Report on Mini Project

Name: Amudha. N

Batch: TN-DA-ANB11

Mail Id: sarah1986aju@gmail.com

Contact No: 9791178867

Tittle: Retail Stores Sales Dataset

Project Domain: Sales and E-Commerce

Data set year Timeline: 2022-2025 Dataset

Submission: 05-01-2025

Mentor: Kumaran. M

Raw Dataset:

https://docs.google.com/spreadsheets/d/1xWpNJOIVkho3buUegv5sMr7TfPCb_D9l/edit?usp=sharing&ouid=110031930354006978938&rtpof=true&sd=true

Cleaned Dataset: https://docs.google.com/spreadsheets/d/1EPuPb8vSb4aTDzc-FrK_-rkwczFGtfkS/edit?usp=sharing&ouid=110031930354006978938&rtpof=true&sd=true

Visuals Dataset:

https://drive.google.com/file/d/18SQyzA5qh6wVPUHhn9hsdTmvlutYHMOF/view?usp=sharing

Mapping Freelancer Retail Store Sales: Data Cleaning, Data Transformation, and Data Visualization

This project aims to analyze and visualize data from a global freelancer dataset to gain insights into demographics, skills, earnings, and performance trends. The dataset contains 12,575 freelancer profiles with information such as freelancer details, transactions, customer IDs, categories, items, prices per unit, quantities, total spend, payment methods, locations, transaction dates, and discounts applied.

Derived Columns for Enhanced Analysis

To enhance the analysis, several derived columns were calculated from the existing data:

- Gross Amount – Calculated as *Price Per Unit × Quantity* to verify total spending before discounts.
- Discount Applied – Discount was applied based on the payment method.
- Transaction Month – Extracted from the transaction date for monthly trend analysis.
- Transaction Year – Extracted to analyze yearly sales patterns.
- Spending Category – Grouped based on total amount spent per transaction.
- Payment Type Group – Categorized into Digital and Non-Digital payments.
- Sales Channel – Derived from location (Online or Offline).
- Average Price per Item – Calculated to compare pricing across categories.

The analysis and visualization help answer questions such as:

- Which product categories generate the highest sales?
- Which items are the most frequently purchased?
- How do sales differ between online and offline locations?
- Which payment methods are most commonly used?
- What is the impact of discounts on total sales?

Tools and Technologies Used

- **Microsoft Excel** – Used for data cleaning, formatting, and basic data transformation.
- **Power BI** – Used for data visualization and creating interactive dashboards.
- **Data Visualization Techniques** – Used to present insights through charts, graphs, and reports.

DATA CLEANING

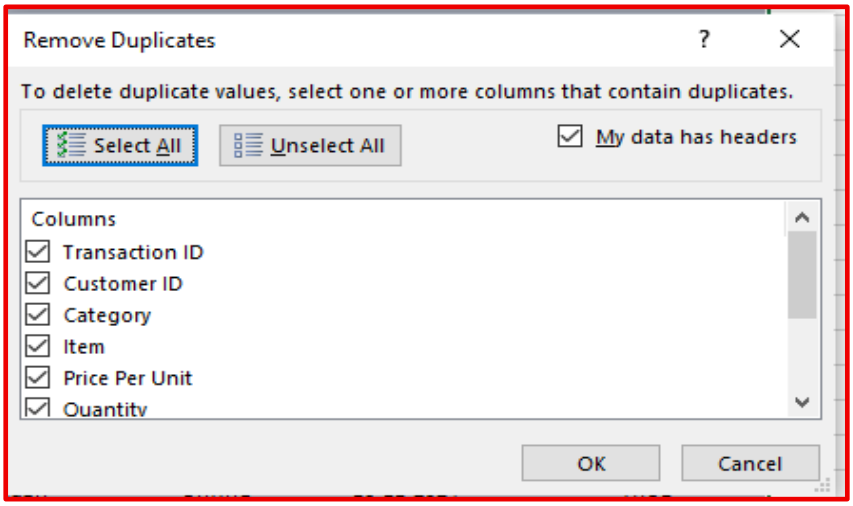
1.Trim the Data: Trimmed the Data to removed extra spaces from all the text fields.

```
=TRIM([@[Category ]])
```

C
Category
Food
Furniture
Furniture
Butchers
Butchers
Milk Products
Food
Patisserie
Milk Products
Beverages
Patisserie
Milk Products
Electric household essentials
Furniture
Beverages
Milk Products
Electric household essentials
Food
Furniture

Category
Food
Furniture
Furniture
Butchers
Butchers
Milk Products
Food
Patisserie
Milk Products
Beverages
Patisserie
Milk Products
Electric household essentials
Furniture
Beverages
Milk Products
Electric household essentials
Food
Furniture
Furniture

2.Remove the Duplicates: After completing this process I started to clean and removed the data in duplicate column.



3.Missing Values in Item: I noticed missing values in the item sequence, but I did not consider them duplicates and filled them sequentially. For instance, if the unit price is 5, the corresponding item number is 1.

Category	Item
Patisserie	Item_10_PAT
Milk Products	Item_17_MILK
Butchers	Item_12_BUT
Beverages	Item_16_BEV
Food	Item_6_FOOD
Patisserie	
Food	Item_1_FOOD
Furniture	
Furniture	Item_16_FUR
Butchers	Item_22_BUT
Butchers	Item_3_BUT
Milk Products	
Food	Item_2_FOOD
Patisserie	Item_24_PAT
Milk Products	Item_16_MILK
Beverages	

Category	Item
Patisserie	Item_10_PAT
Milk Products	Item_17_MILK
Butchers	Item_12_BUT
Beverages	Item_16_BEV
Food	Item_06_Food
Patisserie	Item_10_PAT
Food	Item_01_Food
Furniture	Item_20_FUR
Furniture	Item_16_FUR
Butchers	Item_22_BUT
Butchers	Item_03_BUT
Milk Products	Item_02_MILK
Food	Item_02_Food
Patisserie	Item_24_PAT
Milk Products	Item_16_MILK
Beverages	Item_14_BEV

4.Missing value in price per unit column: For the missing values of price per unit column were calculated using the formula.

```
=[@[Total Spent]]/[@Quantity]
```

Category	Price Per Unit
Patisserie	5
Food	
Food	
Beverages	24.5
Food	11
Beverages	9.5
Patisserie	12.5
Furniture	12.5
Beverages	39.5
Computers ar	12.5
Computers ar	35
Milk Products	30.5
Electric house	27.5
Milk Products	
Milk Products	24.5

Category	Price Per Uni
Patisserie	5
Food	21.5
Food	18.5
Beverages	24.5
Food	11
Beverages	9.5
Patisserie	12.5
Furniture	12.5
Beverages	39.5
ters and electric acce	12.5
ters and electric acce	35
Milk Products	30.5
tric household essen	27.5
Milk Products	9.5
Milk Products	24.5

5.Missing value in Quantity column: For the missing values of quantity column were calculated using Item mode values.

Category	Quantity
Food	8
Furniture	
Furniture	1
Butchers	3
Butchers	9
Milk Products	8
Food	7
Patisserie	6
Milk Products	2
Beverages	
Patisserie	8
Milk Products	10

Category	Quantit
Food	8
Furniture	1
Furniture	1
Butchers	3
Butchers	9
Milk Products	8
Food	7
Patisserie	6
Milk Products	2
Beverages	2
Patisserie	8
Milk Products	10

6. Missing value in Total Spent column: For the missing values of quantity column were calculated using Item mode values.

```
=[@[Price Per Unit]]*[@Quantity]
```

Category	Total Spent
Food	40
Furniture	
Furniture	27.5
Butchers	109.5
Butchers	72
Milk Products	52
Food	45.5
Patisserie	237
Milk Products	55
Beverages	
Patisserie	232
Milk Products	275
Electric house	23
Furniture	

Category	Price Per Uni
Food	5
Furniture	33.5
Furniture	27.5
Butchers	36.5
Butchers	8
Milk Products	6.5
Food	6.5
Patisserie	39.5
Milk Products	27.5
Beverages	24.5
Patisserie	29
Milk Products	27.5
ric household essen	23
Furniture	35

7. Find the Total Sales value in DAX using Power BI

```
1  Total Sales = SUM('Retail_store_sales - Main Proje'[Total Spent])
```

Total Sales Value
245
155
245
335
110
95
215
275
365
125
245
275
155
200
50

8. Find the Total Product Count value in DAX using Power BI

```
1  Total Product = COUNT('Retail_store_sales - Main Proje'[Item])
```

Item
Item_14_Food
Item_08_CEA
Item_14_FUR
Item_20_PAT
Item_5_PAT
Item_04_EHE
Item_12_Food
Item_16_FUR
Item_22_PAT
Item_06_Food
Item_14_Food

8. Find the Total Average sales value in DAX using Power BI

```
1  Average Sales = AVERAGE('Retail_store_sales - Main Proje'[Total Sales Value])
```

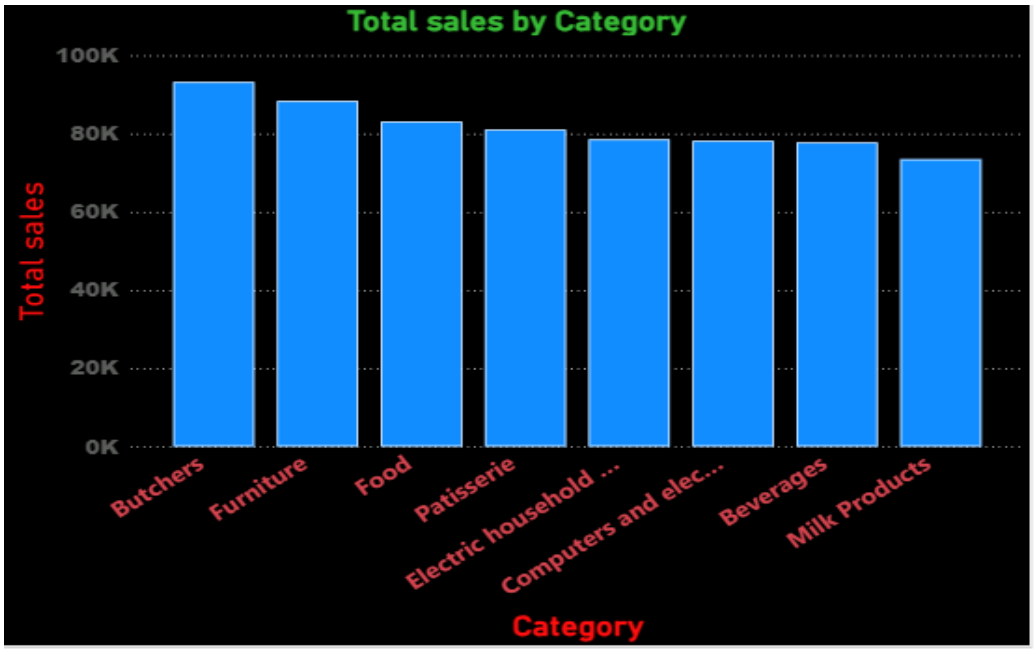
Total Sales Value
245
155
245
335
110
95
215
275

9. Find the Total count of Quantity value in DAX using Power BI

```
1  Sum of Quantity = SUM('Retail_store_sales - Main Proje'[Quantity])
```

Quantity
10
10
10
10
10
10
10

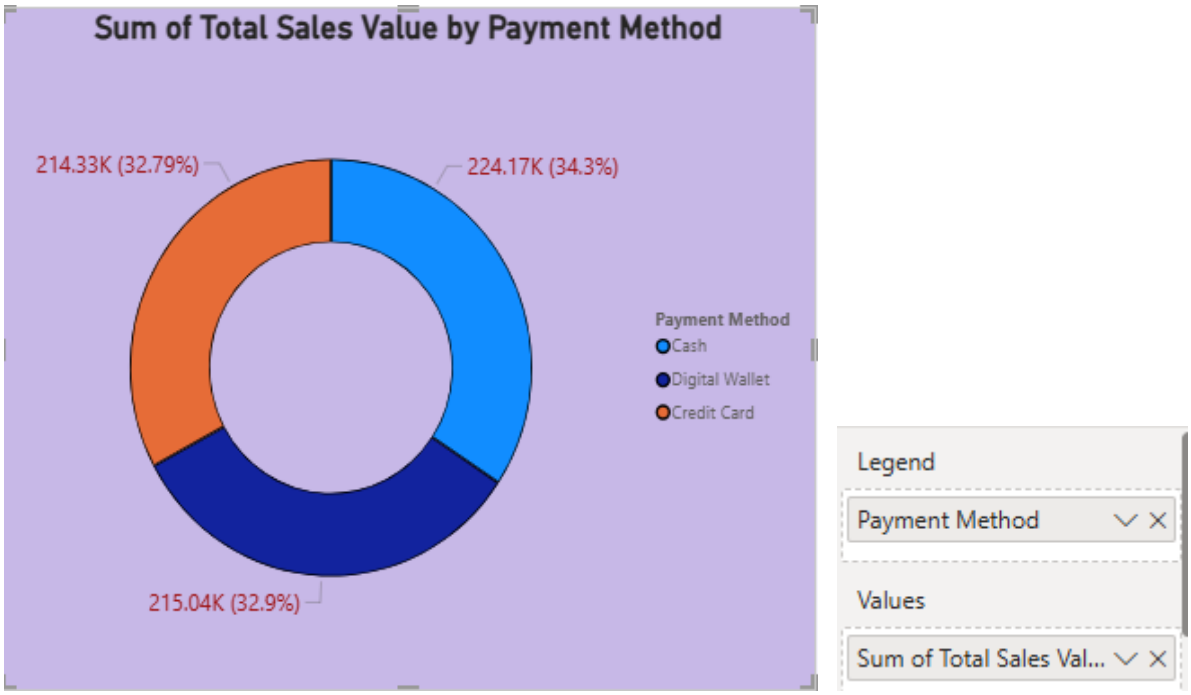
10. **Bar Chart:** The bar chart compares total sales across different product categories. It shows that Butchers has the highest sales, followed by Furniture and Food, while Milk Products has the lowest sales. This indicates which categories contribute most to revenue and which need improvement.



Key Insights

- Butchers' category generates the highest revenue.
- Furniture and Food are strong contributors.
- Milk Products show the lowest sales, indicating improvement opportunity.
- Business should prioritize top-selling categories for stock and promotions.

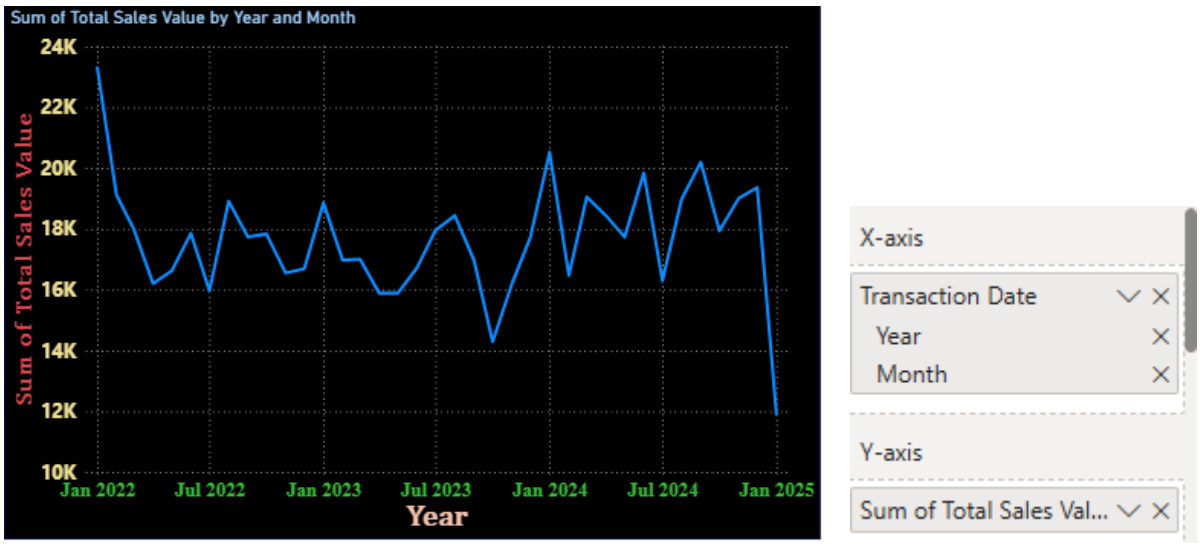
11. Donut Chart: The donut chart displays the percentage distribution of sales by payment method. It shows that Cash, Credit Card, and Digital Wallet are almost equally used, with a slight preference for Cash and Digital Wallets. This indicates balanced customer payment behaviour with growing digital adoption.



Key Insights:

- Cash, Credit Card, and Digital Wallet usage is almost evenly split.
- Slight dominance of Digital Wallet & Cash, showing digital adoption.
- Retailer must support multiple payment options to satisfy customer.

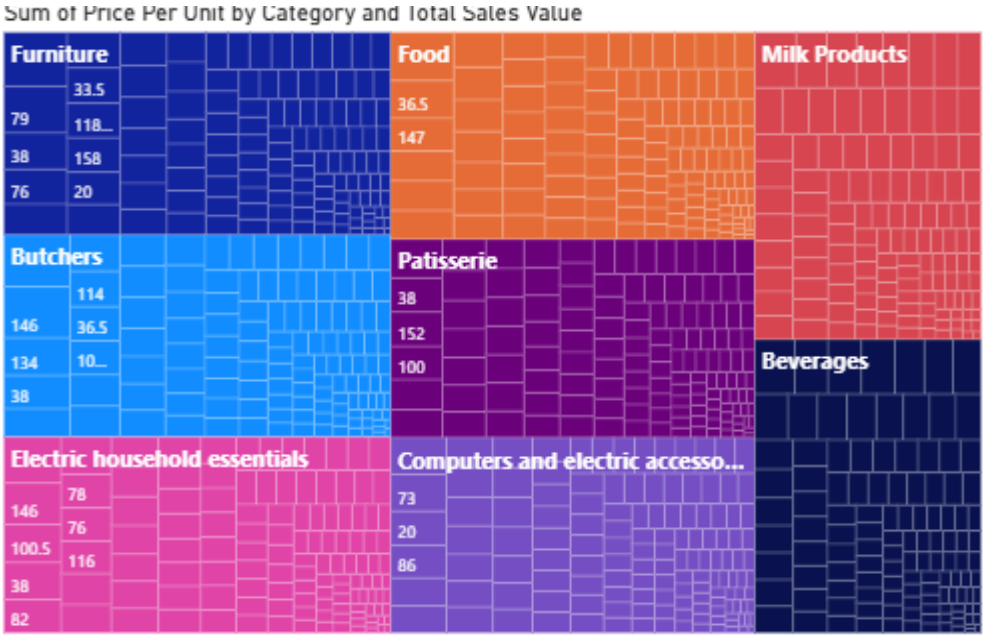
12. Monthly Sales Trend (Line Chart): To analyze sales performance over time and identify trends or seasonality.

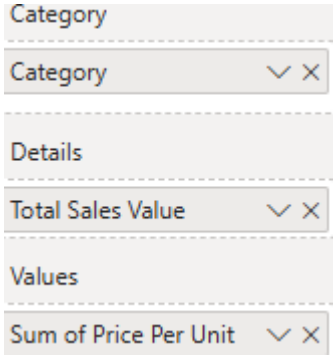


Key Insights:

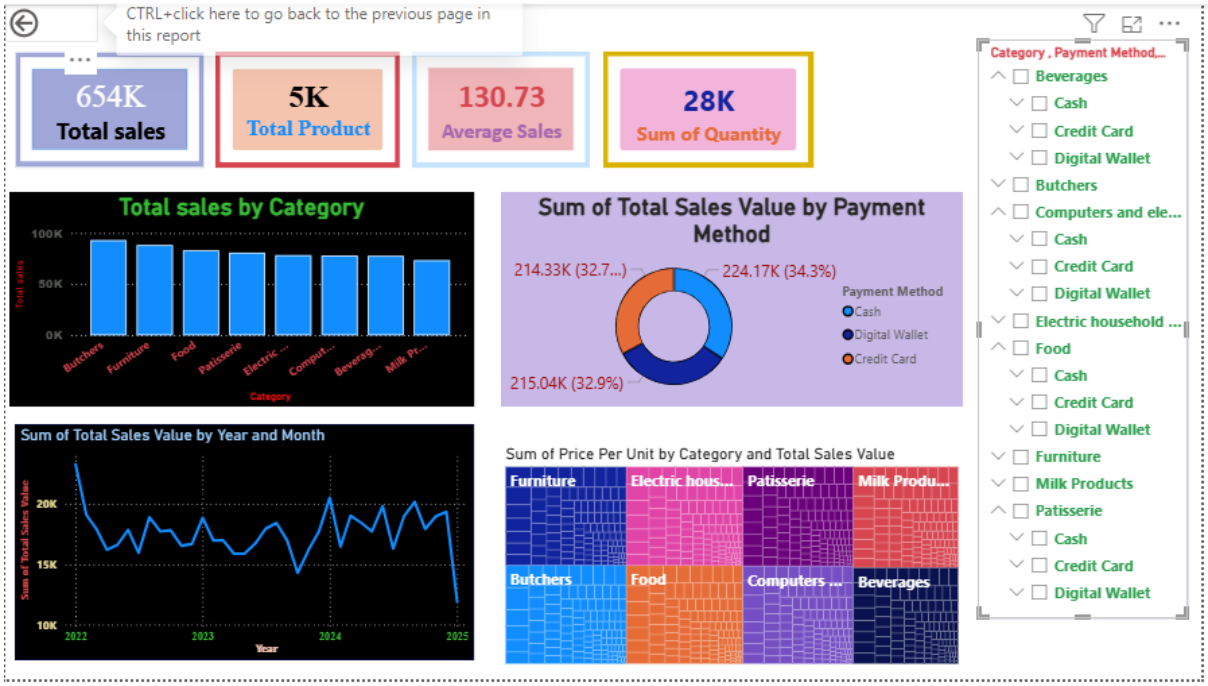
- Sales show fluctuations month by month, not a constant pattern.
- There is a peak around early 2024, indicating higher customer activity.
- A drop towards the latest months suggests seasonal effects or reduced demand.
- Useful for forecasting and planning inventory.

13.Tree Map: To analyze Category performance Total sales value and sum of price per unit identify the trends.





- Food and Milk Products dominate overall sales value, indicating they are the primary revenue drivers despite moderate unit prices.
- Beverages occupy a large area with relatively balanced pricing, showing consistent demand and stable contribution to total sales.
- Furniture and Computers & Electric Accessories show higher price per unit but smaller sales value, suggesting they are high-ticket but lower-volume categories.
- Butchers and Patisserie contribute mid-level sales, reflecting steady but not peak demand, likely dependent on daily/seasonal consumption.
- Electric household essentials have moderate pricing with decent sales share, indicating essential-use products with repeat purchases.
- **Summary:**



The dashboard shows 654K total sales with balanced performance across categories and payment methods. Sales fluctuate over time, and slicers enable quick, interactive analysis by category and payment type.