

Name: **Swagat Kumar Meher**

UNID: **UM20231437**

Project Title: **Supermart Grocery Sales – Retail Analytics using Python and Excel**

### **Report on Supermart Grocery Sales – Retail Analytics**

I obtained this fictional dataset to perform analysis using Python and Excel. It includes grocery order information from a delivery app, focusing on customers in Tamil Nadu, India. This dataset will help me explore customer behaviour and order trends effectively.

### **Objective**

The objective of this project is to analyse grocery order data from a fictional delivery app to gain insights into customer behaviour and order patterns in Tamil Nadu, India. By utilizing Python and Excel for exploratory data analysis and visualization, I aim to identify trends, optimize inventory management, and enhance customer satisfaction through data-driven decisions.

### **Dataset Link**

<https://drive.google.com/file/d/1Vx-lbn11HKofkJasjMZfyigemSu7TOeB/view>

### **Data Cleaning Process**

1. **Date Format Correction:** The “Date” column was initially in the wrong format (either “MM-DD-YYYY” or “M-DD-YYYY”). I converted it to the correct format (“DD-MM-YYYY”).
2. **Data Type Verification:** I checked the data types of all columns to ensure they were appropriate for analysis.
3. **Null Values and Duplicates:** I examined the dataset for any null values and duplicates, addressing any issues found.
4. **File Saving:** After cleaning the data in Excel, I saved the file in CSV format under the new name "Supermart Grocery Sales – Retail.csv".

### **Jupyter Notebook Link**

[Supermart Grocery Sales - Retail Analytics](#)

### **Data Analysis and Analytics**

#### **Visualization**

- **Histogram:** Shows distribution Sales and Profit.
- **Scatter Plot:** Shows correlation between Sales and Profit.
- **Scatter Plot:** Shows correlation of “Sales vs Discount” and “Profit vs Discount”.

- **Dual Bar Chart:** Shows Top 10 Sales and Profit by Order Date.
- **Bar Chart:** Shows Total Sales Category and Total Profit by Category.
- **Bar Chart:** Shows Total Sales Sub Category and Total Profit by Sub Category.
- **Bar Chart:** Shows Top Cities.
- **Bar Chart:** Shows Top 10 Customers.
- **Pie Chart:** Shows distribution of Regions.

## **Key Insights**

### **Distribution of Sales and Profit**

**Sales distribution:** Sales histogram is symmetrical, indicating that the distribution of sales is relatively even across different ranges.

**Profit distribution:** Profit histogram is right-skewed, indicating that the majority of profit values are concentrated on the lower end of the scale, with a tail extending to the right.

### **Correlation between Sales and Profit**

In analysing the relationship between sales and profit using a scatter plot, I found a correlation coefficient of approximately 0.61. This positive correlation indicates a moderate relationship between the two variables, suggesting that as sales increase, profit tends to increase as well.

### **Correlation of “Sales vs Discount” and “Profit vs Discount”**

**Sales vs. Discount:** The correlation coefficient between sales and discount is approximately -0.0055. This value indicates an almost negligible negative correlation, suggesting that there is no significant relationship between sales and the amount of discount offered.

**Profit vs. Discount:** The correlation coefficient between profit and discount is approximately 0.0000175 (or 1.75e-05). Similar to the previous analysis, this value indicates a virtually non-existent relationship between profit and discount.

### **Top 10 Sales and Profit by Order Date**

In the analysis of the top 10 sales and profit figures by order date, a dual axis bar chart was created to visually compare these two key performance metrics.

#### **Top Sales Overview:**

- The highest sales were recorded on September 2, 2018, with a total of Rs. 55,976. This date stands out as a significant sales peak.
- Other notable sales figures include Rs. 53,891 on December 2, 2018, and Rs. 53,561 on September 5, 2017.

### **Top Profit Overview:**

- The highest profit was also on December 2, 2018, at Rs. 15,544.23, highlighting that this date was not only a strong sales day but also a high-profit day.
- Significant profit values include Rs. 14,878.27 on September 5, 2017, and Rs. 14,366.96 on November 10, 2017.

**Correlation Observations:** There is a noticeable alignment between high sales and high profit on specific dates, particularly December 2, 2018.

**Trends and Insights:** The analysis indicates that certain periods, particularly around late 2018, were key for both sales and profits, suggesting effective marketing or seasonal demand.

### **Total Sales Category and Total Profit by Category**

In the analysis of sales and profit by category, separate bar charts were created to visualize the performance of different product categories.

#### **Sales by Category:**

- The leading category in sales is Eggs, Meat & Fish, generating a total of Rs. 2,267,401.
- Snacks closely follows, with sales of Rs. 2,237,546, showcasing strong consumer demand.
- Other notable categories include Food Grains at Rs. 2,115,272 and Bakery at Rs. 2,112,281.

#### **Sales Distribution:**

- The data reveals a competitive landscape among categories, with all categories performing well above the Rs. 2 million marks.
- Beverages and Oil & Masala are at the lower end, yet still contribute significant sales figures.

#### **Profit by Category:**

- Snacks leads in profit with Rs. 568,178.85, indicating a strong margin relative to sales.
- Eggs, Meat & Fish follows closely with a profit of Rs. 567,357.22, demonstrating effective cost management.
- Fruits & Veggies also perform well, generating Rs. 530,400.38 in profit.

#### **Profit Distribution:**

The profit figures show that while some categories like Snacks and Eggs, Meat & Fish lead in both sales and profit, others, such as Oil & Masala, show lower profitability at Rs. 497,895.29 despite decent sales.

**Correlation Between Sales and Profit:** There is a notable correlation between high sales and high profit in categories like Snacks and Eggs, Meat & Fish, indicating effective pricing and cost strategies.

**Focus Areas:** Categories with high sales but lower profit margins (like Oil & Masala) may require a review of pricing strategies or cost structures to enhance profitability.

### **Total Sales Sub Category and Total Profit by Sub Category**

In the analysis of sales and profit by sub category, separate bar charts were created to visualize the performance across various product subcategories.

#### **Sales by Subcategory:**

- Health Drinks lead with sales of Rs. 1,051,439, indicating a strong consumer preference for health-oriented products.
- Soft Drinks follow closely with sales of Rs. 1,033,874, reflecting the popularity of refreshing beverages. Other significant contributors include Cookies at Rs. 768,213 and Breads & Buns at Rs. 742,586.

#### **Sales Distribution:**

- The sales figures show a diverse range of products performing well, with most subcategories exceeding Rs. 500,000 in sales.
- Categories like Organic Staples and Fresh Vegetables are on the lower end but still maintain respectable sales figures, indicating niche markets.

#### **Profit by Subcategory:**

- Health Drinks again lead in profit with Rs. 267,469.79, showcasing not just high sales but also strong profitability.
- Soft Drinks follow with a profit of Rs. 258,135.97, indicating good margins for this popular category.
- Noodles and Breads & Buns also perform well in profit, generating Rs. 193,685.81 and Rs. 190,764.98, respectively.

#### **Profit Distribution:**

- While many subcategories show strong sales, the profit figures reveal that some, like Mutton and Fish, while having decent sales, still generate lower profit margins compared to other categories.
- The lower end of the profit spectrum includes items like Rice and Chicken, which may indicate tighter margins or higher costs associated with these products.

**Correlation Between Sales and Profit:** The strong performance of Health Drinks in both sales and profit suggests effective pricing and demand strategies. This alignment indicates a potentially lucrative area for investment and marketing focus.

**Opportunity for Improvement:** Categories like Fish and Mutton, despite decent sales, could benefit from cost management strategies to enhance profitability.

### **Top Cities by Order Count**

In the analysis of the top 10 cities based on order count, a bar chart was created to visualize the frequency of orders from each city.

### **Top Cities:**

- Kanyakumari ranks highest with 459 orders, indicating a strong customer base and demand in this region.
- Tirunelveli follows closely with 446 orders, showing significant engagement in this city as well.
- Other notable cities include Bodi (442), Krishnagiri (440), and Vellore (435), all of which contribute robustly to the overall order volume.

**Distribution:** The bar chart clearly illustrates a competitive landscape among the top cities, with the counts decreasing incrementally. The numbers suggest a clustering of orders around certain key locations, with a tight range between the top five cities.

### **Insights on Consumer Behaviour:**

- The strong representation of cities in the southern region (like Kanyakumari and Tirunelveli) may indicate regional preferences or effective local marketing strategies.
- Cities like Chennai and Salem, although lower in the top ten, still demonstrate solid order counts (432 and 431, respectively), reflecting their importance in the overall market.

**Market Strategy:** The analysis suggests that businesses should consider targeted marketing strategies in the top cities, especially Kanyakumari and Tirunelveli, to further capitalize on their strong customer bases.

### **Top Customers by Order Count**

In the analysis of the top 10 customers based on order count, a bar chart was created to visualize the frequency of orders placed by each customer.

### **Top Customers:**

- Amrish is the top customer with 227 orders, indicating a strong loyalty or high purchasing frequency.
- Krithika follows closely with 224 orders, showcasing similarly high engagement.
- Arutra and Verma are tied with 218 orders each, highlighting a competitive customer base.

### **Order Count Distribution:**

- The bar chart shows a gradual decline in order counts among the top customers, with the top four customers all exceeding 200 orders.
- The counts for customers from Vidya (215) down to Hussain (208) remain closely clustered, indicating that several customers are highly active in their purchasing behaviour.

### **Insights on Customer Loyalty:**

- The concentration of orders among a few customers suggests a strong reliance on these key individuals for sales, indicating the importance of customer relationship management.
- Engaging these top customers with loyalty programs or personalized marketing could enhance retention and potentially increase their order frequency.

**Marketing Strategies:** Businesses should consider targeted marketing efforts aimed at these top customers to encourage repeat purchases and enhance loyalty.

## **Distribution of Regions**

The pie chart visualizes the distribution of sales across different regions: West, East, Central, South, and North. Here's a detailed analysis of the chart:

### **Regional Representation:**

- Each region's proportion is clearly defined, showcasing the relative sales volume contributed by each area.
- The chart includes an explosion effect for the West region, emphasizing its significance in the overall distribution.

### **Proportional Insights:**

- West Region: The exploded section indicates that this region is a major contributor to sales, suggesting strong market performance and customer engagement.
- Other regions (East, Central, South, North) are represented without explosion but still indicate their respective contributions to total sales.

### **Percentage Distribution:**

- The autopct feature displays the percentage of total sales for each region, providing a clear view of their relative sizes.
- Analysing the percentages helps identify which regions may require more focus or marketing efforts.

**Strategic Focus:** The West region's prominence suggests that businesses may want to invest more resources there, whether through marketing, inventory, or customer service improvements.

**Growth Opportunities:** Regions with smaller percentages could indicate potential growth areas, and targeted strategies could be developed to increase market share in those regions.

## **Conclusion**

The analysis of Supermart's grocery sales data has revealed valuable insights into the performance of various categories, subcategories, customer behaviour, and regional trends. Key findings highlight that categories such as Eggs, Meat & Fish and Snacks not only drive significant sales but also yield substantial profits. This underscores the importance of focusing marketing and operational strategies on these high-performing areas.

Customer analysis shows a strong concentration of orders among a select few individuals, indicating potential for enhanced loyalty programs to further engage these key customers. Geographically, the West region stands out as a dominant contributor to sales, suggesting targeted investments and marketing efforts could yield even greater returns.

Overall, the insights derived from this analysis enable Supermart to make informed, data-driven decisions that can optimize sales strategies and improve customer engagement. By leveraging these findings, the business is well-positioned to enhance profitability and sustain competitive advantage in the grocery retail market. Future analyses could explore deeper customer segmentation and seasonal trends, further refining strategies to meet evolving consumer demands.