Analyzing the growth of FMCG sales in local market for better opportunities and pricing using big data

Diwash Mainali

diwash006@gmail.com

Abstract- This report analyzes sales data from the local retail store selling FMCG goods. This analysis aims to understand the consumer pattern, identifying patterns influencing sales which can help to identify opportunities for optimized pricing and inventory management. The key findings show that the overall sales were high in 2022, possibly due to the COVID-19 pandemic. The sales trend are highest in weekends, festivals and wedding seasons. Analysis of sales prices showed a strong positive correlation with MRP, which is directly correlated with each other. Sales forecasting using historical data revealed cyclical patterns that can help to manage inventory more effectively. By using these insights, we can create optimized pricing and targeted advertisements, which will help improve company profitability.

Keywords- fast-moving consumer goods (FMCG), supply chain management (SCM), big data analytics, demand forecasting, product category analysis, sales patterns, predictive analysis, data-driven decisions, consumer demand, pricing strategy, inventory management, targeted marketing, sales distribution, sales inequality, lorenz curve, data visualization, prophet forecasting

I. Introduction

Fast-moving consumer goods (FMCG), also referred to as consumer packaged goods (CPG), are those products that are generally high in demand, affordable, and sold out in a short period of time (less than a year). These items are considered “fast-moving” as they are used by consumers on a regular basis and are quick to leave the store. These items generally include packaged food, toiletries, stationery, medicines, cleaning & laundry products, plastic goods, less expensive electronics, etc. Some of the FMCG are highly perishable, such as dairy products, fruits, vegetables, meat, etc. Sales of FMCG are most affected by season periods and holidays. There are various international and national/local companies working in the FMCG industry. These companies are always looking for products that consumers use on a daily basis and that are readily available at low prices. Some of the international companies working on FMCG are Nestle, coca-cola, jobs, Tyson Foods, etc. [1] FMCG moves at a very fast pace from production to distribution and finally to consumers. These products are designed to meet the basic needs of consumers, which makes them a fundamental need of the global consumer landscape. FMCG is a high-volume, high-frequence market for business since it generates a significant source of revenue due to constant demand and repeated purchases of different products. This ultimately contributes heavily to the country's GDP and economic growth. It also creates various job opportunities which will contribute to developing countries. There is a huge competition between producers, so there is a very wide range of choices for consumers to choose from, which pushes quality and innovation. Brand loyalty is the main factor for repeated purchases of the products. These products are generally affordable since they are required on a day-to-day basis. They also have a very short life in the store. These products are sold in 2 ways: D2C and B2B. In D2C, the product is directly sold from the producers to consumers, which gives companies greater control over their product presentation, price, and overall customer experience. Since it is directly sold to consumers with no third party in between, it tends to be cheaper. The only problem with this model is it is hard to scale. In B2B, the producers sell their products to different vendors like stores, businesses, or retailers. Then, they have control over the price, presentation, and overall experience of the consumer. These types of products are generally sold in wholesale distribution, which makes them reach a wide area and easy to scale. [2]

A. Problem Statement

Every business has risks and challenges. Similarly, there are also different risks and challenges in FMCG. Companies that are operating in this sector must analyze different risks to ensure long-term success. There is intense competition in the market. If a company is not able to keep up with the trend and demand, then another company will quickly take over, and once the consumer likes the brand, then they will purchase the same brand repeatedly, which might cause problems for other companies. It is also hard for new companies to compete with other well-known companies. FMCG consumers are looking for a product that has a lower price and is of high quality. So, keeping up with the competitor's price might be difficult. There is also an online retail service now. If the company is unable to adapt to online e-commerce services, then the company might get a hard time soon. Companies must manage their inventory to keep up with demand. If they are not able to keep up, then other companies will keep up with demand, and the profit that can be earned will be lost.

B. Aim/Objective of the Report

The aim of this report is to conduct an in-depth analysis of sales data from an FMCG retail store. By carefully analyzing the pattern and trend from the dataset, we can gain valuable insights into different factors that are influencing the consumers’ patterns and

sales volume. Our ultimate objective is to uncover the underlying reasons behind the peak sales period for specific categories and vice versa, which will help us understand the dynamics that are causing these fluctuations.

Additionally, we can discover the relationship between product categories and sales performance, determining the heightened demand during a particular period of time and searching for the reason behind the trend. I also aim to explore the potential influence of external factors on FMCG sales.

C. Contributions of the work connected with Methodology

The methodology that I am going to use for this project is Team Data Science Lifecycle (TDSL). We are not going to build a model; instead, we are going to make sure that data is ready for building models. All the data pre-processing and analyzing parts will be done. The workflow is described below in detail:

- Business Understanding: My business domain is FMCG which stands for "Fast Moving Consumer Goods". They are affordable goods with high demand and are used regularly. For more information regarding FMCG, refer to the introduction section.
  The key stakeholders include the owner/shareholder of the store, employees, suppliers, consumers, community, transportation providers, etc. The needs of each stakeholder are clearly identified and understood.
- Data Acquisition & Understanding: Relevant data sources were identified and collected from the point-of-sale (POS) system. Explanatory data analysis was performed to get insight into purchasing patterns.
- Modeling: Data was prepared to create relevant models from the available data by feature engineering the data. As we have raw data, various steps were performed to prepare data, like clearing out the fields with the null values from the important fields, etc so that we can get a precise analysis and build a model from it. [3]

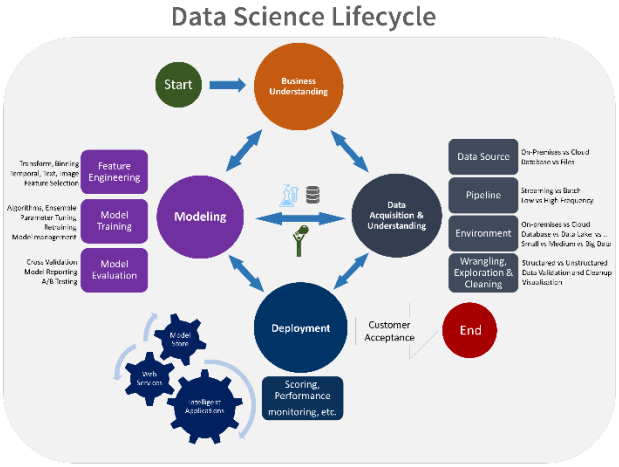


Figure 1: Team Data Science Lifecycle (TDSL)

D. Organization of the report

The report is structured in an easy to read & understand format. The report is made maintaining the IEEE standards. The report starts with Introduction which covers the background of this report. Inside the introduction section there are four subheadings explaining the problem, aim and objectives of the report, methodology used to do the task and finally the report structure is defined.

The introduction section defines and gives an overview of the FMCG industry along with a description of the significance of sales channels and market dynamics. Different identified risks are included in the problem statement. The main objective of this report is to conduct an in-depth analysis of sales data from an FMCG retail store, which is explained in the aim/objective of the report. The goal is to identify the sales pattern & understand consumer behavior along with the reasons behind the sales fluctuations. The TDSL methodology is used to complete the work, which is mentioned in the contributions of the work. Different phases, including business understanding, data acquisition & understanding, and modeling, are discussed.

II. Related Works

Multiple studies have highlighted the growing relevance of Big Data Analysis (BDA) for the Supply Chain Market (SCM). The paper titled "THE IMPACT OF BIG DATA ANALYTICS ON SUPPLY CHAIN MANAGEMENT PRACTICES IN FAST MOVING CONSUMER GOODS INDUSTRY: EVIDENCE FROM DEVELOPING COUNTRIES" by Sazu, M. H., & Jahan, S. A. (2022) explores the significance of BDA in the context of SCM in the FMCG industry. To find the impact of BDA on SCM, the survey was conducted among employees of multinational FMCG companies across various regions, including America, Asia, etc. The survey data was analyzed using systemic situation modeling in order to examine the gathered information and derive useful value from it. The results of the study have shown that the use & adaptation of BDA technologies help to generate extensive financial gain & add a lot of value to the companies. Different examples have been provided showcasing how SCM can be influenced by the use of BDA in real-world scenarios. BDA can have a significant impact on extracting valuable information from the data, which has the ability to influence a company directly. By leveraging BDA effectively, the company can stay ahead of innovations and changes in the FMCG industry, which ultimately enhances the company's competitiveness and results in the company's growth. By utilizing business intelligence insights and analytics, a company can improve its existing SCM methods, reduce cost, enhance inventory management & ultimately increase sales, which results in profit. The paper highlights the need to implement BDA in an

organization to gain a competitive advantage and make decisions based on accurate and timely insights. [4]

The paper titled "INTEGRATED BIG DATA ANALYTICS CONCEPTUAL FRAMEWORK FOR INFORMATION SHARING ACROSS THE FAST-MOVING CONSUMER GOODS INDUSTRY" proposes the conceptual framework for integrating BDA in the FMCG supply chain. It shows the importance of BDA in the current era of the Industrial Revolution. The FMCG supply chain consists of complex, dynamic entities, physical goods information, etc, which results in a generation of massive data due to its high volume, velocity, and variety. The absence of sharing information and integration deprives the FMCG supply chain (SC) of achieving balanced demand and supply, reducing inventory cost & product wastage, improve logistics efficiency. Therefore, the need for BDA has risen, and the need to utilize technologies like Hadoop to process big data and achieve streamlined real-time information, reducing costs across the supply chain. The proposed BDA framework gets its data from different sources like retailers, warehouses, manufacturers, and suppliers. It utilizes big data technology like Hadoop for data processing and storage, analysis, and sequential descriptive. The analytics outcomes are stored in the data management portal for effective visualization & data sharing. The paper also discusses the flow of the application in FMCG SC, which enables data fusion, cost reduction, improved data processing, quality assurance, risk assessment, and data-driven decision-making. It also highlights different challenges in security, privacy, and confidentiality. The BDA framework can be used in the FMCG industry to leverage data integration, analytics, and insights, which improves SC efficiency. [5]

The paper titled "PREDICTIVE ANALYSIS FOR CONSUMER DEMAND PRODUCTS IN FAST MOVING CONSUMER GOODS (FMCG) USING HADOPP FRAMEWORK) explores the application of Hadoop in analyzing data from the FMCG sector. We can process real-time data across various FMCG industries by leveraging the BDA. This paper highlights the importance of using advanced technology like Hadoop to facilitate market analysis and prediction in the FMCG industry. It also highlights the importance of machine learning (ML) techniques in analyzing data generated from FMCG. By using ML algorithms on frameworks like Apache Spark, we can show the demand and supply in a graphical format, making it easier to conclude and finally enhance the decision-making process via understanding consumer preferences. By effectively dividing the consumer into different categories, we can enhance consumer acquisition and conversion, enhance consumer loyalty, optimize marketing, and drive sales growth. To do so, the HDFS-based data pipeline is proposed. It emphasizes the impact of BDA, ML, and predictive analysis on the FMCG industry. By analyzing FMCG sales in the local market using big data, companies can gain valuable insights that can benefit them in many different ways. [6]

The paper titled "REVIEW OF BIG DATA IN FMCG SUPPLY CHAINS: U.S. COMPANY STRATEGIES AND APPLICATIONS FOR THE AFRICAN MARKET" explores the utilization of big data in analyzing the FMCG sales in local markets. In the current level of competition, it has become essential for FMCG companies to use BDA to enhance operational efficiency, pricing, advertisement, consumer engagement, etc. U.S. companies have successfully implemented BDA in order to improve demand forecasting accuracy and streamline logistics management. U.S. companies are in the lead for leveraging data-driven decisions to enhance SC. Big data helps FMCG companies to make more accurate predictions of consumer demand and reduce the risk of stockouts or overstock. It helps to manage the demand and supply ratio in fluctuating market conditions with agility. By integrating IoT devices and RFID systems, U.S. companies where able to track real-time data of goods and were able to enhance overall SC. They used BDA to gain a deep understanding of consumer behavior and created personalized marketing and product offerings. This strategic deployment of data-driven decisions by U.S. FMCG companies has reshaped the industry, which ensures efficiency, resilience, and consumer satisfaction. So, by using BDA, we can forecast real-time logistics and supply accordingly, which ensures profit. [7]

III. Methodology

A. Business Understanding

Those products that do not last longer in the store, i.e., are sold quickly compared to other products. These products are used on a daily basis and generally include items like foods, drinks, soap & shampoo, makeup, household items, etc. These products are sold quickly because people use them on a daily basis, and they have to be restocked regularly. These products are generally cheap compared to other products like electronics and furniture. Since the demand and usage are very high, companies that manufacture these products manufacture them in large quantities, which also helps to keep costs low. Since the number of users is very high and almost all the people use it, it is not possible for a company to sell directly to customers, so they rely on different supply chains to deliver their product to the end user. Since they are distributed to almost every store, it becomes convenient for consumers to buy them from anywhere they want. As almost everyone uses these types of products, there are huge amounts of companies manufacturing the same type of products. Since these products are available in almost every store, it is quite necessary for the company to build customer loyalty in order to keep up with the competition and make a profit. So, companies invest a lot of money in advertising and promoting their products so that they can retain more consumers and build loyalty.

Consumers' behavior, interests, and preferences are always changing, so it is essential for these companies to keep up with the trends and consumer interests. This includes creating a new product with new packaging, changing the recipe of their existing product, or introducing a new recipe, etc. They also need to keep up with the internet trends since they affect consumer preferences the most; for example, people started to become more health conscious as there were a lot of videos on the internet criticizing readily made foods. At this time, if the company starts making more readily made foods, then it will only result in fewer sales and maybe wastage.

Finally, the cost and quality of the product come into play. If any company is able to provide quality products at a cheaper price, then it is obvious that more people are going to buy their products. However, this is not always true; if the company has loyal consumers and provides quality products at reasonable prices (not cheap compared to other FMCG products), then consumers are still going to buy them. For example, for the price of buying one Dettol soap, consumers can buy 2 Lux soaps, but they still prefer Dettol over Lux due to its long-term trust, which, in other words, can be said to be customer loyalty.

So, in summary, the FMCG industry is about producing large amounts of products rapidly, building a brand, distributing the product, and selling products at affordable prices. Maintaining the balance between cost, quality, innovation, and brand loyalty is important to ensure long-term success.

B. Data Acquisition & Understanding

The dataset "98_95_74_FMCGDATA_2024_03_30" was retrieved from the POS system used in the local mart. The source is reliable and is real-world data. The data is complete on its own and includes all the information necessary to analyze the data and build the model.

The dataset includes a wide range of information that is critical for managing operations and analyzing the performance of the store. Each row of the dataset represents a unique sale identified by an invoice number and a sales ID. It includes the date of the order and the date of the invoice, which can be used to track the time taken to process orders. Each sale is linked to a specific product, identified by product ID, name, and some alias. It also includes the sales price and the maximum retail price (MRP) of each product, which can be used to analyze the store's pricing strategy. The product added and updated can be used to track the lifecycle of the product in the store. The sold quantity field represents the quantity of each product being sold in each sale which can be used to analyze sales volume.

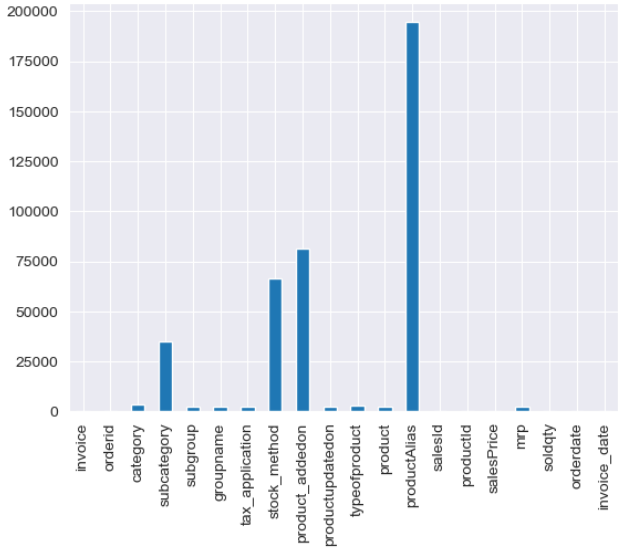
Overall, this dataset provides information that can be used for a wide range of analyses, from operational management to data-based decision-making. It provides a detailed picture of the store sales activity, product range, etc.

C. Modeling

a. Data Preprocessing

Data preprocessing is the process of making the data ready for visualization and making a model. This process involves handling missing values, removing unnecessary columns, etc.

While preprocessing the data, there were a lot of missing values on the dataset.



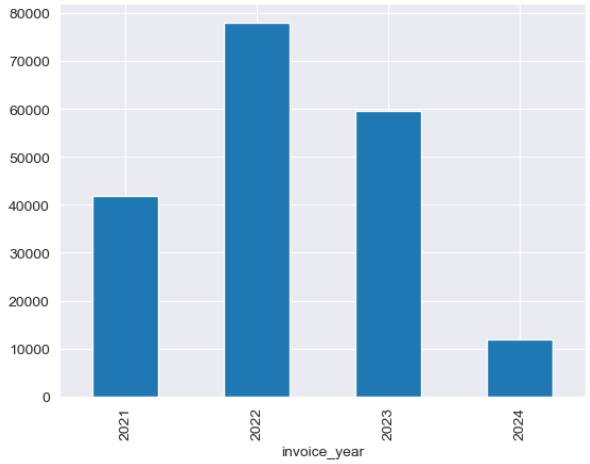
From the above diagram, we can conclude that productAlias has the greatest number of null values, followed by product_addedon, stock_method, and subcategory. To address this, columns named productAlias & stock_method columns are dropped since we won't be using them in our analysis. All the null values of columns category, groupname, tax_appliation, product, mrp & product_addedon are dropped. So now we are left with the dataset, which doesn't have any null values and can be pre-processed further.

Other columns that will not be used are also dropped, which include invoice, ordered, subcategory, subgroup, productupdatedon, salesId, productid & typeofproduct.

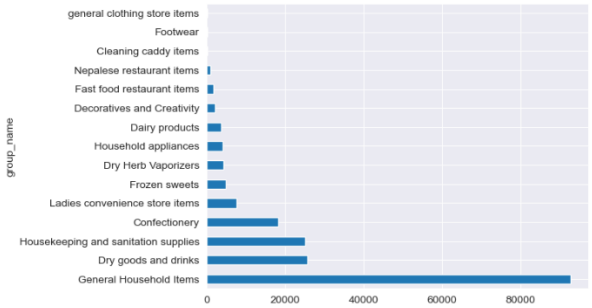
The column groupname consists of a numerical value that is obtained by one-hot-encoding. We need to give each of them a categorical value for meaningful insights. So, the names for groupname are given in a new column named group_name.

b. Data Analysis

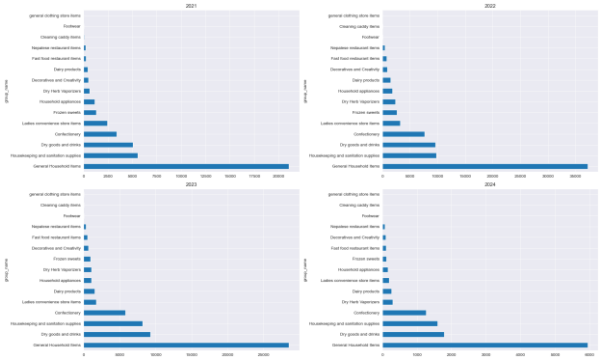
The process of methodically using statistical and/or logical tools to explain and depict, summarize, analyze, and assess data is known as data analysis [8]. Here, we will analyze the data and try to find out why it is occurring.



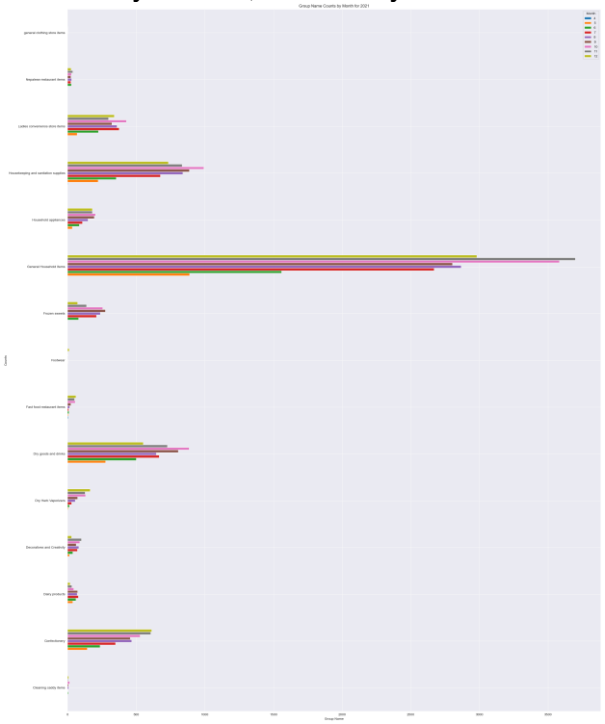
From the above image, we can say that the most sales occurred in 2022, followed by 2023, 2021, and 2024. The years 2021 and 2024 have fewer sales since the data is incomplete for these years and doesn't contain all the months of the year. The reason behind the increase in sales in the year 2022 compared to 2023 might be that the COVID-19 wave was increasing again, so people panicked and bought a lot of stuff in advance to prepare for the lockdown if the government redeclared lockdown.



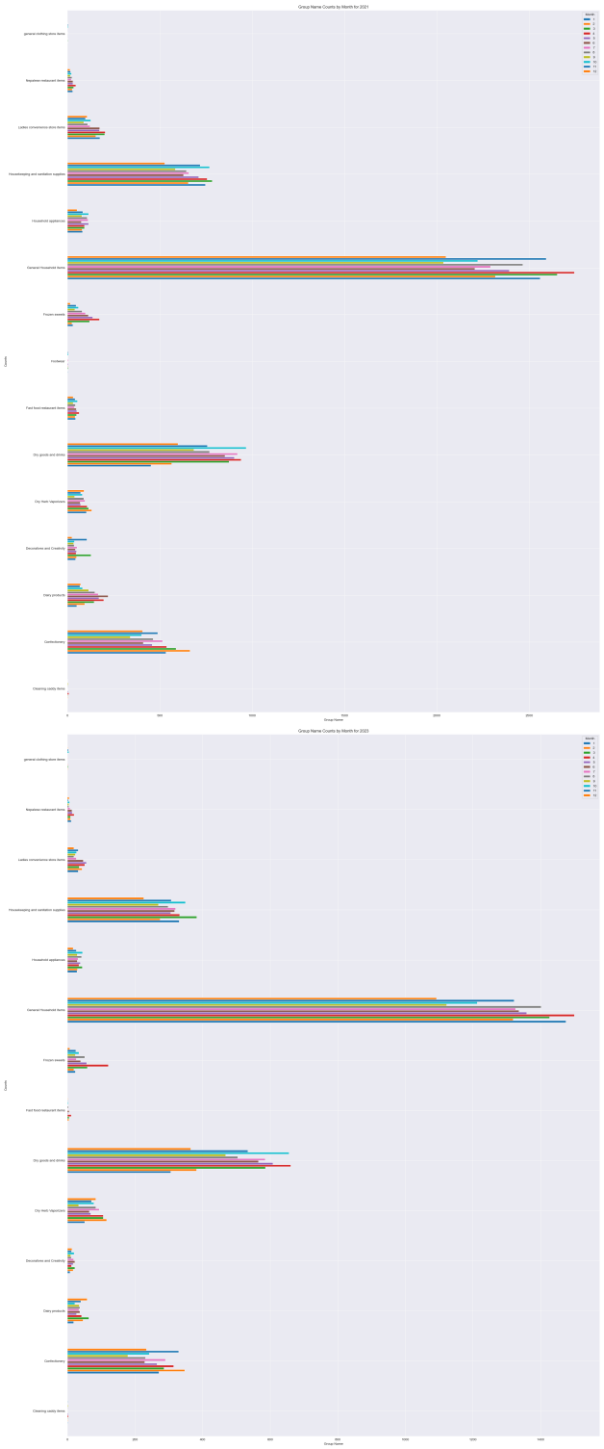
The above diagram shows the number of items sold in the store. General household items are the most sold, followed by drinks and housekeeping & sanitary items. The reason for household items having the most sales is due to the fact that they are used on a daily basis by all the families, and they need to be replaced frequently. People also buy a lot of drinks and dry foods so that they can have them for their afternoon lunch. Footwear has the minimum sales, although it is used on a regular basis because they generally last longer, about 1-2 years in general. Cleaning caddy items also have minimum sales because people don't clean their houses daily, and the majority of Nepalese prefer to use old clothes and water for cleaning. The only things they buy the most include bathroom cleaners, which can be bought at cheaper and in more quantity by going directly to the place where they are manufactured. People also get the job done by using the things that are left in the home, so the purchases of cleaning caddy items are low.



From the above graph of sales in different years, we can conclude that most of the household items were sold in the year 2022, followed by 2023.

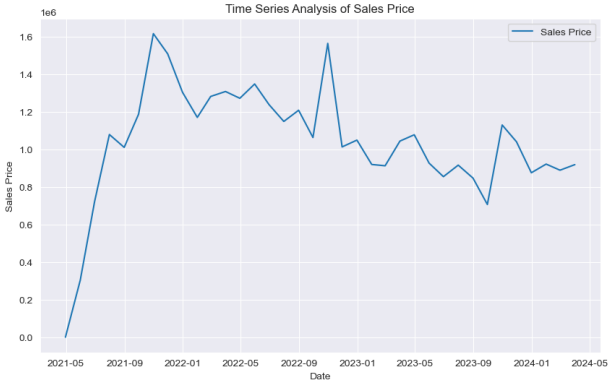


From the above analysis of sales per month of the year 2021, we can see that the sales are highest in November & December, which is Mangsir according to the Hindu calendar. This month, the majority of marriage ceremonies take place in Nepal. Since marriage includes a lot of rituals, household items are highly used. So, it may be one reason for the highest sales of household items in this period of time. From this analysis, we can conclude that the store needs more items in inventory when the Mangsir month is approaching so that they can make a maximum profit by selling household items, as there will be high demand for them during this period.

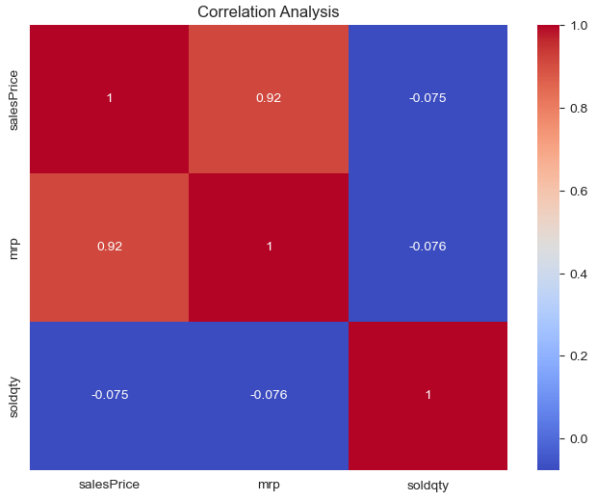


The general household items sales are maximum in the month of April in both 2022 & 2023. As we have complete data for these 2 years, we will assume that the other years may also have the highest sales in 2022 & 2023 until and unless an external factor is affecting it. One of the reasons there is the highest sale in the month of April can be due to the Nepali New Year. People celebrate the new year by visiting different places, meeting family members, and enjoying different varieties of sweets and foods. This all directly affects the sales of the store and results in the maximum number of goods being sold.

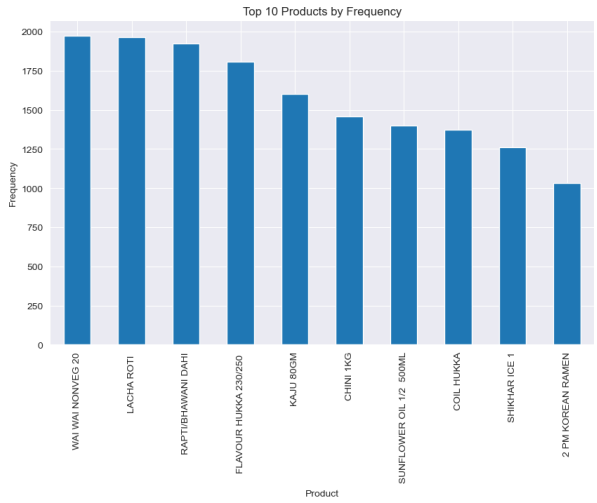
Dry foods and drinks are sold most in the month of October. The biggest festivals of Nepali are Dashain and Tihar, which are typically in Asoj and Kartik following the Hindu calendar, and September-October in the Gregorian Calendar. In Dashain and Tihar, people go to their village to celebrate it. The stored data is from the city where there are fewer people during these festivals, so the overall sales that are shown in the data are low, but the sales will be highest all over the country in this time period. People generally eat homemade dishes and to complement it different drinks and dry foods are bought. So, the reason for the spike in sales of drinks and dry foods during this time period could be the effect of festivals. The housekeeping and sanitary appliances sales also spike during this time period since, to celebrate the festival, people clean their entire house, and after celebrating the festival, the entire mess that was created will also be cleaned out. So, for cleaning purposes, sanitary appliances are bought a lot for these particular festivals.



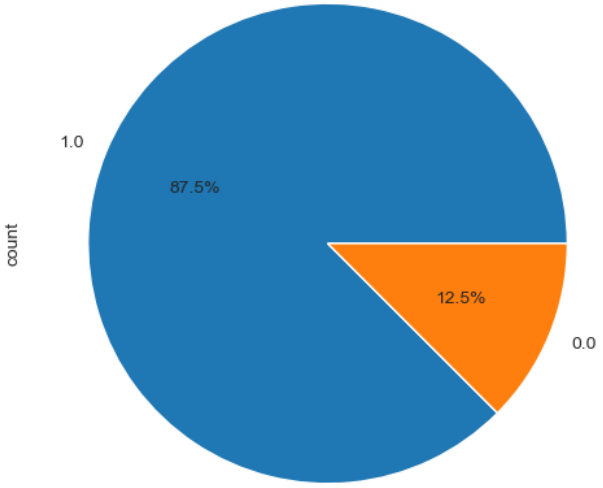
The above graph represents the time series analysis of sales prices over a period from May 2021. There is the most noticeable peak around November 2021 followed by a sharp decline and subsequent fluctuations. By analyzing the overall trend, the sales price appears to be on a downward trajectory, followed by some spikes. The peaks are becoming lower over time. By this, we can conclude that the store is not able to do as well as it used to before (2021-2022). This by be due to the rise of online retail stores. People started to go to local stores less often and started to use online services on a daily basis. So, to get back on the previous track, the store must use online services soon.



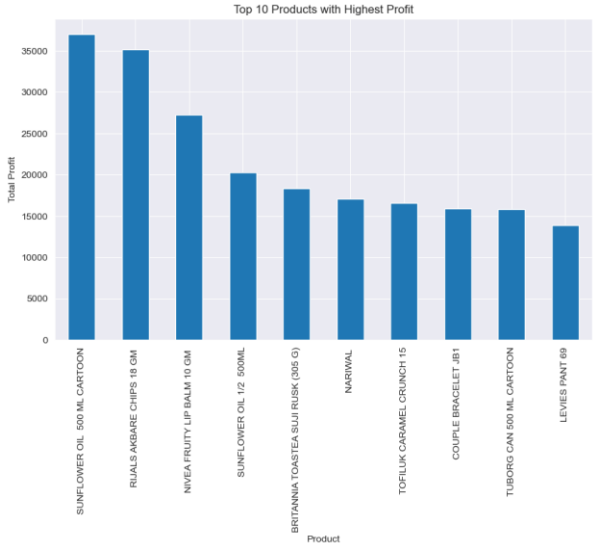
The heatmap represents the correlation between two entities ranging from 1 to -1. 1 represents the perfect positive correlation, 0 represents no correlation, and -1 represents a perfect negative correlation. The correlation value between sales price and MRP is 0.92, which indicates that they are strongly correlated. This indicates that as the MRP increases, the sales price also tends to increase and vice versa. The correlation value between sales price and sold quantity is -0.075, which indicates a weak negative correlation. This indicates that as the sales price increases, the quantity of sold items decreases slightly and vice versa. However, this relation is weak, which implies that other factors may have a more significant impact on the sold quantity than the sales price. The correlation value between MRP and sold quantity is -0.0076, indicating a weak negative correlation. This indicates that as the MRP increases, the sold quantity decreases slightly and vice versa. Similar to the previous relation, this is a weak correlation, which indicates that the sold quantity is not heavily influenced by the MRP alone.
The heatmap shows a strong positive correlation between the sales price and MRP, which indicates that pricing strategies are closely aligned with the retail price.



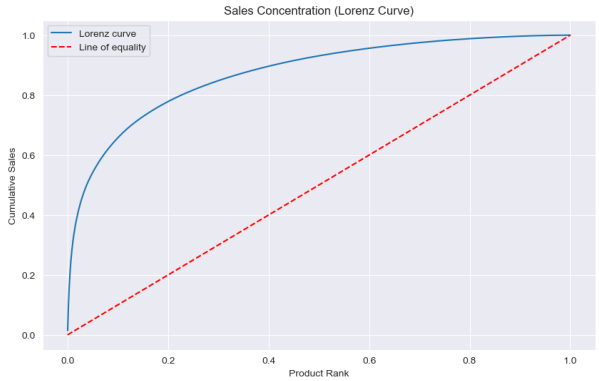
The above bar graph shows that the most sold item is Waiwai noodles. Noodles are widely eaten in lunch, making various recipes for it. They are also used heavily on street foods. Lacha roti is also eaten, especially at lunch. It is a variant of paratha from the North India. People prefer to use readily made foods rather than make them from scratch by themselves, especially if they are busy. This might be one of the reasons for the high sales of lacha roti. It is believed that if we eat curd before going/doing some auspicious thing, the chance of getting success will increase. It also provides a cool feeling in the hot seasons. It is also used in all the ceremonies that are conducted. So, the use of curd is very high, and this might be the reason for the high purchase of curd. Hukka has gained a lot of popularity in the teens recently as a more safe alternative to cigarettes. They also use hukka to look cool, so the hukka is used very widely in clubs, hotels, etc. With the introduction of e-hukkas, i.e., vapes, the flavor of hukka has also been purchased highly. So, this hype may be the reason for the high sales of hukka & flavors. Dry foods like kaju are considered to be healthy and provide us with more energy to do different kinds of things. It is also used heavily in Kheer. This might be the reason for its high sales. Sugar and oil are used almost regularly in our kitchens. The daily usage may be the reason for its high sales. 24.2% of men smoke cigarettes, and 5.9% of females smoke cigarettes. Among them, 17% of men smoke on a daily basis, while the remaining smoke occasionally [9] [10]. The high usage of cigarettes should be the reason for the high sales of cigarettes. The total number of Nepali who are in South Korea in 2022 is 80,161 & The numbers are expected to grow rapidly [11]. While this is only the number of workers, the number of students is not mentioned and is expected to be more than workers. This sudden popularity of Korea has made a trend of Korean foods and has been able to gain growth. This may be the reason for the growth of Korean food.



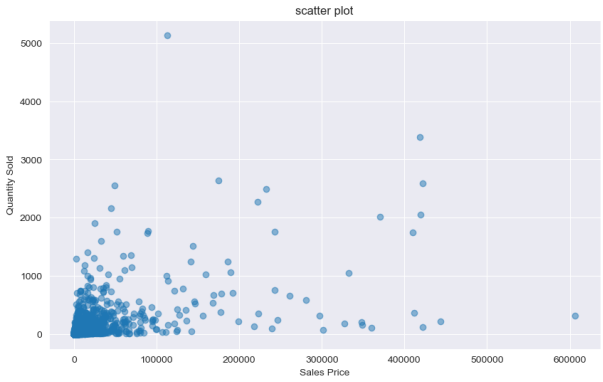
The above pie chart shows the tax-applicable goods. The 87.5% of the goods that are sold are tax applicable, and 12.5% of the goods that are sold don't apply tax. So, the majority of products that are sold apply tax.



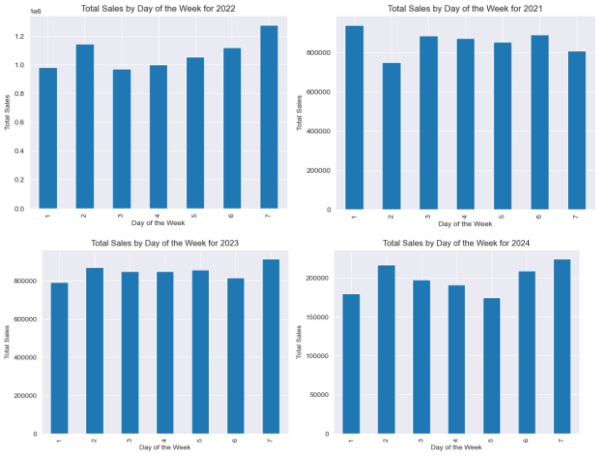
The above chart represents the list of products that make the highest profit for the store. Sunflower oil gives the most profit since it is one of the most used items in the kitchen and needs to be purchased frequently. Chips give the second highest profit since people love chips and have them frequently when they are doing something just to keep their mouths busy. Almost all the products that give the highest profit are food items, and some are accessories that are used on almost a daily basis.



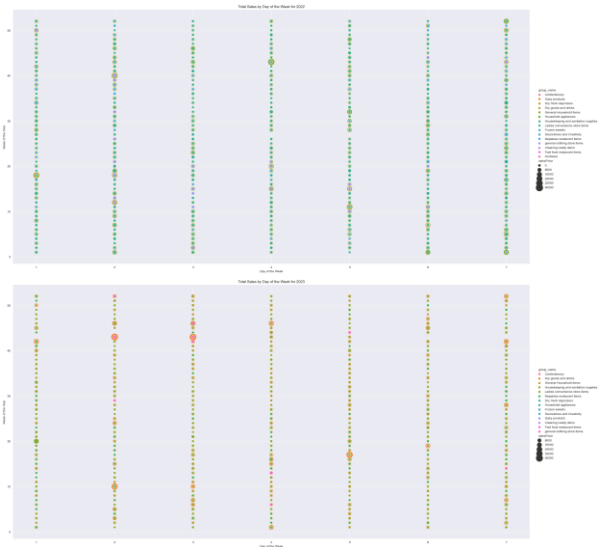
The above curve is the Lorenz curve which represents the cumulative distribution of sales across different products, while the line of equality represents a perfectly equal distribution of sales. The Lorenz curve deviates significantly from the line of equality, which indicates a high level of inequality or concentration in product sales distribution. Initially, the Lorenz curve has a steep slope, which implies that a relatively small portion of products account for a large share of the total sales. This means that the few of the most selling goods dominate the sales. As the curve approaches the end, it tends to flatten, indicating that many products contribute only a small portion of the total sales. The area between the Lorenz curve and the line of equality shows the degree of inequality. The larger the area, the higher the inequality. In the graph, the area is large, which indicates that the level of inequality is high. The deviation of the Lorenz curve from the line of equality suggests that there is room for improvement in sales distribution across products. By analyzing more on the sales data and identifying the factors contributing to occur this inequality the store should be able to implement better strategies to balance sales across different products. [12]



The above graph is the scatter plot. It shows the plot of sales price and quantity sold for different products that are available in the store. In the plot the data points are scattered across the plot which indicates that the wide range of quantities with different sales price are sold for different products. There are higher concentration of data points in the lower left side of the plot which indicates that many products have lower sales price and lower sales quantities. There are few outlies with higher sales prices and quantities. These outliers could represent popular products with higher demand and prices. This plot shows the relation between price and quantity. Higher the price lower the sales and vice versa.



In all four years, the graph shows a consistent sales pattern with higher sales towards the end & start of the week compared to the other days. This indicates that consumer demand and purchasing behavior are high during weekends. In 2021 and 2022, the highest sales occurred on the 7th day & 1st day, while the lowest sales occurred on the 2nd and 3rd day. In 2023 and 2024, the sales peaked on the 7th day, but the lowest was on the 5th and 6th day. So, we can say that consumers purchasing behavior tends to be higher at the end and start of the week.

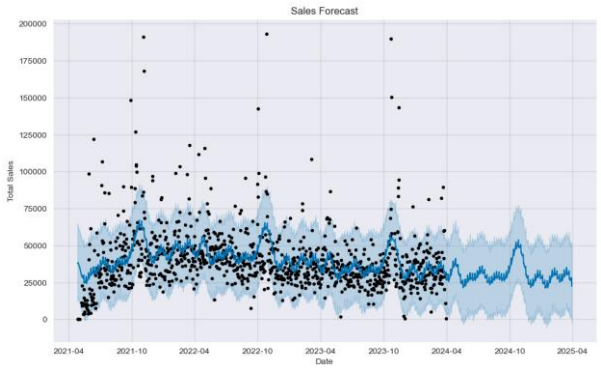


In the first image, we can see distinct sales patterns across different days of the week for most product categories. This also indicates that consumers' demand and purchasing behavior for different products vary significantly depending on the day. Categories like “dry goods and drinks” and “general household items” show relatively high and consistent sales volume throughout the week. These represent essential or frequently purchased items. There are some noticeable sales spikes for certain categories on certain days, like “Seasonal household items” shows a spike on Wednesday. “General clothing store items” shows a spike on Saturday, aligning with typical weekend shopping patterns, “decorative and creativity” also shows spikes on Saturday, showing that people tend to follow their hobby and interest on the weekends & some functions also typically happen on these days, so decorative items are purchased during weekends. Some categories like “Dry Herb Vaporizers” and “Frozen sweets” exhibit relatively low sales volumes consistently across all days.

The sales pattern appears more consistent across days of the week compared to 2022, with fewer distinct spikes for most categories. Similar to 2022, categories like “dry goods and drinks” and “general household items” maintain higher sales levels throughout the week. There are some noticeable spikes, like “household items” on Wednesday, “General clothing store items” on Saturday, and “Decorative and Creativity” on Friday and Wednesday. On the weekends, sales are slightly elevated for certain categories. Categories like “Dry Herb Vaporizers” and “Frozen sweets” exhibit lower sales across all days similar to 2022.

While overall sales patterns remain consistent, certain items are influenced by factors like consumer preferences, external events, etc. Using this data, different targeted advertisements, promotions, etc. can be done.

c. Sales forecasting



The above graph is the sales forecasting done using the Prophet library-developed meta. The forecasted sales trend which is represented by the blue line. It captures the cyclical patterns that are observed in historical data. This projection indicates that future sales are expected to follow similar patterns, with spikes and dropdowns occurring in regular intervals. The forecasted sales volumes appear to follow the historical sales pattern without deviating significantly from the established patterns. However, the forecast smooths out the variability and volatility that were observed in the historical data, underestimating or overestimating actual sales volume during peak periods. The forecast provides guidelines for anticipated sales volume based on historical patterns, and it may not have captured events or outliers accurately. Additionally, the forecast also doesn't account for consumer behavior or other external factors that can impact future sales patterns. By combining the forecast with additional sources of data, we may be able to accurately forecast sales. We should continuously monitor sales and update forecasting models as new data becomes available so that we can get more accurate forecasting.

IV. Result and Discussion

The analysis of sales data of FMCG retail stores has provided valuable insights into consumer behavior, sales patterns & factors influencing the demand for various categories. The major findings of the dataset are discussed below in detail.

- Sales trend and patterns: The analysis has revealed significant differences in sales across different years, with the highest sales record in 2022, followed by 2023; 2021 and 2024 are not included in the highest because they consist of incomplete data. The spike in sales in 2022 could be affected by the COVID-19 pandemic, which led to panic buying and stocking products in case of potential lockdowns. The analysis has also highlighted distinct sales patterns across different months and on different days of the week. Higher sales were observed during weekends, with a spike on Saturdays and Sundays, which indicates that consumer demand is heavily influenced by the day. The findings have shown that consumers engage in their

hobbies during weekends and also buy necessary stuff on the weekends.

- Product category analysis: The analysis has shown that general household items are the most sold items & have dominated the overall sales of the mart followed by dry goods and drinks, and housekeeping and sanitary items. This can be caused due to the fact that these items are essential and used on the regular basis in households. Certain product categories show distinct sales patterns across different months and weeks throughout the year household items experienced a significant spike during November and December (Mangsir according to the Hindu calendar), when the majority of marriage ceremonies take place in Nepal. This can cause an increasing demand for household items during wedding ceremonies and different rituals. Similarly, dry foods and drinks, along with housekeeping appliances, experienced a spike in sales during October, which aligns with the festival season of Dashain and Tihar. These days, people prefer to drink and eat spicy and other things along with meat and clean their houses for the festival which leads towards the increase in demand for dry foods, drinks and cleaning products.
- Sales price and Quantity analysis: The analysis of the correlation between sales price and maximum retail price (MRP) reveals useful insights. A strong correlation i.e., 0.92 was seen between sales price and MRP which indicates that whenever one increases another also increases. A weak negative correlation of -0.075 is seen between sales price and sold quantity which indicates that an increase in sales price results in a slight decline of the quantity of items sold and vice versa; but since it is very weak we cannot exactly say that this will happen for sure. It is similar to MRP and sold quantity i.e., a weak correlation of -0.0076.
- Sales distribution and inequality: The Lorenz curve graph has shown that the Lorenz curve have deviated too much from the line of equality which indicates a high level of inequality/concentration in product sales distribution which implies that a small portion of products are responsible for a large portion of total sales while many products contribute on small portion of the total sales. The scatter plot analysis further validated the conclusion done from the Lorenz curve, with a higher concentration of data points on the lower left side of the graph, which shows the product of lower sales with lower prices. There are also outliers with high sales prices and volume. This inequality in sales distribution shows the opportunity for the store to manage its sales by identifying the key factors contributing to this inequality, which can be later used to do targeting marketing and optimize the store pricing strategies.
- Sales forecasting: The sales forecasting analysis is done by using the Prophet library, which will provide future sales trends. The forecasted sales trend captures the cyclical patterns which are observed in historical data which concludes that the future sales will follow the trend of past data along with similar patterns with spikes and downfall that are occurring in certain intervals. It should be noticed that forecasted output have possibly overestimated or underestimated actual sales volumes and trends. The forecast should be updated on recent data in order to make decisions based on it. To enhance the accuracy of sales forecasting by identifying the variables that are affecting the sales.

V. Conclusion

In conclusion, the FMCG sales data provides a comprehensive understanding of consumer behavior, sales patterns, and different factors that affect consumer behavior and sales. These findings helps to make a decision based on the analysis, which helps companies to optimize their optimizations, pricing strategy, inventory management and targeted marketing by continuously monitoring sales data which will result in profitability of the company. This analysis helps to be ahead of the competition in the FMCG industry.

VI. References

[1] CFI Team, "Corporate Finance Institute," 31 May 2023. [Online]. Available: https://corporatefinanceinstitute.com/resources/valuation/fast-moving-consumer-goods-fmcg/. [Accessed 22 April 2024].

[2] Appinio Research, 24 November 2023. [Online]. Available: https://www.appinio.com/en/blog/market-research/fmcg-fast-moving-consumer-goods. [Accessed 22 April 2024].

[3] Marktab, "What is the Team Data Science Process? - Azure Architecture Center," [Online]. Available: https://learn.microsoft.com/en-us/azure/architecture/data-science-process/overview. [Accessed 22 April 2024].

[4] M. H. Sazu and S. A. Jahan, "THE IMPACT OF BIG DATA ANALYTICS ON SUPPLY CHAIN MANAGEMENT PRACTICES IN FAST MOVING CONSUMER GOODS

INDUSTRY: EVIDENCE FROM DEVELOPING COUNTRIES," *International Journal of Business Reflections,* vol. 3, no. 1, pp. 112-128, 2022.

[5] R. Bhekifa , A. Michael and Y. Sarma, "Integrated Big Data Analytics Conceptual Framework for Information Sharing across the fast-moving consumer goods industry," *2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE,* pp. 1-6, 2022.

[6] M. Thejovathi, D. M. Chandra and S. Rao, "Predictive Analysis For Consumer Demand Products In Fast Moving Consumer Goods (FMCG) Using Hadoop Framework," 2019.

[7] O. I. Oriekhoe, B. I. Ashiwaju, K. C. Ihemereze and U. Ikwue, "REVIEW OF BIG DATA IN FMCG SUPPLY CHAINS: U.S. COMPANY STRATEGIES AND APPLICATIONS FOR THE AFRICAN MARKET," *International Journal of Management & Entrepreneurship Research,* vol. 6, no. 1, pp. 87-103, 2024.

[8] "Responsible Conduct in Data Management," [Online]. Available: https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html. [Accessed 26 April 2024].

[9] "'Smokers are not selfish' campaign to motivate people to quit smoking," 9 February 2024. [Online]. Available: https://kathmandupost.com/health/2024/02/09/smokers-are-not-selfish-campaign-of-motivating-people-to-quit-smoking. [Accessed 28 April 2024].

[10] "Nepal STEPS Survey 2019," 2019. [Online]. Available: https://nhrc.gov.np/wp-content/uploads/2019/11/Tobacco-Fact-Sheet-1.pdf. [Accessed 28 April 2024].

[11] "Nepali workers in top priority in Korea: HRD VP," The Rising Nepal, 6 March 2023. [Online]. Available: https://risingnepaldaily.com/news/23538. [Accessed 28 April 2024].

[12] S. M. Ross, "Introductory Statistics," in *Introductory Statistics*, Academic Press , 2017, pp. 120-127.