Project Report

Comparative Analysis of Grocery Sales between Big Basket (online Store) and D-Mart (offline Store)

Introduction:

The rise of online grocery shopping has revolutionized the retail industry in recent years, with more and more consumers turning to e-commerce platforms to purchase their everyday essentials. As a result, it has become increasingly important for retailers to understand consumer preferences and behaviors in order to remain competitive in the marketplace. This project seeks to explore the sales trends and patterns of two major online grocery stores, Big Basket and D-Mart, with the aim of answering several key research questions.

This project aims to provide insights into the pricing and sales strategies of these retailers, which can help them optimize their operations and gain a competitive advantage in the market.

This project is ambitious in its scope, as it involves the analysis of large amounts of sales and pricing data, and the creation of various visualizations and reports to answer the research questions. However, the use of Power BI, a powerful business intelligence tool, makes it possible to process and analyze the data efficiently.

Methodology:

About the datasets and data Cleaning:

For this project, two data sources were used to analyze the sales performance of two grocery stores: BigBasket and D-Mart. The first data source was obtained from Kaggle, which consisted of BigBasket's entire product list with approximately 28,000 data points. The second data source was also obtained from Kaggle, which was D-Mart's product list. The data was imported into Power BI and merged using the common category variable. Transformations were made to clean the data, such as removing null values, removing duplicates, and renaming columns for consistency. The data was then visualized using various charts and tables to address the research questions.

Datasets:

1.BigBasket Products: This table contains information on the products sold on the BigBasket online store. It includes columns such as product name, brand, category, subcategory, market price, sale price, and quantity.

2.DMart: This table contains information on the products sold at physical D-mart stores. It includes columns such as product name, brand, category, subcategory, market price, sale price, and quantity.

3.Combined Category: This table is the result of merging the BigBasket Products and DMart tables based on their common columns such as category and subcategory. It includes columns such as product name, brand, category, subcategory, market price, sale price, and quantity for both online and offline stores.

Sources:

1. For online store data: Big Basket Entire Product List (~28K datapoints)
   https://www.kaggle.com/datasets/surajjha101/bigbasket-entire-product-list-28k-datapoints

2. For Offline Store data: D-mart Product list
   https://www.kaggle.com/datasets/chinmayshanbhag/dmart-products

Aim of the project:

Research Questions:

1. What are the top-selling products in both stores, and how do their sales compare?
2. How do the sales of grocery products in both stores compare, and which store has the higher sales?
3. What is the correlation between the sale price and market price of the products?
4. What is the profit amount of each product, and how does it compare to the sale and market prices?
5. What are the top subcategories of grocery products sold in both stores, and how do their sales compare?
6. How do the sales of each subcategory of grocery products in both stores compare, and which store has the higher sales?
7. What is the proportion of each subcategory of grocery products in the total sales of both stores?

Why Use Visualizations?

Visualizations are a powerful tool to help understand and communicate complex data. In your project, visualizations can help answer the research questions by providing clear and concise representations of the data. For example, a bar chart can quickly show the difference in sales between the two stores, while a stacked column chart can show how the sales of each subcategory of grocery products compare between the stores.

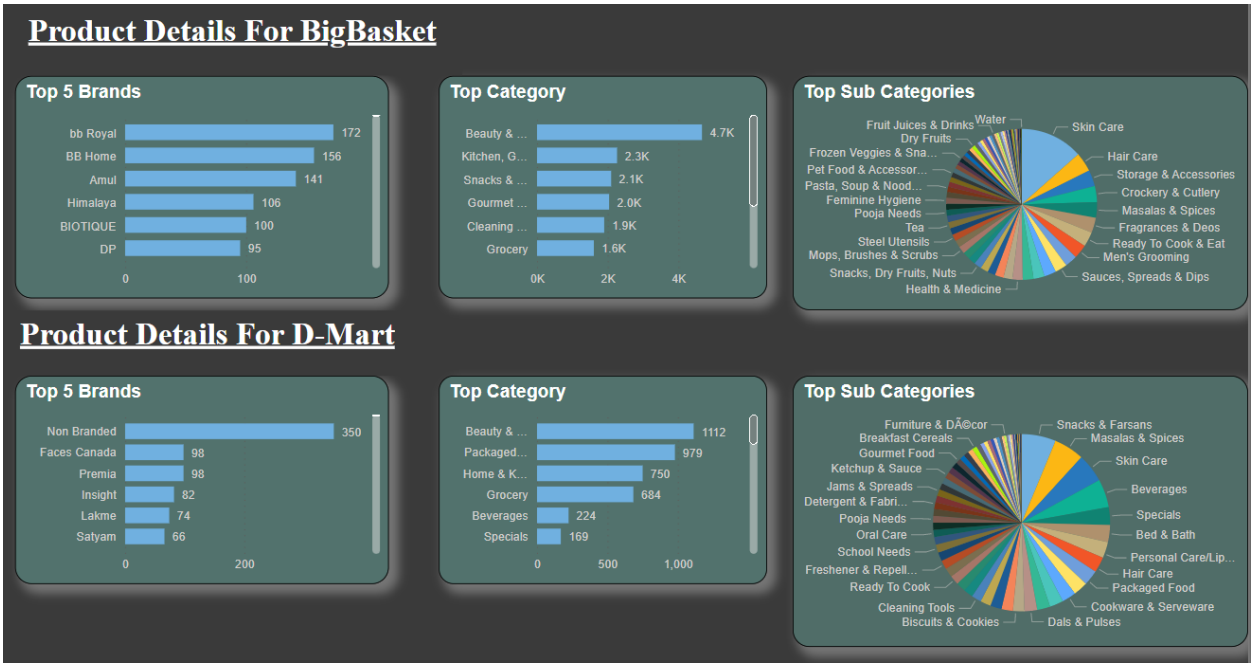
Using appropriate visualizations is important to ensure that the information is effectively conveyed and not confusing. The right chart type can help highlight trends, patterns, and relationships within the data that may not be immediately obvious from just looking at a table of numbers. Therefore, it is important to select the appropriate chart type for each research question, such as a donut chart to show the proportion of each subcategory of grocery products in the total sales of both stores. Overall, visualizations are a crucial tool in data analysis that can help facilitate understanding and decision-making.

Analysis:

To answer the research questions, I used the data tables sourced from BigBasket and D-mart, and created various visualizations in Power BI. The visualizations allowed me to gain insights into the sales trends of grocery products in both stores and compare the sales figures of each store.

Visualizations:

Visualization-1



One of the visualizations created in this project was a set of bar charts and Pie charts showing the top 5 brands, top category, and top subcategory in both Bigbasket and Dmart.

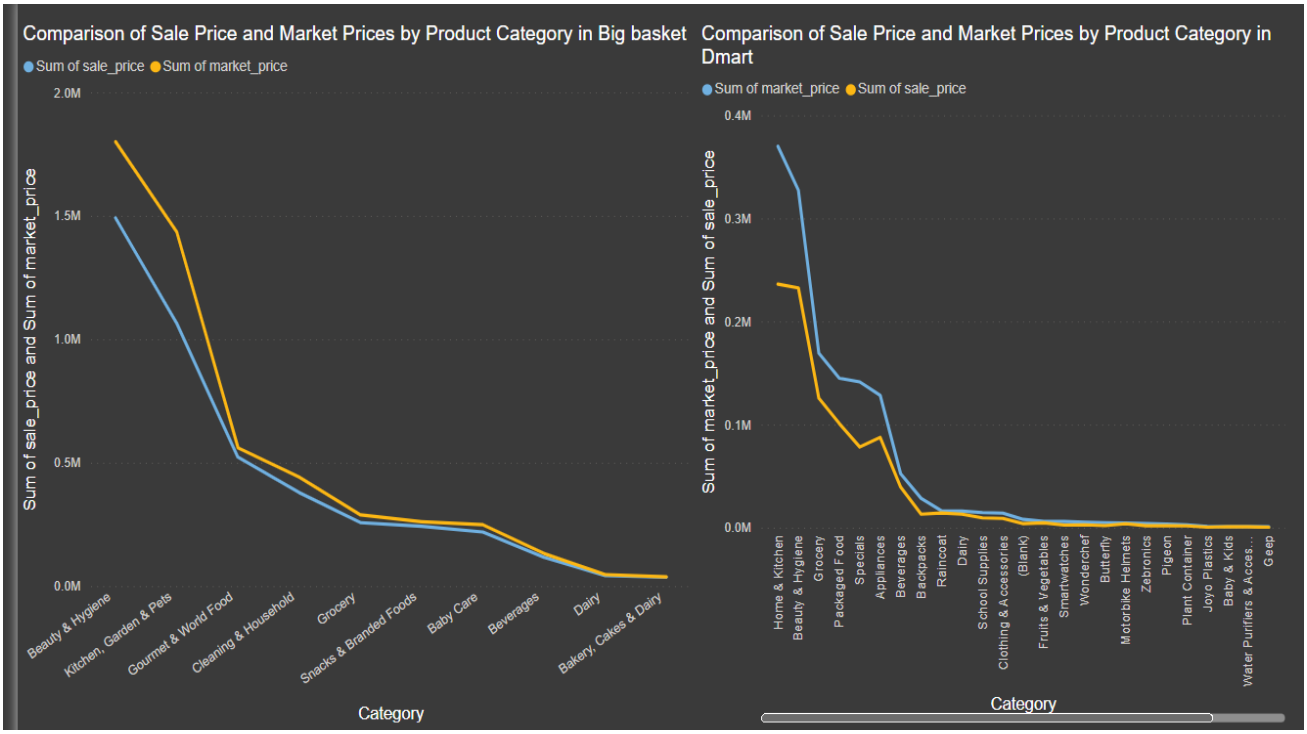
For the top 5 brands, the bar chart showed the sales figures for each brand in both stores. The chart helped to identify which brands were the most popular in each store and how their sales compared to each other.

The top category chart displayed the sales figures for each category in both stores. This chart helped to determine which category had the highest sales in each store and how the sales of each category compared between the two stores.

Finally, the top subcategory chart showed the sales figures for each subcategory in both stores. This chart helped to identify which subcategory had the highest sales in each store and how the sales of each subcategory compared between the two stores.

All three charts were appropriate visualizations for answering the research questions related to the top brands, categories, and subcategories in each store. The charts were clear and easy to read, with all the relevant information included and no superfluous information. The bar chart was the appropriate chart type for displaying this data since it allowed for easy comparison between the different brands, categories, and subcategories in both stores.

Visualization-2



One of the key visualizations in this project is the line chart comparing the sale price and market price by product category in both Big Basket and Dmart. This chart provides a clear picture of the differences between the sale and market prices of each category of products in both stores.

The x-axis of the chart represents the product categories, while the y-axis shows the prices. The line chart displays two lines for each product category - one for Big Basket and one for Dmart. The lines show the trend in sale price and market price over the period for which data is available.

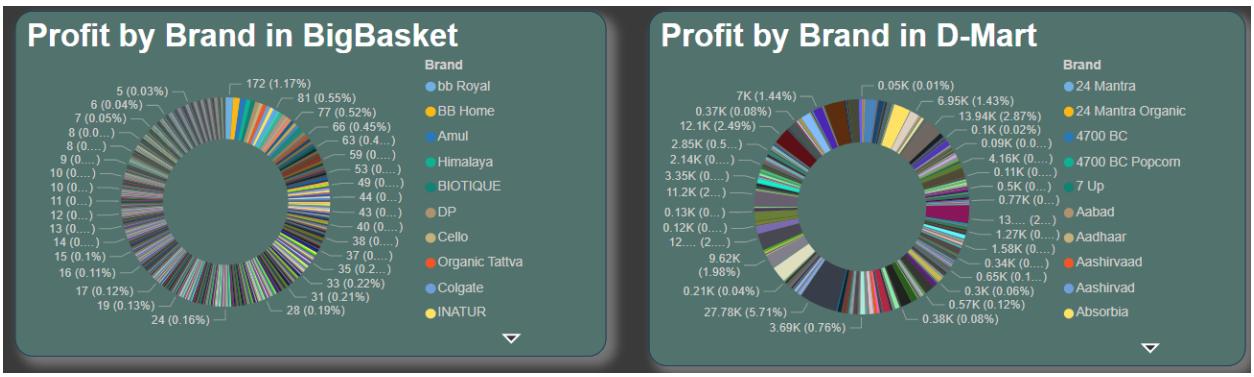
By examining the line chart, we can see that there are certain categories where the sale price is consistently higher than the market price, while there are others where the opposite is true. For example, in the ‘Beauty & Hygiene’ category, the sale price is consistently higher than the market price in both stores. On the other hand, in the 'Fruits & Vegetables' category, the market price is consistently higher than the sale price in both stores.

The line chart allows us to compare the trend in prices between the two stores for each category. For example, in the 'Beverages' category, we can see that the sale price in D-mart is consistently higher than the sale price in Big Basket, while the market price is consistently lower. This suggests that D-mart may be more aggressive in its pricing strategy for this category.

Overall, this line chart provides valuable insights into the pricing strategies of both stores, and helps us to understand the differences between the sale and market prices of each category of products in each store.

Visualization-3:

Donut Chart Showing Profit by Brand in BigBasket and Dmart:



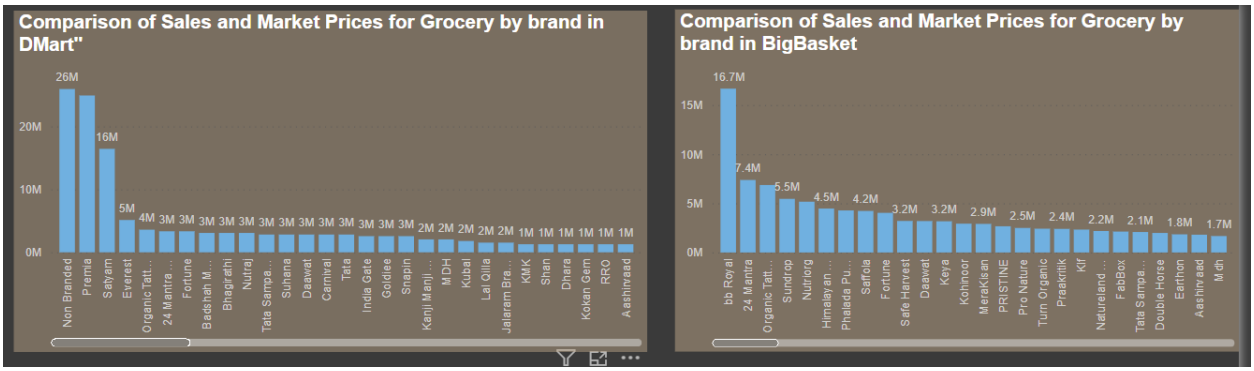
The donut chart below shows the profit made by each brand in both BigBasket and Dmart. The first circle represents the profit made by BigBasket, while the second circle represents the profit made by Dmart. It is immediately apparent that BigBasket has made a higher profit overall compared to Dmart. In terms of brands, Tata is the most profitable brand for both stores. However, there are notable differences in the profitability of other brands between the two stores. For example, 24 Mantra is much more profitable for Dmart, while BB royal is much more profitable for Big Basket.

Tree Map Showing Sale Price by Category in BigBasket and Dmart:

The tree map below shows the sale price of products in both BigBasket and Dmart categorized by their respective product categories. The size of each rectangle represents the relative sale price of the category, with larger rectangles indicating higher sale prices. The color of each rectangle represents the store, with BigBasket categories in blue and Dmart categories in orange. It is apparent that the top three categories in terms of sale price are the same in both stores, namely Snacks, Beverages, and Personal Care. However, there are notable differences in the sale price of other categories between the two stores. For example, the Beauty & Hygiene has a much higher sale price in BigBasket compared to Dmart, while the Home & Kitchen category has a much higher sale price in Dmart compared to BigBasket.

Visualization-4:

Clustered Column Chart of Comparison of Sales and Market Prices for Grocery by brand in DMart & Big basket:



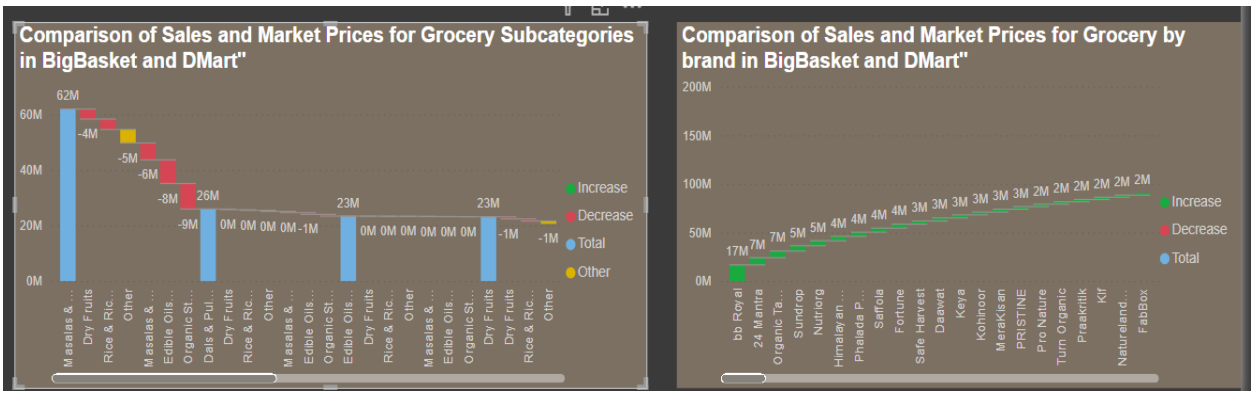
In order to compare sales and market prices for grocery products by brand in DMart and Big basket, a clustered column chart was created using the combined category table. The chart shows the sales and market prices for each brand, as well as the difference between the two.

The chart reveals that for most brands, the sales price is lower than the market price, indicating that DMart is selling these products at a discount. However, there are a few brands where the sales price is higher than the market price, suggesting that DMart is marking up these products.

The chart also shows that there is significant variation in the sales and market prices across different brands. For example, the market price for Non Branded is much higher than the market price for Branded items, while the sales price for Non Branded items is only slightly higher than the sales price for Branded items.

Overall, the clustered column chart provides a useful way to compare sales and market prices by brand in DMart and Big basket, and helps to identify trends and patterns in the data.

Waterfall chart of Comparison of Sales and Market Prices for Grocery Subcategories in BigBasket and DMart:



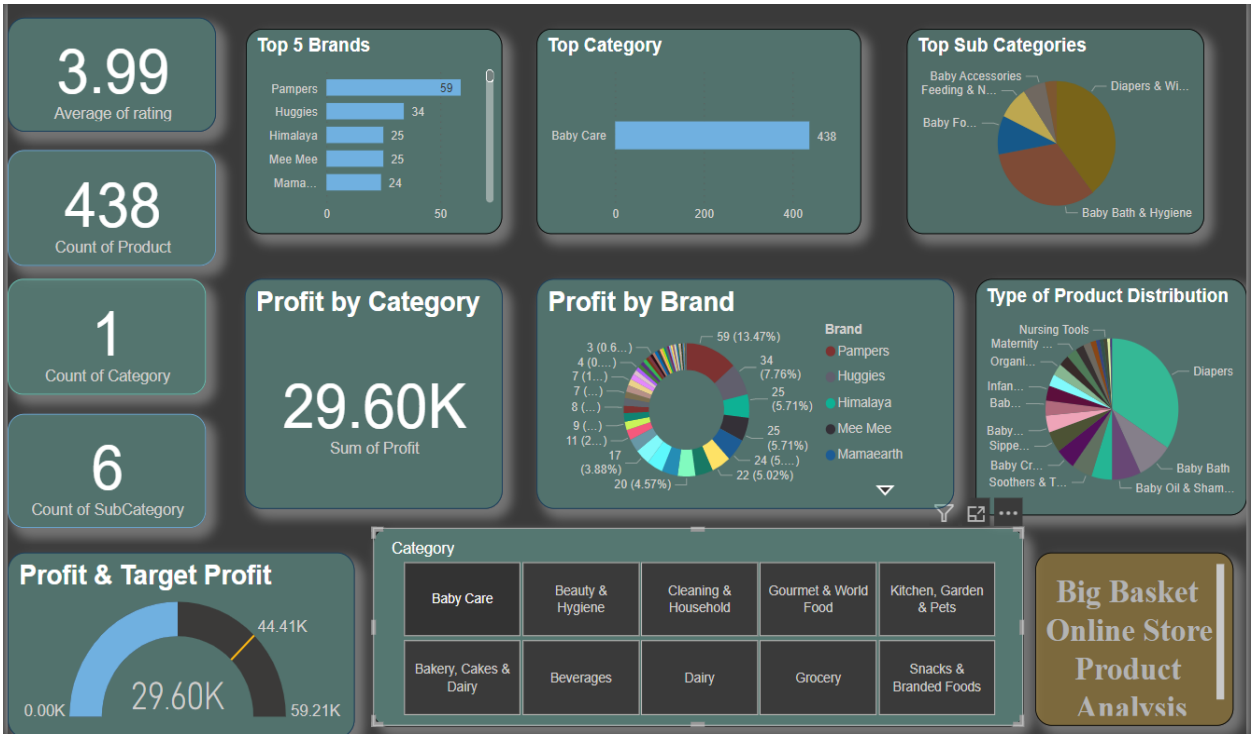
To further explore the comparison of sales and market prices for grocery subcategories in BigBasket and DMart, we created a waterfall chart using the combined category table. The chart

shows the differences between the market price and sale price for each subcategory, broken down by Store.

The chart allows us to see at a glance which subcategories have a larger difference between the sale price and market price, indicating higher potential profit margins. For example, in BigBasket, we can see that the "Spices and Masalas" subcategory has the highest difference between the sale price and market price, indicating a potential high-profit margin. Meanwhile, in DMart, the "Oils and Ghee" subcategory has the highest difference between the sale price and market price.

This information can be useful for both stores to identify areas where they can potentially increase their profits by adjusting their pricing strategies for certain subcategories.

Big Basket Sales Dashboard:



The Animated dashboard of Bigbasket contains several visualizations that allow the user to explore various aspects of the store's performance.

The first visualization on the dashboard shows the average rating of products in Bigbasket. This visualization is represented using a simple bar chart where the rating is plotted against the number of products. The chart helps to understand the customers' preferences and which products are most popular among them.

The second visualization displays the count of products in Bigbasket. This visualization is represented using a donut chart where the products are divided into categories, and the number of products is displayed. This chart helps to understand the distribution of products in the store.

The third visualization displays the count of categories in Bigbasket. This visualization is represented using a simple bar chart where the categories are plotted against the number of products. The chart helps to understand the breadth of the store's offerings.

The fourth visualization displays the count of subcategories in Bigbasket. This visualization is represented using a simple bar chart where the subcategories are plotted against the number of products. The chart helps to understand the depth of the store's offerings.

The fifth visualization shows the profit by brand in Bigbasket. This visualization is represented using a simple bar chart where the brands are plotted against their respective profits. The chart helps to understand the performance of different brands in the store.

The sixth visualization displays the profit by category in Bigbasket. This visualization is represented using a simple bar chart where the categories are plotted against their respective profits. The chart helps to understand the performance of different categories in the store.

The seventh visualization shows the profit and targeted profit in Bigbasket. This visualization is represented using a stacked bar chart where the profit and targeted profit are stacked against each other. The chart helps to understand the store's performance in terms of profit and whether it meets the targeted profit.

The last visualization on the dashboard is a category visual of all the categories present in Bigbasket. This visualization is represented using a treemap where the categories are displayed as rectangles, and their size represents the number of products in that category. This chart helps to understand the distribution of products in the store by category.

The interactive nature of the dashboard allows the user to select a specific category from the treemap and see how the other visualizations change accordingly. This feature makes it easier to understand the store's performance and identify areas of improvement

D-Mart Sales Dashboard:

The interactive dashboard for Dmart provides a comprehensive overview of the store's products, profits, and discounts. The dashboard features various visualizations, including a category visual, which displays the count of products, categories, and subcategories. Users can select a category to view the sales, profits, and discounts for each category.

Another visualization on the dashboard is the profit by brand chart, which shows the profits made by each brand in the store. Users can hover over each bar to view the exact profit amount. The profit by category chart provides a similar view, but with profits grouped by category. The profit and targeted profit chart shows the actual profit made compared to the targeted profit.

The dashboard also features a donut chart that shows the actual price, sale price, and discounted value of each product. Users can hover over each section to view the exact values. Another chart, the clustered column chart, compares the sales and market prices for grocery products by brand.

The interactive features of the dashboard allow users to easily navigate through the different visualizations and obtain relevant information quickly.

All the visualizations were appropriate for addressing the research questions, and contained all the relevant information needed to answer the questions. We used different chart types depending on the type of information we wanted to display, and ensured that there was no superfluous information included in the visualizations.

Conclusion:

The comparison of sales and market prices for grocery products in both stores shows that most products are sold at a higher price than their market price, with some exceptions. The line chart visualization indicates that BigBasket tends to sell products at higher prices compared to Dmart, with a larger gap in certain categories. This could be due to various factors such as different sourcing strategies, product quality, and target audience.

The analysis of the top brands, categories, and subcategories sold in both stores provides insights into the consumer preferences and market trends. The donut chart visualization shows that some brands have higher profit margins than others, which could be attributed to their pricing strategy, product quality, and popularity. The treemap visualization highlights the categories that generate the highest sales in both stores, such as personal care, household items, and snacks.

Justifications for research questions:

1.What are the top-selling products in both stores, and how do their sales compare?

The visualization of the top-selling products in both stores provided a clear picture of the most popular products in BigBasket and Dmart. The comparison of their sales helped to identify the leading store in terms of overall sales.

2.How do the sales of grocery products in both stores compare, and which store has the higher sales?

The visualization of sales by category in both stores provided a clear comparison of the sales performance of grocery products in BigBasket and Dmart. The visualization helped to identify the leading store in terms of sales and to determine the most popular categories of grocery products in both stores.

3.What is the correlation between the sale price and market price of the products?

The line chart visualization of the comparison between sale price and market price by product category in BigBasket and Dmart helped to identify the correlation between these two factors. It showed whether the sale price of a product was influenced by its market price or if it was independent of it.

4.What is the profit amount of each product, and how does it compare to the sale and market prices?

The donut chart visualization of the profit by brand in BigBasket and Dmart helped to identify the profitability of each brand in both stores. The waterfall chart visualization of the comparison of sales and market prices for grocery subcategories in BigBasket and Dmart helped to identify the profit margins of each subcategory and how they compared to the sale and market prices.

5.What are the top subcategories of grocery products sold in both stores, and how do their sales compare?

The visualization of the top subcategories of grocery products sold in both stores provided a clear picture of the most popular subcategories in BigBasket and Dmart. The comparison of their sales helped to identify the leading subcategories in terms of overall sales.

6.How do the sales of each subcategory of grocery products in both stores compare, and which store has the higher sales?

The visualization of sales by subcategory in both stores provided a clear comparison of the sales performance of each subcategory of grocery products in BigBasket and Dmart. The visualization helped to identify the leading store in terms of sales and to determine the most popular subcategories of grocery products in both stores.

7.What is the proportion of each subcategory of grocery products in the total sales of both stores?

The treemap visualization of sale price by category in both stores helped to identify the proportion of each subcategory of grocery products in the total sales of both stores. It helped to determine which subcategories of grocery products contributed the most to the overall sales of both stores.Therefore, the research questions were designed to provide a comprehensive understanding of the sales performance of Bigbasket and Dmart and help stakeholders make data-driven decisions.

Some potential further research questions could be:

1. What is the relationship between product pricing and customer reviews or ratings? Does higher pricing necessarily translate to higher ratings or vice versa?
2. What are the top-selling products in each subcategory, and how do they compare in terms of sales and profitability?
3. Are there any seasonal patterns or trends in the sales and pricing of grocery products in both stores?
4. How does the availability of discounts and promotions impact customer purchase behavior and store profitability?
5. How do the sales and profitability of online grocery stores compare to traditional brick-and-mortar stores, and what are the implications for the industry as a whole?

I have created several advanced visualizations using maps, filters, drill down that can be found in My Power BI workbook which will answer all research questions.