RETAIL SALES ANALYSIS REPORT PROCESS

Mbali Nkosi

FEBRUARY 18, 2025

Table of Contents

Contents

Table of Contents 1
Table of Figures 1
Case Study 3
Tasks 3
Planning And Process 4
Tools 4
Process 4
Findings 7
Data Visualisation 8
Task 8
Basic Depiction of the KPIs to be visualised 8
Total Sales Performance 8
Sales by Region 9
Sales by Category 9
Monthly Sales Trend 10
Customer Insights 10
Top 5 & Bottom 5 Products 11
Final Dashboard 11
Findings 12
Conclusion and Reflection 13

Table of Figures

Figure 1 - Initial Data Set Before Manipulation and Cleaning 4
Figure 2 - Data Set Transformed Into Table Format 4
Figure 3 - Image showing inconsistent data entry 5
Figure 4 – Image showing missing values 5
Figure 5 - Formula used to format the Region column 6
Figure 6 - The formula is applied to "Column1" to ensure we still have the original data we are working with should our formatting be erroneous 6
Figure 7 - The final fixed Region column 6
Figure 8 - Total Sales Performance 8
Figure 9 - Sales by Region 9
Figure 10 - Sales by Category 9
Figure 11 - Monthly Sales Trend 10

Figure 12 - Customer Type Insights 10
Figure 13 - Top 5 & Bottom 5 Products 11
Figure 14 - Final Dashboard 11

Case Study

Scenario: Retail Sales Analysis for 2024 (generated using ChatGPT)

You work as a Data Analyst for **TrendMart**, a nationwide retail chain specializing in electronics, home appliances, and personal gadgets. Your manager has provided you with a **sales dataset** for the 2024 financial year. However, the data is raw and contains errors, duplicates, missing values, and inconsistencies.

Your task is to **clean the dataset** and create a **dashboard** to provide insights into sales performance for management.

Tasks

Tasks:

1. Data Cleaning (Excel & Power Query in Power BI)

- Identify and handle **missing values** (e.g., missing prices, customer IDs, etc.).
- Remove **duplicate records** where applicable.
- Fix **inconsistent date formats** and ensure they align with the fiscal year (Jan–Dec 2024).
- Standardize **product categories and region names** (some might have typos).
- Ensure all monetary values (sales, profit) are formatted correctly.

2. Dashboard & Visualizations (Power BI)

- **Total Sales Performance:** Show total sales revenue, total profit, and number of transactions.
- **Sales by Region:** Which regions performed the best in terms of revenue and profit?
- **Sales by Category:** Which product categories were the most and least profitable?
- **Monthly Sales Trend:** Identify sales trends over the months (e.g., peak seasons, dips).
- **Customer Insights:** Breakdown of returning vs. new customers.
- **Top 5 & Bottom 5 Products:** Based on revenue.

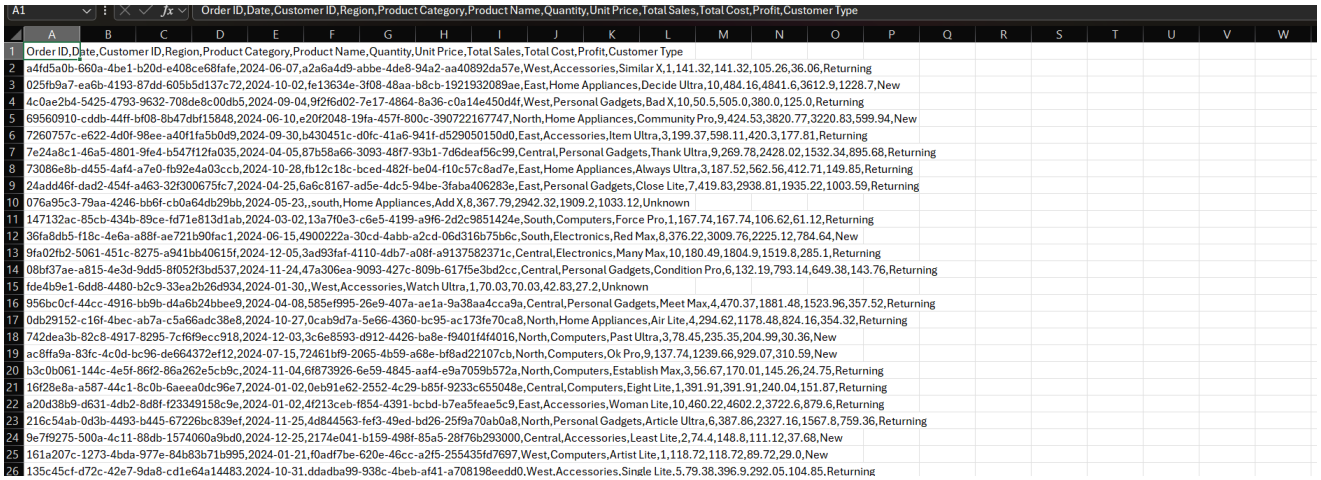
Planning And Process

Tools

I will be using Excel for the main data cleaning process and PowerBI for data visualization and any other further cleaning necessary

Process

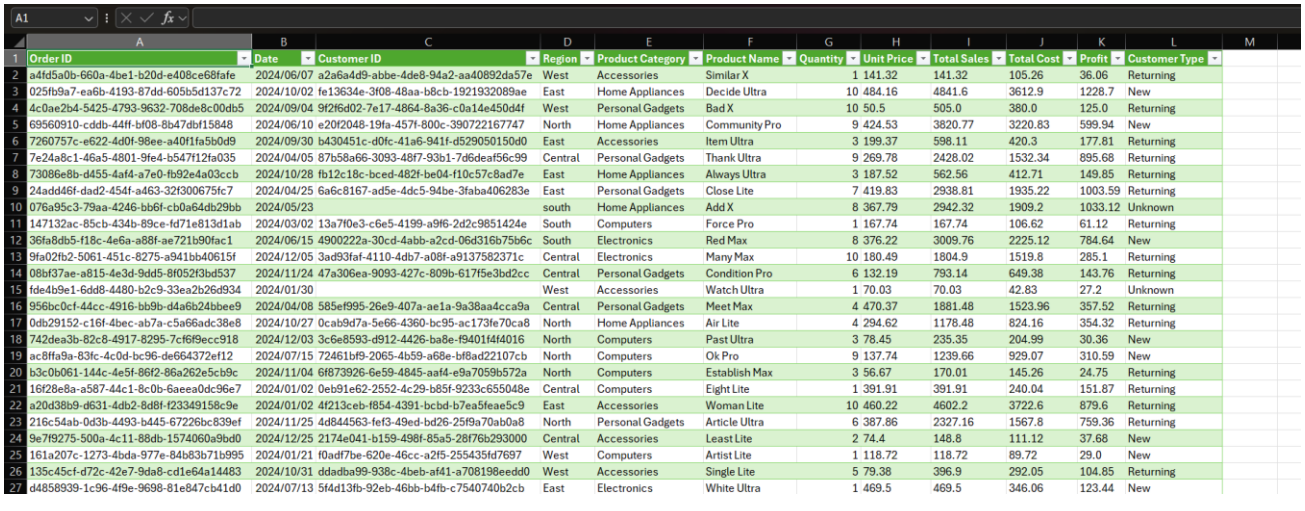
I asked ChatGPT to generate a CSV file for me that contains some unclean data for me to clean and prepare the data for analysis and visualisation. This is the initial format of the CSV file before any data manipulation and cleaning has taken place.



	A
1	Order ID,Date,Customer ID,Region,Product Category,Product Name,Quantity,Unit Price,Total Sales,Total Cost,Profit,Customer Type
2	a4fd5a0b-660a-4be1-b20d-e408ce68fafe,2024-06-07,a2a6a4d9-abbe-4de8-94a2-aa40892da57e,West,Accessories,Similar X,1,141.32,141.32,105.26,36.06,Returning
3	025fb9a7-ea6b-4193-87dd-605b5d137c72,2024-10-02,fe13634e-3f08-48aa-b8cb-1921932089ae,East,Home Appliances,Decide Ultra,10,484.16,4841.6,3612.9,1228.7,New
4	4c0ae2b4-5425-4793-9632-708de8c00db5,2024-09-04,9f2f6d02-7e17-4864-8a36-c0a14e450d4f,West,Personal Gadgets,Bad X,10,50.5,505.0,380.0,125.0,Returning
5	69560910-cddb-44ff-bf08-8b47dbf15848,2024-06-10,e20f2048-19fa-457f-800c-390722167747,North,Home Appliances,Community Pro,9,424.53,3820.77,3220.83,599.94,New
6	7260757c-e622-4d0f-98ee-a40f1fa5b0d9,2024-09-30,b430451c-d0fc-41a6-941f-d529050150d0,East,Accessories,Item Ultra,3,199.37,598.11,420.3,177.81,Returning
7	7e24a8c1-46a5-4801-9fe4-b547f12fa035,2024-04-05,87b58a66-3093-48f7-93b1-7d6deaf56c99,Central,Personal Gadgets,Thank Ultra,9,269.78,2428.02,1532.34,895.68,Returning
8	73086e8b-d455-4af4-a7e0-fb92e4a03ccb,2024-10-28,fb12c18c-bced-482f-be04-f10c57c8ad7e,East,Home Appliances,Always Ultra,3,187.52,562.56,412.71,149.85,Returning
9	24add46f-dad2-454f-a463-32f300675fc7,2024-04-25,6a6c8167-ad5e-4dc5-94be-3faba406283e,East,Personal Gadgets,Close Lite,7,419.83,2938.81,1935.22,1003.59,Returning
10	076a95c3-79aa-4246-bb6f-cb0a64db29bb,2024-05-23,,south,Home Appliances,Add X,8,367.79,2942.32,1909.2,1033.12,Unknown
11	147132ac-85cb-434b-89ce-fd71e813d1ab,2024-03-02,13a7f0e3-c6e5-4199-a9f6-2d2c9851424e,South,Computers,Force Pro,1,167.74,167.74,106.62,61.12,Returning
12	36fa8db5-f18c-4e6a-a88f-ae721b90fac1,2024-06-15,4900222a-30cd-4abb-a2cd-06d316b75b6c,South,Electronics,Red Max,8,376.22,3009.76,2225.12,784.64,New
13	9fa02fb2-5061-451c-8275-a941bb40615f,2024-12-05,3ad93faf-4110-4db7-a08f-a9137582371c,Central,Electronics,Many Max,10,180.49,1804.9,1519.8,285.1,Returning
14	08bf37ae-a815-4e3d-9dd5-8f052f3bd537,2024-11-24,47a306ea-9093-427c-809b-617f5e3bd2cc,Central,Personal Gadgets,Condition Pro,6,132.19,793.14,649.38,143.76,Returning
15	fde4b9e1-6dd8-4480-b2c9-33ea2b26d934,2024-01-30,,West,Accessories,Watch Ultra,1,70.03,70.03,42.83,27.2,Unknown
16	956bc0cf-44cc-4916-bb9b-d4a6b24bbee9,2024-04-08,585ef995-26e9-407a-ae1a-9a38aa4cca9a,Central,Personal Gadgets,Meet Max,4,470.37,1881.48,1523.96,357.52,Returning
17	0db29152-c16f-4bec-ab7a-c5a66adc38e8,2024-10-27,0cab9d7a-5e66-4360-bc95-ac173fe70ca8,North,Home Appliances,Air Lite,4,294.62,1178.48,824.16,354.32,Returning
18	742dea3b-82c8-4917-8295-7cf6f9ecc918,2024-12-03,3c6e8593-d912-4426-ba8e-f9401f4f4016,North,Computers,Past Ultra,3,78.45,235.35,204.99,30.36,New
19	ac8ffa9a-83fc-4c0d-bc96-de664372ef12,2024-07-15,72461bf9-2065-4b59-a68e-bf8ad22107cb,North,Computers,Ok Pro,9,137.74,1239.66,929.07,310.59,New
20	b3c0b061-144c-4e5f-86f2-86a262e5cb9c,2024-11-04,6f873926-6e59-4845-aaf4-e9a7059b572a,North,Computers,Establish Max,3,56.67,170.01,145.26,24.75,Returning
21	16f28e8a-a587-44c1-8c0b-6aeea0dc96e7,2024-01-02,0eb91e62-2552-4c29-b85f-9233c655048e,Central,Computers,Eight Lite,1,391.91,391.91,240.04,151.87,Returning
22	a20d38b9-d631-4db2-8d8f-f23349158c9e,2024-01-02,4f213ceb-f854-4391-bcbd-b7ea5feae5c9,East,Accessories,Woman Lite,10,460.22,4602.2,3722.6,879.6,Returning
23	216c54ab-0d3b-4493-b445-67226bc839ef,2024-11-25,4d844563-fef3-49ed-bd26-25f9a70ab0a8,North,Personal Gadgets,Article Ultra,6,387.86,2327.16,1567.8,759.36,Returning
24	9e7f9275-500a-4c11-88db-1574060a9bd0,2024-12-25,2174e041-b159-498f-85a5-28f76b293000,Central,Accessories,Least Lite,2,74.4,148.8,111.12,37.68,New
25	161a207c-1273-4bda-977e-84b83b71b995,2024-01-21,f0adf7be-620e-46cc-a2f5-255435fd7697,West,Computers,Artist Lite,1,118.72,118.72,89.72,29.0,New
26	135c45cf-d72c-42e7-9da8-cd1e64a14483,2024-10-31,ddadba99-938c-4beb-af41-a708198eedd0,West,Accessories,Single Lite,5,79.38,396.9,292.05,104.85,Returning

Figure 1 - Initial Data Set Before Manipulation and Cleaning

The data was then separated by the comma delimiter and the first row was used as the table headers. This was done to achieve a table structure of the data. As show in the image below.



Order ID	Date	Customer ID	Region	Product Category	Product Name	Quantity	Unit Price	Total Sales	Total Cost	Profit	Customer Type
a4fd5a0b-660a-4be1-b20d-e408ce68fafe	2024/06/07	a2a6a4d9-abbe-4de8-94a2-aa40892da57e	West	Accessories	Similar X	1	141.32	141.32	105.26	36.06	Returning
025fb9a7-ea6b-4193-87dd-605b5d137c72	2024/10/02	fe13634e-3f08-48aa-b8cb-1921932089ae	East	Home Appliances	Decide Ultra	10	484.16	4841.6	3612.9	1228.7	New
4c0ae2b4-5425-4793-9632-708de8c00db5	2024/09/04	9f2f6d02-7e17-4864-8a36-c0a14e450d4f	West	Personal Gadgets	Bad X	10	50.5	505.0	380.0	125.0	Returning
69560910-cddb-44ff-bf08-8b47dbf15848	2024/06/10	e20f2048-19fa-457f-800c-390722167747	North	Home Appliances	Community Pro	9	424.53	3820.77	3220.83	599.94	New
7260757c-e622-4d0f-98ee-a40f1fa5b0d9	2024/09/30	b430451c-d0fc-41a6-941f-d529050150d0	East	Accessories	Item Ultra	3	199.37	598.11	420.3	177.81	Returning
7e24a8c1-46a5-4801-9fe4-b547f12fa035	2024/04/05	87b58a66-3093-48f7-93b1-7d6deaf56c99	Central	Personal Gadgets	Thank Ultra	9	269.78	2428.02	1532.34	895.68	Returning
73086e8b-d455-4af4-a7e0-fb92e4a03ccb	2024/10/28	fb12c18c-bced-482f-be04-f10c57c8ad7e	East	Home Appliances	Always Ultra	3	187.52	562.56	412.71	149.85	Returning
24add46f-dad2-454f-a463-32f300675fc7	2024/04/25	6a6c8167-ad5e-4dc5-94be-3faba406283e	East	Personal Gadgets	Close Lite	7	419.83	2938.81	1935.22	1003.59	Returning
076a95c3-79aa-4246-bb6f-cb0a64db29bb	2024/05/23		south	Home Appliances	Add X	8	367.79	2942.32	1909.2	1033.12	Unknown
147132ac-85cb-434b-89ce-fd71e813d1ab	2024/03/02	13a7f0e3-c6e5-4199-a9f6-2d2c9851424e	South	Computers	Force Pro	1	167.74	167.74	106.62	61.12	Returning
36fa8db5-f18c-4e6a-a88f-ae721b90fac1	2024/06/15	4900222a-30cd-4abb-a2cd-06d316b75b6c	South	Electronics	Red Max	8	376.22	3009.76	2225.12	784.64	New
9fa02fb2-5061-451c-8275-a941bb40615f	2024/12/05	3ad93faf-4110-4db7-a08f-a9137582371c	Central	Electronics	Many Max	10	180.49	1804.9	1519.8	285.1	Returning
08bf37ae-a815-4e3d-9dd5-8f052f3bd537	2024/11/24	47a306ea-9093-427c-809b-617f5e3bd2cc	Central	Personal Gadgets	Condition Pro	6	132.19	793.14	649.38	143.76	Returning
fde4b9e1-6dd8-4480-b2c9-33ea2b26d934	2024/01/30		West	Accessories	Watch Ultra	1	70.03	70.03	42.83	27.2	Unknown
956bc0cf-44cc-4916-bb9b-d4a6b24bbee9	2024/04/08	585ef995-26e9-407a-ae1a-9a38aa4cca9a	Central	Personal Gadgets	Meet Max	4	470.37	1881.48	1523.96	357.52	Returning
0db29152-c16f-4bec-ab7a-c5a66adc38e8	2024/10/27	0cab9d7a-5e66-4360-bc95-ac173fe70ca8	North	Home Appliances	Air Lite	4	294.62	1178.48	824.16	354.32	Returning
742dea3b-82c8-4917-8295-7cf6f9ecc918	2024/12/03	3c6e8593-d912-4426-ba8e-f9401f4f4016	North	Computers	Past Ultra	3	78.45	235.35	204.99	30.36	New
ac8ffa9a-83fc-4c0d-bc96-de664372ef12	2024/07/15	72461bf9-2065-4b59-a68e-bf8ad22107cb	North	Computers	Ok Pro	9	137.74	1239.66	929.07	310.59	New
b3c0b061-144c-4e5f-86f2-86a262e5cb9c	2024/11/04	6f873926-6e59-4845-aaf4-e9a7059b572a	North	Computers	Establish Max	3	56.67	170.01	145.26	24.75	Returning
16f28e8a-a587-44c1-8c0b-6aeea0dc96e7	2024/01/02	0eb91e62-2552-4c29-b85f-9233c655048e	Central	Computers	Eight Lite	1	391.91	391.91	240.04	151.87	Returning
a20d38b9-d631-4db2-8d8f-f23349158c9e	2024/01/02	4f213ceb-f854-4391-bcbd-b7ea5feae5c9	East	Accessories	Woman Lite	10	460.22	4602.2	3722.6	879.6	Returning
216c54ab-0d3b-4493-b445-67226bc839ef	2024/11/25	4d844563-fef3-49ed-bd26-25f9a70ab0a8	North	Personal Gadgets	Article Ultra	6	387.86	2327.16	1567.8	759.36	Returning
9e7f9275-500a-4c11-88db-1574060a9bd0	2024/12/25	2174e041-b159-498f-85a5-28f76b293000	Central	Accessories	Least Lite	2	74.4	148.8	111.12	37.68	New
161a207c-1273-4bda-977e-84b83b71b995	2024/01/21	f0adf7be-620e-46cc-a2f5-255435fd7697	West	Computers	Artist Lite	1	118.72	118.72	89.72	29.0	New
135c45cf-d72c-42e7-9da8-cd1e64a14483	2024/10/31	ddadba99-938c-4beb-af41-a708198eedd0	West	Accessories	Single Lite	5	79.38	396.9	292.05	104.85	Returning
d4858939-1c96-4f9e-9698-81e847cb41d0	2024/07/13	5f4d13fb-92eb-46bb-b4fb-c7540740b2cb	East	Electronics	White Ultra	1	469.5	469.5	346.06	123.44	New

Figure 2 - Data Set Transformed Into Table Format

The data was then cleaned to get rid of data capturing inconsistencies, checks for null values and removal of any columns that would not prove useful in the analysis and visualisation process. Below are some examples of data inconsistencies.

Region
West
East
West
North
East
Central
East
East
south
South
South
Central
Central
West

Figure 3 - Image showing inconsistent data entry

Customer ID
a2a6a4d9-abbe-4de8-94a2-aa40892da57e
fe13634e-3f08-48aa-b8cb-1921932089ae
9f2f6d02-7e17-4864-8a36-c0a14e450d4f
e20f2048-19fa-457f-800c-390722167747
b430451c-d0fc-41a6-941f-d529050150d0
87b58a66-3093-48f7-93b1-7d6deaf56c99
fb12c18c-bced-482f-be04-f10c57c8ad7e
6a6c8167-ad5e-4dc5-94be-3faba406283e
13a7f0e3-c6e5-4199-a9f6-2d2c9851424e
4900222a-30cd-4abb-a2cd-06d316b75b6c
3ad93faf-4110-4db7-a08f-a9137582371c
47a306ea-9093-427c-809b-617f5e3bd2cc

Figure 4 – Image showing missing values

I had then fixed the data entry issues and any inconsistencies. Below showcases how I formatted the Region column

```
=PROPER(TRIM([@Region]))
```

Figure 5 - Formula used to format the Region column

D	E
Column1	**Region**
Central	Central
Central	Central
Central	Central
East	East
West	West
South	South
South	South
North	North
East	East
North	North
North	North
North	North
South	south
Central	Central
North	North
South	South

Figure 6 - The formula is applied to "Column1" to ensure we still have the original data we are working with should our formatting be erroneous

Region
Central
Central
Central
East
West
South
South
North
East
North
North
North
South
Central
North
South
North
South

Figure 7 - The final fixed Region column

Findings

I found that a total of 31 records had missing/blank Customer IDs. They all also had “Unknown” as a Customer Type. This could indicate a few things:

1. These are customers who do not have their information saved on the system and thus their purchases could not be attached to or mapped to a profile.
2. There was an error capturing the customers information and there was trouble finding the correct customer to map the order to so it was decided that the order will not belong to a specific customer

The data set also contained 4 records with missing product names. This does not seem to be a trend in a particular region because there is an instance at each Region. The only possible cause that could possibly be is that the person processing forgot to note down the Product Name down because the Product Categories are listed, this may be an easy fix as we may able to find the product name using the Product Category and Unit Price to a product. Upon further investigation I was unsuccessfully unable to map these Product Names due to the nature and contents of the data set.

Due to the missing content, we may not be able to obtain a true observation or depiction of the most popular products however we will still get a good idea of the leading products.

Data Visualisation

The data will be visualised using PowerBI.

In terms of the data used, it was decided to keep the entries with missing CustomerIDs and Product Names because they still hold valuable information pertaining to sales, profit made by region and some by product name (i.e. the entries that have missing CustomerIDs but have all other data needed). The working set was also transferred to a new workbook to ensure that no data from the original, unformatted dataset was accidentally tampered with or used in the analysis. The only transforming done in PowerBI was to remove duplicate data and ensuring all data types were in the correct.

Task

As mentioned earlier in the document, these are the following KPIs that will be visualised:

- **Total Sales Performance:** Show total sales revenue, total profit, and number of transactions.
- **Sales by Region:** Which regions performed the best in terms of revenue and profit?
- **Sales by Category:** Which product categories were the most and least profitable?
- **Monthly Sales Trend:** Identify sales trends over the months (e.g., peak seasons, dips).
- **Customer Insights:** Breakdown of returning vs. new customers.
- **Top 5 & Bottom 5 Products:** Based on revenue.

Basic Depiction of the KPIs to be visualised

Total Sales Performance

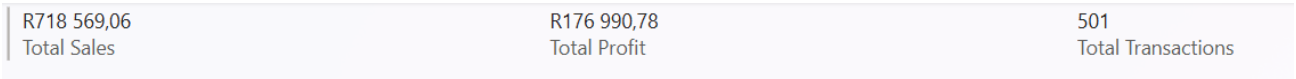


Figure 8 - Total Sales Performance

Sales by Region

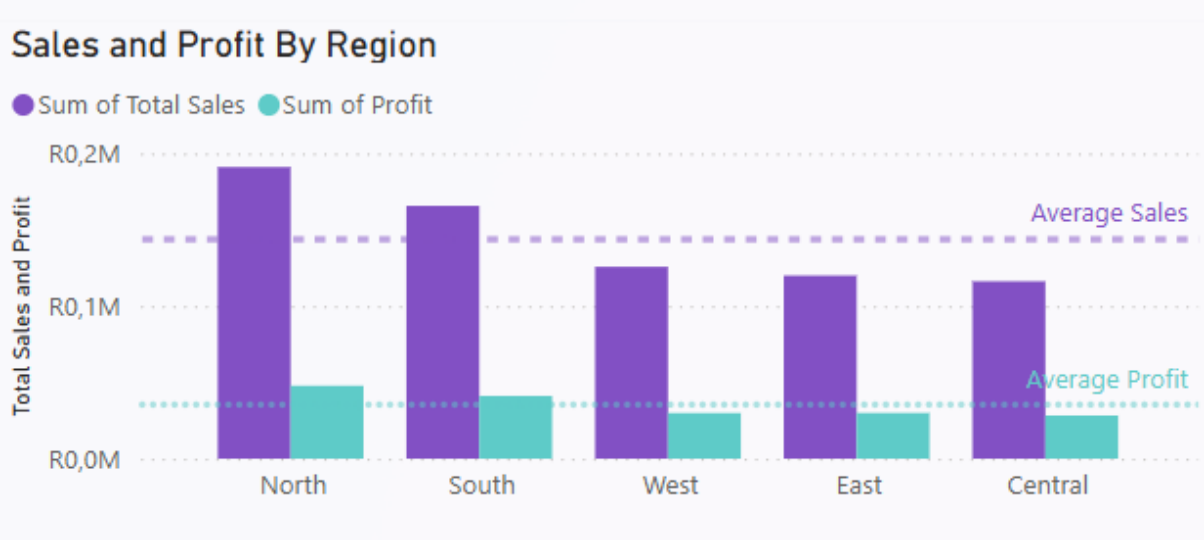


Figure 9 - Sales by Region

Sales by Category

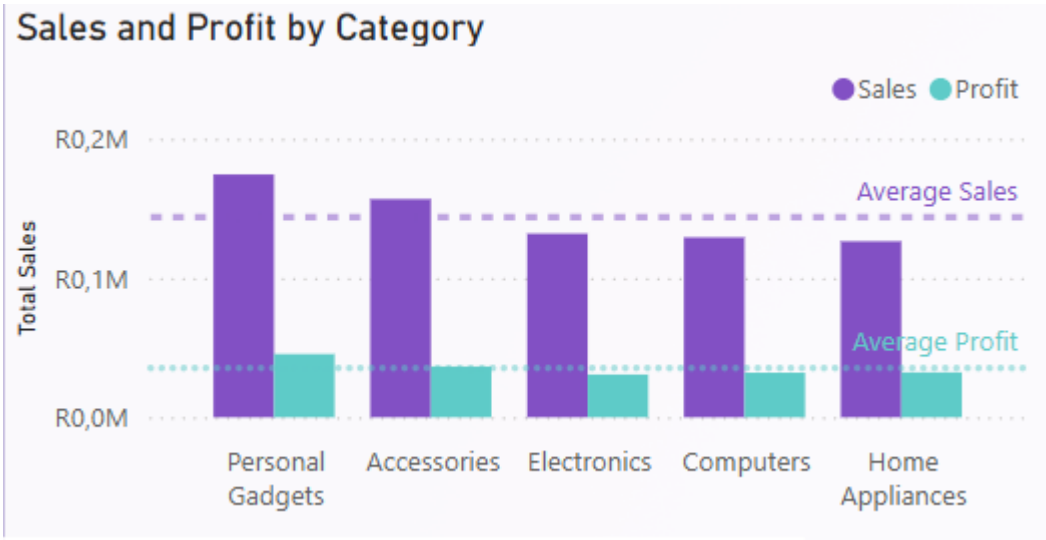


Figure 10 - Sales by Category

Monthly Sales Trend

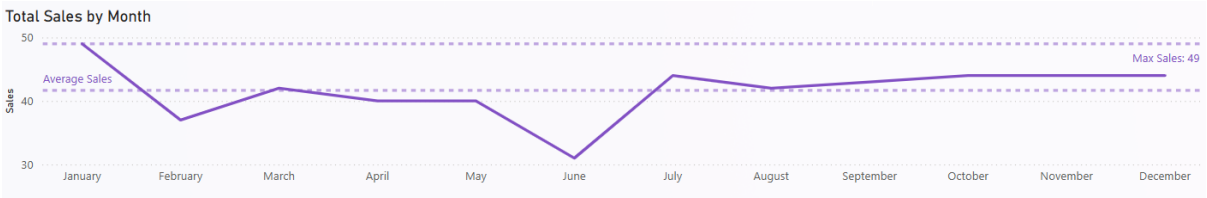


Figure 11 - Monthly Sales Trend

Customer Insights

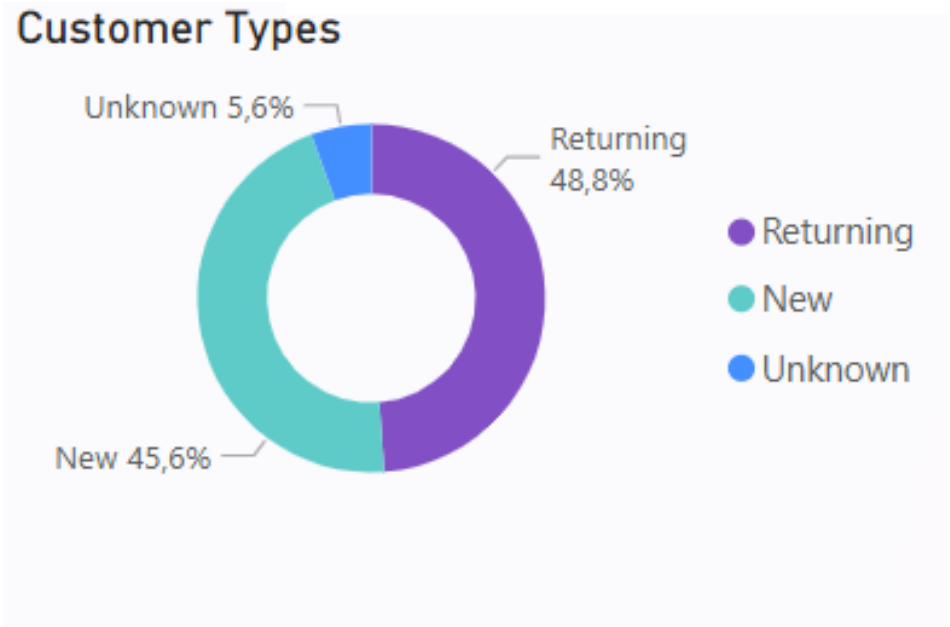


Figure 12 - Customer Type Insights

Top 5 & Bottom 5 Products

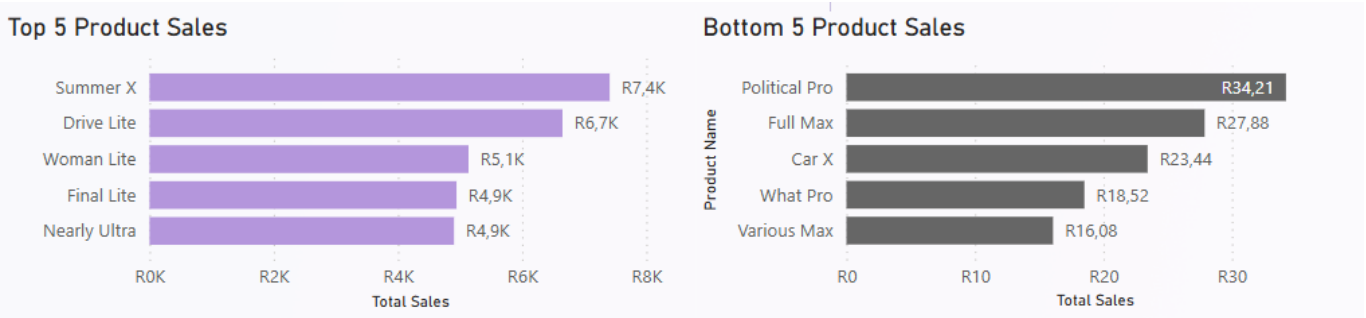
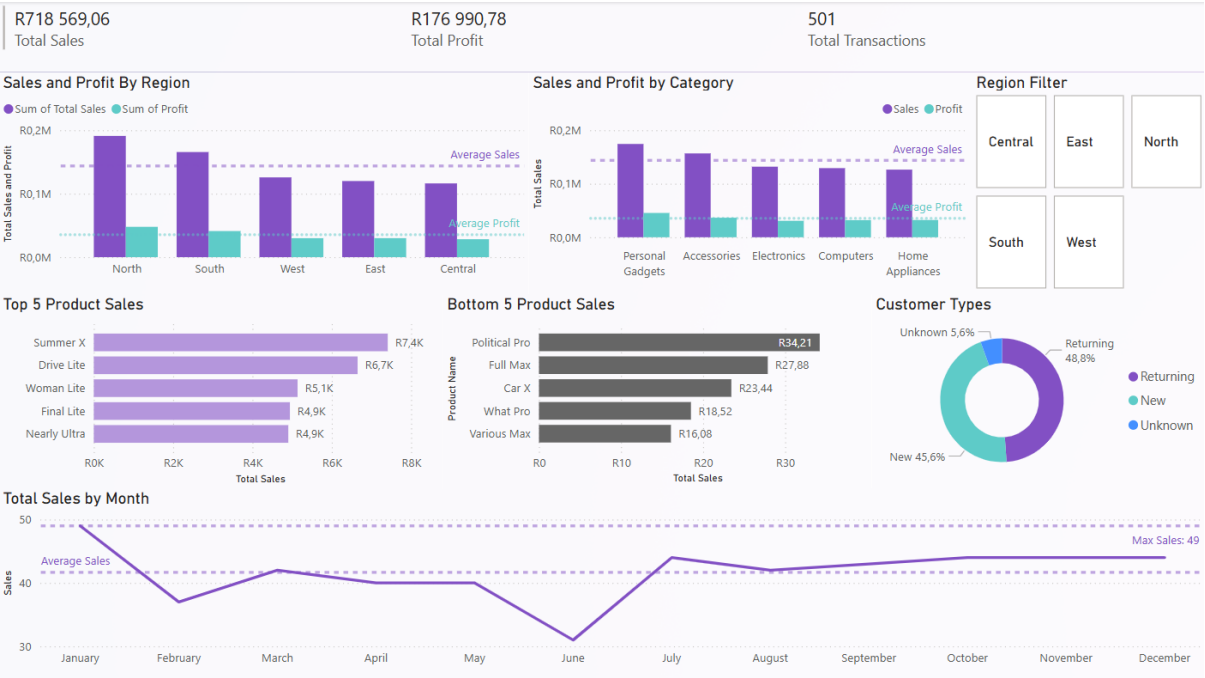


Figure 13 - Top 5 & Bottom 5 Products



Final Dashboard

Figure 14 - Final Dashboard

Findings

Having completed the dashboard, the findings can now be communicated.

Findings before filtering by regions:

- North and South are the two regions with the highest sales and highest profit made
- Overall Personal Gadgets and Accessories are the most purchased product categories with the highest sales and profit figures
- Overall Various Max was the worst performing product and Summer X was the bestselling product.
- Across the whole year, January is the month with the most sales followed by August with the lowest months being December and then May
- Majority of customers are Returning which indicates good customer retention, there is also have a good number of new customers

Findings upon filtering by regions:

- All regions have a good number of returning customers and new customers, the differences are not too glaring
- The Top 5 and Bottom 5 products are extremely different for each region and thus there is no real trend that could be seen in that regard and a further investigation into region specific purchases would be a good idea
- Personal Gadgets seems to be performing well across the regions, all other categories vary quite a lot
- Central saw most sales in February and September and least sales in August, October and December
- East saw most sales in October then July and the least sales in February and May
- North saw the most sales in January and November and the least sales in April and December
- South saw the most sales in December and August and the least sales in February and October
- West saw the most sales in March and December and the least sales in February and November
- The region that made the most sales was North and the Region with the least Sales was Central

Conclusion and Reflection

This was such a fun experience. It was such a learning curve especially because I total control over the chart types used and how the data was communicated. I believe that for my first chart that I have created I did a good job. I could brush up more on the way I use the columns and statistical methods to summarise data.

I believe that the dashboard does communicate the findings well, there could be a lot more finesse especially in the overall colour scheme and how data values are displayed but I am confident that it gets the message across.