



DATA ANALYSIS FOR SALES RETAIL E-COMMERCE

DEPI GRADUATION PROJECT



TEAM MEMBERS

MARINA
BASSEM
FAROUK

JESSICA
ELKESS
BISHOUY

SHERRY
ATEF FAWZY

YOSR
AHMED

HAGER
WAEL
FAROUK



AGENDA

- INTRODUCTION
- PROJECT IDEA
- DATA CLEANING AND PREPROCESSING
- DAX AND MEASURES
- INSIGHTS GENERATIONS
- DATA VISUALIZATION
- CONCLUSION
- Q&A

INTRODUCTION



This project is part of our DEPI graduation requirements and focuses on analyzing sales performance in the retail e-commerce sector.



The aim of the project is to explore sales trends, customer behavior, and key performance indicators using data analysis techniques. Through this analysis, we provide insights that support better decision-making and help improve overall business performance.



project idea

1)problem:

Businesses often struggle to understand their sales performance, customer behavior, and product profitability across different countries and time periods.

2)solution:

We analyzed the dataset using Excel, Power BI, and Tableau, applying pivot tables, dashboards, and interactive visuals to uncover trends in quantity sold, total profit, customer activity, and product performance.



data cleaning:

- 1. Removing Unnecessary Columns**
- 2. Fixing Data Types** We corrected the data types of several columns to ensure
- 3. Cleaning the Date & Time Format**
- 4. Handling Missing Values** :Some fields had missing or blank values.
- 5. Removing Duplicates** :Duplicates were identified in invoice transactions.
- 6. Filtering Out Invalid Records**:We removed rows where:Quantity ≤ 0 (Negative quantities)
- 7. Extracting New Columns**
- 8. Standardizing Country Names**
- 9. Fixing Negative Values**
- 10. Creating Clean Final Dataset**

data cleaning:

	ABC 123	Invoice	ABC 123	StockCode	1 2 3	Quantity	ABC 123	InvoiceDate	1.2	Price	1.2	Customer ID	ABC 123	Country
		● Valid • Error ● Empty		100% 0% 0%	● Valid • Error ● Empty	100% 0% 0%								
1		536365	85123A			6	01/12/2010 08:26:00	ص		2.55		17850	United Kingdom	
2		536365		71053		6	01/12/2010 08:26:00	ص		3.39		17850	United Kingdom	
3		536365	84406B			8	01/12/2010 08:26:00	ص		2.75		17850	United Kingdom	
4		536365	84029G			6	01/12/2010 08:26:00	ص		3.39		17850	United Kingdom	
5		536365	84029E			6	01/12/2010 08:26:00	ص		3.39		17850	United Kingdom	
6		536365		22752		2	01/12/2010 08:26:00	ص		7.65		17850	United Kingdom	
7		536365		21730		6	01/12/2010 08:26:00	ص		4.25		17850	United Kingdom	
8		536366		22633		6	01/12/2010 08:28:00	ص		1.85		17850	United Kingdom	
9		536366		22632		6	01/12/2010 08:28:00	ص		1.85		17850	United Kingdom	
10		536368		22960		6	01/12/2010 08:34:00	ص		4.25		13047	United Kingdom	
11		536368		22913		3	01/12/2010 08:34:00	ص		4.95		13047	United Kingdom	
12		536368		22912		3	01/12/2010 08:34:00	ص		4.95		13047	United Kingdom	
13		536368		22914		3	01/12/2010 08:34:00	ص		4.95		13047	United Kingdom	
14		536367		84879		32	01/12/2010 08:34:00	ص		1.69		13047	United Kingdom	
15		536367		22745		6	01/12/2010 08:34:00	ص		2.1		13047	United Kingdom	
16		536367		22748		6	01/12/2010 08:34:00	ص		2.1		13047	United Kingdom	
17		536367		22749		8	01/12/2010 08:34:00	ص		3.75		13047	United Kingdom	
18		536367		22310		6	01/12/2010 08:34:00	ص		1.65		13047	United Kingdom	
19		536367		84969		6	01/12/2010 08:34:00	ص		4.25		13047	United Kingdom	
20		536367		22623		3	01/12/2010 08:34:00	ص		4.95		13047	United Kingdom	
21		536367		22622		2	01/12/2010 08:34:00	ص		9.95		13047	United Kingdom	
22		536367		21754		3	01/12/2010 08:34:00	ص		5.95		13047	United Kingdom	
23		536367		21755		3	01/12/2010 08:34:00	ص		5.95		13047	United Kingdom	



DAX AND MEASURES:

KPIs of product performance:

```
Average Price = AVERAGE('All sales'[unit price])
```

AVERAGE PRICE: SIMPLE AVERAGE OF THE UNIT PRICE COLUMN.

```
Total Quantity Sold =  
CALCULATE(  
    SUM('All sales'[Quantity])  
    , 'All sales'[Quantity] > 0)
```

Total Quantity Sold: Sum of units sold, filtered to exclude zero quantity entries

```
1 Return Rate = DIVIDE([Total Returns (Abs)], [Total Quantity Sold], 0 )
```

RETURN RATE: RETURNS DIVIDED BY TOTAL QUANTITY SOLD (USING SAFE DIVIDE).

```
Total Returns (Abs) = -[Total Returned Units]
```

TOTAL RETURNS (ABS): THE POSITIVE VALUE OF RETURNED UNITS.

```
Average Order Value = [Total Revenue (All)] / [Total Orders]
```

AVERAGE ORDER VALUE: TOTAL REVENUE DIVIDED BY TOTAL ORDERS.

```
1 Total Orders = DISTINCTCOUNT('All sales'[Invoice])
```

TOTAL ORDERS: COUNTS THE UNIQUE INVOICES (TRANSACTIONS).

```
1 Total Revenue =  
2 CALCULATE(  
3     SUMX('All sales', 'All sales'[Quantity] * 'All sales'[unit price]),  
4     'All sales'[Quantity] > 0)
```

TOTAL REVENUE: CALCULATED BY MULTIPLYING QUANTITY \$\times\$ UNIT PRICE FOR EACH SALE, ONLY INCLUDING POSITIVE SALES.

```
1 Lost Revenue = CALCULATE(SUMX('All sales', 'All sales'[Quantity] * 'All sales'[unit price]), 'All sales'[Quantity] < 0)
```

LOST REVENUE: REVENUE CALCULATION APPLIED SPECIFICALLY TO TRANSACTIONS WHERE QUANTITY IS NEGATIVE (RETURNS).

DAX AND MEASURES:

KPIs of time trend:

```
1 Sales = 'All sales'[unit price] * 'All sales'[Quantity]
```

sum of sales over years and months:

KPIs of countries:

```
1 Sales = 'All sales'[unit price] * 'All sales'[Quantity]
```

sum of sales by country:

INSIGHTS GENERATIONS:

KPIs of product performance:

Which products are driving the most revenue and volume, and what is the overall health (profitability and returns) of the product line?

KPI Cards (Top Row)

These metrics give a full executive overview of the product portfolio, covering the business's scale, value, and performance. Total Revenue (\$21M) and Total Orders (54K) show the overall size of the business, while Total Quantity sold (12M units) reflects the volume of products moved.

Value indicators like Avg Order Value (\$360) and Average Unit Price (\$4.6) describe typical transaction size and product pricing. Finally, business health is highlighted through Lost Revenue (\$-1.53M) and a Return Rate of 9.12% (1M returned units), revealing key inefficiencies and areas that need operational improvement.

Top Selling Products (by Quantity):

This column chart identifies the operational leaders by volume. The "WORLD WAR 2 GLIDERS ASSTD DESIGNS" is the most frequently purchased item (109K units), indicating high customer demand for these products.

Total Selling Products (by Revenue)

The bar chart shows the top-revenue products, with REGENCY CAKESTAND 3 TIER generating the highest revenue (345K), followed by Manual (341K). DOTCOM POSTAGE and WHITE HANGING HEART T-LIGHT HOLDER also perform strongly, while PAPER CRAFT, LITTLE BIRDIE ranks lower.

The visual clearly identifies the main revenue-driving products.

INSIGHTS GENERATIONS:

KPIs of countries:

Which geographical markets contribute the most to sales and revenue, and how does the sales distribution look globally?

Sales and Country Map

This provides a geographic overview, immediately showing that Europe is the core region for the business. It allows for quick visual identification of countries that are either strong contributors or where sales losses are occurring (indicated by the negative range in the sales legend).

Sum of units sold by Country

This bar chart quantifies market volume. It shows that the United Kingdom (UK) is the overwhelmingly largest market by quantity (8.7M units), confirming it as the primary market and operational hub for distribution and sales Volume.

Sum of Sales by Country

This bar chart quantifies market value. The United Kingdom (UK) also dominates in total revenue, reinforcing its status as the most financially valuable market. The next tier of important revenue markets includes EIRE, Netherlands, Germany, and France.



INSIGHTS GENERATIONS:

KPIs of time trend:

How have sales performed over time (years and months), revealing growth patterns, seasonality, and overall trend stability?

Sum of Sales over Years & Months

This Flow Diagram illustrates the path of sales over the period. It shows the massive jump in scale from 2009 to 2010. It also pinpoints the highest-revenue months, with November and December appearing as the most significant contributors across the different years.

Sum of Sales by Year

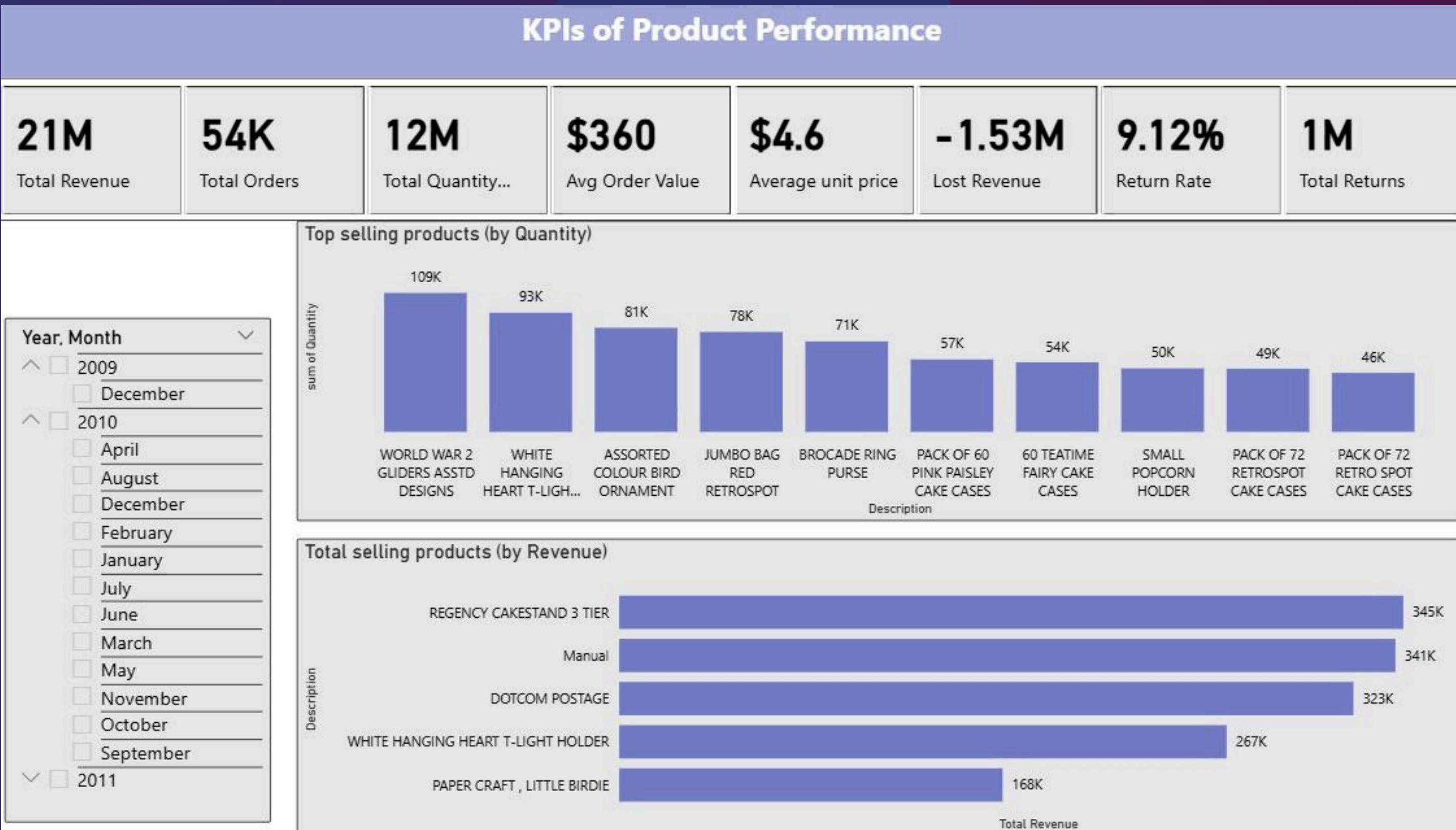
This line chart reveals the overall annual trend. The business had a breakout year in 2010, then entered a stabilization phase in 2011. Management should investigate what drove the huge growth in 2010 and why it wasn't sustained in 2011 – identifying these drivers will help regain growth.

Sum of Sales by Month

This line chart highlights the strong seasonality in the business. Sales peak significantly in the late autumn/winter months, with November (2.88M) and December (2.36M) being the highest-performing months. Sales clearly follow a seasonal pattern. They drop to the lowest point in February. After February, sales slowly rise again and stay steady from May to August. Overall, the highest sales happen from October to December.

DATA VISUALIZATION:

powerbi:



DATA VISUALIZATION:

KPIs of Countries

Country

- Australia
- Austria
- Bahrain
- Belgium
- Bermuda
- Brazil
- Canada
- Channel Islands
- Cyprus
- Czech Republic
- Denmark
- EIRE
- European Community
- Finland
- France
- Germany
- Greece
- Hong Kong
- Iceland
- Israel
- Italy
- Japan
- Korea
- Lebanon
- Lithuania
- Malta
- Netherlands
- ...

Total Revenue	Total Returned Units	Rank
20,814,292.00	-1064078	1

Country and Sales

Microsoft Bing © 2025 TomTom, © 2025 Microsoft Corporation, © OpenStreetMap. Terms

Sum of units sold by Country

Country	Sum of Quantity
United Kingdom	8.7M
Netherlands	0.4M
EIRE	0.3M
Denmark	0.2M
Germany	0.2M

Sum of Sales by Country

Country	Sum of Sales
United Kingdom	16383K
EIRE	616K
Netherlands	549K
Germany	418K
France	328K

DATA VISUALIZATION:

KPIs of Time Trends

Month, Year

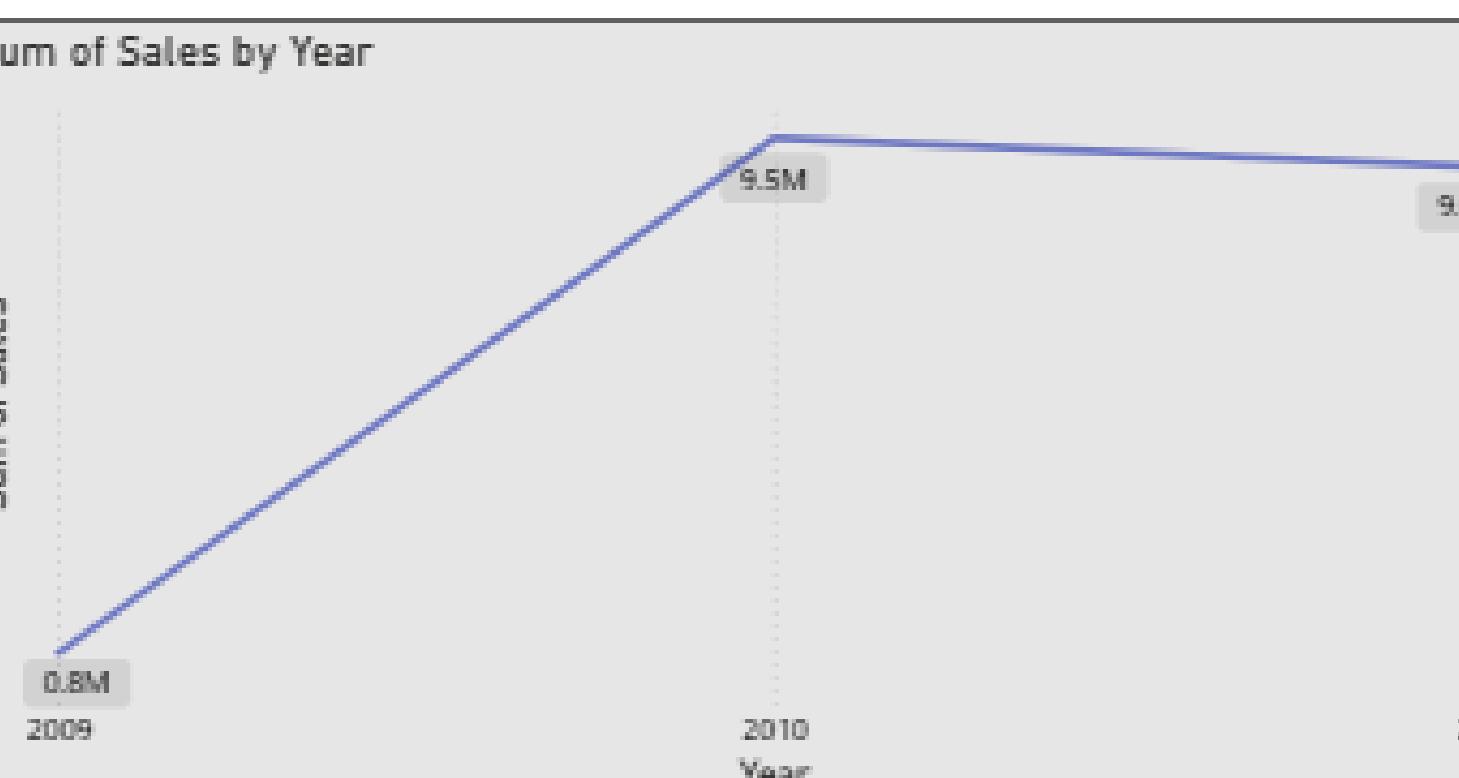
- April
 - 2010
 - 2011
- August
- December
- February
- January
- July
- June

Sum of Sales over Years & Months

Year: 2010 Month: November

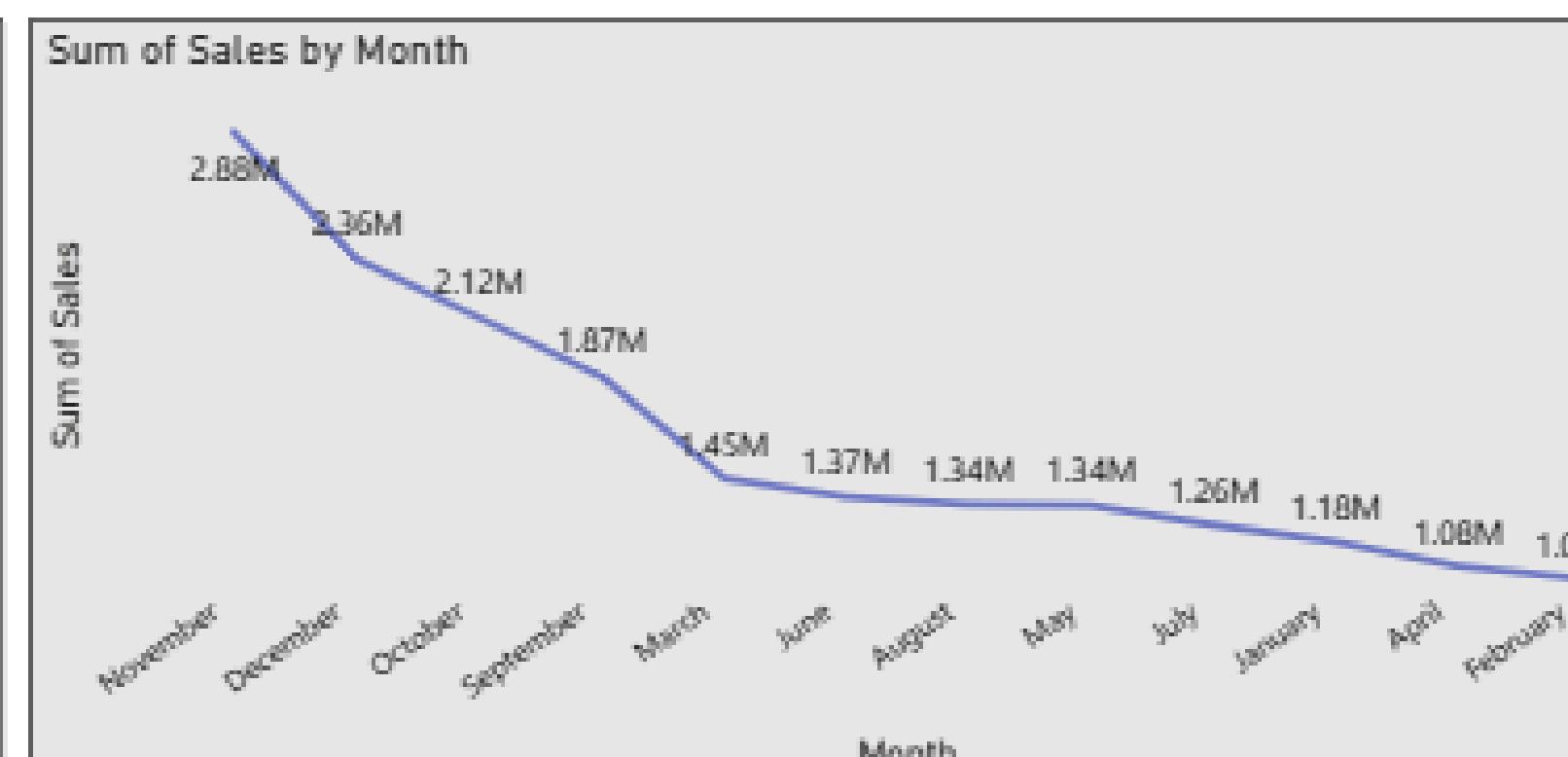
Year	Month	Sum of Sales
2010	November	9,488,595
2011	December	8,998,809
2009	October	799,847

Sum of Sales by Year



Year	Sum of Sales
2009	0.8M
2010	9.5M
2011	9.0M

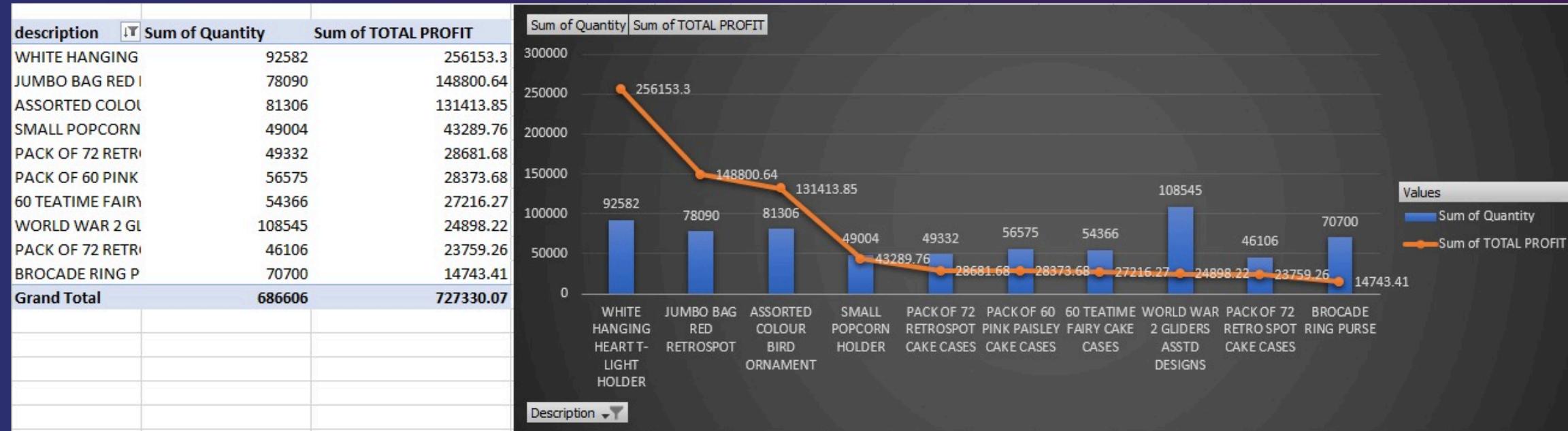
Sum of Sales by Month



Month	Sum of Sales
November	2.88M
December	2.36M
January	2.12M
February	1.87M
March	1.45M
April	1.37M
May	1.34M
June	1.34M
July	1.26M
August	1.18M
September	1.08M
October	1.03M

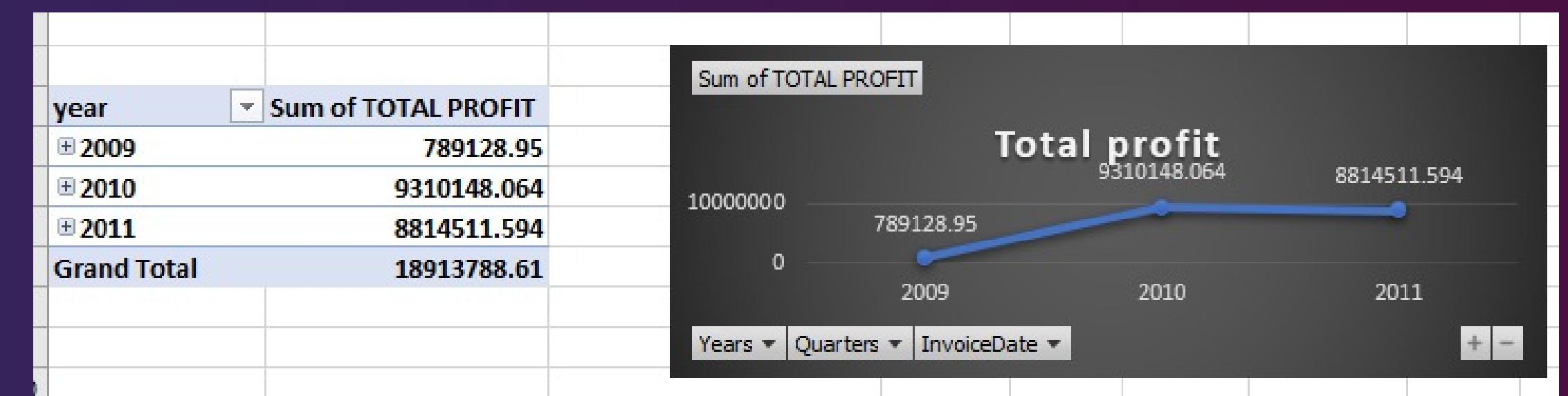
DATA VISUALIZATION:

Excel:



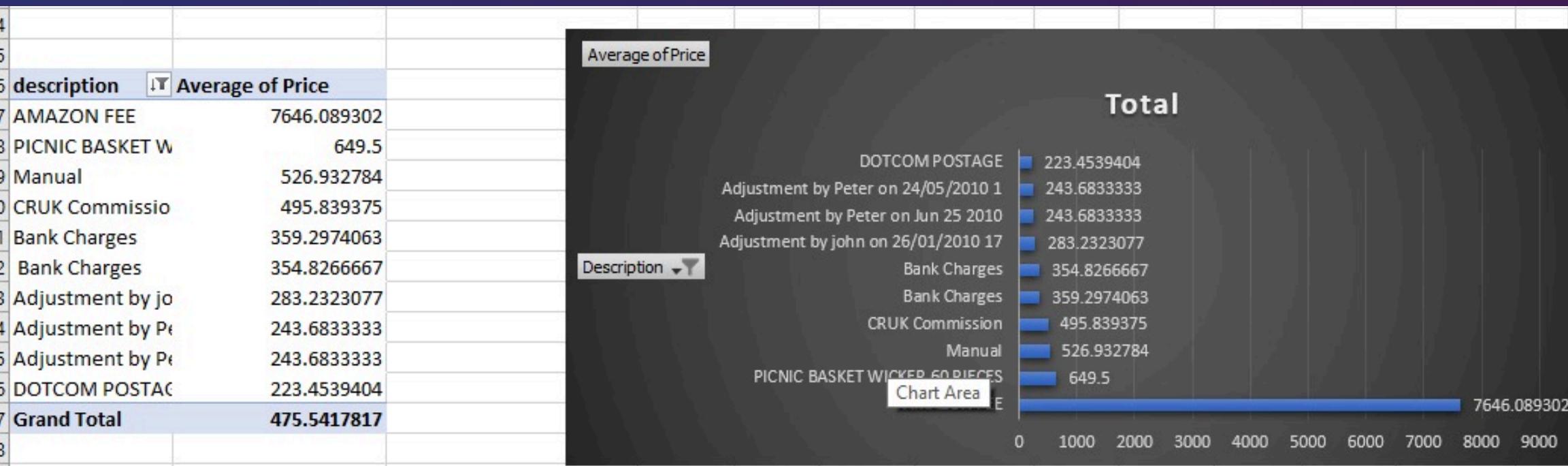
Quantity vs profit per product

Total profit per year



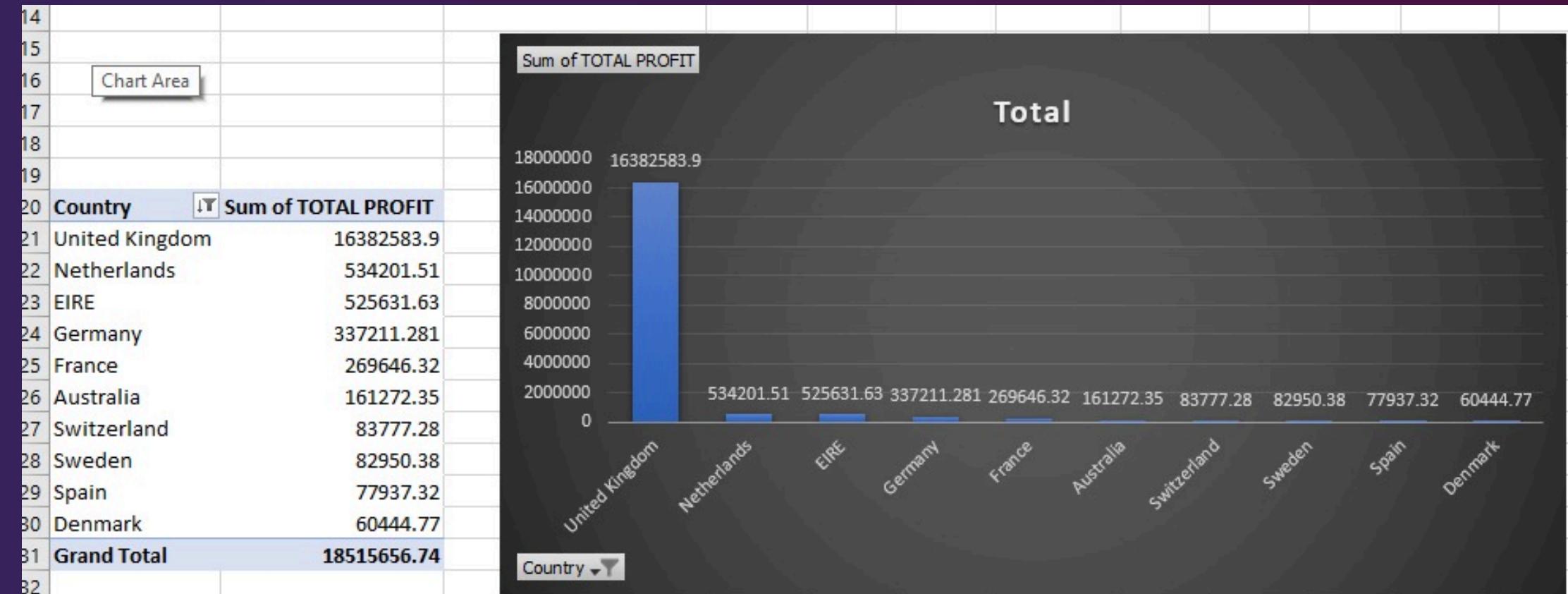
DATA VISUALIZATION:

Excel:



Average price by product

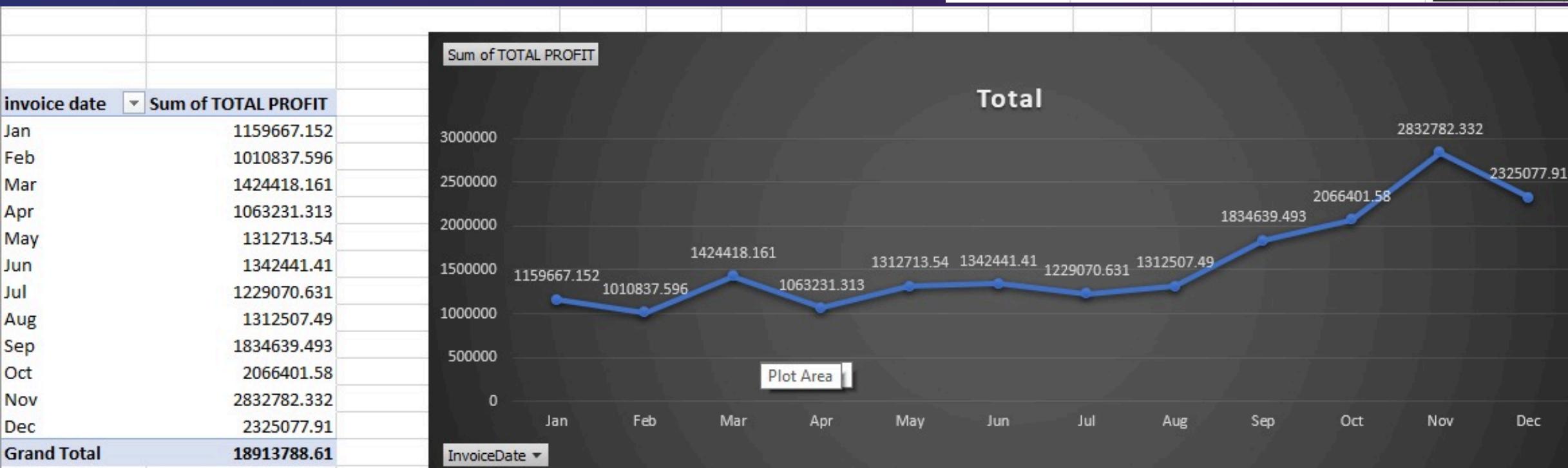
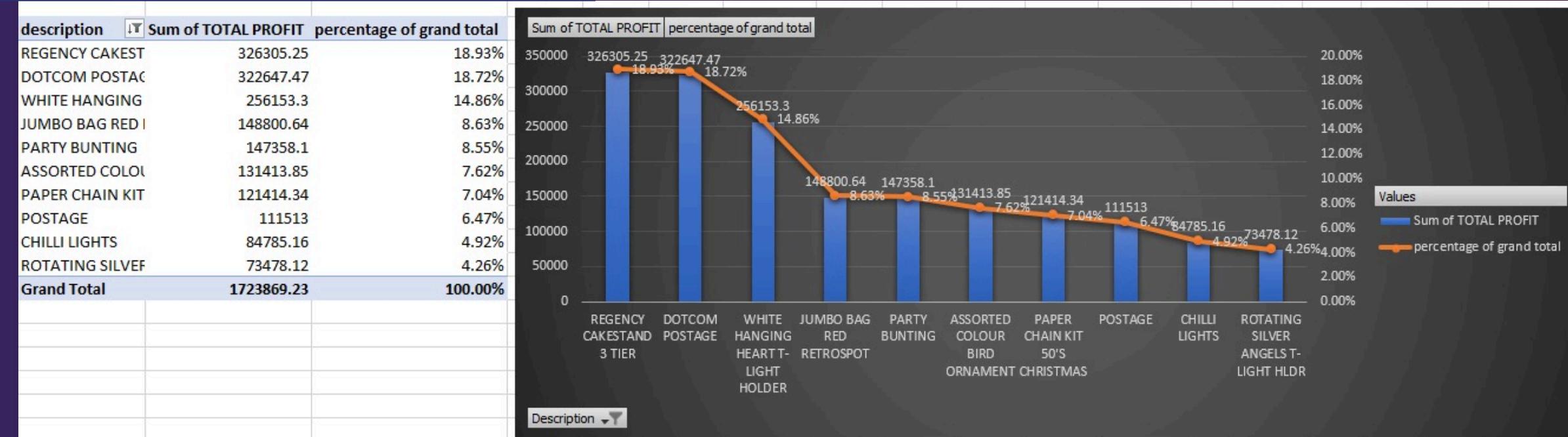
Total profit by country



DATA VISUALIZATION:

Excel:

product profit contribution(%)



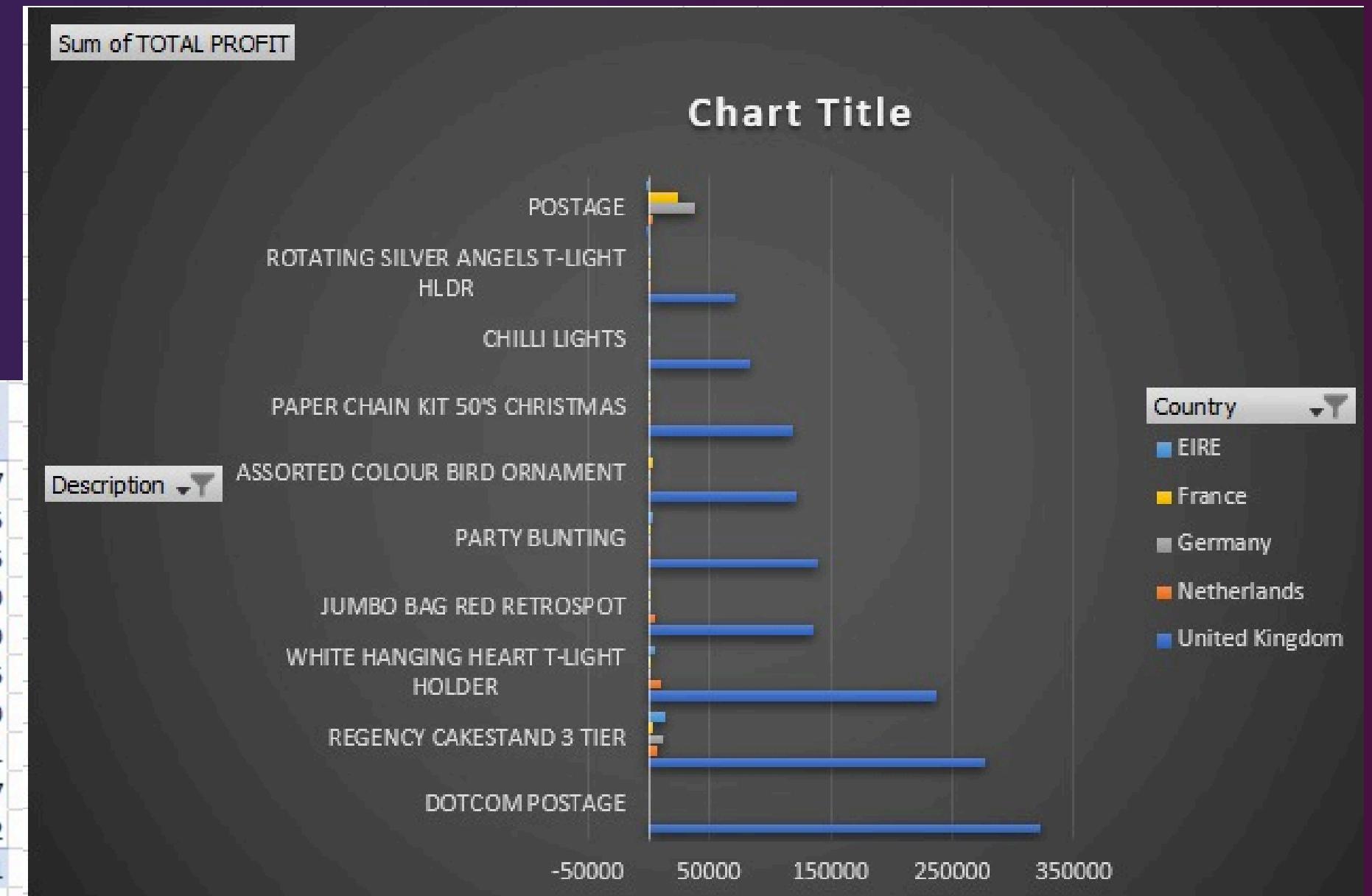
Monthly profit trend over time

DATA VISUALIZATION:

Excel:

total profit by country and product

	Sum of TOTAL PR Column Labels					
	United Kingdom	Netherlands	Germany	France	EIRE	Grand Total
209 DOTCOM POSTAC	322647.47					322647.47
210 REGENCY CAKEST	277436.6		6805.35	12414.15	3270.3	14071.35
211 WHITE HANGING	237531.15		10603.1	123.9	249.75	5100.45
212 JUMBO BAG RED I	136493.45		5613	1638.26	1454.47	395.72
213 PARTY BUNTING	139077.35		666.5	388.5	861	2862.55
214 ASSORTED COLOU	122311.97		1651.04	540.8	3295.88	784.16
215 PAPER CHAIN KIT	118881.79		413.9	88.5	112.1	1099.2
216 CHILLI LIGHTS	83751.21			19.8		834.5
217 ROTATING SILVEF	71129.27		252	321.3	35.7	275.4
218 POSTAGE	-1787.68		3837	37976	24525	-25
219 Grand Total	1507472.58		29841.89	53511.21	33804.2	25398.33
						1650028.21



DATA VISUALIZATION:

3-tableau:



Conclusion

AFTER COMPLETING THE FULL DATA ANALYTICS WORKFLOW—INCLUDING DATA CLEANING, DATA MODELING, DAX MEASURES CREATION, AND BUILDING AN INTERACTIVE POWER BI DASHBOARD—THE E-COMMERCE RETAIL SALES PROJECT SUCCESSFULLY TRANSFORMED RAW TRANSACTIONAL DATA INTO ACTIONABLE BUSINESS INTELLIGENCE.

THROUGH COMPREHENSIVE DATA CLEANING, WE ELIMINATED DUPLICATED INVOICES, FIXED DATE FORMATS, STANDARDIZED PRODUCT CODES, AND ENSURED CONSISTENT CUSTOMER AND COUNTRY FIELDS. THIS IMPROVED DATA RELIABILITY AND SIGNIFICANTLY ENHANCED THE ACCURACY OF CALCULATIONS AND INSIGHTS.

USING THE DATA MODEL AND CUSTOM DAX MEASURES, WE WERE ABLE TO CALCULATE KEY KPI'S SUCH AS TOTAL REVENUE, TOTAL ORDERS, AVERAGE ORDER VALUE, RETURN RATE, AND LOST REVENUE. THESE METRICS PROVIDED A SOLID FOUNDATION FOR UNDERSTANDING OVERALL BUSINESS PERFORMANCE AND IDENTIFYING THE SOURCES OF INEFFICIENCY SUCH AS HIGH RETURN VOLUMES.

ANY
QUESTIONS?



project team + Roles

Team Meambers

Hager wael : Data transformation & Introduction

Yosr Ahmed : Data Extraction & Cleaning

Jessica Elkess Bishouy : PowerBi,Daxes,visualization,insights

Sherry Atef : Pivot tables & visualization using excel

Marina Bassem : Data visualization using tableau



THANK YOU!

DATA ANALYSIS IS KEY TO BUSINESS
GROWTH AND SUCCESS!