BIG DATA AND BUSINESS INTELLIGENCE

INCOURSE ASSESSMENT

REG SUPERSTORE RETAIL ANALYSIS

ABISUGA, REGINA (C2000943)

SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL TECHNOLOGIES

TEESSIDE UNIVERSITY

Table of Contents

BUSINESS REPORT

EXECUTIVE SUMMARY

This paper comprehensively analyses the retail and revenue performance of REG superstore for the period January 2011 to December 2014. The report's primary aim is to assess revenue margins and provide key business suggestions to Executive Management, Board of Directors, various key departments such as Marketing and Communications, Customer Experience management, Sales and Logistics; to boost turnover whilst minimizing cost as this might have an influence on the company's net income in the future.

The report speaks to the following:

- To track KPIs (revenue, turnover, profit) during the four years under evaluation “2011-2014”.
- To provide a detailed customer order profile for loyalty and reward packages to the marketing and Customer Experience units.
- Identifying high-value customer.
- To examine at the best-performing countries and market areas.
- To analyze product-level trends and forecasts

A. BODY

Introduction: Super Store is a global retail company headquartered in the United Kingdom. They sell Furniture, Office Supplies, and Technology items to individuals, corporations, and home offices.

Our objective is to analyse sales data and discover Superstore's strengths, weaknesses, threats, and opportunities for company growth while optimizing profit and minimising costs. The data collection comprises REG Superstore's Sales, Profit, Order, Customer, Product, and Geographic information, as well as data on customer order specifics such as state, area, order date, shipment date, product ordered, and so on. We would develop business questions that would decide what data the dataset might represent. These are the kinds of inquiries we can ask the dataset and get responses to.

Data Source and Selection:

This project's data set was obtained via Kaggle (Global Superstore | Kaggle). It collects retail data from a REG superstore from 2011 – 2014. The Global Superstore would be known to as **REG Superstores** for the purpose of the present study.

As a global outlet, it controls supply of 3 product types to key markets worldwide: technology, furniture, and office supplies. When a consumer puts an order, it is prioritised and prepared for delivery. This helps our business with its customer experience management processes.

The dataset's data structure comprises critical variables for my BI project.

Table & Columns Definition

The data collection has 24 attributes, which are outlined below in the table.

Columns	Description
Row ID	The distinct row identification found in the database
Order ID	The unique number that is generated when an order is placed for a product
Order Date	The date the order is placed
Ship Date	Date product ordered was shipped for onward delivery to the customer
Ship Mode	Mode of shipment used
Customer ID	Customer Identification Number
Customer Name	Full name of customers
Segment	Product segmentation
City	The city where the customers live
State	The state where the cities are situated
Country	The country where these states are based
Postal Code	The postal code of the consumer
Market	Market segmentation based on product sales regions
Region	The geographical region in which the products are marketed
Product ID	Product identification number
Category	The product category
Sub-Category	Sub-Category of products
Product Name	Names of products sold
Sales	The price at which the products are sold
Quantity	The quantity of units sold
Discount	The amount discounted on product's sales.
Profit	The profit generated from a product's sales.
Shipping Cost	The cost of shipping
Order Priority	Holds order categorization depending on its urgency

BI Requirements\Questions:

To thrive as a global retail giant, it is critical to monitor sales performance on periodical basis utilising all relevant indices and performing internal analysis using the SWOT analysis (Strength, Weaknesses, Opportunities and Threats). Top players in this sectors, such as Argos, Sainsbury and Amazon, always take this into account in order to retain a competitive advantage. Power BI will be utilised in this project to present relevant stakeholders in the organization: Executive Management, Board of Directors, Key departments such as Marketing and Communications, Customer Experience management, Sales and Logistics; with a high level retail analysis report for the last four years. The following questions or points are highlighted in the analysis.

- To analyse Revenue margins against Computed Target.
- To review Top products by orders across various market
- To compare regional performance
- To evaluate major influencers and their effect on turnover and profit
- To analyze product-level trends and forecasts

Finding Based on Analysis and Evaluation:

Based on the analysis, the following results were compiled in relation to the Business Intelligence questions posed. Various visualisations were created utilising the existing BI visual tools, as well as importing new ones where appropriate.

- Staples drove the most orders over the 4-year period, at the same time Apple Smart phone drove the most revenue, it drove about 87,000 in revenue.
- December 2014 witnessed a product order surge of with a total month order of 1057 compared to the previous month, which had 1077. There was a decline by 1.86%.
- In comparison to the prior month, revenue/sales increased in December 2014, with a total month revenue of £503,143, which had a £555,279. There was a decline by 9.39%
- Top performing category by Total orders are Office supplies.
- The Central and US markets are the highest performing regions and markets, respectively.
- By the end of 2014, REG Superstore had handled a total of 25,035 orders, representing a 463.9% increase over 2011..
- Global superstores had a total of 1309 consumers in 2011 and 1590 in 2014, which is a 21.5% increase.
- Also, the yearly profit grew by 457.5% closing 2014 with £12.6 million from £2.26M in 2011.

The key findings from the charts must be discussed. The snapshots below give a glimpse into the above-mentioned results and findings.

Retail Analysis Dashboard

- The Key Influencer graph was utilised to obtain perspective into parameters that have an impact on Total profit over time. A significant influencer is the Category, Subcategory, and Product name, which demonstrated metrics that would lead to a reduction in profit.
- **Yearly Profit Outlook**
  After analysing sales/revenue performance, a yearly profit prediction is required. The projection was made after 2014, the previous year under consideration. According to the graph, sales are predicted to plummet by the end of 2015. It does, however, provide upper and lower bounds to account for any ambiguity.
- **Total Profit**
  Total profit was calculated using a sum function.
- **Regional Performance**
  For analysis charts were created to evaluate the Total orders, Income earned across several countries in their Region. Central area outperformed the other product segments throughout the four years under consideration.
- **Average Retail Price**
  Average retail price was calculated using an average function.
- **Gauge Charts**
  Gauge charts were used to check Monthly order and revenue performance. Order Target and Revenue Target were computed using power BI DAX.
- **Sales Growth**

We saw the growth in Sales over the period of years from 2011 to 2014, with 2014 having the most sales signifying increase in turnover for the company.

- **Shipping Cost By Ship Mode**

  Also we must significantly reduce the time it takes from order placement to shipment. This is critical for meeting the customer retention objective while being dynamic and competitive. The shipment cost must be scrutinised in order to get the bare minimum while maintaining delivery criteria. The crucial order priority has the highest transportation cost, according to the images. The waterfall chart was used for this evaluation.

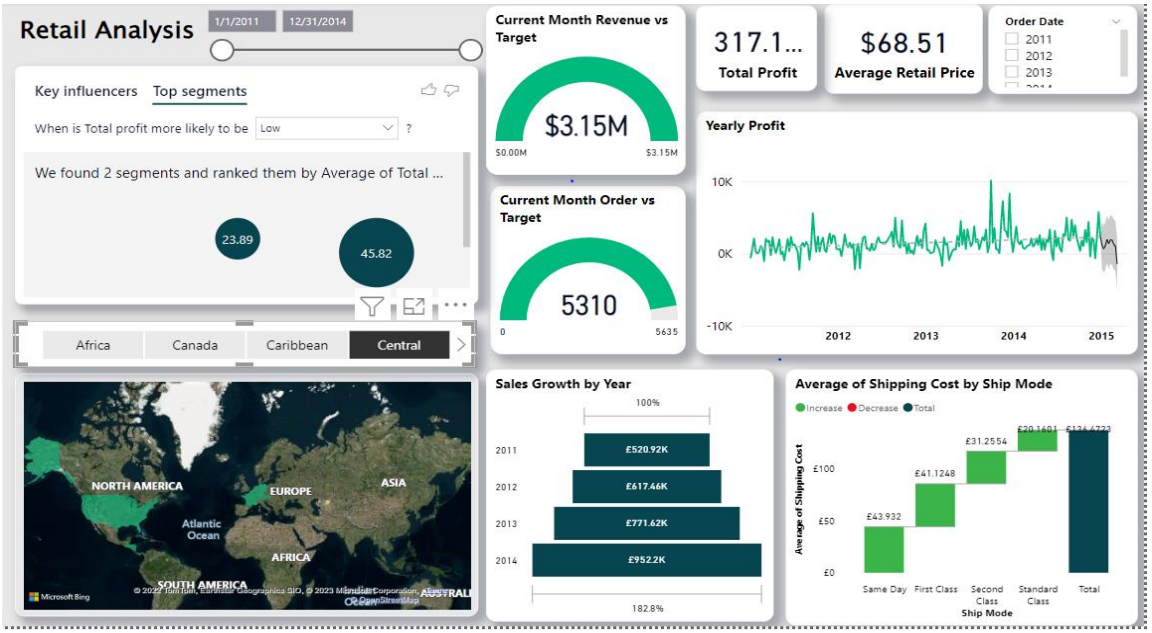


Fig. 1. Retail Analysis Dashboard (Central Region Performance)

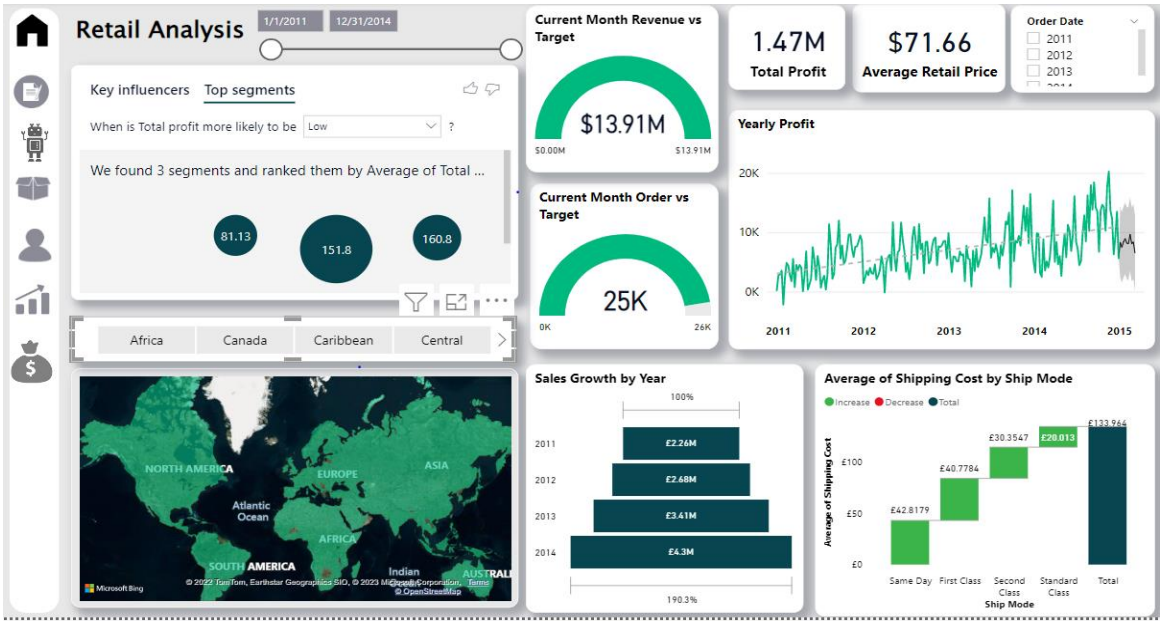


Fig. 2. Retail Analysis Dashboard (Overall Performance)

Artificial Intelligence (AI) Analysis

- **Decomposition Tree**

  This AI graphic was utilised to examine the Total Profit created in order to appraise the

Customer Orders Category , Segment by Country as shown in the below screenshot. We can see that the Top Profit generating Customer is Tamara Chand.

- **Drilldown Radial Chart**
  This animated chart was employed to further explore the Total Orders produced in order to determine the most significant contributor by Category , Segment and SubCategory year by year with slicer provided.

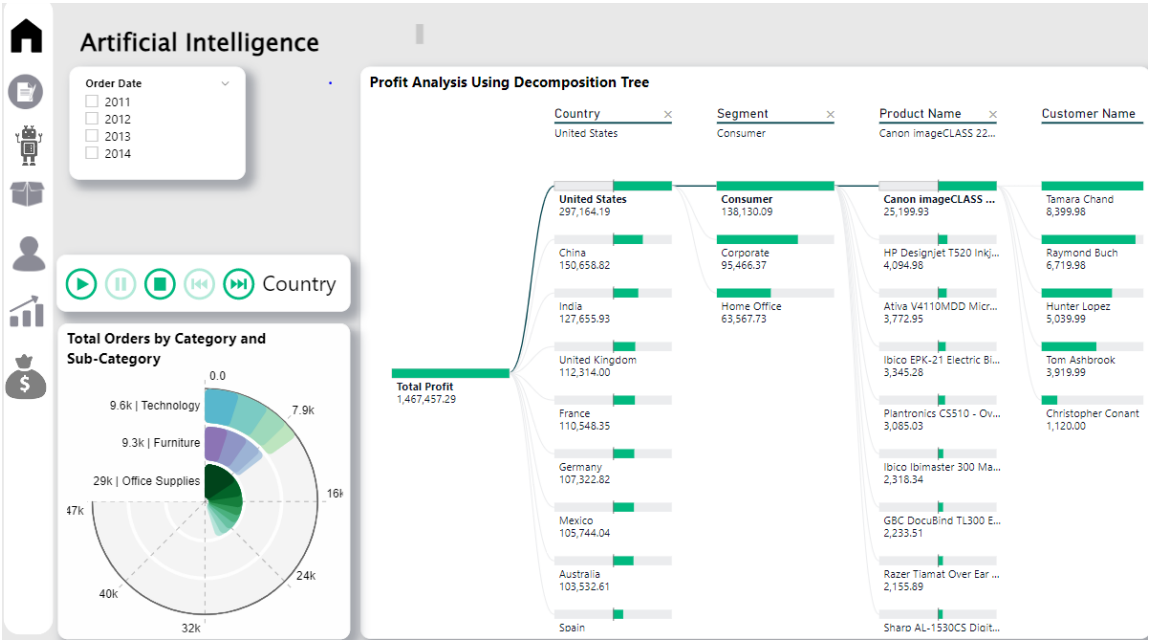


Fig. 3. AI Analysis

Product Detail Analysis

- KPI visuals were also created to compare Current Month Orders and Revenue and Previous Month Orders and Revenue. The Previous Month orders was were estimated using DAX. Previous Month orders = `CALCULATE([Total Orders], DATEADD(Calender_Table[Date],-1,MONTH))`.
- TreeMaps for Category to accurately measure the performance of the product categories of the retail store.
- Top Product - Card visual was used to visualize the Top product by orders. This shows the product performing most in terms of Total orders.
- Matrix visuals were used to evaluate the different products by Total revenue and orders.

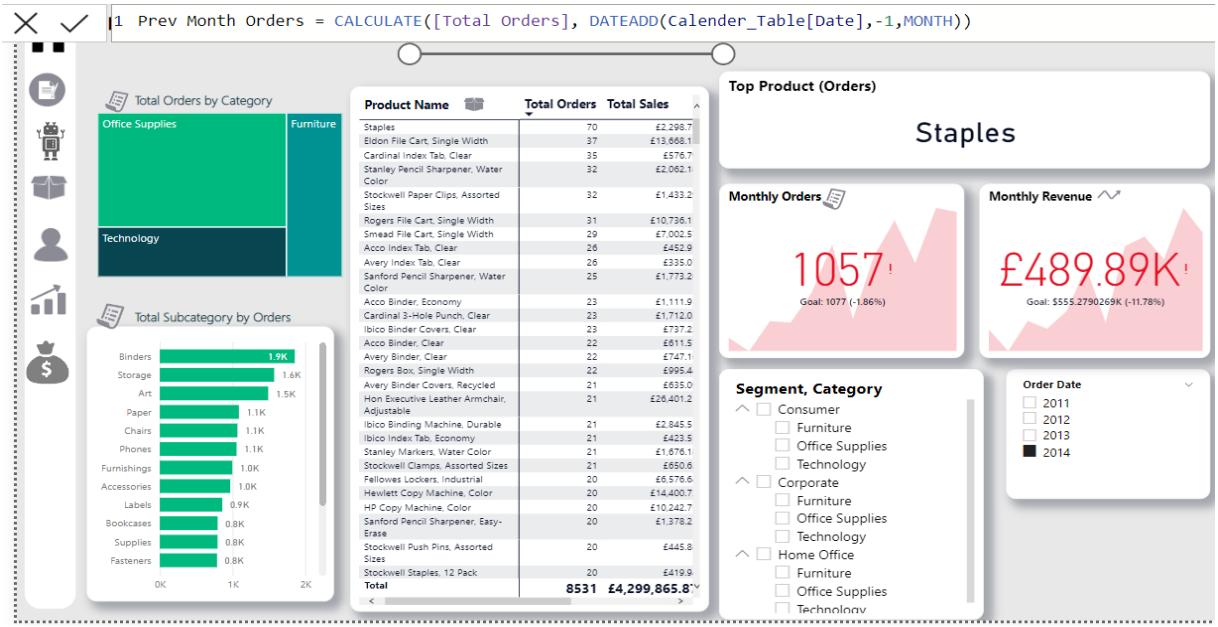


Fig. 4. Product Detail Analysis

Customer Detail Analysis

- Card visuals were used to show Total number of customers
- Matrix visuals were used to evaluate the Customers by Total orders, Revenue Target and Total/Actual Sales.
- The Clustered column chart was used to analyse the customer distribution by different locations, where the United States showed the highest count (846) of Customers.

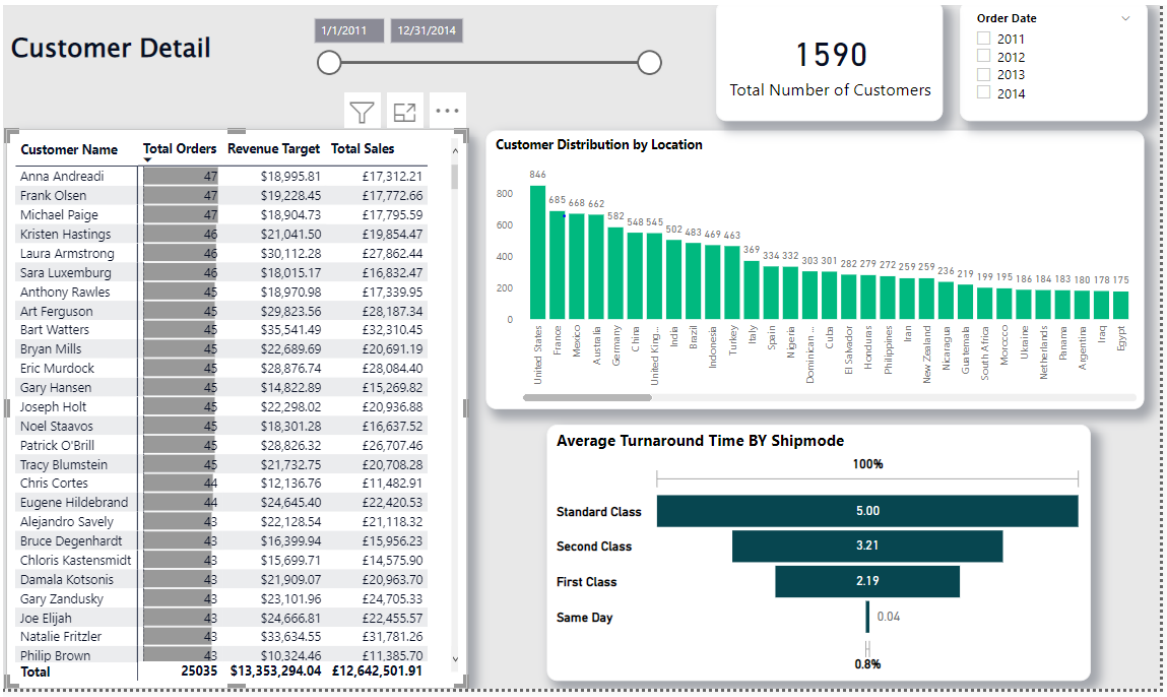


Fig. 5. Product Detail Analysis

Revenue/Sales Analysis

- **Animated Bar and Donut Charts**

These visualisations depicts how sales/revenue performance has changed over time for a number of different product groups and categories. Phones had a strong sales performance in 2014 owing mostly to the proliferation of technology items, which made phones a must-have item for purchasers.

- Card visuals were used to show Total Revenue Target, Total Sales and Total Quantity of Products.
- The Clustered bar chart was used to depict the revenue generated by percentage in different locations (Countries) with United States generating the highest sales in 2014 of 41.37% as it accounts for the country with the highest count of customers.

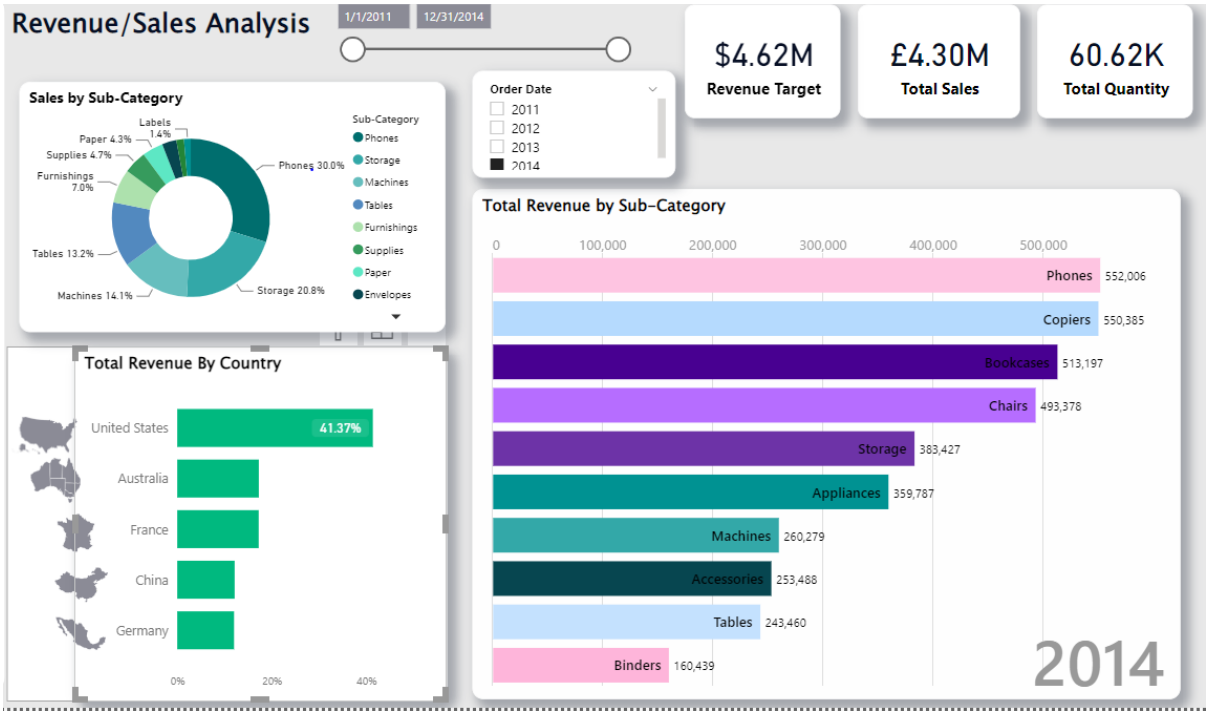


Fig. 6. Revenue/Sales Analysis for 2014

Profit Analysis

- **Infographics Designer – Profit by Category**
  This visualisation illustrates the profit acquired by REG Superstory in the different categories. Technology accounts for the top profit generating category as it has a high selling price thereby generating more profit for the business with the current surge for digital products.
- Card visuals were used to depict the Total Profit made and the Profit margin on Sales
  Based on the Profit Analysis, 2014 has had the highest profit and profit margin over the period in the screenshot below across all visuals.

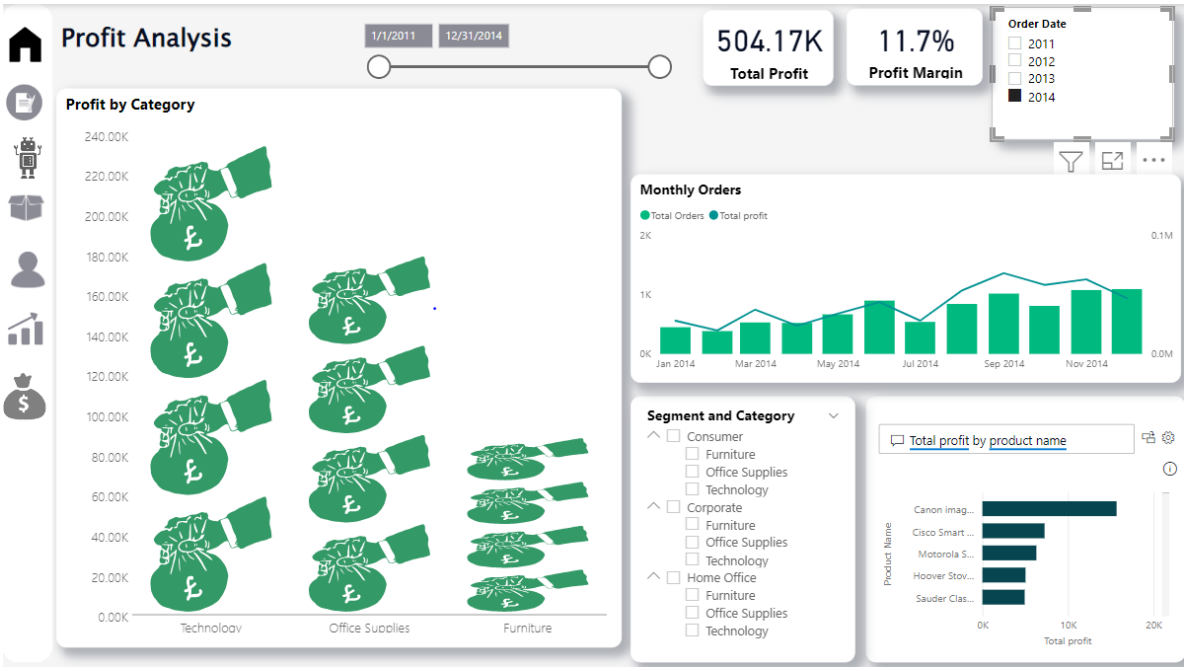


Fig. 7. Profit Analysis for 2014

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

REG Superstores has enormous potential for growth in a bid to restore its competitive advantage in an ever-changing retail market. To do this, focus must be placed on onboarding new consumers and boosting client retention. Prospective market re-entry methods to boost market penetration, visibility, awareness, and presence should be studied in sales regions with low sales turnover. Priority should be given to customer retention and onboarding. On that premise, marketing campaigns of all kinds must be undertaken in order to increase the quickly growing consumer base.

Recommendation

- Boost business visibility by utilising all marketing platforms to reflect the organisations' broader reach for onboarding new customers throughout off-peak and busy cycles.
- Use the product segment and subcategory to profile clients and customise items to their preferences.
- Optimize turnaround times for the various shipping priority services provided.
- Increase business activity and revenue through extending presence in areas such as Africa, The Caribbean, EMEA, and Canada.
- Start a loyalty and reward programme for existing customers who have placed more than 80 orders in the past four years.
- Routinely examine product unit costs to simulate current reality. Unit pricing is critical to sales and has an influence on profitability.

Appendices: BI Design

A. Data Pre-Processing or Data Cleansing

Data pre-processing basically refers to cleaning up the data set so that it is suitable and prepped for analytics. The data set was imported into Power BI , it had 24 attributes and 51,291 rows and before starting the analysis the data is to be cleansed and pre-processed using the following techniques outlined below:

- **Data Quality and Distribution for Removal of Null Values:** The data quality and distribution "View" tab displays inaccuracies within the dataset so that they may be easily removed. It also demonstrates how distinctive each piece of information in each column is, which is critical in the building of data models. The data set was scanned to find columns containing null cells. As a result, the Postal Code attribute was dropped. The column was superfluous because the other geographical identifiers sufficed. Row ID columns were also eliminated since they link to a unique identifier entry in a database and are thus unnecessary for this project.

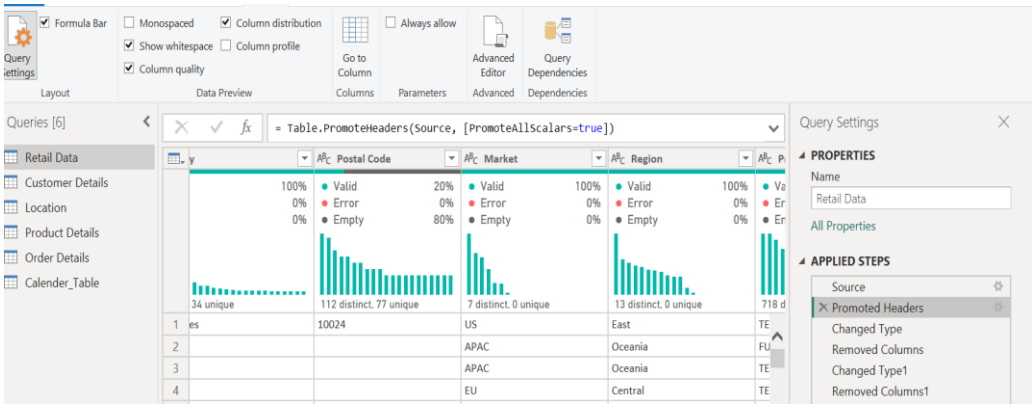


Fig. 8. Data Distribution and Quality

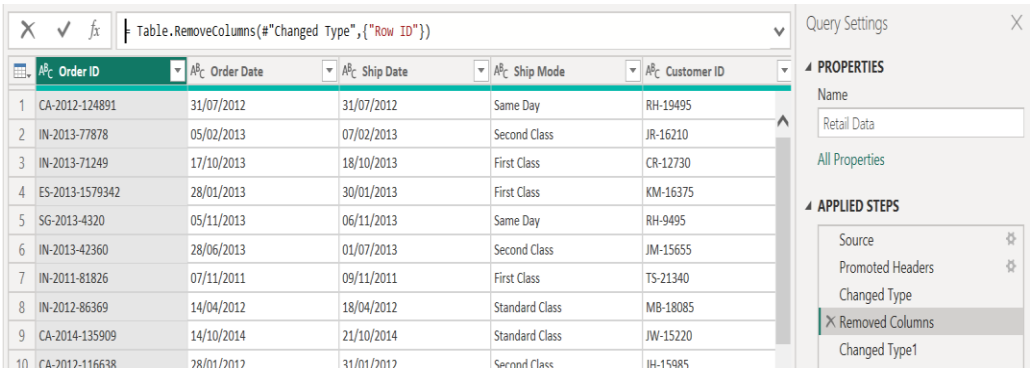


Fig. 9. Removal of Empty Columns

- **Promoted Headers:** The data set was not formatted to come with Headers, to adequately identify each attribute headrs were promoted.

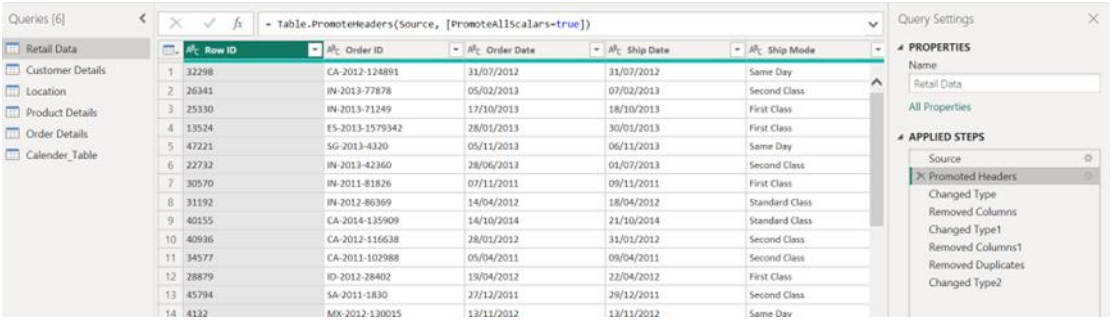


Fig. 10. Promotion of attribute Headers

- **Changing Data Types:** The data type for; Shipping Cost, Sales, Discount and Profit, was altered from decimal number to Currency and Quantity was change to numeric values since Power BI did not correctly identify it.

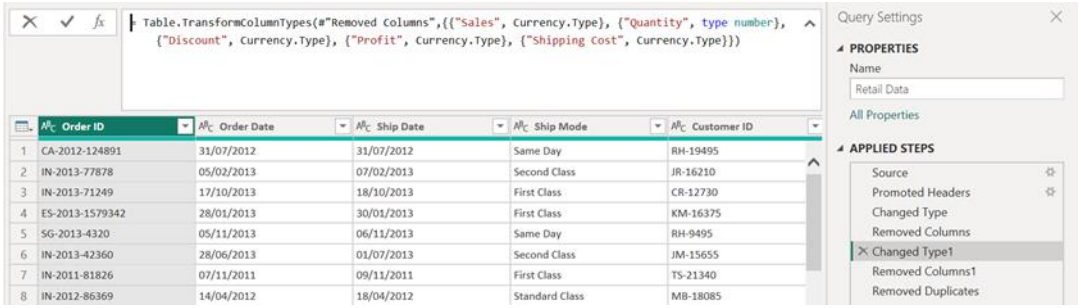


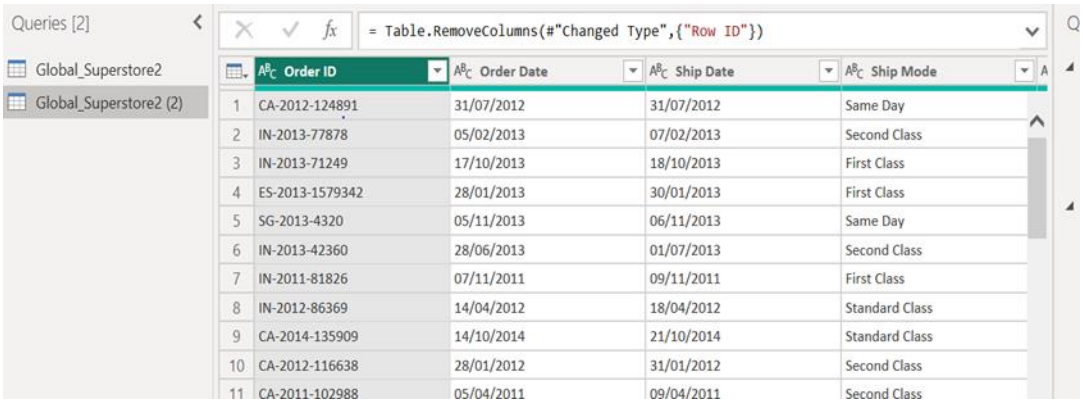
Fig. 11. Change Data types

After the above steps, click on the “close and apply” to put all of your changes into action.

B. Data Modeling using The Star Schema Model - Fact and Dimension Table

The process of splitting our data collection into fact and dimension tables in order to develop and interact effectively for producing meaningful graphics is known as data modelling. This is accomplished by replicating the GlobalSuperstore Data in the power query editor, renaming the GlobalSuperstore table, and removing unneeded columns as

well as duplicates.



	Order ID	Order Date	Ship Date	Ship Mode
1	CA-2012-124891	31/07/2012	31/07/2012	Same Day
2	IN-2013-77878	05/02/2013	07/02/2013	Second Class
3	IN-2013-71249	17/10/2013	18/10/2013	First Class
4	ES-2013-1579342	28/01/2013	30/01/2013	First Class
5	SG-2013-4320	05/11/2013	06/11/2013	Same Day
6	IN-2013-42360	28/06/2013	01/07/2013	Second Class
7	IN-2011-81826	07/11/2011	09/11/2011	First Class
8	IN-2012-86369	14/04/2012	18/04/2012	Standard Class
9	CA-2014-135909	14/10/2014	21/10/2014	Standard Class
10	CA-2012-116638	28/01/2012	31/01/2012	Second Class
11	CA-2011-102988	05/04/2011	09/04/2011	Second Class

Fig. 12. Duplicating the Global Superstore Table to craft out Dimensions Table

Fig. 13. Initial Global Superstore without any relationship

The Global Superstore Data generated the following dimension tables; Customer Details Table, Location Table, Order DetailsTable, Product Details Table were created while the Calender Table was formed using M Language as shown in the below snapshots.

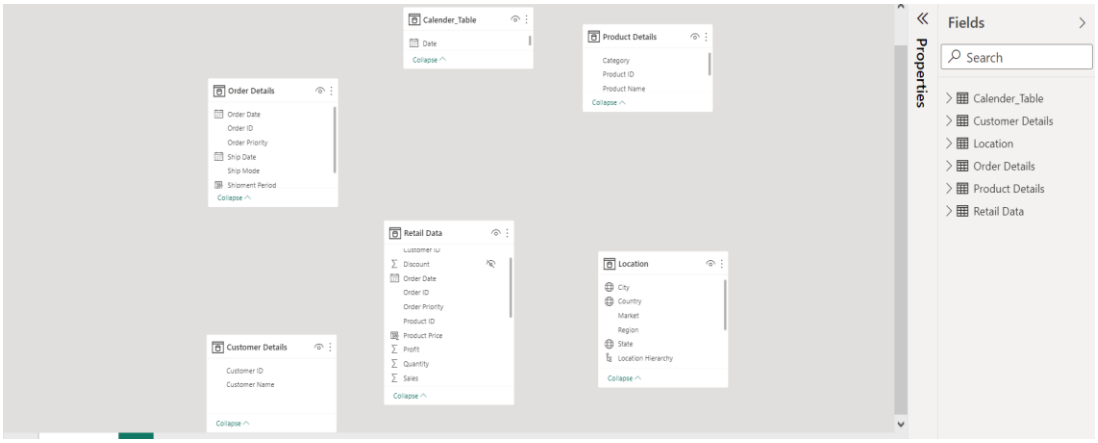


Fig. 14. PowerBI Model page without relationships

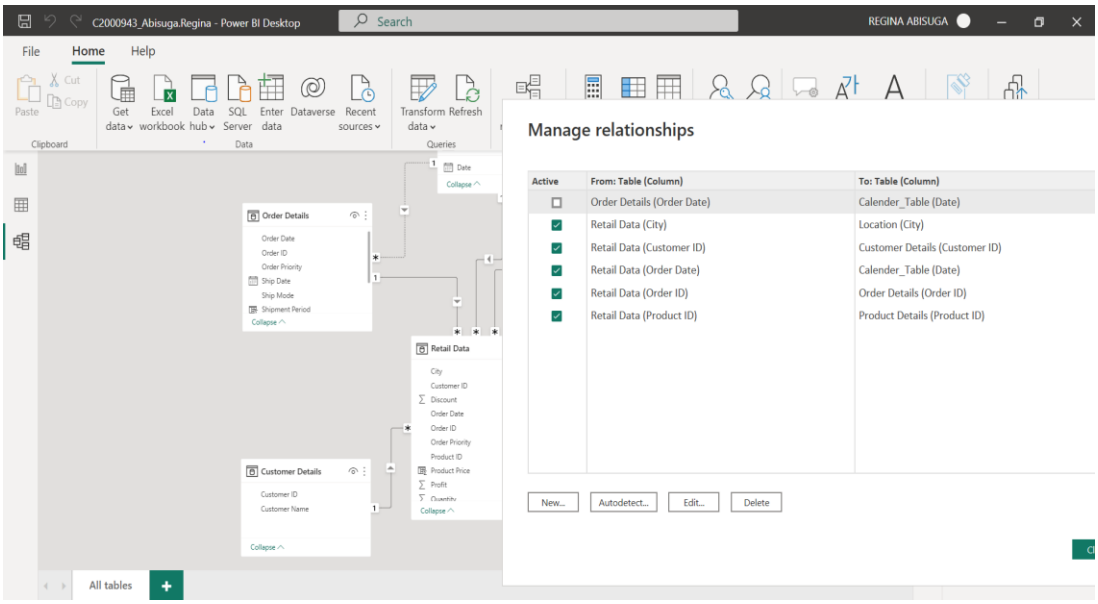


Fig. 15. Model page showing relationships between Facts and Dimension tables

The data model for the Global Superstore Dataset is shown below, with the Retail Data Table serving as the Fact table and the other tables serving as dimension tables. The Retail Data table is in the centre, linked to the dimension tables. The model is a star schema.

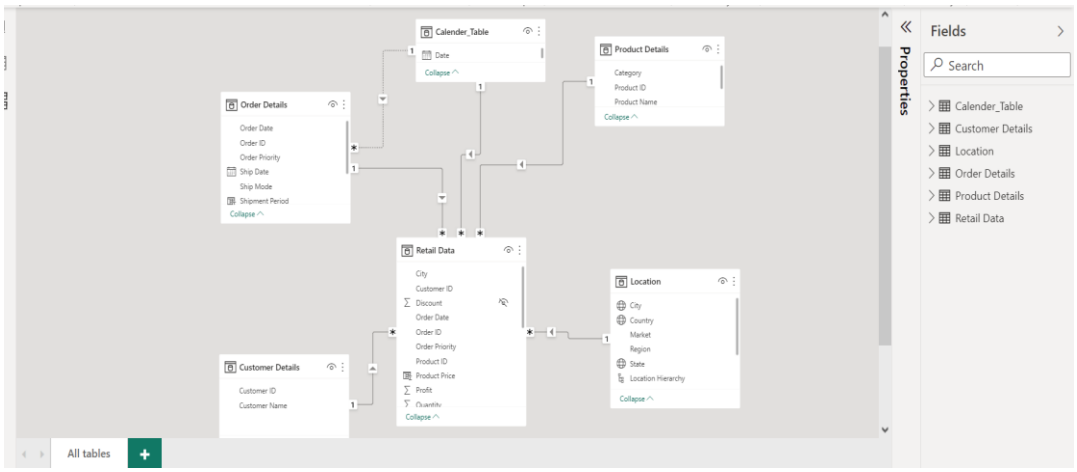


Fig. 16. Star Schema Model

Throughout the data model, we have One to Many "1:*" connections. To see this, hover your mouse over any of the highlighted links.

The Retail table that was developed would function as the fact table. No duplicates are deleted in this example to guarantee that the features in the dimension table are correctly referenced once the relationships are formed.

C. DAX and M Language

- DAX: The table below show the description of Measures and/or Calculated Columns added using DAX the PowerBI project.

Measures and/or Calculated Columns	DAX Formula	Description
Shipment Period	`Shipment Period = 'Order Details'[Ship Date]- 'Order Details'[Order Date]`	The formula provides the timeframe it took for an order to br delivered.
Product Price	`Product Price = 'Retail Data'[Sales] / 'Retail Data'[Quantity]`	Returns the unit price for each product.
Average Retail Price	`Average Retail Price = AVERAGE('Retail Data'[Product Price])`	Returns the average unit price per product
Previous Month orders	`Previous Month orders = CALCULATE([Total Orders], DATEADD(Calender_Table[Date],-1,MONTH))`	Provides the number of orders from the previous month
Previous Month Revenue	`Previous Month Revenue = CALCULATE([Total Revenue], DATEADD(Calender_Table[Date],-1,MONTH))`	Provides the revenue or sales figure from the previous month
Order Target	`Order Target = [Prev Month Orders] * 1.1`	This shows a 10% increase on the number of orders from the previous month
Revenue Target	`Revenue Target = [Prev Month Revenue] * 1.1`	This shows a 10% increase on the revenue or sales from the previous month
Profit Margin	`Profit Margin = 'Retail Data'[Total profit] / 'Retail Data'[Total Sales]`	Returns the profit ratio to sales

❖ M Language: M Language was used to create a calendar table to connect the date fields in the tables. A new, blank query was created (Source > Blank Query) in the Power Query Editor. In the formula bar, generate a starting date by entering a “literal” (in YYYY, MM, DD format): =#date(2011,1,1).

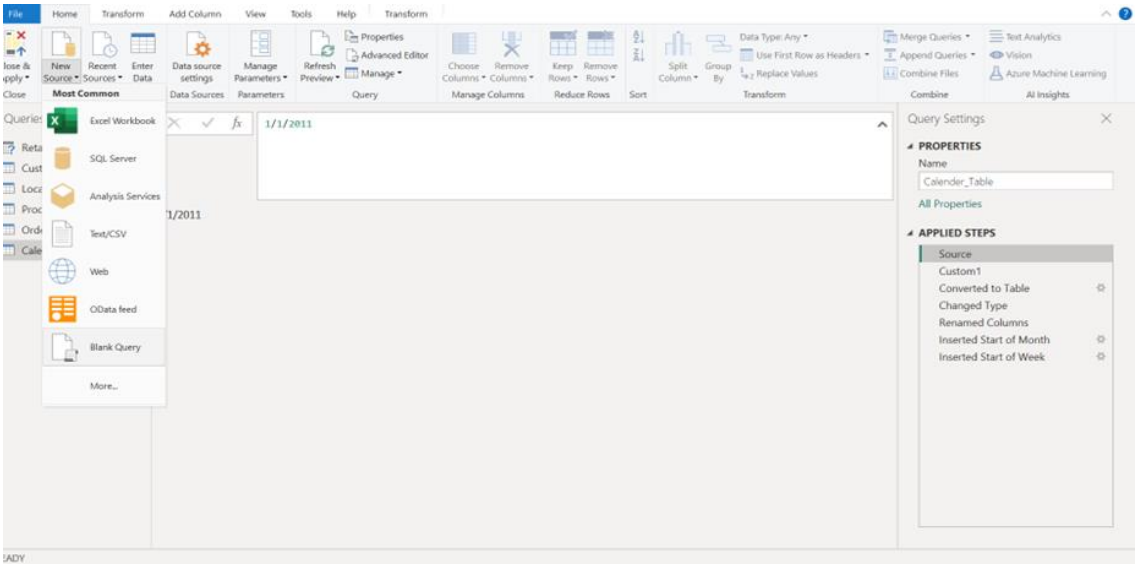


Fig. 17. Creating Blank Query

After which convert the resulting list into a Table (List Tools > To Table) and format the column as a Date.

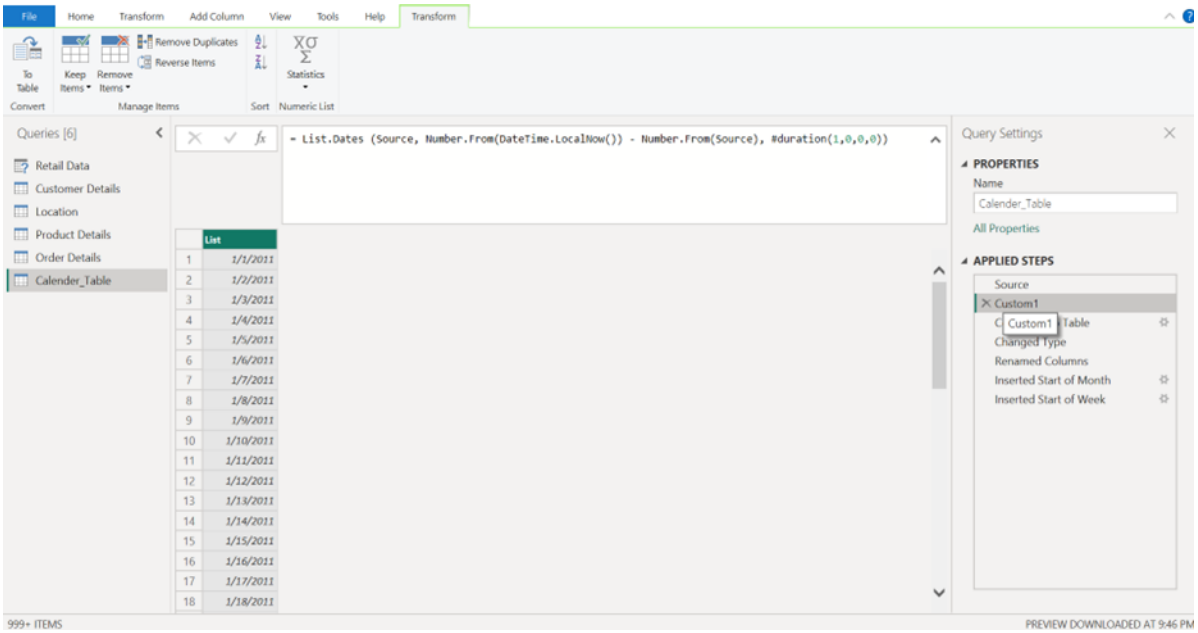


Fig. 18. Creating a list Using M Language

Add calculated Date columns (Start of Week, Start of Month etc.) as necessary using the Add Column tools.

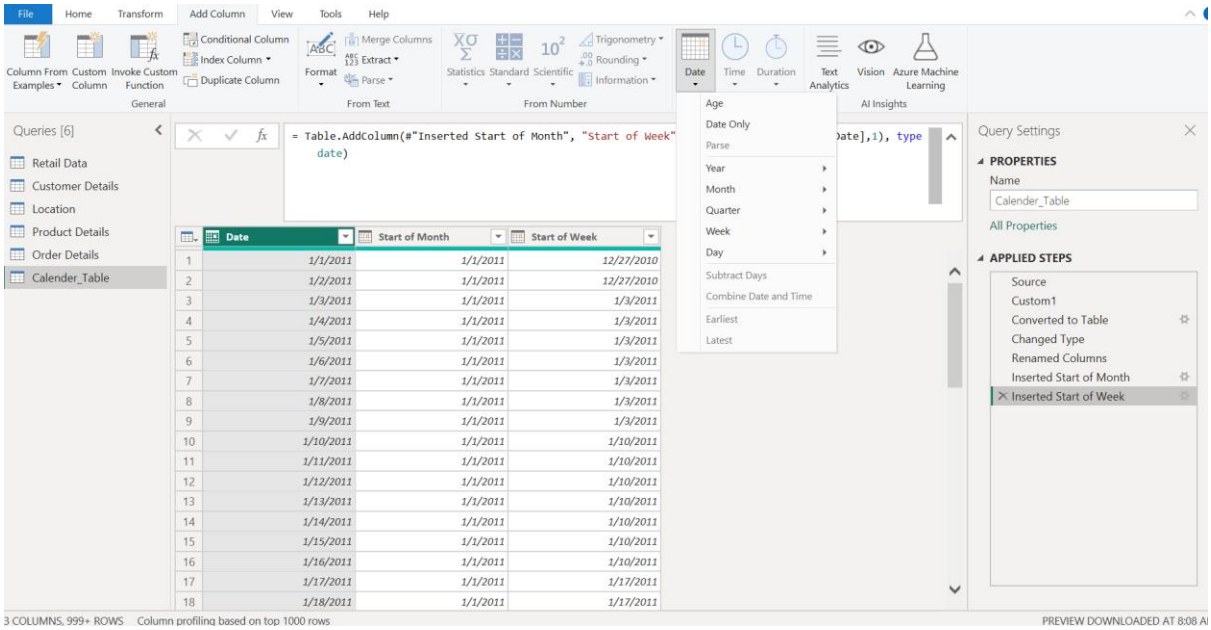


Fig. 19. Adding Calculated Date Columns

D. BI Design Solution

The Business Intelligence solution comprises of six pages for analysis carried out namely;

- Dashboard – Retail Analysis
- Artificial Intelligence (AI) Analysis
- Product Detail Analysis
- Customer Detail Analysis
- Revenue/Sales Analysis
- Profit Analysis

Page navigation buttons using infographics are also included, which make switching between page easier.

The Retail Analysis Dashboard contain various visualisations that depict the following; Key influencers of Sales, Yearly profit made for REG Superstores, Total Profit, Average retail price for all products, Orders and profit generated in various locations, the target revenue/sales against the actual and also the the target monthly orders against the actual.

Fig. 20. Retail Analysis Dashboard

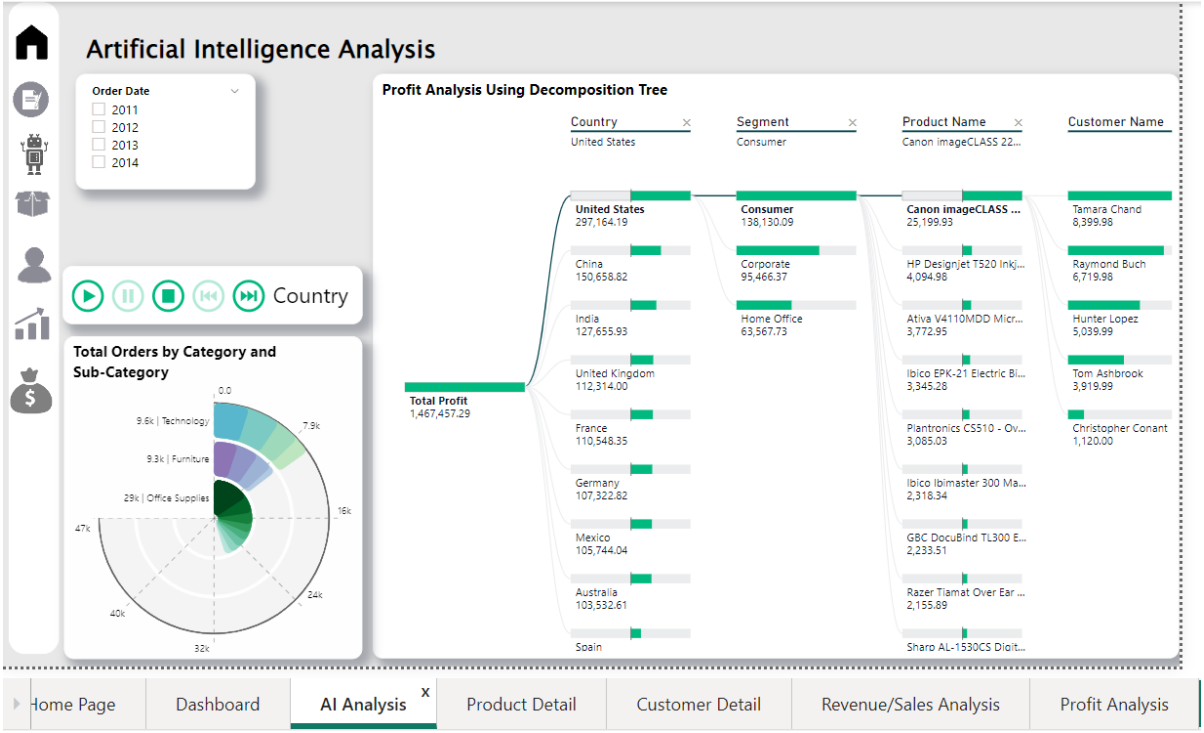


Fig. 21. Artificial Intelligence Analysis

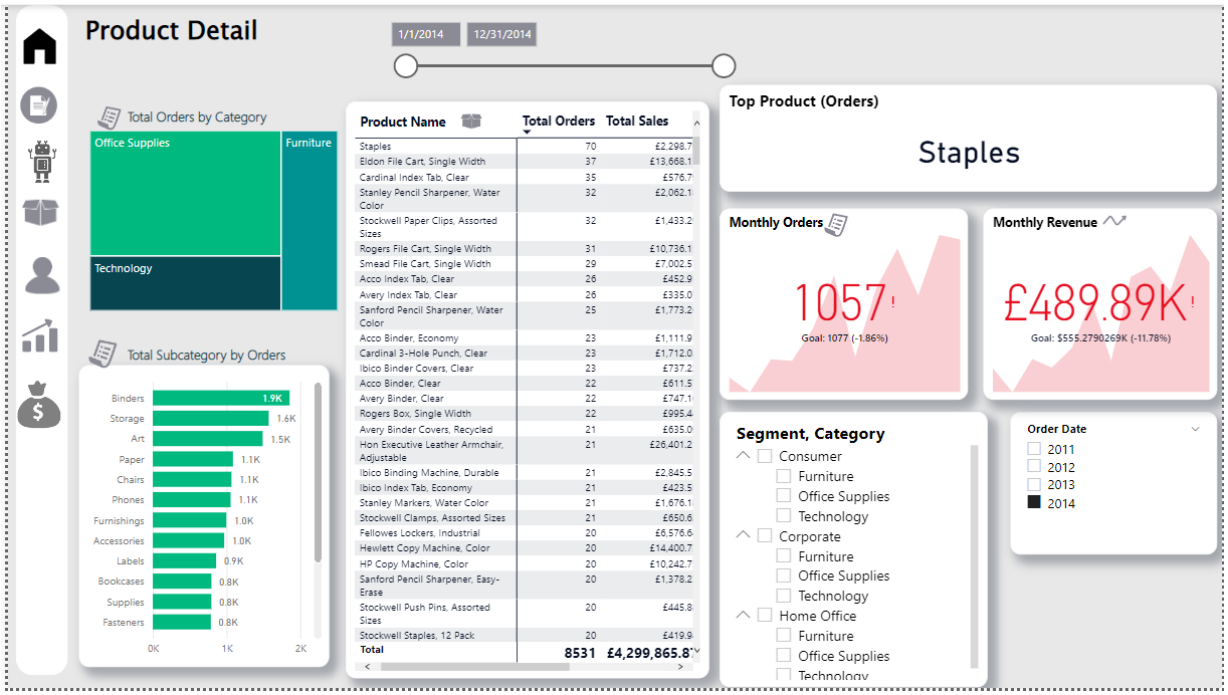


Fig. 22. Product Detail Analysis

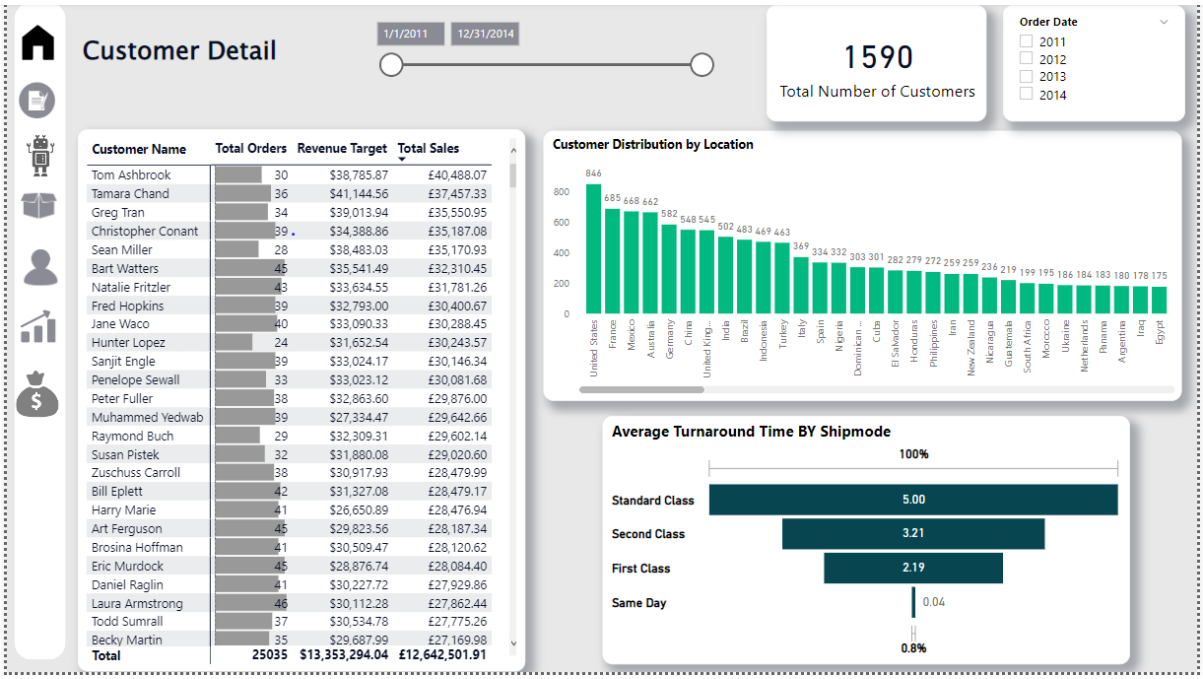


Fig. 23. Customer Detail Analysis

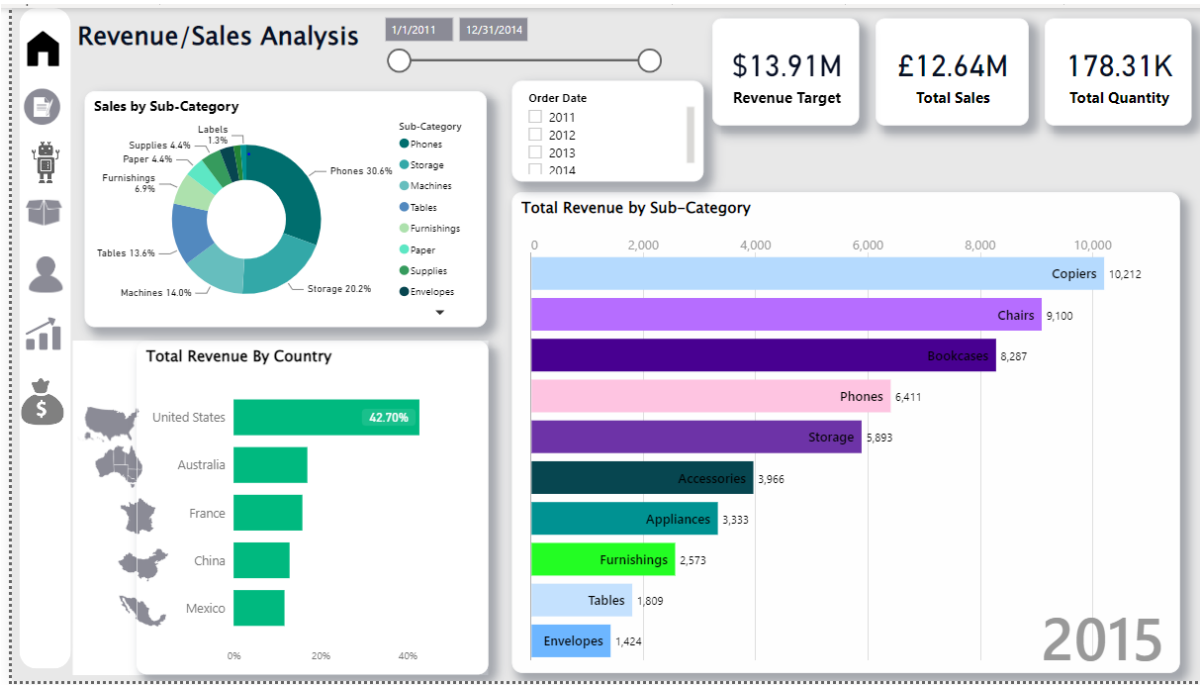


Fig. 24. Revenue/Sales Analysis

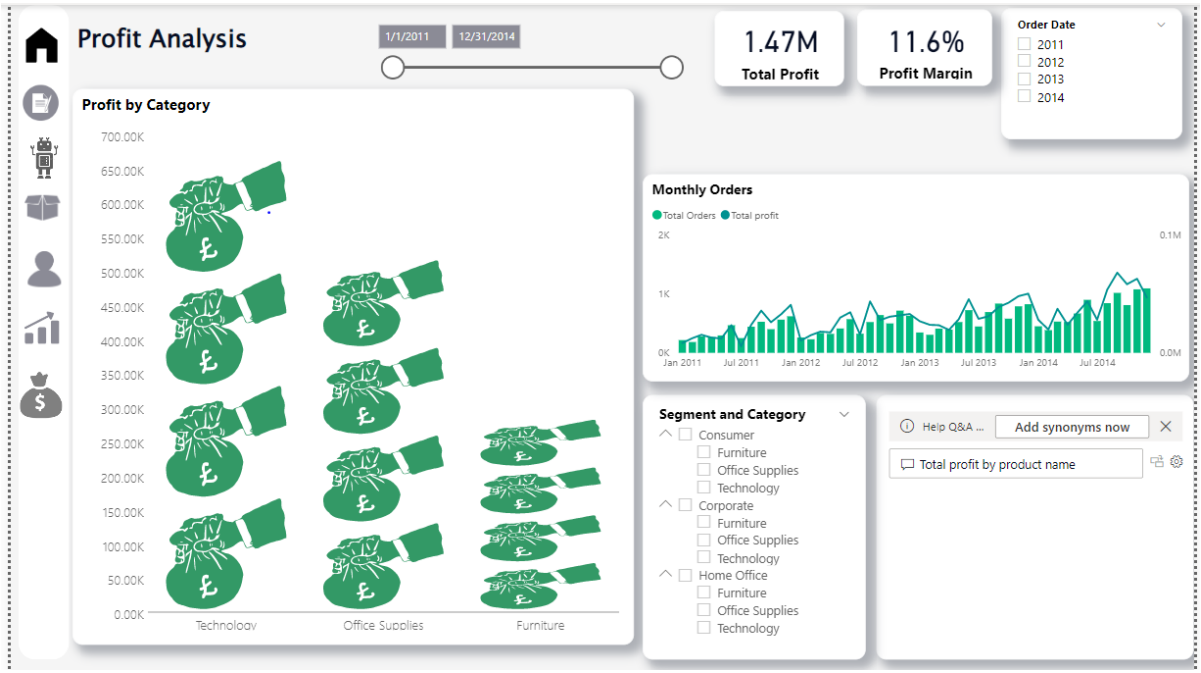


Fig. 25. Profit Analysis

E. Self-Assessment

Report Section	Description	Grade your work from 0 to 100

Report Structure	The report is well-written, and it contains all the relevant sections	100
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	100
Dax and M language	Both DAX and M Language have been extensively used in the report	95
Dashboard Design	The dashboard contains a variety of charts, including advanced ones.	100
Average		**Add below the average of the four cells above: 98.75**

REFERENCES

1. Super Store Sales Analysis (2021). Available at: https://medium.com/analytics-vidhya/exploratory-data-analysis-super-store-cb91c37bcb06
2. Super Store Sales Use-Case Data Analytics and Visualization (2021). Available at: https://medium.com/clique-org/superstore-sales-use-case-data-analytics-and-visualization-62afacd0777