

# School of InfoComm Technology

**Data Exploration & Analysis Assignment**

Diploma in DS

April 2024 Semester

**ASSIGNMENT 2**

(40% of DEA Module)

8th July 2024 – 11th August 2024

**Submission Deadline:**

**Presentation: Week 17 during class**

**Report and files: 11th August 2024 (Sunday), 11:59PM**

|  |  |  |
| --- | --- | --- |
| **Tutorial Group** | **:** | **T01** |
| **Student Name** | **:** | Dylon Wang Yin Chong |
| **Student Number** | **:** | S10263199B |

**Penalty for late submission:**

Marks will be deducted every calendar day after the deadline.

**NO** submission will be accepted after 18th August 2024 (Sunday), 11:59PM.

Contents

[School of InfoComm Technology 1](#_Toc174309438)

[Contents 2](#_Toc174309439)

[**Abstract and Overview** 3](#_Toc174309440)

[**3.1 dashboards to answer business questions** 4](#_Toc174309441)

[**Introduction** 4](#_Toc174309442)

[**Relationship Established** 4](#_Toc174309443)

[**1. Sales by Month** 5](#_Toc174309444)

[**2. Most Popular Products** 6](#_Toc174309445)

[**3. Sales by Location** 8](#_Toc174309446)

[**4. Customer Demographics** 9](#_Toc174309447)

[**Conclusion** 10](#_Toc174309448)

[**3.2 Further Analysis** 11](#_Toc174309449)

[**Overview** 11](#_Toc174309450)

[**Univariate Analysis** 11](#_Toc174309451)

[**Bivariate Analysis** 12](#_Toc174309452)

[**Multivariate Analysis** 14](#_Toc174309453)

[**Tables** 15](#_Toc174309454)

[**Insights and Recommendations** 15](#_Toc174309455)

[**Conclusion** 16](#_Toc174309456)

[**Summary** 17](#_Toc174309457)

# **Abstract and Overview**

This report offers a thorough look at Flipkart's sales data using Power BI dashboards. The main goal is to answer important business questions that help the company make smart and informed decisions. The dashboards created uses advanced ways to show data which gives a clearer picture of how sales are doing, which products are popular, which locations make the most money and profit, and the varying customer behaviours across different regions.

The first part of the report talks about how and why the dashboards were made to answer specific business questions. These dashboards show sales trends, point out the best-selling products, check which locations are profitable and making the most money, and study how customers shop in different places and for different types of products. What we can gather from these dashboards can help Flipkart manage its stock better, come up with better marketing plans, and run its business more efficiently.

The next part digs deeper looking at more parts of the data to find more and conduct further analysis. This section looks at what's driving sales and profits giving a close-up view of what customers like and how this changes in different areas. By using advanced ways to study the data, this analysis gives Flipkart ideas it can use to improve its place in the market and keep growing over time.

To conclude, the report ends with a recap of the main findings and thoughts on the data modelling, exploration, and analysis process. This part brings together the insights from the dashboards and further analysis stressing how data-driven choices matter to reach business goals. It also points out the hurdles faced during the analysis and gives ideas to improve data collection and analysis methods in the future.

Overall, this report shows how Power BI dashboards can turn raw data into useful insights helping Flipkart make smart choices that lead to business success. By looking at sales data, this analysis offers a plan to use data analytics to improve business operations and increase customer satisfaction.

# **3.1 dashboards to answer business questions**

## **Introduction**

This section of the report provides a detailed analysis of the four Power BI dashboards that were created to answer essential business questions for Flipkart. These dashboards offer insights into sales performance, product popularity, regional profitability, and customer demographics. The objective is to enhance Flipkart's sales strategies, optimize inventory management, and improve overall business operations by leveraging data-driven insights.

## **Relationship Established**

* **Primary Key and Foreign Key Relationship:**
  + The **Order ID** serves as a common identifier in both the 'Order Details' and 'List of Orders' tables. This relationship is critical for combining detailed order information with higher-level order attributes. In 'Order Details', it acts as a foreign key, referencing the primary key in 'List of Orders'. This relationship is depicted by the connecting line between the two tables in the diagram when in Power BI.
* **Many-to-Many Relationship:**
  + The relationship between 'List of Orders' and 'Order Details' is a many-to-many relationship. This indicates that each order in 'List of Orders' can have multiple corresponding records in 'Order Details', and similarly, each entry in 'Order Details' can relate to multiple orders in 'List of Orders'.

## **1. Sales by Month**

**Business Question**

**Which period (month) has the most sales?**

**A screenshot of a computer

Description automatically generated**

Analysis and Insights

The first dashboard analyses monthly sales to determine the month with the highest sales volume. Key visualizations included in this dashboard are:

* Bar Chart (Sum of Sales by Month): This chart displays the total sales amount for each month, clearly showing January as the month with the highest sales at 61,439 units sold. This peak indicates a significant increase in consumer purchasing behaviour during this month, which could be due to post-holiday sales or marketing campaigns.
* Line Chart (Sum of Sales by Year, Quarter, and Month): This line chart helps understand the sales trend over time, highlighting significant dips in April 2018 and July 2018. Along with these dips, there were also peaks in January 2019 and March 2019. These dips and peaks suggest that specific months have higher sales due to seasonal promotions, holidays, or other factors influencing consumer spending.
* Waterfall Chart (Sum of Profit by Month): This chart highlights the monthly profit trends, showing December with the highest profit, followed by March and January. This pattern indicates that the end-of-year holiday season is not only a high sales period but also highly profitable, possibly due to holiday shopping and end-of-year clearances.
* Scatter Plot (Category, Sale vs Profit): This scatter plot highlights the relationship between sales amount and profit across different product categories. The plot does reveal that if the sales amount increases, the profit generally increases as well. Although there is variability among the categories, we can see that Electronics (in purple) and Clothing (in red) show a tighter cluster with smaller profit margins, while Furniture (in blue) exhibits a wider range in both sales and profit. This will help Flipkart in identifying the categories that contribute the most to profit and can dial down on areas where pricing or discount strategies could be optimized to improve profitability.
* Tree Map (Sum of Sales by Month): The tree map further emphasizes the sales distribution across months, with larger segments representing higher sales months like January, March and November. This visual provides a quick and intuitive understanding of the sales distribution over the year.

Rationale

Understanding sales seasonality is crucial for optimizing inventory management, staffing, and marketing campaign timings. By identifying peak sales periods, Flipkart can ensure sufficient stock levels, adequate staffing, and targeted promotions to maximize revenue. Additionally, recognizing months with lower sales can help in planning promotions and discounts to boost sales during those periods.

## **2. Most Popular Products**

**Business Question**

**What are the top 5 most popular products (by Category and Sub-category) across different sales channels?**

**A screenshot of a computer

Description automatically generated**

Analysis and Insights

The second dashboard focuses on identifying the most popular products by category and sub-category:

* Donut Chart (Sum of Sales by Category): This chart shows the sales distribution across different product categories, with Electronics (38.3%) and Clothing (32.23%) being the top categories by sales amount. This indicates a significant consumer interest in these categories.
* Funnel Chart (Top 5 Products by Profit): This chart lists the top 5 products by profit, with Printers leading the list, followed by Bookcases and Accessories. This insight helps in identifying high-margin products that contribute significantly to the overall profitability.
* Line Charts (Sum of Sales by Month and Category): These charts display monthly sales trends for Clothing, Electronics, and Furniture. Electronics show the highest variability, indicating seasonality or the impact of promotions on sales within this category.
* Bar Chart (Sum of Sales by Sub-Category): This bar chart identifies popular sub-categories, with Printers and Bookcases topping the list. Along with the added data colours to not only display the sales but also gauge the profit as well. This suggests that office and home essentials are in high demand, guiding inventory and promotional strategies.
* Decomposition Tree (Quantity of Products Sold across Category and Sub-Category): The decomposition tree effectively breaks down the sales data to reveal which categories and sub-categories are driving the most sales in terms of quantity.

Clothing: With the highest quantity sold, strategies can focus on increasing the variety and stock of high-demand items like Saree and Handkerchief.

Electronics: Focus on promoting and stocking high-value items like Printers and Phones to boost overall sales revenue.

Furniture: Investigate the reasons for lower sales mainly in Tables to try and develop strategies to improve their demand.

Rationale

Identifying key products helps target marketing activities and engage customers effectively. By knowing which products are most popular, Flipkart can focus on promoting these items, managing inventory, and negotiating better terms with suppliers. This approach ensures that high-demand products are always available, leading to increased customer satisfaction and loyalty.

## **3. Sales by Location**

**Business Question**

**What are the top and bottom 5 locations in terms of profit?**

**A screenshot of a computer

Description automatically generated**

Analysis and Insights

The third dashboard evaluates regional performance by profit:

* Funnel Chart (Top 5 States by Profit): Maharashtra and Madhya Pradesh are the top states in terms of profit, indicating high profitability in these regions. This could be due to a larger amount of customers, higher consumer spending, effective marketing campaigns, or better logistics.
* Funnel Chart (Bottom 5 States by Profit): Tamil Nadu and Punjab are among the lowest, highlighting areas needing improvement or targeted interventions. Understanding why these states are underperforming can help in devising strategies to boost sales and profitability.
* Map Visualization (Profit by State): This interactive map provides a geographical representation of profitability, making it easier to identify regional trends. States with darker shades of green represent higher profits, visually distinguishing profitable regions from less profitable ones.
* Stacked Column Chart (Cities by Profit): Pune and Indore are the top cities by profit, while Chennai shows the lowest performance boasting a negative amount of profit gained. This insight helps in focusing efforts on underperforming cities to improve overall regional performance.
* Line Chart (Sum of Sales and Sum of Profit by States): There is a visible correlation between the sum of sales amount and sum of profit for most states, with high sales generally translating to high profits. States like Maharashtra and Madhya Pradesh, which have high sales, also show high profits. States with lower sales volumes, like Tamil Nadu and Bihar, tend to have lower or negative profits, indicating either inefficiency or market challenges in these regions.

Rationale

Evaluating branch performance helps make informed decisions, optimize resources, and enhance overall profitability. Understanding regional performance allows Flipkart to allocate resources effectively, target underperforming areas, and replicate successful strategies in other regions. This regional focus ensures that all areas are performing optimally, contributing to the company's overall growth.

## **4. Customer Demographics**

**Business Question**

**How does customer behaviour vary across different regions?**

**A screenshot of a computer

Description automatically generated**

Analysis and Insights

The fourth dashboard provides insights into customer behaviour across regions and product categories:

* Bar Chart (Top 10 Customers by Sales): This chart lists the top 10 customers by sales amount, with Yaani and Pooja being the highest spenders. This insight helps in identifying loyal customers who contribute significantly to sales.
* Map Visualization (Number of Customers by State): This map shows the distribution of customers across different states, with Madhya Pradesh and Maharashtra having the highest customer counts. Understanding customer distribution helps in tailoring marketing strategies to different regions.
* Bar Chart (Number of Customers and Sum of Profit by City): Indore and Mumbai are the cities with the highest customer counts and profits, indicating strong consumer bases in these areas.
* Decomposition Tree (Quantity of Products across Category and Sub-Category by State and City): This visualization shows the flow of product quantities from states to cities, highlighting regional product preferences. It provides a comprehensive view of how products move through different regions.
* Line Chart (Number of Orders by Quarter): This chart shows the order count trends by quarter, with Q1 having the highest order count and Q2-Q3 having the lowest order counts leading to losses in profit. This quarterly view helps in understanding seasonal trends and planning accordingly.

Rationale

Understanding regional customer preferences and behaviour across product categories helps customize the shopping experience and maximize sales. By analyzing customer demographics, Flipkart can tailor their marketing strategies, improve product offerings, and enhance customer satisfaction in different regions. This approach ensures that the company meets the diverse needs of its customer base, leading to higher retention and loyalty.

## **Conclusion**

These Power BI dashboards provide valuable insights into Flipkart's sales performance, product popularity, regional profitability, and customer behavior. By leveraging these insights, Flipkart can optimize its business strategies, improve operational efficiency, and drive overall growth. The dashboards highlight critical KPIs, enabling data-driven decision-making and enhancing the company's competitive edge in the market.

Each dashboard is designed to answer specific business questions, helping Flipkart to understand and respond to various market dynamics effectively. By continuously monitoring and analyzing these metrics, Flipkart can stay ahead of market trends, adapt to changing consumer preferences, and ensure sustained business success.

# **3.2 Further Analysis**

## **Overview**

In this section, we will delve deeper into the data using advanced analytics techniques. The aim is to identify critical trends and patterns that can provide Flipkart with valuable insights to enhance its business performance. We will be utilizing univariate, bivariate, and multivariate analysis, along with tables and data mining techniques. The analysis is supported by various visualizations created in Power BI.

## **Univariate Analysis**

A screenshot of a graph

Description automatically generated

* **Number of Orders by Sub-Category:**

The bar chart in the top-left section displays the count of orders for each sub-category. The top three sub-categories with the highest order counts are:

* Saree (210 orders)
* Handkerchief (198 orders)
* Stole (192 orders)

These categories indicate a strong demand for clothing items, suggesting that this category is a significant revenue driver for Flipkart.

* **Number of Orders by State:**

The bar chart in the top-right section shows the distribution of orders across different states. The top three states with the highest order counts are:

* Madhya Pradesh (101 orders)
* Maharashtra (90 orders)
* Rajasthan (60 orders)

This distribution highlights key geographical areas where Flipkart has a strong customer base. It can guide targeted marketing efforts and logistical planning.

* **Orders by Category:**

The donut chart in the bottom-left section illustrates the proportion of orders by category. Clothing is the most ordered category, followed by Electronics and Furniture. Specifically:

* Clothing: 63.27% (949)
* Electronics: 20.53% (308)
* Furniture: 16.2% (243)

This reinforces the findings from the sub-category analysis and underscores the importance of the Clothing category to Flipkart’s overall sales.

* **Number of Orders by Year, Quarter and Month:**

The line chart in the bottom-right section tracks the count of orders over time. It reveals a noticeable peak in January 2019 (61 orders) and another peak in March 2019 (58 orders). This seasonality trend can be crucial for planning inventory and marketing campaigns around peak sales periods.

## **Bivariate Analysis**

A screenshot of a graph

Description automatically generated

* **Quantity Sold by Profit:**

The scatter plot in the top-left section shows the relationship between the quantity sold and profit. Most data points cluster around zero profit, with some significant outliers showing both high profit and loss. This dispersion indicates variability in profitability across different products and quantities, suggesting the need for more nuanced pricing and inventory strategies.

* **Sum of Sales by Month:**

The bar chart in the top-right section shows the total sales amount for each month. The highest sales are recorded in:

* January (61K)
* February (59K)
* November (48K)

These monthly sales trend further supports the seasonal analysis, indicating higher sales during the winter months. This insight can help Flipkart optimize inventory and marketing strategies to capitalize on these peak periods.

* **Sales and Quantity Sold:**

The scatter plot in the bottom-left section displays the relationship between sales amount and quantity sold. There is a slight positive correlation, as expected, with higher quantities corresponding to higher sales amounts. This visualization helps in understanding the sales dynamics and can inform pricing strategies.

* **Sales and Profit:**

The scatter plot in the bottom-right section illustrates the relationship between sales amount and profit. It shows a positive correlation, indicating that higher sales amounts generally lead to higher profits. However, there are significant outliers with high sales amounts but low or negative profits, highlighting areas where cost management could be improved.

## **Multivariate Analysis**

A screenshot of a computer

Description automatically generated

* **Sum of Sales by Category and Sub-Category:**

The tree map in the top-left section breaks down the total sales amount by category and sub-category. Key insights include:

* Printers (58K) and Phones (46K) are top contributors within the Electronics category.
* Saree (54K) and Trousers (30K) are top contributors within the Clothing category.
* Bookcases (57K) and Chairs (34K) are significant within the Furniture category.

This detailed breakdown helps identify which products drive the most revenue within each category, enabling targeted marketing and inventory optimization.

* **Sum of Sales, Sum of Profit, and Sum of Quantity Sold by City:**

The clustered bar chart in the top-right section compares sales amount, profit, and quantity across different cities. Key findings include:

* Indore leads in total sales amount (79K) and while being second in profit gained (4,159). Indicating that Indore not only has high sales but high profit as well.
* Mumbai and Pune follow with substantial sales amounts (62K and 33K respectively). Along with that, Pune can be seen to be the most profitable city with the highest profit gained (4,539).

This geographical analysis can guide strategic decisions regarding resource allocation and regional marketing campaigns.

* **Sum of Quantity Sold by Sales and Profit:**

The bubble chart in the bottom section visualizes the relationship between quantity sold, sales amount, and profit. The size of the bubbles represents the quantities sold, with larger bubbles indicating more quantity sold. This chart helps identify products with high sales quantities that also yield significant profits, aiding in product prioritization.

## **Tables**

Tables provides a flexible way to summarize and analyse data. By pivoting on different dimensions such as category, sub-category, and geography, deeper insights and trends can be uncovered.

A screenshot of a computer screen

Description automatically generated

* **Sales and Profit by State and Category:**

The table provides a comprehensive overview of sales and profit by state and category. Key insights include high sales in states like Madhya Pradesh, Maharashtra, and Gujarat, with substantial profits from Electronics and Clothing categories. Some states like Nagaland and Punjab show losses in Electronics, indicating areas for cost management improvement.

## **Insights and Recommendations**

1. **Seasonality and Inventory Management:**
   * The analysis shows clear seasonal peaks in sales, particularly in January, February, and December. Flipkart should ensure sufficient inventory during these peak periods to meet demand and avoid stockouts.
2. **Product Focus:**
   * Clothing, especially Sarees and Handkerchiefs, are top-selling categories. Marketing efforts should be intensified for these products. Additionally, high-selling electronics like Printers and Phones should also receive targeted promotions.
3. **Geographical Strategy:**
   * States like Madhya Pradesh, Maharashtra, and cities like Indore and Mumbai are key markets. Flipkart should consider region-specific promotions and logistics enhancements to better serve these areas.
4. **Profitability Improvement:**
   * The scatter plots reveal significant variability in profitability. Flipkart should analyze cost structures and pricing strategies to improve margins, particularly for high-sales but low-profit products.
5. **Customer Segmentation:**
   * Implementing clustering techniques can help in understanding customer segments better, allowing for personalized marketing strategies that can enhance customer engagement and sales.

## **Conclusion**

The further analysis done provides valuable insights into Flipkart’s sales performance. By leveraging these insights, Flipkart can optimize inventory management, enhance marketing strategies, improve profitability, and meet customer needs. The next steps that Flipkart should take involve implementing these recommendations and continuously monitoring the results to adapt and refine strategies in response to evolving market conditions.

# **Summary**

The analysis of Flipkart's sales data through the creation of Power BI dashboards and advanced analytics techniques has provided several key insights that can significantly enhance the company's business performance.

1. **Seasonal Sales Trends:**
   * The data reveals clear seasonal peaks in sales, particularly in January, February, and December. These peaks suggest that Flipkart experiences higher demand during the winter months, possibly due to holiday shopping and seasonal promotions.
   * Recommendation: Flipkart should ensure sufficient inventory during these peak periods to meet demand and avoid stockouts. Marketing campaigns should also be aligned with these seasonal trends to maximize sales.
2. **Product Performance:**
   * The Clothing category, particularly sub-categories like Sarees and Handkerchiefs, emerges as the most ordered category, contributing significantly to the overall sales. Electronics, especially Printers and Phones, also show high sales and profit figures.
   * Recommendation: Intensify marketing efforts for high-selling products in these categories. Targeted promotions and inventory optimization for these products can further boost sales and profitability.
3. **Geographical Distribution:**
   * States like Madhya Pradesh, Maharashtra, and cities like Indore and Mumbai are identified as key markets with high order counts and substantial sales amounts.
   * Recommendation: Consider region-specific promotions and logistics enhancements to better serve these areas. Tailored marketing strategies for these key markets can increase customer engagement and sales.
4. **Profitability Variability:**
   * The scatter plots reveal significant variability in profitability across different products and quantities. There are products with high sales amounts but low or negative profits, indicating potential issues in cost management and pricing strategies.
   * Recommendation: Analyse cost structures and pricing strategies to improve margins. Focus on products with high sales quantities that also yield significant profits for better profitability management.
5. **Customer Segmentation:**
   * Implementing clustering techniques can help in understanding customer segments better. This can lead to personalized marketing strategies, enhancing customer engagement and increasing sales.
   * Recommendation: Use clustering analysis to identify customer segments with similar purchasing behaviour. Develop targeted marketing campaigns for these segments to improve customer satisfaction and loyalty.

**Reflections on Data Modelling, Data Exploration, and Analysis Process**

The process of data modelling, exploration, and analysis was instrumental in uncovering these valuable insights. Here are some reflections on each stage of the process:

1. **Data Modelling:**
   * Data modelling involved structuring the sales data in a way that facilitated easy analysis and visualization. This step was crucial for creating relationships between different data entities and ensuring data integrity.
   * The use of Power BI for data modelling proved to be highly effective. Power BI's capabilities in handling large datasets and creating complex relationships between tables enabled a thorough analysis of Flipkart's sales data.
   * Future Improvement: Continuous refinement of data models to include additional variables and dimensions can provide deeper insights. Incorporating external data sources, such as market trends and competitor analysis, can also enhance the data model.
2. **Data Exploration:**
   * Data exploration involved examining the dataset to understand its structure, quality, and key characteristics. This step helped in identifying missing values, outliers, and data patterns that needed further investigation.
   * Power BI's interactive visualization tools made it easier to explore the data and identify trends and patterns quickly. The ability to drill down into specific data points and filter data dynamically added significant value to the exploration process.
   * Future Improvement: Implementing automated data cleaning and preprocessing techniques can streamline the exploration process. Using advanced data exploration tools, such as machine learning algorithms for anomaly detection, can also enhance the quality of insights.
3. **Analysis Process:**
   * The analysis process involved both univariate and multivariate techniques to uncover critical trends and patterns. Visualizations such as bar charts, scatter plots, tree maps, and bubble charts played a key role in presenting the data in an understandable and actionable format.
   * Advanced analytics techniques, including clustering and time series analysis, provided deeper insights into customer behaviour and sales trends. These techniques enabled the identification of high-value customer segments and seasonal sales patterns.
   * Future Improvement: Integrating more sophisticated analytics tools, such as predictive modelling and machine learning, can further enhance the analysis process. These tools can provide predictive insights and help in making data-driven decisions more effectively.

**Conclusion**

The report ends with a recap of the main findings and thoughts on the data modelling, exploration, and analysis process. The report brings together the insights from the dashboards and further analysis stressing how data-driven choices matter to reach business goals. It also points out the hurdles faced during the analysis and gives ideas to improve data collection and analysis methods in the future.

The thorough review of Flipkart's sales and profitability figures has also produced useful insights to boost business results. These insights can help Flipkart to manage stock better, improve marketing plans, increase profits, and meet customer needs more often to further increase customer satisfaction and loyalty. Building models, exploring data, and doing analysis have shown their worth in making sense of the information and finding key business takeaways. To keep getting better results, Flipkart should keep refining these methods and add new analysis tools.

Overall, this study shows why using data to make choices matters so much in today's tough business world. How well Flipkart uses these insights will play a big role in keeping its place in the market and growing over time.