SCTP
Associate Data Analyst
Capstone Project

Trainer : Swapnil Pandey

Trainee: Low Wei Min

Date : 31 October 2025

Cohort: 43

Content

- Executive Summary / Project Introduction
- Problem Definition
- Objectives
- Workflow
- Scope of Analysis
- Analysis Report (Observations and Implications)
- Python Machine Learning
- Tools and Technologies Summary
- Conclusion and Future Work

Executive Summary

This project analyzes the sales and profitability of an e-commerce company operating in Brazil, Chile, Mexico, and Colombia since August 2019 covering wide range of products from electronics to home goods.

Using data up to December 2022, it examines trends in sales, profit, and units sold, order level and shipping delay. Power BI visualizations highlight growth patterns, customer demographics, and shipping efficiency, while Python machine learning models forecast 2023 profit.

The analysis provides insights into product and country performance, customer behavior, and operational factors, supporting data-driven strategies for sustainable growth.

Problem Statement

Since its inception in August 2019, the e-commerce company had experienced strong initial growth across multiple Latin American markets — Brazil, Chile, Mexico, and Colombia — selling a diverse portfolio of products ranging from electronic goods to home furnishings. However, sales and profitability trends had been stagnant in recent years, particularly around the COVID-19 pandemic period.

Although the company achieved strong growth in 2020, its profit levels had since stablized, raising concerns about post-pandemic market conditions, shifting customer preferences, and operational efficiency (e.g., shipping delays). The management seeks to understand the key factors influencing sales and profitability across countries and product categories, and to forecast future performance to support data-driven strategic decisions.

Project Objectives

Sales and Profitability Analysis

- Analyze yearly trends in units sold, total sales, and profit from 2019 to 2022.
- Compare growth rates across countries (Brazil, Chile, Mexico, Colombia) and product categories (home & living, electronics & gadgets and fashion).

Price and Product Insights

- Examine product-level statistics (median, minimum, and maximum selling prices) for the products.
- Identify which products and price ranges drive the highest profitability.

Operational Performance

- Assess shipping performance by analyzing the delivery delays.

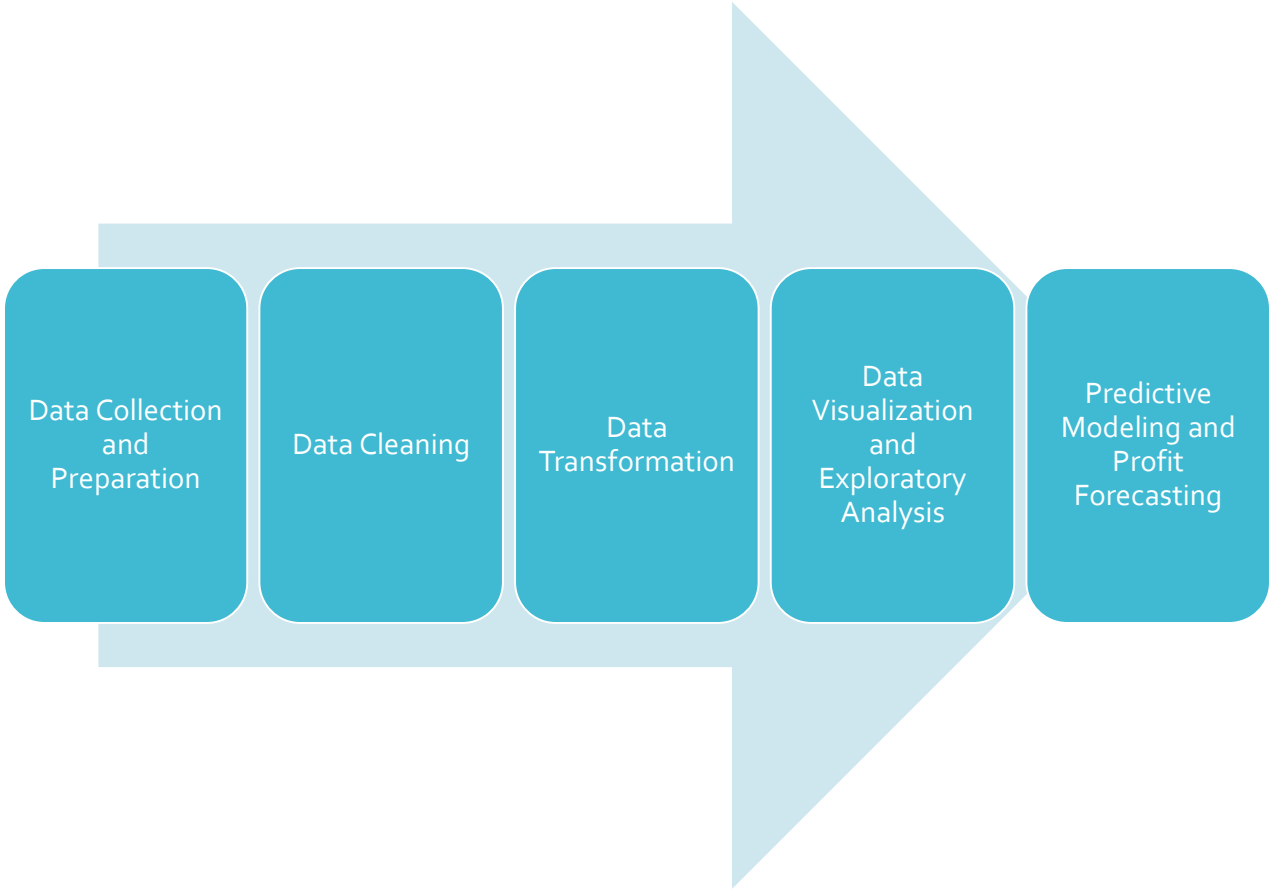
Customer Demographics and Behavior

- Profile customers based on age, income, gender, and country.
- Explore how demographic factors correlate with purchasing patterns.

Profit Forecasting and Post-Pandemic Impact

- Develop a profit forecast for 2023 using Python-based machine learning models.

Workflow


Data Collection and Preparation
Data Cleaning
Data Transformation
Data Visualization and Exploratory Analysis
Predictive Modeling and Profit Forecasting

Data Collection and Preparation

File (csv) : customer

Column Name	Description	Type	Unit
customer_id	Unique ID of the customers.	STRING	-
first_name	First name of the customer.	STRING	-
last_name	Last name of the customer.	STRING	-
gender	Gender of the customer.	STRING	-
age	Age of the customer.	INT	YEARS
country	The country names the customer is from.	STRING	-
income	Income of the customer	INT	-

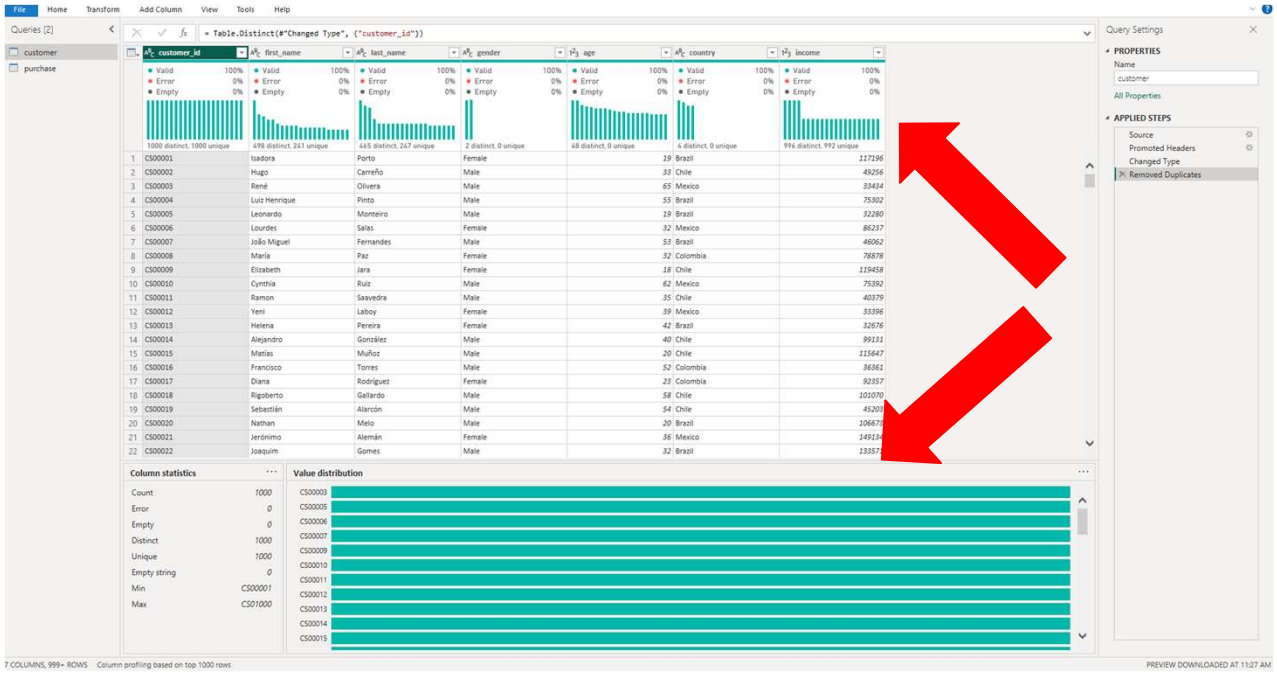
File (csv) : purchase

Column Name	Description	Type	Unit
order_id	Unique order id of the order placed.	STRING	-
customer_id	Unique ID of the customer.	STRING	-
product_name	Product name.	STRING	-
description	The description about the product.	STRING	-
price	Price of the product.	DECIMAL	-
discount	The discount rate on the product.		
tax	Applicable tax on the product.	PERCENTAGE	-
order_date	The date on which order was placed.	DATE	-
quantity	Quantity of the product being ordered.	INT	-
shipping_cost	The cost of shipping.	DECIMAL	-
shipping_date	The date of shipment.	DATE	-

The datasets cover period from 15 August 2019 to 1 January 2023

Data Cleaning

Use data profiling features in Power Query to understand characteristics and quality of source data



Customer dataset

	customer_id	first_name	last_name	gender	age	country	income
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	1000 distinct, 1000 unique	498 distinct, 241 unique	465 distinct, 247 unique	2 distinct, 0 unique	48 distinct, 0 unique	4 distinct, 0 unique	996 distinct, 992 unique
1	CS00001	Isadora	Porto	Female	19	Brazil	117196
2	CS00002	Hugo	Carreño	Male	33	Chile	49256
3	CS00003	René	Olivera	Male	65	Mexico	33434

Column statistics	
Count	1000
Error	0
Empty	0
Distinct	1000
Unique	1000
Empty string	0
Min	CS00001
Max	CS01001

Column statistics	
Count	1000
Error	0
Empty	0
Distinct	997
Unique	994
Empty string	0
Min	Abelard...
Max	Úrsula C...

Column statistics	
Count	1000
Error	0
Empty	0
Distinct	2
Unique	0
Empty string	0
Min	Female
Max	Male

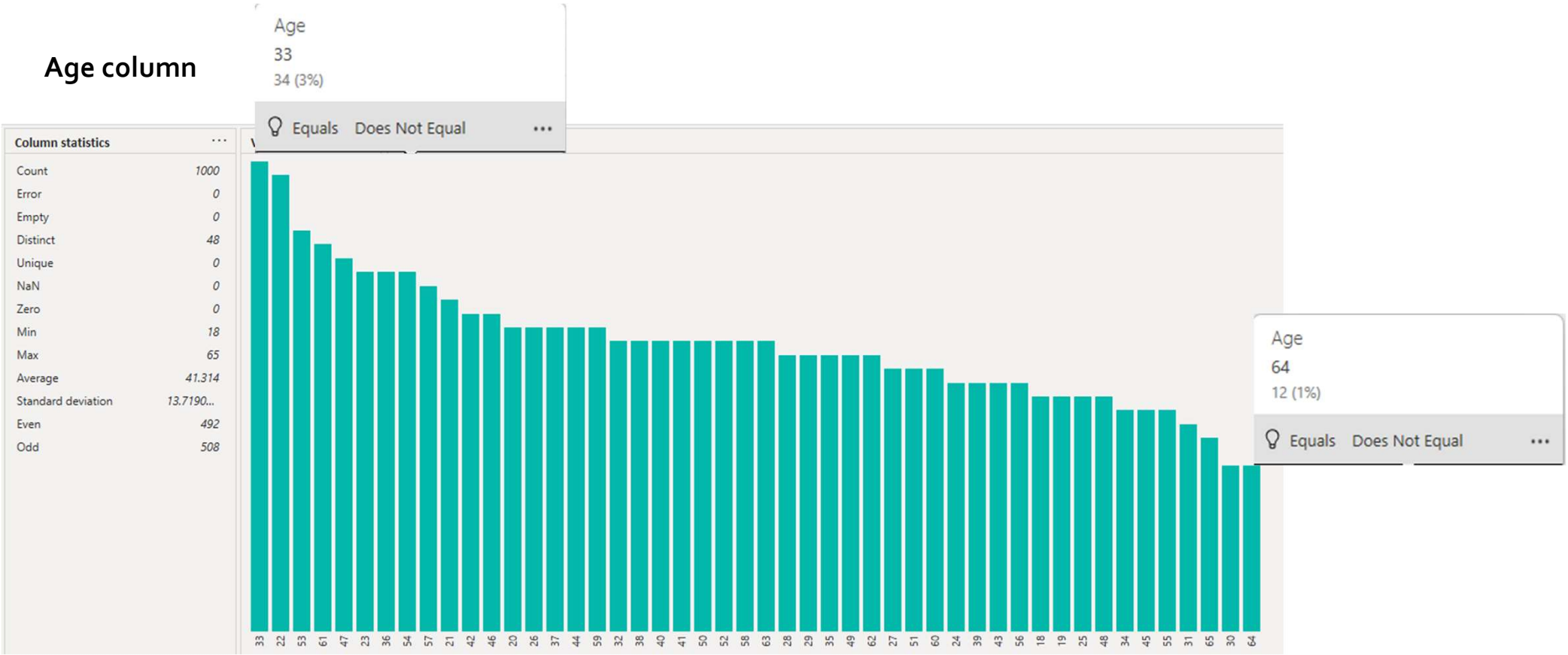
Column statistics	
Count	1000
Error	0
Empty	0
Distinct	2
Unique	0
Empty string	0
Min	Female
Max	Male

Column statistics	
Count	1000
Error	0
Empty	0
Distinct	48
Unique	0
NaN	0
Zero	0
Min	18
Max	65
Average	41.339
Standard deviation	13.7394...
Even	491
Odd	509

Column statistics	
Count	1000
Error	0
Empty	0
Distinct	4
Unique	0
Empty string	0
Min	Brazil
Max	Mexico

Column statistics	
Count	1000
Error	0
Empty	0
Distinct	996
Unique	992
NaN	0
Zero	0
Min	20205
Max	149980
Average	84577.982
Standard deviation	37482.2...

Age column



3% 2% 1%

Country Column

	customer_id	first_name	last_name	gender	age	country	income
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	1000 distinct, 1000 unique	498 distinct, 242 unique	465 distinct, 248 unique	2 distinct, 0 unique	48 distinct, 0 unique	4 distinct, 0 unique	996 distinct, 992 unique
1	CS00078	Abelardo	Vergara	Female	35	Mexico	107684
2	CS00278	Abelardo	Ramírez	Male	45	Chile	117771
3	CS00991	Abigail	Riojas	Male	55	Mexico	38899
4	CS00419	Abraham	Chacón	Male	58	Mexico	32735

Column statistics

Count	1000
Error	0
Empty	0
Distinct	4
Unique	0
Empty string	0
Min	Brazil
Max	Mexico

Value distribution

country
Colombia
280 (28%)
Equals Does Not Equal

country
Chile
255 (25%)
Equals Does Not Equal

country
Brazil
234 (23%)
Equals Does Not Equal

country
Mexico
231 (23%)
Equals Does Not Equal

Income column

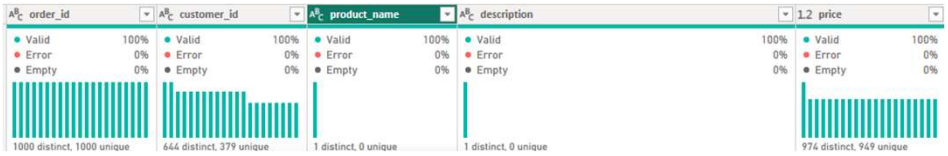
Column statistics

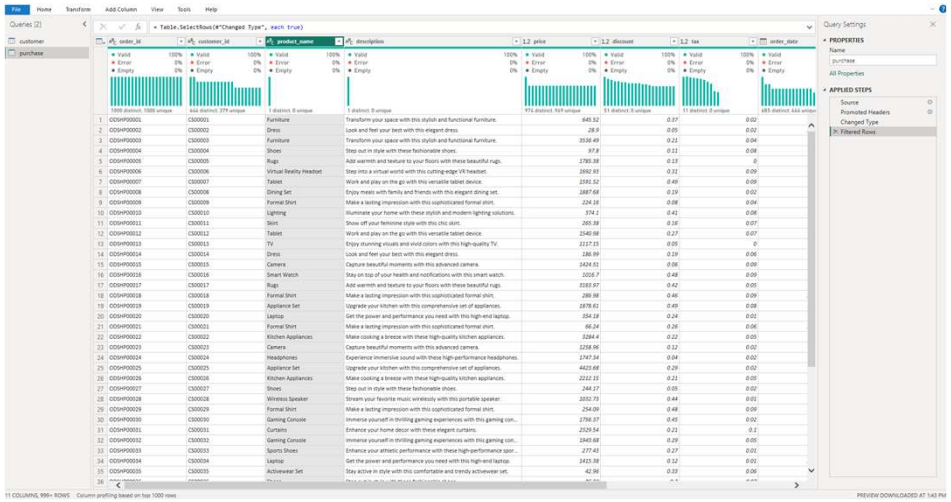
Count	1000
Error	0
Empty	0
Distinct	996
Unique	992
NaN	0
Zero	0
Min	20205
Max	149980
Average	84577.982
Standard deviation	37482.2...
Even	506
Odd	494

Customer dataset

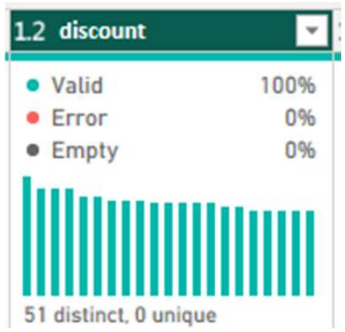
Column name	Findings	What it means? Insights…
All columns	• The data across all columns shows : ➢ Valid 100% ➢ Error 0% ➢ Empty 0%	• Data is valid, no error and no empty record.
customer id	• 1000 distinct and 1000 unique records	• No duplicate record.
first name last name	• First name shows 498 sets of distinct records, with 242 sets of unique records • Last name shows 465 sets of distinct records, with 248 sets of unique records	• There are persons with same names. **Improvement needed**: To merge first name and last name together to create a column 'full name'. It provides clarity on who is the customer.
gender	• 2 distinct sets of data, namely Female and Male.	• There are 505 male customers and 495 female customers. • 50% Male customers and 50% Female customers.
age	• 48 distinct ages. • The customer age ranges from 18 years old (Min – youngest) to 65 years old (Max - oldest). • The average customer's age is 41 years old.	• The value distribution shows that there is a wide spread of customers across different age.
country	• The customers are located in 4 countries : Columbia, Chile, Brazil and Mexico.	• The number of customers in each country is about the same. Columbia has slightly more customers than the rest.
income	• The income of customers ranges from $20,205 (Min) to $149,980.	• There is a fair spread of customers across different income levels.

Purchase Dataset



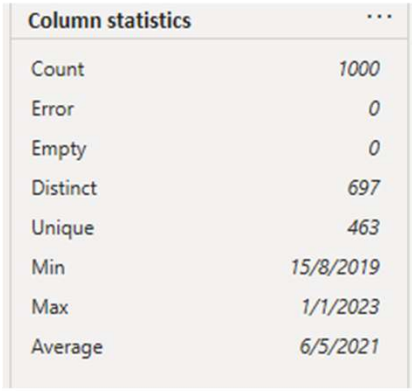


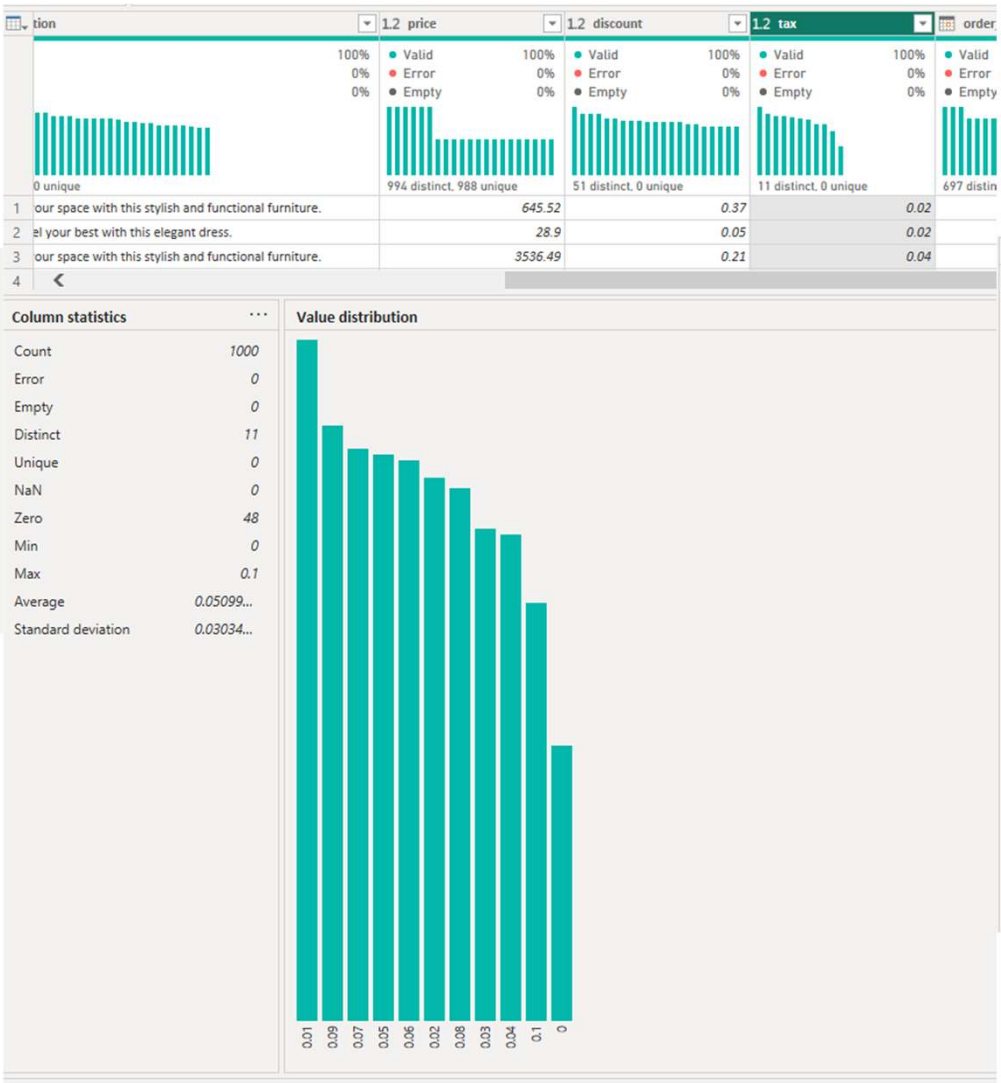
Discount column

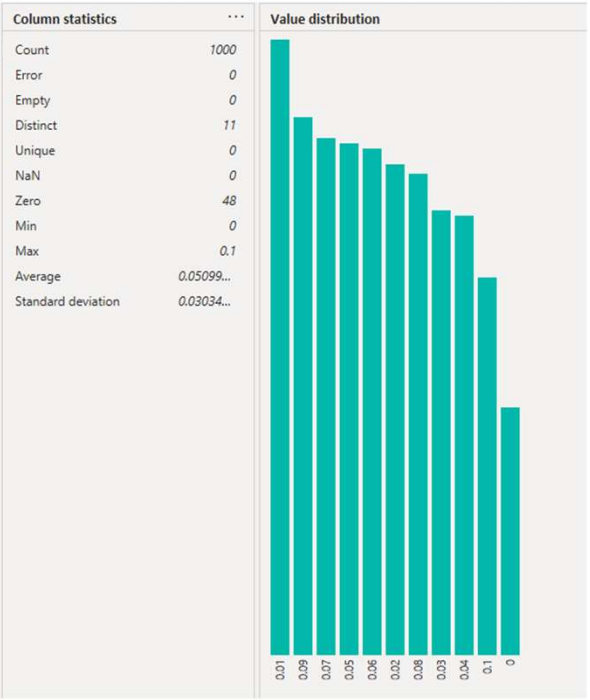


Column statistics	
Count	1000
Error	0
Empty	0
Distinct	51
Unique	0
NaN	0
Zero	6
Min	0
Max	0.5
Average	0.24353...
Standard deviation	0.14498...

Tax column


Column statistics
Count 1000
Error 0
Empty 0
Distinct 697
Unique 463
Min 15/8/2019
Max 1/1/2023
Average 6/5/2021


tion
1.2 price
1.2 discount
1.2 tax
order
100% Valid
0% Error
0% Empty
0 unique
994 distinct, 988 unique
51 distinct, 0 unique
11 distinct, 0 unique
697 distin
1 our space with this stylish and functional furniture. 645.52 0.37 0.02
2 el your best with this elegant dress. 28.9 0.05 0.02
3 our space with this stylish and functional furniture. 3536.49 0.21 0.04
4
Column statistics
Count 1000
Error 0
Empty 0
Distinct 11
Unique 0
NaN 0
Zero 48
Min 0
Max 0.1
Average 0.05099...
Standard deviation 0.03034...
Value distribution
0.01 0.09 0.07 0.05 0.06 0.02 0.08 0.03 0.04 0.1 0


Column statistics
Count 1000
Error 0
Empty 0
Distinct 11
Unique 0
NaN 0
Zero 48
Min 0
Max 0.1
Average 0.05099...
Standard deviation 0.03034...
Value distribution
0.01 0.09 0.07 0.05 0.06 0.02 0.08 0.03 0.04 0.1 0

Date column

Column statistics	...
Count	1000
Error	0
Empty	0
Distinct	697
Unique	463
Min	15/8/2019
Max	1/1/2023
Average	6/5/2021

Quantity column

Column statistics	...
Count	1000
Error	0
Empty	0
Distinct	10
Unique	0
NaN	0
Zero	0
Min	1
Max	10
Average	5.588
Standard deviation	2.82847...
Even	514
Odd	486

Shipping cost column

Column statistics	...
Count	1000
Error	0
Empty	0
Distinct	908
Unique	822
NaN	0
Zero	0
Min	5.03
Max	50
Average	27.6099...
Standard deviation	12.9883...

Shipping date column

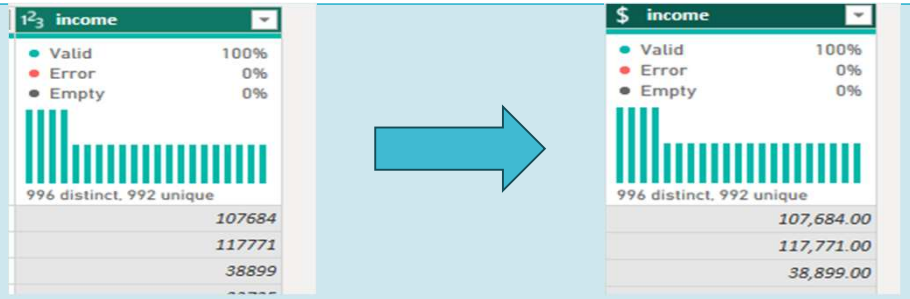
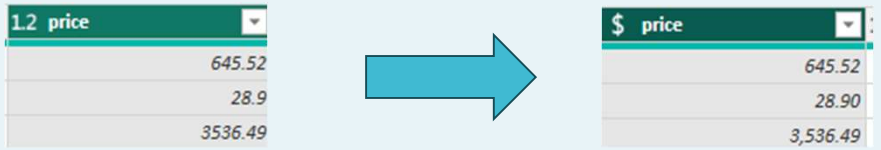
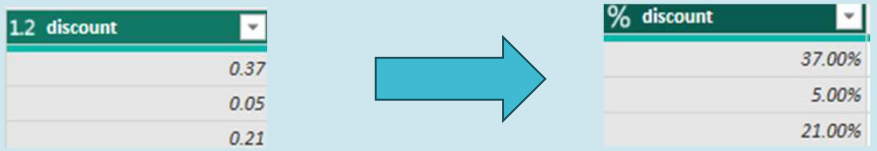
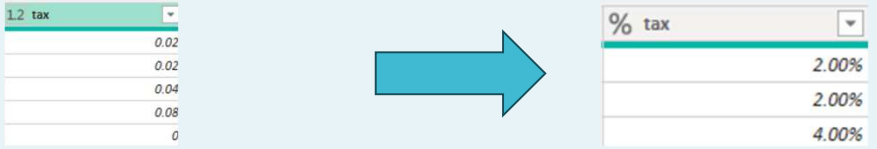
Column statistics	...
Count	1000
Error	0
Empty	0
Distinct	689
Unique	447
Min	17/8/2019
Max	13/1/2023
Average	14/5/2021

Purchase dataset

Column name	Findings	What it means? Insights…
All columns	• The data across all columns shows : ➢ Valid 100% ➢ Error 0% ➢ Empty 0%	• Data is valid, no error and no empty record.
order id	• There is 1000 orders. • The order id range from ODSHP00001 to ODSHP01000.	• No duplicate.
product name	• 30 sets of distinct product name.	• No duplicate. • **Improvement**: To create product category.
description	• 30 sets of distinct description.	• **Improvement**: To remove this column as the data is not needed for this analysis.
price	• Price of product ranges from 20.05 to 4987.1.	• **Improvement**: To change format to Currency and reduce decimal to zero.
discount	• Discount rate ranges from 0 (Min) to 0.5(Max). • The data dictionary does not have this column details.	• There is a wide spread of different discount rates on the products. • **Improvement**: - To change format to Percentage with 2 decimal places. - To add this in Data Dictionary.
tax	• There are 11 distinct sets of tax rates. • Tax rate ranges from 0 (Min) to 0.1(Max). • The Average tax rate as 0.05099, and Standard Deviation as 0.03034. • There are 48 orders with zero tax charged on the products. • A random check revealed that there are incidents where the tax rates are different for same product purchased during same year.	• Need to find out why tax rates are 0% for the 48 orders, and why the tax rates vary for same product purchase during same year. • **Improvement**: - To change format to Percentage with 2 decimal places.

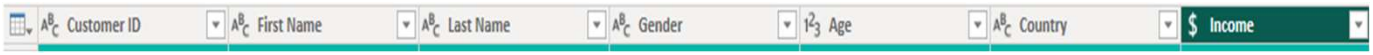
Purchase dataset

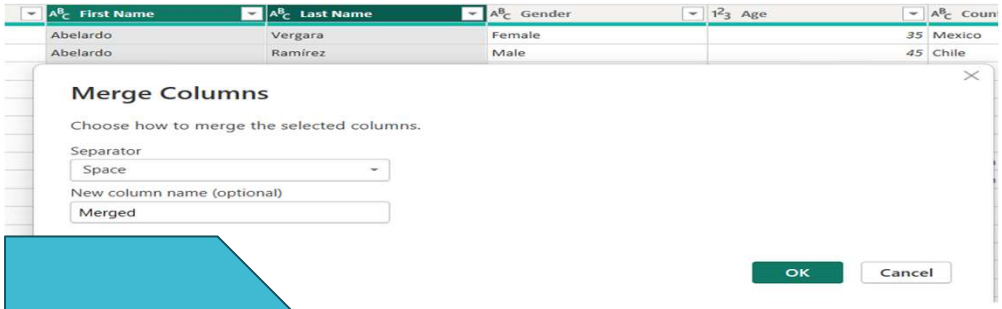
Column name	Findings	What it means? Insights…
order date	• The date ranges from 15/8/2019 to 1/1/2023. • August 2019 and January 2023 data is for partial month. • Data format is Date.	• **Improvement**: To remove 1 January 2023 data when comparing data by year, quarter or month, or when making forecast, to avoid distortion in analysis. • How to exclude 1 January 2023 data? ChatGPT suggests few options. • Option 1: use Power Query 'Transform' to do a filter to exclude 1 January 2023 data. This will remove 1 January 2023 data before modelling is done. In future, if we add full year 2023 data, we have to remove this filter. • Option 2: use Power BI, filter on the Report or Visual feature to exclude 1 January 2023 data on the specific report or visual. • Option 2 is used for this analysis.
quantity	• The units sold ranges from 1 to 10 per order.	
shipping cost	• The shipping cost per order ranges from 5.03 (Min) to 50 (Max). The average shipping cost is 27.6099.	• **Improvement**: To change format to Currency and reduce decimal to zero.
shipping date	• The shipping date data starts from 17 August 2019 and ends on 13 January 2023. • Data format is Date.	• **Improvement**: To remove 1 January 2023 data when comparing data by year, quarter or month, or when making forecast, to avoid distortion in analysis.

Column name	Original data type	Updated data type	Changes made
Income	Whole number	$ fixed decimal place	
Price	Whole number	$ fixed decimal place	
Discount	Decimal	% Percentage	
Tax	Decimal	% Percentage	
Shipping cost	Decimal	$ fixed decimal place	

Configure Customer Query:
- **Rename column**
- **Merge column**

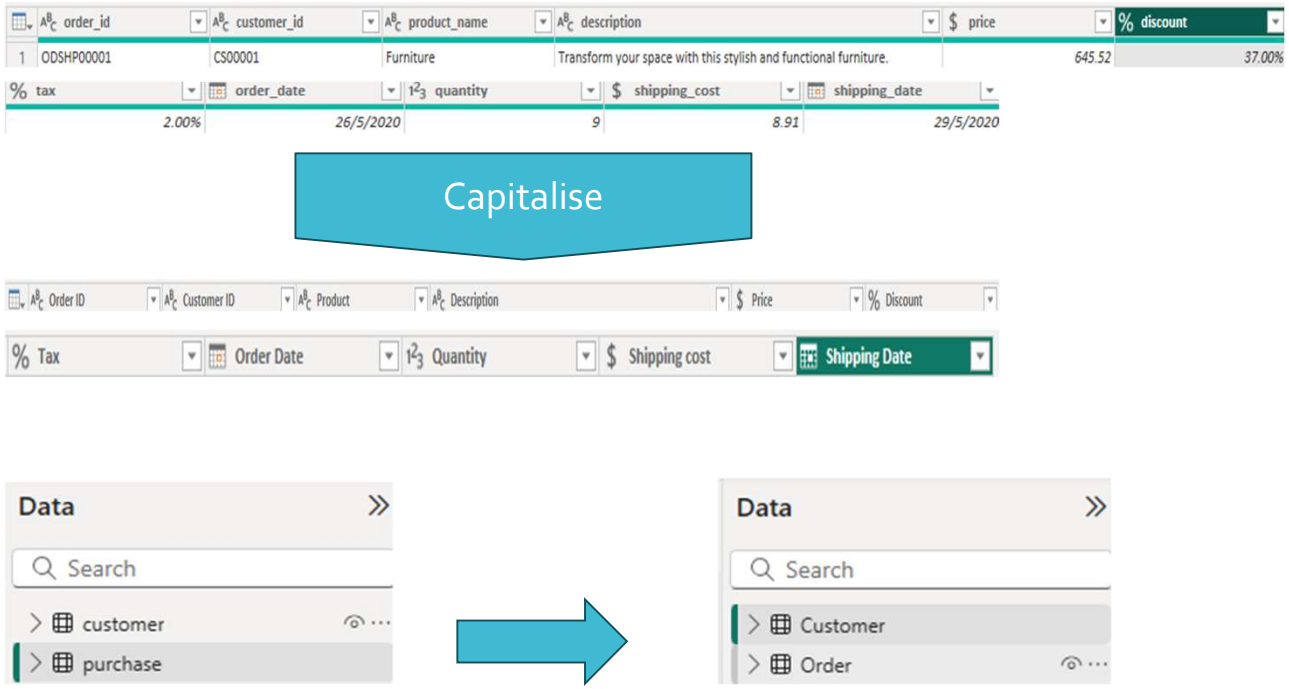
Capitalise the first alphabet





Merge First and Last Name

Configure Purchase Query:
- Rename column
- Rename table



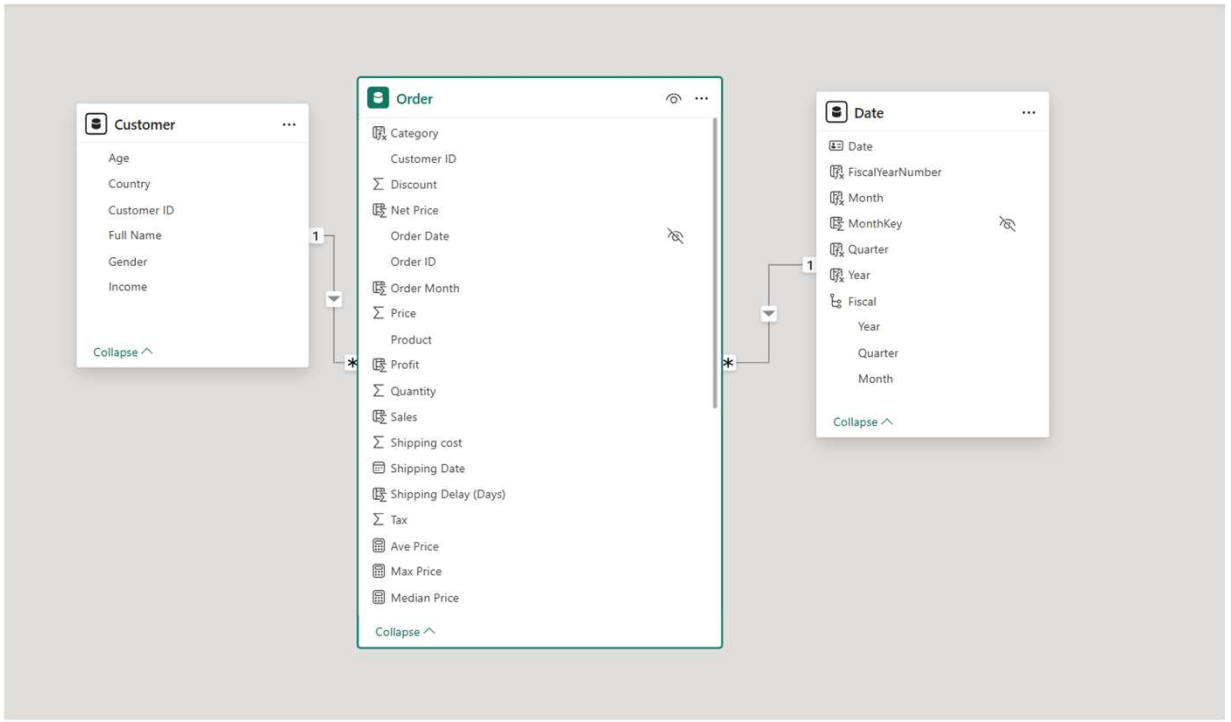
Data Transformation

- Create new measures such as price, orders, quantity (units sold), sales and profit growth/prior year and shipping delays
- Create Date Table with fiscal period (Month, Quarter, Year), Month Key and Fiscal Year Number.
- Refer to attached Snippet file (words document).
  - Date (Time Intelligence)
    - Fiscal (Year, Quarter and Month), Month Key, Fiscal Year Number
  - Quantity
  - Sales
  - Profit
  - Order
  - Shipping Delay (Days)
  - Product Category

Data Model

Relationship

Primary Key

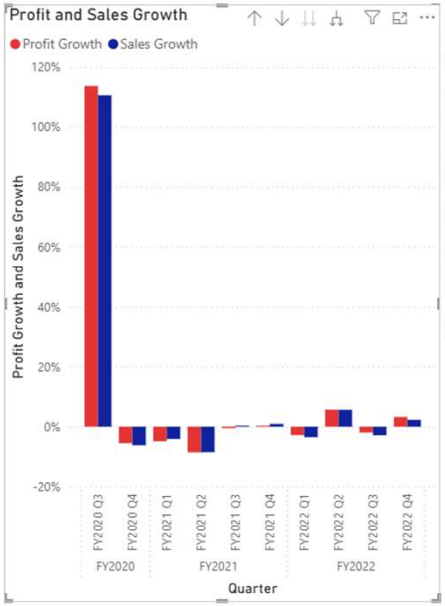


From: table (column)	Relationship	To: table (column)	Status
Order (Customer ID)	* — 1	Customer (Customer ID)	Active
Order (Order Date)	* — 1	Date (Date)	Active

Data Visualization and Exploratory Analysis

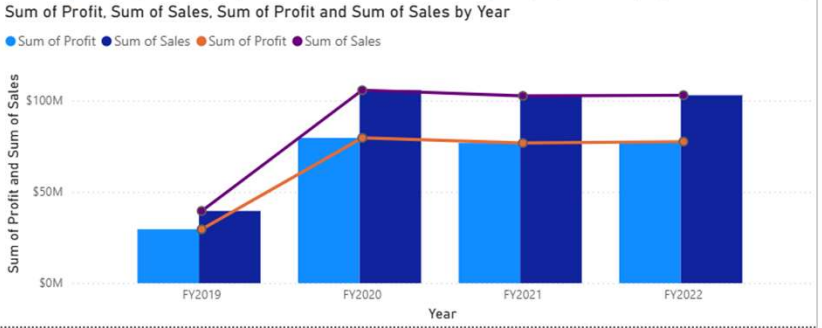
- Developed **interactive dashboards** in **Power BI** to identify patterns and trends:
  - **Sales and Profit Analysis:**
    - Yearly comparison of total sales, profit, and units sold.
    - Visualization by country and by product category.
    - Year-over-year growth rate analysis.
  - **Product-Level Pricing:**
    - Table of the 30 products with their average, minimum, and maximum selling prices.
  - **Operational Performance:**
    - Average **shipping delay days** by country and product type.
  - **Customer Demographics:**
    - Distribution of customers by **age group, gender, income range, and country**.
    - Analysis of how demographic factors influence purchasing patterns.

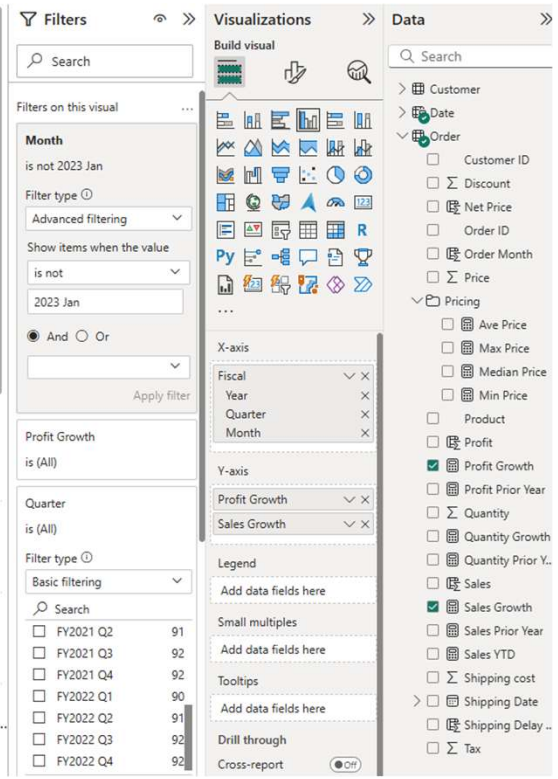
- ✓ Use either Basic or Advanced Filtering Features to remove Jan 2023 data
- ✓ Use DAX to create Sale and Profit Prior Year and Growth measures



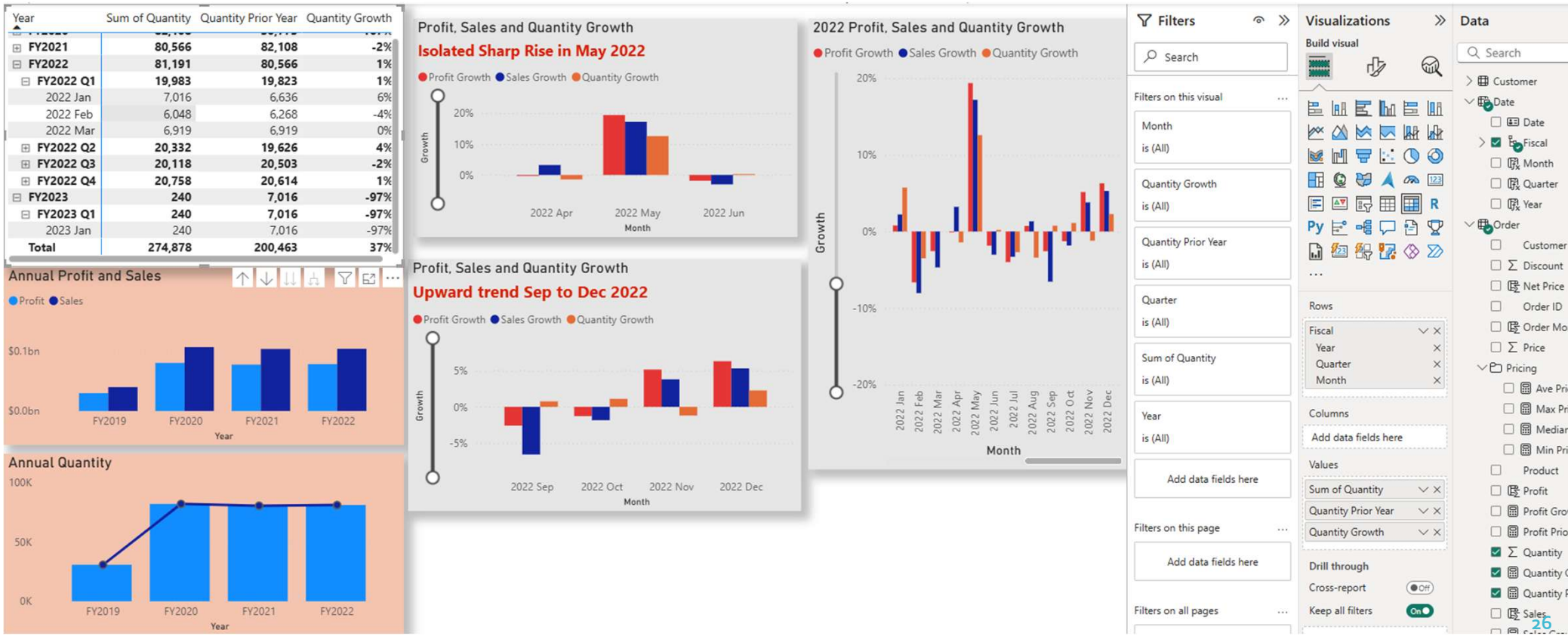
Year	Sum of Sales	Sales Prior Year	Sales Growth
FY2019	$39,545,031		
FY2019 Q3	$12,469,662		
2019 Aug	$4,632,660		
2019 Sep	$7,837,002		
FY2019 Q4	$27,075,368		
FY2020	$105,728,087	$39,545,031	167%
FY2021	$102,603,831	$105,728,087	-3%
FY2022	$102,942,654	$102,603,831	0%
FY2022 Q1	$24,812,070	$25,712,219	-4%
2022 Jan	$8,752,083	$8,562,555	2%
2022 Feb	$7,799,137	$8,482,064	-8%
2022 Mar	$8,260,850	$8,667,600	-5%
FY2022 Q2	$26,350,208	$24,945,646	6%
FY2022 Q3	$25,566,133	$26,315,985	-3%
FY2022 Q4	$26,214,243	$25,629,980	2%
FY2023	$327,219	$8,752,083	-96%
FY2023 Q1	$327,219	$8,752,083	-96%
2023 Jan	$327,219	$8,752,083	-96%
Total	$351,146,822	$256,629,031	37%

Year	Sum of Profit	Profit Prior Year	Profit Growth
FY2019	$29,514,604		
FY2019 Q3	$9,260,729		
2019 Aug	$3,465,169		
2019 Sep	$5,795,560		
FY2019 Q4	$20,253,875		
FY2020	$79,600,805	$29,514,604	169.70%
FY2021	$76,776,468	$79,600,805	-3.55%
FY2022	$77,551,569	$76,776,468	1.01%
FY2022 Q1	$18,602,850	$19,136,750	-2.79%
2022 Jan	$6,464,322	$6,414,512	0.78%
2022 Feb	$5,872,112	$6,290,633	-6.65%
2022 Mar	$6,266,416	$6,431,606	-2.57%
FY2022 Q2	$19,870,139	$18,799,696	5.69%
FY2022 Q3	$19,311,036	$19,694,862	-1.95%
FY2022 Q4	$19,767,544	$19,145,160	3.25%
FY2023	$244,893	$6,464,322	-96.21%
FY2023 Q1	$244,893	$6,464,322	-96.21%
2023 Jan	$244,893	$6,464,322	-96.21%
Total	$263,688,339	$192,356,199	37.08%



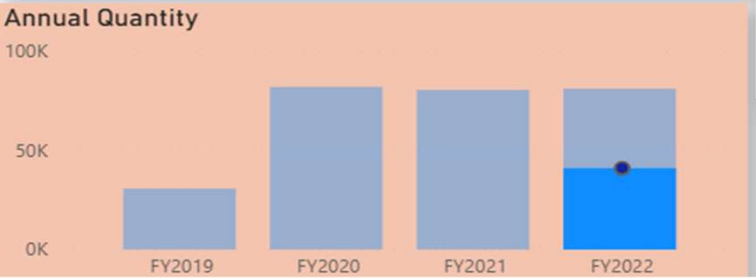


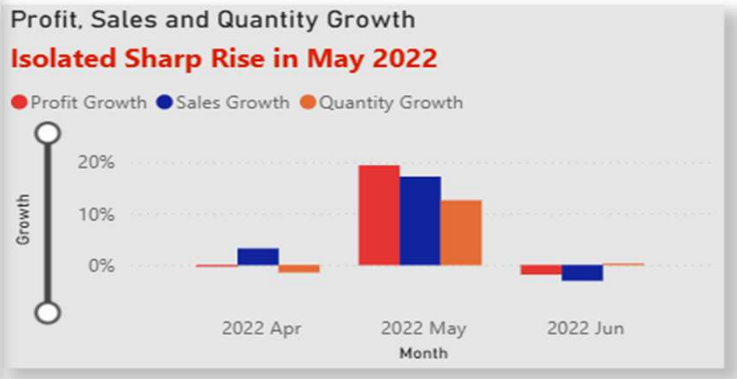
- ✓ Use either Basic or Advanced Filtering Features to exclude Jan 2023 data
- ✓ Use DAX to create Quantity Prior Year and Growth measures
- ✓ Format : Edit Title and Sub-title name to describe visual, change background colour
- ✓ Introduce Zoom slider in visual

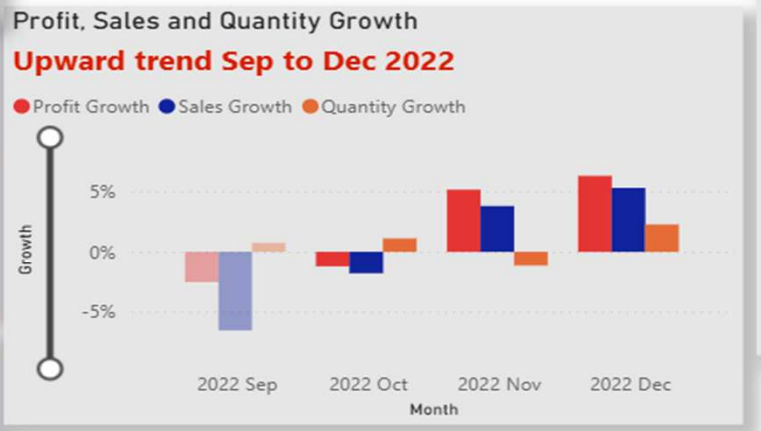


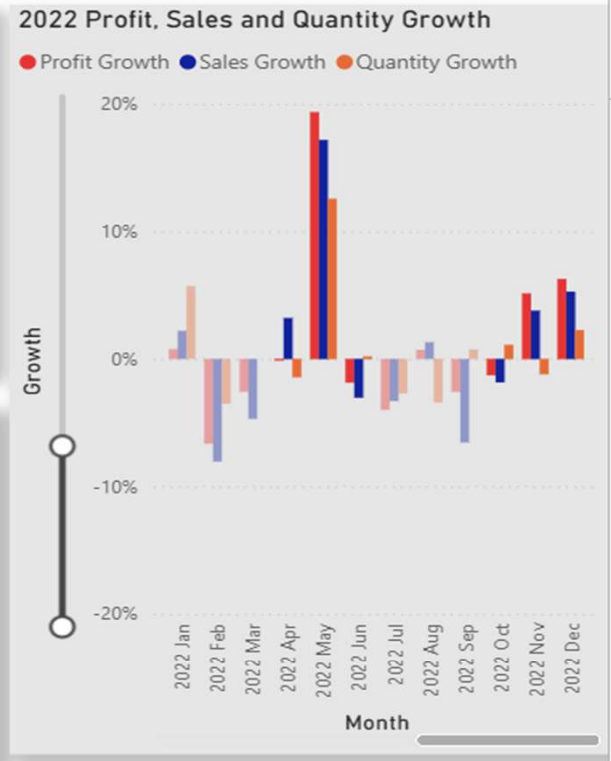
✓ Use CRTL and SELECT to highlight specific months in matrix

Year	Sum of Quantity	Quantity Prior Year	Quantity Growth
FY2022 Q1	19,983	19,823	1%
FY2022 Q2	20,332	19,626	4%
2022 Apr	6,632	6,728	-1%
2022 May	7,064	6,275	13%
2022 Jun	6,636	6,623	0%
FY2022 Q3	20,118	20,503	-2%
FY2022 Q4	20,758	20,614	1%
2022 Oct	6,973	6,897	1%
2022 Nov	6,914	6,998	-1%
2022 Dec	6,871	6,719	2%
FY2023	240	7,016	-97%
FY2023 Q1	240	7,016	-97%
Total	274,878	200,463	37%

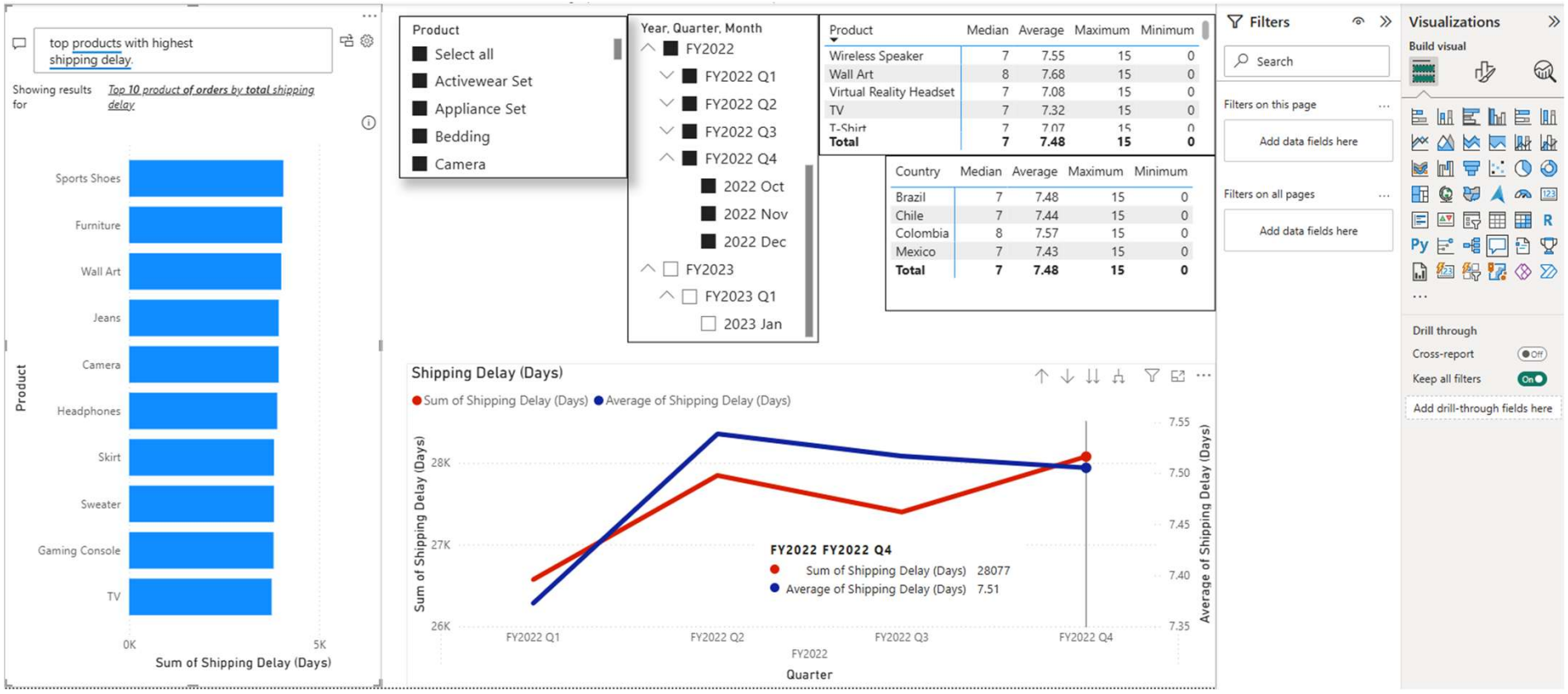








- ✓ Create Product and Year slicers
- ✓ Use Q&A visual to find out products with highest shipping delay
- ✓ Create Median, Average, Maximum and Minimum values in table.
- ✓ Use line chart to analyse the total shipping delay. Average delay is about 7.5 days.



Analysis Report

- Observations
- Implications

Scope of Analysis

- **Actual data period : 15 August 2019 to 1 January 2023**
- **Use historical actual data before 1 January 2023 for comparison and analysis purpose**
- **Definition**:

❑ **Net price** is the price after lessing discount. It is derived as follows:

Price X (Price - Discount)

❑ **Sales** is the gross revenue before deduction of discount. It is derived as follows:

Price X Quantity

❑ **Profit** is the gross revenue after lessing discount. It is derived as follows:

❑ *Net Price X Quantity*

- **Assumptions made**:

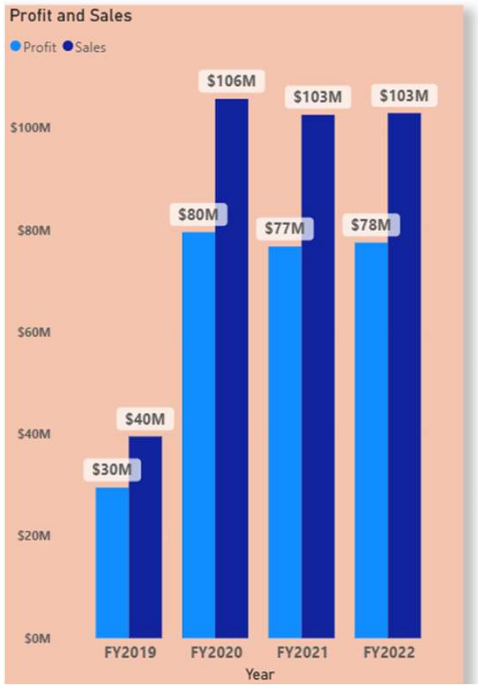
- Currency is in USD.
- Actual data given is correct. For e.g. various tax rates (including zero tax rate) and discount rates on products are correct.
- The discount data given is referring to discount rate.
- The product category and product model codes are not available.
- Price given is before discount, tax and shipping cost.
- Tax refers to the goods and service tax.
- Tax and shipping cost are payable by customers.
- Discount cost is borne by the company selling the product.
- Quantity refers to units sold.

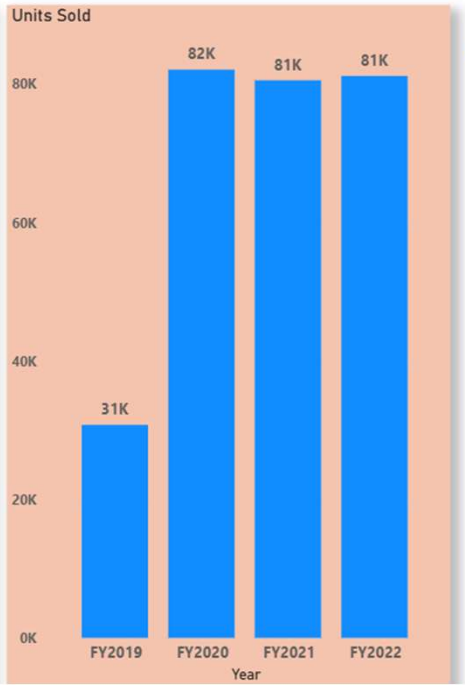
Trend Analysis (2019 – 2022)

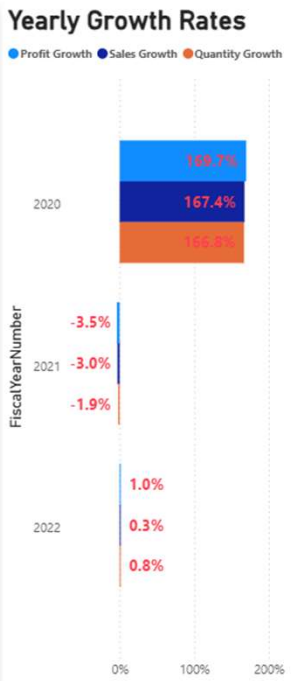
Profit , Sales and Units Sold Trend Analysis (2019 – 2022)

Observations

- The company experienced strong growth in 2020, likely accelerated by increased online shopping at the beginning of the COVID-19 pandemic.
- Profit slightly declined in 2021, witnessed by the negative growth rates.
- 2022 shows marginal recovery, but overall profits remain below the 2020 peak.
- Overall, the profit and sale trends remains stabilized from 2020 to 2022.



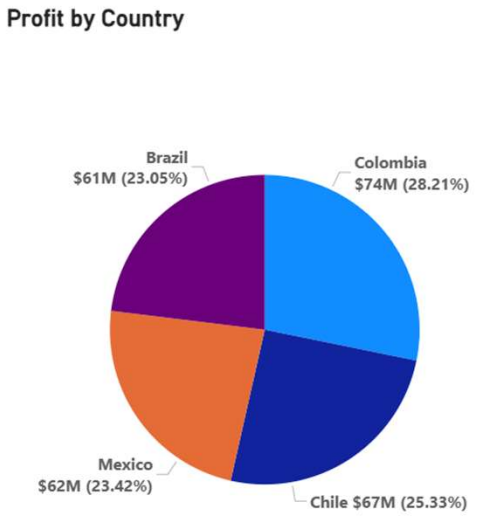


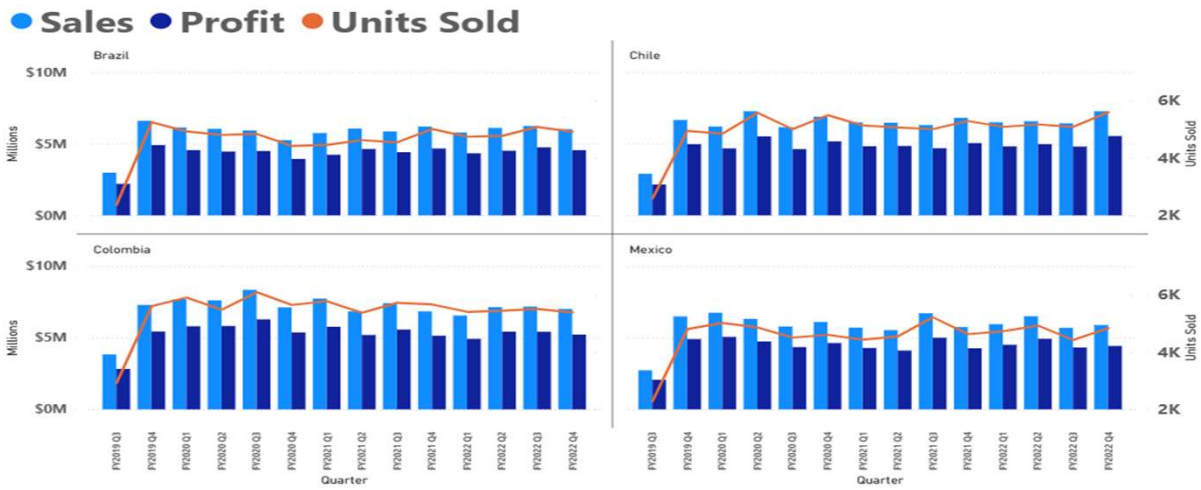


Profit , Sales and Units Sold Trend Analysis (2019 – 2022)

Country-level trends:

- The profit, sales and units sold trends and growth were similar across these countries.
- Columbia contributed to the largest share of profit, followed by Chile, Mexico and Brazil.

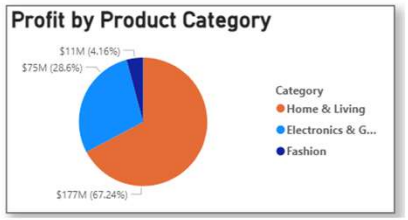


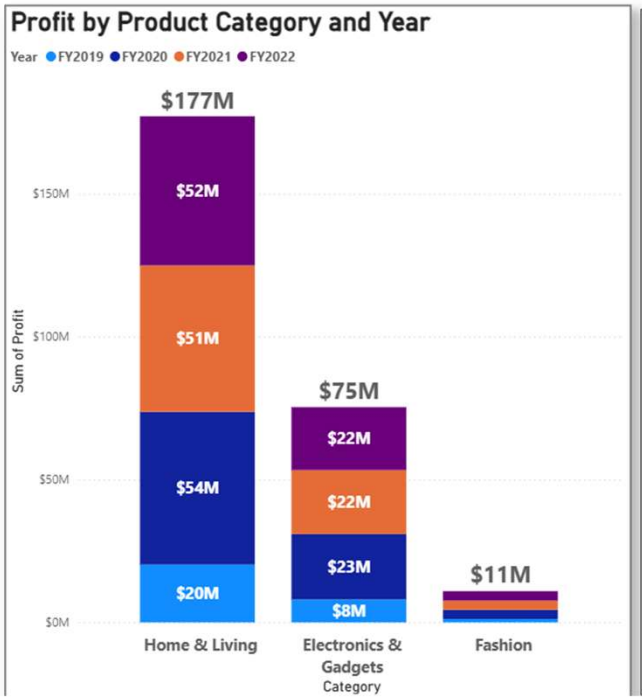


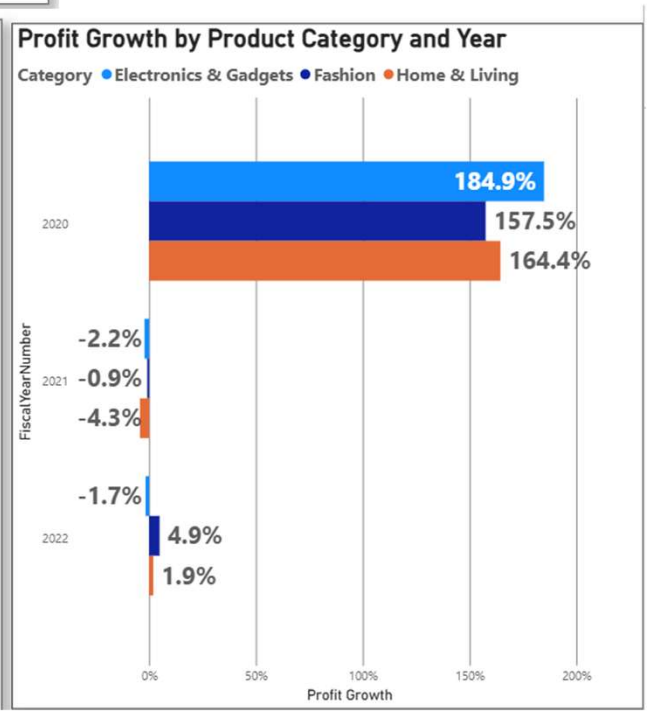
Profit , Sales and Units Sold Trend Analysis (2019 – 2022)

Product-level trends:

- Home & Living products were the top profit drivers, follow by Electronics & Gadgets and Fashion. Home & Living products contributed 67% to total profit.
- All products experienced negative growth rates in 2021, compared with prior year.
- 2022 shows a marginal recovery.



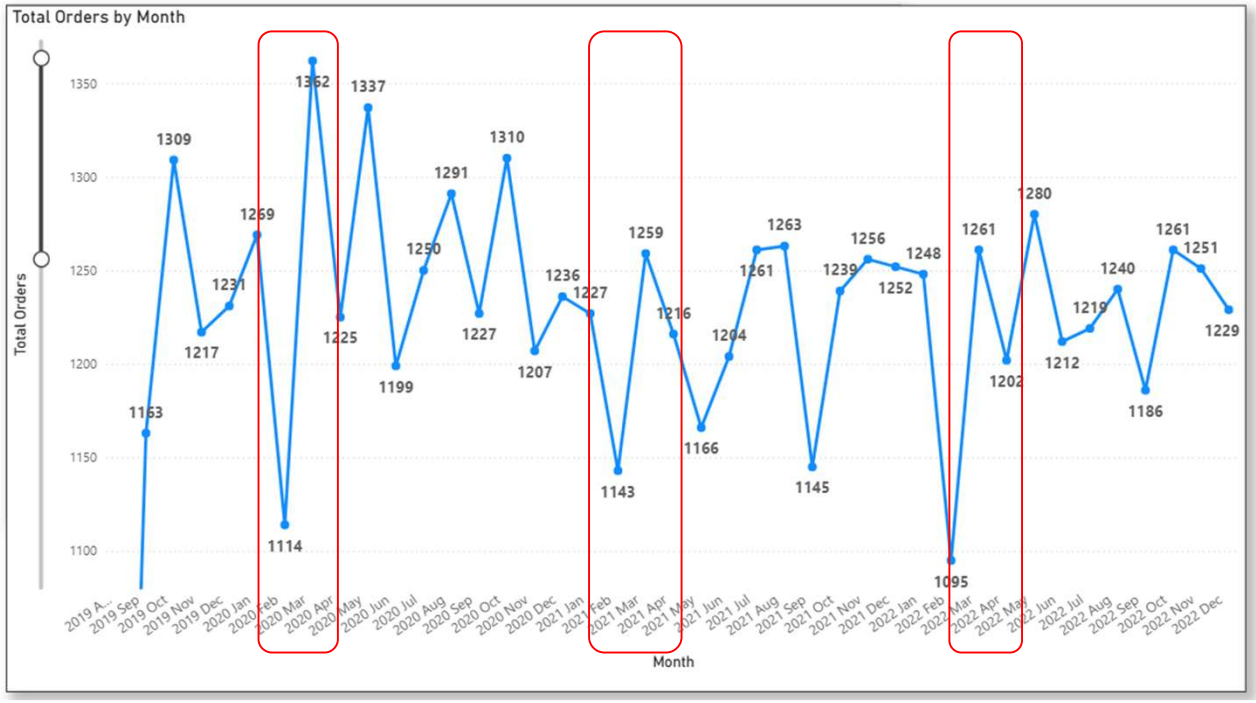




Profit , Sales and Units Sold Trend Analysis (2019 – 2022)

Order-level trends:

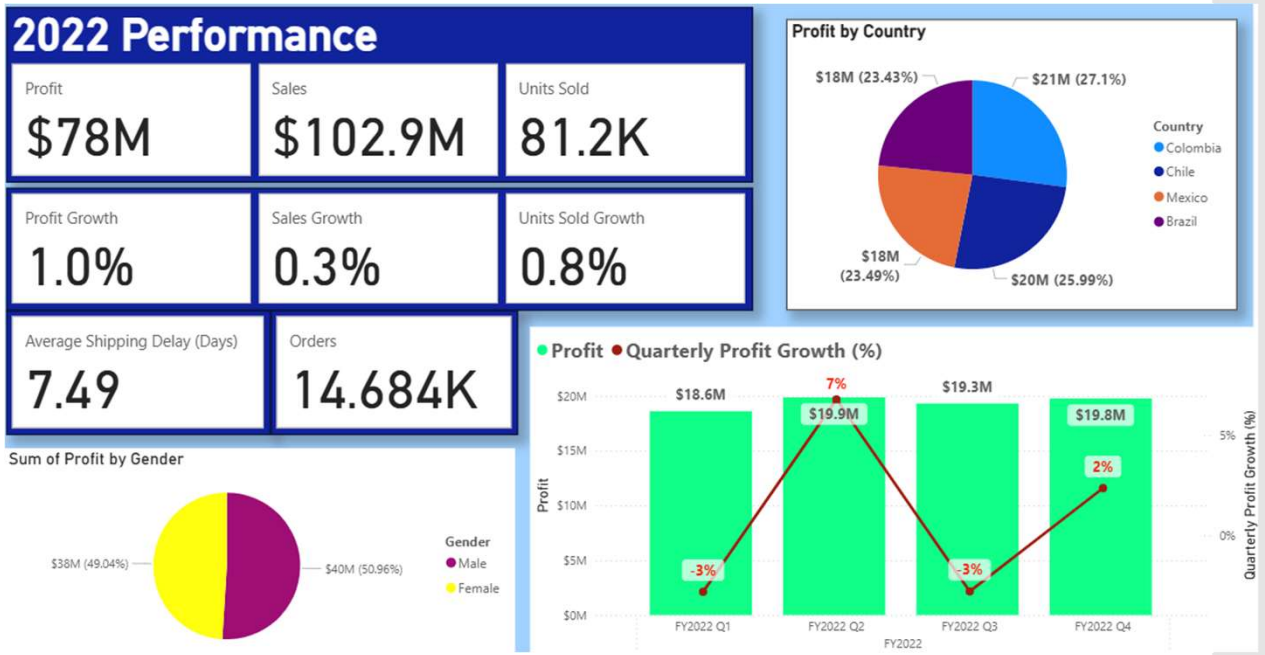
- Orders movement had been volatile. The highest orders level was 1,362 and the lowest was 1,095.
- There was always a sudden drop in orders in February and a rebounce in March each year.
- Overall, there was a downward trend in online orders.



Let’s dive into 2022

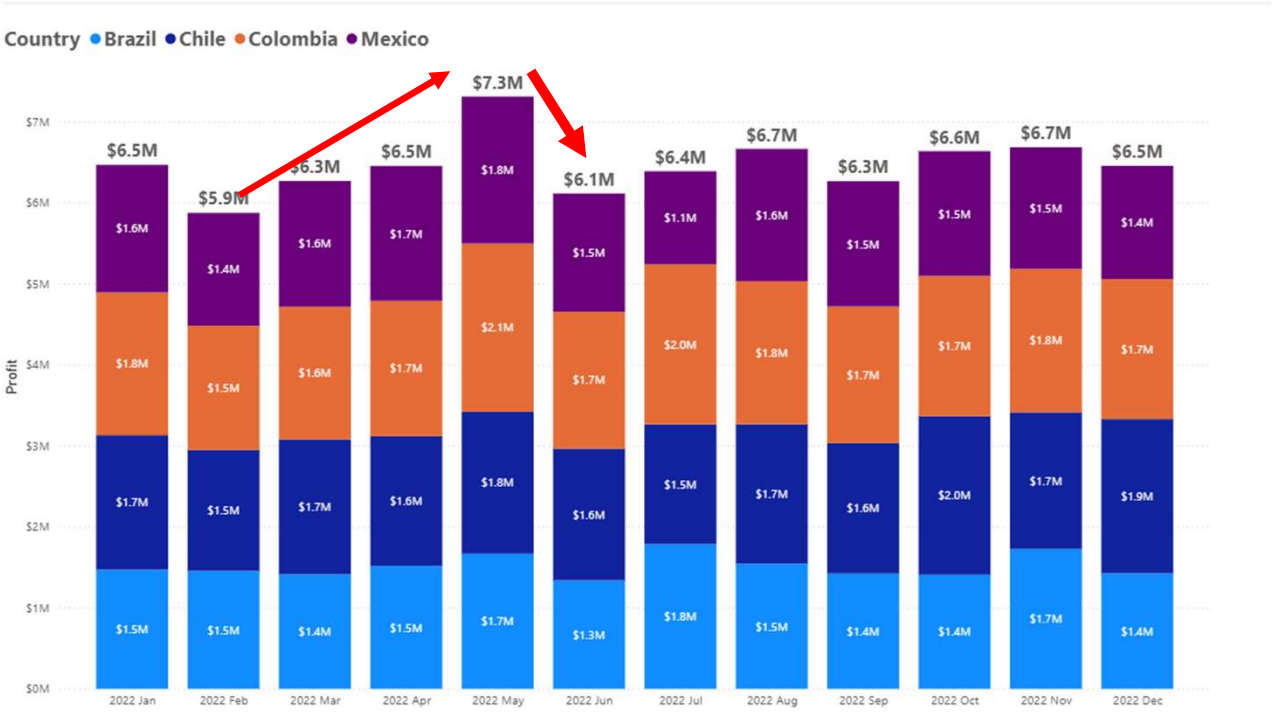
2022 Performance

- The growth compared to prior year was marginal.
- The quarterly growth rates were volatile with sharp growth in Quarter 2.
- The quarterly profit had hovered in the range from $18.6 million to $19.3 million.
- Columbia contributed the largest share of profit.
- The profit earned from male and female customers were about the same.



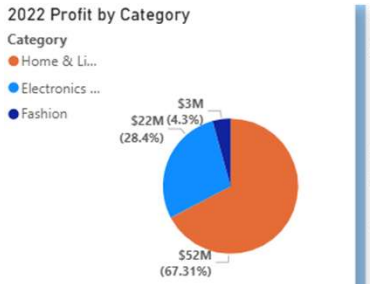
Monthly Profit Trend

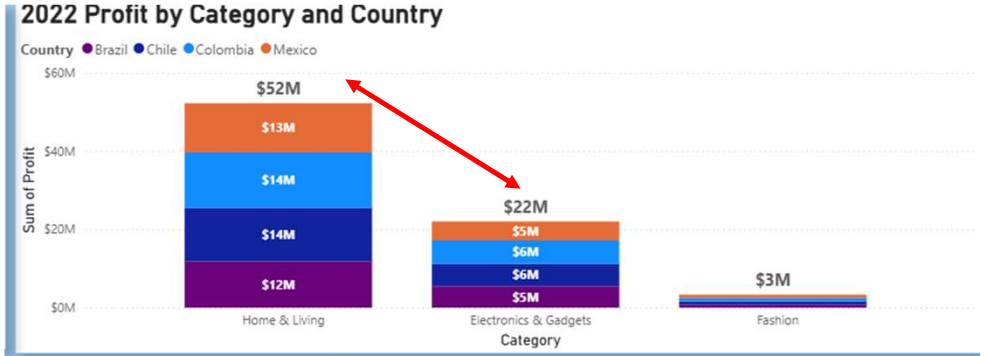
- At country level, the profit movements across the countries were synchronized.
- At month level, the monthly profit varied every month.
- The average monthly profit was $6.5 million.
- In February 2022, the profit was the lowest at $5.9 million.
- It re-bounced in March and reached its peak in May 2022 at $7.3 million.
- This was followed by a sharp plunge in June, with profit dropped by $0.8 million in June 2022.



Products

- The products were grouped into 3 categories:
  - Home & Living products
  - Electronics and gadgets
  - Fashion
- Home & Living products contributed to the largest share of profit, followed by Electronics & Gadgets and Fashion.
- The matrix shows the breakdown of profit range on each type of products:
  - Each Home & Living product type generates profit above $4.9 million up to $5.7 million.
  - Each Electronics and gadgets product type generates profit above $1.8 million to 2.5 million.
  - Each Fashion type generates profit of about $300K.
- In terms of profit growth:
  - Fashion and Home & Living products have slight growth in profit.
  - Although Electronics and gadgets products generate high profit, there is a decline in profit compared to prior year.





Category	Profit
Home & Living	$52,196,796
Curtains	$5,733,125
Wall Art	$5,465,795
Furniture	$5,413,644
Appliance Set	$5,280,567
Lighting	$5,250,940
Dining Set	$5,184,950
Rugs	$5,000,782
Kitchen Appliances	$4,987,371
Bedding	$4,978,656
Home Decor	$4,900,965

Category	Profit
Electronics & Gadgets	$22,023,411
Camera	$2,457,495
Gaming Console	$2,325,078
Headphones	$2,311,283
Laptop	$2,002,093
Smart Watch	$2,196,155
Smartphone	$1,849,250
Tablet	$2,292,650
TV	$2,135,510
Virtual Reality Headset	$2,279,884
Wireless Speaker	$2,174,014

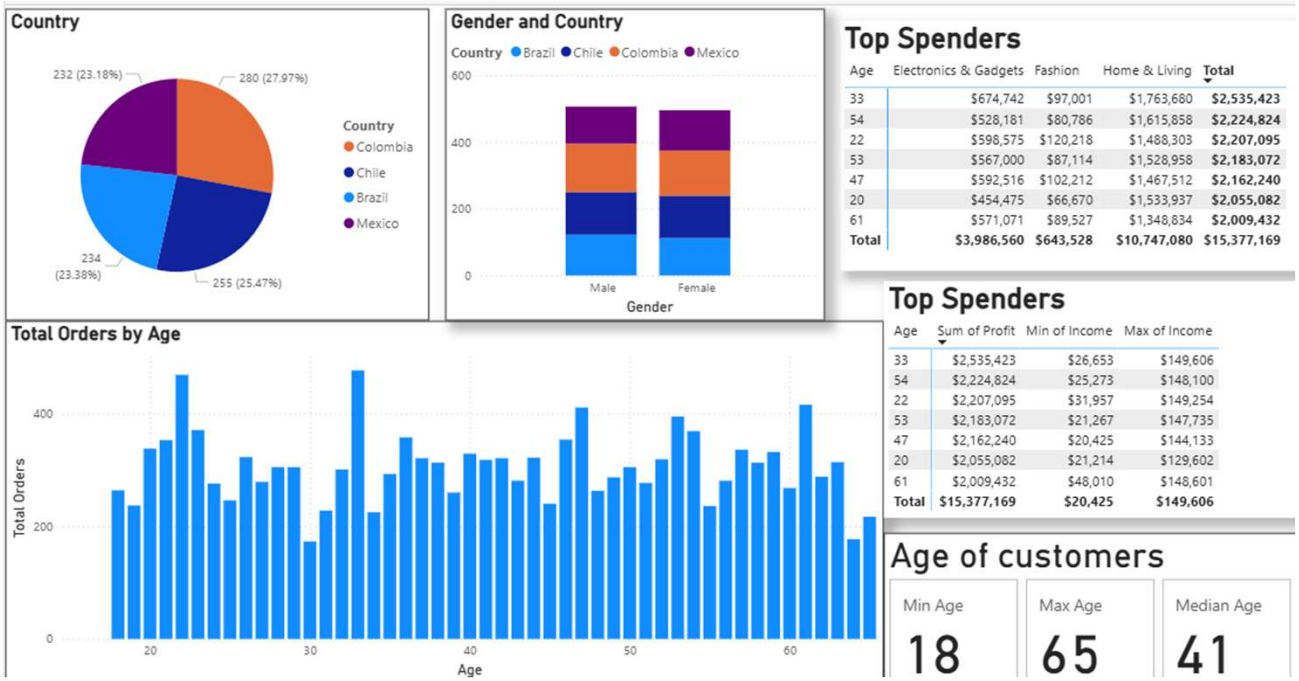
Category	Profit
Fashion	$3,331,363
Activewear Set	$364,500
Jeans	$356,644
Sports Shoes	$352,492
Shoes	$337,065
Sweater	$334,345
Jacket	$334,268
T-Shirt	$334,199
Formal Shirt	$307,709
Dress	$307,239
Skirt	$302,903

Customer Profile

- The customers were spread out evenly across the 4 countries.
- The proportion of male and female customers were almost the same in each country.
- The age of customers were from 18 years old to 65 years old. The median age is 41 years old.

Let's zoom into the Top Spenders.

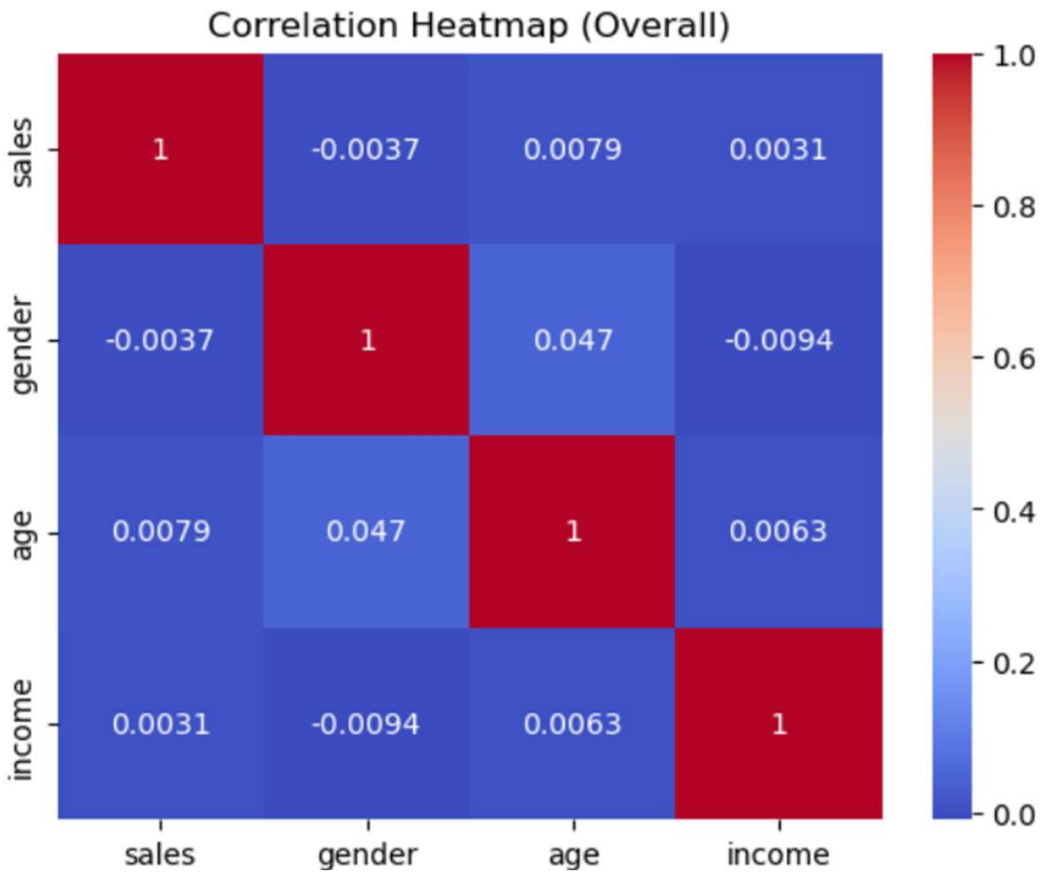
- The column chart shows that there is no correlation between age and orders.
- The matrix shows that:
  - The highest orders came from Age : 33, 54, 22, 53, 47, 20,61.
  - About more than 70% of their total purchases is spent on Home & Living
  - The customer's income level ranges from $20K to $150K.



Customer Profile

There is no correlation between sales and these variables:

- Gender
- Age
- Income



2022
Pricing and discount

Here is the price range of each product.

Category	Product	Median Price	Min Price	Max Price
Home & Living	Appliance Set	$2,438	$104	$4,992
Home & Living	Bedding	$2,426	$103	$4,987
Home & Living	Curtains	$2,687	$103	$4,995
Home & Living	Dining Set	$2,728	$130	$4,999
Home & Living	Furniture	$2,563	$109	$4,988
Home & Living	Home Decor	$2,401	$108	$4,982
Home & Living	Kitchen Appliances	$2,601	$107	$4,991
Home & Living	Lighting	$2,730	$114	$4,987
Home & Living	Rugs	$2,566	$124	$4,991
Home & Living	Wall Art	$2,427	$106	$4,983
Total		**$2,548**	**$103**	**$4,999**

Category	Product	Median Price	Min Price	Max Price
Fashion	Activewear Set	$172	$20	$299
Fashion	Dress	$167	$21	$300
Fashion	Formal Shirt	$157	$20	$298
Fashion	Jacket	$165	$20	$300
Fashion	Jeans	$160	$20	$300
Fashion	Shoes	$160	$21	$300
Fashion	Skirt	$155	$22	$299
Fashion	Sports Shoes	$162	$20	$299
Fashion	Sweater	$158	$20	$300
Fashion	T-Shirt	$164	$20	$299
Total		**$162**	**$20**	**$300**

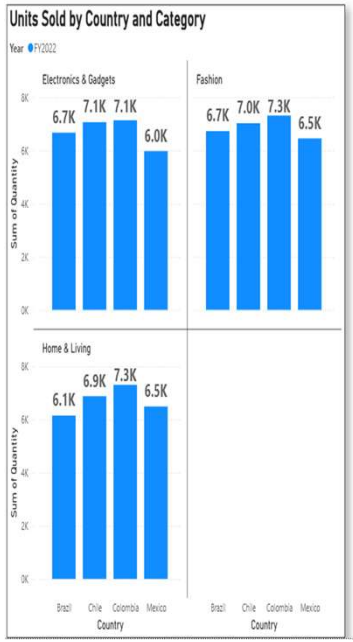
Category	Product	Median Price	Min Price	Max Price
Electronics & Gadgets	Camera	$1,144	$206	$2,000
Electronics & Gadgets	Gaming Console	$1,118	$205	$1,997
Electronics & Gadgets	Headphones	$1,115	$213	$1,999
Electronics & Gadgets	Laptop	$1,094	$202	$1,999
Electronics & Gadgets	Smart Watch	$1,089	$206	$1,999
Electronics & Gadgets	Smartphone	$994	$203	$1,993
Electronics & Gadgets	Tablet	$1,082	$202	$2,000
Electronics & Gadgets	TV	$1,046	$202	$1,999
Electronics & Gadgets	Virtual Reality Headset	$1,127	$200	$1,998
Electronics & Gadgets	Wireless Speaker	$1,107	$201	$1,994
Total		**$1,091**	**$200**	**$2,000**

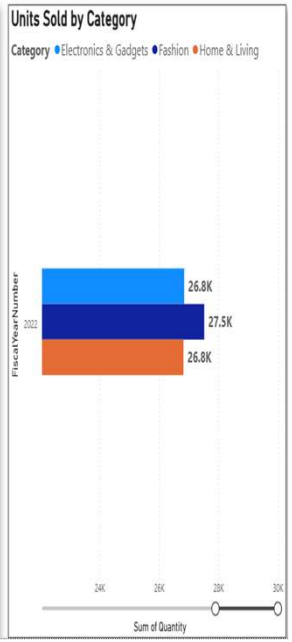
Category	Average Discount
Electronics & Gadgets	25%
Fashion	25%
Home & Living	25%
Total	**25%**

Units Sold in 2022

The total units sold for each category of products were about the same around 27K, but the main contributor of profit came from Home & Living products.

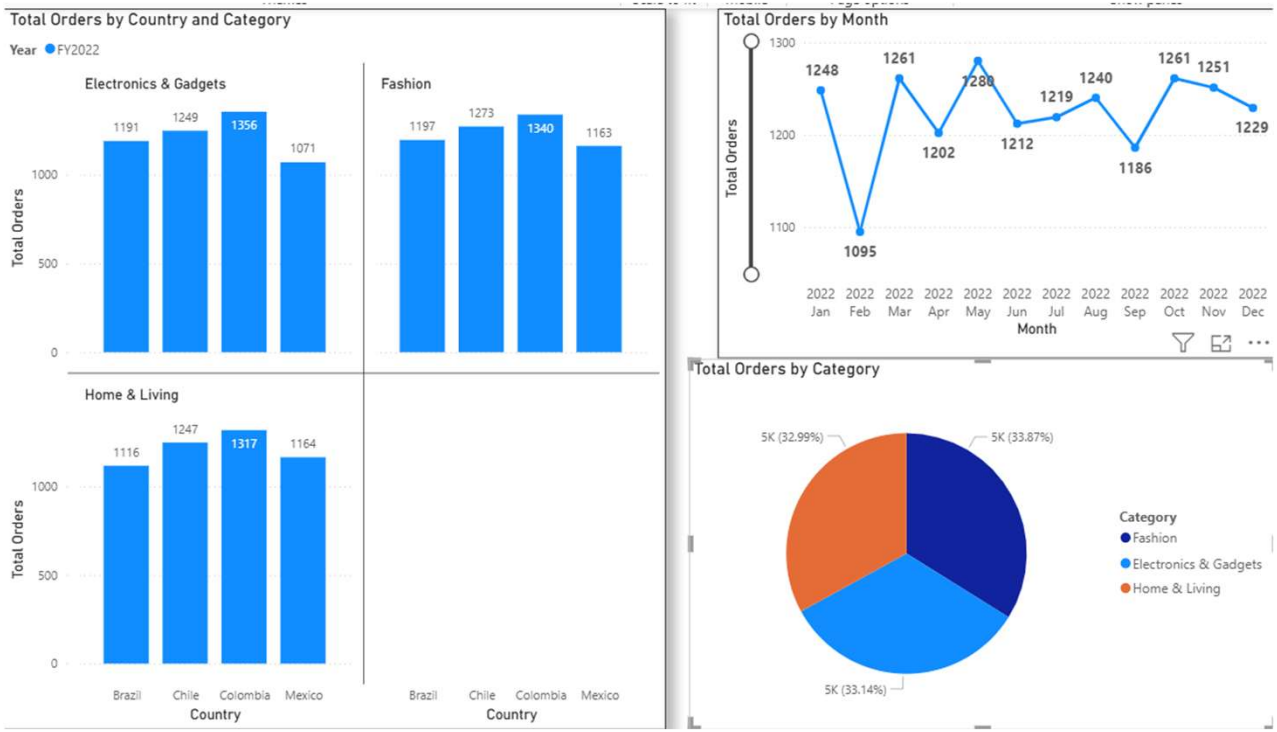
Category	Quantity	Median Quantity	Min Quantity	Max Quantity
Electronics & Gadgets	**26,847**	**6**	**1**	**10**
Camera	2,892	6	1	10
Gaming Console	2,783	6	1	10
Headphones	2,779	6	1	10
Laptop	2,446	5	1	10
Smart Watch	2,698	6	1	10
Smartphone	2,408	6	1	10
Tablet	2,746	6	1	10
TV	2,732	5	1	10
Virtual Reality Headset	2,740	6	1	10
Wireless Speaker	2,623	5	1	10
Fashion	**27,522**	**6**	**1**	**10**
Activewear Set	2,966	6	1	10
Dress	2,523	5	1	10
Formal Shirt	2,704	6	1	10
Jacket	2,770	6	1	10
Jeans	2,897	6	1	10
Shoes	2,735	5	1	10
Skirt	2,611	5	1	10
Sports Shoes	2,861	5	1	10
Sweater	2,690	5	1	10
T-Shirt	2,765	5	1	10
Home & Living	**26,822**	**6**	**1**	**10**
Appliance Set	2,792	6	1	10
Bedding	2,632	6	1	10
Curtains	2,781	5	1	10
Dining Set	2,575	6	1	10
Furniture	2,754	6	1	10
Home Decor	2,670	6	1	10
Kitchen Appliances	2,592	6	1	10
Lighting	2,593	5	1	10
Rugs	2,556	5	1	10
Wall Art	2,877	5	1	10
Total	**81,191**	**6**	**1**	**10**





Total Orders

- Based on the stacked column chart, Columbia has the highest total number of orders.
- The orders for each month ranges from 1,095 to 1,280.
- The total orders for each product category was about 5K, but Home & Living products contributed the largest share of profit.



2022
Shipping Delays

- On average the shipping delay days is 7.48.
- The total shipping delay days is rising. Hence, the company needs to work on reducing its shipping delay on all deliveries.

Product	Sum of Shipping Delay (Days)	Average of Shipping Delay (Days)
Sports Shoes	4,057	7.54
Furniture	4,026	7.99
Wall Art	4,000	7.68
Jeans	3,934	7.71
Camera	3,933	7.68
Headphones	3,895	7.73
Skirt	3,810	7.81
Sweater	3,808	7.62
Gaming Console	3,799	7.61
TV	3,748	7.32
Shoes	3,734	7.51
Curtains	3,725	7.36
Activewear Set	3,707	7.06
Smart Watch	3,682	7.72
Wireless Speaker	3,669	7.55
Rugs	3,643	7.56
Tablet	3,626	7.40
Lighting	3,618	7.60
Formal Shirt	3,615	7.56
T-Shirt	3,587	7.07
Home Decor	3,564	7.58
Appliance Set	3,548	7.24
Jacket	3,505	7.27
Virtual Reality Headset	3,497	7.08
Kitchen Appliances	3,496	7.63
Dining Set	3,483	7.39
Dress	3,394	7.58
Bedding	3,352	7.19
Smartphone	3,235	7.30
Laptop	3,203	7.12
Total	**109,893**	**7.48**

Country	Average	Maximum
Brazil	7.48	15
Chile	7.44	15
Colombia	7.57	15
Mexico	7.43	15
Total	**7.48**	**15**

Implications

Implications

Analysis of 2019 – 2022 trend

- The business started its online sales since August 2019 at the beginning of the COVID-19 pandemic. As a result, there was a sharp rise in profit, sales and units sold from 2019 to 2020 across all products.
- Since 2020, the growth had been stagnant.
- This trend was consistent across all 4 countries: Columbia, Chile, Brazil and Mexico.
- There was no correlation between sales and customer demographic factors such as country, age and income of customers.

Analysis of 2022 performance

- Generally, the Home & Living products were more expensive than the other 2 categories of products.
- In 2022, at country level, the share of profit across the 4 countries are about the same. At product level, while order levels and units sold were about the same, the bulk of profit came from sale of Home & Living products. Electronics & Gadgets products had moderate profit. Fashion had low profit.
- This suggests that higher price products help to contribute to higher profit. In addition, the company needs to improve order level and units sold in order to increase its profit.
- The total shipping delay days was rising. This would affect operation cost and customer satisfaction. This could eventually lead to loss of customers.

Conclusion

- The continued growth will require targeted strategies such as market expansion into other countries, broaden the range of products or cost efficiency improvements.

Python Machine Learning

- **Predictive Modeling and Profit Forecasting**
- **Tools & Libraries Used:**
  - pandas, numpy, matplotlib, seaborn for data processing and visualization.
- **Modeling Process:**
  - **Time Series Analysis**
    - Used historical profit data (2019–2022) to model future profit trends.
    - Tested models such as **Linear Regression** to forecast profit for 2023.
  - **Model Evaluation**
    - Assessed model accuracy using **Root Mean Square Error (RMSE)**.
    - Selected the model with the best predictive performance for final forecasting.

Profit Forecast

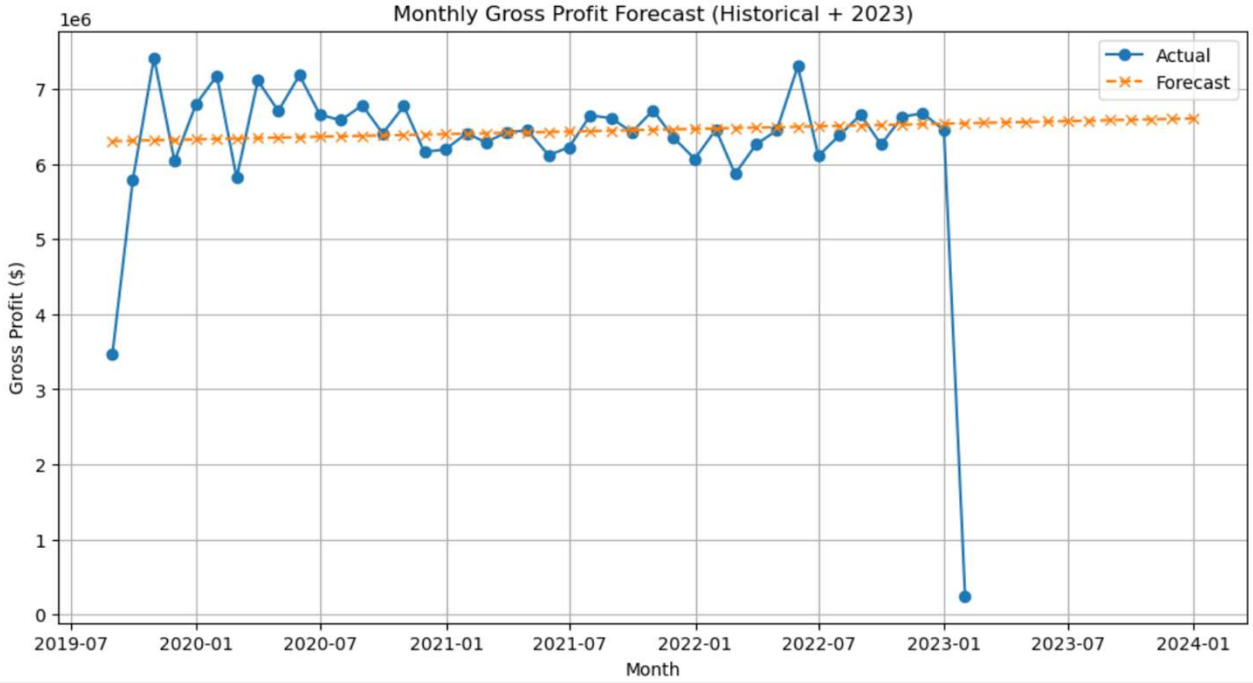
Method	Purpose	Strengths	Limitations
Linear Regression	Predict profit	Simple, interpretable	Only linear patterns
Prophet	Forecast time series	Handles seasonality	Struggles with non-linear features
Random Forest	Predict profit	Captures complex patterns	Requires more computational resources to explain

Why use RMSE?

- RMSE calculates the square root of the average squared differences between predicted and actual values.
- It penalizes larger errors more heavily, making it sensitive to big mistakes.
- Its units match the target variable, so it's easy to understand.

Profit Forecast using Linear Regression

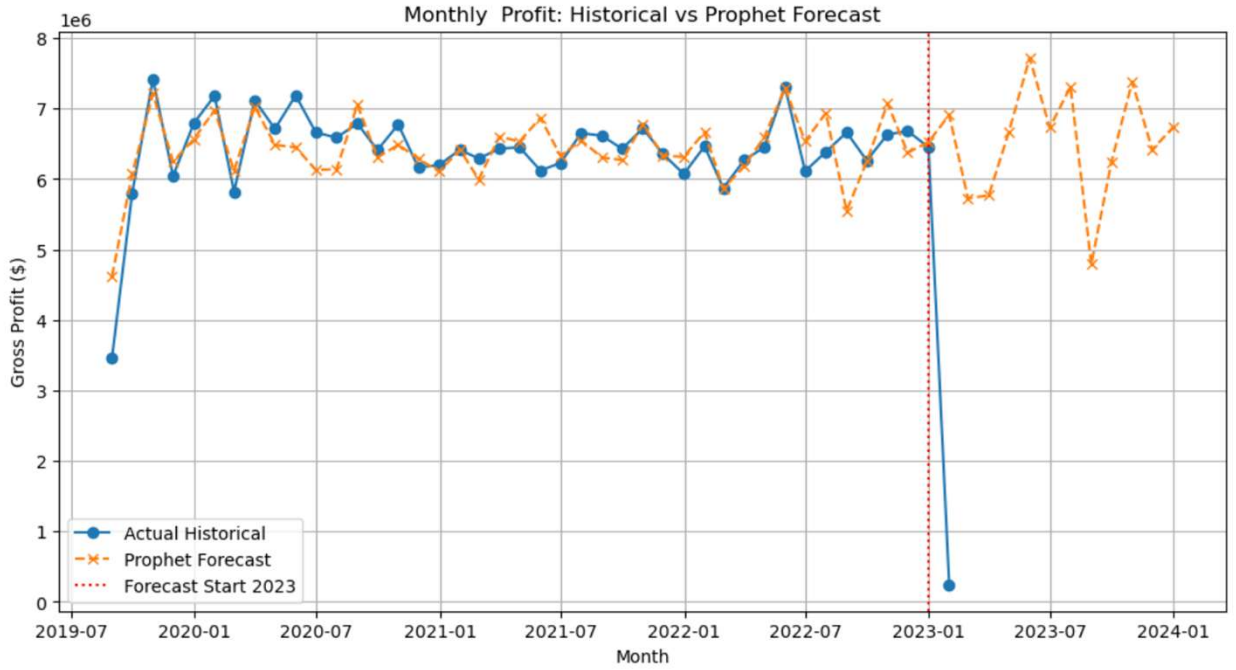
```
RMSE on historical data: $597,350.76

Total Gross Profit for 2023: $78,950,127
```



Profit Forecast using Prophet

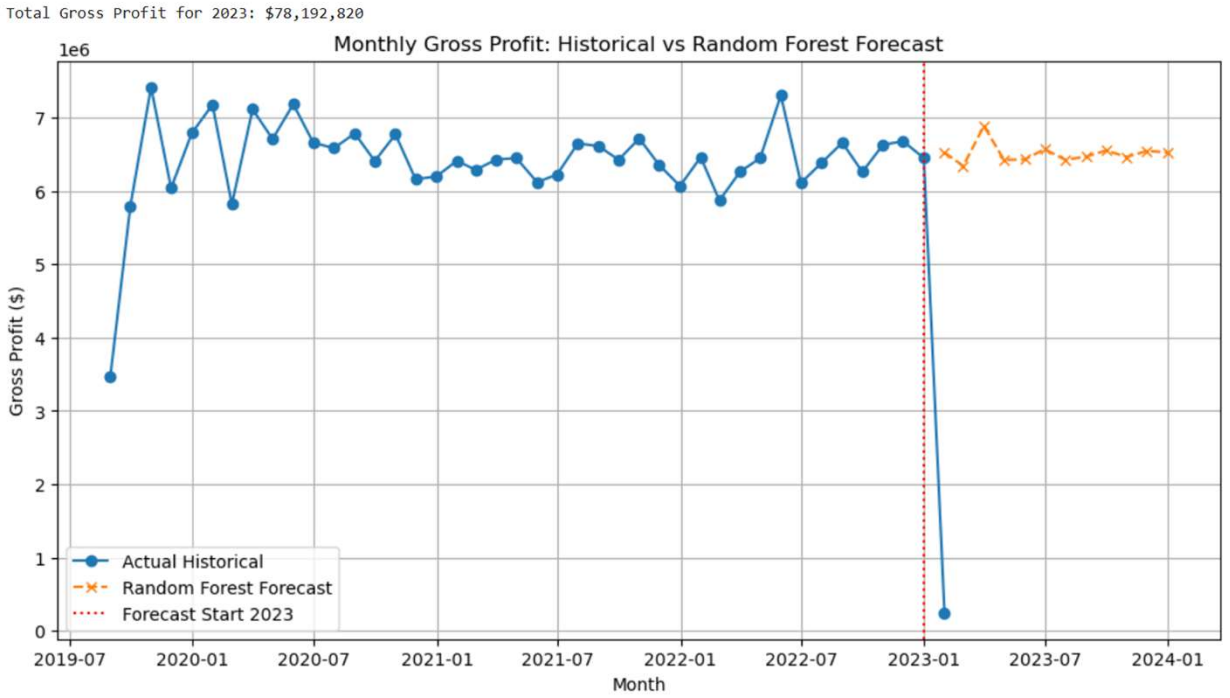
RMSE on historical data: $380,087.56

Total Gross Profit for 2023: $78,392,816



Profit Forecast using Random Forest

RMSE on historical data: $128,974.85



A lower RMSE is better – it indicates predictions are closer to reality. As the monthly profit is about $6.5 million, the error is about 2% of profit (good accuracy).

Profit Forecast using Random Forest

Monthly Gross Profit Forecast for 2023 using Random Forest:

	month_end	predicted_gross_profit_formatted
0	2023-01-31	$6,529,140
1	2023-02-28	$6,343,052
2	2023-03-31	$6,888,899
3	2023-04-30	$6,428,313
4	2023-05-31	$6,432,946
5	2023-06-30	$6,572,540
6	2023-07-31	$6,428,093
7	2023-08-31	$6,471,120
8	2023-09-30	$6,555,060
9	2023-10-31	$6,466,856
10	2023-11-30	$6,549,052
11	2023-12-31	$6,527,749

Total Gross Profit for 2023: $78,192,820

Tools and Technologies Summary

Tool	Purpose
Microsoft Power BI	Interactive visualization, dashboard creation, and trend analysis
Python (pandas etc)	Predictive modeling and statistical forecasting
CSV Files	Data storage and preprocessing
Power Query / DAX	Data cleaning, transformations, and calculated metrics in Power BI

Conclusion and Future Work

- **Key Acheivements**: Combined Power BI visuals with Python forecasts to enable the business to gain insights on its performance and ability to forecast future profit.
- **Potential future enhancements**: To use Python Machine Learning to forecast the customer purchasing behaviour. For example, to develop a model to enable user to input customer's age, income level and country location, to enable the machine to forecast this customer will likely to purchase which product and at what price.

Thank you