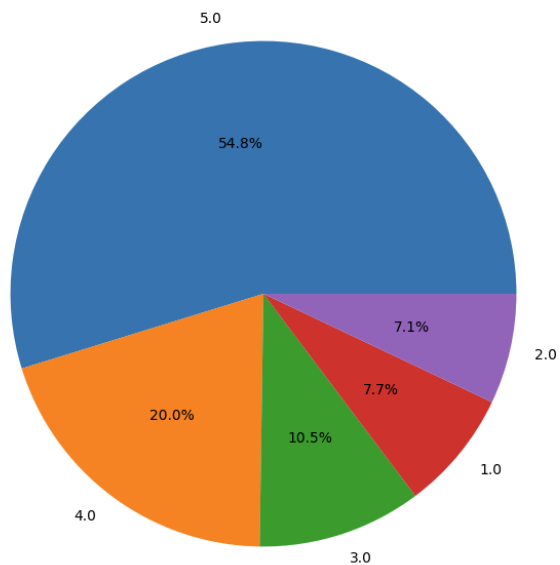
Good



Most commonly used words: Headphones, use, well, good, comfortable, easy, price, love, review, great, music

Bad





Max Reviews

```
Year with Maximum Reviews: 2016  
Number of Reviews: 65933
```

Max Customers

```
Year with Maximum Customers: 2016  
Number of Customers: 53583
```

Part 10

```
=====
Decision Tree
=====
              precision    recall  f1-score   support

     0         0.50         0.52         0.51         9487
     1         0.24         0.28         0.26         6051
     2         0.88         0.86         0.87        50679

 accuracy         0.76         0.76         0.76        66217
 macro avg         0.54         0.55         0.54        66217
weighted avg         0.77         0.76         0.76        66217

=====
```

```
=====
Ada Boost
=====
              precision    recall  f1-score   support

     0         0.41         0.66         0.51         6079
     1         0.07         0.43         0.12         1166
     2         0.97         0.81         0.89        58972

 accuracy         0.79         0.79         0.79        66217
 macro avg         0.48         0.63         0.51        66217
weighted avg         0.90         0.79         0.84        66217

=====
```

```

=====
Multinomial Naive Bayes
=====

```

	precision	recall	f1-score	support
0	0.61	0.64	0.63	9219
1	0.13	0.39	0.19	2239
2	0.95	0.86	0.90	54759
accuracy			0.81	66217
macro avg	0.56	0.63	0.57	66217
weighted avg	0.87	0.81	0.84	66217

```

=====

```

On 2.5% data

```

=====
Gaussian Naive Bayes
=====

```

	precision	recall	f1-score	support
0	0.19	0.23	0.21	194
1	0.46	0.12	0.19	658
2	0.50	0.78	0.61	804
accuracy			0.45	1656
macro avg	0.38	0.38	0.34	1656
weighted avg	0.45	0.45	0.40	1656

```

=====

```

```

=====
K Nearest Neighbours
=====

```

	precision	recall	f1-score	support
0	0.11	0.37	0.17	70
1	0.08	0.23	0.11	56
2	0.95	0.78	0.86	1530
accuracy			0.74	1656
macro avg	0.38	0.46	0.38	1656
weighted avg	0.89	0.74	0.80	1656

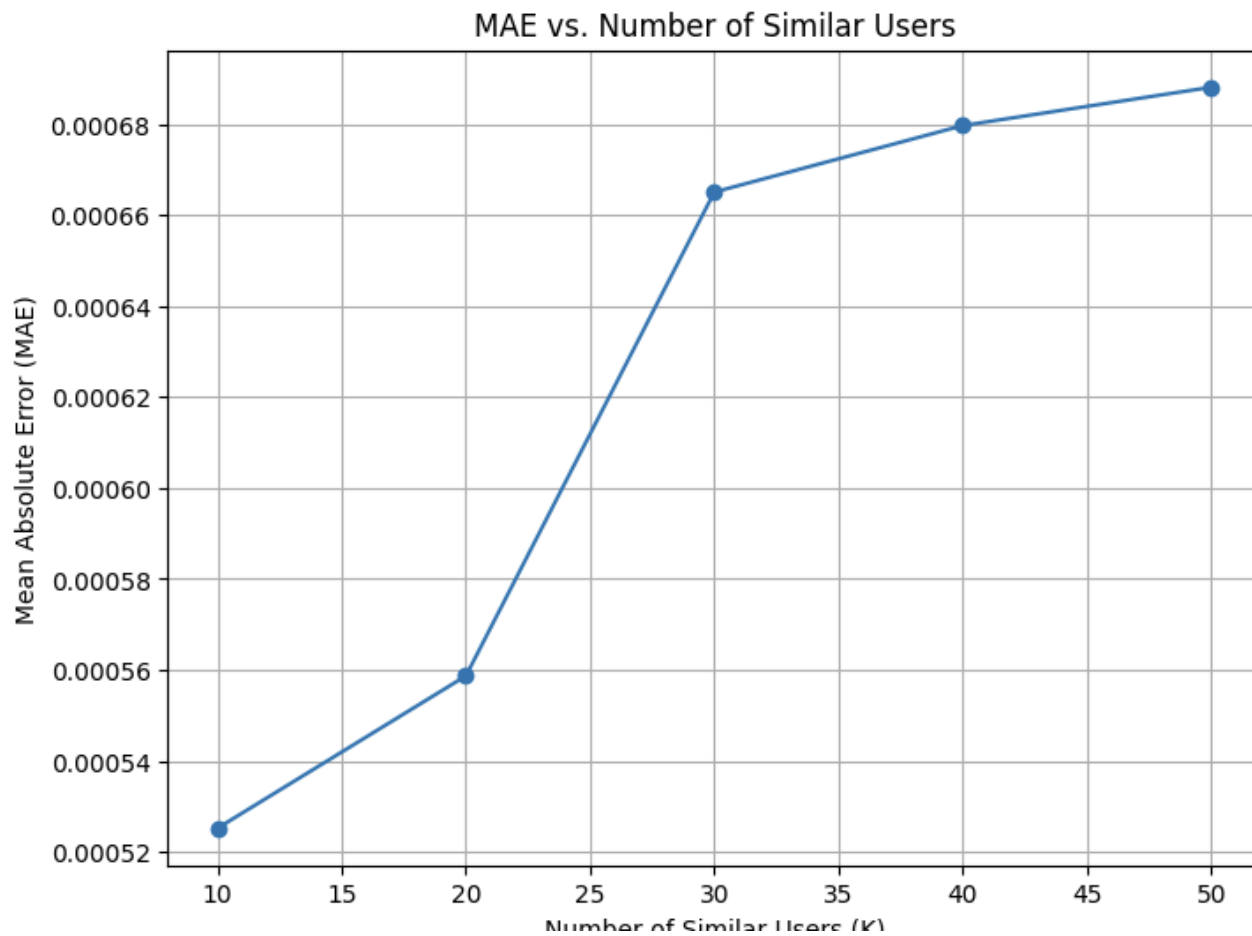
```

=====

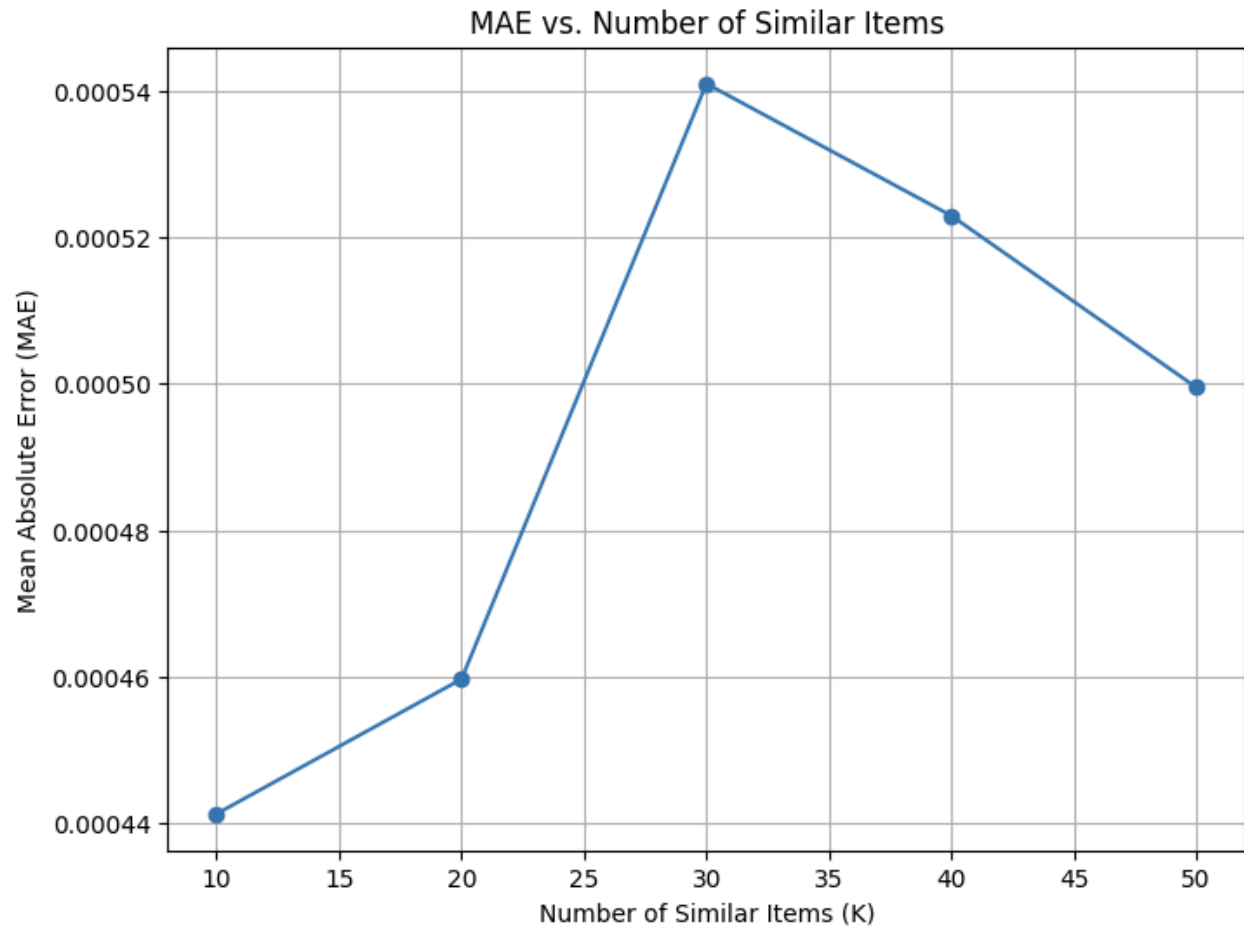
```

Part 11

```
MAE for K=10 similar users: 0.0005251875663042019
MAE for K=20 similar users: 0.0005586874219692478
MAE for K=30 similar users: 0.0006650833194292905
MAE for K=40 similar users: 0.0006797087000827435
MAE for K=50 similar users: 0.0006881384605817375
```



```
<ipython-input-24-f0281d31f29e>:7: RuntimeWarning: invalid value encountered in divide
similarity_score = dot_product / (norm_array1 * norm_array2)
MAE for K=10 similar items: 0.00044123782980138475
<ipython-input-24-f0281d31f29e>:7: RuntimeWarning: invalid value encountered in divide
similarity_score = dot_product / (norm_array1 * norm_array2)
MAE for K=20 similar items: 0.0004596435542870266
<ipython-input-24-f0281d31f29e>:7: RuntimeWarning: invalid value encountered in divide
similarity_score = dot_product / (norm_array1 * norm_array2)
MAE for K=30 similar items: 0.0005409946867921376
<ipython-input-24-f0281d31f29e>:7: RuntimeWarning: invalid value encountered in divide
similarity_score = dot_product / (norm_array1 * norm_array2)
MAE for K=40 similar items: 0.0005229564542129131
<ipython-input-24-f0281d31f29e>:7: RuntimeWarning: invalid value encountered in divide
similarity_score = dot_product / (norm_array1 * norm_array2)
MAE for K=50 similar items: 0.0004994999202198651
```



PART 12

Top 10 products sorted by user sum Ratings:

asin

B004WODP20 13295.0

B00BN0N0LW 13242.0

B00LP6CFEC 10454.0

B00JJ2C0S0 9775.0

B00004T8R2 7392.0

B003LPTAYI 7313.0

B008EPW1MI 6684.0

B00NBEWB4U 6425.0

B000ULAP4U 6179.0

B0007NWL70 5471.0

Name: overall, dtype: float64