

# Actionable Insights of the Consumer Sales Data using MongoDB - Big Data Case Study

## Prologue

The main objective of processing the consumer sales data is to explore the insights (Sales & Profit) from the consumer sales data using MongoDB Big-Data processing database. The relationship between consumer and company is one of the prominent determinants of the progress of any company.

## Introduction to Consumer Data

Manufacturer and marketer of health, beauty products, home and electronic appliances, receives a constant stream of downstream data from retail consumers and market research partners. To drive sales, the company's consumer insights and category management team is charged to produce in-depth reports for more product categories based on the data. Consumer raw data contains all the parameters like consumer id, order id, product id, order category, product sales, order priority, profit etc.

Out of all the parameters, we have to analyze properly the main parameters that affecting on the product sales & profit of the company.

## Handling the Data

Here, we used a MongoDB database to process the consumer sales data. MongoDB is an open source, scalable, high-performance, schema-free, cross-platform document-oriented database system classified as a "No-SQL" database. Map\_Reduce can be used for batch processing of data and aggregation operations. MongoDB requires a data folder to store its files.

We will take the unstructured data into excel and convert it into csv file format. Once csv file is ready, we will convert the csv data into JSON(JavaScript Object Notation) Format. Because the MongoDB database is a Document-Oriented Database, it takes only JSON format data in the form of documents.

And we will create a new collection to put the documents in the collection and MongoDB will process, visualize the documents and performs the Map-Reduce functions on the collection to aggregate the results. The fig.(a) below shows the process of handling the data.



Fig.(a)

## Major Insights to be Found-out

After analyzing and processing the consumer raw data, we get insights of consumer sales and also behavior of the consumer and following are the few of insights came across the process,

1. Total number of product sales
2. Total number of Orders
3. Profit to the company

## Methodology

Here, first we create the database and new collection to store the documents in the collection and next we take the JSON(JavaScript Object Notation) converted data and import into the newly created MongoDB database collection.

To process the data systematically, we perform data visualization by mapping the data using MongoDB Map function and next we perform the Reduce or count operation using MongoDB Reduce function to aggregate the counts and filter the data. Next, we will process the data and processed result will put into the production environment. The below fig.(b) shows MongoDB database model.

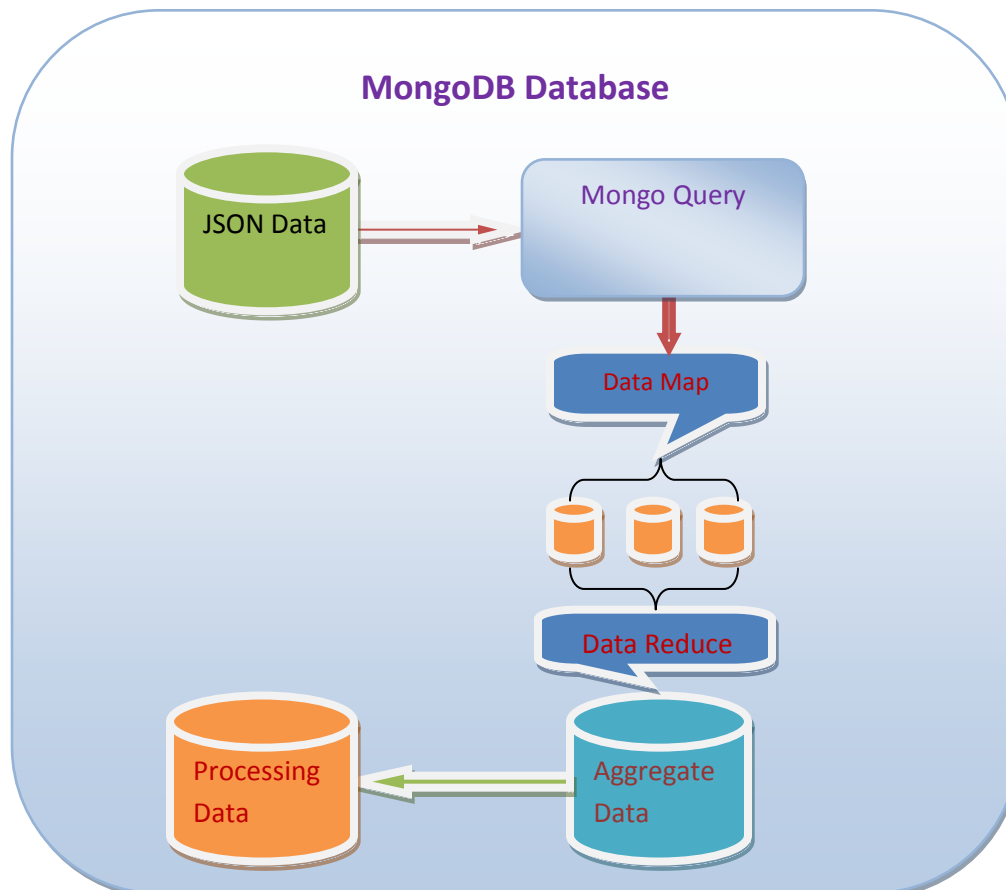


Fig.(b)

## MongoDB Queries

MongoDB queries are Ad hoc, rich and document-based queries. MongoDB supports search by field, range queries, regular expression searches. Queries can return specific fields of documents and also includes user-defined JavaScript functions.

## Results

The input selected was consumer sales data after processing with our database, the results we obtained for individual products are their total number of sales and profit. The results are systematic and anybody can clearly understand the sales and profit. Any number of parameters can be added to the process to analyze the data. Also, based on this result, we can maintain good long-term relationships with consumers.

First, we mapped ( tokenize, filter & count) the entire data based on the data parameters like order id, order date, order priority, order quantity, order price, product category, product sub category and consumer segments and then we reduced or aggregated the mapped data to further processing the data.

Following charts shows the results we obtained after processing the consumer sales data.

1. The Figs.(c) & (d) shows total number of orders based on the order priorities.

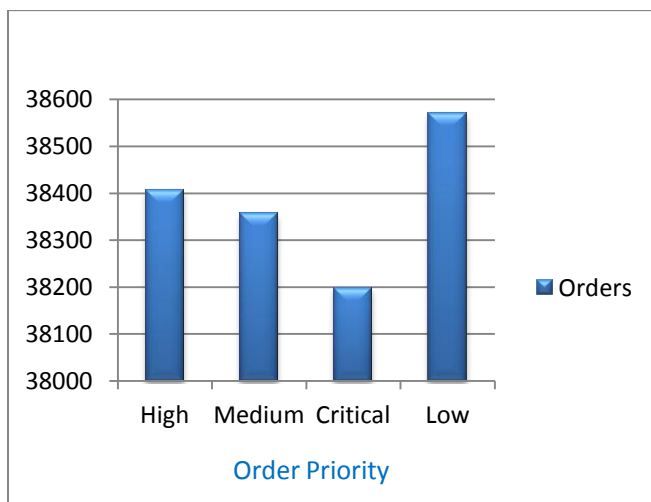


Fig.(c)

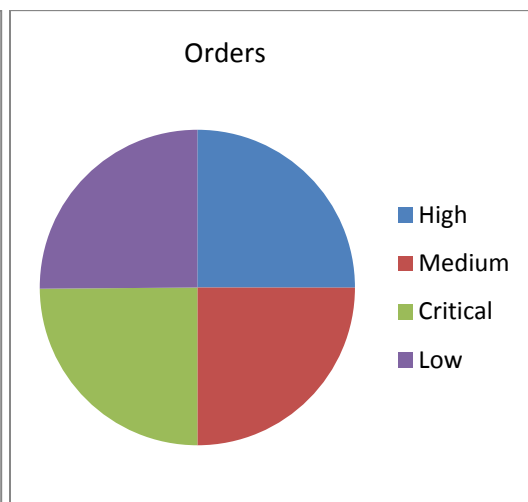


Fig.(d)

2. The Figs.(e) & (f) shows total number of sales based on the product categories.

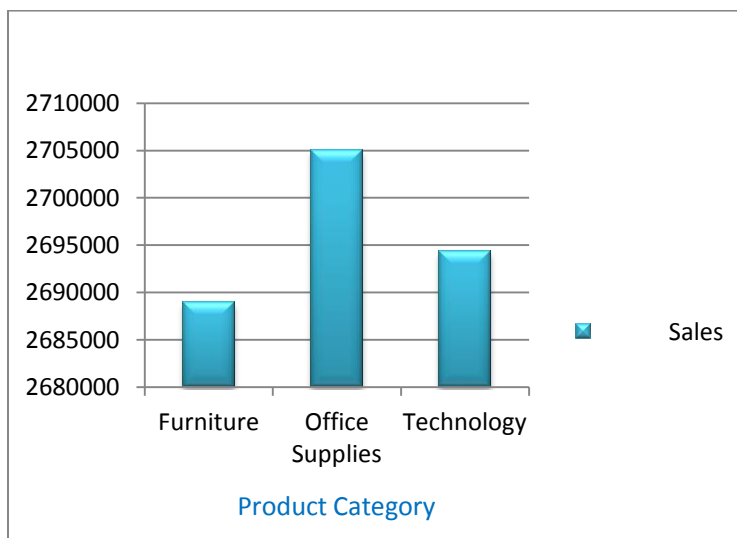


Fig.(e)

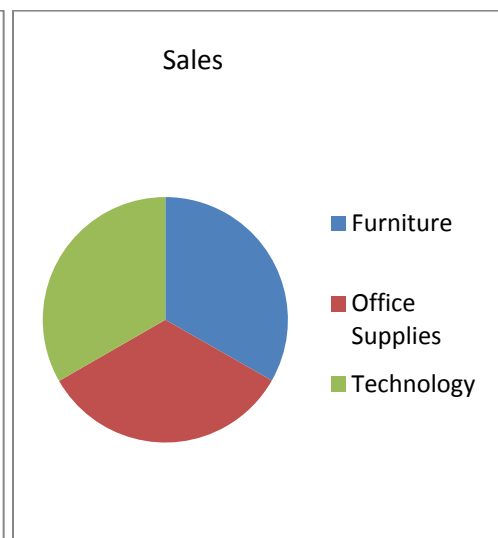


Fig.(f)

3. The Figs.(g) & (h) shows total number of sales based on the product sub-categories.

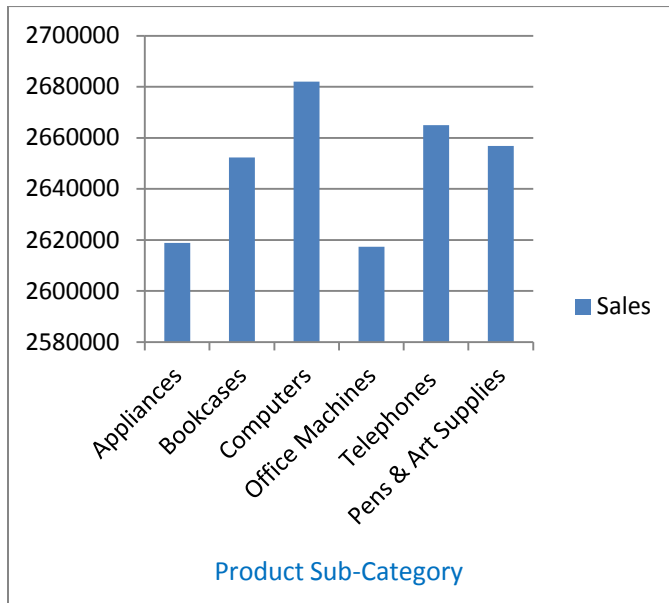


Fig.(g)

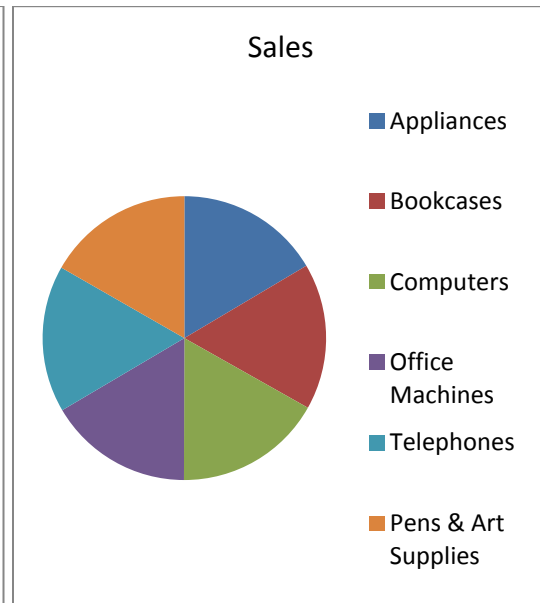


Fig.(h)

4. The Figs.(i) & (j) shows total number of sales based on the consumer segments.

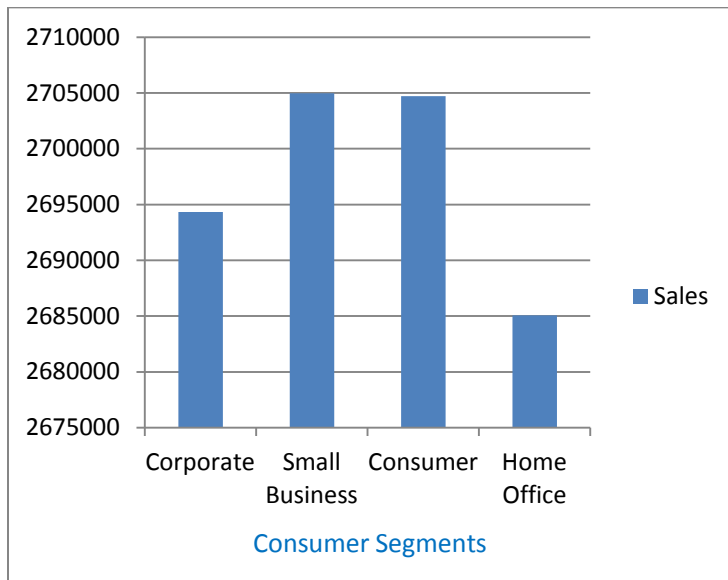


Fig.(i)

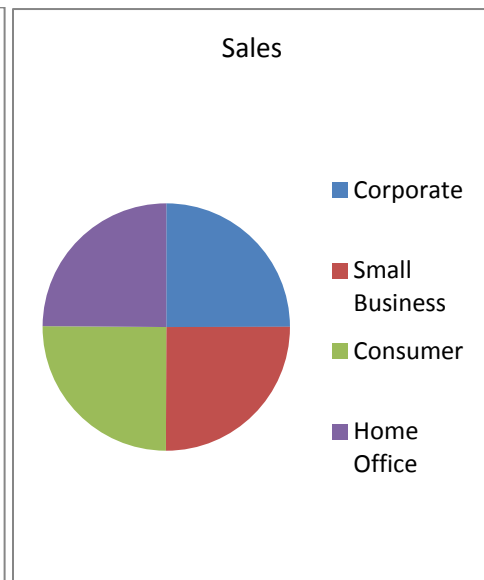


Fig.(j)

5. The Figs.(k) & (l) shows total number of sales & profit based on the discount.

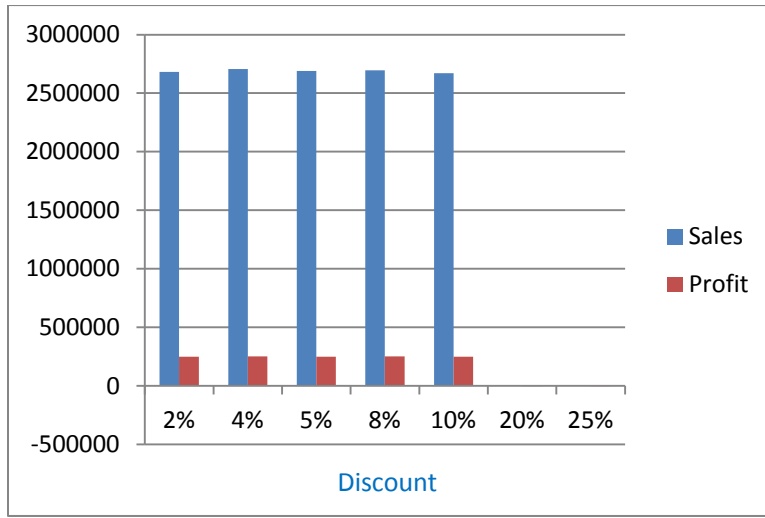


Fig.(k)

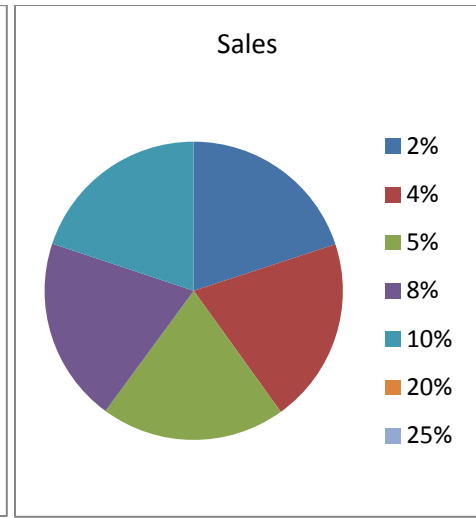


Fig.(l)

## Conclusion

We developed a solution that centralized the company's sales data, including product, market, field/territory and consumer information. The new solution also allowed the company to develop dynamic data sets to display a summary of market trends using real-time data. The consumer information provided was also key to this solution. Within the interface, sales people can now access all of their key consumer information.

## Consumer Delight

The aim of sales and marketing is to know and understand and satisfy the consumer so well the product or service fits him and sells itself. From this methodology, consumer with ease can discover number of products which are in good sales and the trending products and the profit generated from a particular product. Those consumer who are in ambiguity to select a particular product will get clear idea about the product.