Maynooth University
National University of Ireland Maynooth

Scoil Ghnó Ollscoil Mhá Nuad
Maynooth University School of Business

GROUP B

Contributing Team Members:

- Adhish Pillai
- Aswin Muthusamy
- Ganga Hariharan
- Omkar Krishnapurkar
- Yashwanth Duddupuddi

Contents

1. ***Introduction***

9,998 records from a comprehensive US sales dataset will be examined to identify patterns, fix errors, and evaluate the performance of client segments and regions. This will help data-driven initiatives aimed at boosting total sales and accomplishing corporate objectives.

2. ***Description of Analysis done***

We took a few crucial actions during the data preparation process to guarantee the dataset's integrity and suitability for analysis. To preserve uniqueness and correctness, we first eliminated three duplicate entries.

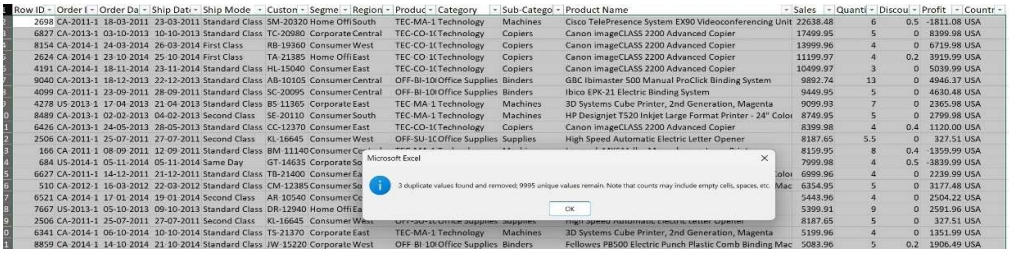

Figure a: Removing Duplicates in Excel

We found a discrepancy where one of the observations has "Technologi" instead of "Technology," for category.

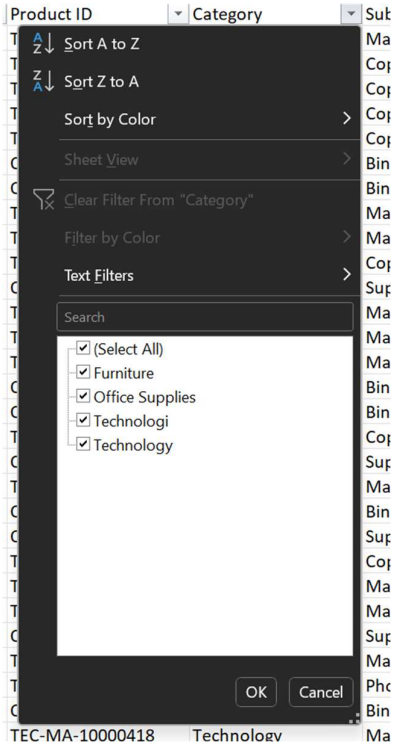

Figure b.1: Filter showing discrepancy

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Segmer	Region	Product ID	Category
CA-2011-145541	14-12-2011	21-12-2011	Standard Class	TB-21400	Consumer	East	TEC-MA-10001127	Technologi

Figure b.2: Record with 'Technologi'

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Segmer	Region	Product ID	Category
CA-2011-145541	14-12-2011	21-12-2011	Standard Class	TB-21400	Consumer	East	TEC-MA-10001127	Technology

Figure b.3: Record changed from 'Technologi' to 'Technology'

Since there was not enough data and reference values to implement a formula to all product IDs, we have decided to manually handle profits with blank data.

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Segmer	Region	Product ID	Category	Sub-Ca	Product Name	Sales	Quantit	Discour	Profit	Price Per Produ	Country
US-2012-126977	17-09-2012	23-09-2012	Standard Class	PF-19120	Consumer	East	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	4228.7	6	0.2		880.98	USA
CA-2011-116246	12-09-2011	17-09-2011	Standard Class	LW-17215	Consumer	East	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	3785.29	6	0.1		700.98	USA

Figure c: Product IDs with blank values.

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Segmer	Region	Product ID	Category	Sub-Ca	Product Name	Sales	Quantit	Discour	Profit	Price Per Produ	Country
CA-2014-118892	18-08-2014	23-08-2014	Standard Class	TP-21415	Consumer	East	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	4416.17	9	0.3	-630.88	700.98	USA
CA-2011-116246	12-09-2011	17-09-2011	Standard Class	LW-17215	Consumer	East	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	3785.29	6	0.1		700.98	USA
CA-2013-122903	28-05-2013	30-05-2013	Second Class	LA-16780	Corporate	Central	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	3504.9	5	0	700.98	700.98	USA
CA-2014-102204	02-05-2014	07-05-2014	Standard Class	CJ-12010	Consumer	South	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	2803.92	5	0.2	0	700.98	USA
CA-2012-164777	27-01-2012	29-01-2012	Second Class	SC-20305	Consumer	West	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	2803.92	5	0.2	0	700.98	USA
CA-2012-139731	15-10-2012	15-10-2012	Same Day	JE-15745	Consumer	Central	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	2453.43	5	0.3	-350.49	700.98	USA
CA-2013-136406	16-04-2013	18-04-2013	Second Class	BD-11320	Consumer	West	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	1121.57	2	0.2	0	700.98	USA
CA-2011-143903	20-07-2011	24-07-2011	Standard Class	KM-16375	Home Offi	Central	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	981.37	2	0.3	-140.2	700.98	USA

Figure d: For Product ID: FUR-CH-10002024

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Segmer	Region	Product ID	Category	Sub-Ca	Product Name	Sales	Quantit	Discour	Profit	Price Per Produ	Country
CA-2014-118892	18-08-2014	23-08-2014	Standard Class	TP-21415	Consumer	East	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	4416.17	9	0.3	-630.88	700.98	USA
CA-2011-116246	12-09-2011	17-09-2011	Standard Class	LW-17215	Consumer	East	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	3785.29	6	0.1	420.59	700.98	USA
CA-2013-122903	28-05-2013	30-05-2013	Second Class	LA-16780	Corporate	Central	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	3504.9	5	0	700.98	700.98	USA
CA-2014-102204	02-05-2014	07-05-2014	Standard Class	CJ-12010	Consumer	South	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	2803.92	5	0.2	0	700.98	USA
CA-2012-164777	27-01-2012	29-01-2012	Second Class	SC-20305	Consumer	West	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	2803.92	5	0.2	0	700.98	USA
CA-2012-139731	15-10-2012	15-10-2012	Same Day	JE-15745	Consumer	Central	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	2453.43	5	0.3	-350.49	700.98	USA
CA-2013-136406	16-04-2013	18-04-2013	Second Class	BD-11320	Consumer	West	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	1121.57	2	0.2	0	700.98	USA
CA-2011-143903	20-07-2011	24-07-2011	Standard Class	KM-16375	Home Offi	Central	FUR-CH-10002024	Furniture	Chairs	HON 5400 Series Task Chairs	981.37	2	0.3	-140.2	700.98	USA

Figure e: For Product ID: FUR-CH-10002024 Profit Calculated

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Segmer	Region	Product ID	Category	Sub-Ca	Product Name	Sales	Quantit	Discour	Profit	Price Per Produ	Country
CA-2012-117086	08-11-2012	12-11-2012	Standard Class	QJ-19255	Corporate	East	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	4404.9	5	0	1013.13	880.98	USA
US-2012-126977	17-09-2012	23-09-2012	Standard Class	PF-19120	Consumer	East	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	4228.7	6	0.2	158.58	880.98	USA
US-2012-150630	17-09-2012	21-09-2012	Standard Class	TB-21520	Consumer	East	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	3083.43	7	0.5	-1665.05	880.98	USA
CA-2013-108987	09-09-2013	11-09-2013	Second Class	AG-10675	Consumer	Central	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	2396.27	4	0.32	-317.15	880.98	USA
US-2014-109316	09-06-2014	11-06-2014	Second Class	MG-17680	Home Offi	West	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	1497.67	2	0.15	140.96	880.98	USA

Figure f: Figure e: For Product ID: FUR-BO-10004834

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Segmer	Region	Product ID	Category	Sub-Cat	Product Name	Sales	Quantit	Discour	Profit	Price Per Produ	Countr
CA-2012-117086	08-11-2012	12-11-2012	Standard Class	QJ-19255	Corporate	East	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	4404.9	5	0	1013.13	880.98	USA
US-2012-126977	17-09-2012	23-09-2012	Standard Class	PF-19120	Consumer	East	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	4228.7	6	0.2		880.98	USA
US-2012-150630	17-09-2012	21-09-2012	Standard Class	TB-21520	Consumer	East	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	3083.43	7	0.5	-1665.05	880.98	USA
CA-2013-108987	09-09-2013	11-09-2013	Second Class	AG-10675	Consumer	Central	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	2396.27	4	0.32	-317.15	880.98	USA
US-2014-109316	09-06-2014	11-06-2014	Second Class	MG-17680	Home Offi	West	FUR-BO-10004834	Furniture	Bookcases	Riverside Palais Royal Lawye	1497.67	2	0.15	140.96	880.98	USA

Figure g: Figure e: For Product ID: FUR-BO-10004834 Profit Calculated

*Then we used **Tableau Prep Builder** to efficiently complete the cleaning process.*

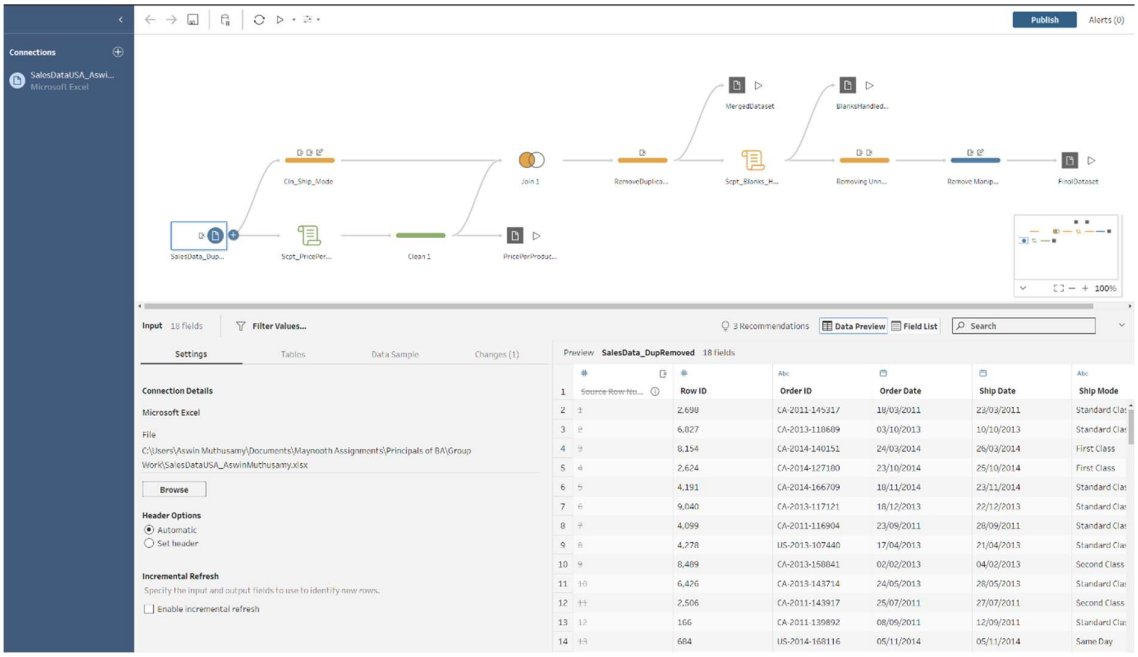

Figure h: Complete Tableau Prep Flow

We developed a Python script that produced a distinct dataset to determine the pricing per product for every unique product ID. This dataset was then combined with the original using a Left Outer Join based on "Product ID." We were able to create a new "Price per Product" column as a result. While in original dataset we are adding a dataDifference calculated field to correct the ship mode based on the below,

dateDifference	***Ship Mode***
= 0	*Same Day*
= 1	*First Class*
= 2	*Second Class*
>= 3	*Standard Class*

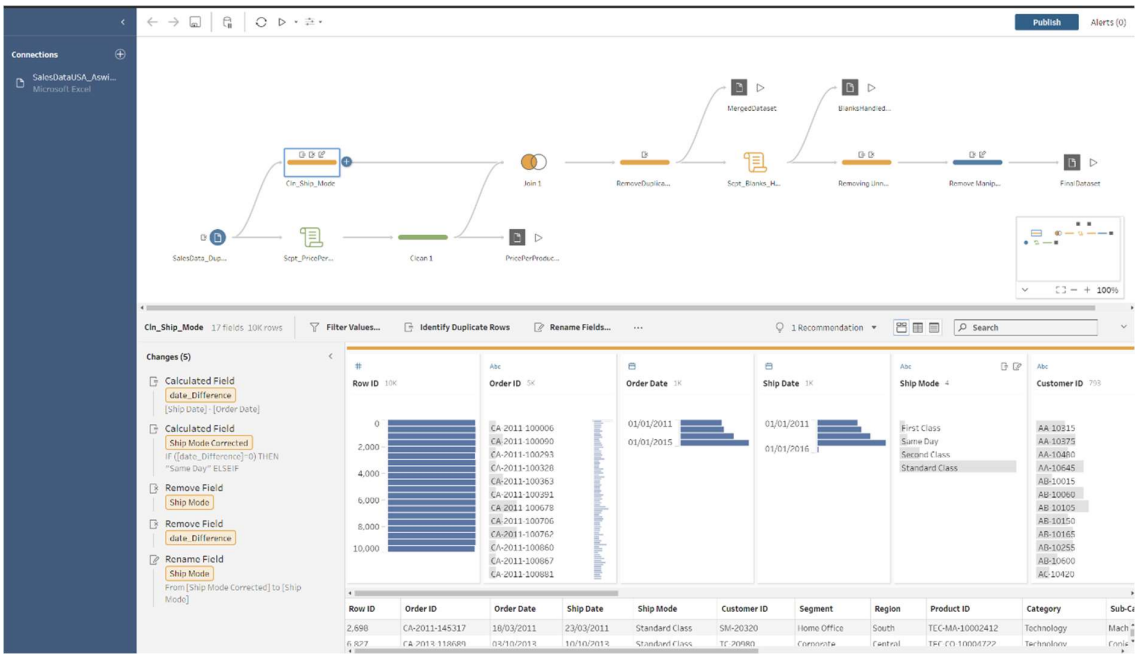

Figure i: Cleaning Ship Mode

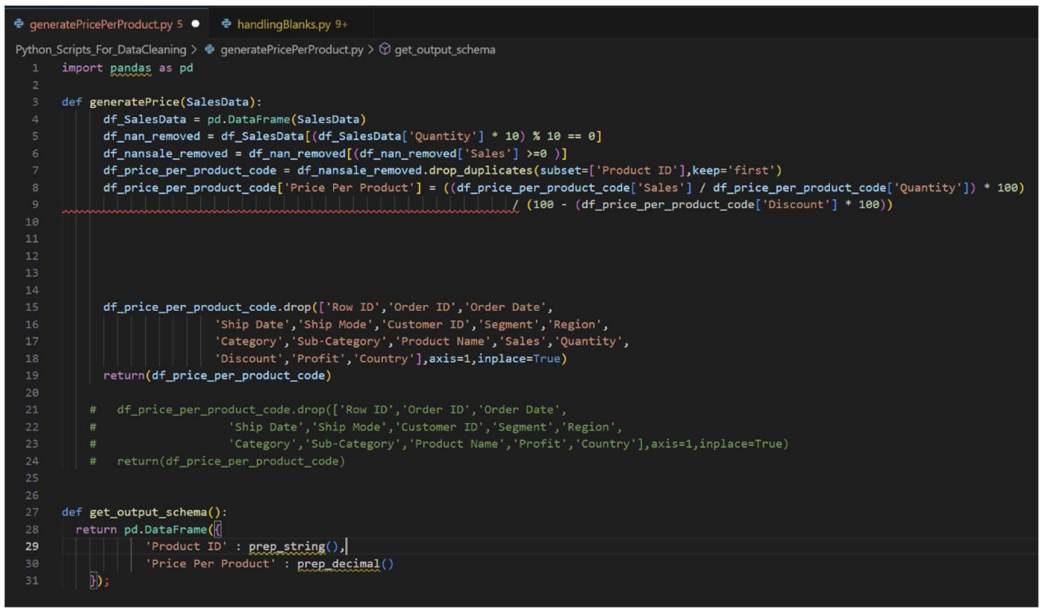

Figure j: Script that Generates the new 'Price Per Product' dataset

To eliminate blanks and manage decimal values in the "Quantity" and "Sales" columns, we additionally included another Python script. This script will handle blank and decimal values for "Quantity" and the blank values for 'Sales'. We

also removed 'Product IDs' with no 'Sales' or reference values to Calculate 'Price Per Product'.

generatePricePerProduct.py 5 ● handlingBlanks.py 9+ ●

Python_Scripts_For_DataCleaning > handlingBlanks.py > handleBlanks > df_blank_sales_and_qty_handled

```python
import pandas as pd

def handleBlanks(MergedSalesData):
    #Handling Blanks in Sales
    df_merged_salesData = pd.DataFrame(MergedSalesData)
    df_merged_salesData.loc[df_merged_salesData['Sales'].isna(),'Sales'] =
    ((df_merged_salesData['Price Per Product'] * df_merged_salesData['Quantity']) * (100 - (df_merged_salesData['Discount'] * 100))) / 100
    df_blankSales_handled = df_merged_salesData

    #Handling Decimal Values in Quantity
    df_blankSales_handled.loc[(df_blankSales_handled['Quantity'] * 10) % 10 != 0 , 'Quantity'] = 0

    #Handling Blanks and Decimal Values in Quantity
    df_blankSales_handled.loc[df_blankSales_handled['Quantity'].isna() | df_blankSales_handled['Quantity'] == 0 , 'Quantity'] =
    (df_blankSales_handled['Sales'] * 100) / (df_blankSales_handled['Price Per Product'] * (100 - (df_blankSales_handled['Discount'] * 100)))

    df_blank_sales_and_qty_handled = df_blankSales_handled

    #Eliminating Data with No Sales value and no reference value to calculate Price per product
    df_final = df_blank_sales_and_qty_handled.dropna(subset='Price Per Product')

    return(df_final)


def get_output_schema():
    return pd.DataFrame({
            'Row ID' : prep_int(),
            'Order ID' : prep_string(),
            'Order Date' : prep_date(),
            'Ship Date' : prep_date(),
            'Ship Mode' : prep_string(),
            'Customer ID' : prep_string(),
            'Segment' : prep_string(),
            'Region' : prep_string(),
            'Product ID' : prep_string(),
            'Category' : prep_string(),
            'Sub-Category' : prep_string(),
            'Product Name' : prep_string(),
            'Sales' : prep_decimal(),
            'Quantity' : prep_decimal(),
            'Discount' : prep_decimal(),
            'Profit' : prep_decimal(),
            'Price Per Product' : prep_decimal(),
            'Country' : prep_string()
        });
```

Figure k: Script that handles blanks in 'Sales' and 'Blanks'

Finally, we constructed calculated fields to round values of "Sales," "Quantity," "Profit," and "Price per Product" columns to two decimal places.

The Excel worksheet was then renamed to "Refined Dataset" which will now allow us to extract insightful information from the data.

3. ***Visualizations and Insights:***

In the visualization section, we created a Power Bi Dashboard. Below are the insights and type of visualizations used-

- ***Sales per Month over the years****: From 2011 to 2014, the "Sales per Month over the Years" figure shows a steady increase trend. Notable sales peaks usually occur in late-year months like October and December, reflecting seasonal demand trends. Sales in the early months of the year, especially January, are frequently lower, indicating both seasonality in consumer behavior and overall company expansion during the studied time.*

- ***Monthly Sales Trend:*** *With sales peaking in March, September, and October and reaching a maximum of $0.31 million in October, the "Monthly Sales Trend" chart clearly shows seasonality. January, February, and May are the early and mid-year months with lower sales volumes. These recurrent peaks point to regular demand cycles, which can guide year-round inventory and marketing plans.*

- ***Total Sales by Region:*** *Sales in the Central, East, South, and West regions are shown in a stacked bar chart. This graph illustrates the regions that contribute the most to sales, demonstrating that some regions—such as the West—have the largest sales. Sales are strongest in the West, which may be the company's primary market for future expansion.*

- ***Average Shipping Time for each month:*** *Shipping times, which range from 3.7 to 4.2 days, are comparatively stable, according to the "Average Shipping Time for Each Month" figure. Because of effective*

logistics, January and February have the greatest averages (4.2 days), while November has the lowest (3.81 days). April and September peaks could point to locations for delivery optimization due to strong demand or operational difficulties.

- ***Total no. of orders based on shipping time:*** *According to the "Total Number of Orders Based on Shipping Time" figure, the majority of orders are delivered within two to four days, with Standard Class and 3-day delivery being the most common. This implies that consumers value dependable, reasonably priced delivery over speed. Promoting First Class and Same Day delivery could draw clients looking for quicker solutions, while highlighting Standard and Second-Class options may increase customer satisfaction.*

- ***Correlation between Avg. Shipping speed and total sales:*** *Faster shipping is positively correlated with better sales, according to the "Correlation Between Average Shipping Speed and Total Sales" graphic. Reduced shipping durations are associated with higher sales, indicating that effective delivery can boost client satisfaction and spur revenue expansion, underscoring the need of optimizing delivery procedures.*

- ***Profit & Sales by Segment and Region:*** *With technology accounting for 48.96% of sales, the office category is exceptionally profitable. Office supplies and furniture come in second and third. By directing optimal discount tactics and focused investments in high-performing regions and segments, minimal discounts optimize revenue.*

- ***Sales by Category:*** *The percentage of sales for each product category—furniture, office supplies, and technology—is displayed in a pie chart. The graphic shows that technology products account for the largest portion of total sales. Since technology items account for the largest portion of sales, the company may benefit from concentrating on growing this market.*

- ***Sales by Discount:***

  The dashboard's "Sales by Category" shows technology leading at 48.96% of revenue, followed by furniture (37.95%) and office supplies (13.09%), guiding strategic focus on technology marketing, inventory, and growth opportunities.

- ***Team Collaboration***

Team members worked closely together throughout the project, each contributing in a different way.

Sr. No	***Team Member***	***Lead Role***	***Supporting Role***
1.	***Aswin Muthusamy***	***Data Cleaning, Scripting, Creating Dashboard***	***Proof-reading***
2.	***Ganga Hariharan***	***Creating the report, Drawing Insights from Refined Data***	***Proof-reading***
3.	***Omkar Krishnapurkar***	***Creating the Report, Scheduling Meetings***	***Creating Dashboards***
4.	***Adhish Pillai***	***Preparing Presentation***	***Drawing Insights from Refined Data***
5.	***Yashwanth Duddupuddi***	***Preparing Presentation***	

Each member's committed efforts made the project successful, demonstrating an overall and cooperative effort throughout the process.

4. ***Conclusion***

To facilitate strategic planning and data-driven decision-making, this study sought to extract useful insights from sales data. After carefully cleansing the data, we used Power BI to develop an interactive dashboard that allowed us to examine important trends like client profitability, geography sales, and product category success.

Important conclusions emphasized the relationship between discount tactics and profit margins as well as how client demographics and geographic location affect sales growth. These findings highlight the necessity of specialized advertising, focused marketing, and improved pricing.
All things considered, this study offers a basis for improving sales and profitability, allowing the company to hone its pricing, distribution, and segmentation initiatives to increase productivity and financial results.