**A PROJECT REPORT ON**

**“AN ANALYSIS OF THE SALES OF ‘A MART’ SUPERMARKET”**

Submitted in partial fulfillment of the requirements for the course on

**DATA ANALYTICS**

**By**

**Dr. Ghali Krishna Harshitha**

**Abhishek Jogu**

**Ishika Garg**

**Manpreet Kaur**



**INDIAN INSTITUTE OF MANAGEMENT**

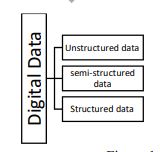
**ROHTAK**

|  |  |  |
| --- | --- | --- |
| Chapters | Description | Page No. |
| Chapter-1 | Introduction | 02 |
| Chapter-2 | Review of Literature | 07 |
| Chapter-3 | Methodology | 09 |
| Chapter-4 | Data Analysis and Interpretation | 10 |
| Chapter-5 | Findings, summary, conclusion and suggestions for future analysis | 16 |

**CHAPTER – 1**

**INTRODUCTION**

Data is a precious and irreplaceable asset of all the enterprises, irrespective of its size. The need of the hour is to understand, manage, process, and take the data for analysis to draw valuable insights. Digital data can be structured, semi-structured or unstructured data.



*Source: https://mu.ac.in/wp-content/uploads/2021/01/BIG-DATA-ANALYTICS.pdf*

***Structured Data:*** The data that follows a pre-defined schema/structure is structured data. This is the data which is in an organized form (e.g., in rows and columns) and be easily used by a computer program. Relationships exist between entities of data, such as classes and their objects. About 10% data of an organization is in this format. Data stored in databases is an example of structured data.

***Semi-structured:*** Semi-structured data is also referred to as self-describing structure.This is the data which does not conform to a data model but has some structure. However, it is not in a form which can be used easily by a computer program. About 10% data of an organization is in this format; for example, HTML, XML, JSON, email data etc.

***Unstructured data:*** This is the data which does not conform to a data model or is not in a form which can be used easily by a computer program. About 80% data of an organization is in this format; for example, memos, chat rooms, PowerPoint presentations, images, videos, letters. researches, white papers, body of an email, etc.

Data analytics has extensively spread across the industries in a very short span with its enhanced capability for processing large amounts of data and efficient storage. This data provides insights into the operations, customers, markets, competitors, etc. enabling them to improve their decision making capacity, improve efficiency, enhance customer experiences, increase revenue, reduce costs, mitigate risks and stay ahead of the competition. Data analytics also helps to simulate scenarios, analyse the impact of changes in various variables and stay atop the competitors.

**Classification of Analytics**

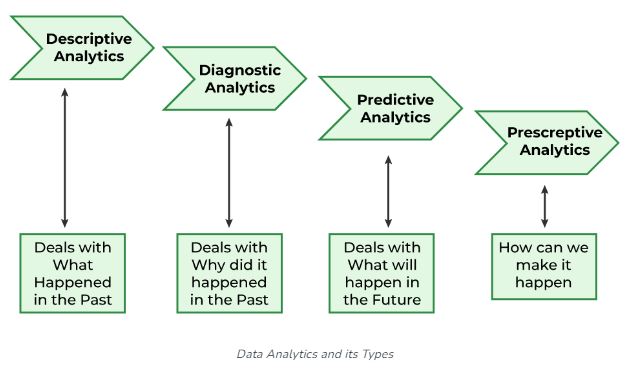
The techniques used for classification of Data are:

***Supervised Learning:*** Supervised learning algorithms require labelled data. This means the algorithm is provided with a dataset that has already been categorised or labelled with class labels. The algorithm then uses this label to learn how to distinguish between different class objects in the data. Once trained, it can use its predictive power to classify new datasets.

***Unsupervised Learning:*** Unsupervised learning algorithms do not require labelled data. Instead, they use clustering and dimensionality reduction techniques to identify patterns in the dataset without any external guidance. These algorithms help segment customers or identify outlier items in a dataset.

***Deep Learning:*** Deep learning is a subset/division of machine learning technologies that use artificial neural networks. These algorithms are capable of learning from large datasets and making complex decisions. Deep learning can be used for tasks such as image classification, natural language processing, and predictive analytics.

**Types of Data Analytics**

****

***Descriptive Analytics***

Descriptive analytics looks at data and analyze past event for insight as to how to approach future events. It looks at past performance and understands the performance by mining historical data to understand the cause of success or failure in the past. Almost all management reporting such as sales, marketing, operations, and finance uses this type of analysis.

The descriptive model quantifies relationships in data in a way that is often used to classify customers or prospects into groups. Unlike a predictive model that focuses on predicting the behavior of a single customer, Descriptive analytics identifies many different relationships between customer and product.

Common examples of Descriptive analytics are company reports that provide historic reviews like:

* Data Queries
* Reports
* Descriptive Statistics
* Data dashboard

***Diagnostic Analytics***

In this analysis, we generally use historical data over other data to answer any question or for the solution of any problem. We try to find any dependency and pattern in the historical data of the particular problem.

For example, companies go for this analysis because it gives a great insight into a problem, and they also keep detailed information about their disposal otherwise data collection may turn out individual for every problem and it will be very time-consuming. Common techniques used for Diagnostic Analytics are:

* Data discovery
* Data mining
* Correlations

***Predictive Analytics***

Predictive analytics turn the data into valuable, actionable information. predictive analytics uses data to determine the probable outcome of an event or a likelihood of a situation occurring. Predictive analytics holds a variety of statistical techniques from modeling, machine learning, data mining, and game theory that analyze current and historical facts to make predictions about a future event. Techniques that are used for predictive analytics are:

* Linear Regression
* Time Series Analysis and Forecasting
* Data Mining

Basic Corner Stones of Predictive Analytics

* Predictive modeling
* Decision Analysis and optimization
* Transaction profiling

***Prescriptive Analytics***

Prescriptive Analytics automatically synthesize big data, mathematical science, business rule, and machine learning to make a prediction and then suggests a decision option to take advantage of the prediction.

Prescriptive analytics goes beyond predicting future outcomes by also suggesting action benefits from the predictions and showing the decision maker the implication of each decision option. Prescriptive Analytics not only anticipates what will happen and when to happen but also why it will happen. Further, Prescriptive Analytics can suggest decision options on how to take advantage of a future opportunity or mitigate a future risk and illustrate the implication of each decision option.

For example, Prescriptive Analytics can benefit healthcare strategic planning by using analytics to leverage operational and usage data combined with data of external factors such as economic data, population demography, etc.

**Usage of Data Analytics**

There are some key domains and strategic planning techniques in which Data Analytics has played a vital role:

***Improved Decision-Making*** – If we have supporting data in favour of a decision, then we can implement them with even more success probability. For example, if a certain decision or plan has to lead to better outcomes then there will be no doubt in implementing them again.

***Better Customer Service –*** Churn modeling is the best example of this in which we try to predict or identify what leads to customer churn and change those things accordingly so, that the attrition of the customers is as low as possible which is a most important factor in any organization.

***Efficient Operations*** – Data Analytics can help us understand what is the demand of the situation and what should be done to get better results then we will be able to streamline our processes which in turn will lead to efficient operations.

***Effective Marketing*** – Market segmentation techniques have been implemented to target this important factor only in which we are supposed to find the marketing techniques which will help us increase our sales and leads to effective marketing strategies.

**Future Scope of Data Analytics**

***Retail:*** To study sales patterns, consumer behavior, and inventory management, data analytics can be applied in the retail sector. Data analytics can be used by retailers to make data-driven decisions regarding what products to stock, how to price them, and how to best organize their stores.

***Healthcare:*** Data analytics can be used to evaluate patient data, spot trends in patient health, and create individualized treatment regimens. Data analytics can be used by healthcare companies to enhance patient outcomes and lower healthcare expenditures.

***Finance:*** In the field of finance, data analytics can be used to evaluate investment data, spot trends in the financial markets, and make wise investment decisions. Data analytics can be used by financial institutions to lower risk and boost the performance of investment portfolios.

***Marketing:*** By analyzing customer data, spotting trends in consumer behavior, and creating customized marketing strategies, data analytics can be used in marketing. Data analytics can be used by marketers to boost the efficiency of their campaigns and their overall impact.

***Manufacturing:*** Data analytics can be used to examine production data, spot trends in production methods, and boost production efficiency in the manufacturing sector. Data analytics can be used by manufacturers to cut costs and enhance product quality.

***Transportation:*** To evaluate logistics data, spot trends in transportation routes, and improve transportation routes, the transportation sector can employ data analytics. Data analytics can help transportation businesses cut expenses and speed up delivery times.

**CHAPTER – 2**

**REVIEW OF LITERATURE**

**Kornelia Batko and Andrzej Slezak (2022)** in their research article analyzed the possibilities of using Big Analytics in healthcare. The research included both critical analysis of literature and presentation of selected results of direct research on the use of Big Data Analytics in medical facilities. The direct research was carried out based on research questionnaire and conducted on a sample of 217 medical facilities in Poland. Literature studies have shown that the use of Big Data Analytics can bring many benefits to medical facilities, while direct research has shown that medical facilities in Poland are moving towards data-based healthcare because they use structured and unstructured data, reach for analytics in the administrative, business and clinical area. The research positively confirmed that medical facilities are working on both structured data and unstructured data. The following kinds and sources of data can be distinguished: from databases, transaction data, unstructured content of emails and documents, data from devices and sensors. However, the use of data from social media is lower as in their activity they reach for analytics, not only in the administrative and business but also in the clinical area. It clearly shows that the decisions made in medical facilities are highly data-driven. The results of the study confirm what has been analyzed in the literature that medical facilities are moving towards data-based healthcare, together with its benefits.

**Md. Saifur Rahman and Hassan Reza (2022**) in their study offered an overview of recent works in social media, data science, and machine learning to gain a wide perspective on social media big data analytics. They explained why social media data are significant elements of the improved data-driven decision-making process. They proposed and built the “Sunflower Model of Big Data” to define big data and bring it up to date with technology by combining 5 V's and 10 Bigs. They discovered the top ten social data analytics to work in the domain of social media platforms. A comprehensive list of relevant statistical/machine learning methods to implement each of these big data analytics were discussed in their work. “Text Analytics” is the most used analytics in social data analysis to date. They created a taxonomy on social media analytics to meet the need and provide a clear understanding. Tools, techniques, and supporting data type were also discussed in the research work. As a result, researchers will have an easier time deciding which social data analytics would best suit their needs.

**Rahul Reddy Nadikattu (2020),** explored the role of data science and data analytics with special reference to the area of business management. Author stated data analytics provide operational insights into complex business situations. He expressing that looking into the historical data from a modern perspective, finding new and challenging business scenarios and applying methodologies to find a better solution are the prime concerns of a data analyst.

**Acharjya. D.P. and Kauser Ahmed, P. (2016)** explored the potential impact of big data challenges, open research issues and various tools associated with it. They stated that every big data platform has its individual focus. Some of them are designed for batch processing whereas some are good at real-time analytic. Each big data platform also has specific functionality. Different techniques used for the analysis include statistical analysis, machine learning, data mining, intelligent analysis, cloud computing, quantum computing, and data stream processing.

**Chun-Wei Tsai, et.al., (2015)** in their paper reviewed studies on data analytics from traditional data analysis to recent big data analysis. KDD process is used as the framework for these studies and is summarized into three parts: input, analysis and output. The discussions were focused on the performance-oriented and results-oriented issues. The paper focused on the data analysis of KDD from the platform/framework to data mining. The open issues on computation, quality of end result, security and privacy are then discussed to explain which open issues might be faced. Last but not least, the paper is aimed at helping the audience of the paper to find the solutions to welcome the new age of big data.

**Bibliography**

* **Acharjya. D.P. and Kauser Ahmed, P. (2016),** A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, pp. 511-518.
* **Chun-Wei Tsai, Chin-Feng Lai, Han Cieh Chao and Athanasios V. Vasilokas (2015),** Big Data Analytics: A Survey, Journal of Big Data 2, Article 21, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0030-3.
* **Kornelia Batko and Andrzej Slezak (2022),** The Use of Big Data Analytics in Healthcare, Journal of Big Data, 9, article number: 3, <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00553-4>.
* **Md. Saifur Rahman and Hassan Reza (2022**), A Systematic Review Towards Big Data Analytics in Social Media, Big Data Mining and Analytics, Vol. 5, no. 3, pp. 228-244.
* **Rahul Reddy Nadikattu (2020),** Research on Data Science, Data Analytics and Big Data, International Journal of Engineering, Science and Technology, Volume 9, Issue 5, pp. 99-105.

**CHAPTER – 3**

**METHODOLOGY**

**Meaning:** Methodology is a way of explaining how a researcher intends to carry out their research. It's a logical, systematic plan to resolve a research problem. A methodology details a researcher's approach to the research to ensure reliable, valid results that address their aims and objectives. It encompasses what data they're going to collect and where from, as well as how it's being collected and analyzed.

**Objectives of the Study:**

* To identify the relationship between Demand and Supply.
* To examine the differences in the ordering pattern of male customers and female customers.
* To evaluate the status of the orders.
* To find out the shares of each state in percentages.
* To identify the age group wise orders received in male and female customers.
* To study the channel-wise distribution of orders received.

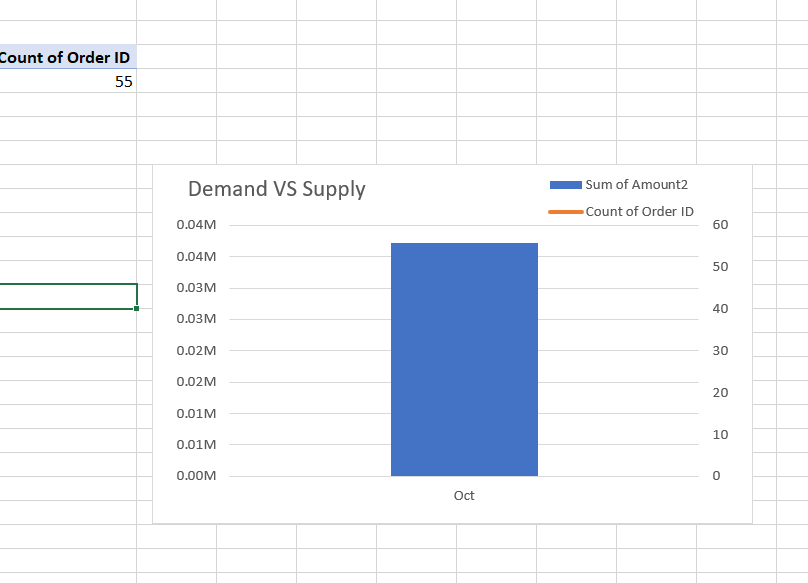
**Data Used for the Study:** The data used for the study comprises of sales data of A Mart Super Market for the year 2022.

**Tools used for the Analysis of Data:** The data are analyzed using Microsoft Excel with the help of tools such as pivot table, charts, vlookup, hlookup, solver, etc.

**CHAPTER 4**

**DATA ANALYSIS AND INTERPRETATION**

**Demand vs. Supply**



**Demand (Sum of Amount2):** $37,216

**Supply (Count of Order ID):** 55 orders

**Analysis:**

**Demand Analysis:**

The total demand for the month of October is $37,216. This represents the total sales revenue or the sum of all amounts sold during this period.

**Supply Analysis:**

There were 55 orders placed during October. This indicates the number of transactions or orders processed within that time frame.

**Interpretation:**

**Trend Analysis:** Through this data we can compare with previous months or the same period in previous years to identify any growth or decline in demand and supply. Understanding seasonal patterns or changes over time can be valuable for business planning and decision-making.

**Further Analysis:** Through this we can dig deeper into the data, exploring factors influencing demand fluctuations (e.g., promotions, seasonal effects, customer behavior) or supply chain efficiencies (e.g., fulfillment speed, inventory management)

**Male vs. Female**

A screenshot of a computer

Description automatically generated

**Female:** Total sales amount - $33,443

**Male:** Total sales amount - $3,773

**Analysis:**

**Gender-based Sales Analysis:**

The total sales attributed to female customers are significantly higher ($33,443) compared to sales from male customers ($3,773).

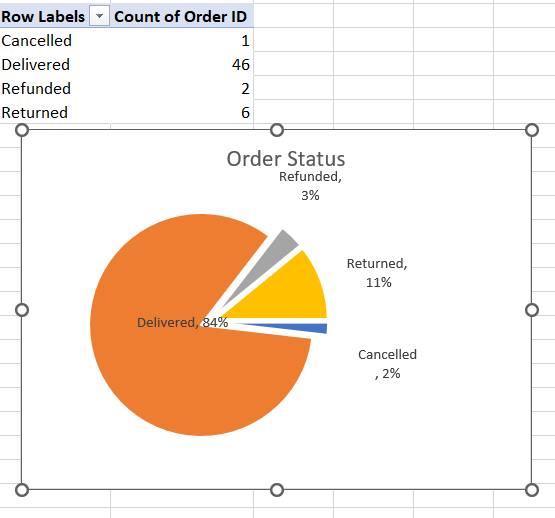
**Interpretation:**

**Gender Disparity in Sales:** There seems to be a substantial difference in purchasing behavior between genders. Females contribute significantly more to the total sales compared to males. This might suggest that products or services offered cater more to female preferences or that the store has a larger female customer base.

**Targeted Marketing or Product Strategies:** This disparity could prompt a closer examination of marketing strategies, product offerings, or even store layout and design to better cater to the preferences and needs of male customers. Understanding what drives higher sales among females might also inform targeted marketing efforts.

**Customer Segmentation and Personalization:** Insights from this data could lead to strategies for better customer segmentation and personalization. Tailoring products, promotions, or experiences based on gender-specific preferences might further enhance sales and customer satisfaction.

**Order Status**



**Analysis:** The analysis of the status of orders revealed that,

* The majority of orders (84%) have been successfully delivered.
* A small percentage of orders were cancelled (2%) or refunded (3%).
* Around 11% of orders were returned, which is a considerable portion of the total orders.

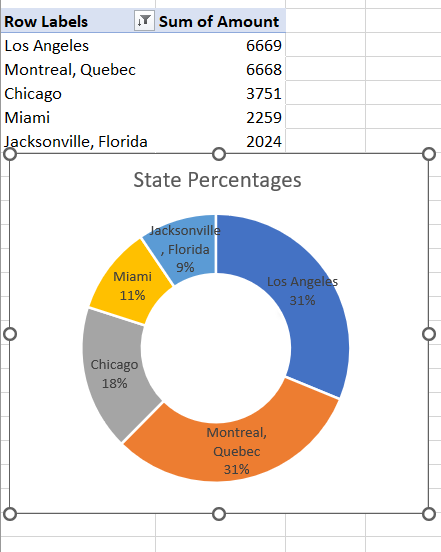
**Interpretation:**

**Delivery Success Rate:** The high percentage of delivered orders indicates an effective delivery process and customer satisfaction in receiving their orders.

**Returns and Refunds:** The percentages of returns (11%) and refunds (3%) signify a significant proportion of customers encountering issues or dissatisfaction with received products. This could prompt an investigation into product quality, accurate descriptions, or customer service to reduce returns and refunds.

**Cancelled Orders:** The low percentage of cancelled orders suggests a relatively low incidence of customers canceling their orders before fulfillment, which could be a positive sign in terms of order processing efficiency.

**State Percentages**



**Analysis:** The analysis of shares of states in percentages revealed,

* Angeles and Montreal have similar percentages of sales, contributing around 31% making them the top-selling locations in terms of revenue.
* Chicago follows with a moderate percentage of 18%.
* Miami and Jacksonville have relatively lower percentages of sales, around 11% and 9% respectively.

**Interpretation:**

Top Performing Locations: Los Angeles and Montreal are the primary contributors to sales, indicating strong market presence or customer preference in these areas.

**Regional Variances:** Variations in sales percentages across different locations might suggest varying market demands, customer behaviors, or effectiveness of marketing strategies in different regions.

**Strategic Focus:** Identifying the regions contributing the most or the least could guide strategic decisions, such as increasing marketing efforts in regions with lower sales percentages or reinforcing successful strategies in high-performing areas.

**Age and Gender**

A screenshot of a computer

Description automatically generated

**Gender and Age Group Distribution:** The analysis of the data to find the distribution of customers according to gender and age showed,

* Among females, the highest percentage of orders comes from the Adult category (49.09%), followed by Teenagers (29.09%), and Seniors (10.91%).
* For males, the majority of orders are from the Adult category (5.45%), followed by Teenagers (5.45%). There are no recorded orders from Seniors (0.00%).

**Interpretation:**

**Female Preferences:** The highest percentage of orders from adult females might indicate that this demographic group contributes significantly more to sales compared to other female age groups. Teenage females also make a substantial contribution to orders.

**Male Preferences:** In contrast, there's a lower overall percentage of orders from males, with the majority coming from adult males and teenage males, but at notably lower percentages compared to their female counterparts.

**Market Segmentation:** Understanding these demographic differences can help tailor marketing strategies, product offerings, and promotions to better suit the preferences and behaviors of each demographic group. It might be beneficial to explore why there are no recorded orders from senior males or very few from senior females, as it could uncover untapped market potential or indicate a need for targeted outreach to that demographic.

**Channel**

A pie chart with numbers and text

Description automatically generated

**Analysis:** The analysis of data on channel used suggests that all the orders in this dataset are associated with or fulfilled through Amazon. This could imply that the store or platform in question is primarily using Amazon as its sole or exclusive sales channel.

**Interpretation:**

**Exclusive Platform Usage:** Relying entirely on Amazon as the sales platform might have both advantages and limitations. While it offers access to a large customer base and the benefits of Amazon's infrastructure, it could also mean dependency on a single platform, risking vulnerability to any changes in Amazon's policies, fees, or market dynamics.

**Strategic Implications:** Understanding the exclusive use of Amazon as the sales channel could prompt considerations for diversification and risk mitigation. Exploring other platforms or establishing an independent sales channel might provide more stability and control over the business.

**Efficiency Considerations:** Operating solely on Amazon might simplify operations but might also limit opportunities for branding, customer experience customization, or direct customer engagement outside of Amazon's ecosystem.

**CHAPTER 5**

**FINDINGS, SUMMARY, CONCLUSION AND SUGGESTIONS FOR FUTURE ANALYSIS**

**Key Findings**

**Demand and Supply:**

In October, the store recorded a demand of $37,216 across 55 orders, indicating the total revenue generated within that timeframe.

**Gender-Based Sales:**

Females contributed significantly more to total sales, with $33,443 in sales compared to males' $3,773.

**Order Status:**

Most orders were successfully delivered (46 out of 55). A small portion was cancelled (1), refunded (2), or returned (6).

**Sales by Location:**

Los Angeles and Montreal emerged as the top-selling locations, each contributing around 31% of the total sales, followed by Chicago (18%).

**Age and Gender Distribution:**

Adult females made the highest percentage of orders (49.09%), followed by Teenage females (29.09%), while adult males made up a smaller percentage (5.45%).

**Amazon as Sales Channel:**

All orders (100%) were attributed to Amazon, suggesting exclusive reliance on this platform for sales.

**Summary**

The data reveals a few crucial aspects:

* Stronger sales from females, especially adults and teenagers, highlight their significant contribution to revenue.
* Locations like Los Angeles, Montreal, and Chicago are primary sales contributors, indicating potential regional market strengths.
* There's a disparity between gender-based sales, with males constituting a smaller portion.
* The exclusive use of Amazon as the sole sales channel might offer advantages but also poses risks of dependency.

**Conclusion**

**Market Segmentation:**

Tailoring strategies and products towards female demographics, particularly adults and teenagers, could yield more substantial returns.

**Regional Focus:**

Considering the success in Los Angeles and Montreal, further investment or marketing efforts in these areas might be beneficial.

**Diversification Strategy:**

Sole reliance on Amazon presents risks; exploring alternative sales channels or platforms could mitigate risks associated with dependency.

**Customer Experience and Returns:**

Addressing reasons for returns and refunds could enhance customer satisfaction and minimize revenue loss.

**Suggestions for Future Analysis**

Continuous monitoring of sales trends, customer preferences, and potential expansions into diverse markets could drive sustained growth.

This data analysis provides insights into current sales patterns, gender, age group preferences, and regional strengths while signaling areas for improvement and future strategic directions.