

**INFO 7250 Engineering Big-Data
Systems
Final Project Report on
'Amazon Product Review'
Dataset**

**SUMANTH HAGALAVADI GOPALAKRISHNA
Northeastern University
MS-IS
NUID: 001824723**

SUMMARY:

This project is based on the Analysis of the 'Amazon Product Review' dataset. The dataset consists of reviews of products from Amazon. The data span a period of more than 10 years, including all ~500,000 reviews up to 2015. Reviews include product id and user information, and product ratings.

Map reduce analysis are done using Hadoop, Mongo dB and Pig. All the statistics are shown using Spring MVC application.

DATASET LINK:

<https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt>

SAMPLE PRODUCT REVIEW DATASET:

sample_fr																
Home Insert Draw Page Layout Formulas Data Review View																
AutoSave OFF																
Calibri (Body) 12 A A																
B I U																
Wrap Text																
Merge & Center																
General																
\$ %																
Conditional Formatting Format as Table Cell Styles Insert																
A1 fx marketplace																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	marketplace	customer_id	review_id	product_id	product_pare	product_title	product_category	star_rating	helpful_vote	total_votes	vine	verified_purc	review_head	review_body	review_date	
2	FR	14952	R32VYUWDI	552774294	362925721	The God Deli Books		5	0	0	N	Y	a propos de	je conseille f	2/13/13	
3	FR	14952	R3CCMP4EVI	B004GJXQ20	268067011	A Game of T Digital_Ebook_Purchase		5	0	0	N	Y	wow	ce magnifiq	8/3/14	
4	FR	17564	R14NAE6UG	B00GIGG564	256731097	Huion H610 f PC		3	1	3	N	Y	Ca fait le job	Je dirais qu'i	7/7/15	
5	FR	18940	R2E7QEWSC	B00CW7KK9H	977480037	Withings Pul Sports		4	0	1	N	Y	Fidele a desc	Je l'ai depuis	6/16/14	
6	FR	20315	R26E6i47GQ	B002LG5KIK	827187473	Prometheus Video DVD		2	3	5	N	N	dv©cevat	je m'attenda	6/10/13	
7	FR	20842	R1RJMTSNCI	B00004YLIU	678427290	Kid A Music		2	0	0	N	Y	10'	Ne disait pas	4/15/13	
8	FR	20913	R2P2XF84YE	B00AYHK7RU	108403123	The Next Day Music		5	0	3	N	Y	the next day	du bon bowi	6/7/13	
9	FR	21490	RJKULSX2Y5I	B00CJ3VSUK	252503117	G.I. Joe 2 : C Video DVD		5	0	0	N	Y	tres bon film	trv's bon filr	11/7/14	
10	FR	24196	R3UYEU75C	B000FUMOT	48021829	Amistad Video DVD		5	2	3	N	Y	Excellent	Un sujet dv©	11/16/12	
11	FR	24196	R1TKJ7XF53f	B000UMXCT	181554537	The Collectic Music		5	0	1	N	Y	C'est parfait	Un coffret d'	4/23/13	
12	FR	24196	R3S9JNS21Q	B0044JV1K6	386772628	Braddock - Iv Video DVD		5	0	2	N	Y	Excellent act'	Un grand cla	8/2/13	
13	FR	24196	R2Y4O06QM	2226249672	753674225	Au revoir Iv't Books		5	2	3	N	Y	Excellent act'	Livre offert p	10/3/13	
14	FR	24196	R3PS3P7G1Z	B000A3IF8G	757843869	The Essential Music		5	0	0	N	Y	Excellent act'	Un artiste in	8/21/14	
15	FR	24207	R1NFWFM2I	B00B4V18MC	351620851	Epic Adventu Mobile_Apps		5	0	0	N	Y	jeux	Trop cool ce	1/31/14	
16	FR	24207	R3QRIH5CXC	B00CW6DAK	87169065	Super Ferme Mobile_Apps		4	4	6	N	Y	ajouter	Ajouter moi	7/29/14	
17	FR	24629	RBCSM1KTYI	B0031ES8P6	724566333	Ty - Ty40819 Toys		4	0	0	N	Y	Trv's bien	Bonne idv©e	1/13/14	
18	FR	24844	R2BIWHFIS7	B007AR7OG	414300413	First lady of Music		5	3	4	N	Y	Great	Just great co	12/17/12	
19	FR	25935	R2VNGSD09	B00XO12REY	271191040	The Monsant Music		1	1	9	N	Y	commentairi	Pas beaucou	7/21/15	
20	FR	26425	RTGC85CCO	B007ZMHB8	893659638	Checkerboari Video DVD		5	0	0	N	Y	Muddy Wate	Superbe sess	11/22/12	
21	FR	26552	R128Q27ZKT	446310786	994527755	To Kill a Moc Books		5	0	0	N	Y	Beautiful!	I know it's a	1/19/15	
22	FR	26552	R35CGHFHFG	B00D1VKZ3N	16141818	Bridget Jone: Digital_Ebook_Purchase		2	2	2	N	Y	What a disa	Waiting for s	1/22/15	
23	FR	28091	RP272TPN9Y	B00649LT06	992326428	Avengers [BI Video DVD		4	1	2	N	Y	Excellent	Produit comr	3/16/13	
24	FR	28091	R3J10C10DS	B0055ODYG	539923046	Made In Ger: Music		4	0	1	N	Y	Top	Produit comr	3/16/13	
25	FR	28091	R1N4WBV1I	B007K60MQ	251269761	Battleship Video DVD		4	0	1	N	Y	Envoi parfait	Produit comr	3/16/13	
26	FR	28104	R3SCHUSHIN	B00785P2QC	162354237	Virtual City P Mobile_Apps		5	3	4	N	Y	bravo	Super jeux di	1/20/13	
27	FR	28104	R2XMPUVFJ	B00C5R2J9I	690824381	Hill Climb Ra Mobile_Apps		5	11	15	N	Y	bien mais ...	Alors il est v	11/29/13	
28	FR	28143	RKLDQ1WDC	B00005UMY	698862122	Exodus Music		4	1	3	N	Y	Un CD de r	Un CD de r	1/21/13	
29	FR	28189	R1CL34PJLR	B00HDEEC9Z	803981632	Le Hobbit - L Video DVD		5	0	1	N	Y	genial	film super px	10/7/14	
30	FR	28448	R3L6J8KA59f	B00B0ID7D8	912727710	LEGO Calend Toys		2	0	1	N	Y	erreur de m	J'ai fait l'erre	1/24/14	
31	FR	28465	R2I1383CZ4C	B00004VXUC	735114749	The Big Lebco Video DVD		3	1	4	N	Y	The Big Lebco	Je n'ai pas e	4/13/15	
32	FR	28465	R2PAKHGCS	B00005UMY	514940909	O'Brother Video DVD		5	0	0	N	Y	Film : O' Bro	Trv's, trv's	5/24/15	
33	FR	28557	R1MFYKSGCI	B001I45M5A	63233157	James Bond Video DVD		4	0	0	N	Y	Sympa	Un 007 moir	11/17/14	
34	FR	28692	R1BSSQ563E	6486118	586496321	A Song of Ice Books		5	0	0	N	Y	Excellent box	Le meilleur c	9/24/13	
35	FR	28700	RSOWF1UJJ	B00I7VMIC2	551407899	True Detectiv Video DVD		5	1	1	N	Y	une sv©rie c'	J'ai regardV©	7/6/14	

DATA CLEANING AND FILTERING:

Since the combined dataset was very huge, Random Filtering is used to get random data from the dataset.

DATA STATISTICS AND ANALYSIS:

Following are the analysis done using MapReduce:

- Average review of the product based on product category
 - Mongo DB
- Percentage of verified purchases for each product
 - Hadoop
- Standard deviation and median based on Rating for the products
 - Hadoop
- Top 100 products based on Total votes
 - Hadoop
- Total votes based on year
 - Hadoop
- Filtering data for Review Headline which has “Five Stars”
 - Pig

SPRING MVC WEB APPLICATION:

Spring MVC is used to show the statistics for each analysis which has been done using different Map Reduce jobs.

Choose which analytics you want to see

Total Votes based on year 

←

→

↻

localhost:9090/controller/

Average review of product category

Percentage of Verified purchases

Standard Deviation and Median

Top 100 products

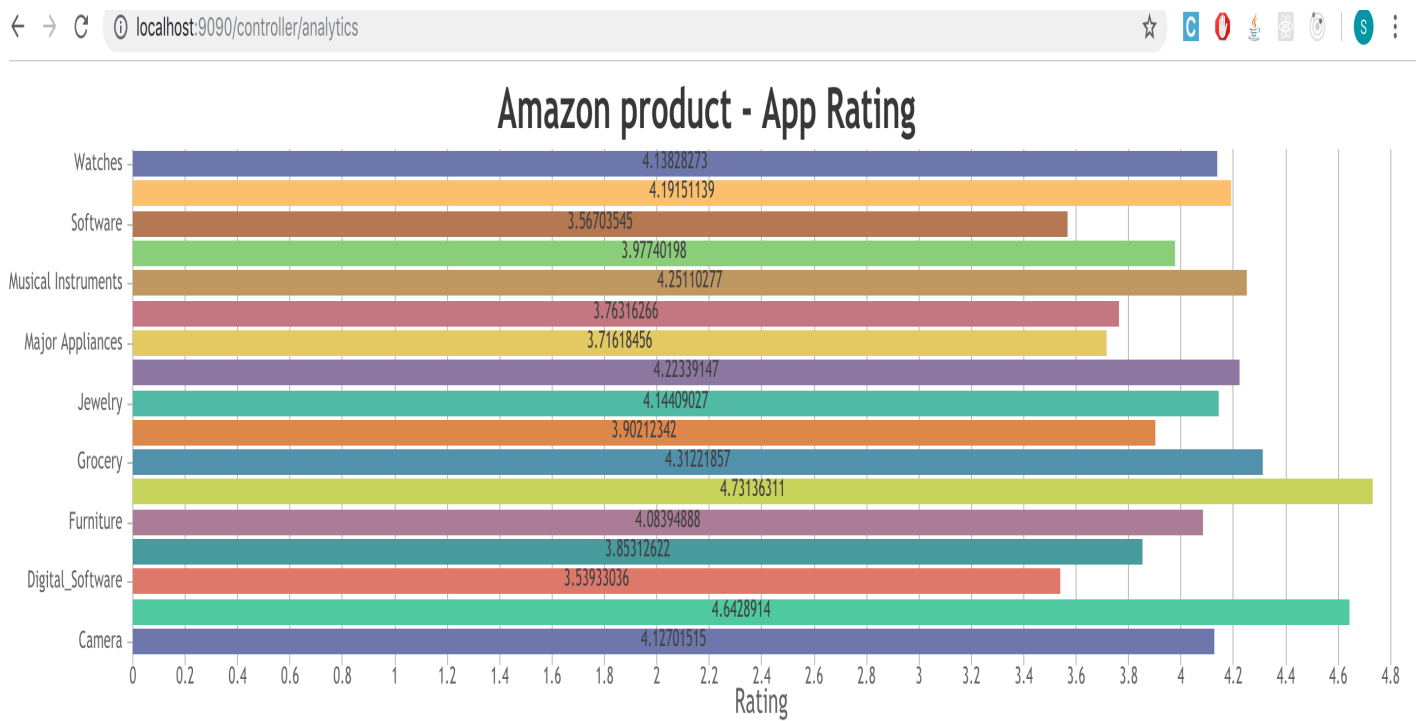
✓ Total Votes based on year

Data with Review headline Five Star

Submit

Analytics you want to see

1. AVERAGE REVIEW OF THE PRODUCT BASED ON PRODUCT CATEGORY



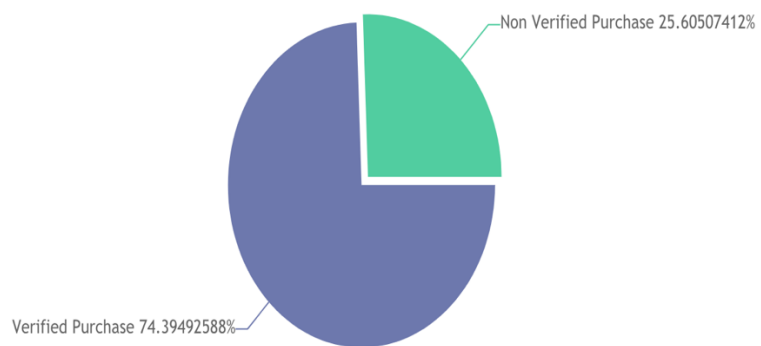
2. PERCENTAGE OF VERIFIED PURCHASES FOR EACH PRODUCT

← → ↻ ⓘ localhost:9090/controller/analytics

Please enter the Product name

← → ↻ ⓘ localhost:9090/controller/percentage

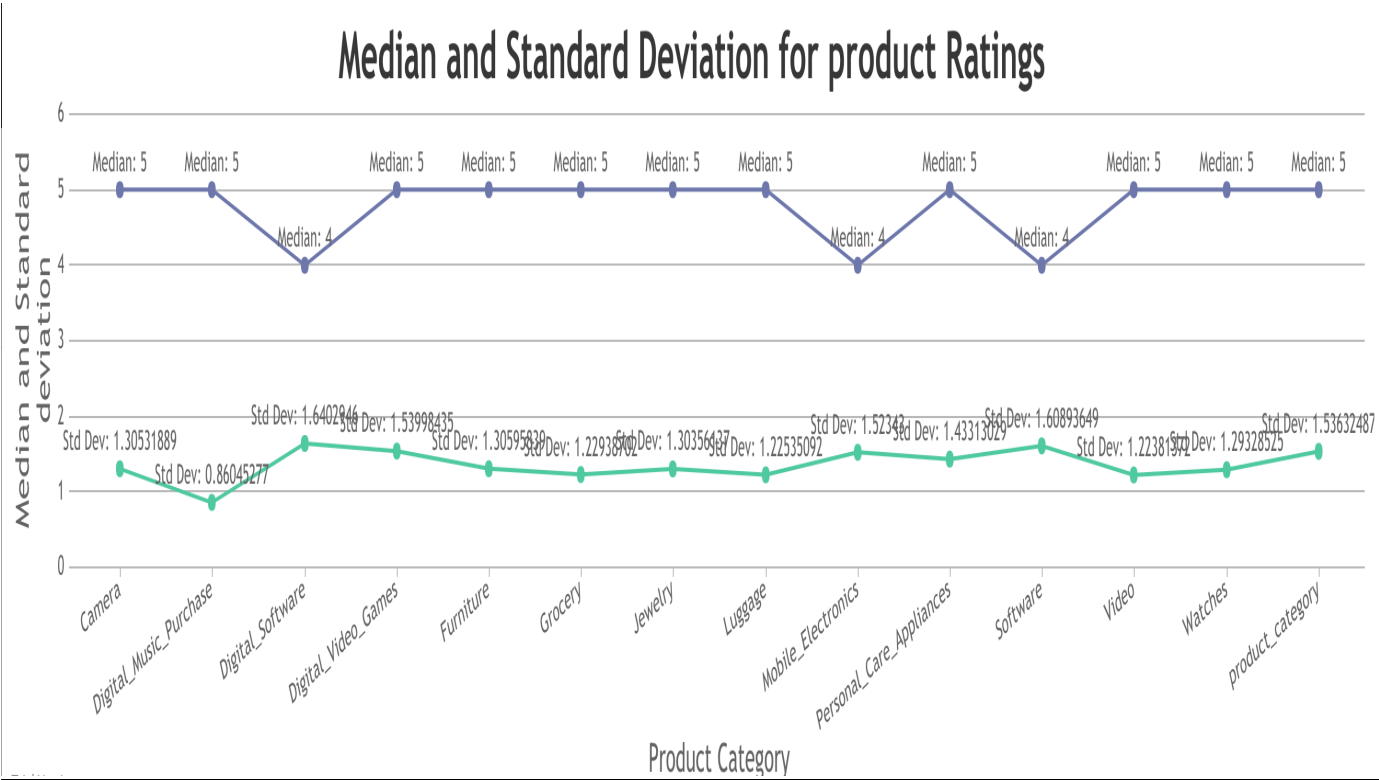
Percentage of Verified and Non Verified products



Trial Version

● Verified Purchase ● Non Verified Purchase

3. STANDARD DEVIATION AND MEDIAN BASED ON RATING FOR THE PRODUCTS

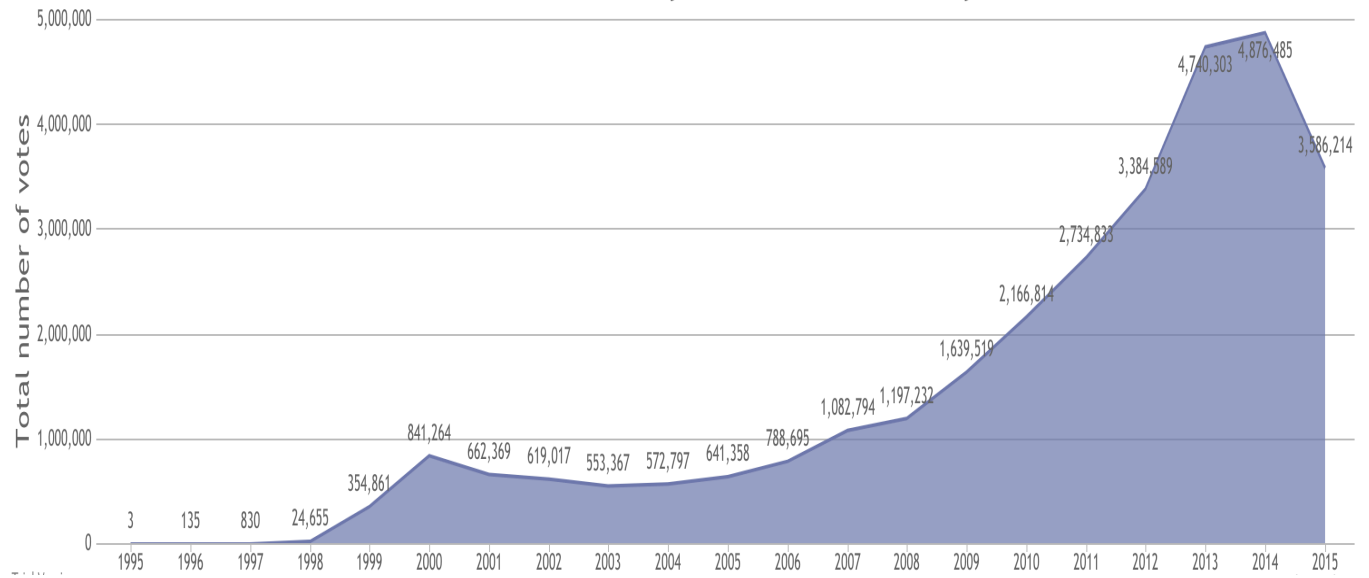


4. TOP 100 PRODUCTS BASED ON TOTAL VOTES

Product Id	Total Votes
BT00CTOUNS	2253
B00NG7K2RA	2268
B006ZP8UOW	2269
B004TJ6JH6	2274
B007VGGFZU	2313
B006P88VSE	2324
B0046ZLW1G	2354
B000V1MLBE	2356
B0055DXL1C	2365
B00NIYJF6U	2381
B004Q0PT3I	2394
B001XURPQS	2406
B00BGGDVOO	2410
B004P8K24W	2411
B002BHSKAC	2416
B00M2OGS08	2433
B00L403O8U	2434
B00M2OGS08	2436
B000EVOSE4	2457
B005CLPP84	2463
B002VA464S	2485
B009B0MZ8U	2505
B007HD9YYQ	2510
B00VSITBJO	2514
B004391DK0	2518
B00032G1S0	2525

5. TOTAL VOTES BASED ON YEAR

Total number of votes by customer in all the years



6. FILTERING DATA FOR REVIEW HEADLINE WHICH HAS “FIVE STARS”

Marketplace	Customer Id	Review Id	Product Id	Product Parent	Product Title	Product Category	Total Votes	Review Headline
US	13394189	R3BPETL222LMIM	B009B7F6CA	873028700	Melissa & Doug Water Wow Coloring Book - Vehicles	Toys	8	Five Stars
US	18778586	RDJJS7QYB6XNR	B00EDBY7X8	122952789	Monopoly Junior Board Game	Toys	3	Five Stars
US	41137196	R2RDOJQ0WBZCF6	B00407S11Y	383363775	Fun Express Insect Finger Puppets 12ct Toy	Pen	11	Five Stars
US	433677	R2B8VBEPB4YEZ7	B00FGPU7U2	780517568	Fisher-Price Octonauts Shellington's On-The-Go Pod Toy	Pen	10	Five Stars
US	6762003	R1H1HOVB44808I	B00PXWS1CY	996611871	Holy Stone F180C Mini RC Quadcopter Drone with Camera 2.4GHz 6-Axis Gyro Bonus Battery and 8 Blades	Pen	1	Five Stars
US	12612039	R1JS8G26X4RM2G	B00D4NJSJE	408940178	"The Game of Life Money and Asset Board Game, Fame Edition"	Car	1	Five Stars
US	15048896	R3N01IESYEW01	B001CTOC8O	278247652	Star Wars Clone Wars Clone Trooper Child's Deluxe Captain Rex Costume	Bat	1	Five Stars
US	7360347	R2BUV9QJ12A00X	B00DOQCWF8	226984155	Minecraft Animal Toy (6-Pack)	Pen	1	Five Stars
US	47781982	R1DVQ4P3WJR42	B00WTGGGRO	162262449	Pokemon - Double Dragon Energy (97/108) - XY Roaring Skies	Car	1	Five Stars
US	32866903	R300I65NW30Y19	B000TFLAZA	149264874	Baby Einstein Octoplush	Car	0	Five Stars
US	41168357	R3T73PQZZ9F6GT	B00CAEEDC0	72805974	Seat Pets Car Seat Toy	Car	0	Five Stars
US	11916403	R36L8VKT9ZSUY6	B00JYV9J1M	771795950	Winston Zeddemore & Ecto-1: Funko POP! Rides x Ghostbusters Vinyl Figure	Car	0	Five Stars
US	16245463	R23URALWA7IHVL	B00IGXV9UI	765869385	Disney Planes: Fire & Rescue Scoop & Spray Firefighter Dusty	Car	0	Five Stars
US	47781982	R3AHZWWOL0IAV0	B00GNDY4OU	438056479	Pokemon - Gyarados (31/113) - Legendary Treasures	Car	0	Five Stars
US	44928701	R1ORWPFQ9EDYA0	B000HZZT7W	967346376	LCR Dice Game (Red Chips)	Car	0	Five Stars
US	125518	R2MW3TEBPWKENS	B00MZ6BR3Q	145562057	Monster High Haunted Student Spirits Porter Geiss Doll	Bat	0	Five Stars
US	18461411	R2SDXLTLF920OH	B00VPPX92W	705054378	Teenage Mutant Ninja Turtles T-Machines Tiger Claw in Safari Truck Diecast Vehicle	Pen	0	Five Stars
US	43880421	R2ATXF4Q3Q0YW	B00000JS5S	341842639	Hot Wheels 48- Car storage Case With Easy Grip Carrying Case	Pen	0	Five Stars
US	52006292	R2D90RQ3V8LH	B00519PJTW	493486387	100 Foot Multicolor Pennant Banner	Pen	0	Five Stars
US	128540	R24VKWVWUMV3M3	B004S8F7QM	829220659	Cards Against Humanity	Bat	0	Five Stars

CONCLUSION:

The MapReduce algorithms performed in the project are java codes with implementation on Hadoop, HDFS, Mongo DB, pig. Spring MVC is used to show the statistical analysis of the data.