



Optimization of Inventory, Financial Conditions and Customer Retention of “G Mega Mart”

| Pushpak Ruhil |

| 21F2001180 |

| Business Data Management Project Mid-term Submission|

| September 2022 Term |

Executive Summary

This project analysed organizational data in excel sheets, including customer information, stock information, and sales information. The data was processed and cleaned for accuracy. Descriptive statistics were used to summarize the data, including the number of invoices, bill amount, the number of items, the value of items, and brand and category information. The descriptive statistics used were Min, Max, Count, Average, Median, and Standard Deviation. According to the use case, the sample vs population standard deviation formula was used.

The analysis process followed Google's six-step data analysis process, which includes the steps of asking, preparing, processing, analysing, sharing, and acting. I had done my analysis till the “analysing” phase, and partially the “share” phase.

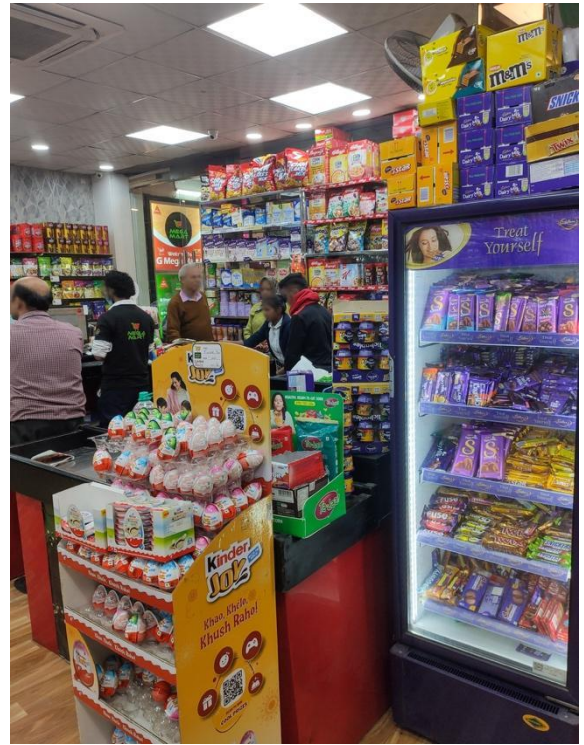
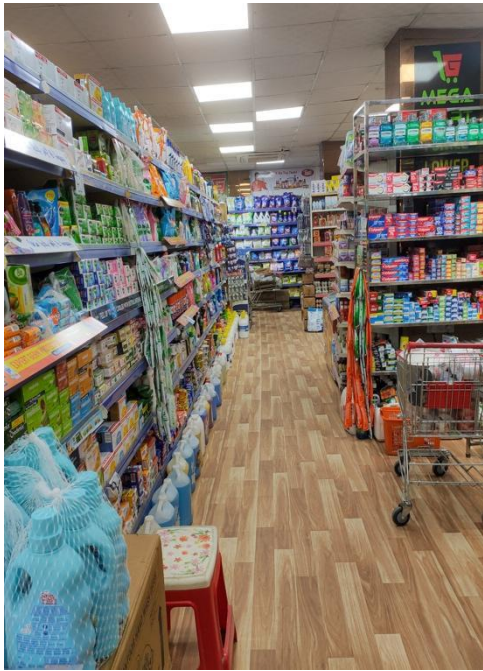
Detailed insights gained from the analysis would be shared in the final report.

Proof Of Originality

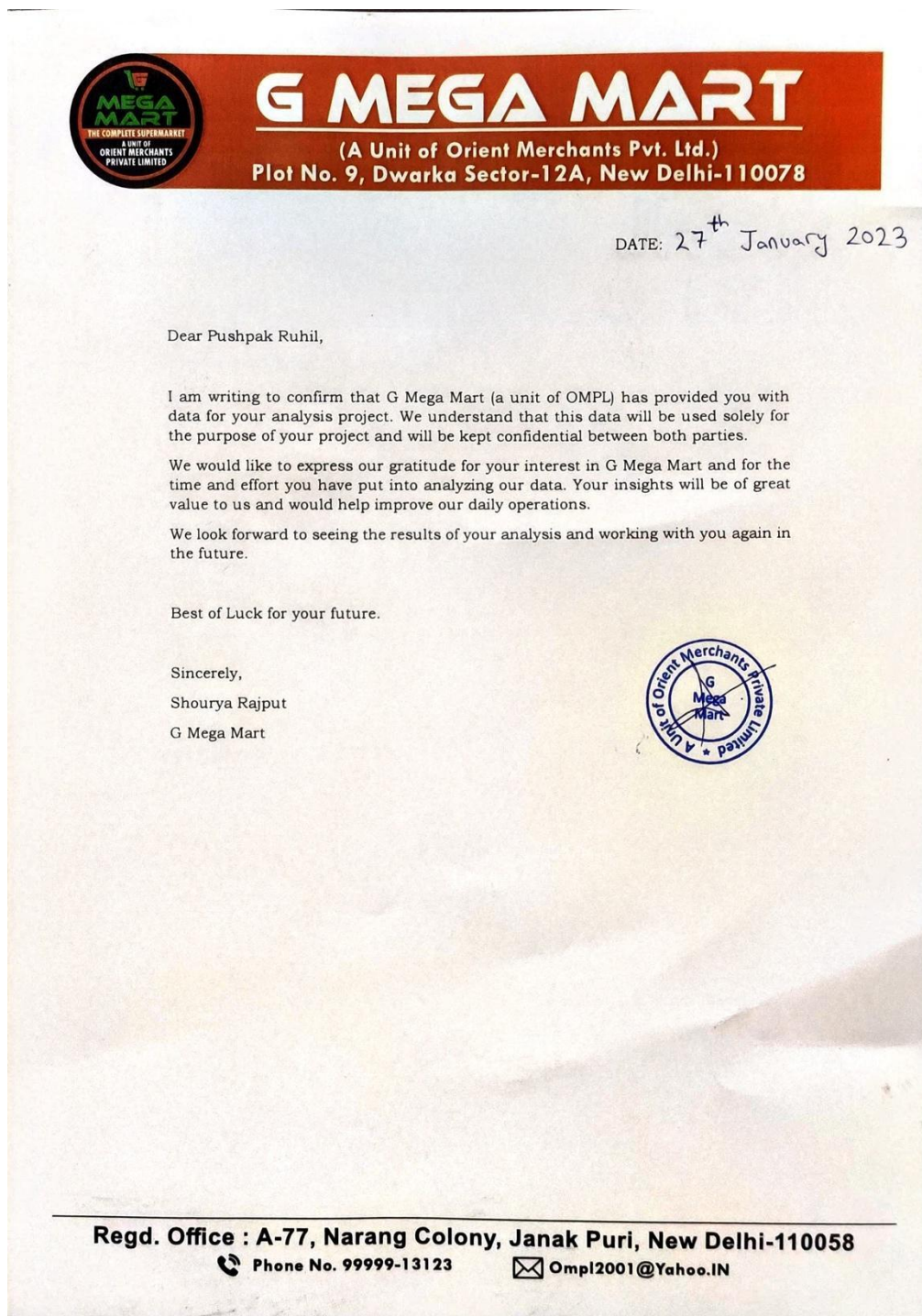
Here is different proof of the originality of the data as well as my project.

Photos:





Letter from the organization:



Video of interaction:

https://drive.google.com/file/d/13CuzM6U6OoOwmZ2Np7MjFdIrS2L-nFu8/view?usp=share_link

Metadata and Descriptive Statistics

The organization granted me access to multiple excel sheets containing their company's data. I began my analysis by backing up all the data files for safekeeping, and then performed descriptive statistics.

METADATA:

- 1) The "Customer_analysis" sheet included customer information such as mobile number, name, and information about the number of purchases and invoices. The data was cleaned to fix data types and remove duplicates.
- 2) The "Stocks_analysis" sheet contained information about stocks such as SKU name, quantity/packs, MRP, cost+GST, total quantity value, and barcode, organized by company.
- 3) The "Stocks with category-analysis" sheet was similar to the previous one, with the addition of a brand name, category, and sub-category of the SKUs.
- 4) The "Data company 22-23-analysis" sheet contained month-wise sales data for various SKUs among all available SKUs. The file was cleaned to separate the month-wise data into different sub-sheets.
- 5) Finally, the "Mama Earth oct 22-analysis" data file contained stock-related information for all the products of the company "Mama Earth," including attributes such as item name, opening stock, inward stock, sales, and closing stock.

DESCRIPTIVE STATISTICS:

Descriptive statistics is a branch of statistics that summarizes data. I attempted to summarize the data for each dataset using basic concepts of descriptive statistics.

Customer-related data:

Number of Invoices	
Total no. of invoices	12182
Customer with min. invoices	1
Customer with max. invoices	57
Mean	2.2584
Median	1
Standard Deviation	3.609

Here, the standard deviation has been calculated using the sample formula.

Bill Amount	
Minimum bill amount	₹ 1.00
Maximum bill amount	₹ 2,75,716.00
Mean	₹ 2,394.22
Median	₹ 899.00
Standard Deviation	₹ 5,381.25

Here, the standard deviation has been calculated using the sample formula.

Sales-related data:

Quantity of items	
Total number of items	9663
Minimum quantity of any item	1
Maximum quantity of any item	1513
Mean	9.2079
Median	4
Standard Deviation	30.318

Here, the standard deviation has been calculated using the population formula.

Value of items	
Minimum	₹0.00
Maximum	₹ 1,20,120.00
Mean	₹ 941.7492
Median	₹ 457.63
Standard Deviation	₹ 2,366.758

Here, the standard deviation has been calculated using the population formula.

Brands	
Total number of brands	964
Minimum number of items of a brand	1
Maximum number of items of a brand	846
Mean	10.0615
Median	3
Standard Deviation	31.0984

Here, the standard deviation has been calculated using the population formula.

Categories	
Total number of categories	16
Minimum number of items of a category	6
Maximum number of items of a category	2812
Mean	10.0615
Median	3
Standard Deviation	31.0984

Here, the standard deviation has been calculated using the population formula.

Analysis processes and methods

I will begin by explaining my approach to the analysis, the functions used, and the methods applied, and then present the insights I have gained thus far.

I have been using Google's data analysis process for my analysis, which says that it is a 6-step process:

Ask, prepare, process, analyse, share and act.

- During the ASK phase, relevant questions were posed to stakeholders.
- The data was then collected (prepared) and processed. These steps include collecting the data and then cleaning it, removing outliers, etc.
- The processed data was then analyzed using various techniques in Excel.
- The share and act phases are the next step (during my final submission).

My first step was to create a backup of the original data files. After that, I started working on the data, sheet-wise.

So, let's begin with the walkthrough, sheet-wise:

Customer-related data:-

- The main objective to analyse this file was to find customer retention.
- This file was first pre-processed. I had assumed that the mobile number would be a primary key here. There were 2 columns – Mobile and Code; these 2 columns were essentially mobile numbers, so I had to merge them both due to some blank values in either of the columns.
- The number of invoices was a text field with the format of “no. of invoices: xxx”. So, I had to extract just the number from this field and convert it to a number data type.
- Conditional formatting was used to identify duplicate mobile numbers. They were merged together with the following rule: add the number of invoices. This merging was achieved with the help of pivot tables.
- This pivot table was used for the descriptive analysis as well as the main analysis.
- For the descriptive statistics, the following formulae used were:
 - MAX
 - MIN
 - AVERAGE
 - COUNTA/COUNT
 - MEDIAN
 - STDEV.S
- The customer retention rate turned out to be 30.70%

Sales-related data:-

- Fortunately, the given data was already clean. There were no duplicates, no outliers, and no errors that I could come up with.

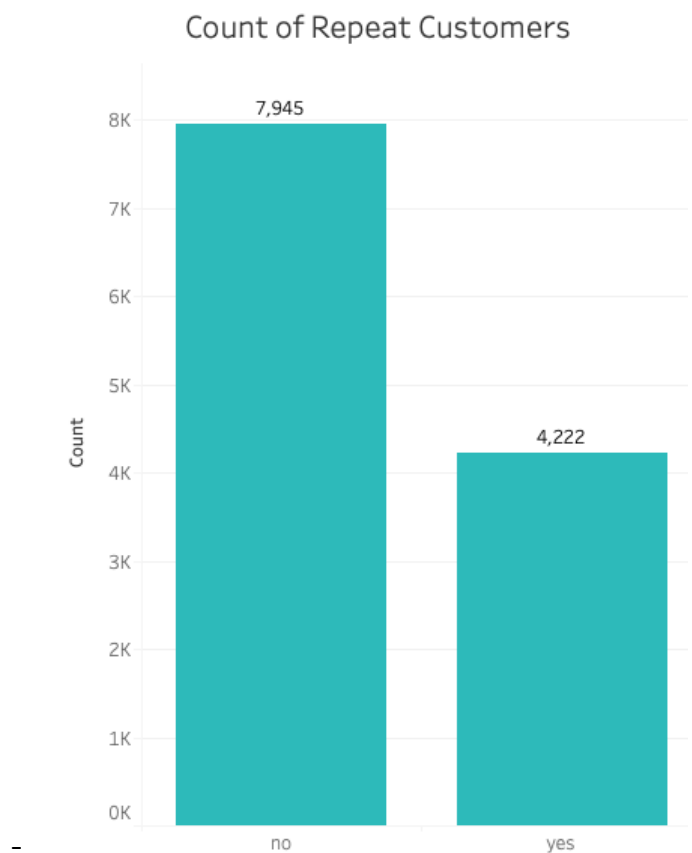
- For my analysis, I created pivot tables.
- The pivot tables had the items, categories and sub-categories as the rows, according to the need. The columns included the count, quantities, values and price.
- Similar set of formulae was used here as well for the descriptive statistics.
- There are a total of 16 unique categories.

So, to summarize the analysis now, below are the findings from the data, till now.

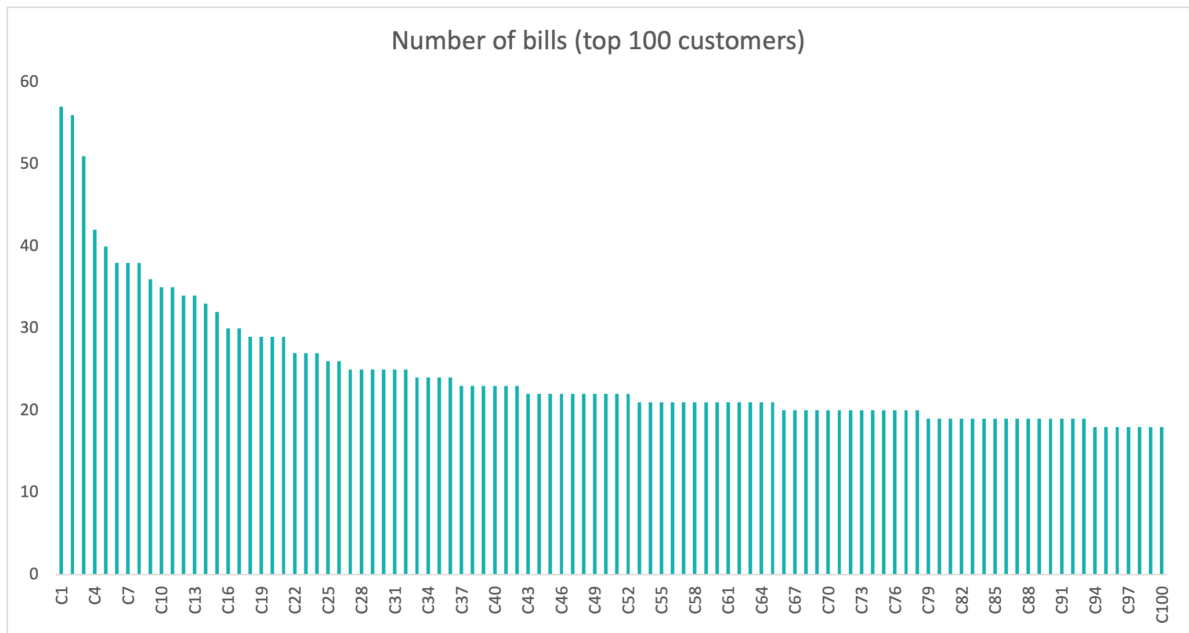
FINDINGS AND INSIGHTS:

Customer-related data:

- Total number of bills recorded by the business is equal to **27513**.
- The maximum number of bills from a single customer is equal to **57**.
- Of the total of **12167** registered customers, there are **7945** registered customers who never came back to the store. Note that there are customers who did not provide their phone numbers and are out of the analysis.
- The bill amount of the customer with the maximum bills is **₹3,723.00**
- While the bill amount of the customer with the 2nd most bills is **₹41,147.00**
- The average bucket size for the registered customers is **₹1,097.66**. This is mainly due to numerous outliers. There are customers with a single bill with a high-value purchase.



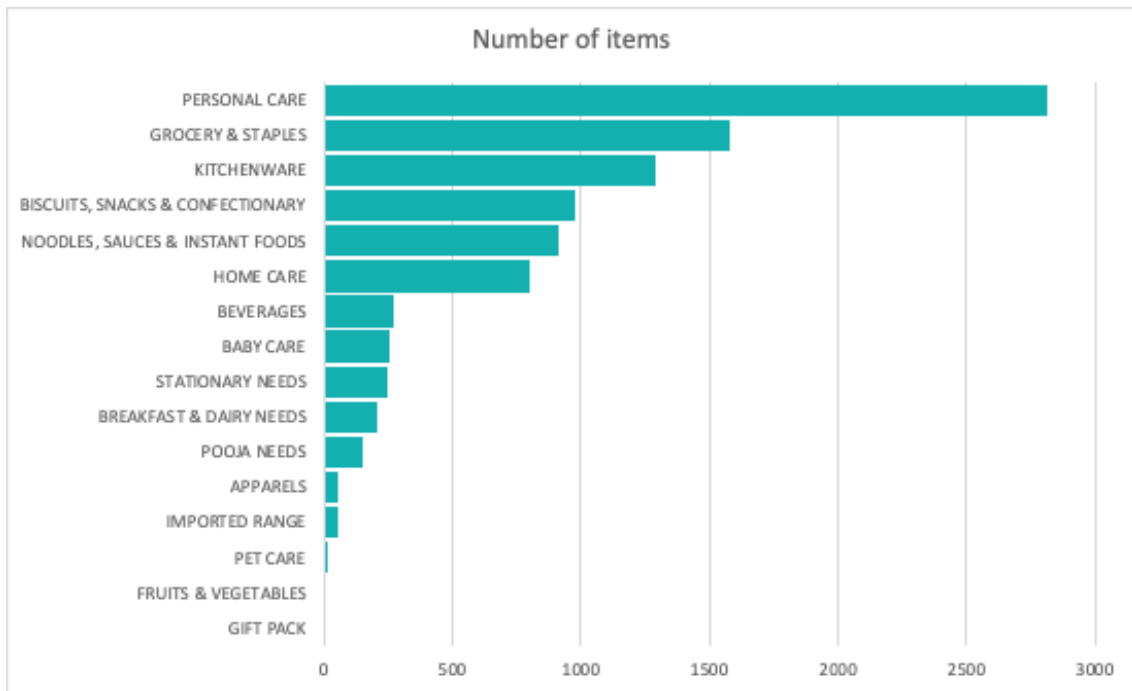
Count of repeating vs non-repeating customers



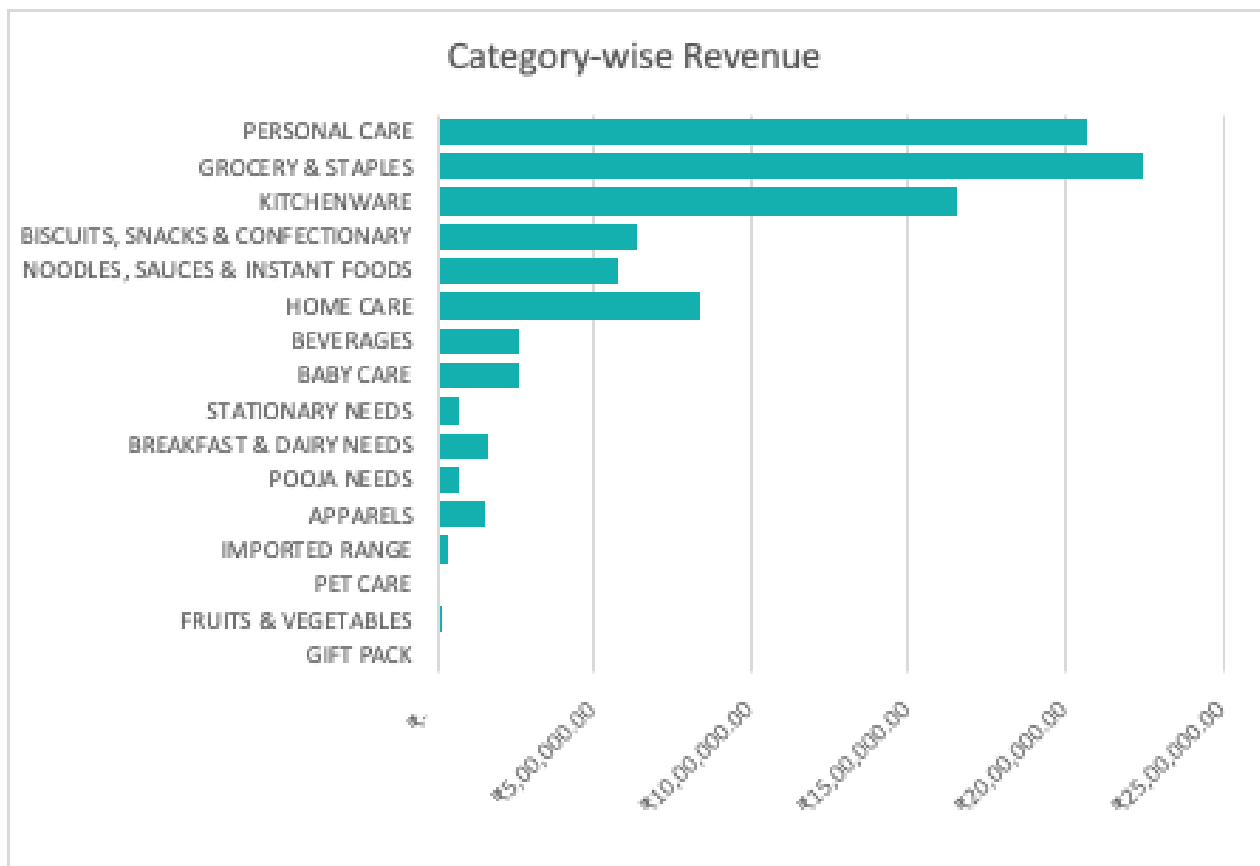
Number of bills of the top 100 customers. C-n is the customer with n-th maximum bills.

Sales-related data:

- Most SKUs are present for the “**Personal Care**” category.
- Under the “Personal Care” category, most SKUs are for the “**Skin Care**” sub-category.
- The category responsible for maximizing the revenue is “**Grocery and Staples**”.
- “**Badam Salora 1kg**” is the product responsible for regenerating the maximum revenue of worth ₹1,20,120/-
- Also, this particular SKU is of the business’ home brand, i.e., “G MEGA MART”, which is a good thing considering that home-brand would also constitute to maximizing the profit.
- There is a total of **151** SKUs which generate no revenue at all. These SKUs are just sitting in the inventory costing the company resources.
- Apart from this, there is a total of **902** SKUs which individually generate revenue of less than ₹100/-
- The business has spent ₹5991.5/- on **1513** 250ML water bottles from July 22-January 23, which is generally kept for customers and is free for them.

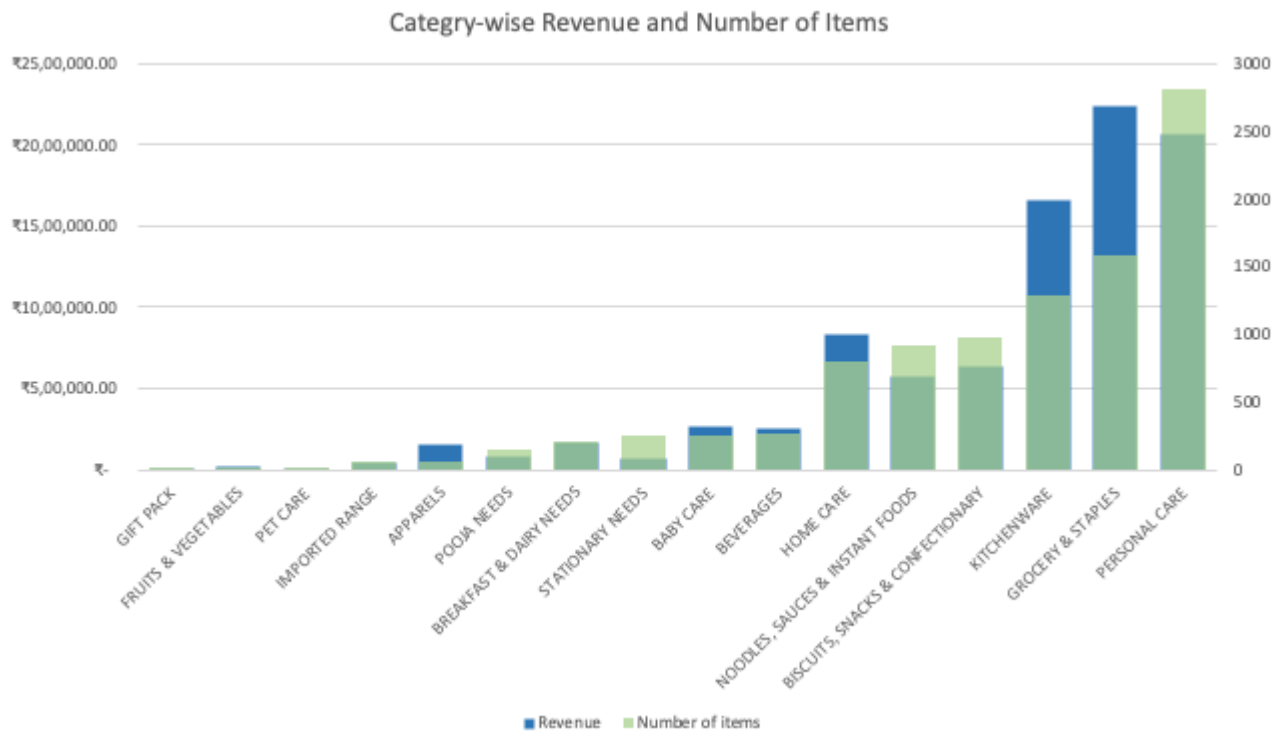


Category-wise number of SKUs



Category-wise revenue generated

The above 2 graphs can be merged together to get something as follows:



Category-wise revenue and number of SKUs

The above graph shows us the number of SKUs vs the revenue generated by each category. Here, it is evident that the maximum revenue is generated by groceries and staples even though the number of SKUs in this category is significantly less than what we have for personal care.