

Retail Sales Analysis Report Process

Mbali Nkosi



February 18, 2025

# Table of Contents

Contents

[Table of Contents 1](#_Toc190779663)

[Table of Figures 2](#_Toc190779664)

[Case Study 3](#_Toc190779665)

[Tasks 3](#_Toc190779666)

[Planning And Process 4](#_Toc190779667)

[Tools 4](#_Toc190779668)

[Process 4](#_Toc190779669)

[Findings 7](#_Toc190779670)

[Data Visualisation 8](#_Toc190779671)

[Task 8](#_Toc190779672)

[Basic Depiction of the KPIs to be visualised 8](#_Toc190779673)

[Total Sales Performance 8](#_Toc190779674)

[Sales by Region 9](#_Toc190779675)

[Sales by Category 9](#_Toc190779676)

[Monthly Sales Trend 10](#_Toc190779677)

[Customer Insights 10](#_Toc190779678)

[Top 5 & Bottom 5 Products 11](#_Toc190779679)

[Final Dashboard 11](#_Toc190779680)

[Findings 12](#_Toc190779681)

[Conclusion and Reflection 13](#_Toc190779682)

# Table of Figures

[Figure 1 - Initial Data Set Before Manipulation and Cleaning 4](#_Toc190778763)

[Figure 2 - Data Set Transformed Into Table Format 4](#_Toc190778764)

[Figure 3 - Image showing inconsistent data entry 5](#_Toc190778765)

[Figure 4 – Image showing missing values 5](#_Toc190778766)

[Figure 5 - Formula used to format the Region column 6](#_Toc190778767)

[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](#_Toc190778768)

[Figure 7 - The final fixed Region column 6](#_Toc190778769)

[Figure 8 - Total Sales Performance 8](#_Toc190778770)

[Figure 9 - Sales by Region 9](#_Toc190778771)

[Figure 10 - Sales by Category 9](#_Toc190778772)

[Figure 11 - Monthly Sales Trend 10](#_Toc190778773)

[Figure 12 - Customer Type Insights 10](#_Toc190778774)

[Figure 13 - Top 5 & Bottom 5 Products 11](#_Toc190778775)

[Figure 14 - Final Dashboard 11](#_Toc190778776)

# 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

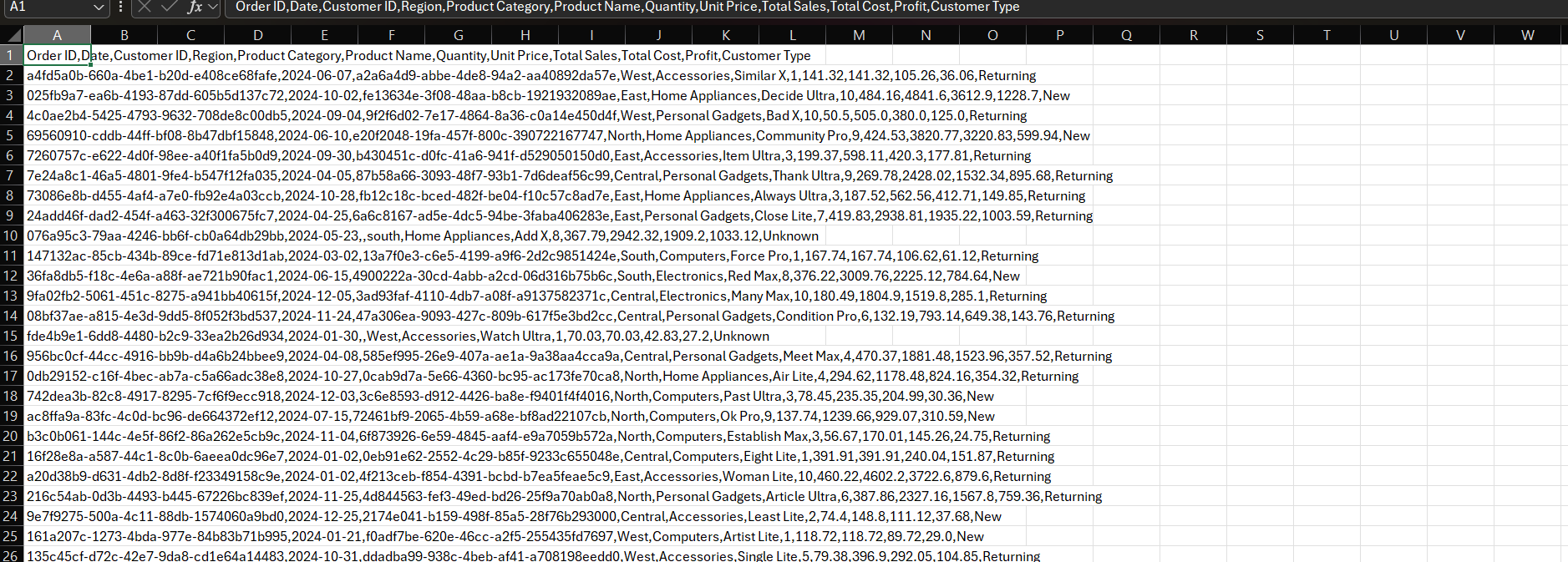
I asked ChatGPT to generate a CSV file for me that contains some unclean data for me to clean and prepare the data for analysis and visualisation. This is the initial format of the CSV file before any data manipulation and cleaning has taken place.

Figure - Initial Data Set Before Manipulation and Cleaning

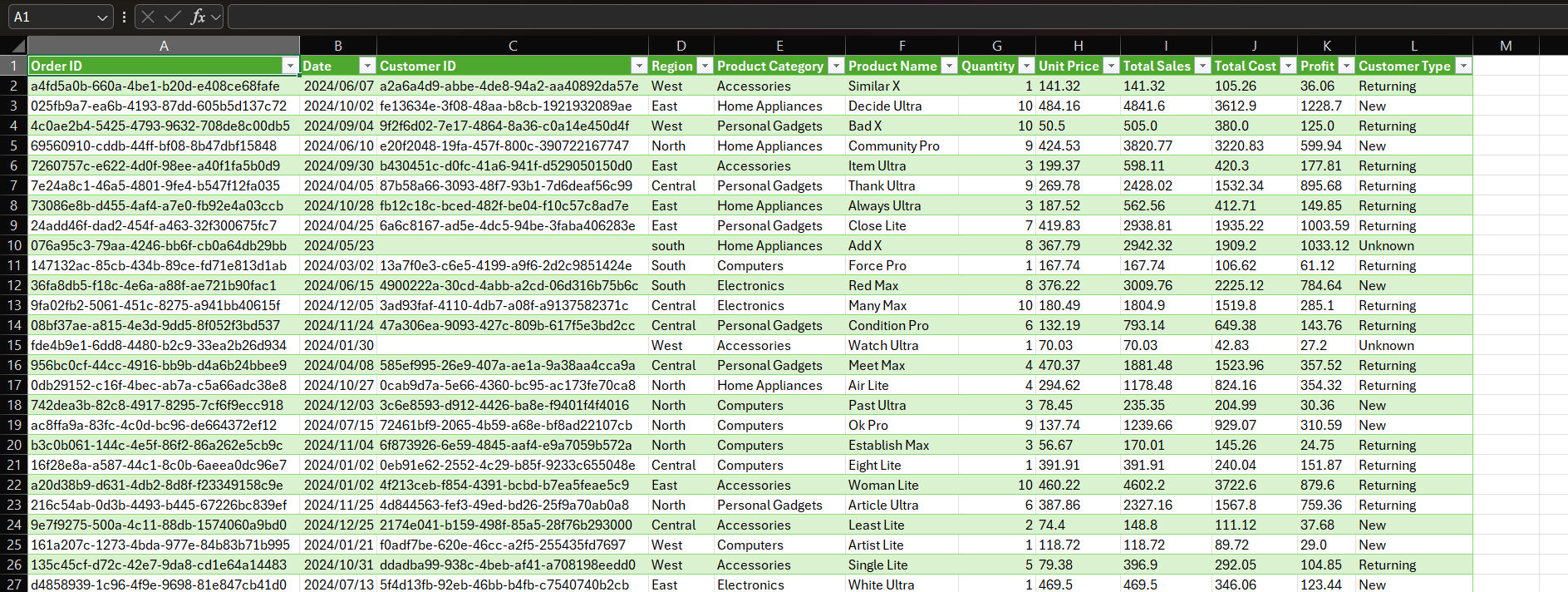
The data was then separated by the comma delimiter and the first row was used as the table headers. This was done to achieve a table structure of the data. As show in the image below.

Figure - 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.

A screenshot of a phone

AI-generated content may be incorrect.

Figure - Image showing inconsistent data entry

A screenshot of a computer

AI-generated content may be incorrect.

Figure – Image showing missing values

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



Figure - Formula used to format the Region column

A screenshot of a computer

AI-generated content may be incorrect.

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

A screenshot of a phone

AI-generated content may be incorrect.

Figure - 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 - Total Sales Performance

### Sales by Region

A graph with different colored bars

AI-generated content may be incorrect.

Figure - Sales by Region

### Sales by CategoryA graph of different colored squares AI-generated content may be incorrect.

Figure - Sales by Category

### Monthly Sales Trend A line graph with purple lines AI-generated content may be incorrect.

Figure - Monthly Sales Trend

### Customer Insights

A chart with numbers and a circle

AI-generated content may be incorrect.

Figure - Customer Type Insights

### Top 5 & Bottom 5 Products

A screenshot of a graph

AI-generated content may be incorrect.

Figure - Top 5 & Bottom 5 Products

## A screenshot of a computer AI-generated content may be incorrect.Final Dashboard

Figure - 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.