

UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEM



FINAL PROJECT

**TOPIC: COMPREHENSIVE LOGISTICS DATA
ANALYSIS AND VISUALIZATION WITH
TABLEAU**

Subject: Data Visualization

Group:

Lecturer:

Ho Chi Minh,

MEMBERS OF GROUP

No.	Student ID	Full name	Performance
1	K214160990	Phạm Công Nguyễn Khôi	100%

ACKNOWLEDGEMENTS

COMMITMENT

LIST OF FIGURES

Figure 3.1 Data Analysis Workflow	26
Figure 3.2 Data Processing Workflow Diagram	27
Figure 3.3 Extract data from necessary columns	28
.Figure 3.4 Duplicate the data tables	29
Figure 3.5 Data Preprocessing	30
Figure 3.6 Join the data	31
Figure 3.7 Data Modeling Diagram for CSV Files	31
Figure 4.1 Dashboard Overview	38
Figure 4.2 Sales Dashboard Overview	39
Figure 4.3 Logistics Dashboard	41
Figure 4.4 Customer Dashboard	42
Figure 4.5 Order Management Dashboard	44

LIST OF TABLES

Table 3.1 Order Item Details	32
Table 3.2 Product Details	33
Table 3.3 Order Details	34
Table 3.4 Departments Details	35
Table 3.5 Customer Details	35
Table 3.6 OrderDate Details	36
Table 3.7 ShippingDate Details	37
Table 3.8 OrderProcess Details	37

TABLE OF CONTENT

CHAPTER 1. ABOUT PROJECT	11
1.1. Business case for the project	11
1.2. Objectives of the project	12
1.2.1. General Objective	12
1.2.2 Specific Objectives	12
1.3. Research Objects	13
1.4. Scope of the project	13
1.5 Value and desired outcome of the project	13
1.6 Structure of project	14
CHAPTER 2: THEORETICAL BASIS	15
2.1 Data Modeling and Schema Design	15
2.1.1 Data Modeling	15
2.1.2 Data Schema	15
2.1.2 Tableau data model	16
2.2 Data Preparation by using Tableau Prep Builder	16
2.3 Dashboard Design Principles	16
2.4 Data Integration and Pipeline Management	18
2.5 Presentation and Reporting Techniques	19
2.6 Theoretical Concepts in the Logistics Domain	22
2.6.1 Introduction to Logistics	22
2.6.2 Key Metrics in Logistics	22
2.6.3 The Role of Data Analytics in Logistics	23
3.1 Overview of the Dataset	25
Summary	25
Purpose of Use	25
Scope of the Data	25
3.2 Workflow	27
3.2 Requirements analysis	28
3.3 ETL Process	29
3.4 Data Modeling	33
3.4.1. Order Items Table	33
3.4.2 Products Table	34
3.4.3 Orders Table	35
3.4.4 Departments Table	36
3.4.5 Customer Table	36
3.4.6 OrderDate Table	37
3.4.7 ShippingDate Table	38

3.4.8 OrderProcess Table	38
CHAPTER 4: RESULTS – DATA ANALYTICS AND VISUALIZATION	40
4.1. Report and dashboard systems (Report tree)	40
4.1.1 Overview	40
4.1.2 Sales	41
4.1.3 Logistic	43
4.1.4 Customer	44
4.1.5 Order Management Dashboard	46
CHAPTER 5: CONCLUSION AND FUTURE WORKS	48
5.1. Results	48
5.2. Limitations	48
5.3. Future works	49
REFERENCES	50

ABSTRACT

This final project, titled "Comprehensive Logistics Data Analysis and Visualization with Tableau," aims to enhance business performance management through advanced data visualization. Utilizing the DataCo Supply Chain dataset, we transformed a complex, multi-column dataset into streamlined tables using Tableau Prep. This enabled the creation of five targeted dashboards: Overview, Sales, Logistics, Customers, and Order Management.

Each dashboard provides specific insights: the Overview Dashboard offers key metrics like total sales and profit; the Sales Dashboard analyzes revenue trends and regional performance; the Logistics Dashboard focuses on shipping efficiency; the Customers Dashboard highlights customer behavior and retention; and the Order Management Dashboard examines product performance and profitability.

While the project successfully delivers actionable insights, it faces limitations such as data quality issues and a restricted scope to the 2015-2018 period. Future improvements include integrating recent data, enhancing analytical methods, and improving user training.

This project demonstrates Tableau's capability to convert raw data into valuable business insights, enabling companies to make informed decisions, optimize operations, and achieve sustainable growth.

CHAPTER 1. ABOUT PROJECT

1.1. Business case for the project

In the competitive landscape of today's markets, businesses must prioritize efficient and accurate business performance management. These factors are crucial for maintaining operational efficiency and ensuring customer satisfaction. Traditional systems often struggle to provide timely and accurate information, leading to inefficiencies. Adopting a dashboard system built with Tableau presents a strategic approach to addressing these challenges, leveraging advanced technology to enhance management and decision-making capabilities.

Effective business performance management is essential for minimizing risks and optimizing business operations. The dynamic nature of the market and the variables in business operations require a system that not only provides real-time information but also integrates seamlessly with existing operational frameworks to enhance decision-making processes. The need for a solution that offers scalability, flexibility, and real-time data access underscores the importance of this business case.

The Tableau platform offers a compelling suite of features that make it an ideal choice for businesses aiming to enhance business performance management. Central to its strengths is powerful data visualization, where Tableau enables efficient data display and analysis. This allows businesses to gain a comprehensive view of key business metrics across all aspects, thereby maintaining optimal operations and responding to market demands. Additionally, Tableau's data-driven model offers significant cost efficiency by reducing the need for large capital investments in IT infrastructure. Businesses benefit from Tableau's scalable services, paying only for the resources they use, which is particularly advantageous in fluctuating market conditions.

Moreover, Tableau enhances operational agility through real-time data processing and advanced analytical tools, allowing businesses to rapidly adapt to market changes. This capability is essential for efficiently managing business operations and adjusting strategies promptly. The platform's scalability and flexibility further empower businesses, enabling them to adjust operations based on demand patterns and explore new business opportunities without the limitations of physical infrastructure. Finally, Tableau's commitment to security and regulatory compliance ensures that sensitive customer and operational data are protected, meeting industry standards and fostering a secure business environment. These features collectively make Tableau a robust and strategic choice for businesses looking to optimize business performance management.

Implementing Tableau for enhanced business performance management is a strategic initiative that aligns with the goals of reducing operational costs, avoiding mismanagement, and improving customer satisfaction. This business case is pivotal for companies looking to leverage technology to streamline operations and respond adeptly to market changes. Tableau's comprehensive suite of tools and scalable solutions provides a robust platform for businesses to optimize their business performance management practices. The adoption of Tableau not only supports day-to-day operational needs but also contributes to long-term business growth and sustainability in a competitive marketplace. This approach not only optimizes resource allocation but also positions the company for future success by enhancing operational flexibility and responsiveness.

1.2. Objectives of the project

1.2.1. General Objective

The general objective of the project is to build a comprehensive dashboard system using Tableau to manage and visualize business performance, thereby supporting strategic decision-making for the enterprise. This system will provide real-time information, support in-depth analysis, and offer an overview of key business metrics.

1.2.2 Specific Objectives

Within the framework of this project, we have set specific objectives aimed at optimizing management and business performance through the deployment of a dashboard system on the Tableau platform

Firstly, develop an overview dashboard to allow managers to grasp the business situation as a comprehensive picture. This dashboard will integrate data from various sources, providing a broad and insightful view of business operations, thereby helping managers make accurate and timely decisions.

Secondly, develop a logistics dashboard to clearly assess the status of orders. This dashboard will provide detailed information on the status and progress of orders from receipt to delivery, helping to identify and address logistical issues promptly.

Beside that, build a customer dashboard to help the business better understand each customer group. This dashboard will analyze and display shopping trends, preferences, and behaviors of customers, thereby supporting the development of more effective marketing and sales strategies.

More, design a dashboard about ordered items, helping the business understand more about the products sold. This dashboard will provide detailed information on

quantities, sales, and best-selling products, supporting inventory management and production planning.

Finally, construct a sales dashboard to help the business grasp important business metrics such as revenue, costs, and profits. This dashboard will be an effective tool for monitoring business performance, evaluating the effectiveness of strategies, and making necessary adjustments.

By implementing these specific objectives, the project will significantly improve the company's management and operational capabilities, contributing to sustainable and long-term development in an increasingly competitive environment.

1.3. Research Objects

In this study, the primary research subjects include small and medium-sized enterprises (SMEs) in the retail sector, particularly those that are implementing information technology to enhance business performance. The research also extends to data management and business analytics platforms, focusing on the dashboard systems developed on Tableau. The study aims to analyze and evaluate the effectiveness of technology implementation in management and operations, through monitoring, analyzing, and displaying data efficiently.

The goal is to gain a deeper understanding of how these tools can assist businesses in making strategic decisions, improving workflow processes, and enhancing their competitive stance in the market. Additionally, the study seeks to explore the challenges and barriers that businesses might encounter while deploying and using these tools in real scenarios.

1.4. Scope of the project

- Space: University of Economics and Law
- Time: 15/07/2024 - 05/08/2024

1.5 Value and desired outcome of the project

This project aims to deliver significant value to businesses by providing an effective business performance management system through the deployment of dashboards on the Tableau platform. The core value of the project lies in its ability to enhance strategic decision-making based on accurate and continuously updated data, thereby boosting business efficiency and optimizing operational processes.

The expected outcomes from implementing this project include a substantial improvement in the management and operation of business activities. The designed dashboards will enable managers to not only monitor the overall business situation but also delve deeper into shopping trends, logistics performance, and a thorough understanding of customer behavior. This will support timely and accurate

decision-making, helping businesses to quickly adapt to market changes and enhance their competitive edge.

Additionally, the implementation of these solutions will also help businesses reduce operational costs by decreasing reliance on traditional IT infrastructure and increasing resource flexibility. This not only supports effective cost management but also contributes to the sustainable and long-term development of the business in today's competitive environment.

By successfully achieving these objectives, the project will bring substantial benefits to the company, not only in optimizing daily business operations but also in establishing a solid foundation for long-term and sustainable growth.

1.6 Structure of project

This project has 5 chapters:

Chapter 1: About the Project - This section presents the context and significance of business performance management, describing why the deployment of a dashboard system on the Tableau platform is necessary.

Chapter 2: Theoretical Basis - Analyzes theories and research models related to performance management and information technology, specifically the application of Tableau in supporting business decision-making.

Chapter 3: Study Framework - Details the research methodology, subjects, and specific objectives of the project, as well as the methods for data collection and analysis.

Chapter 4: Results – Data Analytics and Visualization - Presents analyses from the collected data and displays it through dashboards, highlighting important findings and insights from the data.

Chapter 5: Conclusion and Future Works - Summarizes the findings, evaluates the effectiveness of the project, and proposes next steps for system improvement and future research expansion.

CHAPTER 2: THEORETICAL BASIS

2.1 Data Modeling and Schema Design

2.1.1 Data Modeling

A data model is a framework that outlines the information, data, and relationships that form the basis of visual representations, providing a comprehensive overview of an organization's data. The goal of this model is to facilitate the easy extraction of information by organizations, based on the structured data presentation. This enables businesses to have a direct understanding of their data and to devise strategies to enhance their overall revenue. As data volumes grow, the need for data models becomes critical for every business. The advent of management software, equipped with modern technology, aids in data management and storage activities, optimizing effectiveness through smart analytics features and minimizing errors during database software enhancement, thereby driving organizational success.

2.1.2 Data Schema

Database schema design is the process of creating a detailed plan that describes how data is organized, stored, and managed within a database system. It defines the structure of the data, including tables, fields, data types, indexes, primary keys, and foreign keys, as well as the relationships and constraints between entities within the system. This design process, also known as data modeling, often employs visual diagrams to illustrate the database structure. This helps ensure data integrity and supports the data management requirements of the organization. A database schema serves as the foundation for building and maintaining applications, ensuring that data is organized logically and efficiently to meet the system's usage needs.

The Significance of a Data Schema

A database schema plays a crucial role in ensuring the validity of data by managing normalization processes and preventing data duplication. It supports the enforcement of schema constraints, adhering to ACID properties (atomicity, consistency, isolation, and durability), thus protecting data from errors and ensuring its accuracy and reliability. The database schema helps maintain consistency throughout the entire data system by clearly defining constraints and relationships between tables, preventing duplication and conflicts within the data. Additionally, a well-designed schema facilitates the easy addition of new tables to the database and the handling of large amounts of data in growing tables, ensuring scalability as data volumes increase without compromising performance or management. A well-structured database schema enhances performance by ensuring faster data retrieval, reducing the time needed to manipulate tables, and improving the overall system efficiency. This enables applications and users to access and use data more effectively. With a clear and well-documented structure, the database schema simplifies maintenance, allowing for updates or repairs without affecting other parts of the database, thereby minimizing

risks and disruptions. Moreover, the database schema secures sensitive data, granting access only to authorized individuals and protecting critical information from security threats and unauthorized access through specific access rights and data encryption.

2.1.2 Tableau data model

In the modern landscape of data management, Tableau has developed a robust and flexible data model divided into two primary layers: the Logical and Physical layers.

- Logical Layer: This is the initial layer users interact with on the Data Source page canvas. It facilitates the combination of tables through relationships known as "noodles." These relationships enable users to define how tables are connected without direct joins, maintaining clarity and simplicity in the data structure.
- Physical Layer: This layer manages the integration of data between physical tables using joins and unions. Logical tables may contain multiple physical tables, and their combination results in a more detailed data structure reflected at the row level of the merged table.

This model not only enhances data analysis capabilities but also ensures the flexibility needed to meet diverse user analysis requirements.

2.2 Data Preparation by using Tableau Prep Builder

To manage and analyze data effectively, begin by connecting to various data sources such as files, servers, or Tableau extracts. Utilize the capability to merge information from multiple sources by dragging and dropping tables into the flow pane. Within this pane, engage in various operations including filtering, renaming, and joining which facilitate the cleaning and structuring of your data.

As you manipulate data, Tableau Prep visually represents each step in a customizable flow chart, allowing for ongoing review and adjustments. This visual tracking is crucial for ensuring accuracy and efficiency throughout the data preparation process.

Upon completing the data flow, execute it to apply these operations across the entire dataset. Tableau Prep's integration with other Tableau products enhances its utility, enabling users to perform actions such as extracting data, publishing to servers, and scheduling updates without leaving the platform.

Additionally, Tableau Prep temporarily stores frequently used data in a secure .hyper file, and permanently saves certain operations within a Tableau flow (.tfl) file. This ensures both the security and integrity of data management processes, facilitating effective analysis and decision-making within organizations.

2.3 Dashboard Design Principles

Introduction to Dashboard Design

A dashboard is a crucial tool for visualizing data and providing information quickly and clearly to users. Effective dashboard design must adhere to fundamental principles to optimize the user experience and improve work performance.

The Importance of User-Centric Design

User-centric design is a key factor in developing dashboards. It ensures that the dashboard is not only visually appealing but also easy to use, meeting the specific needs of the end users. A well-designed dashboard helps users quickly grasp information and make accurate decisions.

Key Principles in Dashboard Development:

Clear and simple

- Clear: Information on the dashboard needs to be presented clearly and understandably. Using appropriate charts, tables, and graphs helps users easily access and comprehend the data. Avoid using too much information in one chart to prevent clutter. Important information should be placed in the most visible positions.
- Simple: Avoid using too many colors, images, or unnecessary details. Simplicity in design helps minimize distractions and enhances the focus on key information. Limit the use of complex graphic effects and concentrate on conveying information in the most visual and understandable way possible.

Consistency

- Consistency in design: Use the same font style, colors, and formatting for different sections of the dashboard to create a cohesive visual experience. This helps users avoid confusion when moving from one part of the dashboard to another.
- Consistency in data: Ensure that the data is presented in a consistent and easy-to-understand manner. Use the same units of measurement and number formatting to avoid confusing users. For example, if you use millions as a unit in one chart, all related charts should also use this unit.

Actionability

- Provide actionable information: The dashboard should present information that helps users make decisions easily. Key Performance Indicators (KPIs) and critical data should be highlighted. For example, use colors to emphasize metrics that exceed thresholds or important trends.
- Action guidance: Provide specific suggestions or actions based on the current data to assist users in the decision-making process. For example, if a sales metric is declining, the dashboard can suggest measures for improvement.

Interactivity

- User interaction: The dashboard should include interactive elements such as data filtering, time adjustments, or parameter changes so users can customize and search for specific information according to their needs. For example,

allow users to select a specific time period to view sales data for that timeframe.

- Immediate feedback: When users interact with the dashboard, changes or results should be displayed immediately to enhance the user experience. This helps users avoid long waits and quickly find the necessary information.

By applying these principles, you will create effective, intuitive, and user-friendly dashboards that maximally support the decision-making and data analysis process.

2.4 Data Integration and Pipeline Management

The seamless and continuous integration of data is crucial for ensuring the integrity and continuity of data processing. When data is smoothly integrated from various sources, businesses can generate accurate and timely reports, thus supporting strategic business decisions. An efficient data integration process helps minimize errors, enhance data reliability, and improve analytical capabilities. This is particularly important in today's business environment, where data is a valuable asset that determines the success of business strategies.

To ensure a smooth data flow from Tableau Prep to Tableau, several key steps need to be followed. The first step is data preparation in Tableau Prep. This stage includes tasks such as cleaning the data, which involves removing missing or invalid values, and transforming the data to reformat it as needed. Additionally, merging data from different sources is performed to create a unified and complete dataset. These operations not only ensure that the data is accurately prepared but also optimize the subsequent analysis process.

After preparation, data validation and verification are essential steps. At this stage, it is crucial to ensure that the processed data in Tableau Prep meets quality and accuracy standards. Logical checks are also conducted to ensure that the data does not contain logical errors such as inconsistencies or redundancies. This verification process plays a vital role in ensuring that the data imported into Tableau is reliable and suitable for analysis.

Next, the processed data is exported from Tableau Prep. The data can be saved as a file or directly imported into Tableau Server or Tableau Online. Choosing the appropriate file format (such as .hyper or .csv) is also important to optimize performance when importing into Tableau. The file format affects not only processing speed but also the accessibility and usability of the data in Tableau.

When importing data into Tableau, the first step is to connect Tableau to the data source prepared in Tableau Prep. It is necessary to configure connection options and

refresh the data to ensure that the data in Tableau is always up to date. This helps maintain the continuity and accuracy of the data, allowing users to access the latest information for analysis and decision-making.

After the data is imported into Tableau, the process of creating reports and dashboards begins. Using Tableau tools to create visual charts and dashboards helps users easily understand and analyze the data. Customizing report elements to meet business needs and ensure user-friendliness is also important. A visual and comprehensible report enables users to quickly grasp important information and make timely decisions.

Finally, monitoring and maintaining the data integration process is indispensable. It is important to oversee the refresh and integration process to detect and address issues promptly. Regular system checks and maintenance also help ensure that the data integration process operates smoothly and efficiently. Maintaining system stability gives businesses confidence in using data for strategic decisions.

By accurately and fully implementing these steps, businesses can maintain a continuous and seamless data flow from Tableau Prep to Tableau, thereby optimizing the analysis process and data-driven decision-making. Data becomes not just a support tool but a critical factor determining the success of a business in today's competitive environment.

2.5 Presentation and Reporting Techniques

The Importance of Effective Data Presentation

Effectively presenting data is crucial to ensuring that recipients can understand and use the data to make decisions. A well-prepared report or presentation helps convey key findings and creates a positive impact on the organization's decision-making process. When data is presented clearly and understandably, it not only helps viewers quickly grasp the information but also increases the practical applicability of the data.

Key Components of a Comprehensive Report:

Executive Summary

- **Purpose:** Summarize the key points of the report, helping readers quickly grasp important information without reading the entire report.
- **Content:** Includes the report's objectives, key findings, and action recommendations.
- **Details:** The executive summary should be concise, about 1-2 pages, and clearly state the most important points for readers to quickly understand the information. For example, if the report is about business performance, the summary should highlight key metrics such as revenue, profit, and significant changes compared to the previous period.

Introduction

- **Purpose:** Provide context and objectives of the report.
- **Content:** Describe the problem to be solved, the research method, and the scope of the report.
- **Details:** This section should clearly explain why the research was conducted, the specific research questions, and the limitations of the study. For example, if the study is about consumer behavior, the introduction should explain the reason for choosing this topic and the scope of the research, such as the subjects, time frame, and location.

Methodology

- **Purpose:** Explain how data was collected and analyzed.
- **Content:** Describe the tools and techniques used during the research process, data collection methods, and analysis methods.
- **Details:** Include specific steps, software tools used, and analysis techniques so readers can assess the validity and accuracy of the methodology. For example, if using surveys, this section should detail the questionnaire design, sampling method, and data analysis techniques.

Findings

- **Purpose:** Present the results of the data analysis process.
- **Content:** Include charts, graphs, and tables illustrating key findings.
- **Details:** Present results logically and systematically, using visual tools for easy understanding. Each finding should be clearly explained and linked to the research objectives. For example, if it is found that revenue increased in the last quarter, explain the reasons and factors contributing to this increase.

Analysis

- **Purpose:** Explain the significance of the findings.
- **Content:** Provide insights into trends, patterns, and relationships in the data.
- **Details:** Analyze the findings in depth, explaining why these results are important and what they mean for the research problem. State assumptions and limitations in the analysis. For example, if a relationship between product price and customer satisfaction is found, clearly explain this correlation and suggest improvement solutions.

Conclusion

- **Purpose:** Summarize the main points of the report.
- **Content:** Emphasize key findings and their implications for the organization.
- **Details:** The conclusion should be concise, highlighting the most important findings and their significance. Suggest further research directions if necessary. For example, if concluding that the current marketing strategy is effective, propose continuing or expanding this strategy.

Recommendations

- **Purpose:** Provide action proposals based on the findings and analysis.

- **Content:** Suggest specific steps the organization should take to improve the current situation or capitalize on opportunities.
- **Details:** Recommendations should be specific, practical, and feasible. Clearly explain the rationale and basis for each proposal and provide detailed implementation steps. For example, if recommending enhanced employee training, specify the reasons and implementation methods, including training content, duration, and estimated costs.

Appendices

- **Purpose:** Provide supplementary information to support the main sections of the report.
- **Content:** Include detailed data, calculation formulas, and reference materials.
- **Details:** Appendices should be logically organized and easy to reference, helping readers verify and quickly consult the information. For example, if the report uses econometric models, the appendices should include formulas and detailed results tables.

Best Practices for Presenting Dashboards and Results to Stakeholders:

Know Your Audience

- **Understanding:** Know who will view and use your report to tailor the presentation accordingly.
- **Customization:** Adjust content and presentation style to fit the audience's level of understanding and needs.
- **Details:** Understand the audience's roles, expertise, and expectations to present effectively. For example, presenting to executives may differ from presenting to a technical team. Executives may focus on strategic metrics and financial impacts, while the technical team cares about detailed analysis and methodologies.

Use Effective Data Visualization

- **Charts and Graphs:** Choose appropriate types of charts and graphs to illustrate key findings.
- **Simplification:** Avoid using too many details, focusing on conveying information clearly and concisely.
- **Details:** Use charts such as bar charts, line charts, and pie charts to visually represent data. Avoid excessive use of colors and keep the design clean and easy to read. For example, when presenting revenue trends, a line chart will clearly show fluctuations over time.

Storytelling

- **Narrative:** Build a coherent story to guide the audience through your findings and analysis.
- **Interaction:** Use interactive elements to engage attention and make the presentation more lively.

- **Details:** A compelling story helps the audience remember and understand the information presented. Use real-world examples, case studies, and illustrative images to increase persuasiveness. For example, when presenting sales performance improvement, use the story of a specific salesperson who applied new techniques and achieved success.

Highlight Key Points

- **Emphasis:** Use colors, font sizes, and positioning to highlight important information.
- **Summary:** Provide brief summaries for each section to help the audience quickly grasp key points.
- **Details:** Use headings, bullet points, and graphic elements to highlight important points. Ensure the most important information is always in a visible position. For example, use red to highlight underperforming metrics or potential risks.

Practice Before Presenting

- **Rehearsal:** Practice presenting in advance to gain confidence and better prepare.
- **Feedback:** Gather feedback from others to improve your presentation before the formal one.
- **Details:** Practice in front of a mirror or with a small group to test your presentation and receive feedback. This helps you gain confidence and identify issues needing correction before the official presentation. For example, record or videotape the rehearsal to self-assess and improve.

By following these principles, you can create effective reports and presentations that convey important information and support the organization's decision-making process. A well-prepared report or presentation not only helps viewers understand the issues but also builds trust and credibility for the presenter.

2.6 Theoretical Concepts in the Logistics Domain

2.6.1 Introduction to Logistics

Logistics is the process of planning, implementing, and controlling the efficient, effective forward, and reverse flow and storage of goods, services, and related information between the point of origin and the point of consumption in order to meet customer requirements. It encompasses a broad spectrum of activities including transportation, warehousing, inventory management, order processing, packaging, and customer service.

Logistics plays a pivotal role in supply chain management by ensuring the smooth flow of goods and services, from raw material procurement to final delivery to the end customer. Efficient logistics operations contribute to reduced costs, improved customer satisfaction, and enhanced overall supply chain performance.

2.6.2 Key Metrics in Logistics

Overview of Critical Logistics Metrics: lead time, on-time delivery rate, shipping cost, inventory turnover, and order accuracy.

- Lead Time: The time elapsed between the placement of an order and the receipt of the goods.
- On-Time Delivery Rate: The percentage of orders delivered on or before the promised delivery date.
- Shipping Cost: The total cost incurred for transporting goods from the origin to the destination.
- Inventory Turnover: The number of times inventory is sold and replaced within a specific period.
- Order Accuracy: The percentage of orders filled correctly, without errors in quantity or product.

Logistics metrics provide valuable insights into the efficiency and effectiveness of logistics operations. By monitoring and analyzing these metrics, organizations can identify areas for improvement, make data-driven decisions, and enhance overall supply chain performance.

2.6.3 The Role of Data Analytics in Logistics

Data analytics plays a transformative role in logistics by enabling managers to derive valuable insights from vast amounts of data. These insights facilitate informed decision-making, streamline operations, cut costs, and enhance customer service. By leveraging data analytics, logistics managers can identify patterns, predict trends, and make proactive adjustments to their logistics processes. The result is a more agile, efficient, and customer-centric logistics operation.

- Route Optimization: Data analytics helps in determining the most efficient routes for transportation. By analyzing traffic patterns, weather conditions, and historical delivery data, logistics managers can identify the quickest and most cost-effective paths. This not only reduces fuel consumption but also shortens delivery times, leading to significant cost savings and higher customer satisfaction.
- Demand Forecasting: Accurate demand forecasting is crucial for optimizing inventory levels and production planning. Data analytics tools analyze past sales data, market trends, and seasonal patterns to predict future customer demand. With precise demand forecasts, companies can maintain optimal inventory levels, reduce stockouts, and minimize excess inventory, resulting in lower holding costs and improved cash flow.
- Inventory Management: Effective inventory management is essential for a smooth supply chain. Data analytics can identify slow-moving items, predict stock requirements, and determine the optimal reorder points. By analyzing

sales trends, lead times, and supplier performance, logistics managers can optimize stock levels, reduce carrying costs, and ensure timely replenishment of inventory. This leads to a more efficient use of warehouse space and a reduction in inventory holding costs.

- **Shipment Tracking:** Real-time shipment tracking is a critical component of modern logistics. Data analytics enables the monitoring of goods in transit, providing visibility into the location and status of shipments at any given time. This transparency helps in reducing transit times, identifying potential delays, and taking corrective actions promptly. Enhanced visibility also improves customer satisfaction by providing accurate delivery information and allowing customers to track their orders in real-time.

The integration of data analytics into logistics processes provides a competitive edge by enabling organizations to operate more efficiently and responsively. By harnessing the power of data, logistics managers can:

- **Enhance Decision-Making:** Make informed decisions based on real-time data and predictive insights.
- **Improve Operational Efficiency:** Streamline operations, eliminate bottlenecks, and optimize resource utilization.
- **Reduce Costs:** Identify cost-saving opportunities, optimize inventory levels, and reduce waste.
- **Elevate Customer Service:** Provide timely and accurate information, improve delivery reliability, and enhance the overall customer experience.

In conclusion, data analytics is a vital tool for modern logistics management. It offers a multitude of benefits, from operational improvements to cost reductions and enhanced customer satisfaction. By effectively applying data analytics, organizations can achieve a more responsive, efficient, and customer-centric logistics operation.

CHAPTER 3: STUDY FRAMEWORK

3.1 Overview of the Dataset

Summary

The DataCo Supply Chain dataset is a comprehensive dataset containing detailed information about orders, products, customers, and the delivery process of a company. With a total of 180,519 rows of data and various columns of information, this dataset provides a rich foundation for analyzing and optimizing different aspects of the supply chain.

Purpose of Use

This dataset is used to:

- Analyze supply chain performance.
- Optimize shipping and delivery processes.
- Assess customer satisfaction.
- Forecast sales and profits.
- Identify risks and propose mitigation measures.

Scope of the Data

The dataset includes the following detailed information:

Transaction and Shipping Information:

- Type: Transaction type (DEBIT, TRANSFER, CASH, PAYMENT).
- Days for shipping (real): Actual number of days to ship goods.
- Days for shipment (scheduled): Scheduled number of days to ship goods.
- Benefit per order: Profit from each order.
- Sales per customer: Sales from each customer.
- Delivery Status: Delivery status (Advance shipping, Late delivery, Shipping on time).
- Late_delivery_risk: Risk of late delivery (0 or 1).

Product and Category Information:

- Category Id: Product category ID.
- Category Name: Product category name.
- Product Card Id: Product ID.
- Product Category Id: Product category ID.
- Product Description: Product description.
- Product Image: Product image.
- Product Name: Product name.
- Product Price: Product price.
- Product Status: Product status.

Customer Information:

- Customer City: Customer's city.
- Customer Country: Customer's country.

- Customer Email: Customer's email.
- Customer Fname: Customer's first name.
- Customer Id: Customer ID.
- Customer Lname: Customer's last name.
- Customer Password: Customer's password.
- Customer Segment: Customer segment.
- Customer State: Customer's state.
- Customer Street: Customer's street address.
- Customer Zipcode: Customer's postal code.

Order Information:

- Order City: Order city.
- Order Country: Order country.
- Order Customer Id: Order customer ID.
- Order date (DateOrders): Order date.
- Order Id: Order ID.
- Order Item Cardprod Id: Product ID in the order.
- Order Item Discount: Discount for each product in the order.
- Order Item Discount Rate: Discount rate for each product in the order.
- Order Item Id: Product ID in the order.
- Order Item Product Price: Product price in the order.
- Order Item Profit Ratio: Profit ratio for each product in the order.
- Order Item Quantity: Quantity of products in the order.
- Sales: Sales.
- Order Item Total: Total amount for each product in the order.
- Order Profit Per Order: Profit per order.
- Order Region: Order region.
- Order State: Order state.
- Order Status: Order status.
- Order Zipcode: Order postal code.

Shipping Information:

- Shipping date (DateOrders): Shipping date.
- Shipping Mode: Shipping mode (Standard Class, Second Class, First Class, Same Day).

Department Information:

- Department Id: Department ID.
- Department Name: Department name.

Market and Geographic Information:

- Latitude: Customer's geographic latitude.
- Longitude: Customer's geographic longitude.
- Market: Market.

3.2 Workflow

In this section, we describe the overall workflow of our study. First, we embarked on a search for datasets that met our group's criteria, ensuring that they contained sufficient features to support the creation of five dashboards using Tableau. After a thorough search, we identified the DataCo Supply Chain dataset as the most suitable for our needs. We then proceeded to read and understand the dataset comprehensively. This involved examining the condition of each data column, categorizing them into sub-tables, and connecting them to form a cohesive database schema or model. This schema was visualized to facilitate a clearer understanding and to outline the steps necessary for data cleaning.

With our preparations complete, we launched Tableau Prep, imported the dataset, and retained the necessary columns for each table. We then meticulously cleaned each table within Tableau Prep. The specifics of this cleaning process will be detailed in section 3.3, ETL Process. Before proceeding to create the dashboards, we conducted research to identify the key metrics commonly used in this field and determined which types of charts would be most appropriate for visualizing these metrics. This step ensured that our dashboards would effectively communicate the most relevant information.

After creating a complete pipeline and modeling the data, we transferred the output data into Tableau. This allowed us to proceed with creating charts for the five dashboards we had initially planned. The details and specifics of these dashboards will be elaborated upon in Chapter 4. This comprehensive workflow ensured that our data was thoroughly prepared and accurately represented in our visualizations, facilitating effective analysis and decision-making.

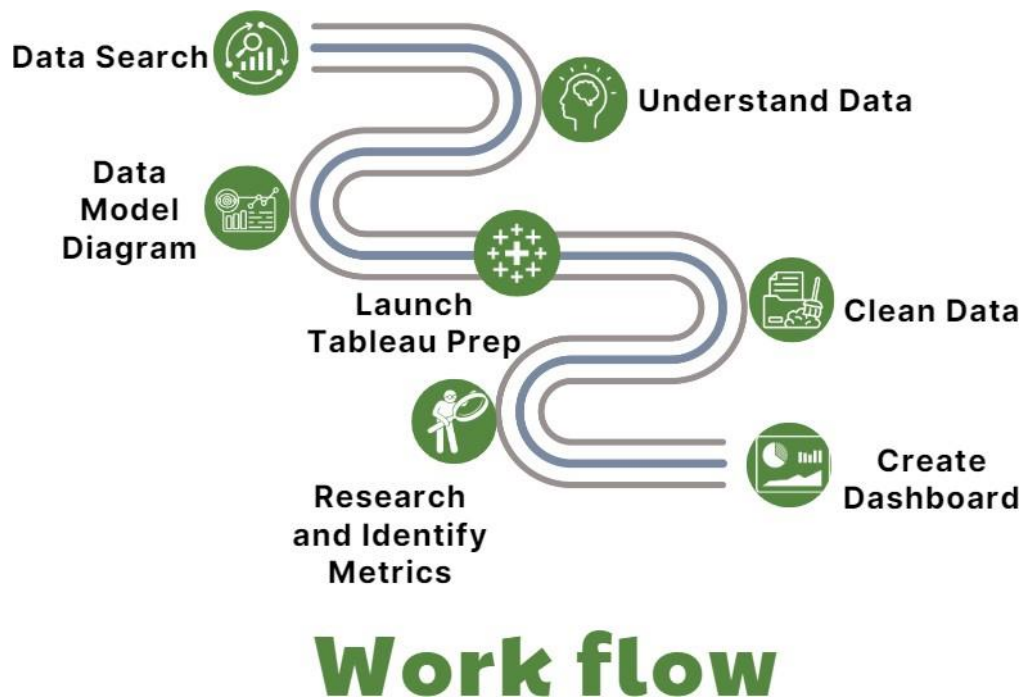


Figure 3.1 Data Analysis Workflow

3.2 Requirements analysis

Identifying Necessary Requirements for the Study

To ensure that our study meets the set objectives, we identified several necessary requirements, including:

- **Data Accuracy and Integrity:** The data must be complete and accurate, free of missing or invalid values. This involves removing duplicate entries, handling missing values, and verifying the logical consistency of the data.
- **Data Diversity:** The data should encompass various aspects of the supply chain to create comprehensive and diverse dashboards. This includes information on transactions, shipping, products, customers, and financial elements.
- **Data Connectivity:** The data tables must be logically linked to form a coherent database schema that supports analysis. This schema facilitates a clear understanding of the relationships among different data points.
- **Data Cleaning and Transformation Capability:** The data must be cleaned and transformed to meet the requirements of Tableau dashboards. This includes standardizing data, creating calculated fields, and converting data formats as necessary.

Objectives to be Achieved

Our study aims to achieve several specific objectives:

- **Creation of 5 Effective Dashboards:** Use Tableau to create five dashboards that effectively visualize key aspects of the supply chain, enabling users to easily understand and analyze the data.
- **Identification of Key Metrics:** Identify and focus on the most critical metrics in the supply chain domain, such as delivery times, late delivery risk, profit margins, sales, and customer satisfaction.
- **Improvement of Operational Efficiency:** Through data analysis, propose improvements to enhance the operational efficiency of the supply chain, including reducing delivery times and optimizing profits.
- **Providing Decision-Support Information:** Equip managers and stakeholders with the necessary information to make accurate and timely business decisions.

Issues and Challenges to Address

Throughout our study, we identified several issues and challenges that need to be addressed:

- **Inconsistent Data Quality:** Data from various sources may vary in quality, necessitating processing and standardization to ensure consistency.
- **Large Volume of Data:** Handling and analyzing a large volume of data requires robust computational power and optimized methods to ensure performance.
- **Diverse Data Sources:** Data may come from multiple sources, requiring integration and synchronization to create a cohesive database.
- **Missing Data:** Some data values may be missing or unavailable, necessitating methods to handle these gaps.
- **Selecting Appropriate Metrics and Charts:** Choosing the most suitable metrics and charts to effectively visualize the data, ensuring that the information is conveyed clearly and understandably.

3.3 ETL Process



Figure 3.2 Data Processing Workflow Diagram

Step 1: Extract data from necessary columns

In the initial steps, our group will extract all necessary columns from the tables mentioned in section 3.4 of the document. These tables and columns are determined based on the analysis requirements and business objectives.

The screenshot shows a data transformation tool interface. At the top, a workflow diagram illustrates the process: 'DimShippingDate' is processed through 'Duplicated', 'Clean 1', 'Join 7', and 'Join 11'. Below the diagram, a table titled 'DimShippingDate' displays 4 fields and 181 rows. The table columns are 'shipping date (D...', 'YEAR', 'Month', and 'Day'. The data rows show dates from 02/03/2018 to 01/18/2018, with corresponding year, month, and day values.

shipping date (D...	YEAR	Month	Day
02/03/2018, 10:56:00 PM	2.018	2	3
01/18/2018, 12:27:00 PM	2.018	1	18
01/17/2018, 12:06:00 PM	2.018	1	17
01/16/2018, 11:45:00 AM	2.018	1	16
01/15/2018, 11:24:00 AM	2.018	1	15
01/19/2018, 11:03:00 AM	2.018	1	19
01/16/2018, 10:42:00 AM	2.018	1	16
01/16/2018, 10:21:00 AM	2.018	1	16
01/16/2018, 10:00:00 AM	2.018	1	16
01/15/2018, 09:39:00 AM	2.018	1	15
01/19/2018, 09:18:00 AM	2.018	1	19
01/18/2018, 08:57:00 AM	2.018	1	18
01/17/2018, 08:36:00 AM	2.018	1	17
01/16/2018, 08:15:00 AM	2.018	1	16
01/16/2018, 07:54:00 AM	2.018	1	16
01/16/2018, 07:33:00 AM	2.018	1	16
01/18/2018, 07:12:00 AM	2.018	1	18
01/16/2018, 06:51:00 AM	2.018	1	16
01/16/2018, 06:30:00 AM	2.018	1	16
01/13/2018, 06:09:00 PM	2.018	1	13
01/13/2018, 05:48:00 PM	2.018	1	13
01/18/2018, 05:27:00 AM	2.018	1	18
01/17/2018, 05:06:00 AM	2.018	1	17
01/16/2018, 04:45:00 AM	2.018	1	16
01/16/2018, 04:24:00 AM	2.018	1	16
01/19/2018, 04:03:00 AM	2.018	1	19
01/18/2018, 03:42:00 AM	2.018	1	18

Figure 3.3 Extract data from necessary columns

Step 2: Duplicate the data tables

After extracting data from the necessary columns, the team will proceed to duplicate each table. This helps protect the original data and facilitates the subsequent steps without affecting the initial data.

The screenshot shows the Tableau Prep interface. At the top, a workflow diagram includes steps: 'Duplicated', 'Clean 1', 'Join 7', and 'Join 11'. Below the diagram, a data table is displayed with the following columns: 'shipping date (D...', 'YEAR', 'Month', and 'Day'. The table contains 20 rows of unique data, each starting with 'Unique' in the first column.

	shipping date (D...	YEAR	Month	Day
Unique	05/01/2015 07:32:00 AM	2,015	5	1
Unique	12/01/2017 05:20:00 AM	2,017	12	1
Unique	04/02/2017 12:36:00 AM	2,017	4	2
Unique	07/02/2015 06:50:00 PM	2,015	7	2
Unique	07/02/2015 09:49:00 PM	2,015	7	2
Unique	09/02/2015 01:10:00 PM	2,016	9	2
Unique	05/03/2016 08:04:00 PM	2,016	5	3
Unique	01/04/2017 03:26:00 PM	2,017	1	4
Unique	01/05/2016 01:40:00 PM	2,016	1	5
Unique	04/06/2016 12:03:00 AM	2,016	4	6
Unique	05/06/2015 03:25:00 PM	2,015	5	6
Unique	05/07/2015 04:54:00 AM	2,015	5	7
Unique	11/07/2016 05:21:00 AM	2,016	11	7
Unique	12/07/2016 01:12:00 PM	2,016	12	7
Unique	01/08/2016 04:44:00 AM	2,016	1	8
Unique	01/10/2018 02:02:00 PM	2,018	1	10
Unique	06/10/2016 12:00:00 AM	2,016	6	10
Unique	01/12/2015 09:37:00 AM	2,015	1	12
Unique	02/13/2016 07:08:00 AM	2,016	2	13
Unique	06/13/2015 07:33:00 PM	2,015	6	13
Unique	12/14/2017 03:33:00 AM	2,017	12	14
Unique	01/15/2018 01:57:00 AM	2,018	1	15
Unique	06/15/2017 05:24:00 AM	2,017	6	15
Unique	10/16/2015 12:23:00 PM	2,015	10	15
Unique	12/16/2016 09:17:00 PM	2,016	12	15
Unique	12/16/2017 07:04:00 AM	2,017	12	15
Unique	01/16/2015 06:58:00 AM	2,015	1	16

Figure 3.4 Duplicate the data tables

Step 3: Data Preprocessing

Next, the team will use the Data Clean function in Tableau Prep for data preprocessing. The data preprocessing process includes the following steps:

- + Check and remove null and duplicate values: Columns containing null or duplicate values will be identified and processed to ensure data consistency and accuracy.
- + Remove unnecessary columns: As mentioned in section 1.2.4 of the BI8 document, columns such as Benefit per order, Sales per customer, Product Category Id, Order Customer Id, Order Item Cardprod Id, and Order Item Product Price will be removed. These columns are identified as unnecessary or redundant and will be eliminated to clean the data.
- + Transform data formats: Convert columns with inappropriate formats to the required formats, such as converting date columns to the DateTime format.
- + Correct and standardize data: Fix data errors such as incorrect labels, inconsistent formatting, and standardize data to ensure consistency and ease in later analysis.
- + Specific example: If the 'Category Id' column contains incorrect values, such as Category Id 37 labeled as 'electronics' but is actually 'golf', we will correct this value to the proper label 'golf'.

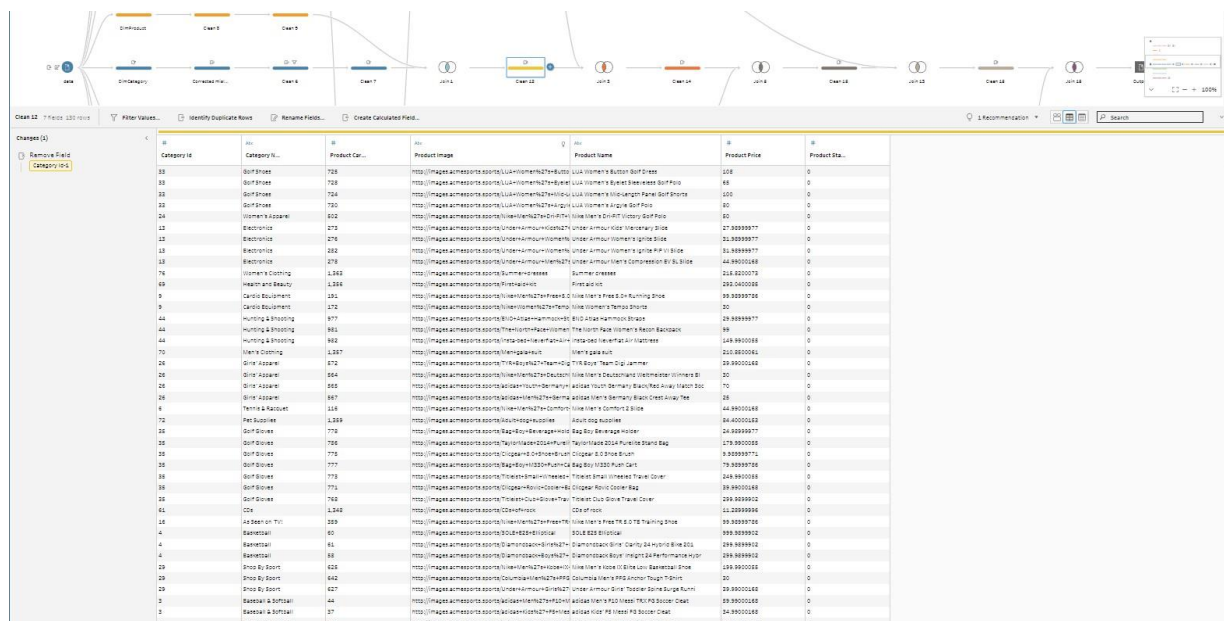


Figure 3.5 Data Preprocessing

Step 4: Join the Data

After data preprocessing, the team will proceed to join the data tables together to form a complete dataset for subsequent analysis steps. The data joining process is conducted as follows:

- + Identify primary and foreign keys: Before performing the join, we need to clearly identify the primary and foreign keys of the tables. For example, in the relationship between the 'Customer' and 'Order Process' tables, 'Customer Id' will be the primary key in the 'Customer' table and the foreign key in the 'Order Process' table.
- + Perform the join: Using tools in Tableau Prep, we will execute joins such as inner join, left join, or right join, depending on the requirements of the analysis. The goal is to merge data from different tables based on the identified keys.
- + Check the join results: After joining, it is crucial to check the results to ensure that the data has been correctly combined and there are no errors. This check involves verifying the number of records, the accuracy of the joined columns, and ensuring there are no unwanted lost or duplicate records.
- + Specific example: If we have two tables, 'Orders' and 'Customers', joining based on 'Customer Id' will allow us to obtain customer information along with their orders. This facilitates the analysis of purchasing behaviors and customer characteristics.

- + Perform the join: Using tools in Tableau Prep, we will execute joins such as inner join, left join, or right join, depending on the requirements of the analysis. The goal is to merge data from different tables based on the identified keys.

- + Check the join results: After joining, it is crucial to check the results to ensure that the data has been correctly combined and there are no errors. This check involves verifying the number of records, the accuracy of the joined columns, and ensuring there are no unwanted lost or duplicate records.

- + Specific example: If we have two tables, 'Orders' and 'Customers', joining based on 'Customer Id' will allow us to obtain customer information along with their orders. This facilitates the analysis of purchasing behaviors and customer characteristics.

Order ID	Date	Type	Days for ship	Days for shipment	Delivery Sta.	Late	Category ID	Category No.	Customer ID	Department
2	2	TRAILER	4	4	Shipping on time	0	40	Flaming	11,000	T
4	4	TRAILER	4	4	Shipping on time	0	40	Flaming	2,400	T
8	8	TRAILER	4	4	Shipping on time	0	40	Flaming	4,200	T
12	12	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	8,841	T
16	16	TRAILER	4	4	Shipping on time	0	40	Flaming	2,000	T
20	20	TRAILER	4	4	Shipping on time	0	40	Water Sports	7,001	T
24	24	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	2,969	T
28	28	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	5,009	T
32	32	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	11,862	T
36	36	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	7,087	T
40	40	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	8,868	T
44	44	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	7,001	T
48	48	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	8,868	T
52	52	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	8,870	T
56	56	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,247	T
60	60	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	6,770	T
64	64	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,247	T
68	68	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	10,000	T
72	72	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	9,792	T
76	76	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	2,077	T
80	80	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	4,961	T
84	84	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	8,867	T
88	88	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,247	T
92	92	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	8,862	T
96	96	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	11,000	T
100	100	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	9,004	T
104	104	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	11,000	T
108	108	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,877	T
112	112	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,888	T
116	116	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	6,000	T
120	120	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	12,000	T
124	124	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	10,000	T
128	128	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	10,071	T
132	132	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
136	136	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
140	140	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
144	144	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
148	148	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
152	152	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
156	156	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
160	160	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
164	164	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
168	168	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
172	172	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
176	176	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
180	180	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
184	184	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
188	188	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
192	192	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
196	196	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T
200	200	TRAILER	4	4	Shipping on time	0	40	Camping & Hiking	1,000	T

Figure 3.6 Join the data

3.4 Data Modeling

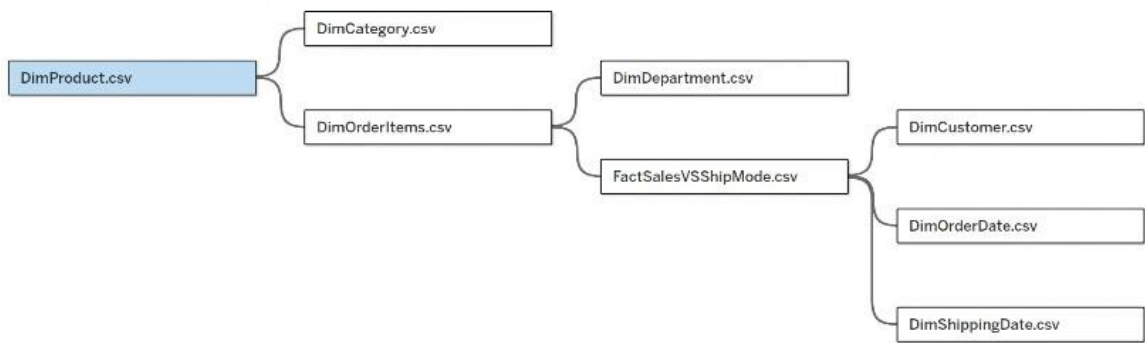


Figure 3.7 Data Modeling Diagram for CSV Files

3.4.1. Order Items Table

Table 3.1 Order Item Details

Column Name	Data Type	Description
Order Id	INT	Unique identifier for each order.
Order Item Cardprod Id	INT	Unique identifier for each product item in the order.
Order Item Discount	FLOAT	The discount amount applied to each product item in the order.
Order Item Discount Rate	FLOAT	The discount rate applied to each product item in the order.
Order Item Id	INT	Unique identifier for each product item in the order.
Order Item Product Price	FLOAT	The price of each product item in the order.
Order Item Profit Ratio	FLOAT	The profit ratio for each product item in the order.
Order Item Quantity	INT	The quantity of each product item in the order.
Order Item Total	FLOAT	The total amount for each product item in the order.
Order Profit Per Order	FLOAT	The profit amount per order.
Product Card Id	INT	Unique identifier for the product.
Sales	FLOAT	Total sales amount for each product item.

Purpose: The `OrderItems` table provides detailed information about each item in an order, including quantity, price, discount, and total amount. It is essential for analyzing the specifics of sales transactions and understanding individual product performance within orders.

Linked Tables:

- `Orders`: Links via `Order Id` to provide overall order context.
- `Products`: Links via `Product Card Id` to provide product details.
- `Departments`: Indirectly linked through products to categorize sales by department.
- `DimOrderDate`: Links via `Order Date` to analyze orders over time.
- `DimShippingDate`: Links via `Shipping Date` to analyze shipping performance.

3.4.2 Products Table**Table 3.2 Product Details**

Column Name	Data Type	Description
Category Id	INT	Unique identifier for the product category.
Category Id	INT	Unique identifier for the product.
Product Image	VARCHAR	Image of the product.
Product Name	VARCHAR	Name of the product.
Product Price	FLOAT	Price of the product.
Product Status	VARCHAR	Status of the product (e.g., available, discontinued).

Purpose: The Products table provides product information, including name, price, and status. This table is crucial for linking sales data with specific products to analyze product performance and availability.

Linked Tables:

- `OrderItems`: Links via `Product Card Id` to provide detailed product-level sales data.
- `Departments`: Links via `Category Id` to categorize products by department.

3.4.3 Orders Table**Table 3.3 Order Details**

Column Name	Data Type	Description
Order Id	INT	Unique identifier for each order.
Order Date	DATE	Date when the order was placed.
Customer Id	INT	Unique identifier for the customer who placed the order.
Order Status	VARCHAR	Status of the order (e.g., pending, completed).
Total Sales	FLOAT	Total sales amount for each order.

Purpose: The Orders table provides an overview of orders, including order dates, customer information, and total sales. It is key for tracking overall sales performance and customer order history.

Linked Tables:

- `OrderItems`: Links via `Order Id` to provide item-level details for each order.
- `Customers`: Links via `Customer Id` to provide customer-related information.
- `DimOrderDate`: Links via `Order Date` to analyze orders over time.
- `DimShippingDate`: Links via `Shipping Date` to analyze shipping performance.

3.4.4 Departments Table

Table 3.4 Departments Details

Column Name	Data Type	Description
Department Id	INT	Unique identifier for each department.
Department Name	VARCHAR	Name of the department.

Purpose: The Departments table provides information about different departments, categorizing and analyzing sales by department.

Linked Tables:

- `Products`: Links via `Category Id` to categorize products by department.
- `OrderItems`: Indirectly linked through products to provide departmental sales data.

3.4.5 Customer Table**Table 3.5 Customer Details**

Column Name	Data Type	Description
Customer City	VARCHAR	City where the customer is located.
Customer Country	VARCHAR	Country where the customer is located.
Customer Email	VARCHAR	Email address of the customer.
Customer Fname	VARCHAR	First name of the customer.
Customer Id	INT	Unique identifier for each customer.
Customer Lname	VARCHAR	Last name of the customer.
Customer Password	VARCHAR	Password for the customer's account.
Customer Segment	VARCHAR	Segment to which the customer belongs (e.g., regular, VIP).
Customer State	VARCHAR	State where the customer is located.
Customer Street	VARCHAR	Street address of the customer.
Customer Zipcode	VARCHAR	Postal code of the customer.

Purpose: The Customers table provides detailed customer information, including contact details, segments, and addresses. This table is crucial for customer segmentation, understanding buying behavior, and tracking customer interactions.

Linked Tables:

- `Orders`: Links via `Customer Id` to provide order-related information and analyze customer order history.

3.4.6 OrderDate Table

Table 3.6 OrderDate Details

Column Name	Data Type	Description
Order Date	DATETIME	Date and time when the order was placed.
Year	INT	Year of the order date.
Month	INT	Month of the order date.
Day	INT	Day of the order date.

Purpose: The DimOrderDate table provides detailed date information for orders, including the year, month, and day. It is used to analyze orders based on different time dimensions.

Linked Tables:

- `Orders`: Links via `Order Date` to provide date-related analysis of orders.
- `OrderItems`: Indirectly linked through orders to provide date-related analysis at the item level.

3.4.7 ShippingDate Table

Table 3.7 ShippingDate Details

Column Name	Data Type	Description
Shipping Date	DATETIME	Date and time when the shipping occurred.
Year	INT	Year of the shipping date.
Month	INT	Month of the shipping date.
Day	INT	Day of the shipping date.

Purpose: The DimShippingDate table provides detailed date information for shipping, including the year, month, and day. It is used to analyze shipping performance based on different time dimensions.

Linked Tables:

- `Orders`: Links via `Shipping Date` to provide date-related analysis of shipping.
- `OrderItems`: Indirectly linked through orders to provide date-related analysis at the item level.

3.4.8 OrderProcess Table

Table 3.8 OrderProcess Details

Column Name	Data Type	Description
Order Id	INT	Unique identifier for each order.
Shipping Cost	FLOAT	The cost associated with shipping the order.
Days for shipping (real)	INT	The actual number of days taken to ship the goods.
Days for shipment (scheduled)	INT	The scheduled number of days to ship the goods.
Delivery Status	VARCHAR	The status of the delivery (e.g., on-time, late).
Late_delivery_risk	INT	The risk of late delivery (0 or 1).

Purpose: The `Order Process` table provides detailed information about the shipping and delivery process, including shipping costs, actual and scheduled shipping days, delivery status, and the risk of late delivery. This table is crucial for monitoring and optimizing logistics performance.

Linked Tables:

- `Order`: Links via `Order Id` to provide overall context for the order's shipping performance.
- `DimShippingDate`: Links via `Shipping Date` to analyze shipping performance over time.

CHAPTER 4: RESULTS – DATA ANALYTICS AND VISUALIZATION

4.1. Report and dashboard systems (Report tree)

4.1.1 Overview

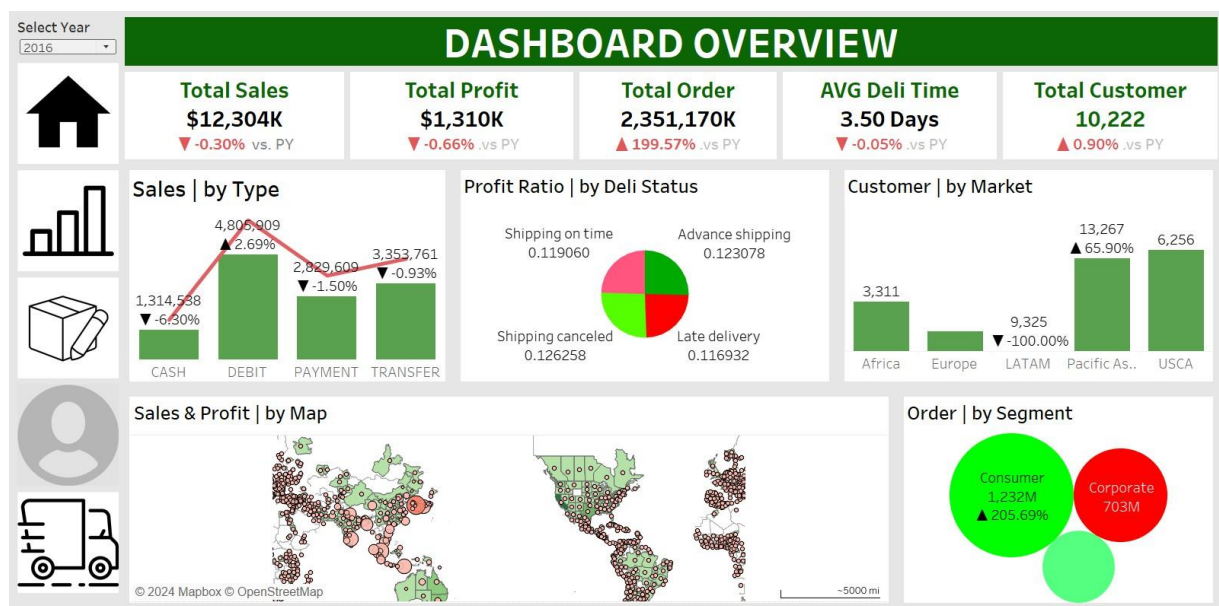


Figure 4.1 Dashboard Overview

This dashboard provides an overview of key performance metrics for the year 2016. Each chart is designed to offer insights into different aspects of sales and customer behavior, aiding in strategic decision-making and operational improvements.

Total Sales, Total Profit, Total Order, AVG Deli Time, Total Customer: These metrics give a quick snapshot of overall business performance. They are crucial for understanding the company's financial health and operational efficiency, and for comparing current performance with the previous year to identify trends and areas for improvement.

Sales by Type: This bar and line chart shows the breakdown of sales by payment type (Cash, Debit, Payment, Transfer). It helps in understanding customer payment preferences and can guide decisions on promoting certain payment methods or addressing issues with less popular ones.

Profit Ratio by Delivery Status: The pie chart illustrates the profit ratio based on different delivery statuses (Shipping on time, Advance shipping, Late delivery, Shipping canceled). This chart helps in identifying the impact of delivery performance on profitability, emphasizing the importance of timely and efficient delivery processes.

Customer by Market: The bar chart shows the number of customers from different market regions (Africa, Europe, LATAM, Pacific Asia, USCA). It highlights the geographical distribution of customers, aiding in regional market analysis and the development of targeted marketing strategies to boost customer acquisition and retention in specific areas.

Sales & Profit by Map: The maps display the geographical distribution of sales and profit. These visualizations help identify high and low performing areas, enabling the company to allocate resources more effectively and focus on regions with greater potential for growth.

Order by Segment: The pie chart shows the distribution of orders by customer segments (Consumer, Corporate). It is useful for understanding which segments contribute the most to orders, guiding targeted marketing efforts and product offerings to maximize engagement and sales in each segment.

Each chart in this dashboard provides valuable insights into different facets of the business, from financial performance and customer distribution to payment preferences and delivery efficiency. By leveraging these insights, managers can make informed decisions to optimize operations, enhance customer satisfaction, and drive business growth. The ability to filter data by year (2015-2018) further aids in trend analysis and strategic planning over time.

4.1.2 Sales

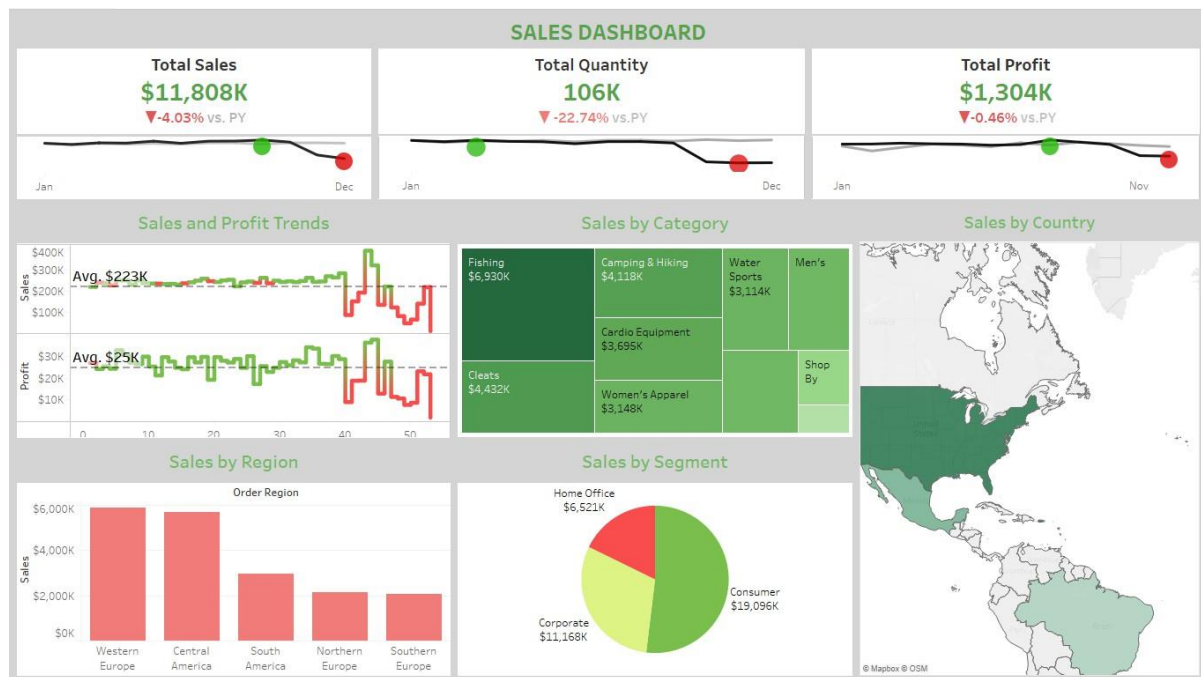


Figure 4.2 Sales Dashboard Overview

This sales dashboard provides a comprehensive view of sales performance from 2015 to 2018. Each chart plays a critical role in supporting analysis and decision-making for better business outcomes.

Total Sales, Total Quantity, Total Profit: These metrics provide a quick overview of the total revenue, total units sold, and total profit for the selected year. They help users quickly assess the overall performance of the company for that year and compare it to the previous year to identify trends in growth or decline.

Sales and Profit Trends: The line charts display monthly trends in revenue and profit throughout the year. These charts are essential for identifying months with high or low performance, recognizing seasonal trends, and evaluating the impact of marketing campaigns or promotions.

Sales by Category: The tree map shows revenue distribution across different product categories. This visualization helps in identifying top-performing categories, allowing businesses to focus resources on these areas or adjust strategies for underperforming categories.

Sales by Country: The map chart illustrates revenue by country. This helps identify strong and weak markets geographically, guiding strategic decisions for market expansion or optimization of sales strategies in specific countries.

Sales by Region: The bar chart displays revenue across major regions (continents). It provides a broad overview of revenue contribution from different parts of the world, aiding in evaluating regional performance and directing strategic development efforts accordingly

Sales by Segment: The pie chart shows revenue by customer segments (Consumer, Corporate, Home Office). This chart is useful for identifying which customer segments generate the most revenue, helping optimize marketing and sales strategies for each segment.

Filter Select Year (2015-2018): This filter allows users to select the year for which they want to view specific data. It is crucial for comparing sales performance across different years, identifying long-term trends, and adjusting business strategies to fit each period.

The charts in this dashboard provide not only a comprehensive overview of sales performance but also detailed insights that support strategic decision-making. By leveraging these charts, managers can make data-driven decisions to optimize business efficiency and guide long-term development.

4.1.3 Logistic

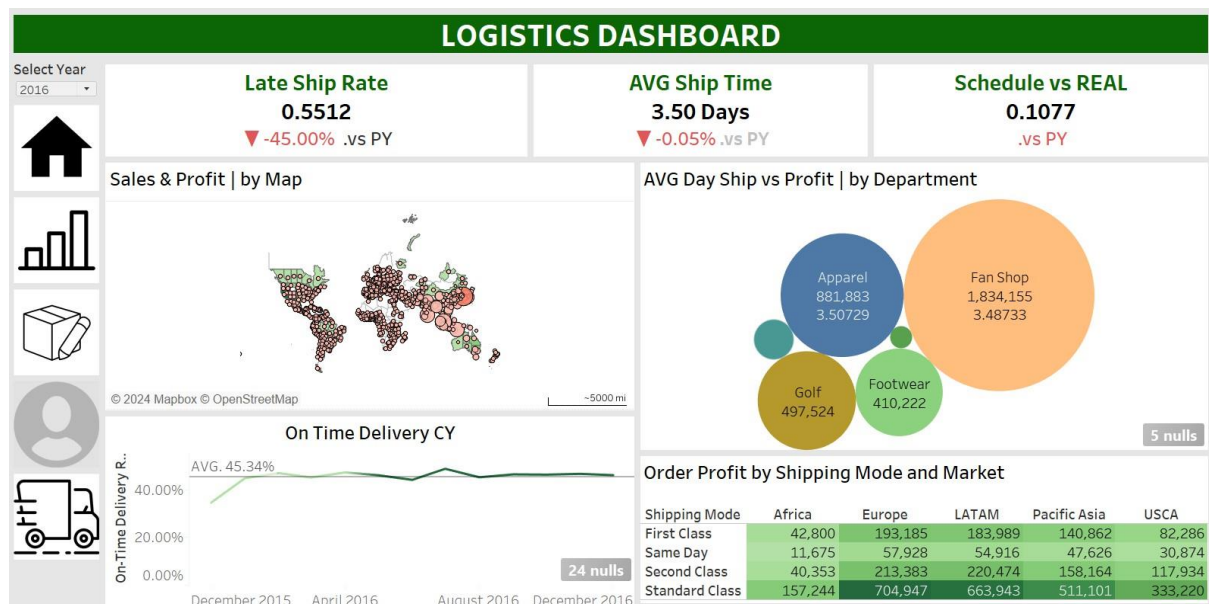


Figure 4.3 Logistics Dashboard

The Logistics Dashboard for the year 2016 provides detailed insights into shipping performance, helping to optimize logistics operations. Each chart is designed to highlight specific aspects of the logistics process, from shipping rates to delivery times and departmental performance.

Late Ship Rate, AVG Ship Time, Schedule vs REAL: These key metrics provide a snapshot of logistics efficiency:

- **Late Ship Rate:** Indicates the proportion of shipments that were late, highlighting areas where the delivery process needs improvement.
- **AVG Ship Time:** Shows the average time taken for shipments, helping to assess and optimize delivery speed.
- **Schedule vs REAL:** Compares scheduled shipping times against actual times, identifying discrepancies and potential delays in the logistics process.

Sales & Profit by Map: The geographical map displays sales and profit distribution across different regions. This helps identify high and low performing areas in terms of logistics, guiding efforts to improve delivery efficiency and customer satisfaction in specific regions.

On Time Delivery CY: The line chart tracks on-time delivery performance over time. Monitoring this trend helps assess improvements or declines in delivery punctuality, enabling timely interventions to maintain or enhance performance.

AVG Day Ship vs Profit | by Department: The bubble chart shows the relationship between average shipping days and profit for different departments (Apparel, Fan Shop, Golf, Footwear). This chart helps identify which departments balance quick shipping times with high profitability, guiding resource allocation and process improvements in lagging departments.

Order Profit by Shipping Mode and Market: The table provides a detailed breakdown of profit by shipping mode (Delivery Mode) and market region. This data is critical for understanding the cost-effectiveness of different shipping methods across various markets, enabling strategic decisions to optimize shipping modes based on profitability and regional performance.

Each chart in this logistics dashboard serves a specific purpose in analyzing and improving the logistics operations of the company. From assessing shipping efficiency and punctuality to understanding regional performance and departmental profitability, these visualizations provide actionable insights. By leveraging these insights, logistics managers can make data-driven decisions to enhance delivery performance, reduce late shipments, and optimize shipping strategies for better overall efficiency and customer satisfaction. The year filter (2015-2018) allows for trend analysis over time, further aiding in long-term strategic planning.

4.1.4 Customer

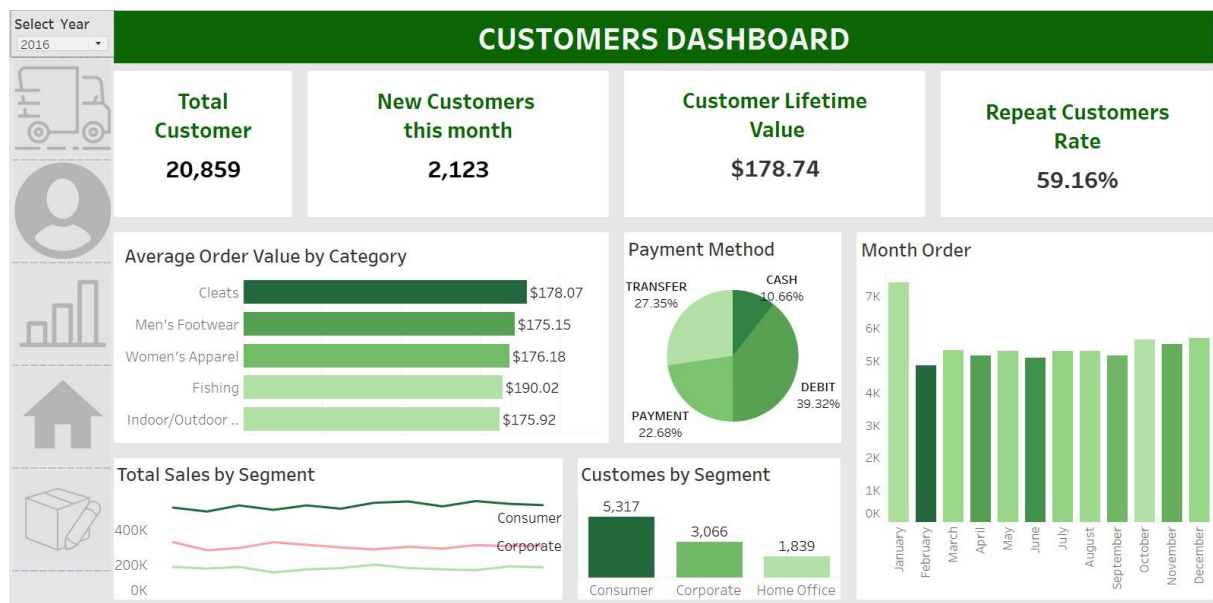


Figure 4.4 Customer Dashboard

The Customers Dashboard for the year 2016 provides detailed insights into customer behavior and performance metrics. Each chart is tailored to analyze various aspects of the customer base, helping to optimize customer relationship management and improve business strategies.

Total Customer, New Customers this month, Customer Lifetime Value, Repeat Customers Rate: These key metrics offer a quick overview of the customer base:

- Total Customer: Shows the total number of customers, providing a sense of the market size.
- New Customers this month: Highlights the number of new customers acquired in the current month, indicating the effectiveness of marketing efforts.
- Customer Lifetime Value (CLV): Represents the average revenue generated by a customer over their lifetime, helping to understand the long-term value of customers.
- Repeat Customers Rate: Indicates the percentage of customers who return for additional purchases, reflecting customer satisfaction and loyalty.

Average Order Value by Category: This bar chart displays the average order value for different product categories. It helps in identifying which categories generate higher revenue per order, guiding marketing and inventory strategies to focus on high-value categories.

Payment Method: The pie chart shows the distribution of payment methods used by customers (Cash, Debit, Payment, Transfer). Understanding payment preferences can help tailor payment options to customer needs and streamline the payment process.

Month Order: The bar chart illustrates the number of orders placed each month. This visualization helps in identifying seasonal trends and peak periods, aiding in inventory management and promotional planning.

Total Sales by Segment: The line chart tracks sales trends across different customer segments (Consumer, Corporate). This chart helps in understanding segment-specific sales performance, guiding targeted marketing efforts and product offerings.

Customers by Segment: The bar chart shows the number of customers in each segment (Consumer, Corporate, Home Office). This helps identify the most significant customer segments, allowing for better resource allocation and strategic focus.

Each chart in this customers dashboard plays a vital role in analyzing customer-related data, from acquisition and retention to payment preferences and order value. These insights enable managers to make informed decisions to enhance customer satisfaction, boost sales, and optimize marketing strategies. The ability to filter data by year (2015-2018) further supports trend analysis and strategic planning over time.

4.1.5 Order Management Dashboard

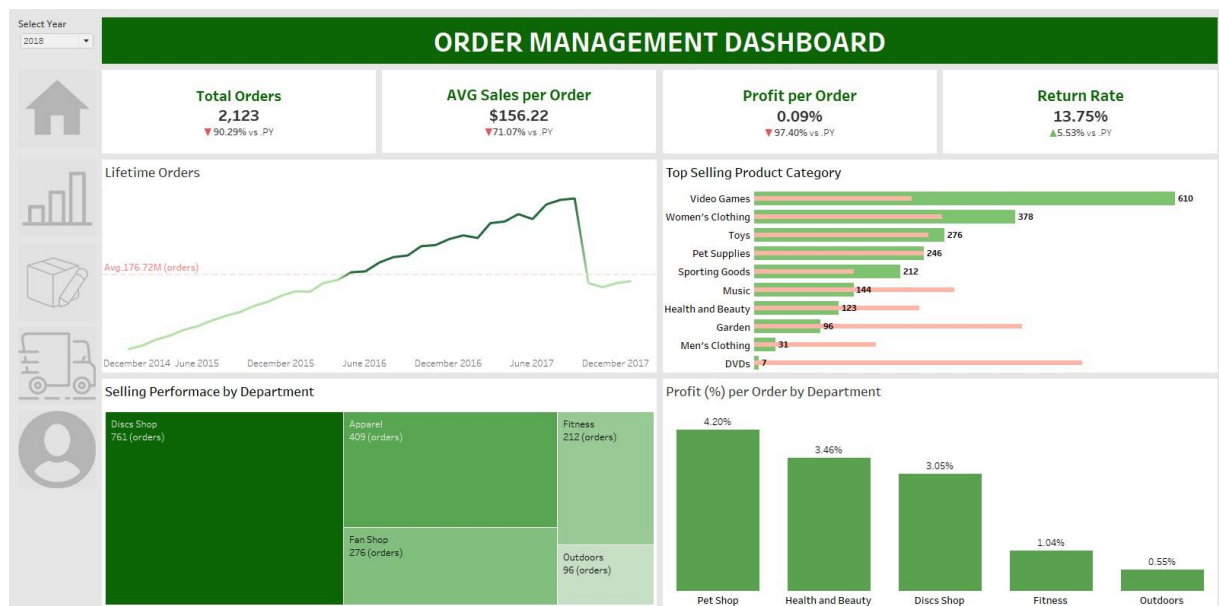


Figure 4.5 Order Management Dashboard

Based on the information presented at the top of the dashboard, it is evident that 2018 was a challenging year for the business, as indicated by significant declines in critical performance metrics, despite the financial year not yet concluding. The total number

of orders dramatically decreased to 2,123, marking a 90.29% reduction, while profit per order plummeted to just 0.09%, a drastic drop of 97.40%. Additionally, the average sales per order fell to \$156.22, a 71.07% decrease from the previous year. Moreover, the return rate escalated to 13.75%, an increase of 5.53%, signaling potential issues in product or service satisfaction.

These declines can be attributed to a range of factors, including heightened competition, shifts in consumer behavior, or internal inefficiencies such as suboptimal inventory management and ineffective pricing strategies. The increase in the return rate, in particular, warrants a detailed analysis to determine its causes, which may involve issues related to product quality, customer service inadequacies, or rigid return policies.

To address these challenges, the company must undertake a thorough investigation into the root causes of these downturns and reassess its current business strategies. Enhancing product quality and customer service should be prioritized to help mitigate the high return rate. Implementing new technologies for better inventory management and demand forecasting could also play a crucial role in cost reduction and revenue enhancement. Furthermore, a strategic review of pricing and promotional tactics is essential to maintain competitiveness and appeal to both new and existing customers. This comprehensive approach will not only help in stabilizing the business in the short term but also lay a strong foundation for sustainable long-term growth.

This dashboard provides insights into the sales performance and profitability of various departments and product categories. A detailed analysis reveals that the cumulative total of orders from 2015 to early 2018 showed steady growth from 2015 to 2017. However, there was a sharp decline in 2018, with the average number of orders dropping to 176.72 million. This sudden decrease could be due to factors such as market saturation, increased competition, or shifts in consumer shopping trends.

The best-selling product categories show that Video Games lead with 610 orders, significantly surpassing other categories, indicating strong growth in this sector. Women's Clothing is second with 378 orders, also showing growth compared to the previous year. Other categories like Toys, Pet Supplies, and Sporting Goods also have notable sales, but they do not achieve the impressive growth seen in Video Games and Women's Clothing.

Regarding sales performance by department, Discs Shop leads with 761 orders, followed by Apparel with 409 orders and Fan Shop with 276 orders. This indicates that Discs Shop is highly effective in attracting customers and increasing sales.

Although Discs Shop has the highest sales volume, Pet Shop leads in profitability with a 4.20% profit margin per order, followed by Health and Beauty at 3.46% and Discs Shop at 3.05%. This suggests that Pet Shop and Health and Beauty are more profitable due to higher profit margins or lower operating costs compared to other departments.

Conclusions drawn from the data suggest that the business should focus on Video Games and Women's Clothing to drive sales and enhance activities for Pet Shop and Health and Beauty to maximize profit per order. Investigating the causes of the sharp decline in order volume in 2018 is necessary to implement timely corrective measures, such as improving marketing strategies or adjusting products to better meet customer needs. Overall, this dashboard provides a comprehensive view of sales and profitability, helping the business make strategic decisions to improve operational performance.

CHAPTER 5: CONCLUSION AND FUTURE WORKS

5.1. Results

Before creating the dashboards in Tableau, our team utilized Tableau Prep for data preprocessing. From an original dataset containing 53 columns, we extracted and transformed the data into several smaller, more manageable tables. This preprocessing step involved cleaning the data, removing unnecessary columns, and ensuring the data was structured in a way that facilitated effective analysis.

We successfully created five comprehensive dashboards, each serving a specific purpose:

- **Dashboard Overview:** This dashboard provided a quick snapshot of key performance metrics, such as total sales, total profit, total orders, average delivery time, and total customers. It also included visualizations to understand sales by type, profit ratio by delivery status, customer distribution by market, and more.
- **Sales Dashboard:** This dashboard offered a detailed analysis of sales performance from 2015 to 2018. It included metrics like total sales, total quantity, and total profit. Visualizations highlighted sales trends, sales by category, sales by country and region, and sales by customer segment.
- **Logistics Dashboard:** This dashboard focused on shipping performance, providing insights into metrics such as late ship rate, average ship time, and schedule versus actual shipping times. It also included analyses of sales and profit by map, on-time delivery trends, and profit by shipping mode and market.
- **Customers Dashboard:** This dashboard provided insights into customer behavior and performance metrics. It included key metrics like total customers, new customers acquired each month, customer lifetime value, and repeat customer rate. Visualizations also showed average order value by category, payment methods, monthly orders, and customer distribution by segment.
- **Order Management Dashboard:** This dashboard analyzed the sales performance and profitability of various departments and product categories. It included metrics on total orders, average sales per order, and return rates. Visualizations highlighted the best-selling product categories, sales performance by department, and profit margins per order.

5.2. Limitations

Despite our efforts to create comprehensive and insightful dashboards, there were certain limitations:

- **Data Quality:** The original dataset contained some inconsistencies and missing values, which may have affected the accuracy of our analysis.
- **Scope:** The analysis was limited to the available data from 2015 to 2018. More recent data could provide a better understanding of current trends.
- **Generalization:** The dashboards were created based on the specific dataset provided, and the findings may not be universally applicable to all businesses or industries.
- **Complexity:** Some visualizations might be complex for stakeholders unfamiliar with data analysis, potentially requiring additional training or simplification.

5.3. Future works

To build upon this project, several future works can be undertaken:

- Data Integration: Incorporating more recent data and integrating data from additional sources could enhance the comprehensiveness and relevance of the analysis.
- Advanced Analytics: Implementing advanced analytics techniques, such as machine learning for predictive analytics, could provide deeper insights and more accurate forecasts.
- User Training: Providing training sessions for stakeholders to better understand and utilize the dashboards could improve decision-making processes.
- Dashboard Enhancement: Continuously improving the dashboards based on user feedback and evolving business needs to ensure they remain relevant and useful.
- Automation: Developing automated data refresh and reporting systems to ensure that the dashboards always reflect the most current data, facilitating real-time decision-making.

By addressing these limitations and pursuing future works, the project can evolve to provide even greater value and support for strategic decision-making in business operations.

REFERENCES

“About Tableau Prep - Tableau.” *Tableau Help*,

https://help.tableau.com/current/prep/en-us/prep_about.htm.

Jain, Sandeep. “Tableau Tutorial.” *GeeksforGeeks*, 4 July 2024,

<https://www.geeksforgeeks.org/tableau-tutorial/>.

“The Tableau Data Model.” *Tableau*,

https://help.tableau.com/current/pro/desktop/en-us/datasource_datamodel.htm.

“Tableau Tutorial.” *Tutorials Point*,

<https://www.tutorialspoint.com/tableau/index.htm>.