Brand Switching Analysis using Data Analytics to Derive Consumer Behaviour

Ayush Tripathi,
Bachelors of Technology in Computer Science and
Engineering, Raj Kumar Goel Institute of Technology,
Ghaziabad (A.K.T.U.),
Uttar Pradesh, India

Zatin Gupta
Assistant Professor at Computer Science and Engineering
Department, Raj Kumar Goel Institute of Technology,
Ghaziabad (A.K.T.U.),
Uttar Pradesh, India

Abstract:- In this paper, we present Brand Switching Analysis using Data Analytics to derive various patterns of consumers' actions within a retail store. The paper demonstrates multiple steps involved in conducting this analysis – Data Cleansing, Data Visualization, Data Segregation, and Data Representation, using various technologies. The patterns and the inferences established from the research can be harnessed by the brands and retail stores for outlining their marketing strategy and targeting their potential customers.

Keywords:- Data Cleaning; Data Cleansing; Data Visualization; Data Segregation; Data Mining; Brand Switching; Data Analysis; Consumer Behaviour; Marketing Strategy; Sales.

I. INTRODUCTION

In our world, more and more brands and products are launched every day. Many retail stores observe the attrition of customers, and the reasons could be numerous – for example - increased choices and new brand launches in competition, better marketing promotions offered by the competitors, change in the placement of brand in a store display. These phenomena where a customer moves from purchasing from one brand of a product to buying a different brand of the same product is known as Brand Switching.

Retailers have huge data collected over a period of time, which can be used to learn about the customer's purchasing habits. Data Mining (turning raw data into useful information) and Data Analytics (analyzing raw data to make conclusions about that information) can be effectively used to analyze a large amount of business data and look for patterns in large batches of data. Through this process, businesses can learn more about their customers to develop more effective marketing strategies, increase sales, and decrease costs.

A retail store must predict it's customers' switching behaviors to sustain and retain its loyal customers. The Brand Switching Analysis can help brands to strategies their marketing to retain and attract potential customers, and enhance sales.

Our goal in this paper is to process the available data* into a meaningful format and conduct Brand Switching Analysis on the processed data to derive patterns of customer purchasing and *Data was downloaded from Kaggle switching behavior for a chosen brand (Paper Chain) within a retail store.

Further, the paper demonstrates the step by step process of data mining and analytics and the inferences derived from it.

The paper also reveals the limitations and the steps that can be taken in the future to scale the research further.

II. RESEARCH PROCESS AND INFERENCES

The dataset practiced in this study has been downloaded from Kaggle. The dataset comprises various information on customers purchasing within a retail store.

➢ *Data Storage*
The raw data* was in CSV format. Opensource technology - HDFS was used to store this data.

➢ *Data Cleansing*
The stored data were cleaned to remove all the null values. The data cleansing was essential to remove the null values (customer id, invoice numbers, brand id, null columns) present in the columns for accurate results.

➢ *Data Wrangling*
The dates in the raw dataset were in the string format. Timestamps and whitespaces were removed from the column values, and the dataset was converted in usable date format (YYYY-MM-DD).

Further, for comparison, the data was needed in two comparable periods. Example – YoY, Quarterly, Monthly, Event vs. Event. In this case, the contrast was made on YoY.

➢ *Data Segregation*
Segregation of data in two usable chunks - 2011 vs 2010. (refer Figure 1 and Figure 2).

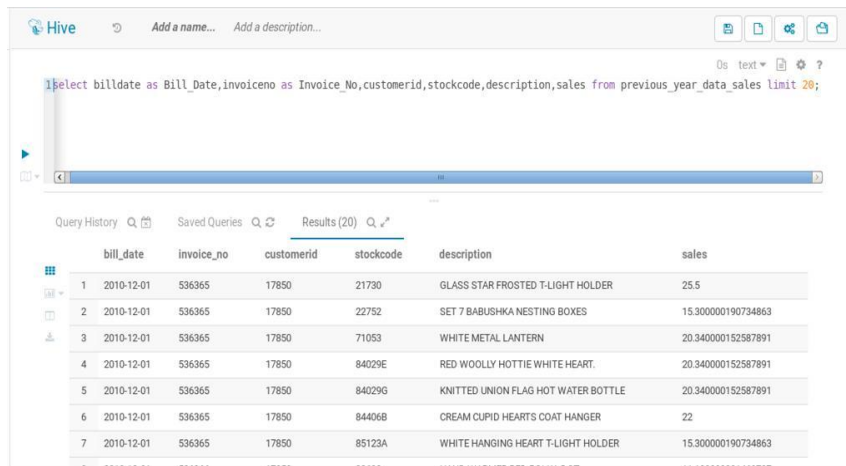The process of Data Cleansing, Wrangling and Segreagtion was done in Hive using HQL.

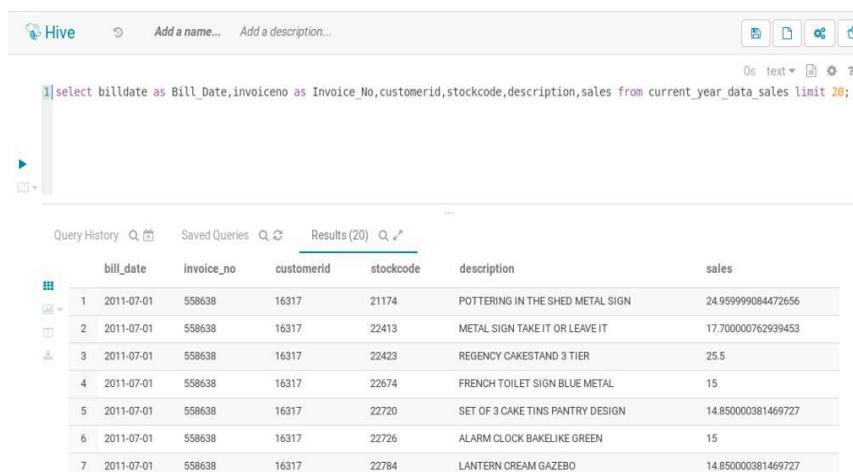Fig 1:- Data of 2010 (after Data Cleansing, Wrangling and Segregation)

Fig 2:- Data of 2011 (after Data Cleansing, Wrangling and Segregation)

> *Data Analysis and Comparison*

Data were further processed to get it into a format that can be used by Data Visualization tools for meaningful representations. A comparison of the Data was conducted between the two periods (2010 vs. 2011) to analyze the consumers in chunks - retained consumers, lost consumers, and new customers.

A connection was established with the Business Intelligence tool, and

Data Representation was done using Bar Charts to display consumer segregation of the chosen brand.

Further, An analysis was done on the brand Paper Chain (refer Figure 3). A significant switching between brands was observed between 2010 vs. 2011.

Fig 3:- Data Analysis conducted for Lost, New, and retained customers using the filter selection of the brand.
In this Bar Graph, the Customer movement pattern between 2010 vs. 2011 is shown in context to the paper chain brand.

➢ *Data Representation and Analysis*

The Data represented in Grid and Table formats (refer Figure 4 and 5) can further help analysts to spring various meaningful patterns. The below representations (refer Figure 4 and 5) clearly shows that:

1. The YoY Sales decreased in 2011 as compared to 2010 when examined in context to the lost and new customers.
2. The YoY Sales in the case of retained customers has increased in 2011 vs. 2010.
3. Overall, YoY Sales has marginally increased in 2011 due to the retained customers.
4. The number of lost customers is more than new customers in 2011.

Sales YoY

Customer Seggregation	Sales 2011	Sales 2010
Lost Customer		3852
New customer	3578	
Retained customer	4991	4322

Fig 4:- Sales YoY – Lost, new and retained customers (for Paper chain brand)

The Analysis depicts that the brand should focus on retaining its lost customers to enhance sales.

Similarly, the **Grid Analysis** (refer Figure 5) for a particular brand shows how the brand (Paper Chain Kit Retrospot, in this context) is Performing within the retail.

Grid Analysis

BrandID	Analyzed Brand	Customer Seggregation	Customer Count
22083	PAPER CHAIN KIT RETROSPOT	New customer	129
		Lost Customer	122
		Retained customer	43

Fig 5:- Grid Analysis corresponding to a Brand. In this context the brand is Paper Chain

As a retail brand owner, the curiosity lies in knowing where the customers are going to - a different store or - a distinct brand within the store. Similarly, for the new customers, the retailer is curious about where the customers are coming from and why.

The further study was conducted on the processed data to analyze the customer switching in detail, and Heat Maps (Figure 6, Figure 7) were used for Data Analysis and Representation.

The lost consumers were picked, and information on where the lost customers were buying in 2010 was derived using the Heap Map Analysis (refer Figure 6)

Similarly, Heat Map Analysis was conducted for new customers (refer Figure 7). The size of the Heat Map boxes for each brand is based on the number of consumers lost and gained, respectively. Besides, using the filter on the right side, various sample sizes can be compared to see the Heat Map.

The Heat Map Analysis conducted provides various information such as – where the consumer was previously buying, which brand they have switched to, etc.

Fig 6:- Heat Map representing the Lost Consumer Analysis (sample brand – 50) ; The data in the white box showcases the consumer who switched from one brand to the other

Fig 7:- Heat Map representing the New Consumer Analysis (sample brand – 50) ; The data in the white box showcases the consumer who switched from the other brand to the paper chain kit retrospot

III. TOOLS USED FOR DATA ANALYSIS

Hadoop Distributed File System – HDFS, The Hadoop Distributed File System (HDFS) is the primary data storage system used by Hadoop applications. It employs a NameNode and DataNode architecture to implement a distributed file system that provides high-performance access to data across highly scalable Hadoop clusters.

HDFS is a key part of the many Hadoop ecosystem technologies, as it provides a reliable means for managing pools of big data and supporting related big data analytics applications.

For this research paper we worked on Cloudera distribution on a Virtual Box.

HIVE is an open-source data warehousing solution built on top of Hadoop. Hive supports queries expressed in a SQL-like declarative language - HiveQL, which are compiled into map-reduce jobs that are executed using Hadoop. In addition, HiveQL enables users to plug in custom map-reduce scripts into queries.

The language includes a type system with support for tables containing primitive types, collections like arrays and maps, and nested compositions of the same.

Tableau is business intelligence software that helps people visualize and understand their data.

In this paper, bar graphs, heat maps were created using Tableau. Heat maps are a visualization where marks on a chart are represented as colors. As the marks "heat up" due their higher values or density of records, a more intense color is displayed. These colors can be displayed in a matrix / crosstab, which creates a highlight table, but can also be displayed on a geographical map or even a customized image – such as a webpage used to show where users are clicking.

IV. LIMITATIONS AND DIRECTIONS FOR THE FUTURE RESEARCH

This research paper was conducted on the available dataset, and hence the inference from the Brand Switching Analysis conducted is purely based on the accuracy of this data. In addition, the dataset available was only one and a half years for a retail store. Hence the patterns derived were based on this period. As the available data was limited, concluding future predictions couldn't be obtained.

Further, the hierarchal data wasn't available. Therefore, patterns couldn't be derived to find switching between the categories within the store.

While this study has explored the lost and new customers' brand switching behaviors within the store, future studies could also track the sales generated, marketing promotions impacts. Moreover, brand switching patterns can be studied on online platforms.

Moreover, finding an extensive dataset can enhance the study by conducting a study of hierarchical data to derive brand switching patterns between categories, which can support the brand to find the loyal customers to a particular category. Brands can then do targeted promotions to enhance sales.

Additionally, other comparisons based on periods, seasons, events, and festivals can be conducted to predict trends and support brands' work on their approaches.

REFERENCES

[1]. Drew Fudenberg and Jean Tirole | Customer Poaching and Brand Switching | The RAND Journal of Economics Vol. 31, No. 4 (Winter, 2000), pp. 634-657

[2]. Michael S. Morgan and Chekitan Dev (1994) | Cornell University School of Hotel Administration | An Empirical Study of Brand Switching for a Retail Service | Pg. 2- 8

[3]. Arvind Sahay and Nivedita Sharma (2010) | Brand Relationships and Switching Behaviour for Highly Used Products in Young Consumers | Vikalpa • Volume 35, NO 1, January – March 2010

[4]. Carl Steinbach (March 2013) | committer and Project Management Committee member | Apache Hadoop* Community Spotlight (2013) | Intel IT Center

[5]. Daniel G. Murray | Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software

[6]. Dataset Source: Kaggle: https://www.kaggle.com/sanjeet41/online-retail

[7]. Ashish Thusoo, Joydeep Sen Sarma, Namit Jain, Zheng Shao, Prasad Chakka, Ning Zhang, Suresh Antony, Hao Liu, Raghotham Murthy (2010) | Facebook Data Infrastructure Team, USA | Hive - a petabyte scale data warehouse using Hadoop | 2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)

[8]. Tableau | Website https://www.tableau.com/

[9]. Hibernate | Community Document | Chapter 15. HQL: The Hibernate Query Language | https://docs.jboss.org/hibernate/orm/3.5/reference/en/html/queryhql.html

[10]. Ryan Sleeper | data visualization evangelist |Tableau 201: How to Make a Heat Map | Evolytics Evolving Analytics | https://evolytics.com/blog/tableau-201-make-heat-map/

[11]. Cambridge Dictionary | https://dictionary.cambridge.org/us/dictionary/english/brand-switching

[12]. TechTargetNetwork | Hadoop Distributed File System (HDFS) | https://searchdatamanagement.techtarget.com/definition/Hadoop-Distributed-File-System-HDFS