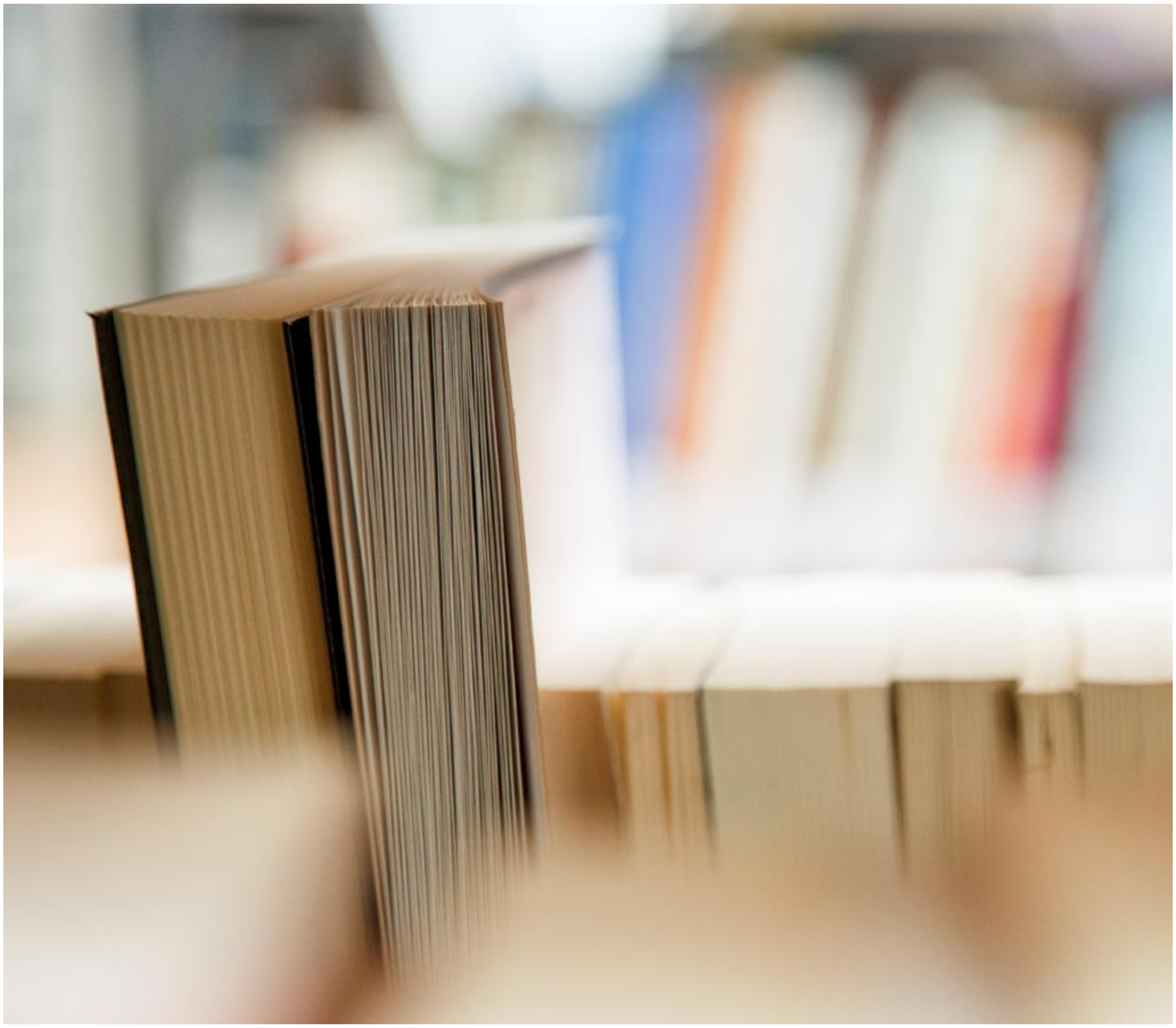


CIS8045–Term Project

Working with the Amazon dataset

SRAVANI KOTHI
SHREYA PATIL
SHILPA VERMA
ANANNYA DAS

4/26/18

UNSTRUCTURED DATA MGMT

Table of Contents

1. Design

1.1 MongoDB Schema design

Our MongoDB schema has the following collections: Product, Reviews, Reviewer.

Unique Index were created on review Text.

Below queries show the creation of each collection (sample), queries used to create subsets, data type constraints, a sample document as well as the index creation.

The first step is to create the collections that have indexes/constraints built in them. Here we create the constraints for collections metadata/product and review

I. CREATING CONSTRAINTS

```
db.createCollection({“product”,{
Validator:
{“price”: {$gt: 0}}}
})
```

```
db.createCollection({“review”,{
Validator:
{“overall”: {$in: [1,2,3,4,5] }}
}})
```

II. CREATING INDEXES

For text search, index on reviewText is created.

```
db.review.createIndex({text: “reviewText”})
```

III. IMPLEMENTING EMBEDDING

Given that when a product is viewed on Amazon all the details along with the reviews are viewed. The most efficient way of presenting the information on Amazon is to embed the review collection into the Product collection. Reviewer will be as a separate collection.

IV. PROPOSED SCHEMA

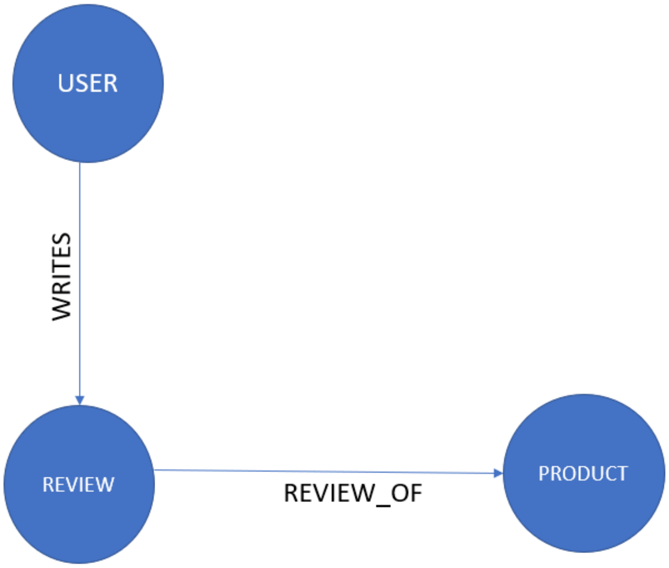
The schema will now be:

```
Product:
{
	"_id" : ObjectId("5ad7e205a9be863492df696e"),
	"asin" : "0000037214",
	"related" : {
		"also_viewed" : [
			"B00JO8II76",
			"B00DGN4R1Q",
			"B00E1YRI4C"
		]
	},
	"title" : "Purple Sequin Tiny Dancer Tutu Ballet Dance Fairy Princess Costume Accessory",
	"price" : 6.99,
	"salesRank" : {
		"Clothing" : 1233557
	},
	"imUrl" : "http://ecx.images-amazon.com/images/I/31mCncNuAZL.jpg",
	"brand" : "Big Dreams",
	"categories" : [
		[
			"Clothing, Shoes & Jewelry",
			"Girls"
		],
		[
			"Clothing, Shoes & Jewelry",
			"Novelty, Costumes & More",
			"Costumes & Accessories",
			"More Accessories",
			"Kids & Baby"
		]
	],
“Reviews”: [
        {
            "_id" : ObjectId("5ad60c80ce94169d49704dd1"),
            "reviewerID" : "A27IQHDZFQFNGG",
            "reviewerName" : "Caitlin",
            "helpful" : [  3,  4  ],
            "reviewText" : "Really good. Great gift for any fan of green tea! Just so expensive
        to purchase candy from across the sea.",
            "overall" : 4,
            "summary" : "Yum!",
            "unixReviewTime" : 1381190400,
            "reviewTime" : "10 8, 2013"
        }
    ]
}
```

```
Reviewer: {
        {
            "_id" : ObjectId("5ad60c80ce94169d49704dd1"),
            "reviewerID" : "A27IQHDZFQFNGG",
            "reviewerName" : "Caitlin",

        }
}
```

1.2 NEO4J SCHEMA DESIGN

PROPOSED PSEUDO DATABASE SCHEMA FOR Neo4j



The graph model consists of 3 nodes and 2 edges/relationships. Their descriptions are as below:

I. NODES AND PROPERTIES

USER	REVIEW	PRODUCT
• _id •reviewerID •reviewerName •review_count •reviewing_since •reviewerRanking	• _id •reviewerId •asin •reviewerName •helpful •reviewText •overall •summary •unixReviewTime • reviewTime	• _id •asin •related •title •price •salesRank •imUrl •brand •categories

II. RELATIONSHIPS

WRITES
REVIEW_OF

Justification: To be able to query users sometimes review as a gang etc, we have made 3 nodes User, Review and Product respectively.

2. Basic Understanding of The Data

2.1 MongoDB Queries and Insights:

a. What is the overall number of products

```
db.grocery_and_gourmet_food.distinct('asin').length
```

b. The overall number of reviews

```
db.grocery_and_gourmet_food.count()
```

c. The overall number of reviewers

```
db.grocery_and_gourmet_food.distinct('reviewerID').length
```

d. The overall number of reviews with ratings less than 3

```
db.grocery_and_gourmet_food.aggregate([ {$match: {overall:{ "$lt": 3 } } }, {$count: "RatingBelow3"} ])
```

e. The overall number of reviews with ratings more than 3

```
db.grocery_and_gourmet_food.aggregate([ {$match: {overall:{ "$gt": 3 } } }, {$count: "RatingAbove3"} ]
```

f. The average number of reviews per product

```
db.grocery_and_gourmet_food.aggregate( [ { $group: { "_id": "$asin", avgRating: {$avg: "$overall"} } } ] )
```

g. The date of the first review per category

```
db.grocery_and_gourmet_food.find({}, {reviewTime: 1, _id: 0}).sort({"unixReviewTime": 1}).limit(1)
```

h. The top 10 most prolific reviewers

```
db.grocery_and_gourmet_food.aggregate([{$group: {_id:"$reviewerID", noOfReviews:{$sum:1}}},{"$sort":{"noOfReviews":-1}},{$limit:10}], {allowDiskUse: true, cursor: {} })
```

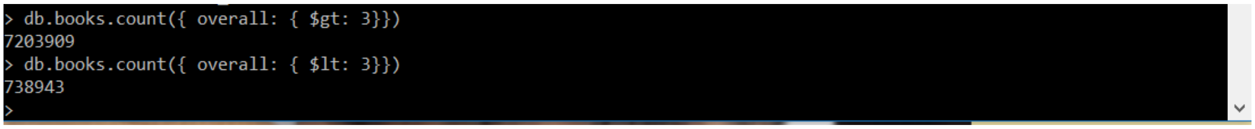
i. The top 10 most verbose reviewers

```
db.grocery_and_gourmet_food.aggregate([ { $group: {_id: "$reviewerID", "review_length": {$sum: {"$strLenCP":"$reviewText"}}}  }, {$sort: {review_length: -1} }, {$limit: 10} ],{allowDiskUse: true} )
```

j. Report and interpret your findings (e.g., if you find that there are more positive (>3) reviews than negative (<3), what implication does that have?, is the trend the same across product categories types?)

Category 1: Grocery_and_gourmet_food: >3: 120044, <3: 13696

Category 2:



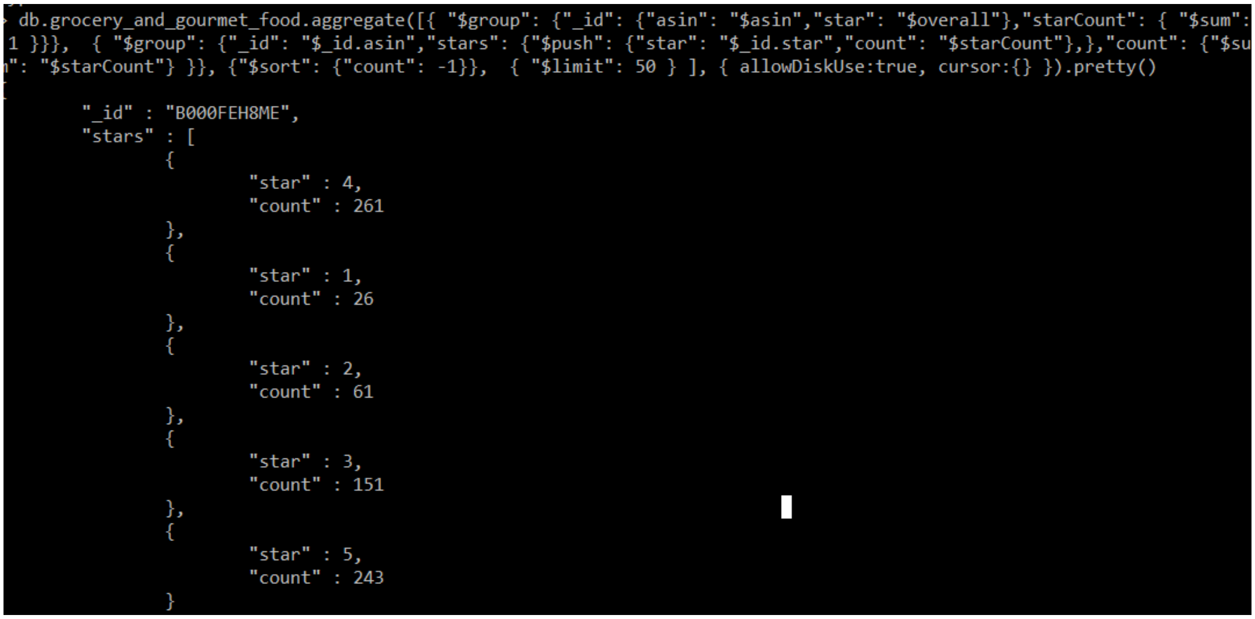
```
> db.books.count({ overall: { $gt: 3}})
7203909
> db.books.count({ overall: { $lt: 3}})
738943
>
```

This imply that there could be a possibility of unusual activity and that sum of these ratings could be fake or forced ratings.

3. Analytics

3.1 Review Histogram

db.grocery_and_gourmet_food.aggregate([{ "$group": {"_id": {"asin": "$asin","star": "$overall"},"starCount": { "$sum": 1 }}}, { "$group": {"_id": "$_id.asin","stars": {"$push": {"star": "$_id.star","count": "$starCount"},},"count": {"$sum": "$starCount"} }}, {"$sort": {"count": -1}}, { "$limit": 50 }], { allowDiskUse:true, cursor:{} })



```
> db.grocery_and_gourmet_food.aggregate([{ "$group": {"_id": {"asin": "$asin","star": "$overall"},"starCount": { "$sum":
 1 }}}, { "$group": {"_id": "$_id.asin","stars": {"$push": {"star": "$_id.star","count": "$starCount"},},"count": {"$su
m": "$starCount"} }}, {"$sort": {"count": -1}}, { "$limit": 50 } ], { allowDiskUse:true, cursor:{} }).pretty()
{
        "_id" : "B000FEH8ME",
        "stars" : [
                {
                        "star" : 4,
                        "count" : 261
                },
                {
                        "star" : 1,
                        "count" : 26
                },
                {
                        "star" : 2,
                        "count" : 61
                },
                {
                        "star" : 3,
                        "count" : 151
                },
                {
                        "star" : 5,
                        "count" : 243
                }
```

This query is to display the count of each rating given by reviewers to a product. This is useful for a customer to analyse how good the product is based on the distribution of ratings.

3.2 LIST OF TOP 10 MOST RECENT REVIEWS

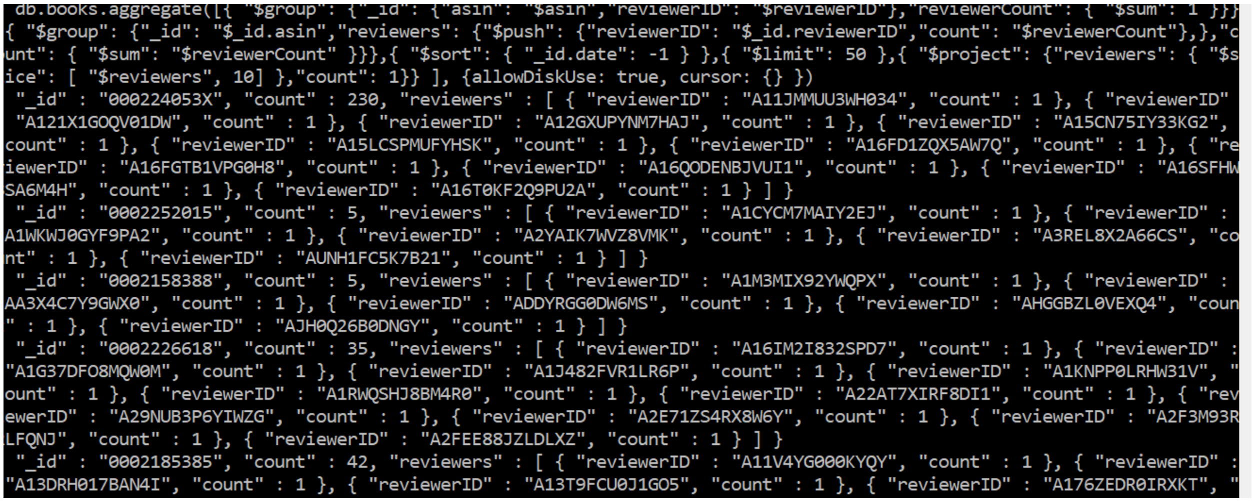
db.books.aggregate([{ "$group": {"_id": {"asin": "$asin","reviewerID":

"$reviewerID"},"reviewerCount": { "$sum": 1 }}},{ "$group": {"_id":

"$_id.asin","reviewers": {"$push": {"reviewerID": "$_id.reviewerID","count":

"$reviewerCount"},},"count": { "$sum": "$reviewerCount" }}},{ "$sort": { "_id.date": -1 } },{

"$limit": 50 },{ "$project": {"reviewers": { "$slice": ["$reviewers", 10] },"count": 1}}], {allowDiskUse: true, cursor: {} })

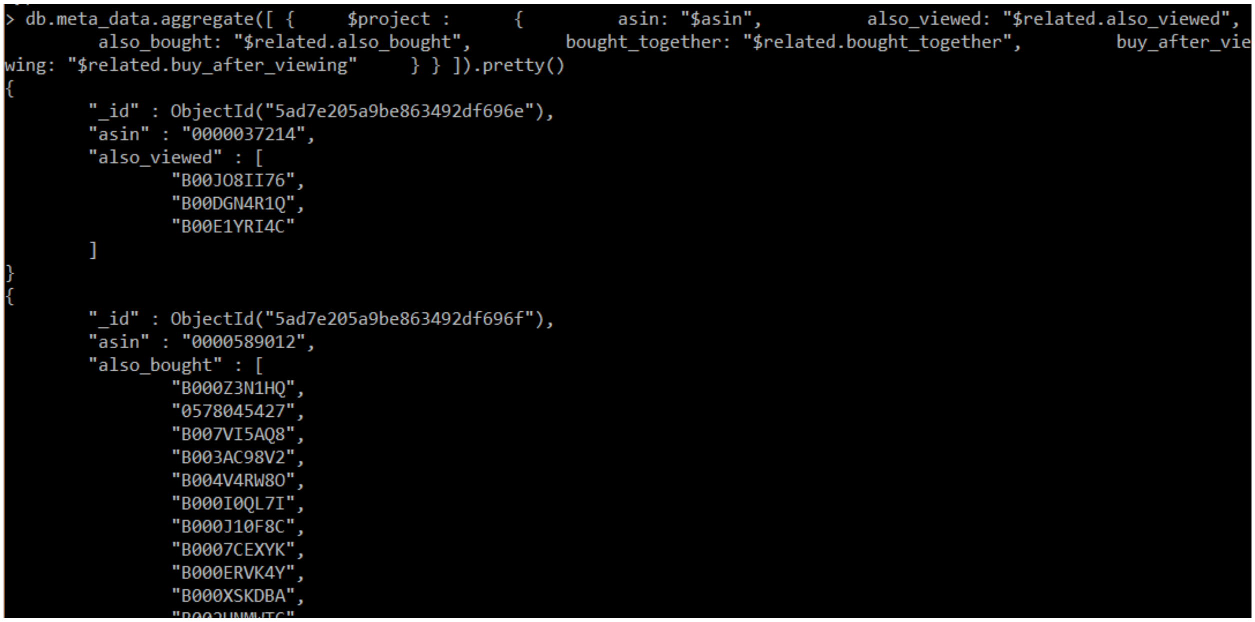


```
 db.books.aggregate([{ "$group": {"_id": {"asin": "$asin","reviewerID": "$reviewerID"},"reviewerCount": { "$sum": 1 }}}
{ "$group": {"_id": "$_id.asin","reviewers": {"$push": {"reviewerID": "$_id.reviewerID","count": "$reviewerCount"},},"c
unt": { "$sum": "$reviewerCount" }}},{ "$sort": { "_id.date": -1 } },{ "$limit": 50 },{ "$project": {"reviewers": { "$s
ice": [ "$reviewers", 10] },"count": 1}} ], {allowDiskUse: true, cursor: {} })
 "_id" : "000224053X", "count" : 230, "reviewers" : [ { "reviewerID" : "A11JMMUU3WH034", "count" : 1 }, { "reviewerID"
 "A121X1GOQV01DW", "count" : 1 }, { "reviewerID" : "A12GXUPYNM7HAJ", "count" : 1 }, { "reviewerID" : "A15CN75IY33KG2",
count" : 1 }, { "reviewerID" : "A15LCSPMUFYHSK", "count" : 1 }, { "reviewerID" : "A16FD1ZQX5AW7Q", "count" : 1 }, { "re
iewerID" : "A16FGTB1VPG0H8", "count" : 1 }, { "reviewerID" : "A16QODENBJVUI1", "count" : 1 }, { "reviewerID" : "A16SFHW
SA6M4H", "count" : 1 }, { "reviewerID" : "A16T0KF2Q9PU2A", "count" : 1 } ] }
 "_id" : "0002252015", "count" : 5, "reviewers" : [ { "reviewerID" : "A1CYCM7MAIY2EJ", "count" : 1 }, { "reviewerID" :
A1WKWJ0GYF9PA2", "count" : 1 }, { "reviewerID" : "A2YAIK7WVZ8VMK", "count" : 1 }, { "reviewerID" : "A3REL8X2A66CS", "co
nt" : 1 }, { "reviewerID" : "AUNH1FC5K7B21", "count" : 1 } ] }
 "_id" : "0002158388", "count" : 5, "reviewers" : [ { "reviewerID" : "A1M3MIX92YWQPX", "count" : 1 }, { "reviewerID" :
AA3X4C7Y9GWX0", "count" : 1 }, { "reviewerID" : "ADDYRGG0DW6MS", "count" : 1 }, { "reviewerID" : "AHGGBZL0VEXQ4", "coun
" : 1 }, { "reviewerID" : "AJH0Q26B0DNGY", "count" : 1 } ] }
 "_id" : "0002226618", "count" : 35, "reviewers" : [ { "reviewerID" : "A16IM2I832SPD7", "count" : 1 }, { "reviewerID" :
"A1G37DFO8MQW0M", "count" : 1 }, { "reviewerID" : "A1J482FVR1LR6P", "count" : 1 }, { "reviewerID" : "A1KNPP0LRHW31V", "
ount" : 1 }, { "reviewerID" : "A1RWQSHJ8BM4R0", "count" : 1 }, { "reviewerID" : "A22AT7XIRF8DI1", "count" : 1 }, { "rev
ewerID" : "A29NUB3P6YIWZG", "count" : 1 }, { "reviewerID" : "A2E71ZS4RX8W6Y", "count" : 1 }, { "reviewerID" : "A2F3M93R
LFQNJ", "count" : 1 }, { "reviewerID" : "A2FEE88JZLDLXZ", "count" : 1 } ] }
 "_id" : "0002185385", "count" : 42, "reviewers" : [ { "reviewerID" : "A11V4YG000KYQY", "count" : 1 }, { "reviewerID" :
"A13DRH017BAN4I", "count" : 1 }, { "reviewerID" : "A13T9FCU0J1GO5", "count" : 1 }, { "reviewerID" : "A176ZEDR0IRXKT", "
```

A product can have good reviews and ratings in the past. A customer wants to know the reviews and how the product is from someone who has bought it recently. Therefore this metric is useful.

3.3 ALSO VIEWED, ALSO BOUGHT

```
db.meta_data.aggregate([

{

   $project :

   {

      asin: "$asin",

      also_viewed: "$related.also_viewed",

      also_bought: "$related.also_bought",

      bought_together: "$related.bought_together",

      buy_after_viewing: "$related.buy_after_viewing"

   }

}
]).pretty()
```

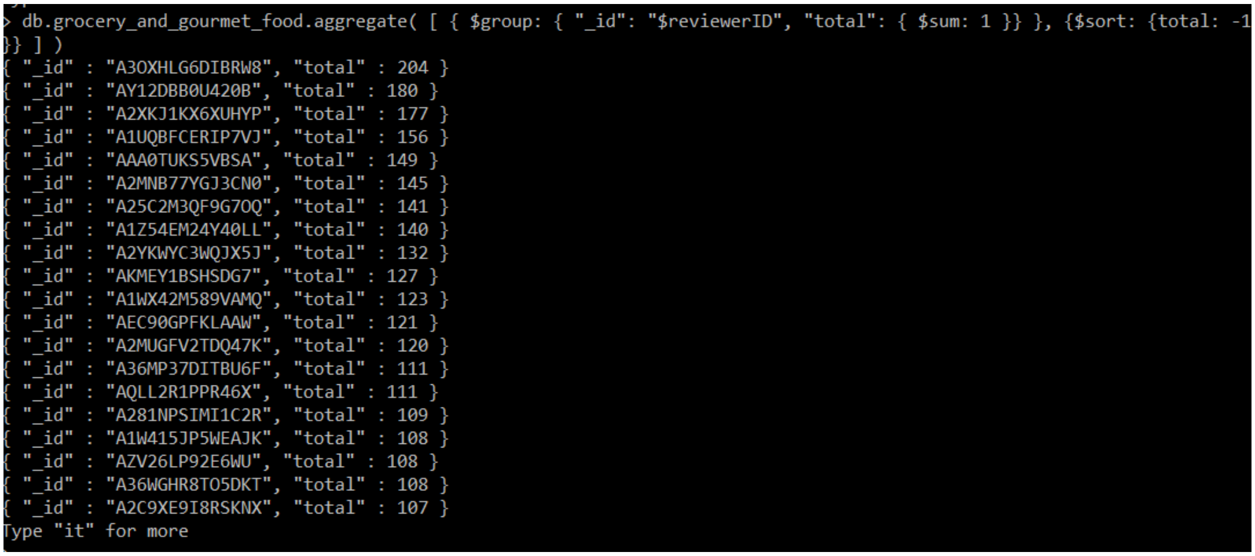


```
> db.meta_data.aggregate([ {     $project :      {         asin: "$asin",          also_viewed: "$related.also_viewed",
         also_bought: "$related.also_bought",         bought_together: "$related.bought_together",         buy_after_vie
wing: "$related.buy_after_viewing"     } } ]).pretty()
{
        "_id" : ObjectId("5ad7e205a9be863492df696e"),
        "asin" : "0000037214",
        "also_viewed" : [
                "B00JO8II76",
                "B00DGN4R1Q",
                "B00E1YRI4C"
        ]
}
{
        "_id" : ObjectId("5ad7e205a9be863492df696f"),
        "asin" : "0000589012",
        "also_bought" : [
                "B000Z3N1HQ",
                "0578045427",
                "B007VI5AQ8",
                "B003AC98V2",
                "B004V4RW8O",
                "B000I0QL7I",
                "B000J10F8C",
                "B0007CEXYK",
                "B000ERVK4Y",
                "B000XSKDBA",
                "B002UNMWTG"
```

This metric is useful to improve the sales..

3.4 REVIEWER REVIEWS COUNT

```
db.grocery_and_gourmet_food.aggregate( [ { $group: { "_id": "$reviewerID", "total": {
$sum: 1 }} }, {$sort: {total: -1}} ] )
```

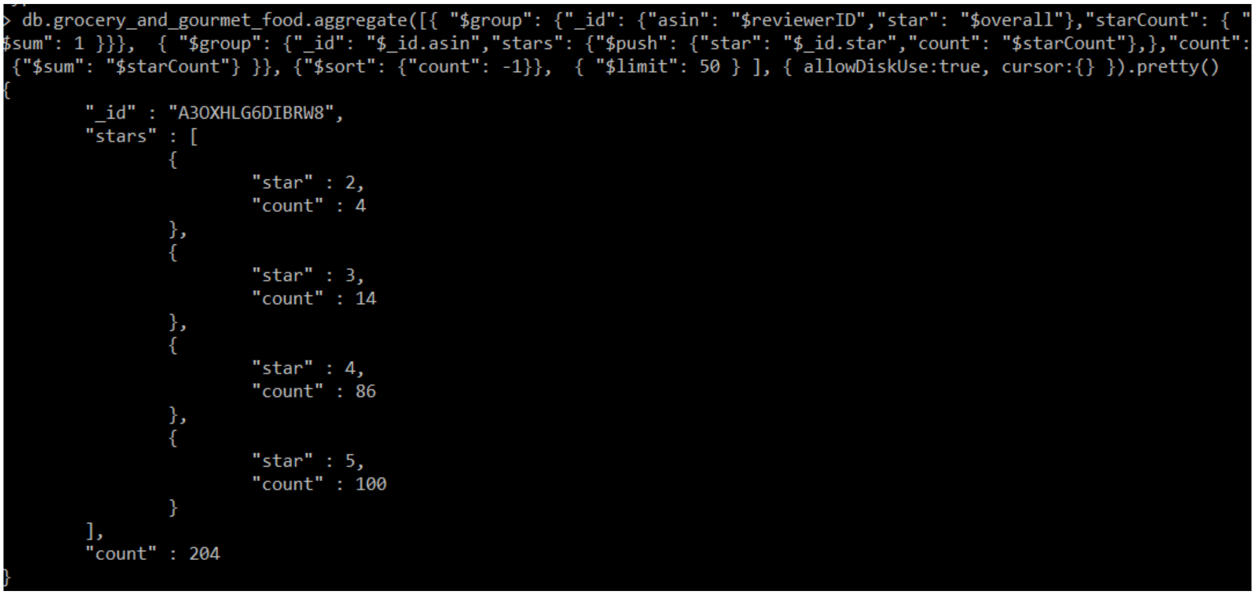


```
> db.grocery_and_gourmet_food.aggregate( [ { $group: { "_id": "$reviewerID", "total": { $sum: 1 }} }, {$sort: {total: -1
}} ] )
{ "_id" : "A3OXHLG6DIBRW8", "total" : 204 }
{ "_id" : "AY12DBB0U420B", "total" : 180 }
{ "_id" : "A2XKJ1KX6XUHYP", "total" : 177 }
{ "_id" : "A1UQBFCERIP7VJ", "total" : 156 }
{ "_id" : "AAA0TUKS5VBSA", "total" : 149 }
{ "_id" : "A2MNB77YGJ3CN0", "total" : 145 }
{ "_id" : "A25C2M3QF9G7OQ", "total" : 141 }
{ "_id" : "A1Z54EM24Y40LL", "total" : 140 }
{ "_id" : "A2YKWYC3WQJX5J", "total" : 132 }
{ "_id" : "AKMEY1BSHSDG7", "total" : 127 }
{ "_id" : "A1WX42M589VAMQ", "total" : 123 }
{ "_id" : "AEC90GPFKLAAW", "total" : 121 }
{ "_id" : "A2MUGFV2TDQ47K", "total" : 120 }
{ "_id" : "A36MP37DITBU6F", "total" : 111 }
{ "_id" : "AQLL2R1PPR46X", "total" : 111 }
{ "_id" : "A281NPSIMI1C2R", "total" : 109 }
{ "_id" : "A1W415JP5WEAJK", "total" : 108 }
{ "_id" : "AZV26LP92E6WU", "total" : 108 }
{ "_id" : "A36WGHR8TO5DKT", "total" : 108 }
{ "_id" : "A2C9XE9I8RSKNX", "total" : 107 }
Type "it" for more
```

This metric is useful to find the total number of reviews a reviewer has given till data and showed on the reviewer page.

3.5 REVIEWER RATING HISTOGRAM

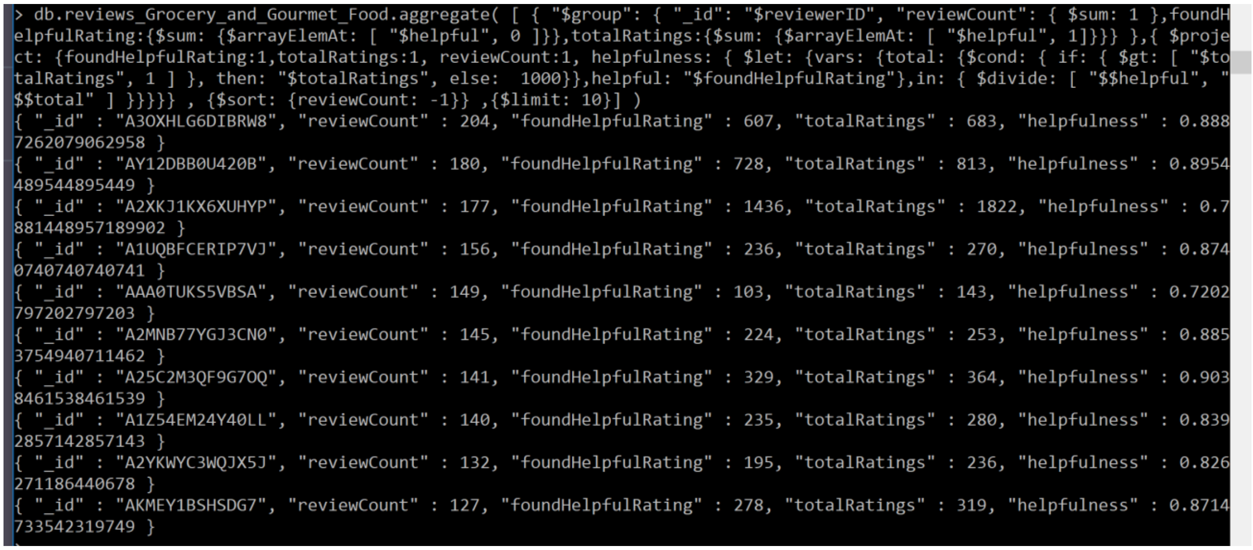
db.grocery_and_gourmet_food.aggregate([{ "$group": {"_id": {"asin": "$reviewerID","star": "$overall"},"starCount": { "$sum": 1 }}}, { "$group": {"_id": "$_id.asin","stars": {"$push": {"star": "$_id.star","count": "$starCount"},},"count": {"$sum": "$starCount"} }}, {"$sort": {"count": -1}}, { "$limit": 50 }], { allowDiskUse:true, cursor:{} }).pretty()



```
> db.grocery_and_gourmet_food.aggregate([{ "$group": {"_id": {"asin": "$reviewerID","star": "$overall"},"starCount": { "
$sum": 1 }}},  { "$group": {"_id": "$_id.asin","stars": {"$push": {"star": "$_id.star","count": "$starCount"},},"count":
 {"$sum": "$starCount"} }}, {"$sort": {"count": -1}},  { "$limit": 50 } ], { allowDiskUse:true, cursor:{} }).pretty()
{
        "_id" : "A3OXHLG6DIBRW8",
        "stars" : [
                {
                        "star" : 2,
                        "count" : 4
                },
                {
                        "star" : 3,
                        "count" : 14
                },
                {
                        "star" : 4,
                        "count" : 86
                },
                {
                        "star" : 5,
                        "count" : 100
                }
        ],
        "count" : 204
}
```

3.6 Helpfulness Rating of reviewers with most number of reviews -

db.reviews_Grocery_and_Gourmet_Food.aggregate([{ "$group": { "_id": "$reviewerID", "reviewCount": { $sum: 1 },foundHelpfulRating:{$sum: {$arrayElemAt: ["$helpful", 0]}},totalRatings:{$sum: {$arrayElemAt: ["$helpful", 1]}}} },{ $project: {foundHelpfulRating:1,totalRatings:1, reviewCount:1, helpfulness: { $let: {vars: {total: {$cond: { if: { $gt: ["$totalRatings", 1] }, then: "$totalRatings", else: 1000}},helpful: "$foundHelpfulRating"},in: { $divide: ["$$helpful", "$$total"] }}}}} , {$sort: {reviewCount: -1}} ,{$limit: 10}])



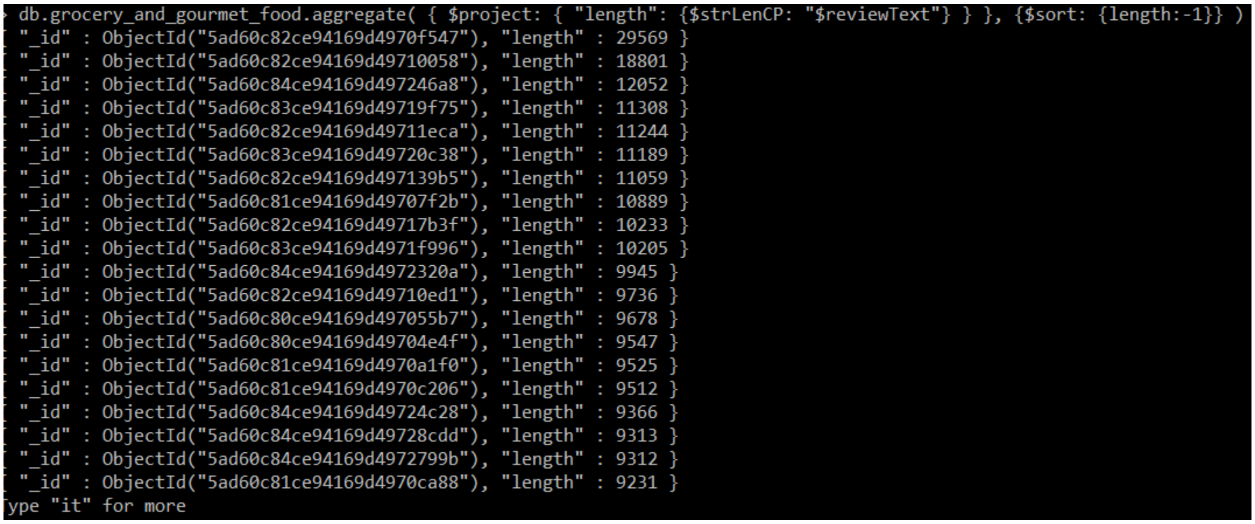
```
> db.reviews_Grocery_and_Gourmet_Food.aggregate( [ { "$group": { "_id": "$reviewerID", "reviewCount": { $sum: 1 },foundH
elpfulRating:{$sum: {$arrayElemAt: [ "$helpful", 0 ]}},totalRatings:{$sum: {$arrayElemAt: [ "$helpful", 1]}}} },{ $proje
ct: {foundHelpfulRating:1,totalRatings:1, reviewCount:1, helpfulness: { $let: {vars: {total: {$cond: { if: { $gt: [ "$to
talRatings", 1 ] }, then: "$totalRatings", else:  1000}},helpful: "$foundHelpfulRating"},in: { $divide: [ "$$helpful", "
$$total" ] }}}}} , {$sort: {reviewCount: -1}} ,{$limit: 10}] )
{ "_id" : "A3OXHLG6DIBRW8", "reviewCount" : 204, "foundHelpfulRating" : 607, "totalRatings" : 683, "helpfulness" : 0.888
7262079062958 }
{ "_id" : "AY12DBB0U420B", "reviewCount" : 180, "foundHelpfulRating" : 728, "totalRatings" : 813, "helpfulness" : 0.8954
489544895449 }
{ "_id" : "A2XKJ1KX6XUHYP", "reviewCount" : 177, "foundHelpfulRating" : 1436, "totalRatings" : 1822, "helpfulness" : 0.7
881448957189902 }
{ "_id" : "A1UQBFCERIP7VJ", "reviewCount" : 156, "foundHelpfulRating" : 236, "totalRatings" : 270, "helpfulness" : 0.874
0740740740741 }
{ "_id" : "AAA0TUKS5VBSA", "reviewCount" : 149, "foundHelpfulRating" : 103, "totalRatings" : 143, "helpfulness" : 0.7202
797202797203 }
{ "_id" : "A2MNB77YGJ3CN0", "reviewCount" : 145, "foundHelpfulRating" : 224, "totalRatings" : 253, "helpfulness" : 0.885
3754940711462 }
{ "_id" : "A25C2M3QF9G7OQ", "reviewCount" : 141, "foundHelpfulRating" : 329, "totalRatings" : 364, "helpfulness" : 0.903
8461538461539 }
{ "_id" : "A1Z54EM24Y40LL", "reviewCount" : 140, "foundHelpfulRating" : 235, "totalRatings" : 280, "helpfulness" : 0.839
2857142857143 }
{ "_id" : "A2YKWYC3WQJX5J", "reviewCount" : 132, "foundHelpfulRating" : 195, "totalRatings" : 236, "helpfulness" : 0.826
271186440678 }
{ "_id" : "AKMEY1BSHSDG7", "reviewCount" : 127, "foundHelpfulRating" : 278, "totalRatings" : 319, "helpfulness" : 0.8714
733542319749 }
```

This metric is useful in deciding a reviewer's ranking i.e., if a reviewer has written 100 reviews in total and 90 people have found the reviews he wrote as helpful, then the helpfulness of his reviews is determined to be 0.9. Based on this we decide the ranking of a reviewer and show the ranking on reviewer page.

3.7 Text-based and a non-text based definition of helpful reviews

I. Text-based:

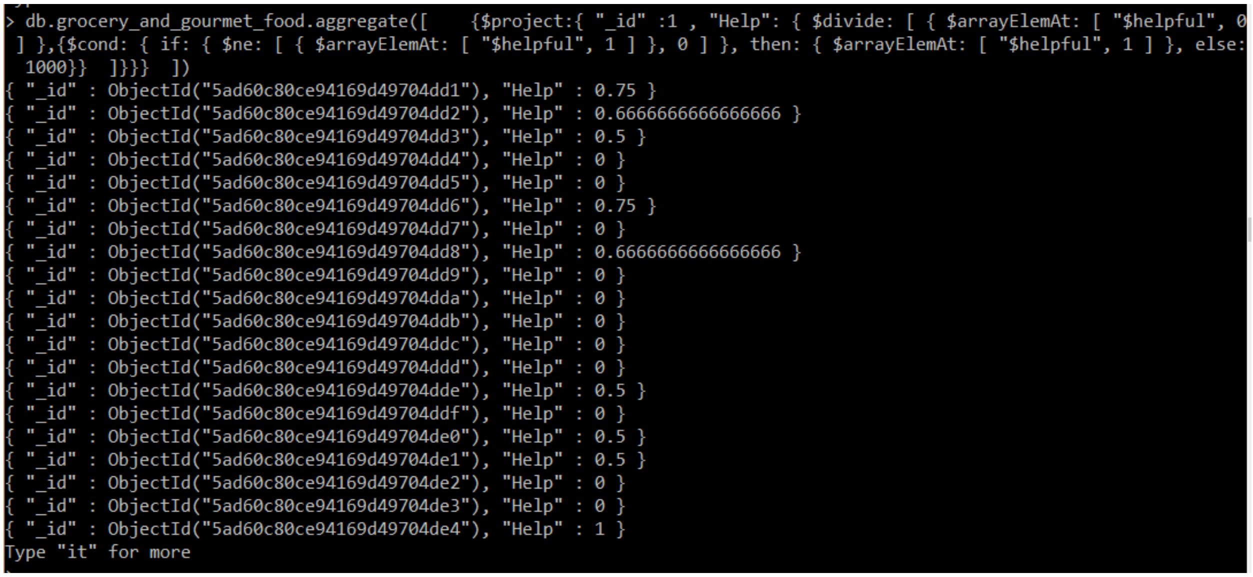
db.grocery_and_gourmet_food.aggregate({ $project: { "length": {$strLenCP: "$reviewText"} } }, {$sort: {length:-1}})



```
 db.grocery_and_gourmet_food.aggregate( { $project: { "length": {$strLenCP: "$reviewText"} } }, {$sort: {length:-1}} )
 "_id" : ObjectId("5ad60c82ce94169d4970f547"), "length" : 29569 }
 "_id" : ObjectId("5ad60c82ce94169d49710058"), "length" : 18801 }
 "_id" : ObjectId("5ad60c84ce94169d497246a8"), "length" : 12052 }
 "_id" : ObjectId("5ad60c83ce94169d49719f75"), "length" : 11308 }
 "_id" : ObjectId("5ad60c82ce94169d49711eca"), "length" : 11244 }
 "_id" : ObjectId("5ad60c83ce94169d49720c38"), "length" : 11189 }
 "_id" : ObjectId("5ad60c82ce94169d497139b5"), "length" : 11059 }
 "_id" : ObjectId("5ad60c81ce94169d49707f2b"), "length" : 10889 }
 "_id" : ObjectId("5ad60c82ce94169d49717b3f"), "length" : 10233 }
 "_id" : ObjectId("5ad60c83ce94169d4971f996"), "length" : 10205 }
 "_id" : ObjectId("5ad60c84ce94169d4972320a"), "length" : 9945 }
 "_id" : ObjectId("5ad60c82ce94169d49710ed1"), "length" : 9736 }
 "_id" : ObjectId("5ad60c80ce94169d497055b7"), "length" : 9678 }
 "_id" : ObjectId("5ad60c80ce94169d49704e4f"), "length" : 9547 }
 "_id" : ObjectId("5ad60c81ce94169d4970a1f0"), "length" : 9525 }
 "_id" : ObjectId("5ad60c81ce94169d4970c206"), "length" : 9512 }
 "_id" : ObjectId("5ad60c84ce94169d49724c28"), "length" : 9366 }
 "_id" : ObjectId("5ad60c84ce94169d49728cdd"), "length" : 9313 }
 "_id" : ObjectId("5ad60c84ce94169d4972799b"), "length" : 9312 }
 "_id" : ObjectId("5ad60c81ce94169d4970ca88"), "length" : 9231 }
ype "it" for more
```

II. NON-TEXT BASED:

```
db.grocery_and_gourmet_food.aggregate([    {$project:{ "_id" :1 , "Help": { $divide: [ {
$arrayElemAt: [ "$helpful", 0 ] },{$cond: { if: { $ne: [ { $arrayElemAt: [ "$helpful", 1 ] }, 0 ]
}, then: { $arrayElemAt: [ "$helpful", 1 ] }, else:  1000}}  ]}}}  ])
```



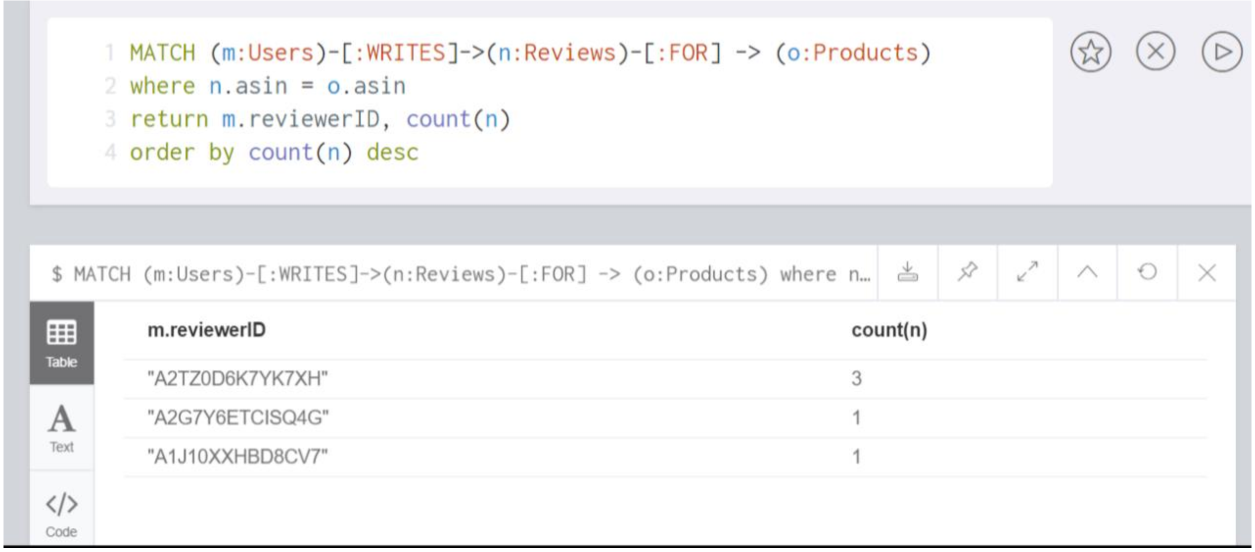
```
> db.grocery_and_gourmet_food.aggregate([    {$project:{ "_id" :1 , "Help": { $divide: [ { $arrayElemAt: [ "$helpful", 0
 ] },{$cond: { if: { $ne: [ { $arrayElemAt: [ "$helpful", 1 ] }, 0 ] }, then: { $arrayElemAt: [ "$helpful", 1 ] }, else:
  1000}}  ]}}}  ])
{ "_id" : ObjectId("5ad60c80ce94169d49704dd1"), "Help" : 0.75 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd2"), "Help" : 0.6666666666666666 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd3"), "Help" : 0.5 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd4"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd5"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd6"), "Help" : 0.75 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd7"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd8"), "Help" : 0.6666666666666666 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dd9"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dda"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704ddb"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704ddc"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704ddd"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704dde"), "Help" : 0.5 }
{ "_id" : ObjectId("5ad60c80ce94169d49704ddf"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704de0"), "Help" : 0.5 }
{ "_id" : ObjectId("5ad60c80ce94169d49704de1"), "Help" : 0.5 }
{ "_id" : ObjectId("5ad60c80ce94169d49704de2"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704de3"), "Help" : 0 }
{ "_id" : ObjectId("5ad60c80ce94169d49704de4"), "Help" : 1 }
Type "it" for more
```

We consider the helpfulness of a review text-based will be based on the length of the text they write.

Non-text-based: For each review, there is number of users that found the review useful and not useful. Therefore, to calculate the helpfulness of a review we use the no. of people who found the review useful/ total number of users who gave their opinion.

This is a very useful metric that we can rely on to show the reviews in the order of helpfulness as they are voted by other reviewers.

3.8 MULTIPLE REVIEWS FOR ONE PRODUCT BY A USER.



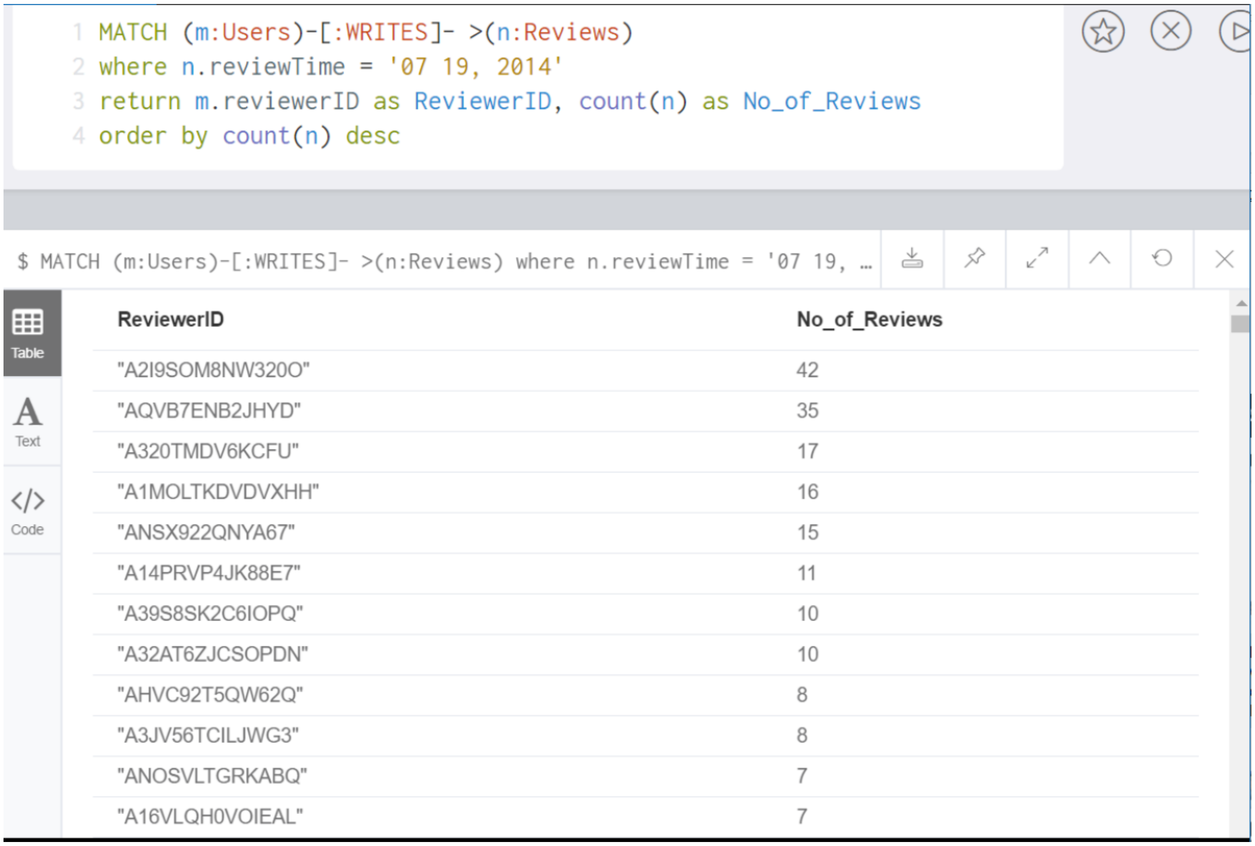
```
MATCH (m:Users)-[:WRITES]->(n:Reviews)-[:FOR] -> (o:Products)
where n.asin = o.asin
return m.reviewerID, count(n)
order by count(n) desc
```

```
$ MATCH (m:Users)-[:WRITES]->(n:Reviews)-[:FOR] -> (o:Products) where n…
```

m.reviewerID	count(n)
"A2TZ0D6K7YK7XH"	3
"A2G7Y6ETCISQ4G"	1
"A1J10XXHBD8CV7"	1

Generally, a user doesn’t write multiple reviews for the same product. There can be cases when a user might write multiple reviews but that may be rare. And that user’s reviews may not be reliable.

3.9 NUMBER OF REVIEWS BY A PERSON ON A DAY

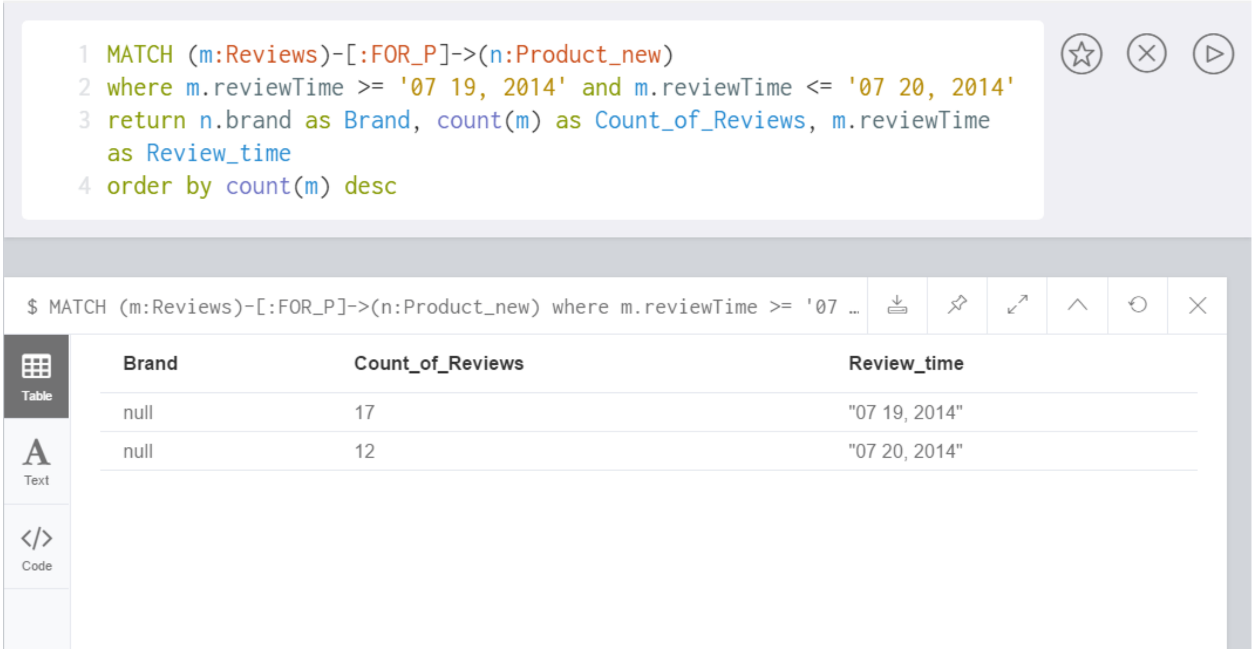


```
MATCH (m:Users)-[:WRITES]- >(n:Reviews)
where n.reviewTime = '07 19, 2014'
return m.reviewerID as ReviewerID, count(n) as No_of_Reviews
order by count(n) desc
```

```
$ MATCH (m:Users)-[:WRITES]- >(n:Reviews) where n.reviewTime = '07 19, …
```

ReviewerID	No_of_Reviews
"A2I9SOM8NW320O"	42
"AQVB7ENB2JHYD"	35
"A320TMDV6KCFU"	17
"A1MOLTKDVDVXHH"	16
"ANSX922QNYA67"	15
"A14PRVP4JK88E7"	11
"A39S8SK2C6IOPQ"	10
"A32AT6ZJCSOPDN"	10
"AHVC92T5QW62Q"	8
"A3JV56TCILJWG3"	8
"ANOSVLTGRKABQ"	7
"A16VLQH0VOIEAL"	7

3.10 ITEMS GETTING HIGH NUMBER OF REVIEWS IN A SPAN OF TWO DAYS



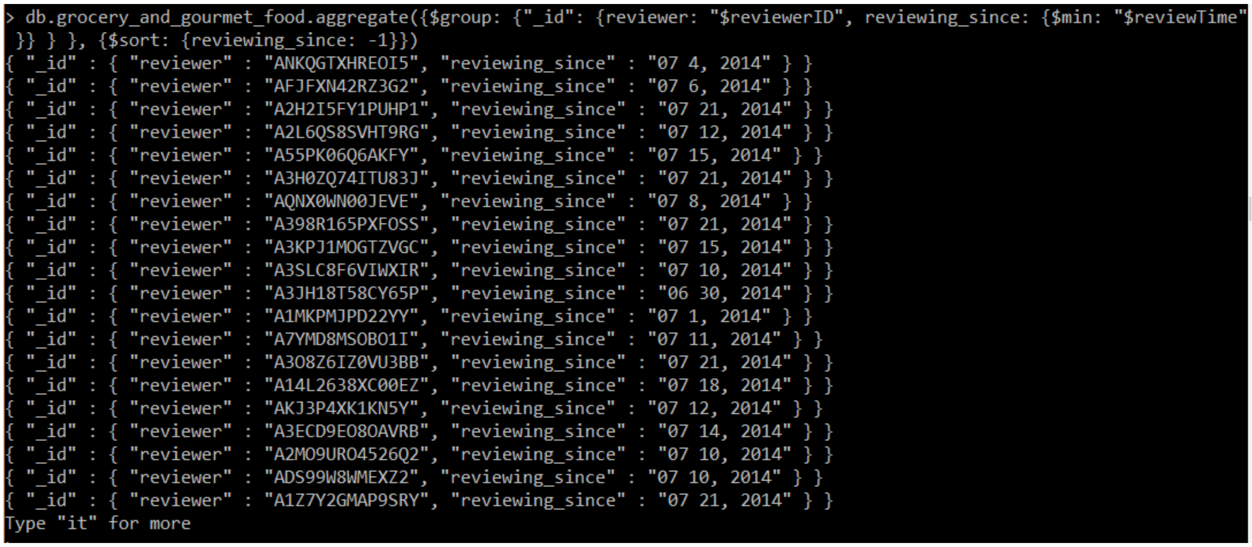
```
MATCH (m:Reviews)-[:FOR_P]->(n:Product_new)
where m.reviewTime >= '07 19, 2014' and m.reviewTime <= '07 20, 2014'
return n.brand as Brand, count(m) as Count_of_Reviews, m.reviewTime
as Review_time
order by count(m) desc
```

```
$ MATCH (m:Reviews)-[:FOR_P]->(n:Product_new) where m.reviewTime >= '07 …
```

Brand	Count_of_Reviews	Review_time
null	17	"07 19, 2014"
null	12	"07 20, 2014"

3.11 REVIEWER REVIEWING SINCE

db.grocery_and_gourmet_food.aggregate({$group: {"_id": {reviewer: "$reviewerID", reviewing_since: {$min: "$reviewTime" }} } }, {$sort: {reviewing_since: -1}})

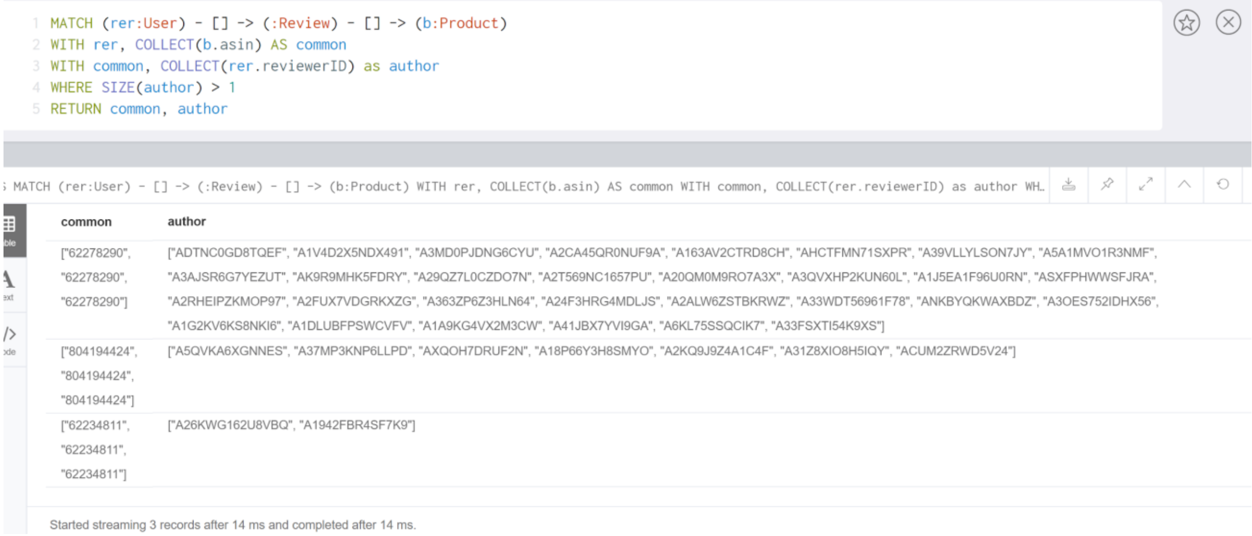


```
> db.grocery_and_gourmet_food.aggregate({$group: {"_id": {reviewer: "$reviewerID", reviewing_since: {$min: "$reviewTime"
 }} } }, {$sort: {reviewing_since: -1}})
{ "_id" : { "reviewer" : "ANKQGTXHREOI5", "reviewing_since" : "07 4, 2014" } }
{ "_id" : { "reviewer" : "AFJFXN42RZ3G2", "reviewing_since" : "07 6, 2014" } }
{ "_id" : { "reviewer" : "A2H2I5FY1PUHP1", "reviewing_since" : "07 21, 2014" } }
{ "_id" : { "reviewer" : "A2L6QS8SVHT9RG", "reviewing_since" : "07 12, 2014" } }
{ "_id" : { "reviewer" : "A55PK06Q6AKFY", "reviewing_since" : "07 15, 2014" } }
{ "_id" : { "reviewer" : "A3H0ZQ74ITU83J", "reviewing_since" : "07 21, 2014" } }
{ "_id" : { "reviewer" : "AQNX0WN00JEVE", "reviewing_since" : "07 8, 2014" } }
{ "_id" : { "reviewer" : "A398R165PXFOSS", "reviewing_since" : "07 21, 2014" } }
{ "_id" : { "reviewer" : "A3KPJ1MOGTZVGC", "reviewing_since" : "07 15, 2014" } }
{ "_id" : { "reviewer" : "A3SLC8F6VIWXIR", "reviewing_since" : "07 10, 2014" } }
{ "_id" : { "reviewer" : "A3JH18T58CY65P", "reviewing_since" : "06 30, 2014" } }
{ "_id" : { "reviewer" : "A1MKPMJPD22YY", "reviewing_since" : "07 1, 2014" } }
{ "_id" : { "reviewer" : "A7YMD8MSOBO1I", "reviewing_since" : "07 11, 2014" } }
{ "_id" : { "reviewer" : "A3O8Z6IZ0VU3BB", "reviewing_since" : "07 21, 2014" } }
{ "_id" : { "reviewer" : "A14L2638XC00EZ", "reviewing_since" : "07 18, 2014" } }
{ "_id" : { "reviewer" : "AKJ3P4XK1KN5Y", "reviewing_since" : "07 12, 2014" } }
{ "_id" : { "reviewer" : "A3ECD9EO8OAVRB", "reviewing_since" : "07 14, 2014" } }
{ "_id" : { "reviewer" : "A2MO9URO4526Q2", "reviewing_since" : "07 10, 2014" } }
{ "_id" : { "reviewer" : "ADS99W8WMEXZ2", "reviewing_since" : "07 10, 2014" } }
{ "_id" : { "reviewer" : "A1Z7Y2GMAP9SRY", "reviewing_since" : "07 21, 2014" } }
Type "it" for more
```

This is a metric that we will show on the representative reviewer's page.

3.12 REVIEWERS REVIEW AS A MOB (REVIEWING THE COMMON SET OF PRODUCTS). DO YOU FIND THIS BEHAVIOR IN THIS DATASET? RUN THE QUERIES AND DERIVE THE RESULTS.

```
MATCH (rer:User) - [] -> (:Review) - [] -> (b:Product)

WITH rer, COLLECT(b.asin) AS common

WITH common, COLLECT(rer.reviewerID) as author

WHERE SIZE(author) > 1

RETURN common, author
```



```
MATCH (rer:User) - [] -> (:Review) - [] -> (b:Product)
WITH rer, COLLECT(b.asin) AS common
WITH common, COLLECT(rer.reviewerID) as author
WHERE SIZE(author) > 1
RETURN common, author
```

```
; MATCH (rer:User) - [] -> (:Review) - [] -> (b:Product) WITH rer, COLLECT(b.asin) AS common WITH common, COLLECT(rer.reviewerID) as author WH…
```

common	author
["62278290", "62278290", "62278290"]	["ADTNC0GD8TQEF", "A1V4D2X5NDX491", "A3MD0PJDNG6CYU", "A2CA45QR0NUF9A", "A163AV2CTRD8CH", "AHCTFMN71SXPR", "A39VLLYLSON7JY", "A5A1MVO1R3NMF", "A3AJSR6G7YEZUT", "AK9R9MHK5FDRY", "A29QZ7L0CZDO7N", "A2T569NC1657PU", "A20QM0M9RO7A3X", "A3QVXHP2KUN60L", "A1J5EA1F96U0RN", "ASXFPHWWSFJRA", "A2RHEIPZKMOP97", "A2FUX7VDGRKXZG", "A363ZP6Z3HLN64", "A24F3HRG4MDLJS", "A2ALW6ZSTBKRWZ", "A33WDT56961F78", "ANKBYQKWAXBDZ", "A3OES752IDHX56", "A1G2KV6KS8NKI6", "A1DLUBFPSWCVFV", "A1A9KG4VX2M3CW", "A41JBX7YVI9GA", "A6KL75SSQCIK7", "A33FSXTI54K9XS"]
["804194424", "804194424", "804194424"]	["A5QVKA6XGNNES", "A37MP3KNP6LLPD", "AXQOH7DRUF2N", "A18P66Y3H8SMYO", "A2KQ9J9Z4A1C4F", "A31Z8XIO8H5IQY", "ACUM2ZRWD5V24"]
["62234811", "62234811", "62234811"]	["A26KWG162U8VBQ", "A1942FBR4SF7K9"]

Started streaming 3 records after 14 ms and completed after 14 ms.

3.13 SAS

I. Identification of Fake Reviews through SAS

A. Analysis

On Amazon, customer comments can help a product surge in popularity. The online retail giant says that more than 99 percent of its reviews are legitimate because they are written by real shoppers who aren't paid for them.

A Washington Post examination found that for some popular product categories, such as Bluetooth headphones and speakers, the vast majority of reviews appear to violate Amazon's prohibition on paid reviews. Such reviews have certain characteristics, such as repetitive wording that people probably cut and paste in.

Input for the analysis is Books.csv

B. Product Reviews: Not as Unbiased as You Think

Do you trust every online product review you read? Including those glowing five-star reviews? What about the angry one-star reviews?

Or perhaps only verified purchases are credible? The reality is, deciding which consumer reviews to trust or not trust has become so difficult for shoppers.

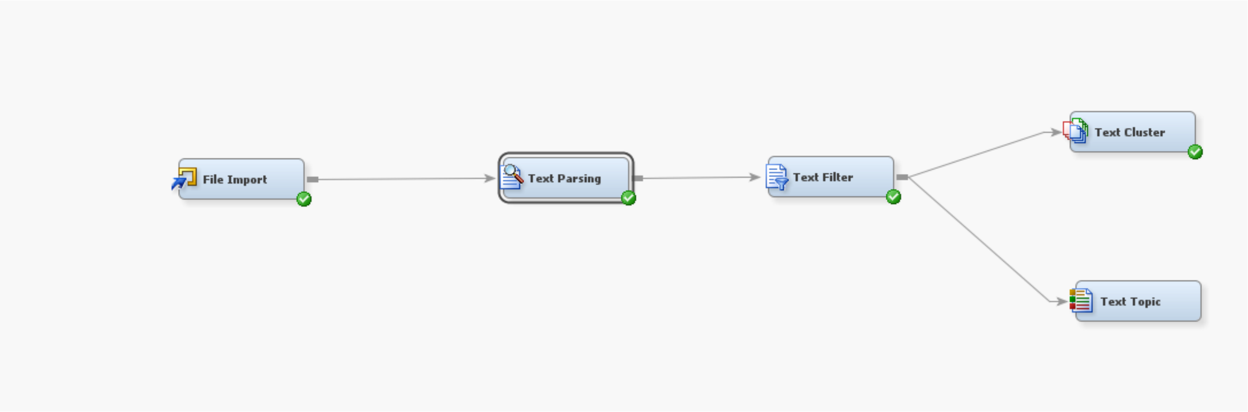
With SAS we attempt to help reviewers… review the reviews!

Review the reviews!

Fake reviews are usually those people who have not made any use of a service, i.e., buy a product, visit a restaurant etc. This can happen if someone is trying to either promote their own products or to demote their competitors'.

II. PROCESS

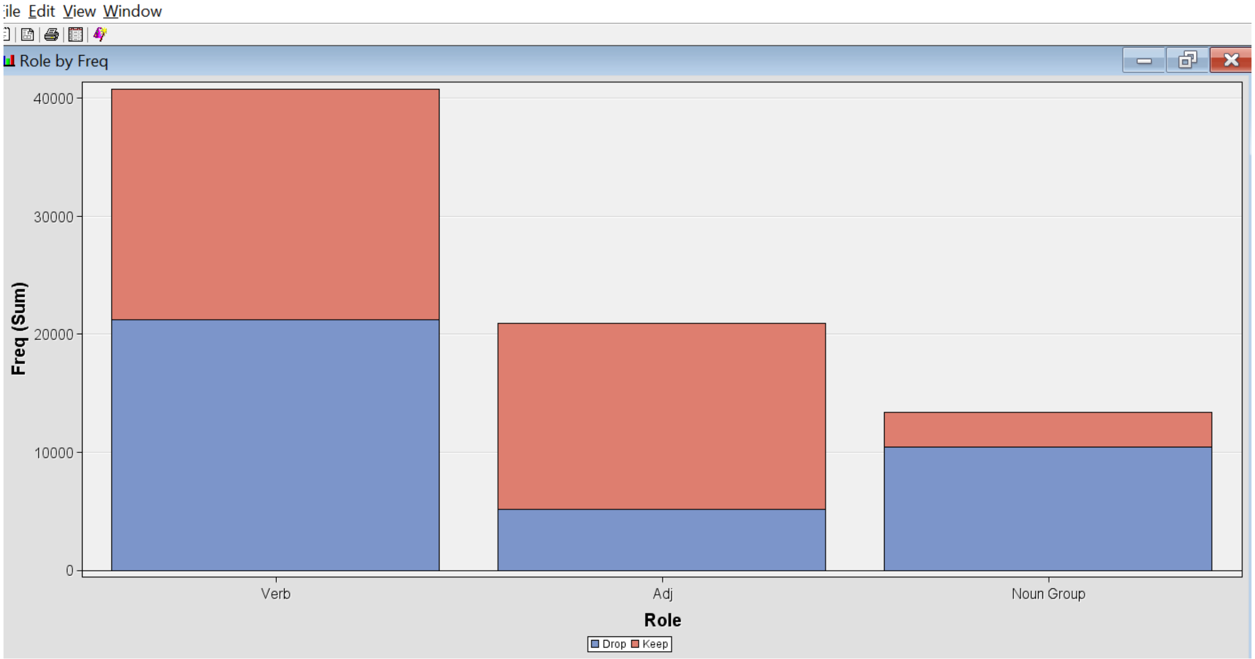
Adding the dataset as input to SAS Enterprise miner we parsed the text in the Review Text field. Next we filtered it using the text filter node.



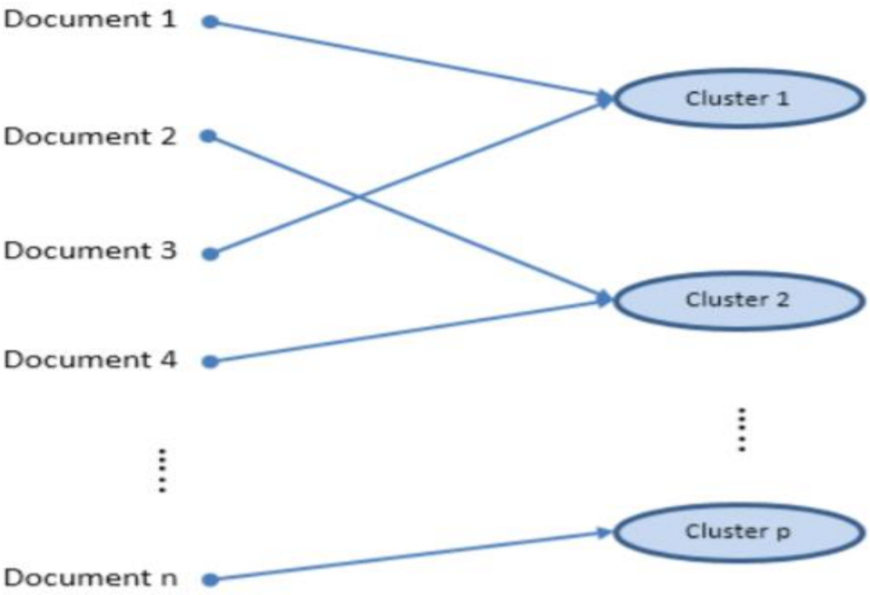
We propose this model for prediction of fake reviews

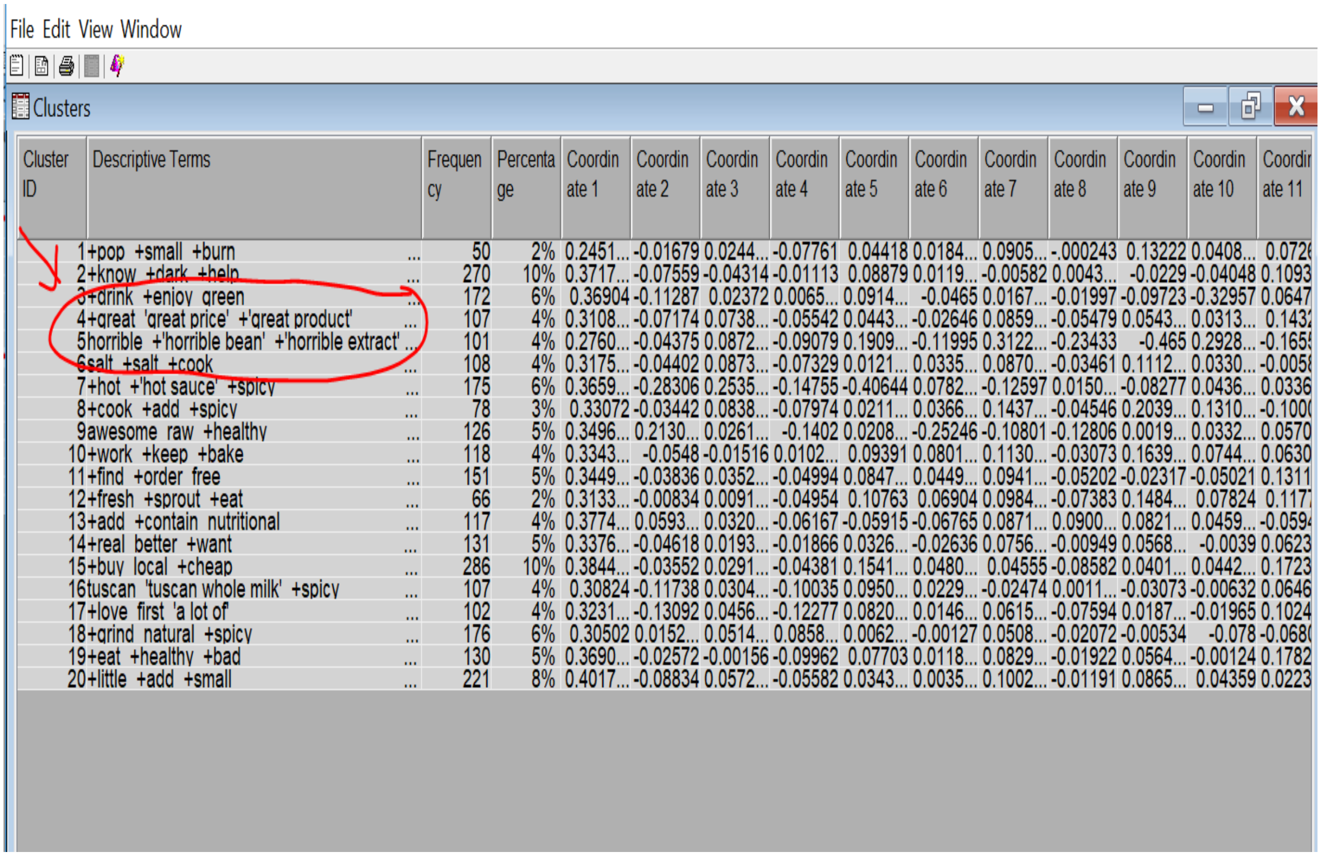
Step 1 : We parse the documents

Step 2: After Parsing the documents, in the filter part we only take the adjectives and adverbs and leave all other text from reviewText.



In the above picture, we analyse only the adjectives and adverbs .We have dropped other irrelevant words from our analysis.

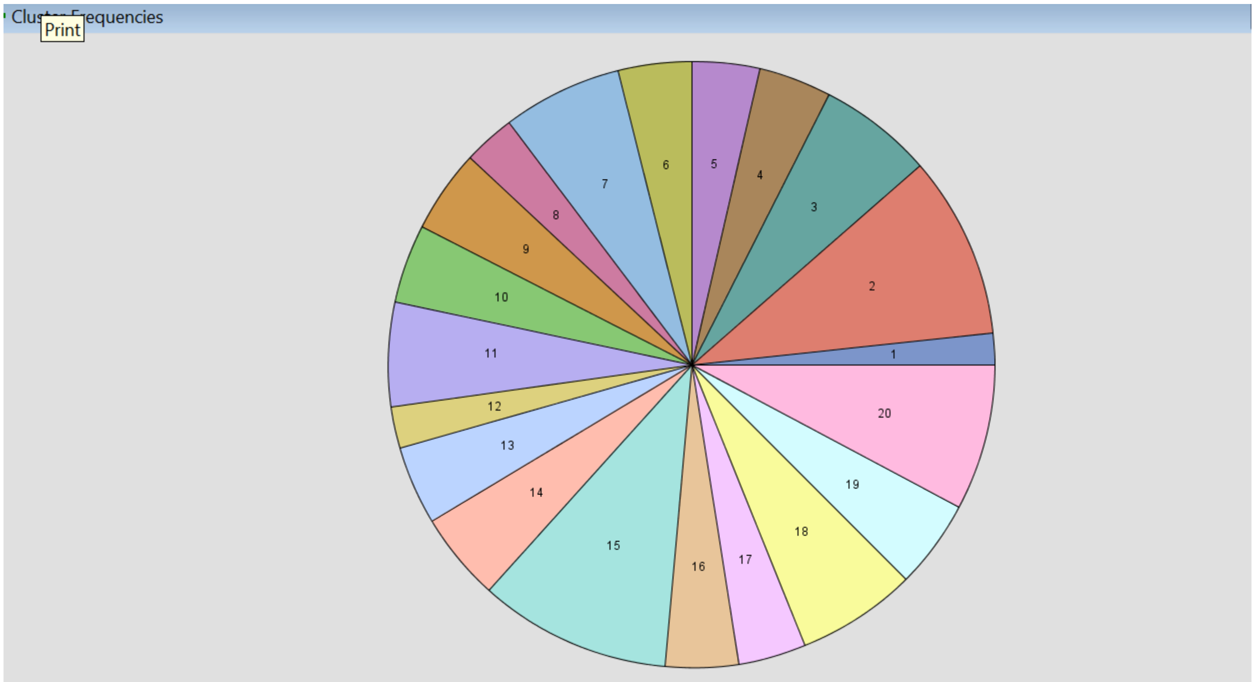




Cluster ID	Descriptive Terms	Frequency	Percentage	Coordinate 1	Coordinate 2	Coordinate 3	Coordinate 4	Coordinate 5	Coordinate 6	Coordinate 7	Coordinate 8	Coordinate 9	Coordinate 10	Coordinate 11
1	+pop +small +burn ...	50	2%	0.2451...	-0.01679	0.0244...	-0.07761	0.04418	0.0184...	0.0905...	-.000243	0.13222	0.0408...	0.072
2	+know +dark +help ...	270	10%	0.3717...	-0.07559	-0.04314	-0.01113	0.08879	0.0119...	-0.00582	0.0043...	-0.0229	-0.04048	0.1093
3	+drink +enjoy green ...	172	6%	0.36904	-0.11287	0.02372	0.0065...	0.0914...	-0.0465	0.0167...	-0.01997	-0.09723	-0.32957	0.0647
4	+great 'great price' +'great product' ...	107	4%	0.3108...	-0.07174	0.0738...	-0.05542	0.0443...	-0.02646	0.0859...	-0.05479	0.0543...	0.0313...	0.143
5	horrible +'horrible bean' +'horrible extract' ...	101	4%	0.2760...	-0.04375	0.0872...	-0.09079	0.1909...	-0.11995	0.3122...	-0.23433	-0.465	0.2928...	-0.165
6	salt +salt +cook ...	108	4%	0.3175...	-0.04402	0.0873...	-0.07329	0.0121...	0.0335...	0.0870...	-0.03461	0.1112...	0.0330...	-0.005
7	+hot +'hot sauce' +spicy ...	175	6%	0.3659...	-0.28306	0.2535...	-0.14755	-0.40644	0.0782...	-0.12597	0.0150...	-0.08277	0.0436...	0.0336
8	+cook +add +spicy ...	78	3%	0.33072	-0.03442	0.0838...	-0.07974	0.0211...	0.0366...	0.1437...	-0.04546	0.2039...	0.1310...	-0.100
9	awesome raw +healthy ...	126	5%	0.3496...	0.2130...	0.0261...	-0.1402	0.0208...	-0.25246	-0.10801	-0.12806	0.0019...	0.0332...	0.0570
10	+work +keep +bake ...	118	4%	0.3343...	-0.0548	-0.01516	0.0102...	0.09391	0.0801...	0.1130...	-0.03073	0.1639...	0.0744...	0.0630
11	+find +order free ...	151	5%	0.3449...	-0.03836	0.0352...	-0.04994	0.0847...	0.0449...	0.0941...	-0.05202	-0.02317	-0.05021	0.1311
12	+fresh +sprout +eat ...	66	2%	0.3133...	-0.00834	0.0091...	-0.04954	0.10763	0.06904	0.0984...	-0.07383	0.1484...	0.07824	0.117
13	+add +contain nutritional ...	117	4%	0.3774...	0.0593...	0.0320...	-0.06167	-0.05915	-0.06765	0.0871...	0.0900...	0.0821...	0.0459...	-0.059
14	+real better +want ...	131	5%	0.3376...	-0.04618	0.0193...	-0.01866	0.0326...	-0.02636	0.0756...	-0.00949	0.0568...	-0.0039	0.0623
15	+buy local +cheap ...	286	10%	0.3844...	-0.03552	0.0291...	-0.04381	0.1541...	0.0480...	0.04555	-0.08582	0.0401...	0.0442...	0.1723
16	tuscan 'tuscan whole milk' +spicy ...	107	4%	0.30824	-0.11738	0.0304...	-0.10035	0.0950...	0.0229...	-0.02474	0.0011...	-0.03073	-0.00632	0.0646
17	+love first 'a lot of' ...	102	4%	0.3231...	-0.13092	0.0456...	-0.12277	0.0820...	0.0146...	0.0615...	-0.07594	0.0187...	-0.01965	0.1024
18	+grind natural +spicy ...	176	6%	0.30502	0.0152...	0.0514...	0.0858...	0.0062...	-0.00127	0.0508...	-0.02072	-0.00534	-0.078	-0.068
19	+eat +healthy +bad ...	130	5%	0.3690...	-0.02572	-0.00156	-0.09962	0.07703	0.0118...	0.0829...	-0.01922	0.0564...	-0.00124	0.1782
20	+little +add +small ...	221	8%	0.4017...	-0.08834	0.0572...	-0.05582	0.0343...	0.0035...	0.1002...	-0.01191	0.0865...	0.04359	0.0223

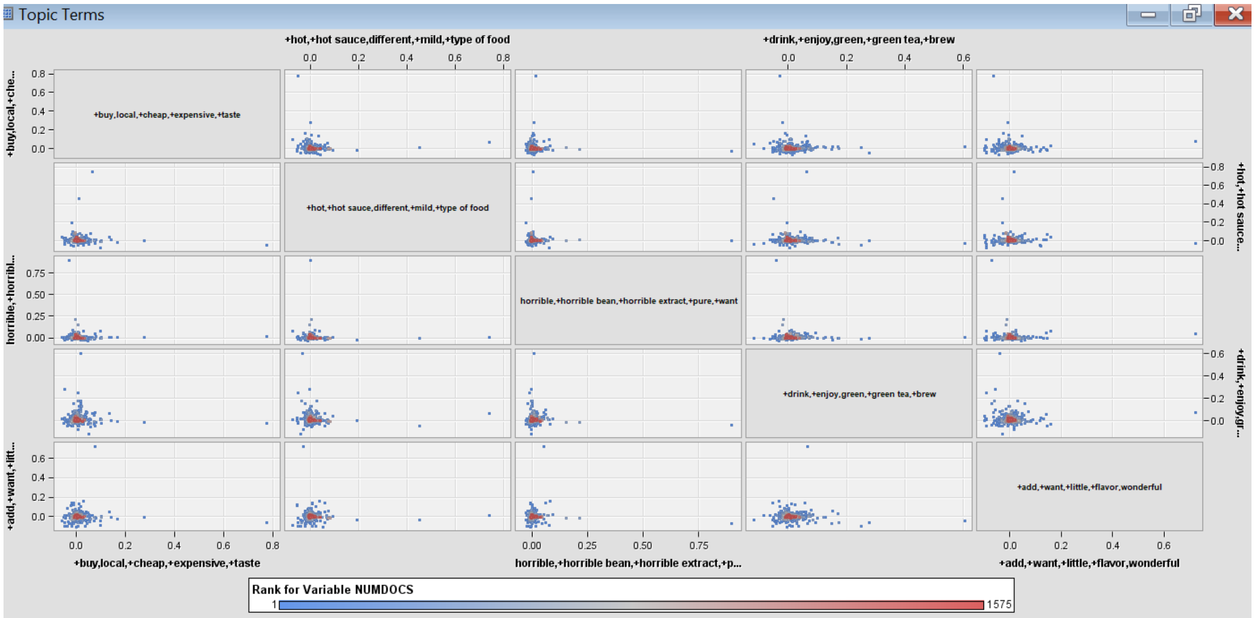
In the next step of Text cluster we identify the groups where the word are extreme negative or extreme positive, the reviews that contain these set of words(cluster of words) are the ones which may potentially be made by a fake reviewer.

Also, the reviews just describe a product with an adjective and no supplemental description of the product hence potentially not being very useful for the customer.



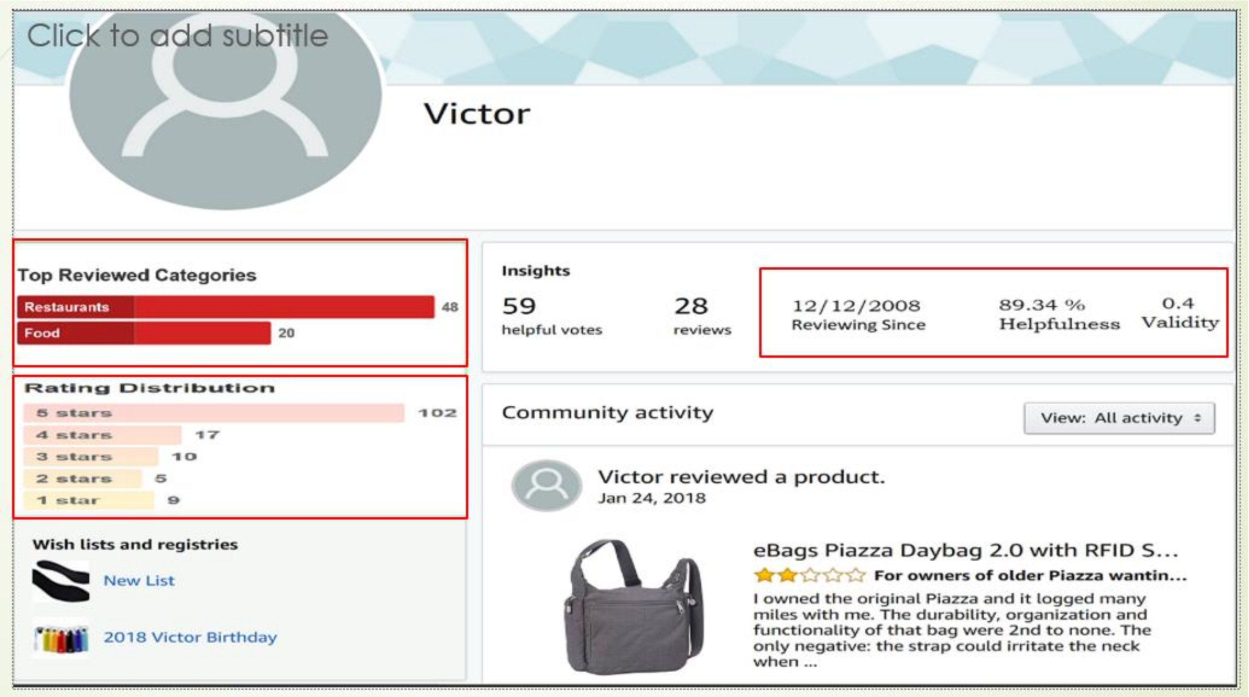
This pie chart represents the frequencies of the clusters which we are considering to be containing fake reviews. They are clusters 4 and 5.

The two categories “great + great price + great product” and “horrible + horrible bean + horrible extract” are the ones containing fake reviews.



4. Design Template

For representative customer:



Click to add subtitle

Victor

Top Reviewed Categories

Restaurants	48
Food	20

Insights

59 helpful votes

28 reviews

12/12/2008 Reviewing Since

89.34 % Helpfulness

0.4 Validity

Rating Distribution

5 stars	102
4 stars	17
3 stars	10
2 stars	5
1 star	9

Community activity

Victor reviewed a product.
Jan 24, 2018

eBags Piazza Daybag 2.0 with RFID S...

For owners of older Piazza wantin...

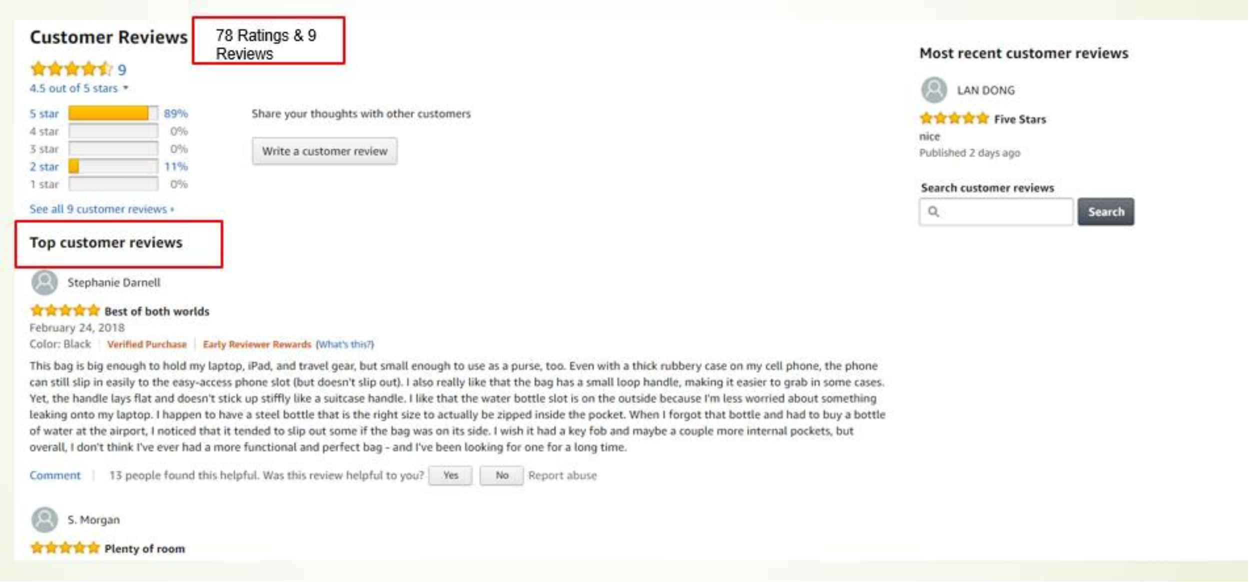
I owned the original Piazza and it logged many miles with me. The durability, organization and functionality of that bag were 2nd to none. The only negative: the strap could irritate the neck when ...

Wish lists and registries

New List

2018 Victor Birthday

For representative product:



Customer Reviews

78 Ratings & 9 Reviews

9
4.5 out of 5 stars

5 star	89%
4 star	0%
3 star	0%
2 star	11%
1 star	0%

See all 9 customer reviews ›

Share your thoughts with other customers

Top customer reviews

Stephanie Darnell

Best of both worlds
February 24, 2018
Color: Black | Verified Purchase | Early Reviewer Rewards (What's this?)

This bag is big enough to hold my laptop, iPad, and travel gear, but small enough to use as a purse, too. Even with a thick rubbery case on my cell phone, the phone can still slip in easily to the easy-access phone slot (but doesn't slip out). I also really like that the bag has a small loop handle, making it easier to grab in some cases. Yet, the handle lays flat and doesn't stick up stiffly like a suitcase handle. I like that the water bottle slot is on the outside because I'm less worried about something leaking onto my laptop. I happen to have a steel bottle that is the right size to actually be zipped inside the pocket. When I forgot that bottle and had to buy a bottle of water at the airport, I noticed that it tended to slip out some if the bag was on its side. I wish it had a key fob and maybe a couple more internal pockets, but overall, I don't think I've ever had a more functional and perfect bag - and I've been looking for one for a long time.

Comment | 13 people found this helpful. Was this review helpful to you?

S. Morgan

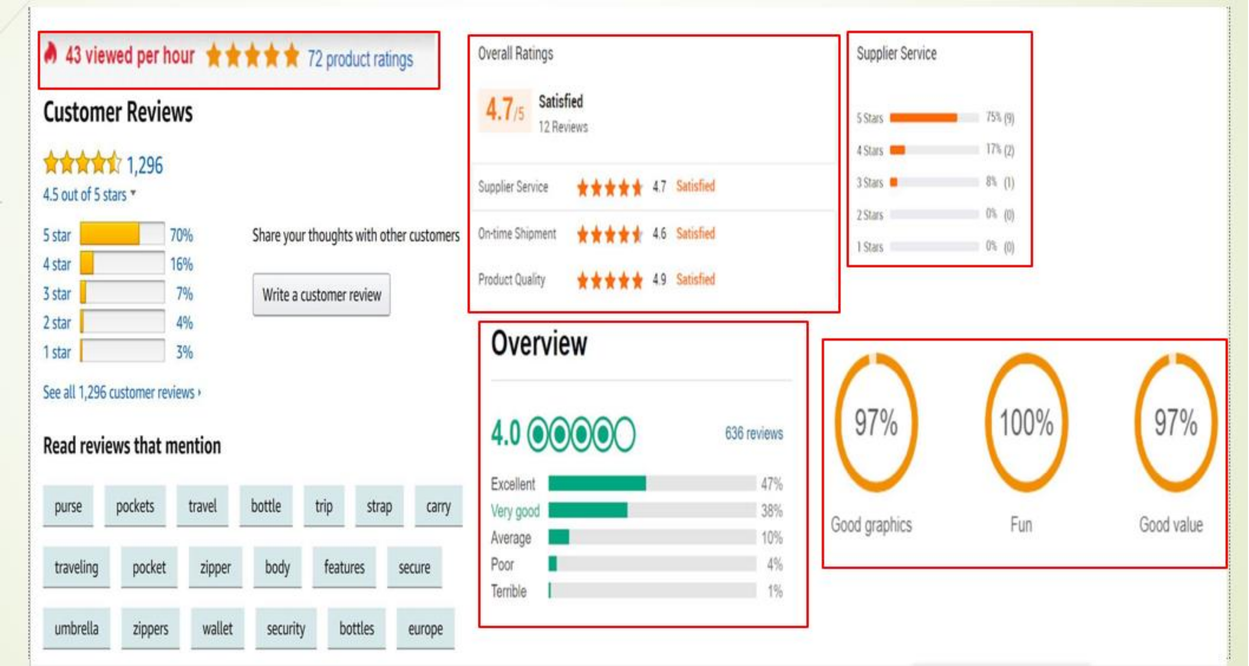
Plenty of room

Most recent customer reviews

LAN DONG

Five Stars
nice
Published 2 days ago

Search customer reviews



43 viewed per hour ★★★★★ 72 product ratings

Customer Reviews

★★★★☆ 1,296

4.5 out of 5 stars

5 star	70%
4 star	16%
3 star	7%
2 star	4%
1 star	3%

See all 1,296 customer reviews ›

Share your thoughts with other customers

Write a customer review

Read reviews that mention

purse · pockets · travel · bottle · trip · strap · carry · traveling · pocket · zipper · body · features · secure · umbrella · zippers · wallet · security · bottles · europe

Overall Ratings

4.7/5 **Satisfied** 12 Reviews

Supplier Service	★★★★★	4.7	Satisfied
On-time Shipment	★★★★★	4.6	Satisfied
Product Quality	★★★★★	4.9	Satisfied

Supplier Service

5 Stars	75% (9)
4 Stars	17% (2)
3 Stars	8% (1)
2 Stars	0% (0)
1 Stars	0% (0)

Overview

4.0 ●●●●○ 636 reviews

Excellent	47%
Very good	38%
Average	10%
Poor	4%
Terrible	1%

Good graphics	Fun	Good value
97%	100%	97%