

VIETNAM NATIONAL UNIVERSITY OF HCM CITY
UNIVERSITY OF ECONOMICS AND LAW
FACULTY OF INFORMATION SYSTEM



BUSINESS DATA ANALYTICS

Final Report:

**SALES & MARKETING
ANALYTICS REPORT**

Instructor: Nguyen Van Ho

Group 10

No	Member	Student ID
1	Phan Văn Hiếu	K194111605
2	Nguyễn Thị Quỳnh Như	K194111621
3	Lương Trường Phước	K194111624
4	Nguyễn Thị Thanh Tâm	K194111627
5	Nguyễn Trần Huyền Trang	K194111634

Ho Chi Minh City, 20th January 2023

ACKNOWLEDGEMENT

The project is the result of the learning process as well as a summary of the valuable knowledge that our group has gained in the Business Data Analytics subject.

To complete this thesis, group 10 would like to express our deepest thanks to Mr. Nguyen Van Ho - Instructor. By your enthusiasm, we have gained useful knowledge in the field of data science as well as using NI tools to visualize data; knowing how to manage data, make statistics and reports in businesses. Additionally, during the learning process, our group also cultivated many soft skills such as teamwork skills, presentation skills, etc. However, during the implementation of this project, due to limited knowledge and not much time to implement, the group's project was unable to avoid errors. The group is looking forward to receiving your understanding.

We wish you a good health and success in work and life. Thank you very much.

Group 10

TABLE OF CONTENT

Chapter 1: OVERVIEW OF PROJECT	1
1.1. Reason for the topic	1
1.2. Project objective	1
1.3. Subject and Research scope of the project	1
1.4. Used Tools	1
1.5. Research Implication	2
1.6. Structure of the project	2
Chapter 2: THEORETICAL BASIC	3
2.1 Overview of business analytics	3
2.2 Overview of business intelligence	3
2.2.1 Benefits of BI	3
2.2.2 Process of building BI for business	4
2.3 Data analysis	4
2.3.1 What is data analysis	5
2.3.2 Data analysis theory	5
2.3.3 Data analysis method	5
2.4. Visualization	6
2.5. Data analysis tools	6
2.5.1. Python	6
2.5.2. Tableau	6
Chapter 3: ANALYSIS OF USER REQUIREMENTS AND DATA DESCRIPTION	8
3.1. Business issue understanding	8
3.1.1. Business objective:	8
3.1.2. Gather required information:	8
3.1.3. Determine appropriate analysis method:	8
3.1.4. Scope of work:	9

3.2. Data understanding	9
3.2.1. Collect initial data:	9
3.2.2. Identify data requirements:.....	9
3.2.3. Determine data availability:	10
3.2.4. Explore data and characteristics:	11
3.3. Overview of the database	12
3.3.1. Data table	12
3.3.2. Data Relationship	15
Chapter 4: DATA ANALYSIS AND RESULTS	16
4.1. Data preprocessing	16
4.2. The structure of the Reporting dashboard	18
4.3. Analyze and visualize data	19
4.2.1 Business overview analysis	19
4.3.2 Sales & Profit analysis.....	20
4.3.3 Customer analysis.....	22
4.3.4. Product Analysis.....	24
4.3.5. Shipment Analysis.....	25
4.4. Discuss and evaluate the results of data analytics	27
4.4.1. Discussion.....	27
4.4.2. Evaluation.....	39
Chapter 5: CONCLUSION.....	44
5.1. Results	44
5.1.1. Advantages	44
5.1.2. Limitations.....	44
5.2. Future works	44
APPENDIX	45
REFERENCES	46

TABLE OF FIGURE

Figure 1. Data relationship model	15
Figure 2. Checking data by Info Statement	17
Figure 3. Checking duplicated data	17
Figure 4. Showing duplicated result	17
Figure 5. Deleting duplicate result	17
Figure 6. Describing dataset	18
Figure 7. Save data	18
Figure 8. Data visualization of General analysis.....	19
Figure 9. Sales and Profit dashboard	21
Figure 10. Data visualization of Customer analysis.....	23
Figure 11. Product Dashboard.....	25
Figure 12. Data visualization of Shipment analysis	26
Figure 13. Sales & Profit Dashboard.....	30
Figure 14. Sales and Profit in 2012	31
Figure 15. Sales and Profit in 2014	32
Figure 16. Statistics of Machines in Technology	36
Figure 17. Sales, Profit and Profit ratio by Products of Furniture.....	36
Figure 18. Sales, Profit ratio and Discount by Category of Furniture.....	37
Figure 19. Sales, profit and quality by mmanufacturer of sub-category	37
Figure 20. Statistics of Office suppliers category	38

TABLE OF TABLE

Table 1. Data requirement	10
Table 2. Table of data	14
Table 3. General analysis.....	28
Table 4. Overview of Sales and profit from 2011 - 2014.....	29
Table 5. Metric of customer analysis	33
Table 6. Metrics of product analysis	35
Table 7. Shipment analysis	39
Table 8. Evaluation table	43

Chapter 1: OVERVIEW OF PROJECT

1.1. Reason for the topic

In today's modern era, data plays an important role in any business due to its various benefits. Data can be used to measure or record a wide range of business activities - both internal and external. While the data itself may not be very informative, it is the basis for all reporting and as such is crucial in business. They also allow enterprises to establish baselines, benchmarks, and performance goals for business development.

Data analytics is essential because it helps businesses optimize their performances. Implementing it into the business model means companies can help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data. A company can also use data analytics to analyze business situations, make better business decisions and help analyze customer behaviors and satisfaction.

By making a "*Sales and Marketing Analytics Report*", we would like to explore, visualize and analyze data in many aspects of business situations to recommend many valuable and practical Marketing solutions for contributing to building development strategies for business.

1.2. Project objective

The objective of making Sales and Marketing analytics is to better understand business situations and increase the performance goals of the business.

- Data analysis of some departments such as Sales & profit, products, customer, shipment, etc
- Build a dashboard to visualize business situations.
- Propose recommendations for the enterprise development strategies
- Build some directions to develop the topic in the future.

1.3. Subject and Research scope of the project

- Subject: Data of Super Stores database (Furniture, Office Supplies, Technology)
- Research scope: This project explores data in some domains Sales & Marketing.

1.4. Used Tools

- Tableau

1.5. Research Implication

After finishing the Data analytics, Super Stores can:

- Review the statistical data of business situation from 2011 - 2014
- Identify customers segment, and potential locations,... based on sales and profit
- Build development strategies and change them promptly.
- Suggest some ways to optimize the Sales and Marketing performance

1.6. Structure of the project

Structure of report:

- Chapter 1: Overview of the project
Presenting an overview of the topic, including the reason for forming the topic, topic goal, Subject and research scope of the project, research implications and structure of the topic.
- Chapter 2: Theoretical basic
Presenting theoretical bases of data analytics, business intelligence, models, solutions, and benefits of BI.
- Chapter 3: Analysis of user requirements and data description
Presentation of the implementation process including requirements, original data description, and data analysis.
- Chapter 4: Data analysis and Results
Visualize requirements for generalization.
- Chapter 5: Conclusion
Presentation of the results and give some recommendations for the company

Chapter 2: THEORETICAL BASIC

2.1 Overview of business analytics

Business analytics uses data to create mathematical models to help organizations make decisions that bring value or are in their best interest. There is much data out there for organizations to use, though what data they choose and why they use it will vary from industry to industry. Data is such an essential tool for organizations because the data itself is objective, and grounded in fact. It's this objectivity that can be most helpful for organizations in determining what decisions they need to make to create the most value for them.

There are a few types of business analytics, all of which use data to accomplish a goal. Predictive analytics is focused on identifying probable outcomes. Prescriptive analytics recommends actions to an organization to help reach organizational goals. And descriptive analytics identifies an organization's historical trends. Most business analytic professionals use either predictive or descriptive analytics.

2.2 Overview of business intelligence

Business intelligence combines business analytics, data mining, data visualization, data tools and infrastructure, and best practices to help organizations make more data-driven decisions. In practice, you know you've got modern business intelligence when you have a comprehensive view of your organization's data and use that data to drive change, eliminate inefficiencies, and quickly adapt to market or supply changes. Modern BI solutions prioritize flexible self-service analysis, governed data on trusted platforms, empowered business users, and speed to insight.

BI should not be confused with 'business analytics'. Business intelligence is descriptive and uses metrics to generate clear snapshots of business performance. Meanwhile, business analytics is predictive and describes what organizations should do in the future to generate better outcomes.

2.2.1 Benefits of BI

Almost every other company these days requires business intelligence tools to stay in the competition. Companies of all sizes and stages need BI tools to make smarter and more informed business decisions, manage, and analyze business data. BI tools help in

visualizing the reports generated and maximize revenue outcomes. The primary benefit of business intelligence tools is to provide a solution that is suitable for your business. It customizes and offers a solution that is most suitable to the company and is relevant to your business goals. With business intelligence and business analytics, your company can have an account of all the impacts it has had (whether they're good or bad.) Data analytics plays a vital role in helping your company devise the best plan possible. Business intelligence systems help companies move towards the right direction and make smart choices from the business data collected and analyzed. The benefits of BI are plentiful. The most advantageous of them are:

- Faster decision making
- Getting key insights
- Improving customer satisfaction
- Greater operational efficiency
- Bigger profits

2.2.2 Process of building BI for business

Although business intelligence is utilized in different ways and for different purposes by individual companies, the process is fairly uniform throughout all industries and typically unfolds as follows:

- Data from various sources — including internal company data and external market data — is collected, integrated, and then stored; because “big data” is commonly used, data is commonly stored in what’s called a data warehouse, created by a data engineer
- Data sets are created and prepared for data analysis, often by creating data analysis models
- Data analysts run queries against the data sets or models
- The results of queries are used to produce visualizations in the form of charts, graphs, histograms, or other visual representations, along with BI dashboards and reports
- Decision-makers utilize the data visualizations and reports to help them in making decisions; they may also use their BI dashboard to probe further into the data for more information.

2.3 Data analysis

2.3.1 What is data analysis

Data analysis is the process of cleaning, changing, and processing raw data and extracting actionable, relevant information that helps businesses make informed decisions. The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.

2.3.2 Data analysis theory

Data analysis theory:

- John Tukey's Theory of Exploratory Data Analysis (EDA): analyze data sets to summarize their key characteristics, often using statistical graphics and other data visualization methods.
- Confirmatory Data Analysis (CDA): Concentrate on using traditional statistical tools to assess the data and challenge any assumptions you made during EDA.
- Grounded Theory of Analysis: collection and analysis of information at the same time.
 - Stage 1: Collect a satisfying amount of information
 - Stage 2: Analyze all collected data. Data is indexed and relevancy discovered.This process automatically repeats as more data is collected and analyzed.

2.3.3 Data analysis method

- Multidimensional cubes: Managers and employees can view data in different aspects through rotate, slice, drill down operations on the data block
- Time series: Data will be recorded over a set period of time rather than intermittently or randomly recording data points to uncover data differences over time. It is one of the most effective methods to help companies predict or forecast the future based on historical data.
- Data mining: Analyze large amounts of information to distinguish trends and patterns in order to uncover customer behaviors, habits, and trends based on the analyzed results.
- Optimization models: include three elements: the objective function, decision variables and business constraints to identify the best solutions out of a given set of solutions

2.4. Visualization

Data visualization is a way of representing information from metrics to visual elements such as charts, graphs and maps, thereby making it easier for departmental staff managers to capture information. Data visualization makes it easy for users to access and understand trends, outliers, and patterns in data.

There are some types of chart which have their own objective:

- Specific value: Single value, table, highlight table.
- Comparison: Single line, multiples line, bar chart, group bar, group bar chart, bullet chart.
- Relationship: Scatter plot, bubble chart, word cloud.
- Composition: Tree-map, pie chart, donut chart.
- Distribution: Box plot, scatter plot, histogram.
- Geographic Filled map, symbol map.

2.5. Data analysis tools

2.5.1. Python

Python is a high-level programming language, used in application, web, and software development. In addition, Python performs well in data analysis thanks to the following outstanding features.

- Easy to learn: Python is a good choice for those who are just starting out with data analysis because it is quite easy to learn and understand, and the syntax is clear.
- Flexibility: Thanks to its flexibility and high applicability (can be used with any other application development tool), Python is used in many fields and projects.
- Supports a wide range of libraries: Python supports a wide range of libraries available for data analysis and data science, helping analysts save a significant amount of money. In addition, Python libraries are always evolving and updating with new features that make data analysis easier

2.5.2. Tableau

Tableau is an excellent business intelligence and data visualization application for reporting and analyzing huge amounts of data. It was founded in America in 2003, and in June 2019, Salesforce acquired Tableau. It assists users in producing a variety of graphs, maps, dashboards, and stories for the purpose of visualizing and analyzing data to aid in business decision-making.

Features of Tableau:

- Tableau supports powerful data discovery and exploration that enables users to answer important questions in seconds
- Users without relevant experience need not have any prior programming knowledge in order to begin building visualizations with Tableau.
- It can link to a number of data sources that are not supported by other BI products. Users of Tableau can combine and produce reports from many datasets.
- A centralized location to manage all published data sources inside an organization is supported by Tableau Server.

Chapter 3: ANALYSIS OF USER REQUIREMENTS AND DATA DESCRIPTION

3.1. Business issue understanding

3.1.1. Business objective:

With growing demands and cut-throat competition in the market, the business stakeholders want to know our expertise in order to determine what will work best for their business. In order to keep the business growing and increase the profit better, they should constantly have an overall view of how their sales are and easily conclude insights from it. Then, the stakeholders would be able to have an objective decision-making process.

To have a better understanding of the business situation and decision-making process, the information the stakeholders are seeking would be:

- Current business situation
- Customer analysis
- Product category analysis
- Sales and profit time series
- Shipment analysis
- Correlation analysis between Discount and Sales
- Insight: which products, regions, categories and customer segments they should target or avoid.

3.1.2. Gather required information:

- The target customers of the report are the board of directors, executives, business strategy developers and investors.
- They are people who hold senior positions in the company, without in-depth knowledge of data, statistics, and numbers. Therefore, they need a specific, clear report that can easily identify the results and grasp the insights of the report.
- Besides, the time of the executive board is also limited, it is necessary to pay attention to the presentation time, consider the main content.

3.1.3. Determine appropriate analysis method:

From the business objective, we decided to use Descriptive Analytics to analyze data reports on the business situation of the business. Thus, finding insights and issues

in the business to make decisions and deliver an overview of the operational situation for managers and investors.

In addition, we may use Diagnostic Analytics to explore data, notice correlations between variables, and give reasons for business problems.

3.1.4. Scope of work:

- To build dashboards that are intuitive, legible with basic variables in the dataset.
- Expected output: Build dashboard that visualizes the relationship between variables based on Areas, Product categories, Customer segment and Correlation between Discount and Sales by product category
- Special requirements: minimize time spent on the executing board. Focus on the main content that describes the business situation. In this way, stakeholders who know little about data analytics can make future decisions.

3.2. Data understanding

3.2.1. Collect initial data:

The Superstore dataset is data from a global retail chain that sells office supplies, Furniture and Technology Stuff. This data is downloaded from Tableau Public Resource page. The dataset includes data on the business situation of Superstore businesses from 2011 to 2014.

Link: <https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls>

3.2.2. Identify data requirements:

Our deliverables are expected to include the following sections:

Deliverables	Data requirements
General dashboard (Current business situation)	<ol style="list-style-type: none">1. Period (Timeline)2. Total orders3. Total Products sold4. Total Sales of business5. Total profit6. Profit ratio7. Sales chart over year

	8. Profit chart over year
Area analysis	<ol style="list-style-type: none"> 1. Country 2. Region 3. Sales 4. Profit
Customer analysis	<ol style="list-style-type: none"> 1. Customer name 2. Customer Segment 3. Sales 4. Profit
Product category analysis	<ol style="list-style-type: none"> 1. ProductCategory 2. Detail Category 3. Sales 4. Profit
Sales and profit time series	<ol style="list-style-type: none"> 1. Order date 2. Sales 3. Profit 4. Profit Ratio
Shipment Analysis	<ol style="list-style-type: none"> 1. Shipment type 2. Customer segment 3. Region
Correlation analysis between Discount and Sales by product category	<ol style="list-style-type: none"> 1. Product category 2. Discount 3. Sales 4. Quantity

Table 1. Data requirement

3.2.3. Determine data availability:

- For the current business situation, we will use the Order date variable to calculate metrics such as product, order, sales, profit,... Total order is based on the unique value in the dataset. Variable of Quantity will be used to show total product. Sales

and Profit in the dataset will be used to calculate the total sales and total profit of business. Also, these variables and order date can be utilized to generate charts of sales and profits over years. Profit ratio will be shown on the dashboard as well.

- Regarding Area Analysis, in order to show Sales and profit by area, we choose variables of City, Region, Sales and Profit in the dataset.
- To meet the requirement for customer-based business performance analysis. We select the data in the Customer name, Segment (customer), Sales, Profit columns to support descriptive analysis.
- The data in the dataset related to Category, Sub-Category, Quantity, Sales and Profit will be utilized to analyze the business situation by product category.
- The data in the Order date, Sales, Profit, Profit Ratio columns of the dataset will be selected to show the change in Sales as well as Profit over each time period.
- Regarding the analysis of the shipping array, we will use the Ship Mode, Segment (customer) and Region data available in the dataset.
- In addition, the customer has a requirement to show the correlation between Sales and Discount data, so we will choose the variables of Sales, Profit, Category (product), quantity and maybe there will need to be data about Segments (customer).

3.2.4. Explore data and characteristics:

- Order ID: Unique Order ID for each Customer.
- Order Date: Order Date of the product.
- Ship Date: Shipping Date of the Product.
- Ship Mode: Shipping Mode specified by the Customer.
- Customer Name: Name of the Customer.
- Segment: The segment where the Customer belongs.
- Country: Country of residence of the Customer.
- City: City of residence of the Customer.
- Manufacturer: The place of manufacturer of the product
- State: State of residence of the Customer.
- Postal Code: Postal Code of every Customer.
- Region: Region where the Customer belongs.
- Category: Category of the product ordered.
- Sub-Category: Detailed category of the product ordered.

- Product Name: Name of the Product
- Sales: Sales of the Product.
- Quantity: Quantity of the Product ordered.
- Discount: Discount provided.
- Profit: Profit/Loss incurred.
- Number of Records: Number of records per order
- Profit Ratio: the earning profits over a period of time

3.3. Overview of the database

3.3.1. Data table

No	Type	Table name	Description	Data quantity
1	Fact	fact_sales	sales_order_number (PK), product_id_key (FK), order_id_key(FK), ship_id_key (FK), location_id_key (FK), discount_id_key (FK), date_id_key (FK), sales_amount, total_sales, total_profit, profit_ratio, total_product_cost	<ul style="list-style-type: none"> • 9995 records • Key: 100% valid, 0% error, 0% empty
		fact_product	product_id_key (FK), product_name, total_product_cost, total_product_sales, total_product_profit	<ul style="list-style-type: none"> • 1841 records • Key: 100% valid, 0% error, 0% empty
		fact_order	order_id_key (FK), customer_id_key (FK), product_name, order_date (FK), Quantity, Discount (FK)	<ul style="list-style-type: none"> • 5009 records • Key: 100% valid, 0% error, 0% empty
		fact_customer	customer_id_key (FK), location_id_key (FK), total_sales, age, gender, modified_date	<ul style="list-style-type: none"> • 793 records • Key: 100% valid, 0% error, 0% empty

2	Dim	dim_customer	customer_id_key (PK), dim_customer_id (NK), first_name, last_name, gender, job, phone, email, street, location_id (FK), segment_id (FK) created_by, modified_by, created_date, modified_date	<ul style="list-style-type: none"> • 9995 records • Key: 100% valid, 0% error, 0% empty
		dim_segment	segment_id_key (PK), segment, product_id (FK)	<ul style="list-style-type: none"> • 3 records • Key: 100% valid, 0% error, 0% empty
		dim_location	location_id_key (PK), postal_code, city, state, country, region	<ul style="list-style-type: none"> • 631 records • Key: 100% valid, 0% error, 0% empty
		dim_subcategory	sub-category_id_key (PK), sub- category_name, created_date, modified_date	<ul style="list-style-type: none"> • 17 records • Key: 100% valid, 0% error, 0% empty
		dim_product	product_id_key (PK), product_name, category_id_key (FK), sub- category_id_key (FK), created_date, modified_date, unit_price, SKU_number, manufacturer_id_key (FK)	<ul style="list-style-type: none"> • 1841 records • Key: 100% valid, 0% error, 0% empty
		dim_category	category_id_key (PK), category_name, created_date, modified_date	<ul style="list-style-type: none"> • 3 records • Key: 100% valid, 0% error, 0% empty
		dim_date	date_id_key (PK), actualdate, year, month, day, day_number_of_week, name_of_day, quarter_of_year	<ul style="list-style-type: none"> • Key: 100% valid, 0% error, 0% empty

		dim_order	order information: order_id_key (PK), customer_id_key (FK), product_id_key (FK), order_quantity, total_order_price, order_date, ship_date (FK), ship_address (FK), ship_postal_code, discount_id_key (FK), location_id (FK), city_id (FK), ship_id, segment_id, modified_date (FK)	<ul style="list-style-type: none"> • 5009 records • Key: 100% valid, 0% error, 0% empty
		dim_manufacturer	manufacturer_id_key (PK), manufacturer_name, location_id_key (FK), category_id_key (FK), product_id_key (FK), modified_date (FK)	<ul style="list-style-type: none"> • 174 records • Key: 100% valid, 0% error, 0% empty
		dim_discount	store discount information: discount_id_key (PK), discount_name, product_id_key (FK), discount_date_id_key (FK), discount_amount, description, created_by, modified_by, created_date, modified_date	<ul style="list-style-type: none"> • Key: 100% valid, 0% error, 0% empty
		dim_shipment	store shipment information: ship_id_key (PK), order_id_key (FK), location_id_key (FK), ship_date_id_key (FK), ship_date, ship_mode, ship_amount, created_date, modified_date	<ul style="list-style-type: none"> • 9995 records • Key: 100% valid, 0% error, 0% empty

Table 2. Table of data

3.3.2. Data Relationship

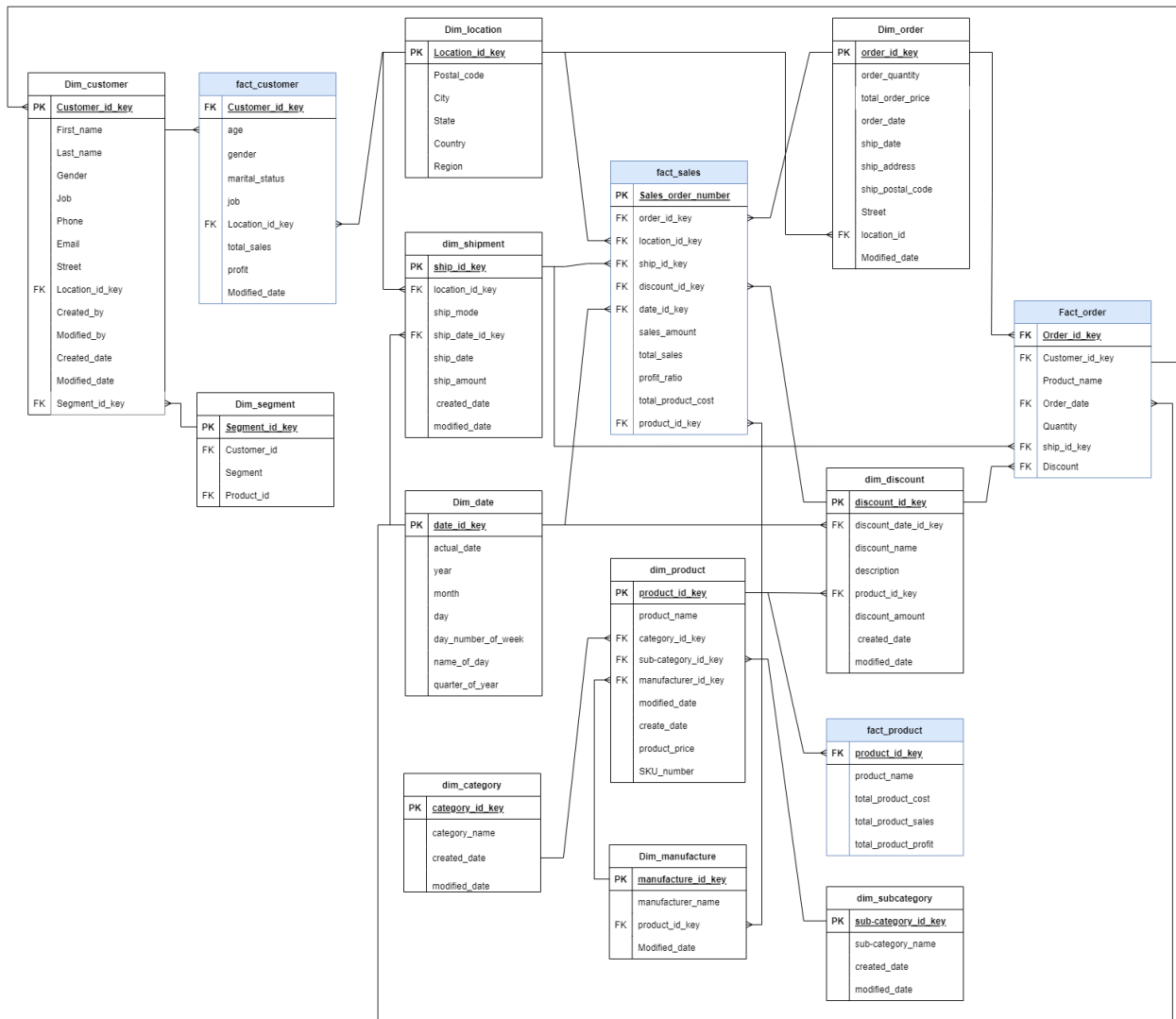


Figure 1. Data relationship model

Chapter 4: DATA ANALYSIS AND RESULTS

4.1. Data preprocessing

4.1.1 Loading Packages

```
- import pandas as pd
```

We used the Pandas library to clean our dataset.

4.1.2 Reading the data

```
- df = pd.read_excel('Sales-Superstore-Dataset.xlsx')
```

4.1.3 Checking Info data

```
- df.info()
```

```
#   Column      Non-Null Count  Dtype
---  -
0   Category    9994 non-null     object
1   City         9994 non-null     object
2   Country      9994 non-null     object
3   Customer Name 9994 non-null     object
4   Manufacturer 9994 non-null     object
5   Order Date   9994 non-null     datetime64[ns]
6   Order ID     9994 non-null     object
7   Postal Code  9994 non-null     int64
8   Product Name 9994 non-null     object
9   Region       9994 non-null     object
10  Segment      9994 non-null     object
11  Ship Date     9994 non-null     datetime64[ns]
12  Ship Mode     9994 non-null     object
13  State        9994 non-null     object
14  Sub-Category 9994 non-null     object
15  Discount      9994 non-null     float64
16  Number of Records 9994 non-null     int64
17  Profit        9994 non-null     int64
18  Profit Ratio  9994 non-null     float64
19  Quantity      9994 non-null     int64
20  Sales         9994 non-null     int64
dtypes: datetime64[ns](2), float64(2), int64(5), object(12)
memory usage: 1.6+ MB
```

Figure 2. Checking data by Info Statement

- The data set is relatively clean, the columns are properly formatted and complete

4.1.3 Checking duplicated data

```
df.duplicated().sum()

1
```

Figure 3. Checking duplicated data

- We have noticed that there is 1 duplicate row in the dataset.

```
df[df.duplicated(keep=False)]
```

	Category	City	Country	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	...	Ship Date	Ship Mode	State	Sub-Category	Discount
385	Furniture	Columbus	United States	Laurel Beltran	Global	2011-04-23	US-2011-150119	43229	Global Leather Highback Executive Chair with P...	East	...	2011-04-27	Standard Class	Ohio	Chairs	
386	Furniture	Columbus	United States	Laurel Beltran	Global	2011-04-23	US-2011-150119	43229	Global Leather Highback Executive Chair with P...	East	...	2011-04-27	Standard Class	Ohio	Chairs	

2 rows × 21 columns

Figure 4. Showing duplicated result

- After reviewing a duplicate row, we found this to be a crawl error because the Order Id cannot be duplicated. So we decided to delete it.

```
df.drop_duplicates(inplace=True)
```

Figure 5. Deleting duplicate result

```
df.describe()
```

	Postal Code	Discount	Number of Records	Profit	Profit Ratio	Quantity	Sales
count	9993.000000	9993.000000	9993.0	9993.000000	9993.000000	9993.000000	9993.000000
mean	55191.576403	0.156188	1.0	28.655959	0.120548	3.789753	229.868208
std	32065.074478	0.206457	0.0	234.267081	0.466944	2.225149	623.279922
min	1040.000000	0.000000	1.0	-6600.000000	-2.750000	1.000000	0.000000
25%	23223.000000	0.000000	1.0	2.000000	0.070000	2.000000	17.000000
50%	56560.000000	0.200000	1.0	9.000000	0.270000	3.000000	54.000000
75%	90008.000000	0.200000	1.0	29.000000	0.360000	5.000000	210.000000
max	99301.000000	0.800000	1.0	8400.000000	0.500000	14.000000	22638.000000

Figure 6. Describing dataset

- The results show that the columns in numerical form are relatively reasonably distributed, with no suspicious outliers.

4.1.4 Save data

```
df.to_excel("Sales_Superstore_fixed.xlsx")
```

Figure 7. Save data

- Finally, we save the edited dataset as a new excel file.

4.2. The structure of the Reporting dashboard

Dashboard included 5 analysis parts:

- General analysis: Basic statistics on revenue and business system of the whole enterprise.
- Sale & Profits analysis: Detailed statistics on the sales and profit of the whole enterprise classified by criteria.
- Customer analysis: Detailed statistics on the enterprise's customer clusters classified by criteria.
- Product analysis: Detailed statistics on the enterprise's product category that are often bought by each customer segment.
- Shipment analysis: Statistics of ship mode that are often bought together by clusters of customers.

4.3. Analyze and visualize data

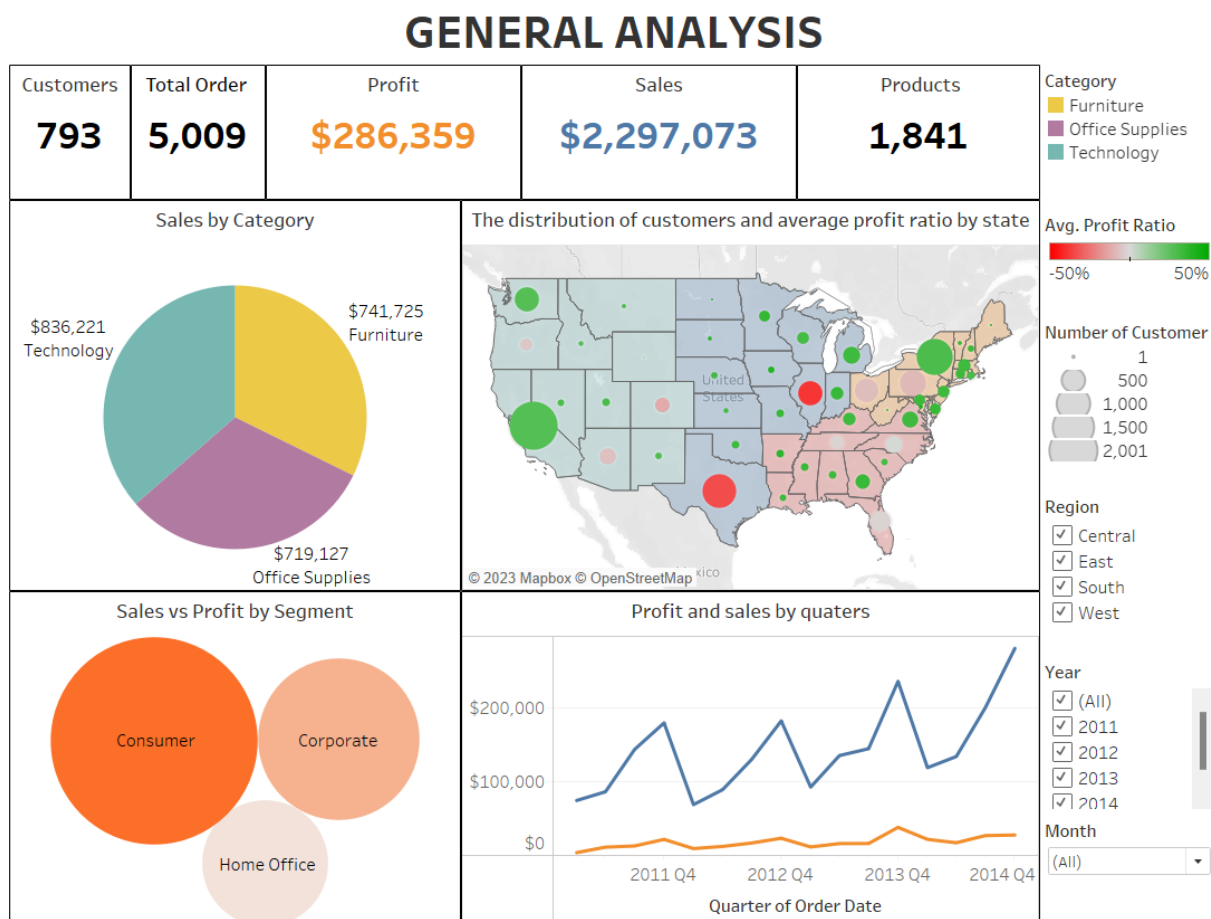
4.2.1 Business overview analysis

4.3.1.1. Data analysis

First, we created an overview dashboard of Sales Superstore by:

- Giving some important metrics about Sales Superstore's business: how many customers, how many orders, how many sold, how many products, profit.
- Statistics of sales & profit and sales & profit by category, by segment, by quarters of the year.
- The statistics can be filtered by years, by months, by region, etc.

4.3.1.2. Data visualization



We use Card visualization to show an overview of some key metrics of Sales Superstore including Customers , Total Order, Profit, Sales and Product and we also

represent the data in charts, graphs. These metrics will change by using filter Year and Region.

Sales by Category statistics: we use the Pie chart because it shows which category has the highest or lowest sales and profit.

Sales and Profit by Segment statistics: we use the Bubbles Chart to compare and show the relationships between Segment circles, by the use of positioning and size.

In addition, we use Symbol Maps to describe the statistics of customers and average profit ratio by state.

Last, we use Line Graph to show the relationship of profit and sales by quarter.

4.3.2 Sales & Profit analysis

4.3.2.1. Data analysis

For Sales and Profit Analysis, there are 3 metrics that need to be determined:

Profit Ratio:

- Description: A key indicator of the financial health of the company, comparing the earnings reported by a business to its sales..
- Formula: Profit ratio = $\text{SUM}([\text{Profit}]) / \text{SUM}([\text{Sales}])$

Sales MoM (%):

- Description: Percentage of sales growth of the current month compared to the previous month in the same year,
- Formula: Profit MoM = $(\text{Current Month Sales Value} - \text{Prior Sales Profit Value}) / \text{Prior Month Profit Value}$

Profit MoM (%):

- Description: Percentage of profit growth of the current month compared to the previous month in the same year.
- Formula: Profit MoM = $(\text{Current Month Profit Value} - \text{Prior Month Profit Value}) / \text{Prior Month Profit Value}$

Sales YoY (%):

- Description: Percentage of sales growth of the total fourth-quarter sales of the current year compared to the total fourth-quarter sales of the previous year.
- Formula: Profit MoM = $(\text{Current Year Total Sales} - \text{Prior Year Total Sales}) / \text{Prior Year Total Sales}$

Profit YoY (%):

- Description: Percentage of profit growth of the total fourth-quarter profit of the current year compared to the total fourth-quarter profit of the previous year.
- Formula: Profit MoM = (Current Year Total Profit – Prior Year Total Profit) / Prior Year Total Profit

4.3.2.2. Data visualization

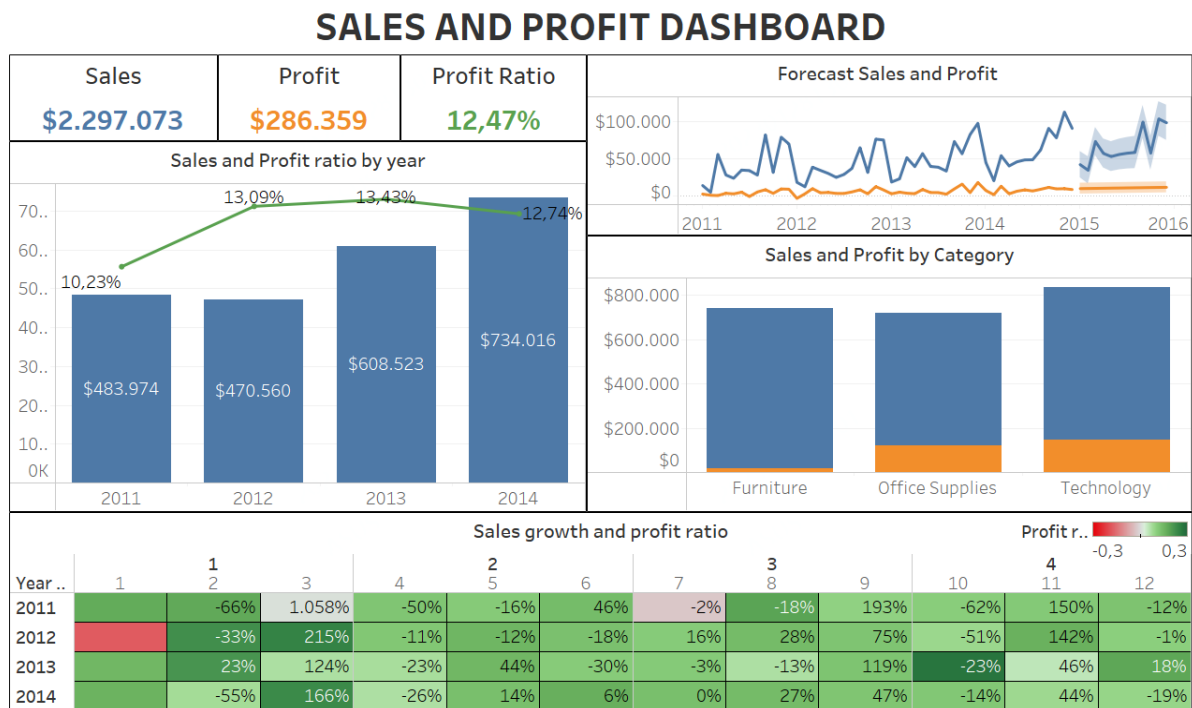


Figure 9. Sales and Profit dashboard

The sales & Profit dashboard contains data of Sales, Profit, Profit Ratio, Month over Month Sales Growth (MoM%), Month over Month Profit Ratio Growth (MoM%), Year over Year Profit Ratio Growth (YoY%), and Year over Year Sales Growth (YoY%). All the data will change when a user chooses a specific year or Category.

There are 3 metrics which are Sales, Profit, and Profit Ratio put on the left corner of the dashboard. These metrics show the overall business situation of the company. For better identification, each metric has different colors consistently throughout the data of this dashboard. Sales are blue; Profit is Orange and Profit Ratio is shown in green.

The mixed chart on the left side illustrates the changes in total sales and profit ratio over years from 2011 to 2014. The columns in the chart show the company's exact sales for each year and the percentage growth over the years. When users mouse over each column, there will be additional data on the percentage change in sales and profits.

The green line shows the annual rate of profit ratio. This chart helps the company have a more comprehensive view of sales and profits over the years to build an appropriate development strategy.

To the right of the dashboard are two charts: a line chart and a column chart. The line chart visualizes the fluctuations in sales and profit in detail over the years, when hovering the mouse over the line, it will display the detailed index for each month. In addition, the chart also calculates and estimates indicators of future sales and profits. On the other hand, the bar chart below shows the sales and profits of the product category. It is easy for users to select a specific product category, the metrics in the dashboard then will change into data of its category.

To add vibrancy to the Sales & Profit dashboard, at the bottom of the dashboard will be a color palette showing the growth/decrease of sales and profit margin month over month for all four years. The percentage shown in the table is the change in MoM sales, and the color will show the profit margin. The color bar of the profit ratio will be in the range of -30% to 30%, the redder the color is, the more negative the indicator receives; the darker the green color will show the higher profit ratio.

4.3.3 Customer analysis

4.3.2.1. Data analysis

For Customer Analysis, there are only 2 key metrics that need to be determined: AST, Profit Ratio (%).

AST:

- Description: Average sales per transaction.
- Formula: $\text{SUM}([\text{Sales}]) / \text{COUNT}([\text{Order ID}])$

Profit Ratio (%):

- Description: Percentage of the company's sales consists of profit.
- Formula: $\text{Profit ratio} = \text{SUM}([\text{Profit}]) / \text{SUM}([\text{Sales}])$

4.3.2.2. Data visualization

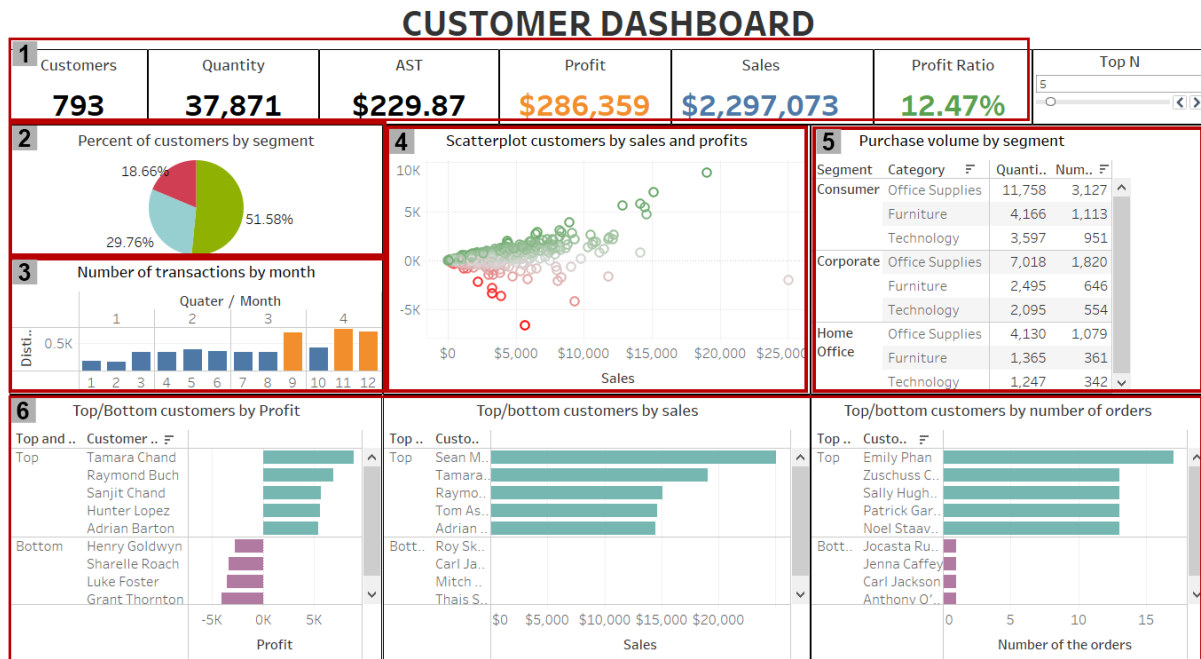


Figure 10. Data visualization of Customer analysis

[1] For the figures that are required to be emphasized, such as: Customer, Quantity, AST, Sales, Profit, Profit Ratio; Card visualizations are utilized to make it simpler and more convenient for viewers to perceive the number.

[2] The pie chart shows the overall proportions of our customer segments.

[3] The graph illustrates how the number of transactions fluctuates by month. Therefore, users can clearly notice the correlation using the column and line combination chart.

[4] Scatterplot customers by sales and profit is used to observe how much each customer generates sales and profit in general.

[5] Data table displays product category information in a grid-like format of rows and columns. It organizes information in a way that's easy to scan so that users can look for patterns.

[6] The dashboard uses Horizontal Bar Charts to show comparison between same-group customers or different-group customers by profit, sales and number of orders. The number of top customers or bottom customers can be displayed in the range from 1 to 50, depending on our setting.

4.3.4. Product Analysis

4.3.4.1. Data Analysis

For Product Analysis, there are 3 key metrics that need to be determined:

Products is the number of distinct counts of product names.

Quantity is the number of products that have been sold.

Manufacturer is the number of distinct counts of manufacturer.

4.3.4.2. Data visualization

In the first row of the dashboard, we have an overview of values that affect **Products**: Products, Quantity, Manufacturer, Discount, Sales, Profit and Profit Ratio.

In the second row, we got the horizontal bar charts and a pie chart. The dashboard uses Horizontal Bar Charts to show comparison between same-group product or different-group product by number of orders. The number of top orders or bottom orders can be set to display in the range from 1 to 50. The pie chart shows the overall proportions of products by category.

And in the last row, we got Scatter charts “Sales, Profit and Profit ratio by Products” indicating how much each product generates sales and profit in general; “Sales, Profit and Quantity by Manufacturer” indicating how much each manufacturer generates sales, profit and quantity. In the middle of this row is a combination of bar chart and line chart. The bar represents **Sales** up to Sub-category of product which is partition by category while the line chart represents **Average discount** of each sub-category.

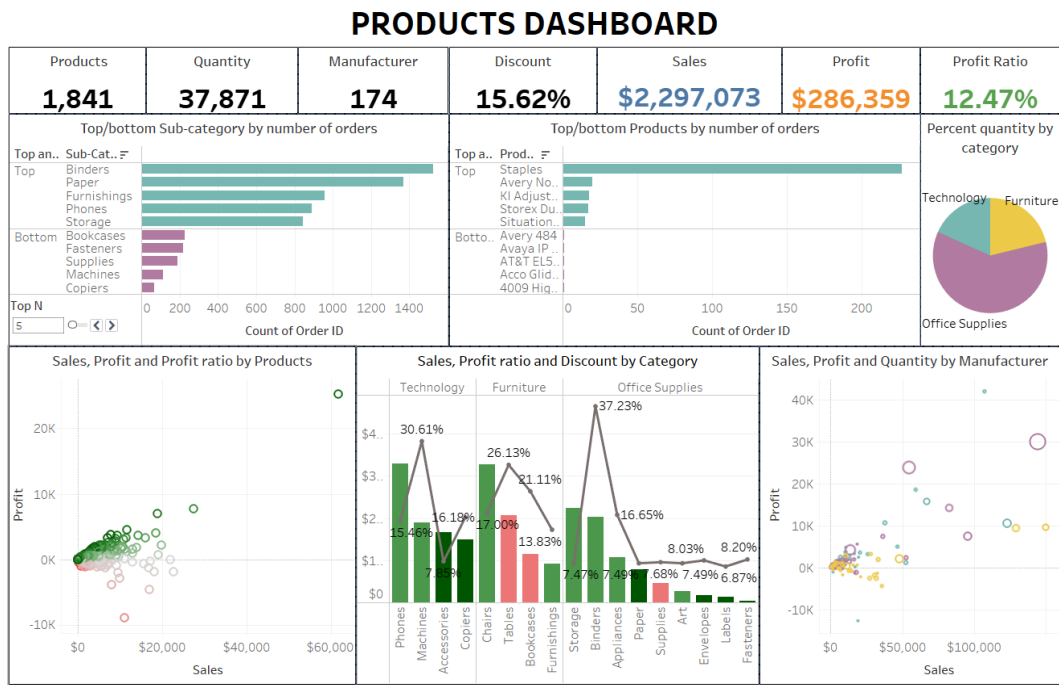


Figure 11. Product Dashboard

4.3.5. Shipment Analysis

4.3.5.1. Data analysis

For Shipment Analysis, we need to determine metric Avg. Ship time:

- Description: average delivery day
- Formula: $AVG([Delivery\ Day])$

4.3.5.2. Data visualization

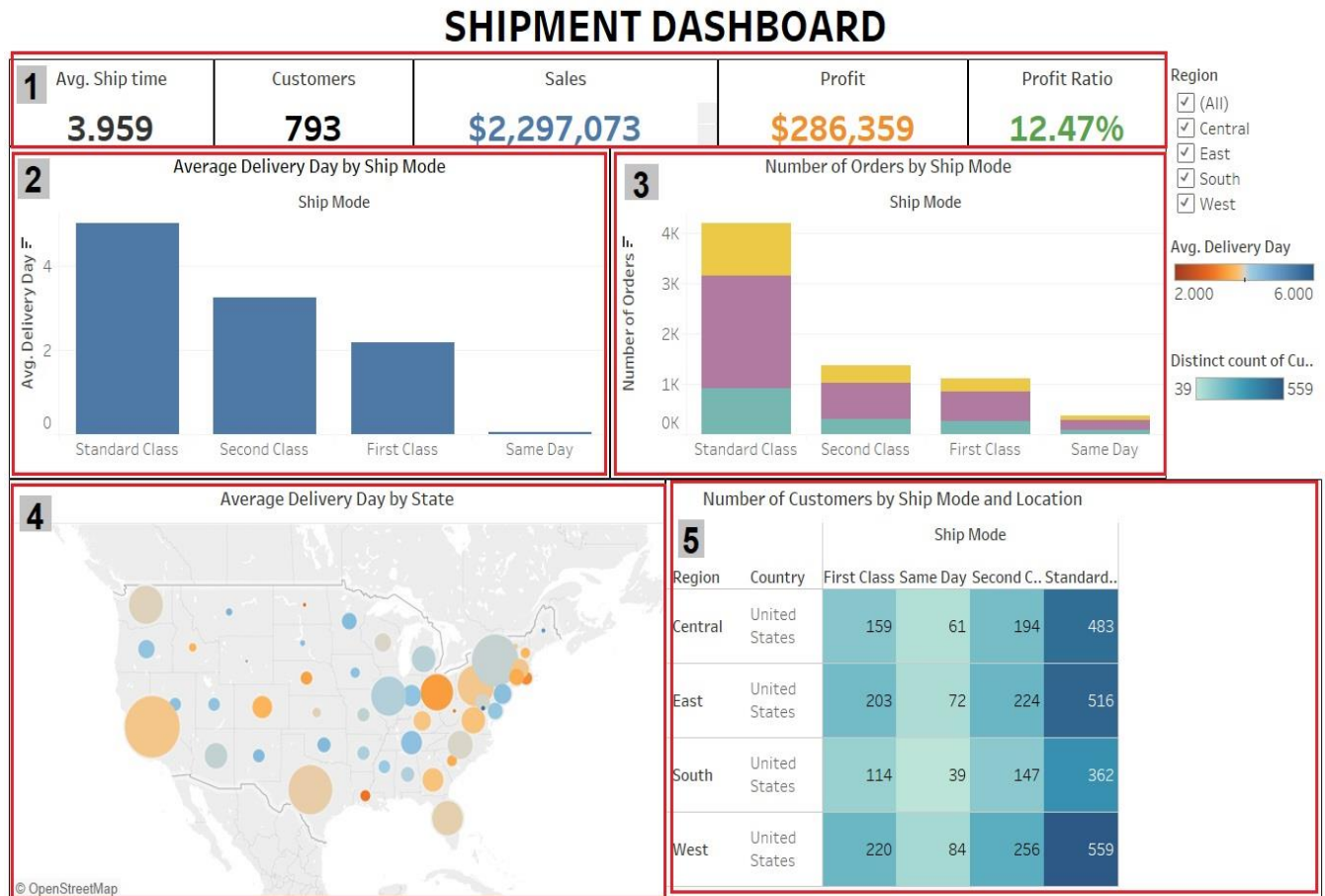


Figure 12. Data visualization of Shipment analysis

[1] This section includes 5 numbers that we want to emphasize to viewers about the shipping time and basic shipping time-related parameters: Average ship time, Number of Customers, Sales, Profit and Profit Ratio.

[2] This chart shows the average delivery day by Ship mode, when using the region filter will show the delivery time for each region.

[3] This chart shows the number of orders by Ship mode and the number orders of each category by Ship mode.

[4] This chart, when viewers filter by region, will show the shipping time by each state in the region. And help viewers know which region/state has the fastest or slowest delivery time.

[5] This is a table of the number of customers who choose to use different types of delivery distributed by region. Based on this table, we can easily find out which Ship mode is used the most or least and which region uses the most, the least type of ship.

The customers in this table are represented by number and chroma, ranging from 39 to 559.

4.4. Discuss and evaluate the results of data analytics

4.4.1. Discussion

4.4.1.1. General analysis

Based on the overview dashboard of Sales Superstore:

- The total number of distinct customers: 793
- Total orders: 5009
- Total of profit: \$286,359
- Total of sales: \$2,297,073
- Products: 1841

Next, to clarify the indicators and the relationship between them, we will show a statistical table for each area.

		Central	East	South	West
Overview		- customers: 629 - total orders: 1175 - sold 1287 products - profit: \$39,719 - sales: \$501,256	- customers: 674 - total orders: 1401 - sold 1406 products - profit: \$91,533 - sales: \$678,553	- customers: 512 - total orders: 822 - sold 1034 products - profit: \$46,721 - sales: \$319,750	- customers: 686 - total orders: 1611 - sold 1494 products - profit: \$108,360 - sales: \$725,514
Sales category	Highest	Technology (\$170,432)	Technology (\$264,994)	Technology (\$148,782)	Furniture (\$252,620)
	Lowest	Furniture (\$163,789)	Office Supplies (\$205,549)	Furniture (\$117,306)	Office Supplies (\$220,881)
Sales by segment	Segment with the highest	Consumer (\$250,040)	Consumer (\$350,947)	Consumer (\$195,602)	Consumer (\$362,908)

	sales				
	Segment with the lowest sales	Home office (\$91,218)	Home office (\$127,184)	Home office (\$74,255)	Home office (\$136,736)

Table 3. General analysis

4.4.1.2. Sales & Profit analysis

For detailed statistics on revenue and revenue growth, we will choose a fixed time frame for all years which is January to make it easier to analyze Month on Month Sales/Profit Growth (MoM%) and Year over Year Sales/Profit Growth (YoY%).

For 2011 sales statistics, the Month on Month Revenue Growth (MoM%) and Year over Year Revenue Growth (YoY%) are completely meaningless as 2011 is the first month that is recorded in the database, so there is absolutely no baseline data of the previous periods recorded for comparison.

Regarding 2011 sales statistics, Year over Year Sales Growth (YoY%) is not recorded as 2011 is the first year in the database. Thus, there is absolutely no baseline data of the previous periods recorded for comparison. The same with January 2011 in data of the Month over Month Sales Growth (MoM%). Therefore only the Sales are meaningful and can be used as the basis for the statistics recorded at later stages. So in 2011 data, we only show the Dashboard results, but do not perform data analysis for the above statistics

Criteria		2011	2012	2013	2014
Overview	Total Sales	\$483.974	\$470.560	\$608.523	\$734.016
	Total Profit	\$49.532	\$61.606	\$81.721	\$93.5
	Profit Ratio	10,23%	13,09%	13,43%	12,74%
	Sales YoY%	-	-2,77%	29,32%	20,62%

	Profit YoY%		-	24,38%	32,65%	14,41%
Sales by Category	Furniture	Total Sales	\$156.907	\$170.518	\$198.910	\$215.390
		Profit	\$5.462	\$3.012	\$6.961	\$3.021
		Profit Ratio	3,48%	1,77%	3,05%	1,4%
	Office Supplies	Total Sales	\$151.782	\$137.248	\$183.531	\$246.566
		Profit	\$22.580	\$25.101	\$35.020	\$39.773
		Profit Ratio	14,88%	18,29%	19,08%	16,13%
	Technology	Total Sales	\$175.285	\$162.794	\$226.082	\$272.060
		Profit	\$21.490	\$33.493	\$39.740	\$50.706
		Profit Ratio	12,26%	20,57%	17,58%	18.64%

Table 4. Overview of Sales and profit from 2011 - 2014

Overview of the business situation of the business, mixed chart of Sales and Profit Ratio in the dashboard of the figure below. Looking at the "Forecast Sales and Profit" line chart, the company will better understand the changing trend of sales and profits year on year. The company's sales tended to be low in Q1 and increased strongly towards the end of the year. The growth is shown in more detail in the table "Sales Growth and profit Ratio", sales increased strongly in March of all 4 years, especially in 2011 with an increase of 1.058%, the growth is also shown in March, September and November every year. In terms of profits, the company generally earns high profits at the beginning and the end of the year. Notably, in 2011 and 2012, the company made a loss but in the following 2 years, it improved significantly.

SALES AND PROFIT DASHBOARD

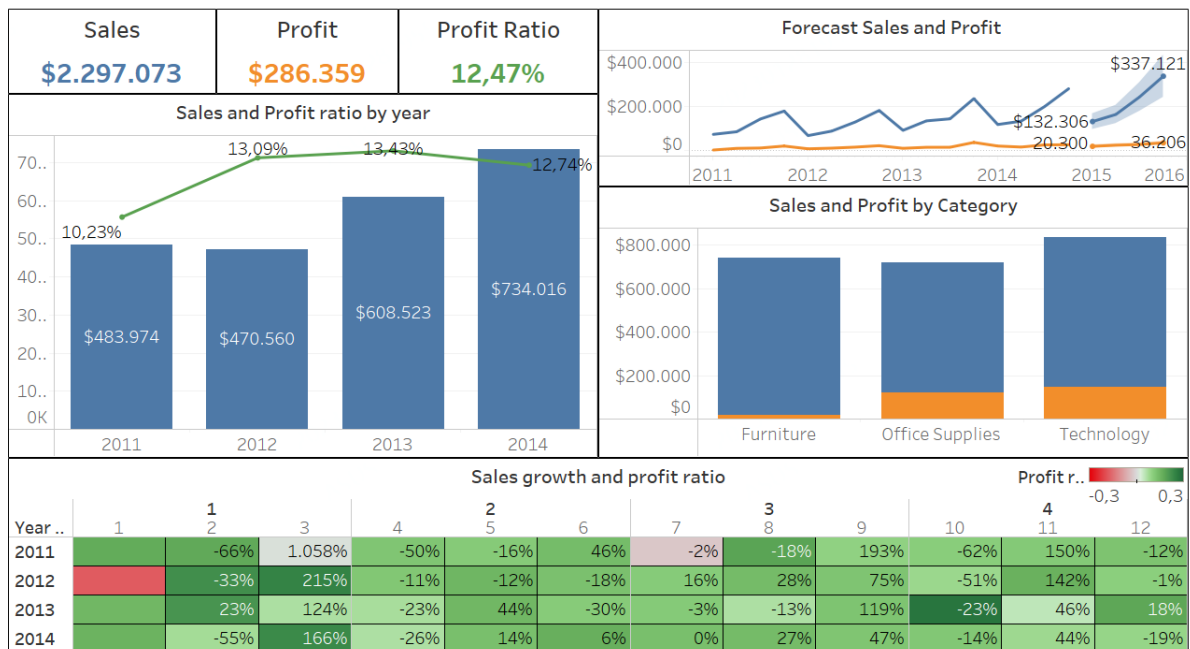


Figure 13. Sales & Profit Dashboard

The data in the "Sales and Profit ratio by year" chart clearly showed that in 2014 the company achieved its highest four-year sales volume of \$734,016. The opposite happened in 2012 with relatively modest sales of \$470,560.

a) Business situation in 2012

In 2012, the company's sales were the lowest over 4 years but the profit and profit ratio was still higher compared to 2011. We can see that in the first January of the year, the company accepted to receive losses in order to make sales. Therefore, the company's profit ratio stopped at a negative place at -18.03%. In February, although sales decreased by approximately 33% compared to the previous month, the company started to achieve a quite high profit with a profit ratio of about 23%.

March is the time when the company attained the best business performance with sales growth of up to 215%, which means the highest profit. This shows that the company's business strategy is showing positive signs. In the following months, sales fluctuated up and down but not significantly. The profit ratios were still positive in the range of 10% - 15%. Notably, at the end of the year, October saw a sharp drop of 51.4% in sales, even though sales increased in September. Then sales increased strongly again

in November with approximately 142% due to Black Friday season help promote the shopping demand of customers.

On the other hand, based on the bar chart "Sales and Profit by Category", all the three categories — Furniture, Office Supplies and Technology made different amount of sales. Furniture had the great sales but making poorest profit. While Technology had lower selling but being the most profitable over 3 categories.

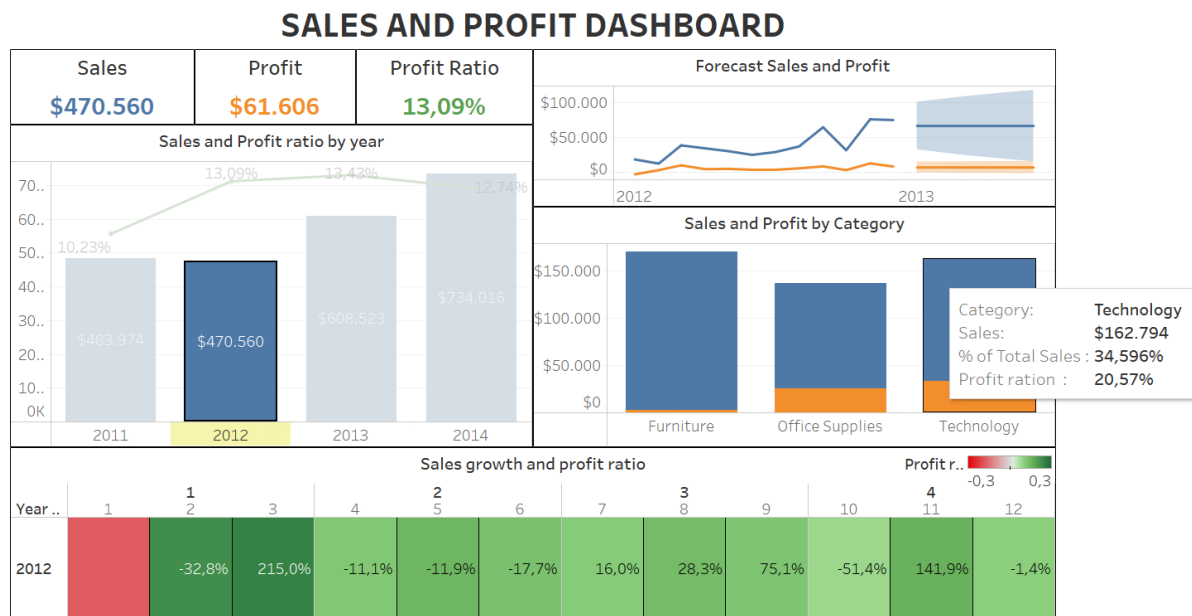


Figure 14. Sales and Profit in 2012

On the other hand, based on the bar chart "Sales and Profit by Category", all the three categories — Furniture, Office Supplies and Technology made different amount of sales. Furniture had the great sales but making poorest profit. While Technology had lower selling but being the most profitable over 3 categories.

b) Business situation in 2014

The bar chart "Sales and Profit Ratio by year" illustrates that the sales performance of 2014 was the best over 4 years, at \$734,016. The "Forecast Sales and Profit" line chart shows that the business situation of 2014 is not too fluctuated. Similar to 2012, the sales trend of 2014 also decreased at the beginning of the year and stabilized in the middle of the year. By the end of the year, sales tend to increase positively again. In addition, the company's profit ratio did not record a negative value, and the profit ratio color band was green for the whole year, indicating that the company earned a relatively stable profit. In summary, March and November are ideal times for sales and profit

growth. The company should keep an eye on these times to up more sales. In addition, October needs some promotions to better improve sales

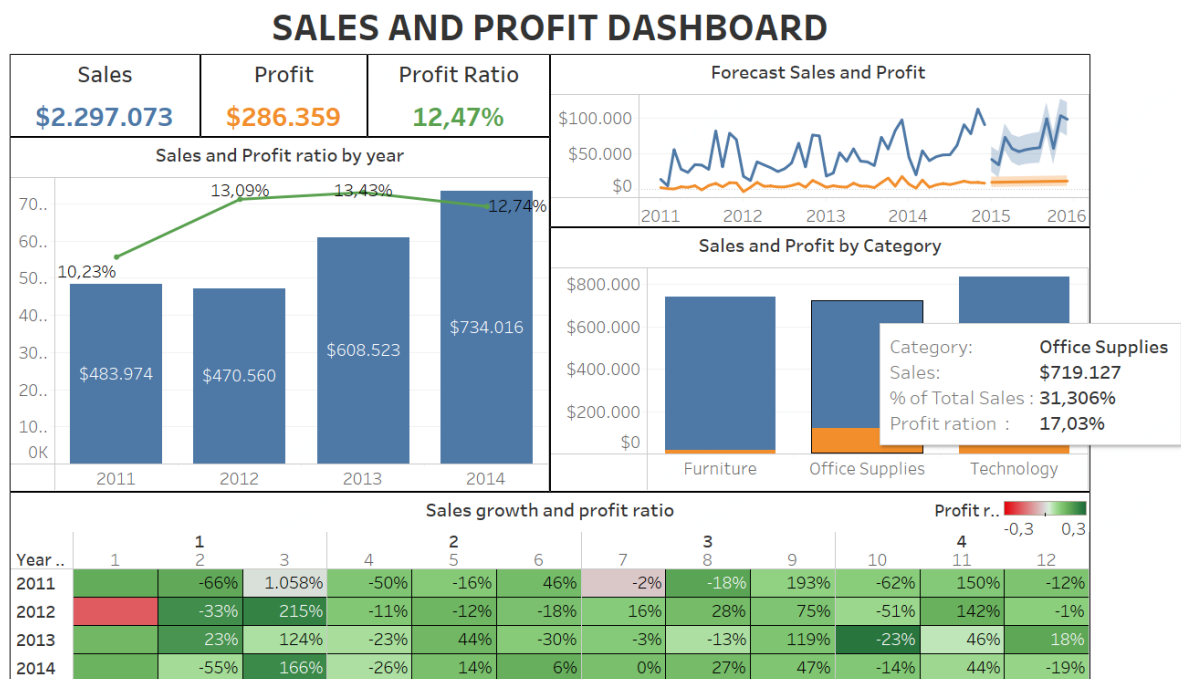


Figure 15. Sales and Profit in 2014

In terms of sales and profits by Category, Technology is best selling and is the most Profitable too. Office Supplies also bring high sales and profits for the company. Furniture, despite having the second highest sales, still has the lowest profit. It seems that this product segment has a high price, so the company has to reduce the price a lot to attract customers to buy.

There's a need to set up a price marketing strategy for the Furniture segment to increase profits for the company. The sales of the furniture segment are quite high but having low profit, means that customers are providing low profit from high sales. The company should think about adjusting the price of the products in this category to see if this marketing price strategy can generate more profit from these sales. If we lower the prices a little and limit the discounts, customers may find reasonable prices, therefore, we don't need to discount a lot but still gain positive profit

4.4.1.3. Customer analysis

These are some overall numbers of customer analysis

- Customer (Total of customers): 793

- Quantity (quantity of orders): 37,871
- AST (Average sales per transaction): \$229.87
- Sales (Total of sales): \$2,297,073
- Profit (Total of profit): \$286,359
- Profit Ratio (Profit/Sales): 12.47%
- Percentage of customers by segment: Consumer (51.58%), Corporate (29.76%), Home Office (18.66%)

More than half of our current customers (51.58%) are categorized as "Consumer". This segment visibly outnumbers others. The remaining customers are "Corporate" or "Home Office", with the figures being 29.76%, and 18.66% respectively.

Segment	Metric
Consumer	<ul style="list-style-type: none"> ● Quantity (quantity of orders): 19,521 ● AST (Average sales per transaction): \$223.75 ● Sales (Total of sales): \$1,161,497 ● Profit (Total of profit): \$134,113 ● Profit Ratio (Profit/Sales): 11.55%
Corporate	<ul style="list-style-type: none"> ● Quantity (quantity of orders): 11,608 ● AST (Average sales per transaction): \$233.84 ● Sales (Total of sales): \$706,183 ● Profit (Total of profit): \$91,965 ● Profit Ratio (Profit/Sales): 13.02%
Home Office	<ul style="list-style-type: none"> ● Quantity (quantity of orders): 6742 ● AST (Average sales per transaction): \$240.96 ● Sales (Total of sales): \$429,393 ● Profit (Total of profit): \$60,281 ● Profit Ratio (Profit/Sales): 14.04%

Table 5. Metric of customer analysis

The ASTs are noticeably equivalent, hence the other metrics vary due to the difference in size between the segments. All three segments have approximate impact so there are no prominent segments.

- Number of transactions by month:

The graph illustrates that the number of transactions by month fluctuates throughout the months. It starts off with about 160-180 transactions in the first two months of the year. Then, the amount of transactions soars over 300 and remains constant for the next six months (from March to August). During the last four months, we witnessed a double increase in the number of transactions, (especially September, November and December). This may be due to the discount, or urgent need during that period. Our business could offer a variety of new products during this quarter to attract our current and new customers.

- Scatterplot customer by sales and profits:

To determine how much each client contributes to overall sales and profit, utilize a scatterplot of customers by sales and profit.

- Purchase volume by segment

In all three segments, “Office Supplies” is the best selling category; furniture and technology respectively follow. This pattern remains unchanged throughout the year.

- Top/Bottom customers by profit, sales, orders:

This list of our important customers that need special attention, customer services, which could be business gifts, exceptional discounts to retain them. In all three top lists, “Consumer” segment outnumbers the others.

4.4.1.4. Product analysis

Based on the Products dashboard, we have an overview of product analysis

- Products (Distinct count of product name): 1,841
- Quantity (quantity of orders): 37,871
- Manufacturer (Distinct count of manufacturer): 174
- Discount (Average discount): 15.62%
- Sales (Total of sales): \$2,297,073
- Profit (Total of profit): \$286,359
- Profit Ratio (Profit/Sales): 12.47%

Office Supplier accounts for more than half of quantity (60.48%). The remaining orders are divided into "Technology" and "Furniture", with the figures being 18.32%, and 21.19% respectively.

Category	Metric
Technology	<ul style="list-style-type: none"> ● Products: 421 ● Quantity: 6,939 ● Manufacturer: 63 ● Avg. Discount: 13,23% ● Sales: \$836,221 ● Profit: \$145,429 ● Profit Ratio: 17.39%
Furniture	<ul style="list-style-type: none"> ● Products: 380 ● Quantity: 8,026 ● Manufacturer: 45 ● Avg. Discount: 17,39% ● Sales: \$741,725 ● Profit: \$18,456 ● Profit Ratio: 2.49%
Office Suppliers	<ul style="list-style-type: none"> ● Products: 1,050 ● Quantity: 22,906 ● Manufacturer: 87 ● Avg. Discount: 15,73% ● Sales: \$719,127 ● Profit: \$122,474 ● Profit Ratio: 17.03%

Table 6. Metrics of product analysis

-Although Technology accounts for 18.32% of total quantity, its profit takes half of total profit. It also has the least Avg. Discount in 3 categories. Two scatter charts of Technology are very positive although there are still some negative profit products.

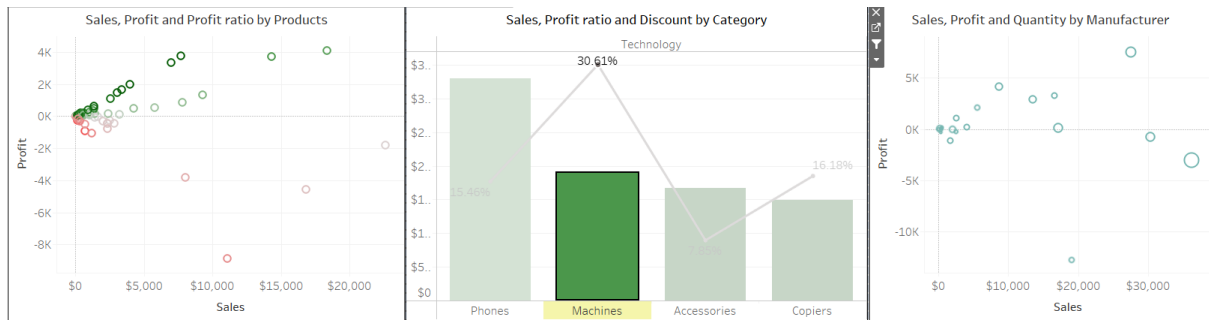


Figure 16. Statistics of Machines in Technology

In Machines sub-category, lower Discount to maximize profit is worth trying.

- Furniture has Sales at \$741,725 but its profit ratio is only 2.49%, much lower than it should have. To figure out this case, we may need to dive deep down in “the Sales, Profit and Profit ratio by Products” dashboard. Based on the dashboard below, nearly half of furniture products have negative profit (the profit ratio range from -1% to -29%)

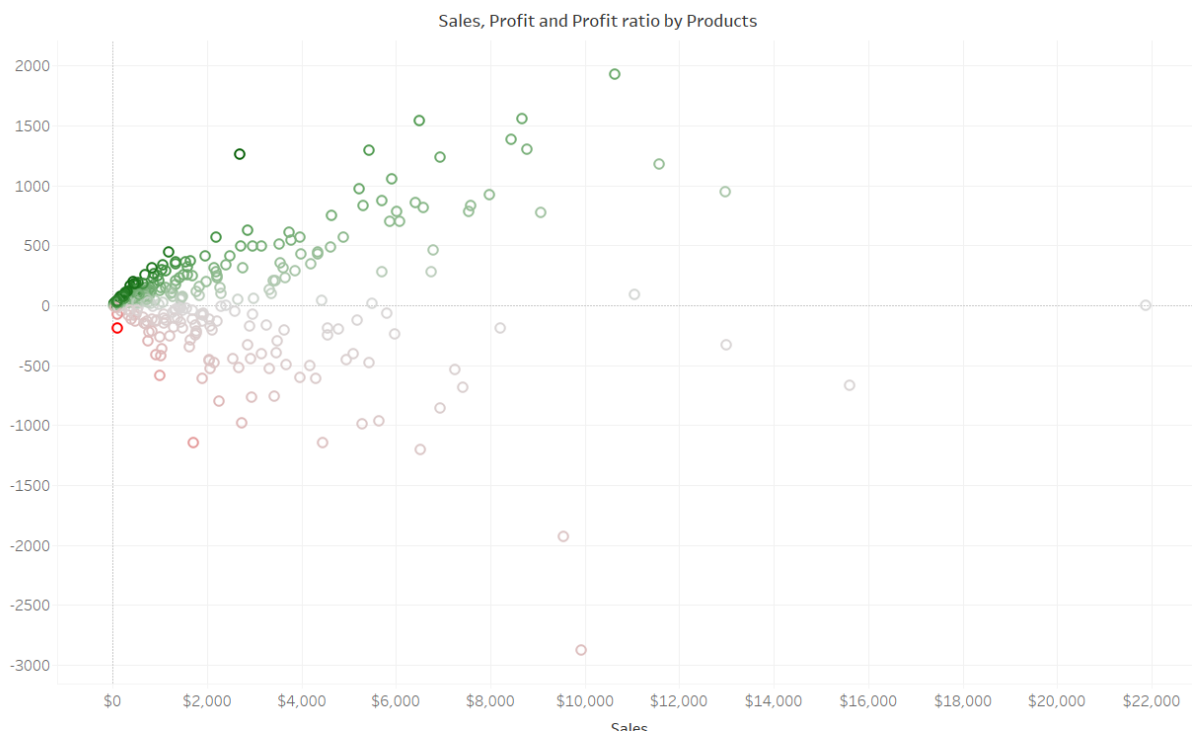


Figure 17. Sales, Profit and Profit ratio by Products of Furniture

Let’s have a glance at “Sales, Profit ratio and Discount by Category” of Furniture, Tables and Bookcases are sub-category that have negative profit in spite of its high average discount. It’s the multi-factor case. Maybe, highly discount, which is thought to help Sales static, is the main factor causing profit to be negative. We should try to lower its Avg. Discount and its stock, reallocate our resources to Chairs and Furnishings.

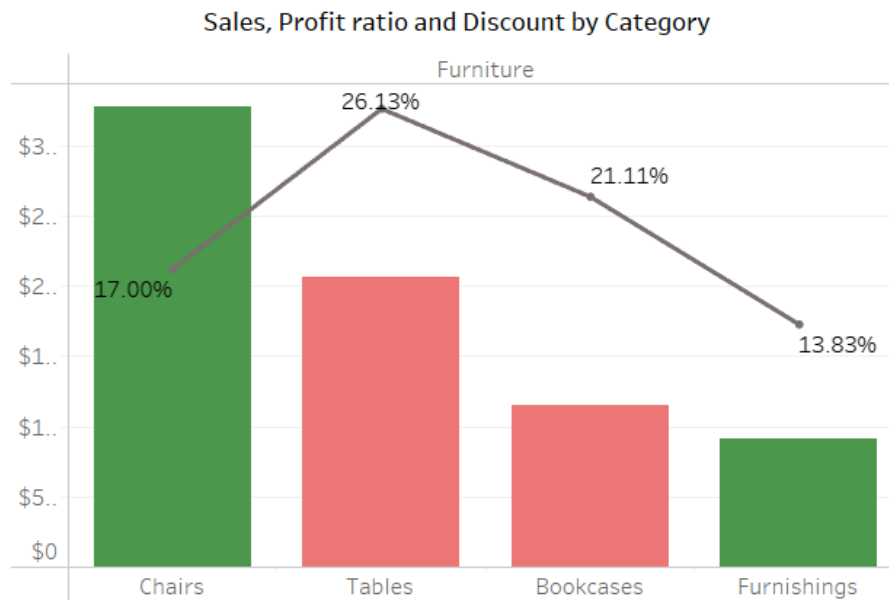


Figure 18. Sales, Profit ratio and Discount by Category of Furniture

In the Table sub-category of Furniture, most of the manufacturers do not work efficiently. Cutting the workforce and reallocating quantity in positive manufacturers may be the solution. Bookcases sub-category should apply the same solution.

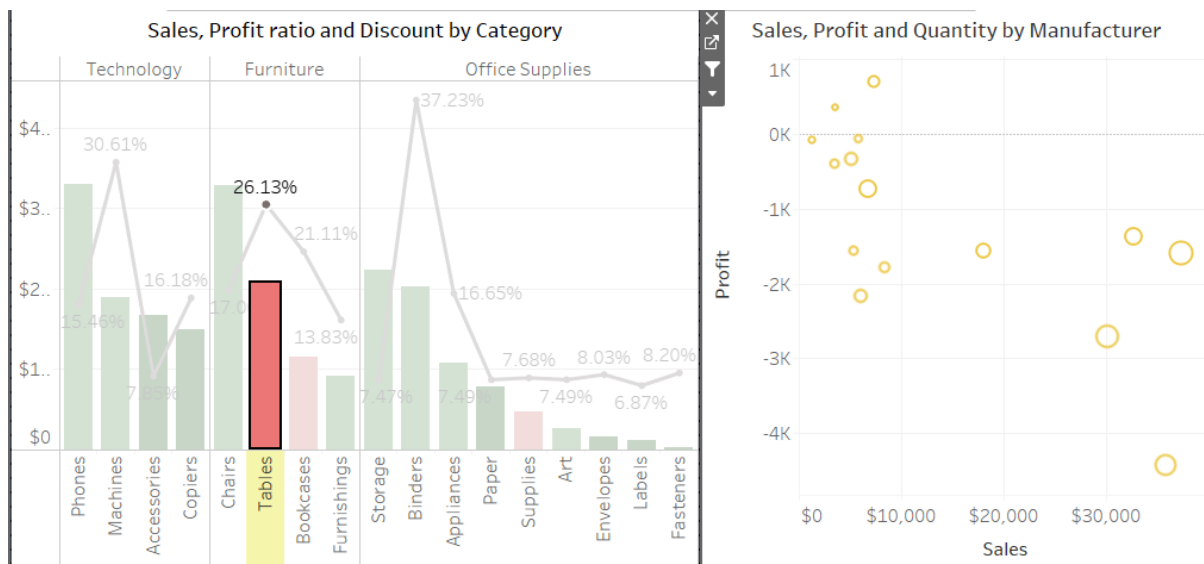


Figure 19. Sales, profit and quality by manufacturer of sub-category

Office suppliers is the category that need the least adjustment. If we must do something about it, my suggestion is still to reallocate resources and to adjust Average discount.

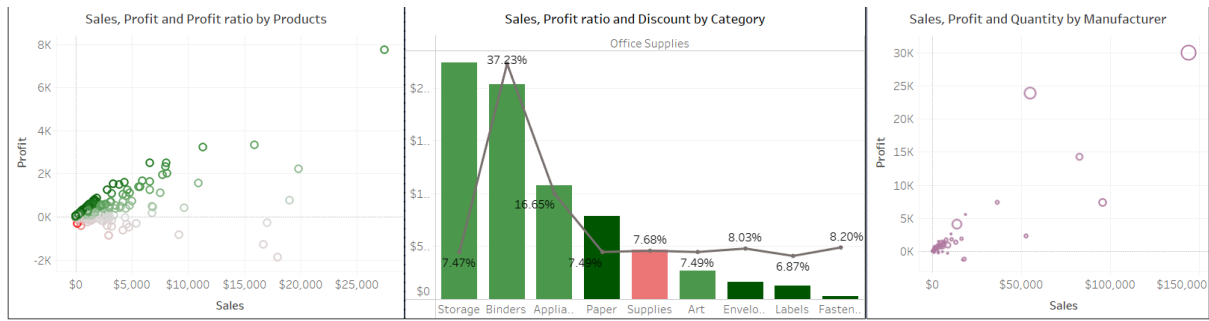


Figure 20. Statistics of Office suppliers category

4.4.1.5. Shipment analysis

Based on the dashboard of Shipment analysis above, we have an overview of Shipment analysis:

- Average delivery time (Avg. ship time): 3.959
- Total of distinct customers (Customers): 793
- Total of sales (Sales): \$2,297,073
- Total of profit (Profit): \$286,359
- Profit Ratio (Sales/Profit): 12.47%

Next, to clarify the indicators and the relationship between them, we will show a statistical table for each region.

Overview		Central	East	South	West
		avg. ship time: 4.058	avg. ship time: 3.91	avg. ship time: 3.96	avg. ship time: 3.93
Average Delivery Day by Ship Mode	Standard class	4.99	4.971	4.998	5.058
	Second class	3.331	3.266	3.198	3.172
	First class	2.218	2.222	2.111	2.097
	Same day	0.033	0	0.012	0.103

Average Delivery Day by State	Fastest	North Dakota: 2.857	West Virginia: 3.0	Louisiana: 3.048	California: 3.868
	Slowest	Oklahoma: 4.485	District of Columbia: 5.7	Tennessee: 4.383	Wyoming: 5.0
Number of Customers by Ship Mode	Highest order	Standard class: 483	Standard class: 516	Standard class: 362	Standard class: 559
	Lowest order	Same day: 61	Same day: 72	Same day: 39	Same day: 84

Table 7. Shipment analysis

Based on the statistical table above, it is easy to see the average delivery time by region and ship mode:

- The most chosen ship mode by customers is the “Standard class”, the number of customers ranging from 300 to more than 500. Besides, “Same day” is the type of ship that is least used by customers, ranging from 30 to 90 customers. Need to find a solution to help ship mode “Same day” be trusted and used more widely by customers.
- The “Same day” ship mode of the west area always completes orders on the same day of ordering (avg. delivery day =0).
- The ship mode “Standard class” of the West region has the highest number of customers but has the longest delivery time in the 4 regions. Business needs to improve delivery quality to achieve customer satisfaction
- In addition, the “Standard class” ship mode of the South region has the fastest delivery time in the 4 regions, but the number of orders is the smallest. Businesses need to implement marketing strategies to attract more customers to achieve business efficiency.

4.4.2. Evaluation

Statistical Analysis	Evaluate	Development Suggestion
General Analysis	Giving an overview of core business indicators help users easily understand the general business situation.	Develop a combination of comparing actual business metrics and business KPIs
Sales and Profit Analysis	<p>- Sales tended to grow strongly after 2012 while the profit ratio decreased slightly, indicating that the company's strategy to boost sales is appropriate and correct. However, the company needs to pay more attention to the strategy for profit</p>	<p>- The company's strategy should focus on both sales and profit growth. It may not be worth focusing any efforts on customers who generate low sales and low profits. Thus, retarget customers as well as potential customers that provide high sales - high profit and low sales - high profit</p> <p>- Geography: Proposing business expansion in high-revenue areas, implementing campaigns to entice customers in low-revenue areas.</p> <p>- Build buying habits for customers in special time periods (March, September, November) to increase sales, then make price adjustments in the following year to get higher profits</p> <p>- The company can test the strategy of creating many promotions in the middle of the year (June, 7) to increase the company's sales</p> <p>- Adjust the selling price of</p>

		product category with high sales but low profits (here is Furniture segment) to observe customer needs and buying behavior
Customer Analysis	<ul style="list-style-type: none"> - Data of customers is too limited to gain detailed insights. - The ASTs are noticeably nearly equivalent. There are no prominent segments because all three segments have almost the same impact. - During the last four months, we witnessed a double increase in the number of transactions, (especially September, November and December). - “Consumer” segment outnumbers the others in the top lists. 	<ul style="list-style-type: none"> - There shall be a necessity for changes in the customer data collections process. Current information is not specific enough to define our actual target segment. - The company should concentrate on attracting more and more customers in all categories in lieu of only focusing on the biggest segments. If there is a noticeable difference in ASTs, the company may take into account the fact that which segments should be our target segments. - Gathering other data to find out the reasons behind large sales in the last quarter. Why do our customers love to purchase during that time (discount, urgent need,...) or is this a common situation for all businesses? - Importing new products, especially office supplies, and offering a variety of goods should be considered and completed before the last quarter. This helps

		<p>to retain customers, captivate new ones, prevent shortage for the biggest demand of the year.</p> <ul style="list-style-type: none"> - Business gifts, exceptional discounts, special offers... should be given to the top customers in order to maintain good relationships with them. - Figuring out which other products that they may like to expand their regular purchase list.
Product Analysis	<ul style="list-style-type: none"> - Furniture category is the one that needs adjustment due to its negative profit sub-category. - The other categories seem okay at first but there is something to do when we dive into the details. Most of the cases, the manufacturers do not work effectively. Besides, the average discount is too high. Perhaps, it's just the temporary solution for long-term inventory. 	<ul style="list-style-type: none"> - Priority should be given to resource reallocation (both changing manufacturer and changing product in the same manufacturer) instead of removing items because the variety of goods is necessary. Keeping diversity of products is how the company keeps its position in the market.
Shipment Analysis	<ul style="list-style-type: none"> - In general, the average delivery time is 3,959 days per order. - The most used ship mode is Standard class (average delivery time is 5,007 days/order), while Same day is the least used (average delivery time is 0.044 days/order). 	<ul style="list-style-type: none"> - Need to adjust the delivery time of "Standard class", especially the west area - Implement marketing plans, attract customers in the south region - Looking for solutions to help customers approach and use ship

	<ul style="list-style-type: none"> - Orders are spread evenly in each area but are more concentrated in the west region, the number of orders is gradually decreasing in the south area. - “Standard class” of the West region has the highest number of customers but has the longest delivery time. - “Standard class” ship mode of the South region has the fastest delivery but the number of orders is the smallest. 	mode “Same day”
--	--	-----------------

Table 8. Evaluation table

Chapter 5: CONCLUSION

5.1. Results

5.1.1. Advantages

- Members actively contribute in meetings, participate fully and complete the project completely.
- Apply analytical methods and learn knowledge to the project effectively.
- Know how to use Tableau proficiently to analyze data.
- Have basic knowledge of data analysis in business.

5.1.2. Limitations

- The number of chart types is limited, there is not a variety of charts in the report.
- The data is small and relatively clean, not including all the actual operations of a business enterprise.
- Not having much experience in designing and using functions, so they have not given new and important insights.
- Tableau's algorithms and advanced functions have not been applied to data analysis.

5.2. Future works

- Our team will cultivate knowledge as well as technical to develop the dashboard to gain deeper insight into the customer's experience. Moreover, the deep learning model is also expected to be applied to the large dataset
- Combine multiple data sources, including more information close to reality.
- Use a variety of additional tools and integrate a variety of charts to most effectively represent data, providing a variety of insights.
- Apply machine learning techniques to see the correlation between many metrics, so that can help build predictive models
- Learn about algorithms, advanced functions, to build a seamless, vivid story that attracts viewers.
- Real-time data analysis.

APPENDIX

Contribution assessment table:

Student ID	Name	Contribution
K194111605	Phan Văn Hiếu	100%
K194111621	Nguyễn Thị Quỳnh Như	100%
K194111624	Lương Trường Phước	100%
K194111627	Nguyễn Thị Thanh Tâm	100%
K194111634	Nguyễn Trần Huyền Trang	100%

REFERENCES

- [1] Master Class (Nov 3, 2021). How to Calculate Year-Over-Year Growth: Pros and Cons of YOY. Retrieved Jan 14, 2023 from <https://www.masterclass.com/articles/year-over-year-growth-explained>
- [2] Wall Street Prep. Month over Month Growth (M/M). Retrieved Jan 14, 2023 from <https://www.wallstreetprep.com/knowledge/month-over-month-growth-rate/>
- [4] Ben Clark. Exploring The Theories of Data Analysis: EDA, CDA, and Grounded Theory. idashboards.com. Retrieved Jan 08, 2023 from: <https://www.idashboards.com/blog/2018/02/28/exploring-the-theories-of-dataanalysis-eda-cda-and-grounded-theory/>
- [4] CFI Team (December 22, 2022). Data Anaytics. Retrieved Jan 08, 2023 from <https://corporatefinanceinstitute.com/resources/data-science/data-analytics/>