

# EXPERIMENTING ADVANCED ANALYTICS TECHNIQUES USING BUSINESS INTELLIGENCE TOOL

**Dr. Dhanalakshmi R V <sup>1</sup>, Dr. Richa Tiwari <sup>2</sup>, Dr. P. Padmaja <sup>3</sup>,  
Dr. K. Balanaga Gurunathan <sup>4</sup> and Dr. Anand Patil <sup>5</sup>**

<sup>1</sup> Associate Professor, Department of Management Studies, New Horizon College of Engineering, Bengaluru. Email: dhanalakshmirv19@gmail.com

<sup>2</sup> Associate Professor, School of Legal Studies, Reva University, Bengaluru. Email: richa.1976@gmail.com

<sup>3</sup> Associate Professor, Department of Management Studies, Nalla Narasimha Reddy Group of Institutions, Hyderabad.

Email: padmajaharinath1406@gmail.com

<sup>4</sup> Professor in Finance and Research Head, School of Commerce and Management, Jain University, Bengaluru. Email: balanagagurunathan@yahoo.com

<sup>5</sup> Associate Professor, School of Business and Management, CHRIST University, Bengaluru. Email: anandp7@gmail.com

DOI: 10.17605/OSF.IO/237A6

## Abstract

Business Intelligence plays a vital role in data analytics domain. It was used in all the fields mostly like weather forecasting, fraud detection etc., We can do some complex analysis like Market Basket Analysis, Cluster Analysis using BI Tools. Data Mining techniques are one of the essential things for researchers to do the advanced level of analysis. This research is about a store sales analysis which is based out of America. In our dataset, we have order priority, shipping cost, customer name, ID, ship mode, customer segment, products and other related data. Using these data, we are going to perform Market Basket Analysis and Cluster Analysis with some Business Intelligence processes like data visualisation. We took data from online source on retail store to perform basics and advanced level analysis.

**Keywords:** Data Mining, Market Basket Analysis, Cluster Analysis, Business Intelligence, Data Visualisation

## INTRODUCTION

Data Mining is well-defined as discovery of the unseen information from the folder and extraction of the outlines from the data. It is used in all the companies now-a-days for storing large amount of data and process the same. Data mining can also be referred as knowledge extraction, pattern analysis and data dredging. Basically, data mining is integrated with many techniques like information retrieval, forecasting, trends analysis, machine learning, statistics, warehousing. In technical terms, it can be referred as collecting data from different viewpoints and presenting it into meaningful statistics. In this research, we are discussing data analysis concepts right from scratch (i.e.,) from life cycle of data science, methods, process and steps of data cleaning, creating dashboard for simple analytics and complex analytics using business intelligence tools and comparing both to get the output.

## LITERATURE REVIEW

Fayyad et.al (1996) [1] in research article "From data mining to knowledge discovery in databases" labelled KDD as "a non-trivial process of knowing legal, novel, potentially useful and finally understandable patterns in data". Particularizing the description of data were any set of effective truths that are accessible in an automatic

form. Designs are models expressed by some code as data subset. The patterns must be effective so that are accurate and can be modelled for any new statistics. Procedure comprises of several stages from research to information development used for getting the result.

Non-trivial denotes that there is sort of extrapolation calculation to distinguished it from the old calculation of values. Fayyad and Stolorz (1997) [2] in their article described KDD as “generalized procedure of uncovering cherished knowledge from data with mining being one among other steps in that process that uses some algorithms for information mining process”.

## OBJECTIVES

The retail industry refers to the sector encompassing businesses involved in the sale of goods or services directly to consumers. It includes a wide range of establishments, such as department stores, supermarkets, e-commerce platforms, and specialty stores, that cater to customer needs and preferences. The objective of utilizing market basket analysis and cluster analysis in the retail industry is to gain insights into customer behaviour, identify product associations, optimize cross-selling and upselling, improve inventory management, segment customers, enhance customer profiling, and optimize marketing strategies.

### Life Cycle of Data Science

The data science life cycle includes the following stages,

- Defining and understanding the Problem
- Collection of Data
- Cleaning and preparing the data
- Exploratory Data Analysis
- Model Building Deployment

### Defining and understanding the Problem

Before starting the analysis, the most essential thing is to understand the business problem by giving the clear definition for providing a data driven decision. This is very important stage because if we missed something here, then the entire analysis will go wrong. This stage involves gathering raw structured and unstructured data. Process includes Data Extraction, Data Entry, Data Architecture, Data Mining and Data Warehousing. Our Business Problem Statement is to find the performance evaluation of Big Bazaar.

### Collection of Data

We collected the data from many raw sources such as databases and flat files. Then we integrated that into a particular format after the data gets transformed into structured and it is also called as ETL process (ETL – Extract, Transform, Load). Some tools for doing ETL process are as follows Talend Studio, DataStage and Informatica.

### Cleaning and Preparing the Data

In this stage, we need to find the missing data, highlighting and removing the duplicate entries, data formatting, validation and evaluation. The steps which are followed during

this process are transformation of data, Outliers handling, Integration of data and Reduction of data.

### Exploratory Data Analysis

In this stage, researchers will explore all the possible analysis using the data which they have. Some of them includes Regression, Text Mining, Qualitative Analysis, Web Mining, Decision Tree, Market Basket Analysis, Cluster Analysis, Decision Trees, Correlation.

### Model Building and Deployment

This stage goes with further algorithm using simple data mining and complex data mining. The model is built by a suitable machine learning algorithm which matches with our data, problem statement and the available resources. ML algorithms includes Supervised and Unsupervised. Supervised includes Regression and Classification. Unsupervised includes Clustering, Association Rule Analysis includes Apriori Algorithm and Hidden Markov Model. Then deploying the data in the form of report or presentation.

### Analysis

Before getting into the analytics part, I'll attach some rows and columns of my dataset for clear understanding.

Order ID	Order Date	Ship Mode	Product Category	Product Sub-Category
88525	01-08-2017	Take Away	Office Supplies	Labels
88522	01-08-2017	Take Away	Office Supplies	Pens & Art Supplies
88523	06-10-2018	Take Away	Office Supplies	Paper
88523	06-10-2018	Take Away	Office Supplies	Scissors, Rulers and Trimmers
88523	06-10-2018	Take Away	Technology	Telephones and Communication
88524	06-10-2018	Take Away	Office Supplies	Paper
88526	10-08-2017	Take Away	Technology	Office Machines
90193	01-01-2018	Take Away	Furniture	Chairs & Chairmats

**Table 1: Data for analysis**

We have around 2,000 customers data in total. Using the data, we need to analyse the product wise sales, Top 10 and Bottom 10 products, Top 10 and Bottom 10 profitable products and other possible analysis using a most powerful BI tool which is called **"Tableau"**.

Tableau software is focussed mainly on Business Intelligence and it is basically an American interactive data visualisation software company. Recently Tableau is acquired by Salesforce.

From the big data of about 2,000 rows and around 15 columns, we found the following Key Performance Indicator using Tableau.

Total Customers – 2,000

Total Revenue – ₹73,95,805

Total Profit - ₹9,39,066

KPI's



Fig 1: KPI

Then we found the sales analysis in terms of state and region in tableau which was shown below,

State wise Sales

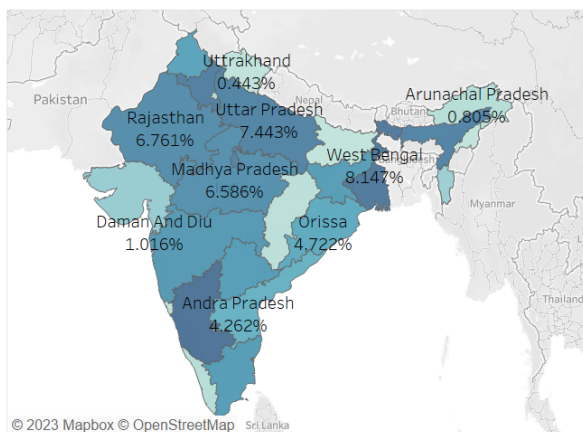


Fig 2: State wise Sales

Region wise Sales

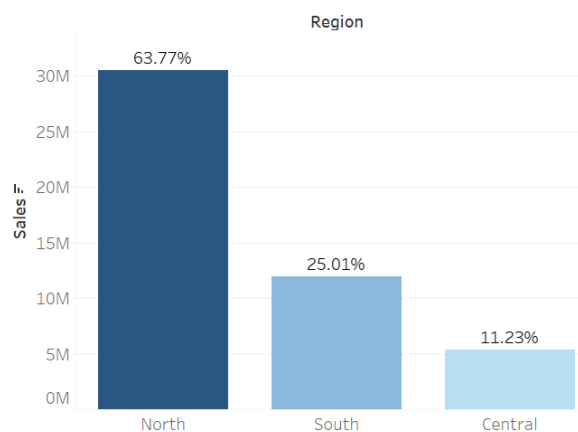
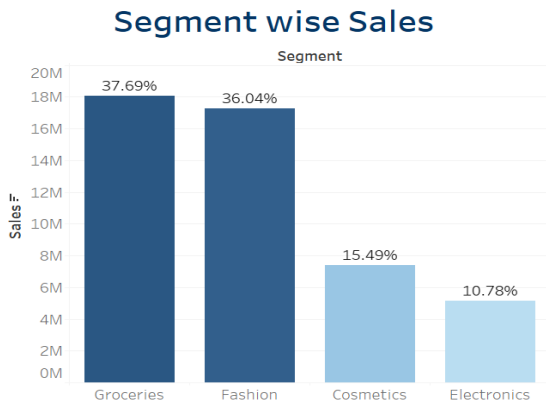
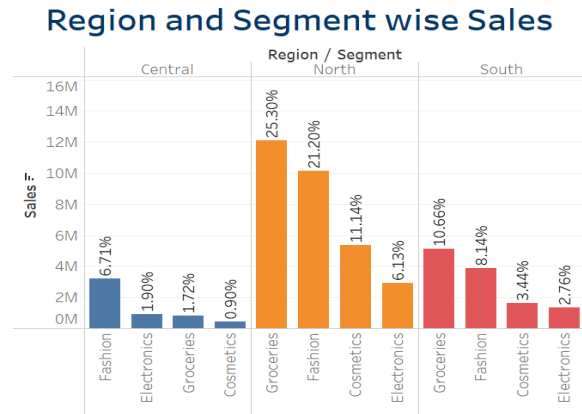


Fig 3: Region wise Sales

From the above picture, we found the sales in terms of state and region. The picture clearly shows that Uttar Pradesh holds a greater number of sales while comparing with other states and Northern region gives more sales while comparing with other region.



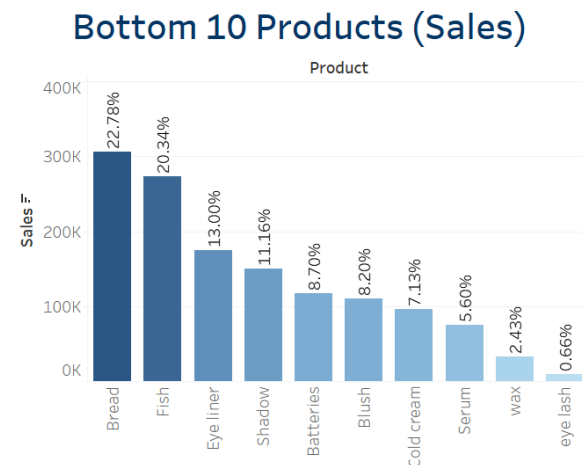
**Fig 4: Segment wise Sales**



**Fig 5: Region and Segment wise Sales**

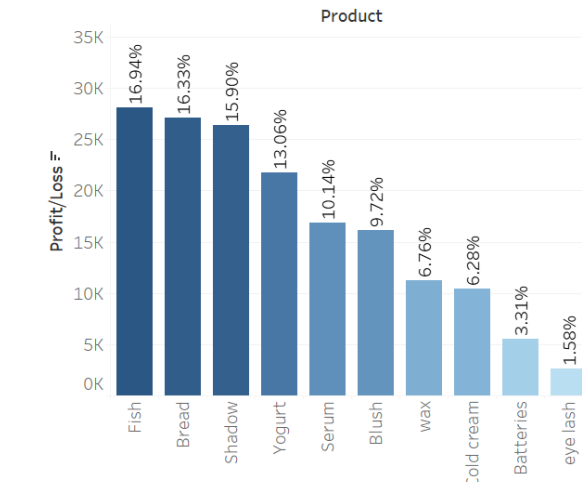
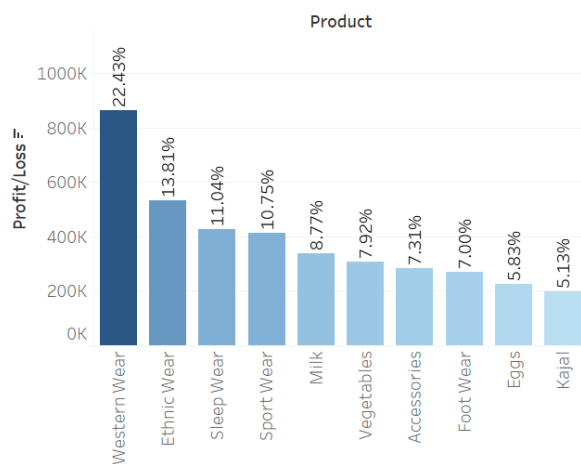
From the above visualisation, we found the segment wise sales and region and segment wise sales.

Groceries segment holds a greater number of sales and Northern region tops the sales.



**Fig 5,6: Top 10 and Bottom 10 Products based on Sales**

**Fig 7,8: Top 10 and Bottom 10 Profitable Products**



**Fig 5,6: Top 10 and Bottom 10 Products based on Sales**

**Fig 7,8: Top 10 and Bottom 10 Profitable Products**

From the above pictures, Western Wear is top product in terms of both sales and profit. But if we see the other products, the condition is different. Hence, we can't conclude our point by analysing top 10 products by sales only. That is the advanced analytics plays a major part especially in retail industry.

Some more things we analysed are,

Sportswear is in 2<sup>nd</sup> position in sales but in profit, it was in 4<sup>th</sup> position. Even the product is good in sales, the profit is low. Therefore, they need to revise their pricing and branding strategy.

For Ethnic wear, the condition is good as it was 3<sup>rd</sup> position in terms of sales but it was in 2<sup>nd</sup> position in profit. So, we can conclude that pricing and marketing strategy is good and recommend to follow the same in future as well.

Some products like Detergents, Lipstick, Egg and Kajal are either in top 10 in terms of sales or in terms of profit. Therefore, retailer needs to analyse and implement the necessary pricing strategy, branding strategy and to understand the customer needs.

Using Business Intelligence tool (Tableau), we can analyse all kind of big data based on our criteria. Hence, we can say the performance of some products are good and some are bad and providing some insights for taking better data driven decisions. As everyone knows, those analysis can be done using Tableau software.

Now, we tried implementing some advanced data mining techniques like Market Basket Analysis and Clustering Analysis in Tableau and successfully did the same. Before seeing the visualisation, let us understand what is Market Basket Analysis and Clustering Analysis in detail.

## Advance Analytics Techniques

### Market Basket Analysis

Market Basket Analysis is otherwise termed as Association Rule Analysis/Mining and Affinity Analysis. This is widely used for analysing the retail data. Market Basket means set of items in the transactions. We can do this analysis using the values like support, confidence and lift. To do the analysis, we need to find the set of items that has an important influence on business. Then we need to collect the information from numerous transactions and finally we can generate rules from counting the transaction. Some advantages of this rule are Easy Interpretation, Easy to start, Flexible data formats and simplicity of the process. Some weakness includes Lumping, Rule selections and analysing the rare items are difficult.

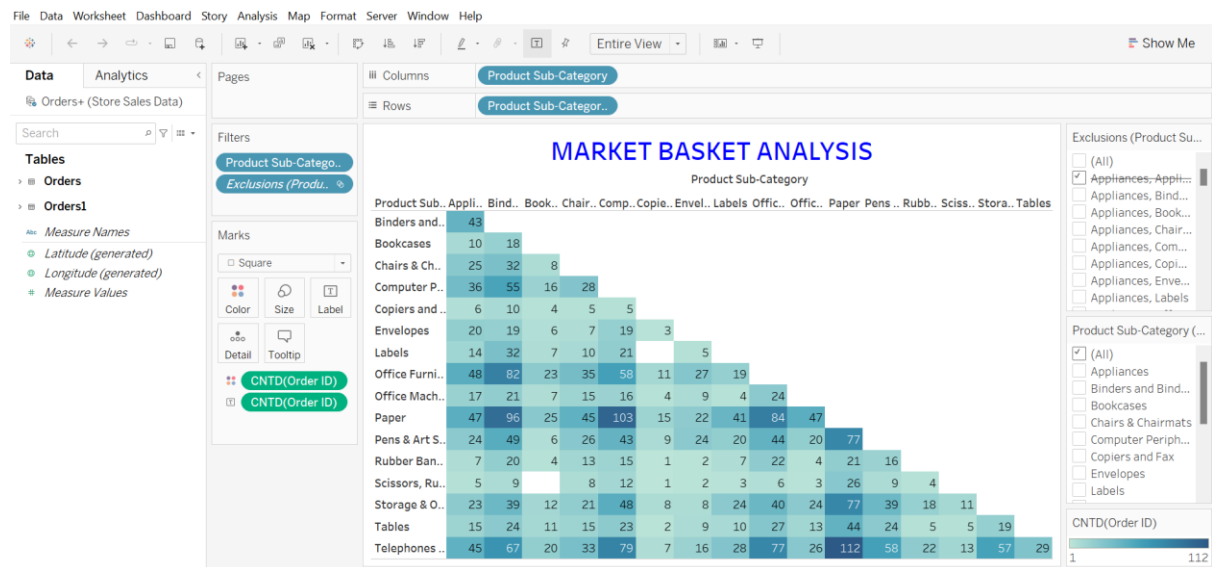
Steps for doing Market Basket Analysis includes,

- (i) Connecting data from MS Excel to Tableau.
- (ii) After dragging the order data from data source, we need to create a duplicate order (in the picture shown below) and interlink the both with some criteria like sub-category and order id.

#### Tables

- > ■ Orders
- > ■ Orders1

(iii) Finally, by dragging product sub category to both rows and columns and labelling the order id, we shall find the market basket analysis for the particular dataset which means which items are frequently brought together.



**Fig 9: Market Basket Analysis in Tableau**

This is the final look of Market Basket Analysis in Tableau which we analysed for our big dataset. From this analysis, we found that a product from technology and a product from office supplies are brought together frequently.

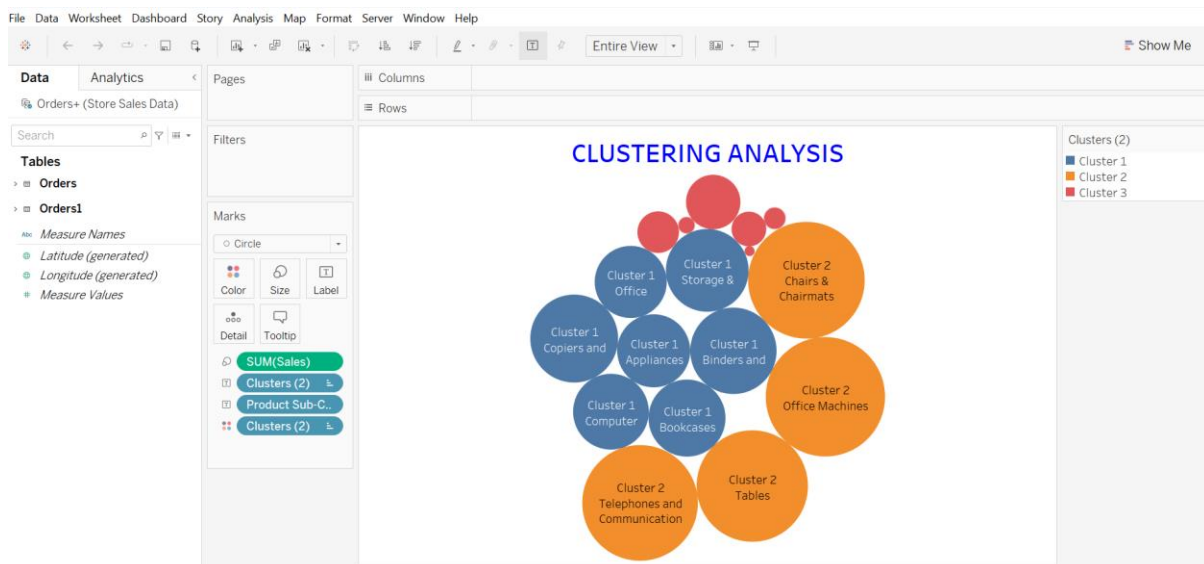
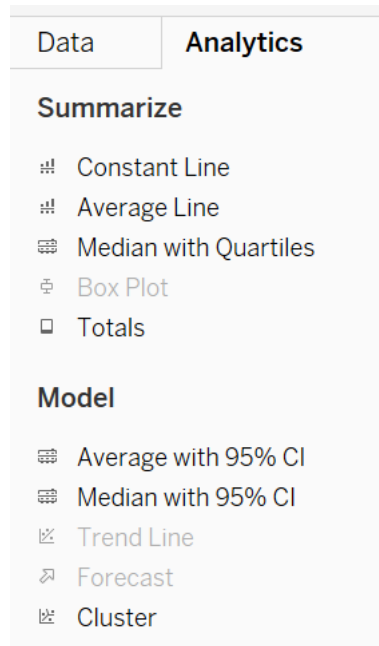
### Clustering Analysis

Clustering analysis is nothing but dividing the product category into various sub category in terms of profit, sales, category, region, products etc., where we have large amount of dataset with diverse categories.

From our dataset, we analysed a cluster technique for products by dividing that into sub-categories.

Steps for doing cluster analysis are,

- (i) Extract the data from MS Excel to Tableau
- (ii) Using Product Category and sales field, we can get a normal bar chart or pie chart.
- (iii) Then from the analytics, we need to select the cluster option and then based on our criteria, we can give either 3, 4 or 5 clusters.



**Fig 10: Clustering Analysis in Tableau**

### Future Scope of the Project

Since we experimented only two advanced analyses in our research, other researchers can try experimenting other techniques like Text Mining, Web Mining, Decision Trees, Hidden Markov Model etc.,

### CONCLUSION

From our analysis, we found various insights on how big giant FMCG companies big bazaar analyses their sales. Right from knowing their Key Performance Indicators, State wise Sales, Region wise Sales, Segment wise Sales, Region and Segment wise Sales, Top 10 Products based on sales, bottom 10 Products based on Sales, Top and Bottom 10 Profitable Products out of all the Products selling in Big Bazaar.



Finally, we did some advanced analysis as well like Market Basket Analysis and Clustering Analysis to know what products are being purchased frequently together and we clustered the products based on the total number of sales.

### References

1. Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 2011.
2. Piatetsky-Shapiro, Gregory. "Knowledge discovery in databases: 10 years after." ACM SIGKDD Explorations Newsletter 1.2 (2000): 59-61.
3. Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." AI magazine 17.3 (1996): 37.
4. Fayyad, Usama, and Paul Stolorz. "Data mining and KDD: Promise and challenges." Future generation computer systems 13.2 (1997): 99-115.
5. Venkatadri, M., and Lokanatha C. Reddy. "A review on data mining from past to the future." International Journal of Computer Applications 15.7 (2011): 19-22.
6. Silwattananusarn, Tipawan, and Kulthida Tuamsuk. "Data mining and its applications for knowledge management: a literature review from 2007 to 2012." arXiv preprint arXiv:1210.2872 (2012).
7. Tomar, Divya, and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare." International Journal of Bio-Science and Bio-Technology 5.5 (2013): 241-266. [16]. Saurkar, Anand V., et al. "A Review Paper on Various Data Mining Techniques." International Journal of Advanced Research in Computer Science and Software Engineering 4.4 (2014).
8. Soni, Shraddha. "A Literature Review on Data Mining and its Techniques." Indian Journal of Applied Research 5.6 (2016).