

K. L. E. SOCIETY'S
**B. V. BHOOMARADDI COLLEGE OF ENGINEERING &
TECHNOLOGY, HUBLI – 580031**

(An Autonomous Institution)



DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING
2016 – 2017

BDA Report on
Analysis of Amazon Food Reviews

Under the guidance of
Ms. Sunita Hiremath

Submitted by

Mr. V T Vishwanath	2BV14IS114
Mr. Raghavendra P	2BV14IS131
Mr. Vijay Kumar	2BV14IS118
Mr. Surendar Singh	2BV14IS112

TABLE OF CONTENTS

1. About	3
2. Motivation	3
3. Problem Statement	3
4. Objectives	3
5. Description of the data	3
6. Dataset	4
7. Data Statistics	4
8. Implementation and analysis	5

ABOUT

The dataset is taken from www.kaggle.org. The dataset is provided by the Amazon Company and all the products are food items. The dataset contains the details of the reviews about the products over time. The dataset contains ProductId for which review details are given; and the same product can have reviews over time.

MOTIVATION

Amazon fine food dataset contains the review of the food items sold by amazon. The buyers review food items. It is very important as it will indicate the user's reaction on each food item. User will give rating and write the summary and detailed review. This review on food item is important as it will indicate the food item is good or bad. From packaging to quality, quantity everything is important. So indirectly these reviews will increase/decrease the marketing shares and number of food items sold on amazon. Now a days text mining is trending and it will indicate the user's reaction by giving it as positive, negative or neutral. This motivates us to analyze the each review and give the sentimental result. And to do the time series analysis which will shows the current rating of particular product. This will help amazon to sell that particular product and this will also increase the user's experience.

PROBLEM STATEMENT

To analyse the review for each food item and provide the status of it in the market over time.

OBJECTIVES

- Providing average sentiment for each product based on the text present in each review.
- Taking most helpful review from each product.
- Providing the details of the scores for each food item (Saying how many 1 star, 2 star reviews till 5 stars).
- Providing the time series analysis for the product.

DESCRIPTION OF THE DATA

- The Amazon Fine Food Reviews dataset consists of 568,454 food reviews Amazon users left up to October 2012.
- Volume of the data – 289MB
- Velocity – Amazon review dataset consists of real time reviews regarding food items by the users.
- Data must be categorized based on the rating provided. Thus the retrieval of the data becomes faster as the data becomes organized.

- Each day more than 5000 food items will be sold and this number increases gradually with increase in company's performance.
- The columns in the dataset are:
- **Id** – column number
- **ProductId** - unique identifier for the product
- **UserId** - unique identifier for the user
- **Profile Name** – User Profile name
- **Helpfulness Numerator** - number of users who found the review helpful
- **Helpfulness Denominator** - number of users who indicated whether they found the review helpful
- **Score** - rating between 1 and 5
- **Time** - timestamp for the review
- **Summary** - brief summary of the review
- **Text** - text of the review

DATASET

A	B	C	D	E	F	G	H	I	J	K	L
Id	ProductId	UserId	ProfileName	HelpfulnessNur	HelpfulnessDe	Score	Time	Summary	Text		
1	B001E4KFG0	A35GXH7AUHU8GV	delmartian	1	1	5	1303862400	Good Quality Dog	I have bought several of the Vitality canned dog food products and have found		
2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertise	Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually s		
3	B000LQOCHO	ABXLMWJIXXAIN	Natalia Corres	1	1	4	1219017600	"Delight" says it a	This is a confection that has been around a few centuries. It is a light, pillowy		
4	B000UA0QIQ	A395BORCF6GVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient in Robitussin I believe I have found		
5	B006K2Z27K	A1UQRSCLF8GW1T	Michael D. Bigl	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wide assortment of yummy taffy. Del		
6	B006K2Z27K	ADT0SRK1MGOEU	Twoapennythi	0	0	4	1342051200	Nice Taffy	I got a wild hair for taffy and ordered this five pound bag. The taffy was all ver		
7	B006K2Z27K	A1SP2KVFXXRUI	David C. Sulliv	0	0	5	1340150400	Great!	Just as got This saltwater taffy had great flavors and was very soft and chewy. Each candy		
8	B006K2Z27K	A3JRGQVEQN31IQ	Pamela G. Will	0	0	5	1336003200	Wonderful, tasty	This taffy is so good. It is very soft and chewy. The flavors are amazing. I wou		
9	B000E7L2R4	A1MZY09TZK08BI	R. James	1	1	5	1322006400	Yay Barley	Right now I'm mostly just sprouting this so my cats can eat the grass. They love		
10	B00171APVA	A21BT40VZCCYT4	Carol A. Reed	0	0	5	1351209600	Healthy Dog Foo	This is a very healthy dog food. Good for their digestion. Also good for small p		
11	B0001PB9FE	A3HDKO70W0QNK	Canadian Fan	1	1	5	1107820800	The Best Hot Sau	I don't know if it's the cactus or the tequila or just the unique combination of i		
12	B0009XLVG0	A2725IB4Y9JEB	A Poeng "Spar	4	4	5	1282867200	My cats LOVE this	One of my boys needed to lose some weight and the other didn't. I put this fo		
13	B0009XLVG0	A327PCT23YH90	LT	1	1	1	1339545600	My Cats Are Not f	My cats have been happily eating Felidae Platinum for more than two years. I		
14	B001GVISJM	A18ECVX2RJHUE	willie "roadie"	2	2	4	1288915200	fresh and greasy!	good flavor! these came securely packed... they were fresh and delicious! i lo		
15	B001GVISJM	A2MUGFV2TDQ47H	Lynrie "Oh HEL	4	5	5	1268352000	Strawberry Twizz	The Strawberry Twizzlers are my guilty pleasure - yummy. Six pounds will be a		
16	B001GVISJM	A1CZX3CP8IKQJ	Brian A. Lee	4	5	5	1262044800	Lots of twizzlers,	My daughter loves twizzlers and this shipment of six pounds really hit the spo		
17	B001GVISJM	A3KLWF6WQ5BNY	Erica Neathery	0	0	2	1348099200	poor taste	I love eating them and they are good for watching TV and looking at movies! It		
18	B001GVISJM	AFKW14U97Z6QO	Becca	0	0	5	1345075200	Love it!	I am very satisfied with my Twizzler purchase. I shared these with others and		
19	B001GVISJM	A2A9X58G2GTBLP	Wolfeel	0	0	5	1324598400	GREAT SWEET CA	Twizzlers, Strawberry my childhood favorite candy, made in Lancaster Pennsy		
20	B001GVISJM	A31V7CL2C13K2U	Greg	0	0	5	1318032000	Home delivered	Candy was delivered very fast and was purchased at a reasonable price. I was		
21	B001GVISJM	A1W0OKGLPR5PV6	mom2emma	0	0	5	1313452800	Always fresh	My husband is a Twizzlers addict. We've bought these many times from Amaz		
22	B001GVISJM	AZOF9E17RGZ8H	Tammy Anders	0	0	5	1308960000	TWIZZLERS	I bought these for my husband who is currently overseas. He loves these, and		
23	B001GVISJM	ARYVQL4N737A1	Charles Brown	0	0	5	1304899200	Delicious product	I can remember buying this candy as a kid and the quality hasn't dropped in all		
24	B001GVISJM	AJ613OLZ2UG7V	Mare's	0	0	5	1304467200	Twizzlers	I love this candy. After weight watchers I had to cut back but still have a cravir		
25	B001GVISJM	A22P2J09NJ9HKE	S. Cabanaugh "	0	0	5	1295481600	Please sell these	I have lived out of the US for over 7 yrs now, and I so miss my Twizzlers!! Whe		

DATASET STATISTICS

Number of reviews	568,454
Number of users	256,059
Number of products	74,258
Users with > 50 reviews	260
Median no. of words per review	56
Timespan	Oct 1999 - Oct 2012

IMPLEMENTATION AND ANALYSIS

Python script using panda's package is used for the implementation of the objectives

1. **Providing average sentiment for each product based on the text present in each review:** "vaderSentiment" library is used for the sentimental analysis of the comments by the user. Providing the time series analysis for the product based on the previous history. Sentiment provides the information of each product saying how good or bad the product is based on the comments given by the users. It signifies the company (Amazon) to improve the products which are not gaining appreciation from the users.

```
1
our prediction: 4
AVERAGE SENTIMENT FOR REVIEW: 0.2254
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.185
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.129
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.1504
our prediction: 4
AVERAGE SENTIMENT FOR REVIEW: 0.2096
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.1633
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.1097
our prediction: 5
AVERAGE SENTIMENT FOR REVIEW: 0.6115
our prediction: 4
AVERAGE SENTIMENT FOR REVIEW: 0.422
our prediction: 4
AVERAGE SENTIMENT FOR REVIEW: 0.4179
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: -0.1061
our prediction: 4
AVERAGE SENTIMENT FOR REVIEW: 0.2678
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.1242
our prediction: 4
AVERAGE SENTIMENT FOR REVIEW: 0.4649
our prediction: 5
AVERAGE SENTIMENT FOR REVIEW: 0.8519
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.0189
our prediction: 4
AVERAGE SENTIMENT FOR REVIEW: 0.3385
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.1146
our prediction: 3
AVERAGE SENTIMENT FOR REVIEW: 0.0793
3
Please select a product from the product list to analyze
['B001E4KFG0' 'B00813GRG4' 'B000LQOCH0' ..., 'B003S1WTCU' 'B004I613EE'
 'B001LR2CU2']
Press e to elaborate, c to continue: _
```

2. **Taking most helpful review from each product:** Most helpful reviews for a particular product is done by considering Helpfulness Numerator and Denominator. Helpfulness Numerator signifies that the review is most viewed by the user. Helpfulness Denominator signifies that the comment for that review is found good.

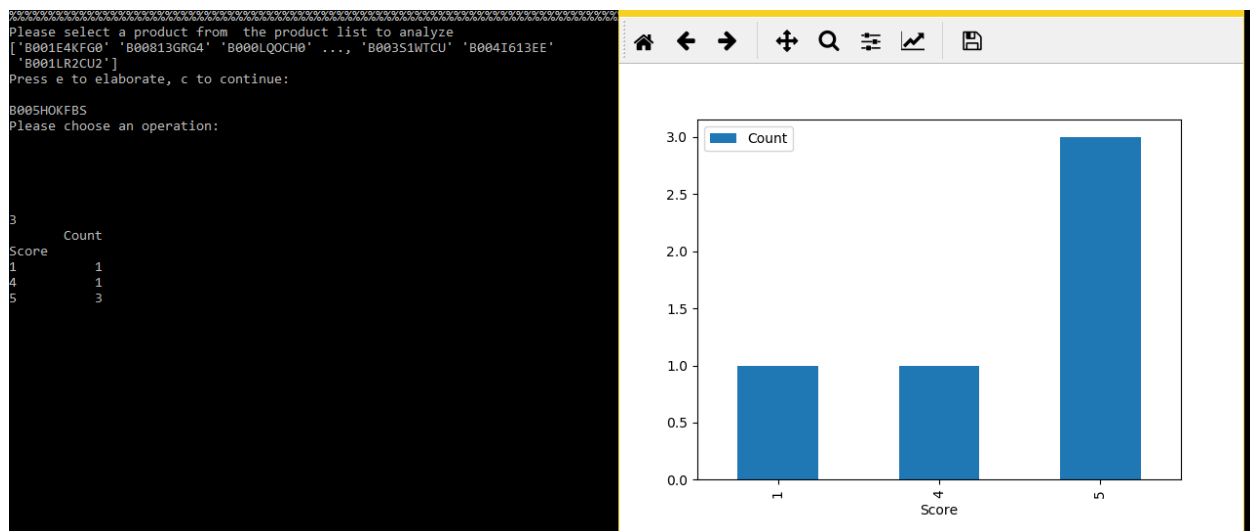
```
['B001E4KFG0' 'B00813GRG4' 'B000LQOCH0' ..., 'B003S1WTCU' 'B004I613EE'
'B001LR2CU2']
Press e to elaborate, c to continue:

B005HOKFBS
Please choose an operation:

1 -> Get overall sentiment for product,
2 -> get helpful review for product,
3 -> Get score plot
4 -> Get time series analysis of score for product
5 -> Quit

2
Fast shipping. One of our favorite gf crackers. They don't crumble easy or have a weird gluten free taste. I also use these to make bread crumbs.
~~~~~
I really don't understand how anyone could eat these crackers. They taste awful when you first put them in your mouth and the aftertaste is even worse. I have just star
ted a gluten free diet so I'm still used to regular crackers and I can tell you that these are no where near the real thing. And now having invested in this three pack,
I'm going to have to toss them all out. I'm usually pretty flexible about flavor and I really hate to throw out food, but these are so bad that the taste actually upse
t my stomach. I had to suck on a mint afterwards to get the flavor out of my mouth. A complete waste of time and money.
~~~~~
Please select a product from the product list to analyze
['B001E4KFG0' 'B00813GRG4' 'B000LQOCH0' ..., 'B003S1WTCU' 'B004I613EE'
'B001LR2CU2']
Press e to elaborate, c to continue:
```

3. **Providing the details of the scores for each food item (Saying how many 1 star, 2 star reviews till 5 stars):** This will show the reaction of users for a particular product.



4. **Providing the time series analysis for the product:** The time series analysis is provided for each product based on the previous year's records. This indicates how the product is seen by the people over time.

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Please select a product from the product list to analyze
['B001E4KFG0' 'B00813GRG4' 'B000LQOCH0' ..., 'B00351WTCU' 'B004I613EE'
 'B001LR2CU2']
Press e to elaborate, c to continue:

B005HOKFBS
Please choose an operation:

4

```

