Retail/Buyer Analytics for Online Shoppers

Papori Kalita
Student
Department of MCA
The Oxford College of Engineering, Bomannahalli, Hosur Road, Bangalore-68

Abstract— In recent times as the power has shifted to customers, the shopping experience has changed greatly. Even while walking on the streets shoppers can search and compare products from any device. They can easily share their reviews on vendors and products they are using through social media sites and can create an impression about the product on other prospective customers. Retailers need to use new ideas to draw attention of the customers to face this new multi-channel environment. Hadoop and Big data allows vendors to associate with customers using various channels at a new level by controlling the enormous size of new data available today. The MapReduce(MapR) Distribution for Apache Hadoop helps retailers save, combine and examine a wide range of online and offline consumers data—online transactions, mouse click data, email, social sites and call centre records—everything in a central database. Retailers can examine this data to generate report on individual consumer nature and preferences, and based on the report retailers can offer personalized recommendations in real time.

Key words: Hadoop, MapReduce(MapR), Social Media Sites.

I. INTRODUCTION

Various innovations of technology forces to change the customer's needs and a delayed period of slow economic growth are making accelerated changes in the retail industry. Retailers can respond to that change using Apache Hadoop, which considers as a powerful tool. Technology gave consumers the opportunity to be smarter and shop online. Online reviews, ratings, and comparisons among products and prices, which helps a customer in making decision while shopping .When it comes time to buy, a customer can purchase the products on online, a mobile device or in a store. Analysis of customers is easy using Hadoop. Retailers store a huge range of data of customers from various sources previously unavailable at a huge scale, such as blogs, online reviews and social media. Hadoop deliver retailers records of customer behavior and preferences. It can also observe customers over multiple medium and touch points to prepare in-store or online personalized suggestions, along with real-time offers.

Retailers can examine this data to generate report on individual consumer nature and preferences, and based on the report retailers can offer personalized recommendations in real time.

The opportunity to achieve competitive advantage from "analytical retail" is enormous. With these analytical tools, retailers can:

- Develop close relationships with customers based on a deep understanding of their behaviors and needs;
- Deliver the targeted advertising, promotions and product offers to customers that will motivate them to buy;
- Balance inventory with demand, so you're never out of stock or carrying excess inventory;
- Charge exactly the price that customers are willing to pay at any moment;
- Determine the best use of marketing investments;
- Locate stores, distribution centers, and other facilities in optimal locations.

II. APPLICATION OF RETAIL ANALYSIS

The application of analysis of stored data can be in many ways:

- In retail environment of high-speed order fulfillment and rising product costs, it is more important than ever for retailers to have a real-time overview of what is happening in the supply chain. This critical information enables retailers to make rapid decisions that reduce transportation costs, boost inventory productivity levels, streamline distribution center processes, and improve supplier performance and negotiations.
- Loss prevention (LP) is a key bottom line issue for any retailer. The combined financial costs associated with shrinkage, theft, and fraud can be substantial. By drilling down and analyzing data, LP managers can spot abnormalities, take immediate action, and significantly cut down on losses associated with inventory shrinkage.
- Retailers face lots of trouble to balance huge size of record levels with in-stock positions while maintaining the variety of data that are proper and fresh. Retailers who are capable of responding to customers wishes while controlling the huge amount of data have a competitive edge in the market. By grasping analytics, Retailers are able to make quicker, more valid decisions which leads to improvement in-stock rates and category management, increase in sales and gross margins, and better vendor association.
- In retail environment, understanding the customer can be a big differentiator for successful businesses. It is critical to leverage every consumer touch point to gain insights on how best to serve and engage customers to maximize profits and improve customer loyalty. With this data, retailers can create more targeted promotions, tailor store assortments to specific clientele, and create a superior, personalized shopping experience for customers at all stages of the buying cycle.

III. DATA COLLECTION

The accuracy of predictive models depends on the quality of the datasets. There are total 440 records in the dataset.
Few number of dataset is given below as for example:

Channel	Region	Fresh	Milk	Grocery						
2	3	12669	9656	7561						
2	3	7057	9810	9568						
2	3	6353	8808	7684						
1	3	13265	1196	4221						
2	3	22615	5410	7198						
2	3	9413	8259	5126						
2	3	12126	3199	6975						
2	3	7579	4956	9426						
1	3	5963	3648	6192						

IV. Data Description

A. Types of Dataset:

a) Nominal:

Values of Data which doesn't share any relationship with other values known as Nominal Dataset.

b) Continuous:

A set of data is said to be continuous if the values belonging to the set can take on ANY value within a finite or infinite interval.

B. Attribute Information:

a) Fresh:

Annual spending on fresh products (Continuous);

b) Milk

Annual spending on milk products (Continuous);

c) Grocery:

Annual spending on grocery products (Continuous);

d) Frozen:

Annual spending on frozen products (Continuous)

e) Detergents Paper:

Annual spending on detergents and paper products (Continuous)

f) Delicatessen:

Annual spending on and delicatessen products (Continuous);

g) Channel:

Customers Channel – Feedback of the customers(Nominal)
Customers Satisfied: 1 or Yes
Customers Unsatisfied: 2 or No

h) Region:

Customers Region Mumbai, Bangalore or Other (Nominal)
Mumbai: 1
Bangalore: 2
Delhi: 3

V. Results And Outputs

A. Partitioning:

A partition function that gives an approximately uniform distribution of data per shared for load-balancing purposes,generally the Map Reduce operation can be held up sitting tight for moderate to wrap up. Between the guide and lessen stages, the information is rearranged keeping in mind the end goal to move the information from the guide hub that delivered it to the shard in which it will be lessened. The mix can at times take longer than the processing time contingent upon system data transfer capacity, CPU speeds, information delivered and time taken by guide and decrease calculations.


```
2015-04-07 20:57:25,678 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2015-04-07 20:57:25,683 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2015-04-07 20:57:25,716 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2015-04-07 20:57:25,717 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
((1,1),59)
((1,2),28)
((1,3),238)
((2,1),18)
((2,2),19)
((2,3),108)
((Channel,Region),1)
grunt>
```

Fig. 1: Result Of Partitioning The Dataset.

B. Grouping:

Grouping of data collects records with the same key from one or more inputs. When you group records, the result is a new records with two columns: "group" and the name of the original relation. The group column has the schema of what you grouped by. If you grouped by an integer column, the type will be int.

Fig. 2: Result for Grouping The Dataset

C. *Descriptive Analytics:*

Descriptive analytics takes a gander at information and dissects past occasions for understanding regarding how to approach what's to come. Enlightening examination takes a gander at past execution and comprehends that execution by mining verifiable information to search for the explanations for past achievement or disappointment. Descriptive models can be used, for example, to categorize customers by their product preferences and life stage. The figure below depicts the overall customer satisfaction with the product that is offered. The number of customers satisfied with the delivering is almost double as compared the unsatisfied
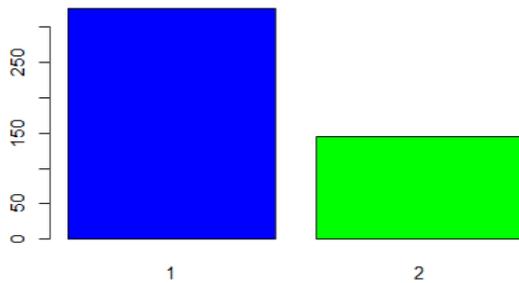
.
Fig. 3: Descriptive Analysis On Channel
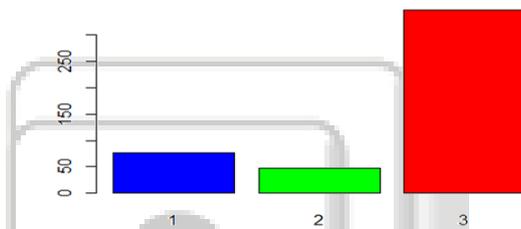
Fig. 4: Descriptive Analysis Of Region

As it is observed in the graph in figure 2 that city 3 has a higher concentration of buyers with higher disposable income. However city 1 and 2 even with similar disposable incomes is observed to have lower market penetration. Targeted advertisements or lucrative offers will provide the initial push for growth in these two cities.

VI. CONCLUSION

Studies of customer behavior in retail environment normally manage distinguishing proof of clients and their purchasing conduct designs. The point of such examination is to discover who purchases where, what, when and how. What's more, such research tries to find out about client reaction to deals advancement gadgets. The aftereffects of these studies are helpful in the arrangement of a variety of advertising issues.

Online shopping is more available than it has ever been in the advanced period with the presentation of tablets, cell phones, simple checkout frameworks and more secure frameworks. Therefore, the e-business industry is producing higher volumes.

The endeavor has been made here to show along what lines and with what techniques studies of customer behavior can be made in retail environment. Possible uses of the results for the solution of marketing problems have been suggested or indicated. To spell out fully and systematically the many practical uses to which such analysis can be put by producers, manufacturers and distributors would require a series of analysis. This analysis is intended to stimulate those who can gain the most from such studies to take advantage of the opportunities for learning more about customer behavior in the market-place.

REFERENCES

[1] Accenture Seamless Retail Capabilities Benchmark Study, December 2012.
[2] See, e.g., Retail Information Systems (RIS), next Steps for Cloud Retailing, May 2012.
[3] Accenture, What's next for Cloud in the CG&S Industry? February 2012.
[4] Seamless Retail Global Customer Survey, Retail Industry, November 2012.
[5] Microsoft case study: "Target: Large Retailer Relies on a Virtual Solution to Deliver Optimal Shopping Experience," March 17, 2011, http://www.microsoft.com/casestudies/Case_Study_Detail.aspx?casestudyid=400000940