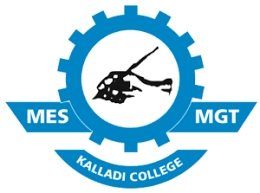
PROJECT REPORT

**Rahul**

**MES KALLADI COLLEGE MANNARKAD**

Mannarkkad, Palakkad, Kerala



An Internship Report

On

**“Data Science and Analytics using Excel & Power BI”**

Along with the project

“VRINDA STORE DATA ANALYSIS”

Submitted in Partial Fulfillment of the requirement for the award of the degree

**BACHELOR OF VOCATION**

**IN**

**DATA SCIENCE & ANALYTICS**

Submitted by

Name: RAHUL K

Reg.No: KIAWBOE024

Under the guidance of

Internal Guide

**Hasfiya KP**

**Dept. of B.Voc Data Science & Analytics**

**COMPANY CERTIFICATE**

**DECLARATION**

I, **RAHU K**, student of II semester B.voc Data science & Analytics a M.E.S Kalladi College Mannarkkad, Palakkad, hereby declare that the Internship work entitled “VRINDA STORE DATA ANALYSIS” has been independently carried out by me under the supervision of **SHABNA**, Head of department of B.voc data science & analytics, and the coordinator **HASFIYA KP**, Assistant Professor, submitted in partial fulfillment of the course requirement for the award of degree **Bachelor of vocation in Data Science & Analytics** of Calicut University, during the year 2023. I further declare that the report has not been submitted to any other University for the award of any other degree.

**PLACE: MANNARKKAD STUDENT NAME: RAHUL K**

**Date: 30 July 2023 REG.NO: KIAWBOE024**

**ABSTRACT**

This data analytics project focuses on addressing business challenges through effective data analysis using Microsoft Excel. The project aims to enhance skills in various data-related tasks such as data cleaning, data processing, data analysis, data visualization, and generating actionable insights.

**Objectives:**

- Utilize Excel to clean and preprocess sales data from Vrinda Store. - Apply data analysis techniques to uncover patterns, trends, and key metrics within the sales dataset. - Create informative visualizations to represent sales data visually and intuitively. - Generate insightful reports that highlight important findings and recommendations for business decision-makers. **Methodologies:**

1. Data Cleaning and Preprocessing: The project will involve cleaning the raw sales data, handling missing values, and ensuring data integrity for accurate analysis.

2. Data Analysis: Various analytical methods will be employed to examine sales trends, customer behavior, popular products, and revenue patterns.

3. Data Visualization: Utilizing Excel's charts and graphs, visual representations of sales data will be created to communicate insights effectively.

4. Report Generation: A comprehensive report will be generated, summarizing the analysis results, visualizations, and actionable recommendations.

**Expected Outcomes:**

- Enhanced proficiency in data manipulation, cleaning, and analysis using Excel.

- Profound understanding of business problem solving through data-driven insights.

- Creation of insightful visualizations and reports that aid decision-making.

- Improved skills in translating raw data into meaningful insights for business stakeholders. **Scope:**

The project's scope encompasses the analysis of sales data from Vrinda Store. It involves extracting, transforming, and analyzing data related to sales transactions, customer behavior, and product performance.

**Dataset:**

The project will utilize sales data from Vrinda Store to showcase the application of data analytics techniques in a real-world business context.

In conclusion, this data analytics project will equip participants with valuable skills in data handling, analysis, visualization, and reporting, while addressing real business challenges faced by Vrinda Store

**TASK PERFORMED**

In this Internship Data analysis using Excel and Power BI.

**Training Program**

The internship is a platform where the trainees are assigned with specific task. In the initial days of the internship, I was trained on the following:

* Cleaning Data using MS Excel
* Analysis using Power Bi

**VRINDA STORE DATA ANALYSIS**

**INTRODUCTION**

"In today's data-driven world, the ability to extract meaningful insights from vast datasets has become a crucial skill across various industries. This project delves into the realm of data analytics, utilizing the versatile tool of Microsoft Excel to explore, analyze, and visualize complex datasets. By employing a combination of data manipulation techniques, statistical analysis, and graphical representation, this project aims to uncover valuable patterns, trends, and correlations hidden within the data. Through the application of Excel's features, this project not only demonstrates the practicality of data analytics but also showcases its potential to drive informed decision-making and enhance business strategies. Join us on this journey as we harness the power of Excel to transform raw data into actionable intelligence."

Excel is a versatile and widely-used tool for creating dashboards due to its user-friendly interface, powerful data manipulation capabilities, and the following reasons:

1. Familiarity:

Excel is a commonly-used software that many individuals, including professionals and students, are familiar with. This familiarity reduces the learning curve for creating dashboards, making it accessible to a broader audience.

2. Ease of Use:

Excel offers a range of intuitive features for data entry, manipulation, and visualization. Its drag-and-drop functionality, formula-based calculations, and user-friendly charts and graphs make it easy to construct visually appealing dashboards without requiring extensive programming skills.

3. Data Manipulation:

Excel excels (pun intended) in handling data. You can easily clean, filter, sort, and transform data using its built-in functions. This is essential for creating meaningful insights in a dashboard, as data often needs to be prepared before visualization.

4. Interactive Visualizations:

Excel provides a variety of charts and graphs that allow for interactive data visualization. From bar charts to pivot tables, these tools enable users to explore data trends, patterns, and outliers, enhancing the dashboard's effectiveness in conveying information.

5. Customization:

Dashboards often need to be tailored to specific needs and branding. Excel offers extensive customization options, allowing you to adjust colors, fonts, layouts, and design elements to match your project's requirements or your organization's branding.

6. Rapid Prototyping:

Excel enables quick creation of prototypes and mockups. This is particularly useful during the planning and development stages, as you can iterate through various dashboard designs and layouts before finalizing the most effective one.

7. Integration:

Excel supports integration with various data sources, including databases and external spreadsheets. This enables you to regularly update your dashboard with new data, ensuring that it remains current and relevant.

8. Cost-effective:

Excel is often readily available on many computers, reducing the need for additional software purchases. This makes it a cost-effective option for creating dashboards without investing in specialized dashboarding software.

In summary, Excel's combination of familiarity, ease of use, data manipulation capabilities, interactive visualizations, customization, rapid prototyping, integration, and cost-effectiveness makes it a compelling choice for creating dashboards that effectively communicate insights from complex datasets.

**OBJECTIVES THAT I WANTED TO ACHIEVE**

**1. Data Exploration**:

The primary objective is to explore the dataset thoroughly, identifying trends, patterns, and potential outliers that could provide valuable insights.

**2. Data Cleaning and Preparation:**

Ensure the dataset is clean, consistent, and formatted correctly to avoid inaccuracies in the analysis. This objective focuses on data preprocessing to improve the quality of insights derived.

**3. Statistical Analysis:**

Utilize Excel's statistical functions to perform relevant analyses, such as calculating means, medians, standard deviations, and correlations. These analyses can provide quantitative insights into the dataset.

**4. Visualization Creation:**

Develop a visually appealing dashboard using Excel's charting and graphing tools. The objective is to present data insights in a clear and easily understandable manner.

**5. Trend Identification:**

Identify and showcase significant trends, changes, or patterns within the dataset over time or across different variables. This could involve creating line charts, bar graphs, or scatter plots.

**6. Comparison and Benchmarking:**

Compare different data points or variables to uncover relationships or differences that might offer actionable insights. Excel's capabilities allow for easy creation of comparative visualizations.

**7. Data-driven Decision Making:**

Empower stakeholders to make informed decisions by presenting insights that drive actionable outcomes. The dashboard should provide clear recommendations based on the data analysis.

**8. Interactive Elements:**

Implement interactive features within the dashboard, such as drop-down menus or slicers, to allow users to customize the view and focus on specific subsets of data.

**9. User Accessibility:**

Ensure that the dashboard is user-friendly and accessible to a wide range of users, regardless of their level of familiarity with Excel or data analysis.

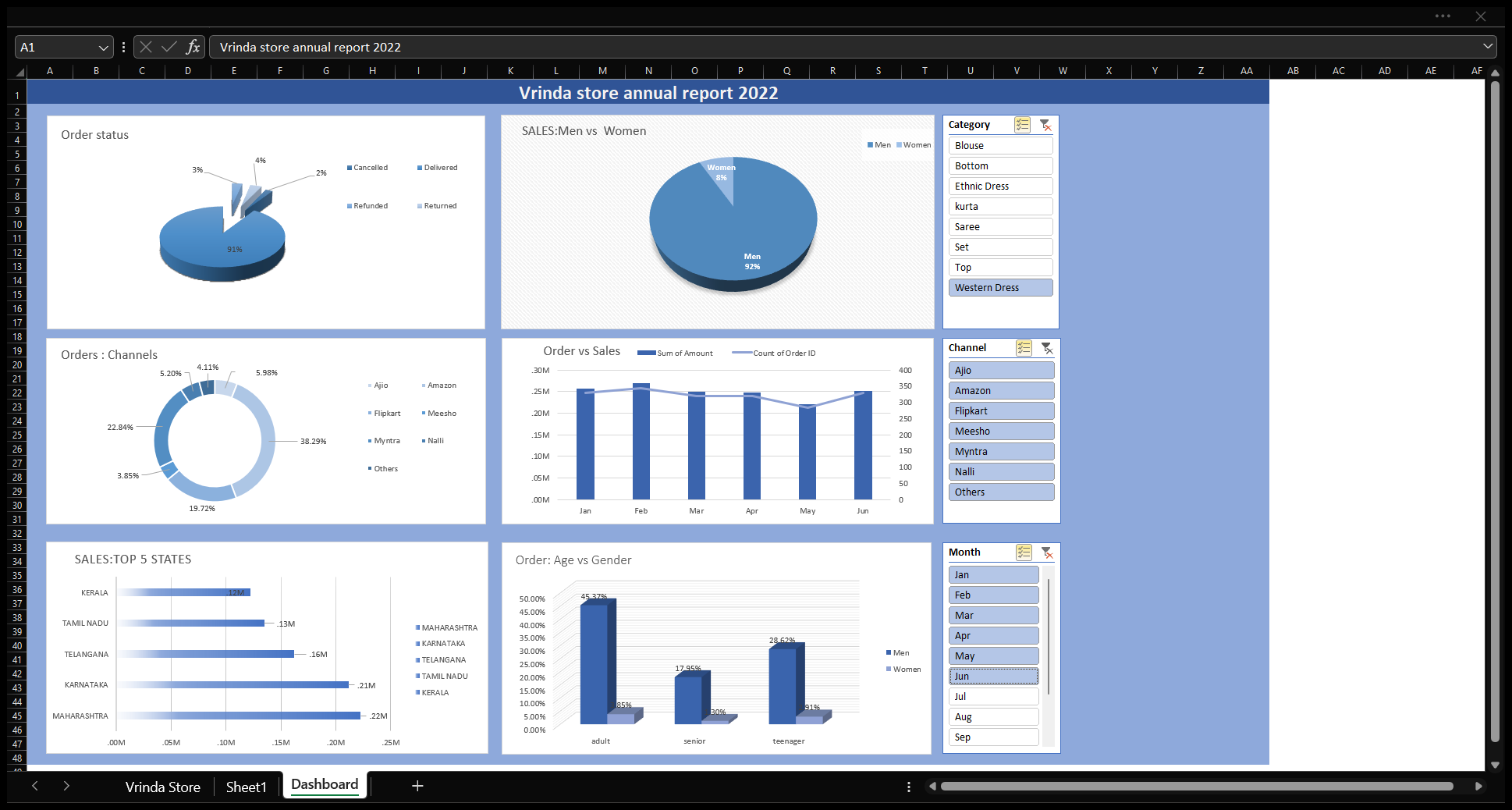
**10. Documentation and Communication:**

Document the process, methodologies, and findings of the analysis to ensure transparency and replicability. Effective communication of insights is crucial for conveying the significance of the data-driven outcomes.

**11. Continual Improvement:**

Plan for future enhancements and refinements to the dashboard based on user feedback and evolving data needs. The objective is to create a dynamic tool that adapts to changing requirements.

**DASHBOARD OVERVIEW**



Vrinda store annual report dashboard

**Dashboard Overview:**

The Excel dashboard I've designed offers a comprehensive analysis of sales and customer data for "Vrinda Store," providing valuable insights into customer behavior, sales trends, and performance across various dimensions. This dashboard serves as a powerful tool for decision-makers to make data-driven choices and optimize strategies.

**Data Source:**

The primary data source is the "Vrinda Store" dataset, encompassing essential information such as order details, customer demographics, order status, channels, product categories, quantities, amounts, and geographical attributes. This rich dataset forms the foundation for generating meaningful insights.

**Main Components/Widgets:**

1. Order Status Distribution (Pie Chart):

This pie chart visually presents the distribution of order statuses (delivered, returned, in progress), allowing stakeholders to quickly gauge the efficiency of order processing and customer satisfaction.

2. Gender-based Sales Breakdown (Pie Chart):

Another pie chart breaks down sales based on gender, providing insights into which gender contributes more to the store's revenue. This information can guide marketing strategies and product offerings.

3. Orders by Channels (Doughnut Chart):

The doughnut chart showcases the distribution of orders across different sales channels (Ajio, Amazon, Myntra, etc.), enabling stakeholders to assess channel popularity and performance.

4. Top 5 States by Sales (Column Chart):

A column chart displays the top 5 states that generate the highest sales, helping identify regional preferences and tailor marketing efforts for these areas.

5. Order Analysis: Age vs. Gender (Bar Chart):

The bar chart visually represents the relationship between customer age, gender, and order amounts. This helps identify spending patterns across different age groups and genders.

Intended Audience:

The dashboard serves the following stakeholders within "Vrinda Store":

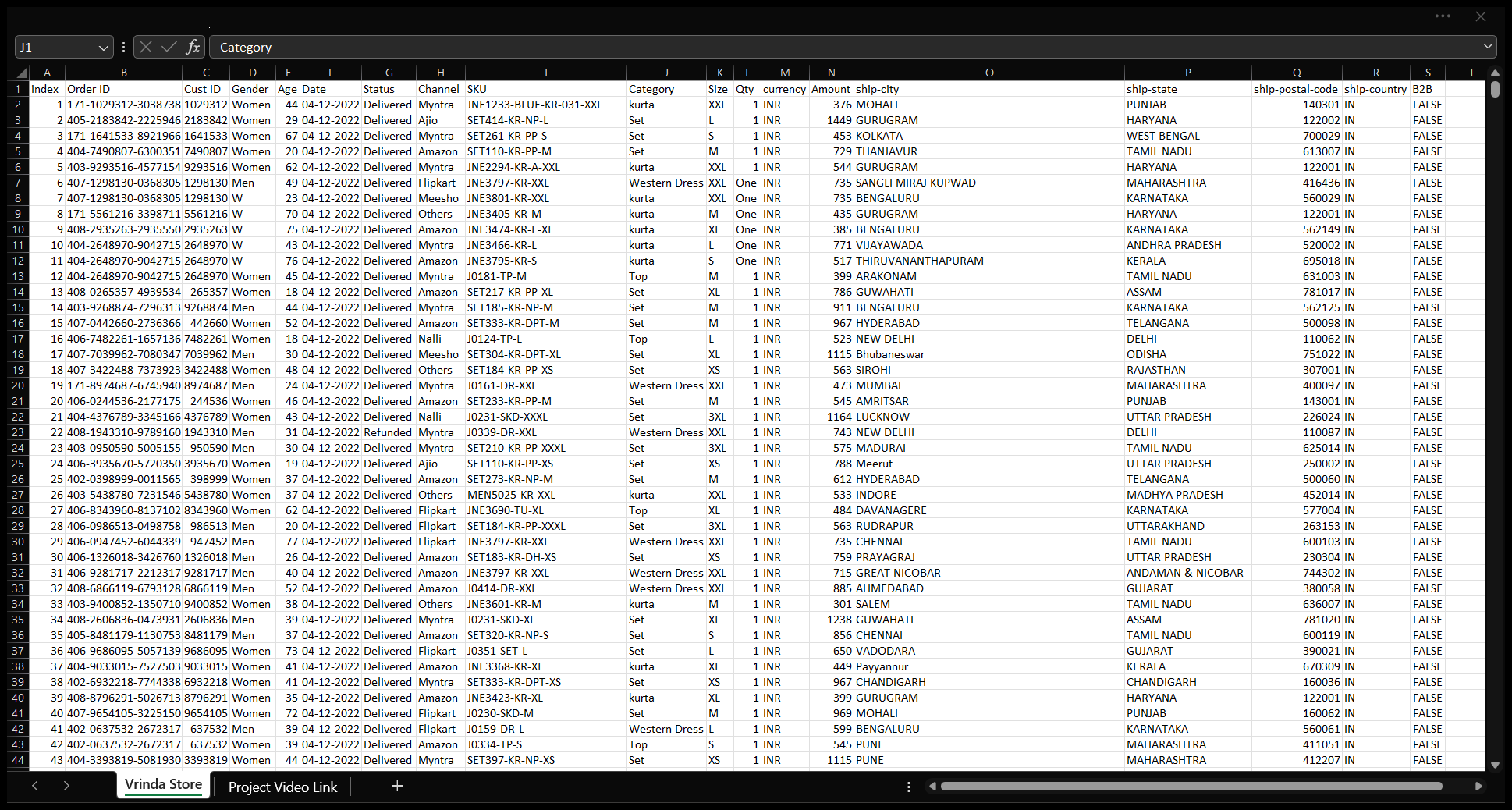
- Sales Managers: To monitor order statuses, channel performance, and regional sales, allowing for quick strategic adjustments.

- Marketing Team: To devise targeted campaigns based on gender insights and optimize marketing efforts on popular sales channels.

- Operations Team: To track order processing and customer satisfaction, ensuring smooth operations.

By leveraging the diverse "Vrinda Store" dataset, this Excel dashboard transforms raw data into actionable insights, enabling the store's team to make informed decisions, enhance customer experiences, and drive growth.

**DATA SOURCE USED**



Sample screenshot of data source used

The data source used to build the dashboard is the "Vrinda Store" dataset, a comprehensive collection of information that forms the basis for the analysis and visualization within the Excel dashboard. This dataset contains a variety of fields that collectively provide insights into customer behavior, sales trends, and order dynamics. Here's a breakdown of the key fields included in the dataset:

**1. Order ID:** A unique identifier assigned to each individual order placed by customers.

**2. Customer ID:** A unique identifier for each customer who made a purchase.

**3. Gender:** The gender of the customer (e.g., Male, Female), providing insights into purchasing behavior by gender.

**4. Age:** The age of the customer, which helps analyze spending patterns across different age groups.

**5. Date:** The date when the order was placed, allowing for time-based analysis and trend identification.

**6. Status:** The status of the order (e.g., Delivered, Returned, In Progress), indicating the progress of each transaction.

**7. Channel:** The sales channel through which the order was made (e.g., Ajio, Amazon, Myntra, Flipkart, etc.), providing insights into the popularity of different platforms.

**8. Category:** The product category associated with the order, helping analyze which types of products are most popular.

**9. Size:** The size of the product ordered, relevant for understanding customer preferences.

**10. Quantity:** The number of units ordered for each product, contributing to total sales.

**11. Amount:** The monetary value of the order, reflecting the revenue generated from each transaction.

**12. Currency:** The currency in which the transaction was conducted.

**13. Ship City:** The city to which the order was shipped, offering insights into geographical distribution.

**14. Ship State:** The state or region to which the order was shipped, providing regional insights.

The dataset's richness enables a multifaceted analysis that covers customer demographics, order details, sales performance, and channel popularity. By utilizing these fields, the Excel dashboard is able to generate visualizations that provide a comprehensive understanding of the store's operations, customer behavior, and sales trends.

DATA CLEANING , TRANSFORMATION AND PREPROCESSING.

To ensure the quality of data, I performed comprehensive data cleaning, followed by transformation to standardize and enhance features, and preprocessing to handle missing values, outliers, and improve model compatibility. This rigorous process yielded high-quality data for more accurate analysis .

**Data Transformation**:

**1. Age Group Categorization:**

- Created a new column to categorize customers into age groups based on their age.

- Defined age groups such as "senior" for ages greater than 50, "adult" for ages greater than 30, and "teenager" for other ages.

- This age categorization will provide a clear way to analyze customer behavior across different age ranges.

**2. Quantity Data Transformation:**

- Transformed textual quantity data (e.g., "one," "two") into numerical values.

- This conversion ensures that the quantity data is in a format suitable for calculations and analysis.

Data Preprocessing:

**3. Month Extraction:**

- Extracted the month from the date column using Excel's functions.

- This enables you to focus on monthly trends and identify any seasonality in sales over time.

Data Cleaning:

**4. Gender Standardization:**

- Modified the gender data to ensure consistency and clarity.

- Replaced "w" with "women" and "m" with "men" to standardize the gender representation across the dataset.

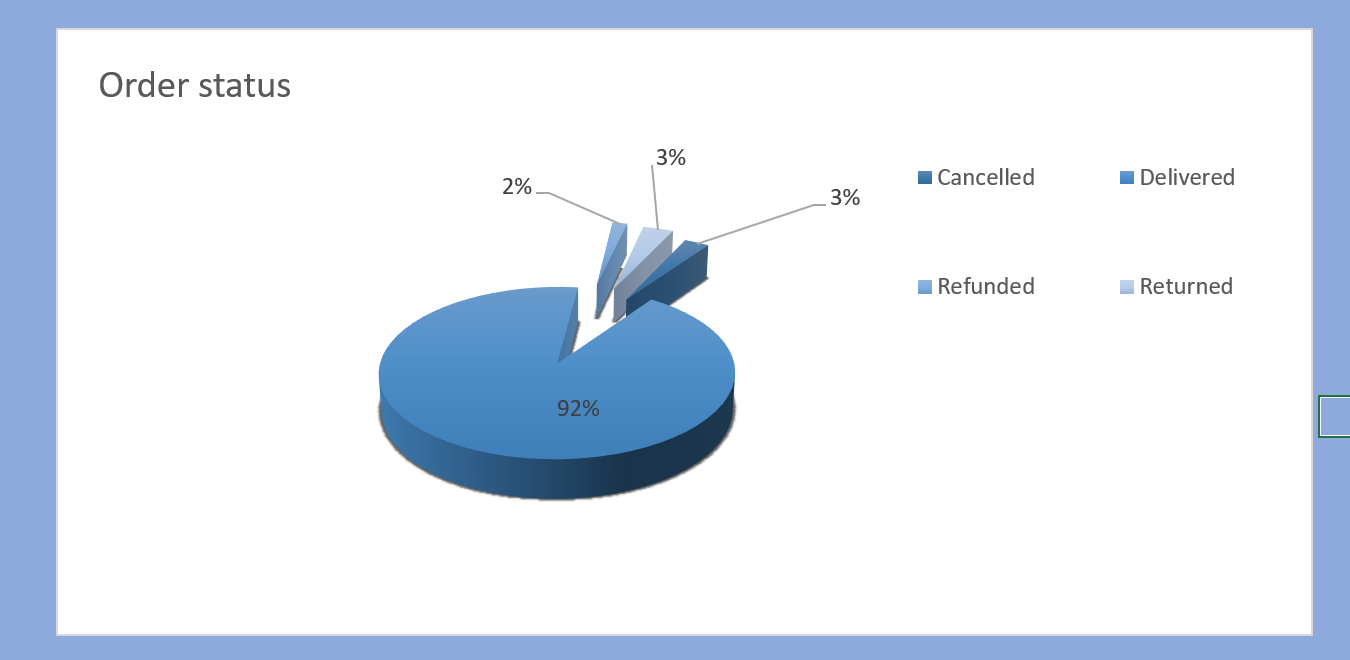
By performing these data cleaning, transformation, and preprocessing steps in Excel, I have improved the quality and usability of the dataset. These actions are essential for creating accurate and meaningful visualizations in my Excel dashboard.

These steps reflect a combination of data cleaning (ensuring consistency), data transformation (changing data format or structure), and data preprocessing (preparing data for analysis) activities that collectively prepare your dataset for analysis and visualization in my Excel dashboard.

**Visualizations used in the dashboard**

Visualizations serve as a powerful tool to comprehend and communicate complex data in a clear and accessible manner. They offer a visual narrative, illuminating trends, patterns, and outliers present within the data. In this exploration, we'll examine a variety of visual representations that provide valuable insights and aid in understanding the underlying story hidden within the numbers."

**Pie chart for order status**

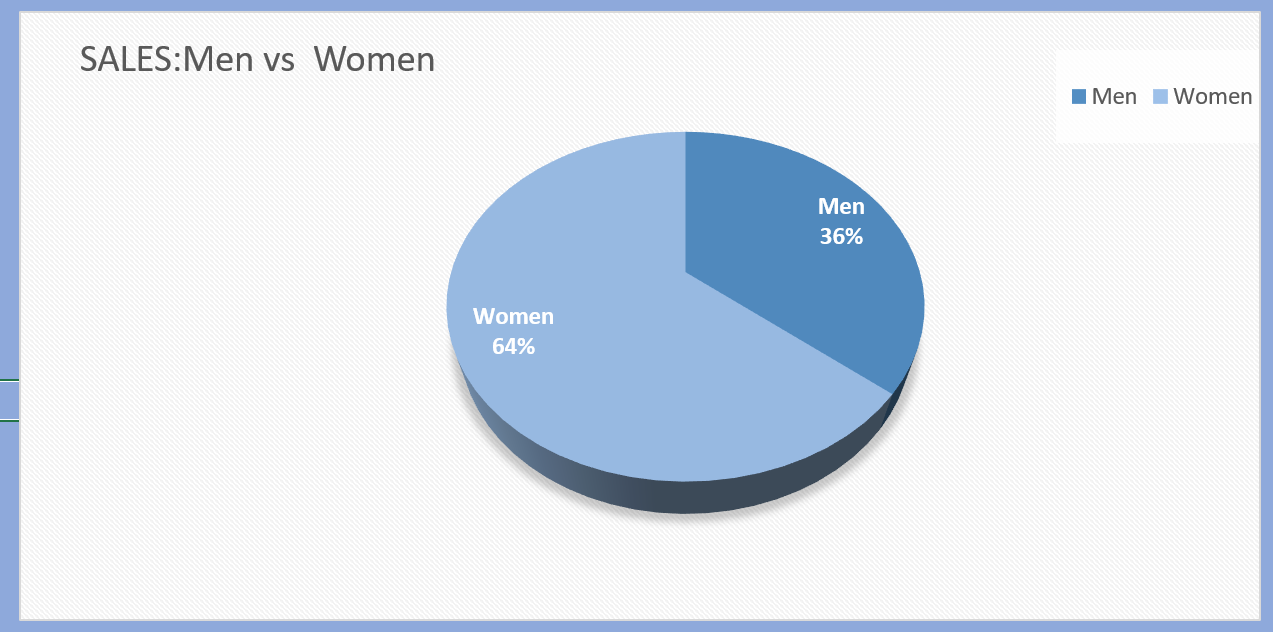
****

A pie chart is effective for showcasing the distribution of categories, making it suitable for representing order statuses.

The clear separation of slices helps viewers quickly understand the proportion of each status relative to the whole, making it an intuitive choice for displaying completion .

* Clearly represents the distribution of different order statuses (delivered, returned, refunded, cancelled).
* Provides an immediate visual overview of how orders are progressing.
* Suitable for displaying the proportion of each status category.

**Pie chart for Sales Men vs Women:**

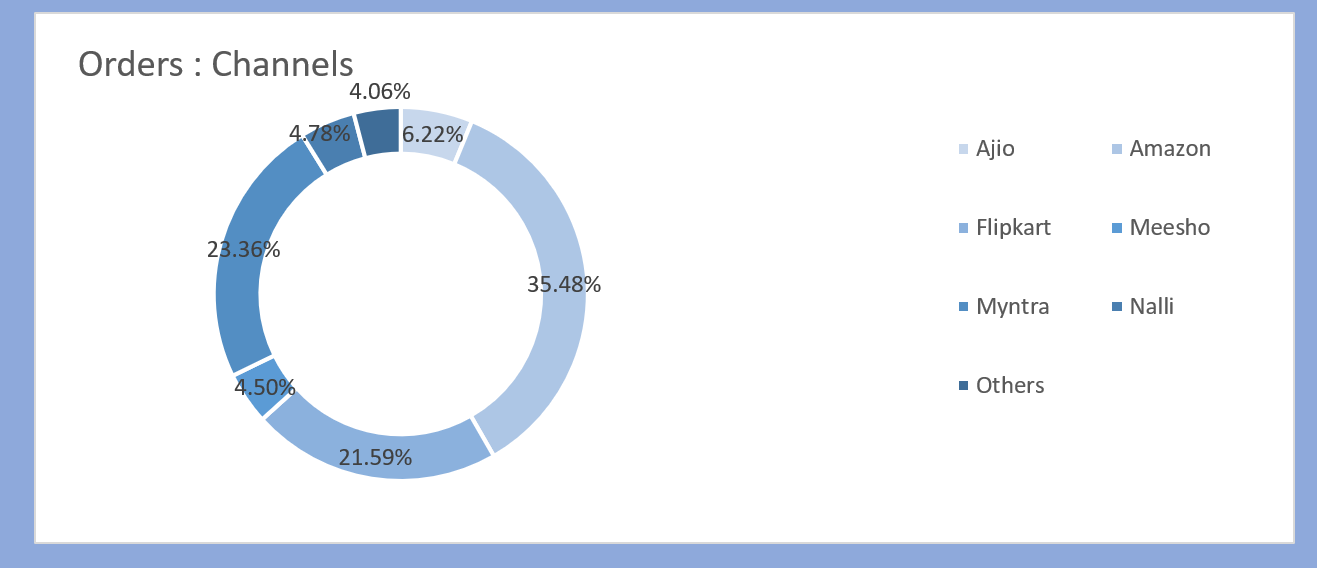
****

**A pie chart is chosen for its simplicity and effectiveness when comparing two categories (men and women).**

**The binary nature of gender makes a pie chart visually appealing and easy to interpret, providing an immediate understanding of the relative contributions of men and women to sales.**

* **Effectively compares sales contributions between men and women.**
* **Visual simplicity allows for quick interpretation.**
* **Ideal for binary comparisons like gender.**

**Doughnut Chart for Orders by Channels:**

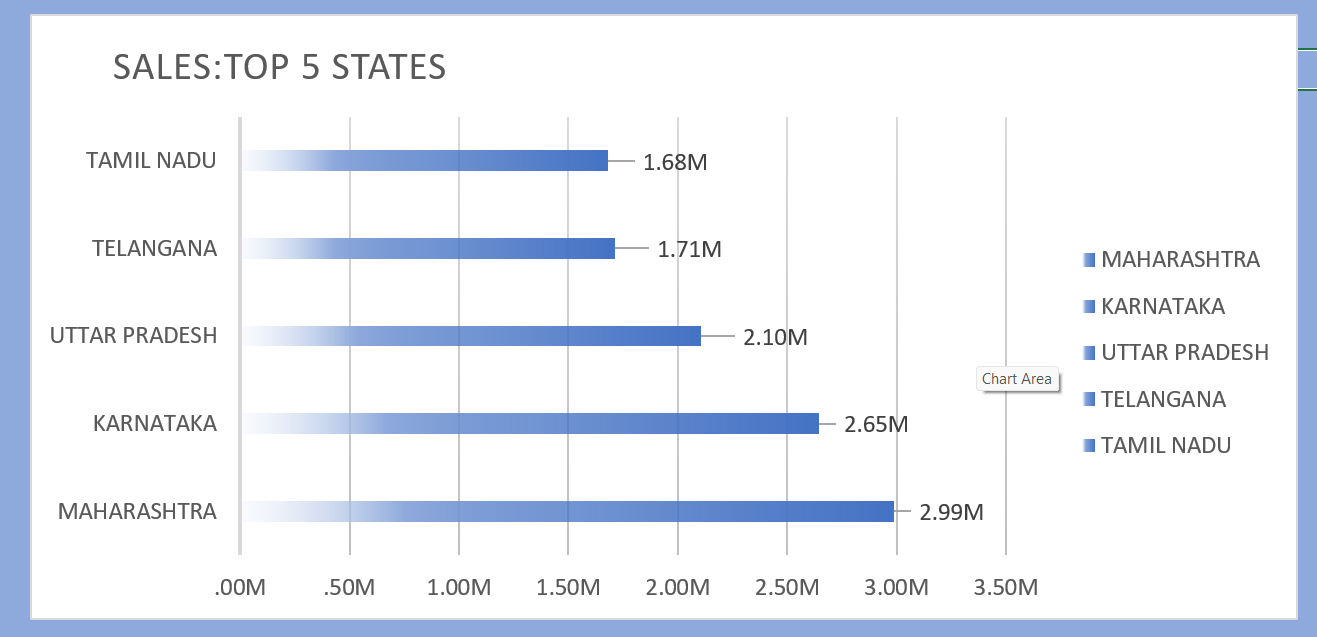
****

**A doughnut chart is suitable for displaying the distribution of orders across different channels.**

**The concentric nature of a doughnut chart allows for a clear visual representation of each channel's share, enabling easy comparison.**

* **Represents the distribution of orders across different sales channels.**
* **Allows for easy comparison between channel contributions.**
* **Provides a clear visual of each channel's share.**

**Column Chart for Sales in Top 5 States:**

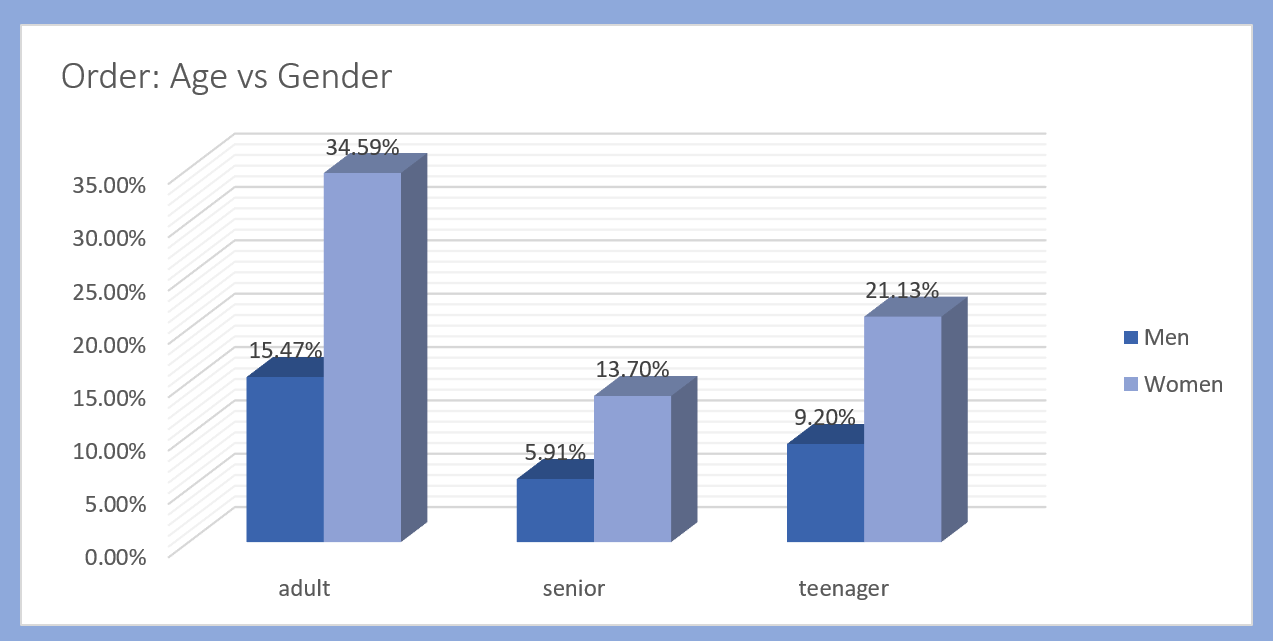
****

**A column chart is chosen to emphasize and compare sales performance across the top 5 states.**

**Columns make it easy to compare the magnitude of sales in different states, and the top 5 selection provides focused insights into regional contributions.**

* **Clearly displays the top 5 states contributing to sales.**
* **Allows for easy comparison of sales performance across states.**
* **Ideal for showcasing variations in regional performance.**

**Bar Chart for Order Analysis: Age vs Gender:**

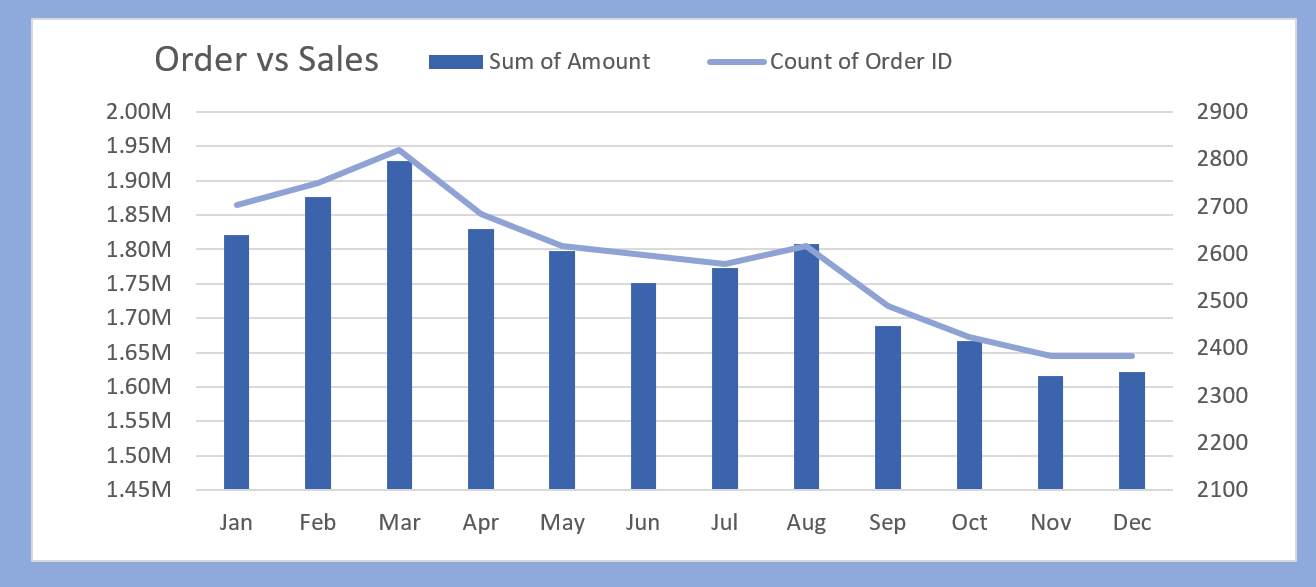
****

**A bar chart is effective for representing the relationship between age, gender, and order amounts.**

**The horizontal bars allow for easy comparison across different age groups and genders, aiding in trend analysis and providing a clear visual representation of the data.**

* **Represents the relationship between customer age, gender, and order amounts.**
* **Ideal for showing comparisons across multiple categories (age groups and genders).**
* **Supports trend analysis.**

**Order vs sales**

****

**Insights and observations**

**1. 92% Orders Delivered:**

**- The high percentage of orders being delivered (92%) indicates a robust order fulfillment process.**

**- This is a positive metric suggesting efficiency in delivering products to customers.**

**2. Women Bought Items More Than Men:**

**- The observation that women make more purchases than men suggests a significant market share and purchasing power among female customers.**

**- Tailoring marketing strategies and product offerings to cater to the preferences of female customers could be a strategic move.**

**3. Amazon Has the Highest Number of Orders:**

**- The dominance of Amazon in terms of the highest number of orders indicates the popularity and effectiveness of the platform.**

**- Focusing marketing efforts or optimizing product visibility on Amazon might be a key strategy.**

**4. Orders and Sales Increasing from January to March, Decreasing Afterwards, with a Slight Uptick in August Compared to March:**

* **The increasing trend in orders and sales from January to March signifies a positive trajectory in the initial quarter.**
* **Subsequent decreases in orders and sales from April to July indicate a decline in performance.**
* **Notably, a slight increase in August compared to March suggests a mid-year uplift, potentially due to seasonal adjustments or strategic modifications.**

**5. Top 1 State is Maharashtra:**

**- Maharashtra being the top state in terms of orders or sales highlights the importance of this region in your business.**

**- Further analysis of customer demographics, preferences, and regional marketing strategies in Maharashtra can provide valuable insights for targeted approaches.**

**Patterns Trends and Outliers**

**Patterns:**

**Seasonal Order and Sales Variation:**

**There is a pattern of increasing orders and sales from January to March, followed by a decrease. This may suggest a seasonal trend, potentially influenced by factors like holidays or promotions.**

**Trends:**

**March Spike in Orders and Sales:**

**The spike in orders and sales in March continues to be a trend, indicating a consistent exceptional performance during that month.**

**Exploring and understanding the reasons behind this March spike could unveil successful strategies applicable to other periods.**

**Mid-Year Increase (July and August):**

**The small increase in orders and sales during July and August forms a trend, deviating from the overall decreasing pattern.**

**Investigating the factors contributing to this mid-year increase could inform strategies for sustaining or replicating success.**

**Outliers:**

**Spike in March:**

**The significant spike in orders and sales in March remains an outlier, standing out from the surrounding months.**

**Identifying and leveraging the success factors during March can provide insights for targeted strategies.**

**Interactivity and Filters**

**Dashboard slicers, including Channel, Category, and Month slicers, enhance the user experience by providing dynamic filtering options. They empower users to customize their data views, fostering personalized and in-depth data exploration.**

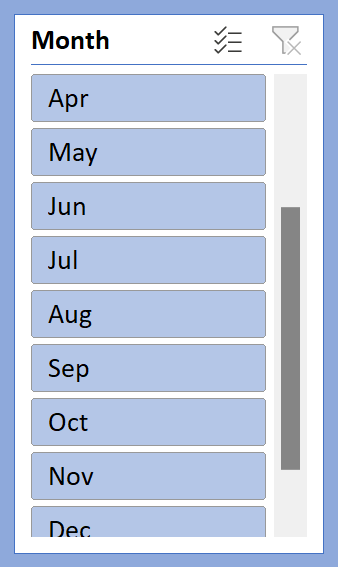
* **These slicers offer a fluid and real-time exploration experience, allowing users to instantly switch between parameters for a comprehensive understanding.**
* **The dynamic nature of slicers aids efficient decision-making, streamlining the process of gaining insights tailored to user-specific interests and requirements.**

**Channel Slicer:**

**Users can focus on specific sales channels (e.g., Jio, Amazon) for a granular analysis. This facilitates a detailed assessment of performance on different platforms, aiding strategic decision-making.**

**Category Slicer:**

**Enables users to filter by distinct product categories (e.g., kurtha, saree). This feature is valuable for narrowing analyses to specific product types, allowing tailored insights and evaluations.**

**Month Slicer:**

**Allows dynamic selection and analysis based on specific months. This flexibility supports focused exploration of sales trends, changes, or anomalies tied to each month, crucial for seasonal analysis and performance assessment.**

how these features enhance the user experience

**1. Channel Slicer:**

Enhanced Focus and Comparison: Users can selectively filter data by specific sales channels, such as Jio, Amazon, or Flipkart. This enhances focus and facilitates detailed comparisons between different platforms, allowing users to assess the individual contribution and performance of each channel.

Platform-Specific Analysis: The channel slicer enables users to zoom in on particular channels, uncovering insights into user behavior, preferences, and sales patterns specific to each platform. This depth of analysis is valuable for tailoring marketing strategies and optimizing performance on different channels.

**2. Category Slicer:**

Targeted Product Analysis: The category slicer allows users to filter data based on distinct product categories like kurtha, saree, set, top, etc. This targeted approach aids in drilling down into specific product lines, enabling users to analyze the performance of each category individually.

Product-Specific Insights: Users can gain insights into the popularity, sales trends, and customer preferences for each product category. This level of granularity is essential for making informed decisions related to inventory management, marketing strategies, and product development.

**3. Month Slicer:**

Temporal Analysis: The month slicer empowers users to dynamically select and analyze data based on specific months. This feature is particularly valuable for conducting temporal analysis, identifying seasonal trends, and understanding variations in sales and customer behavior over time.

Seasonal Patterns: Users can explore seasonal patterns, spikes, or dips in sales during different months. This facilitates the identification of season-specific opportunities or challenges, allowing for strategic planning and adjustments based on the temporal dynamics of the data.

**Overall Benefits:**

**Flexibility and Interactivity:** The combined use of these slicers offers users a highly flexible and interactive experience. Users can seamlessly switch between channels, product categories, and months, tailoring their exploration to specific areas of interest.

**Efficient Decision-Making:** The dynamic nature of slicers streamlines the decision-making process. Users can efficiently explore relevant data subsets, leading to quicker and more informed decisions related to marketing strategies, inventory management, and overall business performance.

**User-Centric Data Exploration:** These slicers put the control in the hands of the user, allowing them to explore the data based on their unique interests and requirements. This user-centric approach enhances the overall experience and ensures that insights are aligned with specific business goals and objectives.

In conclusion, the incorporation of channel, category, and month slicers not only adds granularity to the analysis but also provides users with the tools needed for a comprehensive and user-friendly exploration of the data.

**CONCLUSION**

In the realm of data exploration, the dashboard crafted from the Vrinda Store dataset, adorned with insightful visualizations, has unfurled a tapestry of revelations. Key findings from our journey through this data landscape include:

**Main Findings:**

1. **Dominance of Amazon:**

- The data spotlighted Amazon as the preeminent sales channel, signifying its pivotal role in our business landscape.

2. **Category Dynamics:**

- The nuanced trends within product categories, as revealed by the visualizations, are guiding targeted marketing efforts with precision.

**3. Temporal Patterns Unveiled:**

- The dashboard's revelations on seasonal variations across months serve as a strategic compass, influencing our marketing strategies for different times of the year.

**Impact and Reflection:**

As we reflect on this data-driven expedition, the dashboard emerges not just as a visual interface but as a strategic ally. Its power lies not only in the charts and numbers but in the stories it tells and the directions it provides.

This project marks a transformative moment—a shift toward data-driven decision-making. The dashboard is not merely a project conclusion; it's a launchpad for informed strategies, promising a future where insights gleaned from data steer our course toward continued success and adaptability.