UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT

COURSE: DATA VISUALIZATION

Topic:

BUILDING DASHBOARD SYSTEMS IN TABLEAU DESKTOP BASED ON THE ADVENTUREWORKS 2019 DATABASE

MODULES: SALES AND PURCHASING

Ho Chi Minh City, August 8, 2023

UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT

COURSE: DATA VISUALIZATION

Topic:

BUILDING DASHBOARD SYSTEMS IN TABLEAU DESKTOP BASED ON THE ADVENTUREWORKS 2019 DATABASE

MODULES: SALES AND PURCHASING

GROUP 5:

NO	Name	ID	Faculty	Role	Phone	Email
1	Vu Van Son	K20410669	FIS	Leader	0944244857	sonvv20416c@st.uel.edu.vn
2	Chau Nguyen Huong Thao	K204162000	FIS	Member	0919772802	thaocnh20416c@st.uel.edu.vn
3	Nguyen Tuong Vy	K204162009	FIS	Member	0764187037	vynt20416c@st.uel.edu.vn

Lecturer: **M.S. Nguyen Phat Dat**

Ho Chi Minh City, August 8, 2023

ACKNOWLEDGEMENTS

Throughout this course in general, and this project in particular, our group has received dedicated and enthusiastic support from our lecturer as well as our classmates. We would like to express our heartfelt gratitude and deep appreciation to M.S. Nguyen Phat Dat for being with us throughout this course and during the execution of this final project.

Despite our efforts to carry out the project and achieve the best possible results, it is inevitable that unintended mistakes may arise in the process. We hope to receive your understanding and valuable feedback, both from our lecturer and everyone. This serves as our motivation to continually strive for improvement and refinement in future projects

COMMITMENT

We would like to assure you that our final project is original and based on the research conducted by the entire team. We have also included a list of documents that we referenced in our report and have cited them accordingly.

In the event that any of the information provided is found to be incorrect or misleading, we take full responsibility for any consequences that may arise, and we will work closely with our professors to rectify the situation.

Ho Chi Minh City, May, 5, 2023

Committed person

Group 7

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

NO.	ACRONYMS	MEANINGS
1	BI	Business Intelligence
2	ETL	Extract - Transform - Load
3	KPI	Key Performance Indicator
4	RFM	Recency, Frequency, Monetary
5	AWC	Adventure Works Cycles

CHAPTER 1: INTRODUCTION (done all)

1.1 Business case for the project

AdventureWorks, a prominent player in the retail industry, expertly manages diverse product portfolios across multiple geographical regions. As a forward-thinking enterprise, AdventureWorks recognizes the pivotal role of data in modern business strategies. To amplify operational efficiency and facilitate well-informed decision-making, the company has strategically constructed a robust data warehouse, adeptly integrating data from disparate sources.

In its unwavering pursuit of excellence, AdventureWorks now embarks on a visionary data visualization project, underpinned by Tableau's prowess. This pivotal initiative is laser-focused on the crucial purchasing and sales modules, where a trove of untapped insights lies dormant.

This project represents a pivotal step in AdventureWorks' journey towards becoming a data-driven enterprise. By seamlessly integrating Tableau's ETL capabilities and data visualization prowess, the company is poised to unlock the full potential of its data assets. As a result, AdventureWorks anticipates not only elevating its decision-making prowess but also securing a competitive edge in the dynamic and ever-evolving retail landscape.

1. 1.2 Objectives of the project

1. 1.2.1 General Objective

The general objective is to leverage visual representations of data to gain insights, improve decision-making, enhance efficiency, and identify growth opportunities within the organization.

2. 1.2.2 Specific Objectives

Gain Insights: Visualize purchasing and sales data to gain actionable insights into customer behavior, product performance, inventory management, and supplier relationships.

Improve Decision-Making: Enable stakeholders to make informed decisions by providing intuitive and interactive visualizations that highlight key trends, patterns, and anomalies in purchasing and sales data.

Enhance Efficiency: Streamline processes related to purchasing and sales by identifying bottlenecks, optimizing inventory levels, and reducing supply chain costs through data-driven analysis.

Identify Growth Opportunities: Uncover untapped market segments, product opportunities, and customer preferences to drive revenue growth and competitive advantage.

2. 1.3 Research Objects

The research objects of this business intelligence (BI) project revolve around the purchasing and sales modules in AdventureWorks database. The primary focus is to gain deeper insights and understanding by leveraging data visualization techniques. Key research objects include:

- ❖ Purchasing Trends: Analyzing historical purchasing data to identify trends, patterns, and seasonality in order to optimize procurement processes and supplier relationships.
- ❖ Sales Performance: Exploring sales data to uncover valuable insights about customer behavior, product preferences, and market trends, enabling targeted sales strategies and revenue growth.
- ❖ Inventory Management: Examining inventory levels, turnover rates, and stock-outs to improve operational efficiency, minimize carrying costs, and enhance customer satisfaction.

3. 1.4 Scope of the Project

Our project revolves around the implementation of a robust Business Intelligence (BI) and Data Analytics (DA) solution for AdventureWorks, a fictitious retail entity operating within the retail sector. The key objective of this endeavor is to elevate operational efficiency and strategic decision-making through the utilization of data-driven insights. Project encompasses the development of comprehensive and interactive dashboards tailored to AdventureWorks' Sales and Purchase modules. These dashboards will provide a holistic view of key performance indicators (KPIs) and trends, enabling stakeholders to make informed decisions and optimize operational strategies.

The project will focus on visualizing the company's revenue trends over the years, broken down by Product Category and Sales Territory. Additionally, the visualization will spotlight customer segmentation based on RFM scores and offer insights into purchasing behaviors. The Purchase module will showcase supplier performance, order distribution, and expenditure breakdown by category. The project's user-centric approach ensures intuitive navigation, allowing users to filter and drill down into data for detailed analysis. Data security measures will be explored, and documentation and training will be provided to ensure effective utilization of the visualization tools. The scope aims to empower AdventureWorks with actionable insights, fostering data-driven decision-making and operational efficiency improvements.

4. 1.5 Value and desired outcome of the project

The value and desired outcome of the project are rooted in the overarching goals of enhancing the travel booking application's functionality, improving user experience, and leveraging data-driven insights to drive informed decision-making. By successfully implementing the project, the following key values and outcomes are anticipated:

Enhanced User Experience: The project aims to provide users with a seamless and personalized travel booking experience. By integrating sentiment analysis, users will receive more relevant recommendations based on their sentiments and preferences. This personalization will result in increased user satisfaction and loyalty, contributing to a positive brand image.

Improved Operational Efficiency: Through the automation of sentiment analysis and recommendation generation, operational processes will become more streamlined and efficient. This will reduce manual effort, allowing staff to focus on higher-value tasks, ultimately leading to cost savings and improved resource allocation.

Data-Driven Decision Making: The implementation of data visualization dashboards empowers stakeholders with actionable insights derived from real-time data. This enables more informed and strategic decision-making at all levels of the organization. Leaders can identify trends, patterns, and opportunities to optimize business operations and offerings.

Increased Revenue Generation: By offering personalized recommendations and optimized itinerary planning, the project aims to drive higher conversion rates from recommendations to actual bookings. This will lead to an increase in the number and value of bookings, directly contributing to revenue growth.

Enhanced Customer Engagement: The integration of sentiment analysis will allow AdventureWorks to proactively address customer concerns and feedback. Engaging with customers in a meaningful way will foster stronger relationships and trust, resulting in increased customer retention and word-of-mouth referrals.

1.6 Structure of project

The project structure consists of 6 parts:

- **Chapter 1: Introduction**

In this initial chapter, the project will be introduced, highlighting the AdventureWorks sample database as the foundation. The chapter will outline the project's overarching objectives and aims, emphasizing its significance and relevance in the context of data visualization and Business Intelligence (BI) practices. Essential information about the project scope and key details will also be provided to set the stage for the subsequent chapters.

- **Chapter 2: Theoretical Basis**

This chapter delves into the essential theoretical underpinnings of Business Intelligence solutions, data warehousing, data integration, and data analytics. Through an exploration of core concepts and theories, learners will establish a strong foundation in BI practices, fostering a deeper comprehension of the project's technical aspects and methodologies.

- **Chapter 3: Requirements Analytics and Introduction to BI Solution**

Chapter 3 focuses on a meticulous analysis of project requirements, ensuring alignment with the overarching goals. It introduces the Business

Intelligence solution as a pivotal tool for transforming raw data into actionable insights. Learners will explore the functionalities and capabilities of BI tools, understanding their vital role in addressing data-driven challenges.

- **Chapter 4: Building Data Warehouse and Integrating Data**

This pivotal chapter guides participants through the intricate process of constructing a robust data warehouse. Detailed insights into the Extract, Transform, Load (ETL) process will be provided, shedding light on how data is extracted from various sources, transformed into usable formats, and loaded into the data warehouse. Learners will gain hands-on experience with cutting-edge tools and software that facilitate seamless data integration.

- **Chapter 5: Results - Data Analytics and Visualization**

The heart of the project unfolds in Chapter 5, where learners will witness the tangible outcomes of their efforts. Through comprehensive data analysis and visualization using tools like Tableau, participants will bring data to life, unveiling patterns, trends, and insights that drive informed decision-making. This chapter showcases the power of visual representation in conveying complex information and underscores the project's impact on facilitating AdventureWorks' business strategies.

- **Chapter 6: Conclusion and Future Works**

In the concluding chapter, the project's achievements will be summarized, providing an insightful overview of the journey. A reflective evaluation of the implementation process will be offered, highlighting successes, challenges, and lessons learned. Moreover, this chapter will present forward-looking recommendations for future enhancements and advancements, offering a roadmap for further exploration and optimization of the project's potential.

CHAPTER 2: THEORETICAL BASIS

2. 2.1 Business Intelligence (BI) Fundamentals:

This module serves as an intellectual compass guiding participants through the intricate tapestry of Business Intelligence (BI). At its core, BI represents an ensemble of methodologies, technologies, and processes woven together to illuminate the path of organizational success. This profound exploration unfolds across several key dimensions:

2.1.1 What is BI?

At its foundational essence, Business Intelligence (BI) represents a strategic and methodical approach that orchestrates the entire lifecycle of business information, from its inception through to its utilization as a catalyst for insightful decision-making. This theoretical aspect involves an in-depth exploration of BI's evolutionary trajectory, traversing the historical path from conventional, manual reporting systems to the contemporary data-driven paradigm that harnesses cutting-edge technology.

Within this journey, BI's evolution is characterized by a transformative shift from static data repositories to dynamic data ecosystems. The evolution encompasses the refinement of data collection methods, the establishment of robust data integration mechanisms, and the ascendancy of advanced analytics techniques. Furthermore, the exploration of BI's historical context serves as a lens through which to comprehend its metamorphosis into a strategic asset, transcending its humble origins to become an indispensable driver of organizational growth and competitive advantage.

Understanding the key components of BI becomes imperative in this theoretical exploration. This encompasses the intricate orchestration of data warehousing, a pivotal reservoir that consolidates disparate data sources into a unified repository. The transformative power of Extract, Transform, Load (ETL) processes comes into play, as data is extracted, cleansed, and harmonized to unveil hidden insights. Analytics tools, ranging from descriptive to prescriptive, assume center stage in unlocking the latent potential of data. Moreover, the art of visualization emerges as a potent conduit for conveying complex information with simplicity and clarity.

2.1.2 Objectives

Within the panorama of Business Intelligence, objectives stand as guiding beacons that navigate the course of implementation within organizations. This theoretical dimension unfurls an elaborate tapestry of objectives that reflect the myriad facets of BI's

impact. From the vantage point of enhancing operational efficiency, BI endeavors to streamline processes, minimize redundancies, and amplify resource utilization, thus paving the way for heightened agility and responsiveness.

Optimizing resource allocation assumes paramount significance, with BI wielding the power to illuminate areas of excessive expenditure and opportunities for prudent investment. In the realm of strategic planning, BI serves as an invaluable compass, guiding organizations through the labyrinth of decision-making by fusing data-driven insights with visionary foresight. Customer insights unfurl as a treasure trove within BI's arsenal, where an intimate understanding of consumer behavior and preferences leads to tailored experiences and enriched engagement.

Moreover, BI's objectives span the horizon of foresight, enabling organizations to navigate dynamic market trends with poise and prescience. These objectives, collectively, interlace BI with broader organizational strategies, manifesting its role as an enabler of innovation and sustained growth.

2.1.3 Advantage of BI in enterprises

BI's significance unfurls a captivating narrative of its profound impact within modern business realms. As data assumes the mantle of the world's most coveted resource, BI rises to the occasion as the catalyst that transforms this raw material into a strategic asset. In the crucible of exponential data growth, BI emerges as a beacon of order, offering the tools to harness the deluge of information and distill it into actionable insights.

BI's prowess to facilitate real-time decision-making becomes a cornerstone of its modern-day relevance. The theoretical landscape here encompasses a contemplation of how BI equips organizations to respond with alacrity to swiftly evolving market dynamics. This agility, in turn, paves the way for data-driven strategies that adjust in real-time, positioning organizations to seize fleeting opportunities and navigate challenges with acumen.

Moreover, BI's ability to foster competitive advantages becomes an emblem of its significance. In the relentless pursuit of market dominance, organizations wield BI as a potent weapon to enhance market responsiveness, optimize customer experiences, and unveil uncharted avenues of innovation. The culmination of these factors positions BI as an architect of sustainable growth and prosperity within modern business landscapes.

2.1.4 significance in Modern Business Ecosystems:

Agile Responsiveness: In an era of constant change, Business Intelligence empowers organizations to be proactive navigators of dynamic market dynamics. Swift data analysis and interpretation enable timely adjustments, ensuring that businesses stay ahead of shifts and capitalize on emerging opportunities.

Market Discernment: Business Intelligence delves deep into data to provide organizations with a richer comprehension of market trends, consumer preferences, and untapped opportunities. By unraveling the intricate threads of information, BI equips businesses with the insight needed to make informed and strategic decisions.

Risk Mitigation: Armed with data analysis and scenario dissection capabilities, BI serves as a shield against potential risks. By identifying vulnerabilities and predicting potential pitfalls, organizations can proactively implement measures to mitigate risks and safeguard their operations.

Informed Discernment: Business Intelligence bestows decision-makers with the tools for informed discernment. Precise, pertinent, and current information empowers leaders to make choices backed by data-driven insights, minimizing uncertainty.

Holistic Insight: With a panoramic lens, BI provides a comprehensive view across multifaceted dimensions of an organization – from finance to operations, sales to customer satisfaction. This holistic insight fosters a well-rounded understanding of the entire ecosystem.

Scenario Anticipation: Similar to a strategist envisioning multiple scenarios, BI facilitates scenario anticipation. Through simulations and data modeling, BI empowers decision-makers to explore diverse pathways, enabling foresight-driven choices.

Iterative Enhancement: Just as alchemy pursued mastery through experimentation, BI cultivates a culture of continuous improvement. Data-driven learning fuels iterative enhancements, driving organizations towards operational excellence and innovation.

2.1.5 Tableau

Tableau is a powerful and widely-used data visualization and business intelligence tool that empowers users to create interactive and shareable visualizations, reports, and dashboards from various data sources. It simplifies complex data analysis and enables users to gain valuable insights quickly.

Tableau Desktop:

Tableau Desktop is the primary authoring and development tool within the Tableau ecosystem. It provides a user-friendly interface for creating, designing, and publishing interactive visualizations. Key features include:

Data Connection: Tableau Desktop allows you to connect to a wide range of data sources, including databases, spreadsheets, cloud services, and more.

Drag-and-Drop Interface: The intuitive drag-and-drop interface makes it easy to create charts, graphs, maps, and other visualizations without the need for coding.

Data Transformation: You can perform data transformations directly within Tableau Desktop using calculated fields, parameters, and other functions to clean, reshape, and aggregate data.

Visualization Creation: Design rich and interactive visualizations using a variety of chart types, filters, hierarchies, and formatting options.

Dashboards and Stories: Combine multiple visualizations into interactive dashboards and stories to provide a comprehensive view of your data.

Analytics and Calculations: Perform advanced analytics and calculations using functions and expressions to derive insights from your data.

Real-Time Data: Connect to live data sources to visualize real-time updates and changes.

Tableau Prep Builder:

Tableau Prep Builder is a data preparation tool that complements Tableau Desktop. It allows you to clean, shape, and combine data from various sources before bringing it into Tableau Desktop. Key features include:

Data Profiling: Understand the structure and quality of your data through data profiling and exploration.

Data Cleaning: Identify and address data quality issues, such as missing values, duplicates, and inconsistencies.

Data Transformation: Use a visual interface to transform, pivot, aggregate, and filter data without writing code.

Data Joining: Combine data from multiple sources by defining joins and relationships.

Automated Workflows: Create repeatable and automated data preparation workflows for consistent results.

Output Options: Export clean and transformed data to various formats for analysis in Tableau Desktop.

Data Visualization and ETL Workflow:

Connect Data: Use Tableau Prep Builder to connect to various data sources, perform data profiling, and clean the data.

Data Transformation: Transform and reshape the data as needed using Tableau Prep Builder's visual interface.

Output Clean Data: Export the cleaned data from Tableau Prep Builder to a format suitable for analysis.

Connect in Tableau Desktop: Use Tableau Desktop to connect to the cleaned data and create visualizations, dashboards, and stories.

Visualization Design: Design interactive visualizations using Tableau Desktop's drag-and-drop features.

Dashboard Creation: Combine visualizations into interactive dashboards to convey insights effectively.

Publish and Share: Publish your Tableau workbooks and data sources to Tableau Server or Tableau Online for sharing and collaboration.

Tableau Extract files (.hyper):

A Tableau Extract file with the extension .hyper is a proprietary data file format used by Tableau software. It's specifically designed to store and manage data extracts for optimized querying and visualization within Tableau.

Tableau Extract files are highly compressed, columnar storage files. They store data in a way that makes it efficient for querying and analysis, allowing for faster loading times and more responsive visualizations in Tableau. These files include metadata, data structures, and indexes that enable Tableau to quickly retrieve and display data, even when dealing with large or complex datasets.

Extract files can be created by connecting to a data source, specifying filters, aggregations, and calculations, and then saving the results in this format. They can be refreshed to keep the data up-to-date, and they provide a convenient way to work with data for visualizations and analytics in Tableau, especially when dealing with scenarios where direct querying of the source database might be slow or impractical.

2.2 ETL Process

2.2.1 What is ETL?

ETL stands for Extract, Transform, and Load, as outlined by IBM [7]. This refers to a data integration methodology that merges information from multiple origins, creating a uniform and coherent data repository that can be directed into a data warehouse or another intended platform. Consequently, it can be deduced that ETL serves as a pivotal mechanism for converting data from diverse systems into a format suitable for storage within a data warehouse or data analytics system. The ETL procedure encompasses the collection of data from various sources, the alteration of data to align with the specifications of the data warehouse or data analytics system, and the subsequent loading of the processed data into the warehouse or analytics platform to facilitate analysis and reporting

2.2.2 Why do we need ETL?

ETL is needed because data often exists in different sources with various formats and qualities. ETL helps by collecting important data from these sources, making it consistent, and putting it all together in one central place called a data warehouse. This makes sure the data is accurate, of good quality, and easy to analyze and report. ETL also helps organizations combine data from different sources into one reliable source for making decisions.

2.2.3 ETL Basic Process

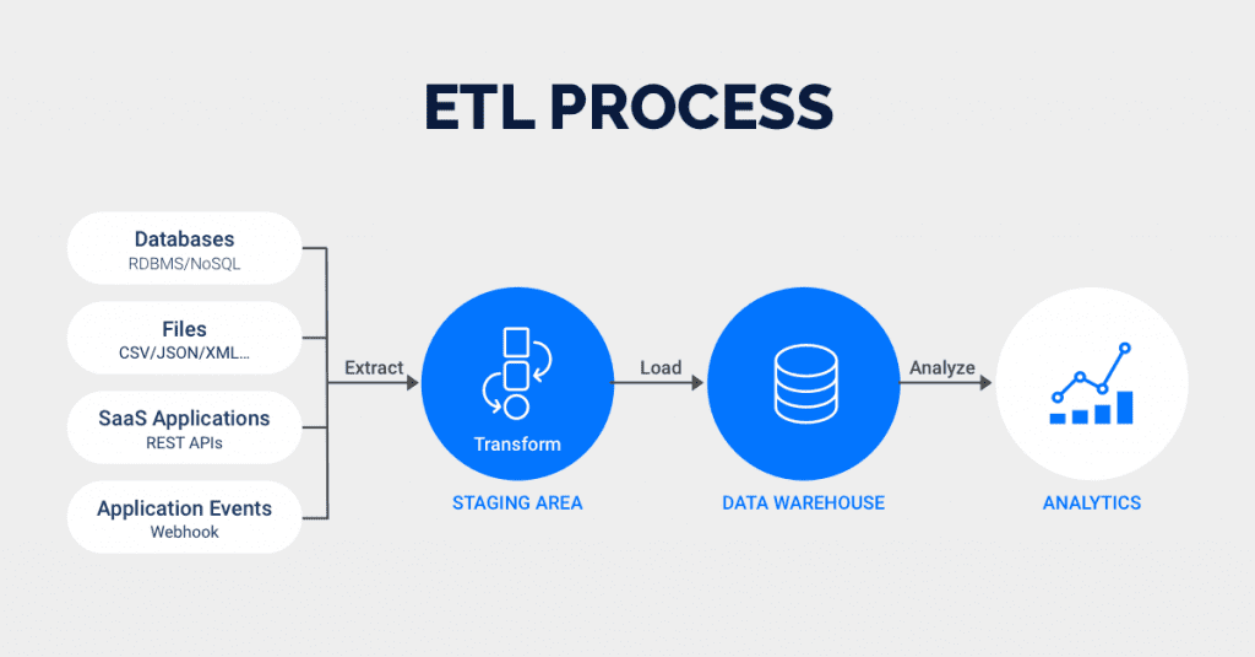


Figure 3: ETL Process (source:rivery.io)

Figure 3 illustrates a simplified representation of the ETL process. The ETL layer is primarily engaged in three key steps: extracting data, transforming it, and loading it. Given that data often originates from disparate internal and external sources that may be disjointed, incomplete, inconsistent, or susceptible to duplication, the extraction step gathers data from various origins and filters out relevant information to support organizational decision-making. In order to maintain consistency and prevent the need for repeated extractions, the extracted data is temporarily stored. Data transformation and cleansing involve converting data formats using business rules and standardizing data definitions to ensure uniformity across the organization. Data cleansing identifies and rectifies data discrepancies according to predefined standards, ensuring the accuracy and comprehensiveness of the data. After the data has been converted and cleaned, it is stored in the data repository to prevent the need for re-conversion in case of data download interruptions. Subsequently, the data is loaded into the designated location.

In this project, we used the Tableau Prep Builder tool to ETL processes. Consequently, there are several differences compared to the basic ETL workflow.

2.3 KPIs

2.4.1 KPIs Definition

Key Performance Indicators (KPIs) are measurable indicators used to assess the efficiency of an organization in implementing its strategic goals. These quantifiable metrics evaluate an organization's performance in accomplishing its objectives. KPIs unify various levels within the organization, from business units to departments and individuals, by establishing clear targets and benchmarks that promote accountability and monitor advancement. KPIs are strategic drivers of value that transcend the mere measurement of insignificant business activities and processes.

A KPI should serve as a practical tool rather than a theoretical concept. It should serve as a fundamental criterion, essentially a language that measures project effectiveness and predicts its likelihood of success. KPIs are rarely standalone; typically, a set of KPIs is required. Lack of consensus on KPI interpretation could lead to disorder rather than focused action.

2.4.2 Categories of KPIs

❖ **Sales module**

- **Revenue:**
    - Gross revenue or top-line revenue:

Revenue is calculated by incorporating tax amount, freight charges, and discounts, resulting in what is known as gross revenue or top-line revenue. Gross revenue reflects the complete sum of money collected from customers, excluding deductions or expenses. This method of revenue calculation provides insight into the overall scale and expansion of the business, yet it does not offer an assessment of the business's profitability or efficiency.

    - Net revenue or bottom-line revenue:

Revenue is calculated by removing taxes, freight, and discounts, giving us net revenue or bottom-line revenue. Net revenue shows the actual earnings after subtracting costs of delivering products or services. This helps gauge business profitability and efficiency but doesn't show the full potential or market demand.

- **RFM score**

RFM score is a method used by businesses to group customers based on their behavior. It considers three aspects: Recency (how recently a purchase was made), Frequency (how often purchases occur), and Monetary Value (how much money is spent). To calculate the RFM score, just add the three scores together. A higher score means a more valuable customer.

$$RFM\ Score = Recency\ Score\ \_\ Frequency\ Score\ \_\ Monetary\ Score \quad (1)$$

Currently, there are various ways to calculate the RFM score. However, in this instance, we employ quartile ratios derived from individual calculations of each factor (Recency, Frequency, Monetary Value) to assign scores. Subsequently, RFM combines and categorizes customers based on these scores.

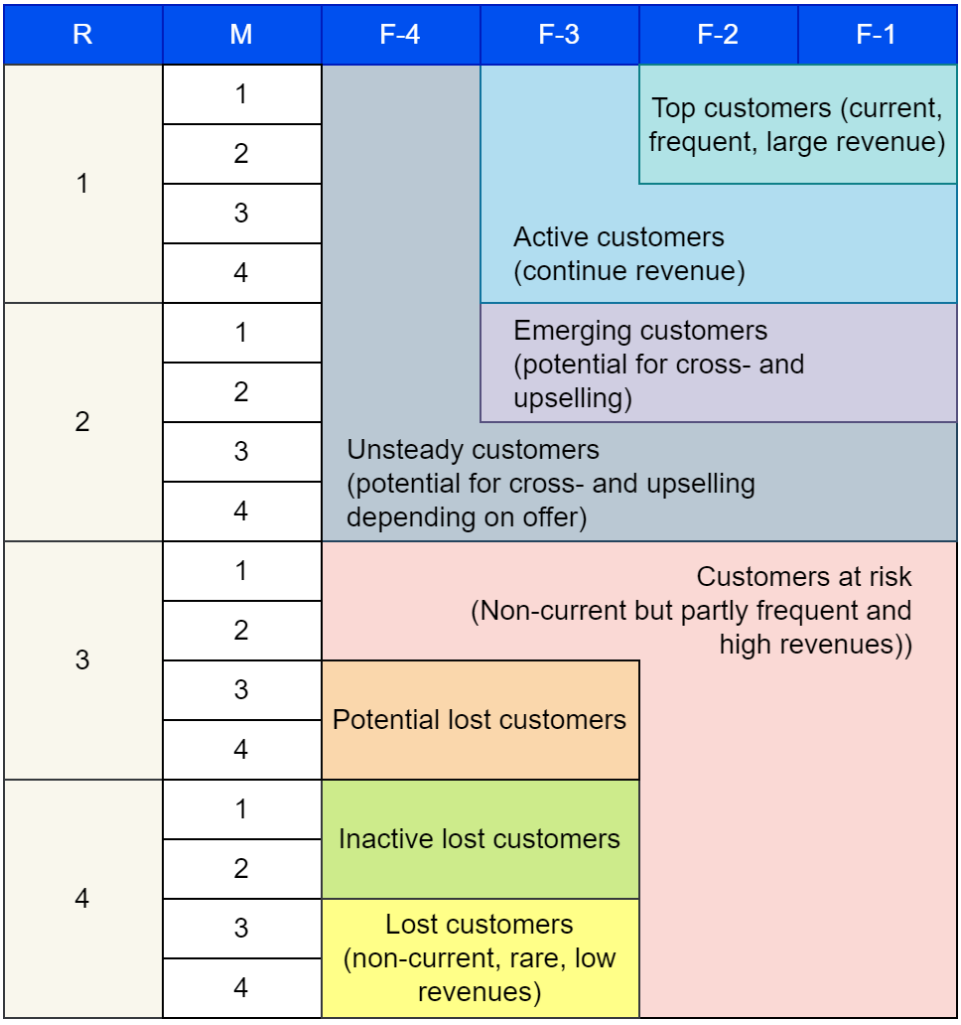


Figure 6: A matrix of the RFM score.

Our aim is to categorize customers into various groups that align with our company's objectives. We consider customers who have made recent, frequent, and

substantial purchases as the most valuable. To achieve this, we use quartiles for Recency, Frequency, and Monetary values, resulting in 64 potential combinations of RFM scores. We then assign each customer segment to one of these combinations based on their resemblance to our technical and business goals. We assign scores from 1 to 4 for each RFM value, with 1 being the highest and 4 the lowest. However, we invert the scoring for Recency, where a higher value receives a lower score. For instance, a customer with an RFM score of 111 is our top-tier customer, while a score of 444 indicates a lost customer.

- **Customer Churn**

Customer Churn is a metric that evaluates the number of customers who discontinue their association with a company within a specific timeframe. Elevated churn rates might signal inadequate customer satisfaction or heightened market competition. The formula to calculate the customer churn rate is:

$$\frac{Number\ of\ Customers\ Lost}{Total\ Number\ of\ Customers} x\ 100 \qquad (2)$$

- **Customer Retention Rate:**

The retention rate is a measure of how many customers or employees remain with a company or product over a given period of time. It is usually expressed as a percentage and can be calculated using different formulas depending on the context. Here are some possible formulas and definitions of retention rate:

$$Retention\ Rate\ = \frac{Ending\ Count - New\ Count}{Starting\ Count} x\ 100 \quad (3)$$

Where:

- Starting Count: The count of customers/users/employees at the beginning of the time period.
- Ending Count: The count of customers/users/employees at the end of the time period.

- **Sales Quota**

The total employee sales quota is established based on the collective sales achieved by the sales team within a specified time frame. This calculation enables the company to assess the individual performance of each salesperson.

❖ **Purchasing module**

- **Vendor rejection rates and costs**

  This metric serve as indicators of the quality of products and services supplied by a vendor. These metrics can be computed using the following formula:

$$Supplier\ rejection\ rate\ and\ cost\ = \frac{total\ of\ the\ rejected\ product}{total\ of\ received\ product} x\ 100 \quad (3)$$

  where:

- total cost of rejected products" is the total cost of rejected products due to quality or other problems among the products received from the vendor.
- "total cost of received products" is the total cost of the product received from the vendor.

- **Supplier Lead Time**

  This is a metric that evaluate the duration between placing an order and the delivery of goods or services by a supplier. The formula to calculate supplier lead time is:

$$Supplier\ Lead\ Time\ = \frac{Delivery\ Date - Order\ Date}{Total\ Number\ of\ Orders} x\ 100 \quad (4)$$

  Where:

  - Delivery Date: The date when the items are delivered by the supplier.
  - Order Date: The date when the order for items was placed.
  - Total Number of Orders: The total count of orders placed.
  - New Count: The count of new customers/users/employees acquired during the time period.

- **Vendor Availability:** This is the percentage of items or services that the vendor can provide on time and as requested. It can be calculated by dividing the receiveqty by the orderqty for each vendor and multiplying by 100. A higher vendor availability indicates a more reliable and consistent vendor.

$$Supplier\ Availability\ = \frac{Receive\ Quatity}{Order\ Quantity}$$

  Where:

- Received Quantity: The quantity of items received from the supplier.
- Order Quantity: The quantity of items initially ordered from the supplier.

2.4.3 The advantages and disadvantages of KPIs

❖ **Advantages**

- *Aligns Business Objectives:* KPIs synchronize business goals with the broader organizational strategy, establishing a transparent comprehension of desired accomplishments and the means of measurement. This promotes a unified approach, minimizing misunderstandings and uncertainties.
- *Identifies Opportunities for Improvement:* KPIs facilitate the identification of areas needing enhancement through the evaluation of performance against defined targets and benchmarks. They offer valuable insights into potential improvements and showcase accomplishments that can serve as models for other segments of the organization.
- *Improves Decision Making:* KPIs offer precise and pertinent data, enabling well-informed decision-making. They aid in recognizing trends and patterns that can inform strategy adjustments and guide data-based choices.
- *Facilitates Performance Management:* KPIs foster a culture of responsibility and monitor advancement towards targets and objectives. They establish a foundation for assessing employee achievements, pinpointing training requirements, and identifying coaching prospects.
- *Accelerates Collaborative Planning:* KPIs promote smooth and collaborative planning throughout the organization by establishing a common comprehension of goals and objectives. This ensures alignment towards shared targets, diminishing redundant work, and enhancing overall efficiency.

❖ **Disadvantages**

- *Over-reliance on KPIs:* Organizations can develop an excessive dependence on KPIs, treating them as the sole evaluate of achievement. This approach may lead to a restricted emphasis on immediate objectives, potentially compromising the organization's long-term prosperity.

- *Incorrect KPI selection:* Picking the wrong KPIs can lead to unhelpful actions and less-than-ideal outcomes. Organizations need to choose KPIs carefully, focusing on what truly matters for their goals and objectives.
- *Disregarding qualitative data:* KPIs often rely on numbers, which may overlook qualitative insights. Qualitative data can offer valuable understanding about an organization's performance and should not be ignored in favor of quantitative data alone.
- *Unrealistic targets:* Establishing unattainable goals can discourage employees and lead to a lack of support for KPIs. Organizations should make certain that KPIs are realistic, matching the abilities of their staff and available resources.

CHAPTER 3: STUDY FRAMEWORK

3.1 Business processes

3.1.1 Sales department

❖ **The purpose of Sales**

The primary objective of the sales division within a company is to facilitate the sale of its products and services. Collaborating with various stakeholders, a salesperson aims to finalize agreements, enhance revenue generation, and cultivate enduring client relationships to promote recurring business and brand loyalty. This involves identifying opportunities in the market to present products and services to both existing and potential customers effectively. Sales professionals create compelling presentations tailored to customer needs, maintaining regular communication to comprehend their preferences. The division also establishes annual budget goals and closely monitors quarterly budget execution rates. In formulating target budgets or forecasts, the unique business potential of each client is considered. Building strong affiliations with current and prospective customers is instrumental in gaining insights into their future aspirations.

❖ **Sales process**

AdventureWorks employs two distinct sales channels: distribution channels for online sales and direct online sales through the company's website. The online sales category encompasses bicycles, clothing, accessories, and parts. Notably, three key product classifications—mountain bikes, road bikes, and touring bikes—are offered under the bicycle segment. Regional sales offices allocate multiple sales territories to sales representatives, each managed by a team leader overseeing multiple sales agents. Distribution channel sales transactions encompass four stages: Order Processing, Packing and Goods Delivery, Invoicing and Debit Recognition, and Payment. Meanwhile, the internet sales channel follows a streamlined two-step process: Order and Payment, followed by Delivery and Invoicing.

3.1.2 Purchasing Department

❖ **The purpose of Purchasing**

The purchasing department's core objective revolves around strategic procurement planning. It aims to efficiently source products through strategic e-sourcing, emphasizing competitive pricing and stringent quality standards. The department plays a pivotal role in deciding whether to procure externally from vendors or pursue internal production during the strategic purchasing phase.

❖ **Purchasing process**

The purchasing division at AdventureWorks is responsible for acquiring essential raw materials and components required for bicycle manufacturing. Additionally, AdventureWorks procures goods for resale, including cycling accessories like water bottles and pumps, as well as cyclist apparel. The AdventureWorks sample database comprehensively documents these items and their respective suppliers. This procurement process ensures a seamless flow of resources, contributing to the company's successful bicycle production and distribution.

3.2 Data source and challenges

3.2.1 Data Source

Transactional Databases: The AdventureWorks 2019 database stands as a robust and invaluable data reservoir for both the sales and purchasing departments. Within its structured framework, a wealth of data pertaining to orders, customers, products, and suppliers resides, providing a rich foundation for analytical insights.

The organizational structure of the AdventureWorks database, as illustrated in Table 2:

Table 2: AdventureWorks 2019 database

Schema	**Description**	**Number of tables**
Human Resources	Employees of the company Adventure Works Cycles.	6
Person	Names and addresses of customers: individual customers, suppliers and employees.	13

Production	Products manufactured and sold by Adventure Works Cycles.	25
Purchasing	Suppliers of products that the company buys.	5
Sales	Customers and data related to the purchase.	18

3.2.2 Challenges

Data Integration Complexity: Integrating diverse data sources from transactional databases, data warehouses, and external systems into Tableau can be intricate. Ensuring accurate and timely data extraction, transformation, and loading (ETL) processes while maintaining data consistency is crucial for meaningful visualizations.

Data Volume and Variety: Dealing with large volumes of sales and purchasing data, including details of orders, products, customers, and suppliers, requires efficient data handling techniques. Overcoming the challenges of handling structured and unstructured data while maintaining performance is essential.

Data Cleansing and Quality: Raw data may contain inconsistencies, errors, and missing values. Ensuring data quality through effective cleansing processes before visualization is a critical step to avoid misleading insights.

Complex Data Relationships: Sales and purchasing data often have complex relationships that need to be appropriately modeled in Tableau. Creating meaningful relationships and hierarchies can be challenging, especially when dealing with multi-level categorizations.

Performance Optimization: Designing dashboards that load quickly and provide responsive interactions, even with complex queries and large datasets, demands optimization techniques such as data aggregation, caching, and efficient use of Tableau's features.

User-Friendly Interface: Designing an intuitive and user-friendly interface that allows users to easily explore and interact with the data without feeling overwhelmed requires a deep understanding of user needs and behaviors.

Choosing the Right Visualizations: Selecting appropriate visualizations to effectively convey insights while avoiding information overload can be challenging. Balancing aesthetics with data representation clarity is crucial.

Cross-Functional Insights: Integrating sales and purchasing data to provide holistic insights across departments demands a clear understanding of business processes and the ability to create cohesive visualizations.

Data Security and Compliance: Ensuring that sensitive sales and purchasing data is appropriately secured and complying with data protection regulations adds another layer of complexity to the project.

3.3. Business Requirements Analysis

3.3.1. Sales Requirements Analysis

This facet delves into discerning the intricate data prerequisites essential for propelling the Sales department's operations and strategic decision-making. The pursuit encompasses the identification of primary data origins, encompassing sales transactions, customer engagements, and promotional campaigns. The range spans determining the specific data categories vital for informed choices, encompassing customer demographics, sales performance metrics, and inventory particulars. A vital culmination involves identifying the cardinal Key Performance Indicators (KPIs) that will be the yardsticks for gauging sales accomplishments and gauging the headway towards sales targets.

3.3.2. Purchasing Requirements Analysis

In analogous fashion, the Purchasing department's data requisites demand meticulous attention to endorse operational efficiency and informed judgments. The exploration entails recognizing pivotal data fountains, spanning purchase orders, supplier accords, and inventory thresholds. This expedition also encompasses ascertaining the exact nature of data required, spanning supplier performance metrics, purchasing history, and inventory turnover insights. Comparable to its Sales counterpart, the Purchasing Requirements Analysis culminates in isolating pivotal Key Performance Indicators (KPIs) pivotal for gauging procurement achievements and charting progress towards stipulated procurement goals.

3.4. IT Requirements Analysis (IT & Infrastructure)

- **Hardware:** The project will require a server with sufficient processing power, memory, and storage to store and process the data. The server should also have a high-speed internet connection to allow for the transfer of data between the server and the Tableau Desktop and Tableau Prep Builder applications.

- **Software:** The project will require the following software:

  SQL Server: This database management system will be used to store the data.

  Tableau Desktop: This software will be used to create interactive dashboards and reports.

  Tableau Prep Builder: This software will be used to clean and transform the data before it is loaded into Tableau Desktop.

- **Infrastructure:** The project will require the following infrastructure:

  A network that connects the server to the Tableau desktop and Tableau Prep Builder applications.

  A storage system for the data.

  A security system to protect the data.

3.5. Workflow

3.5.1. Project Initiation and Planning

Define the project goals, objectives, and scope related to data visualization.

Identify the stakeholders and their requirements for the sales and purchase module dashboards.

Establish the key performance indicators (KPIs) that will drive the design of your visualizations.

Define the data sources, including the AdventureWorks 2019 database, and understand the data structure.

Plan the timeline, resources, and team members involved in the visualization project.

3.5.2. Data Understanding and Preparation

Conduct a comprehensive analysis of the data available in the AdventureWorks 2019 database. This includes understanding the data types, the relationships between the tables, and the quality of the data.

Clean and preprocess the data, addressing any missing values, duplicates, or inconsistencies. This is an important step to ensure that the data is accurate and reliable.

Perform data transformations and aggregations to create the necessary datasets for your visualizations. This involves transforming the data into a format that is suitable for visualization and aggregating the data to get the insights you need.

Use Tableau Prep Builder to create data flows. Tableau Prep Builder is a data preparation tool that can be used to automate the data cleaning and transformation process.

3.5.3. Visualization Design

Use Tableau Desktop to create interactive dashboards and visualizations that are easy to understand and interpret.

Choose the right data visualizations for your data and your audience.

Use color, shape, and size to effectively communicate your data.

Add annotations and labels to your visualizations to provide additional context.

Test your visualizations with stakeholders to ensure that they are meeting their needs.

3.5.4. Visualization Implementation

Share our visualizations with stakeholders and make sure they are accessible to everyone who needs them.

Automate the process of creating and updating your visualizations so that they are always up-to-date.

Use Tableau Server or Tableau Online to publish your visualizations so that they can be accessed by anyone with an internet connection.

3.6. Data visualization

Framework for Using the Right Data Visualizations of our project:

Understand our data: What are we trying to visualize? What are the key points we want to communicate? What is the audience for our visualizations?

Identify the key questions that we want to answer with our visualizations. What do we want to learn from our data? What insights do we hope to gain?

Identify the key metrics or variables that we will use to answer your questions. What data will we need to collect? How will we measure our metrics?

Consider the audience for our visualizations. Who will be looking at our visualizations? What is their level of understanding of the data? What are their needs and expectations?

Choose the right visualization type: There are many different types of data visualizations, each with its own strengths and weaknesses. Choose the visualization type that is best suited for our data and our audience.

Bar charts: Bar charts are good for showing comparisons between categories.

Line charts: Line charts are good for showing trends over time.

Pie charts: Pie charts are good for showing parts-to-whole relationships.

Scatter plots: Scatter plots are good for showing relationships between two variables.

Heatmaps: Heatmaps are good for showing the distribution of data.

Use effective design principles:

Clarity: Our visualizations should be easy to understand and interpret. Use clear fonts and labels, and avoid clutter.

Conciseness: Our visualizations should be clear and concise. Don't overload our visualizations with too much information.

Effective use of color: Use color effectively to highlight important information and to make our visualizations visually appealing.

Adequate labelling: Label our visualizations appropriately so that our audience can understand what they are looking at.

3. CHAPTER 4: DATA PREPARATION

4.1 Designing the Data

4.1.1 Bus Matrix

Table 3: Bus Matrix

	COMMON DIMENSIONS							
BUSINESS PROCESSES	Date	Products	Customers	Employees	Vendors	Shipping	Location	Promotion
Issues Purchase Orders	x	x		x	x			
Supply management	x	x			x			
Sales Quota management	x			x				
Vendors Evaluation	x	x			x	x	x	
Promotion Tracking	x	x	x					x
Sales Forecasting	x	x	x	x	x	x	x	x

4.1.2 Master Data

For Sales module, we got a list of necessary master data as follows:

- *Product data:* contains information about the products that are sold, such as the product name, subcategory, category, stock quantity and other details for the price of each product.
- *Customer data:* contains information about the customers who purchase the products, such as their name, address, contact information, etc. In some cases, customers may not have done any transactions with the firm.
- *Time data:* contains information about the time dimension, such as the date, month, year, etc. This is a calendar recording of any datetime entities.
- *Promotion data:* represents the specific information of promotion applying to any orders. This kind of data may be useful for marketing evaluation later.
- *Employee data:* persons are responsible for making customer orders. Generally, the company will evaluate the work performance of these salespeople and treat them with appropriate policies.

Similarly, Fact Purchasing also includes Time, Employee, and Product product dimensions. On the other hand, this event table will contain vendors' dimensions to store vendor personal information relating to evaluating and beneficial for making orders decision.

4.1.3 Transaction Data

- *Sales transaction data:* contains information about the individual sales transactions, such as the product sold, the customer who purchased the product, the quantity sold, the price per unit, the date of sale, etc.
- *Purchase transaction data:* contains information about the individual purchase transactions, such as the product purchased, the vendor from whom the product was purchased, the quantity purchased, the price per unit, the date of purchase, etc.

4.1.4 Fact and dimension tables

- DimCustomers

Table 4: DimCustomers

Column name	Datatype	Description	Key
CustomerID	nvarchar(50)	The code of customer	PK
FirstName	nvarchar(50)	First name of the person.	

LastName	nvarchar(50)	Last name of the person.	
EmailAdress	nvarchar(50)	Contact email	
EmailPromotion	int	0 = Contact does not wish to receive e-mail promotions, 1 = Contact does wish to receive e-mail promotions from AdventureWorks, 2 = Contact does wish to receive e-mail promotions from AdventureWorks and selected partners.	
PhoneNumber	int	Telephone number identification number.	
TerritoryID	int	The code of the customer territory	
TerritoryName	nvarchar(50)	Customer territory located	
CountryRegionCode	nvarchar(3)	Country Code	
CountryRegionName	nvarchar(50)	Country name	
TransactionFlag	bit	0 = had not recorded transaction, 1 is reversed.	

- DimEmployees

Table 5: DimEmployees

Column name	**Datatype**	**Description**	**Key**
EmployeeID	int	The code of each employee	PK
EmployeeName	nvarchar(50)	Name of a employee	
NationalIDNumber	nvarchar(15)	National ID number of the employee	
JobTitle	nvarchar(50)	Employee’s job title	
BirthDate	date	Employee’s birth day	

Gender	nvarchar(1)	Employee's gender Rule: M = Male, F = Female	
CurrentFlag	bit	Rule: 0 = Inactive, 1 = Active	
HireDate	date	Employee’s hire date	
SalariedFlag	bit	Employee’s salaried flag Rule: 0 = Yes, 1 = No	
VacationHours	smallint	Employee’s hours spending vacation	
SickLeaveHours	smallint	Employee’s hours sick leave	
CountryRegionCode	nvarchar(3)	Country Code	
CountryRegionName	nvarchar(50)	Country Name	
SalesPersonFlag	bit	Define whether an employee is a salesperson. Rule: 0 = False, 1 = True	

- DimProducts

Table 6: DimProducts

Column name	**Datatype**	**Description**	**Key**
ProductID	int	Natural Key	PK
ProductName	nvarchar(50)	English name of product	
FinishedGoodsFlag	bit	0 = Finished goods is False, 1 = Finished goods is True	
ProductColor	nvarchar(15)	Product’ Color	
SafetyStockLevel	smallint	Minimum inventory quantity.	

ReorderPoint	smallint	Inventory level that triggers a purchase order or work order.	
StandardCost	money	Standard cost of the product.	
ListPrice	money	Selling price.	
ProductSize	nvarchar(5)	Product size	
ProductWeight	decimal(8,2)	Product weight.	
DaysToManufacture	int	Number of days required to manufacture the product.	
ProductLine	nvarchar(2)	R = Road, M = Mountain, T = Touring, S = Standard	
ProductClass	nvarchar(2)	H = High, M = Medium, L = Low	
ProductStyle	nvarchar(2)	W = Womens, M = Mens, U = Universal	
ProductSubcategory	nvarchar(50)	Name of product's subcategory	
ProductCategory	nvarchar(50)	Name of product's category	

- DimShipMethods

Table 7: DimShipMethods

Column name	**Datatype**	**Description**	**Key**
ShipMethodsID	int	The code of the shipping method	PK
Name	nvarchar(50)	English name of product	

ShipBase	money	Minimum shipping charge. Default: 0.00	
ShipRate	money	Shipping charge per pound. Default: 0.00	

- DimVendors

Table 8: DimVendors

Column name	**Datatype**	**Description**	**Key**
VendorID	int	The code of vendor	PK
VendorName	nvarchar(50)	English name of company	
ActiveFlag	bit		
CreditRating	tinyint	1 = Superior, 2 = Excellent, 3 = Above average, 4 = Average, 5 = Below average	
PreferredVendorStatus	bit	0 = Do not use if another vendor is available. 1 = Preferred over other vendors supplying the same product. Default: 1	
CountryRegionCode	nvarchar(3)	Country Code	
CountryRegionName	nvarchar(50)	Country Name	
City	nvarchar(30)		

- DimPromotion

Table 9: DimPromotion

Column name	Datatype	Description	Key
PromotionID	int	Natural Key = SpecialOfferID	PK
Description	nvarchar(255)	English name of product	
DiscountPct	smallmoney	Discount precentage. Default: 0.00	
PromotionType	nvarchar(50)	Discount type category	
Category	nvarchar(50)	Group the discount applies to such as Reseller or Customer	

- DimSalesTerritory

Table 10: DimSalesTerritory

Column name	Datatype	Description	Key
TerritoryID	int	The code of territory	PK
TerritoryName	nvarchar(50)	English name of product	
CountryRegion Code	nvarchar(3)	Country Code	
Group	nvarchar(50)	Geographic area to which the sales territory belong	
SalesYTD	money	Sales in the territory year to date. Default: 0.00	
SalesLastYear	money	Sales in the territory the previous year. Default: 0.00	

CostYTD	money	Business costs in the territory year to date. Default: 0.00	
CostLastYear	money	Business costs in the territory the previous year. Default: 0.00	

- DimDate

Table 11: DimDate

Column name	Datatype	Description	Key
FullDateAlternateKey	date	dd/MM/YYYY	PK
DayNumberOfWeek	tinyint	The day number of day in a week. Ex:1, 2, 3,..., 7	
DayNameOfWeek	nvarchar(10)	The day name of each day in a week. Ex: Monday, Tuesday,...	
DayNumberOfMonth	tinyint	The day number of day in a month. Ex:1, 2, 3,..., 31	
MonthName	nvarchar(10)	The name of each month	
MonthNumberOfYear	tinyint	The month number of a year. Ex: 1,2,3,...,12	
CalendarQuarter	tinyint	Quarter of a year	
CalendarYear	smallint		

- FactSales

Table 12: FactSales

Column name	Datatype	Description	Key
SalesOrderKey	int	Unique key	PK

OrderKey	int	Natural Key	
ProductKey	int	ProductKey	
CustomerKey	int		
SalesPersonKey	int		
TerritoryKey	int		
ShipMethodKey	int		
PromotionKey	int		
DateTimeKey	int		
OrderDate	datetime		
DueDate	datetime		
ShipDate	datetime		
Status	bit	Order current status. 1 = In process; 2 = Approved; 3 = Backordered; 4 = Rejected; 5 = Shipped; 6 = Cancelled Default: 1	
OnlineOrderFlag	bit	0 = Order placed by sales person. 1 = Order placed online by customer. Default: 1	
Freight	money	Shipping cost. Default: 0.00	
TaxAmt	money	Tax amount. Default: 0.00	
OrderQty	smallint	Order quantity of product in one order	

UnitPriceDiscount	money	unit price discount of product in one order	
TotalProductCost	money	Sales subtotal. Computed as SUM(SalesOrderDetail.LineTotal)for the appropriate SalesOrderID. Default: 0.00	
TotalPaid	money	Total due from customer. Computed as Subtotal + TaxAmt + Freight. Computed: isnull(([SubTotal]+[TaxAmt])+[Freight],(0))	

- FactPurchasing

Table 13: FactPurchasing

Column name	**Datatype**	**Description**	**Key**
PurchaseKey	int	Unique key	PK
PurchaseOrderKey	nvarchar(50)	English name of product	
EmployeeKey	int	Employee ID	
ShipMethodKey	int	Ship Method ID	
VendorKey	int	Vendor ID	
ProductKey	int	Product ID	
UnitPrice	money	Vendor's selling price of a single product	
OrderQty	smallint		

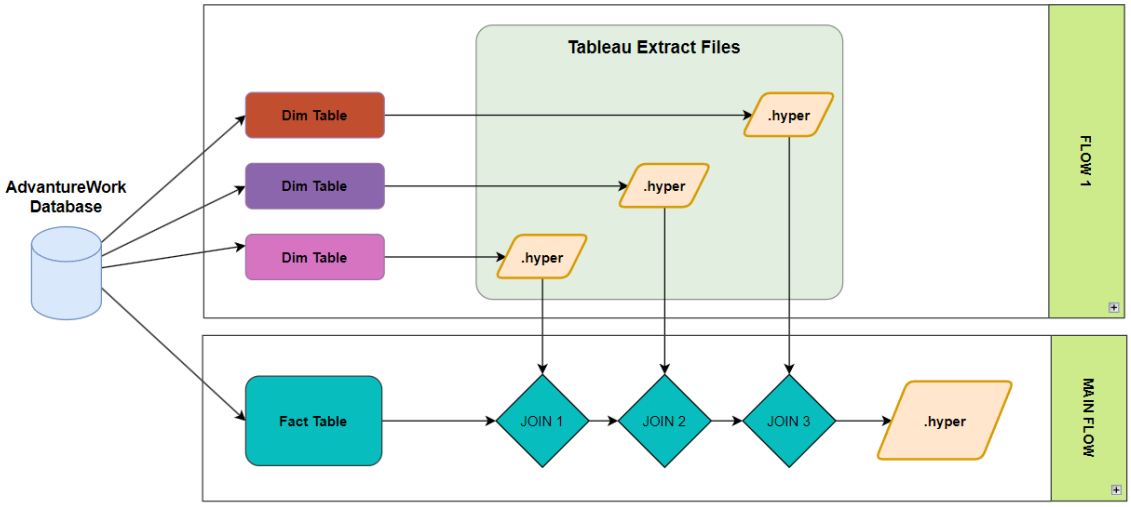
ReceivedQty	decimal(8, 2)	Quantity actually received from the vendor	
RejectedQty	decimal(8, 2)	Quantity rejected during inspection	
AverageLeadTime	int		
LastReceiptDate	datetime	Date the product was last received by the vendor.	
LastReceiptCost	datetime	The selling price when last purchased.	
StandardPrice	money	The vendor's usual selling price	
MinOrderQty	int	The minimum quantity that should be ordered.	
MaxOrderQty	int	The maximum quantity that should be ordered.	
OrderDate	datetime	Dates the sales order was created. Default: getdate()	
ShipDate	datetime	Date the order was shipped to the customer	
Status	tinyint	Order current status. 1 = Pending; 2 = Approved; 3 = Rejected; 4 = Complete Default: 1	
TaxAmt	money	Tax amount. Default: 0.00	
Freight	money	Shipping cost. Default: 0.00	
TotalDue	money	Total due to vendor. Computed as Subtotal +	

		TaxAmt + Freight. Computed: isnull(([SubTotal]+[TaxAmt])+[Freight],(0))	

1. 4.2 ETL processes

We have built this process using Tableau Prep Builder. We decided to choose this platform for ETL for a few reasons. Firstly, because we use Tableau Desktop for data visualization, Tableau Prep Builder becomes a good choice as both platforms share the Tableau ecosystem. Moreover, Tableau is one of the most popular and powerful BI platforms, holding a significant market share in today's market. It also boasts a sizable user community and active discussion forums. Many large businesses as well as individuals opt for Tableau. Specifically, Tableau Prep Builder is a relatively robust data processing tool compared to others such as Microsoft SSIS, Data Factory (Microsoft Azure), etc. Its standout feature is allowing users to manipulate, observe, and analyze data directly with a user-friendly and visually appealing interface. Another reason is that we also use Tableau Server, so using Tableau Prep Builder makes tasks like refreshing and updating data on a daily basis much easier.

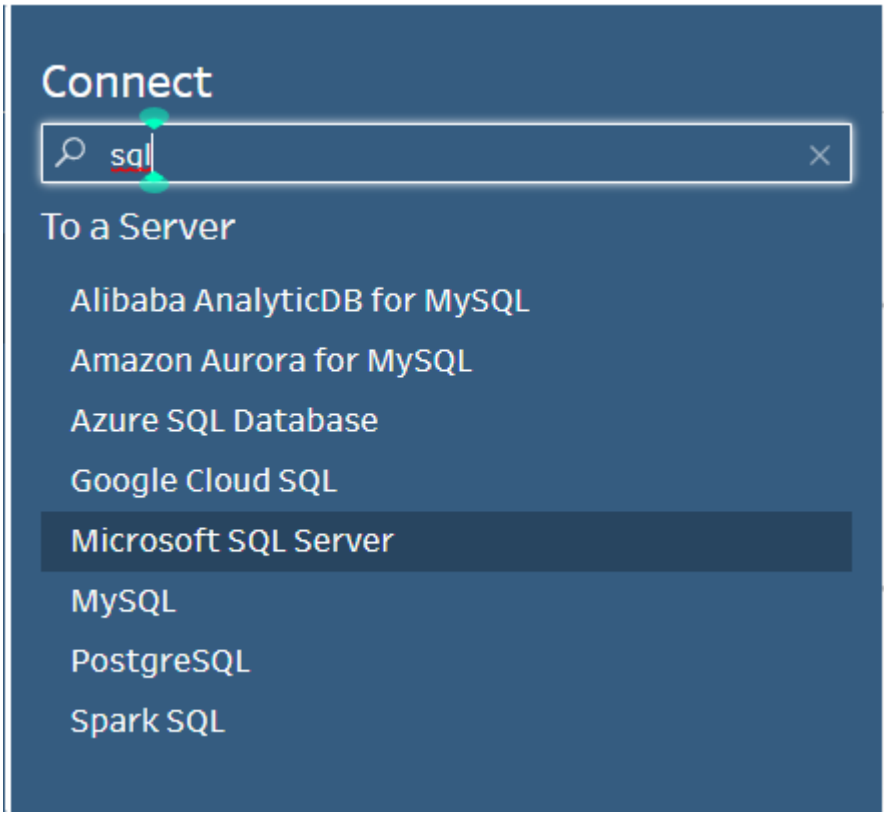
Before commencing the ETL process, we outline a high-level workflow. The ETL process is divided into two smaller flows. Flow 1 involves extracting data from the data source and then exporting it as Tableau Extract files. The Main Flow entails extracting data from the data source into the fact table, followed by joining this data with previously exported Tableau Extract files from Flow 1, using the primary keys of the Dimension tables. Lastly, after completing data modifications, we export the data again as a Tableau Extract file. At this point, the ETL process is complete, and we can utilize Tableau Desktop for data visualization.

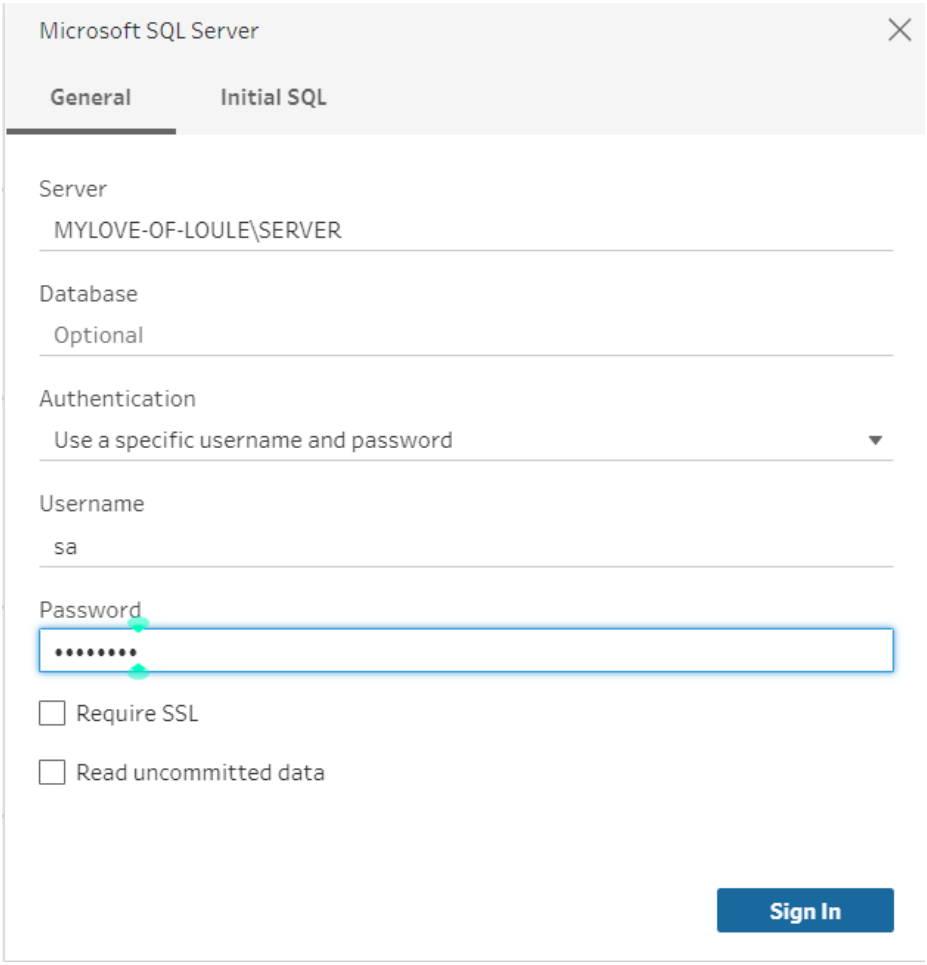


Why do we follow this approach? It's because by doing so, we significantly reduce the time required for directly querying data from the data source when setting up tasks like scheduling automatic data refresh or updates. This method helps optimize the data retrieval process

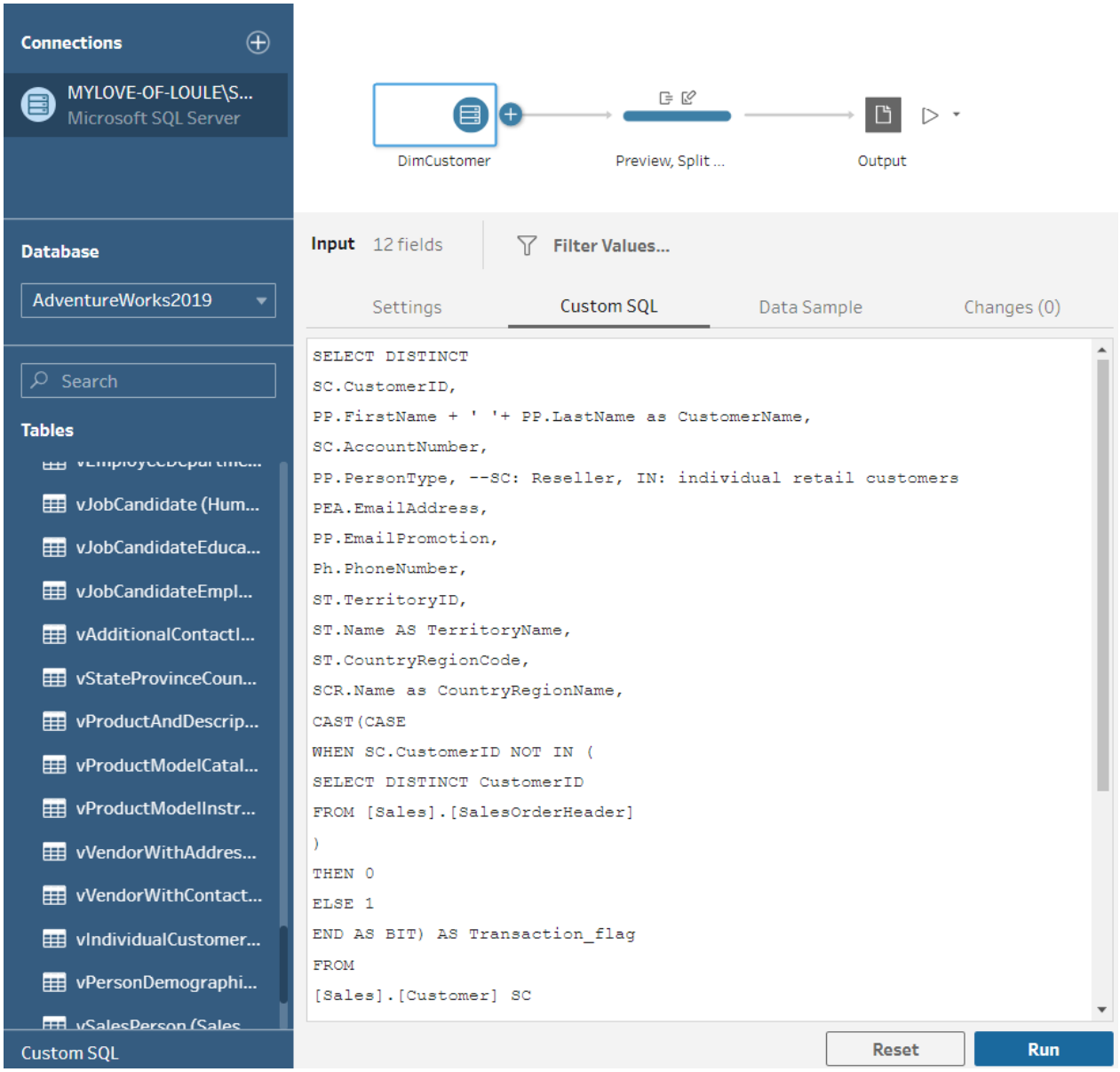
4.2.1 Dimension Table's ETL Process

Firstly, we connect to the data source by clicking on  symbol next to the "**Connections**" label (in this project, the AdventureWorks2019 database downloaded, stored on the desktop, and managed by the Microsoft SQL Server database management system)


Connect
sql
To a Server
Alibaba AnalyticDB for MySQL
Amazon Aurora for MySQL
Azure SQL Database
Google Cloud SQL
Microsoft SQL Server
MySQL
PostgreSQL
Spark SQL



Once successfully connected, we proceed to retrieve data from each table in a separate flow by utilizing SQL query statements through the "**Custom SQL**" option located at the bottom left corner of the screen that show in Figure below.



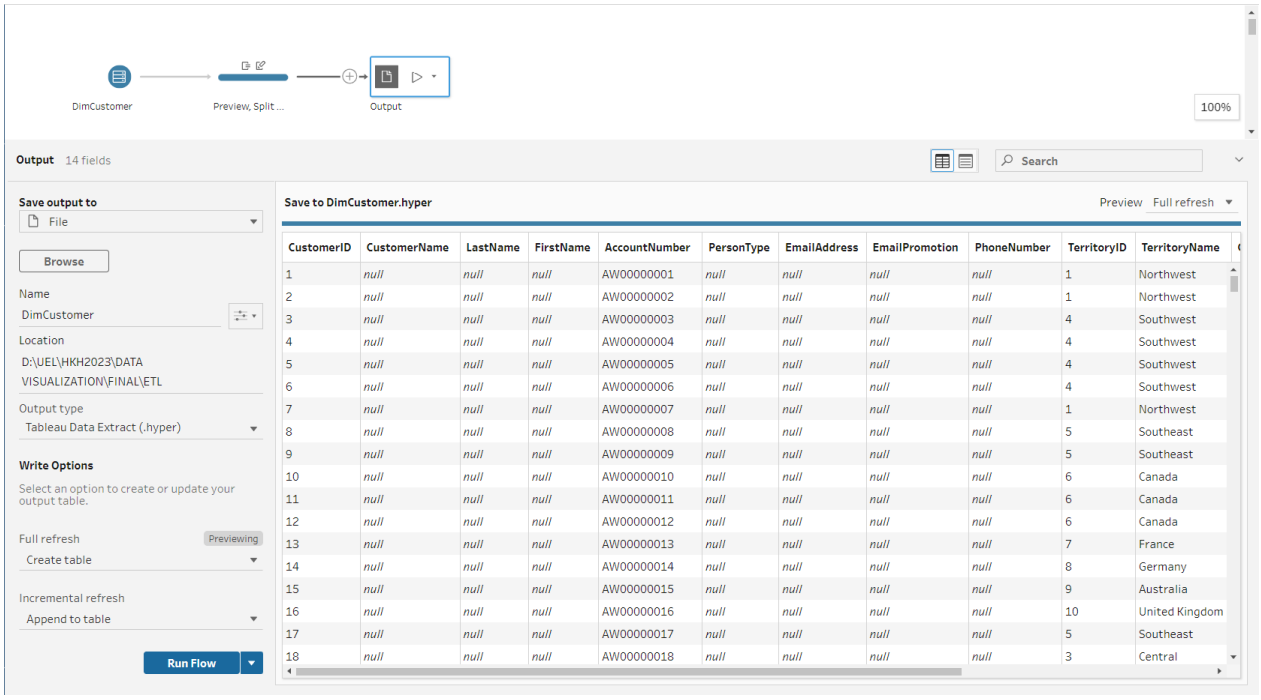
Next, after running the SQL query, you will receive an overview of the data you have just retrieved from the data source.

DimCustomer

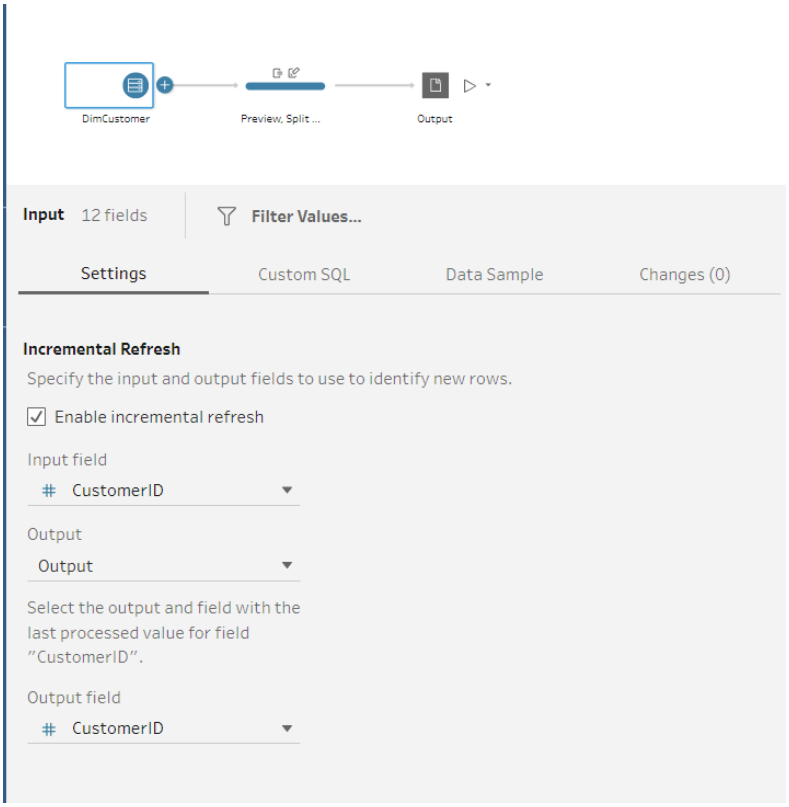
Remove fields you don't need and add filters to limit the data included in your flow. This can improve performance. For more cleaning options or to view your data, add a Clean Step.

Fields included: 12 of 12

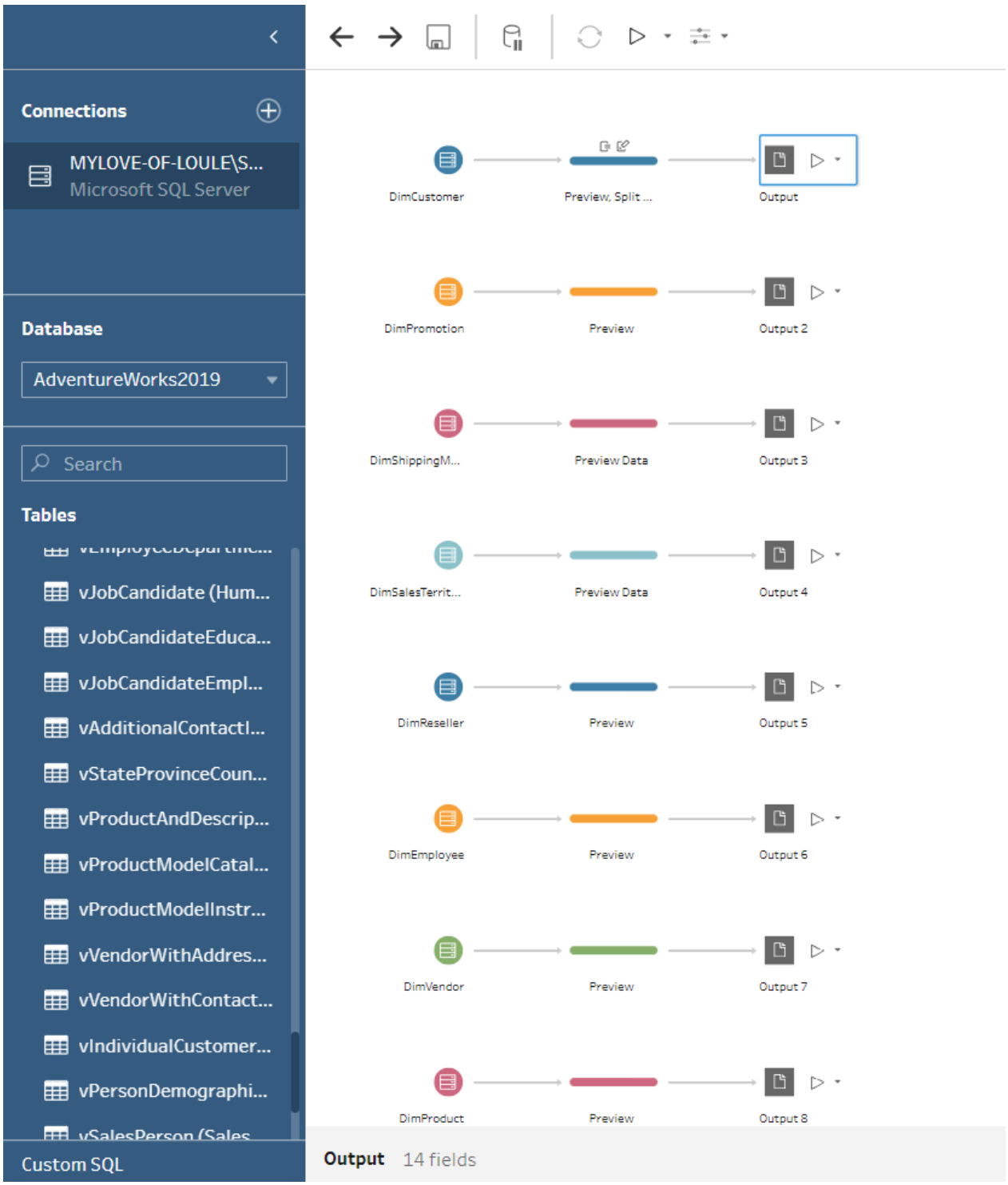
Type	Field Name	Changes	Preview
#	CustomerID		2, 3, 4
Abc	CustomerName		null
Abc	AccountNumber		AW00000002, AW00000003, AW00000004
Abc	PersonType		null
Abc	EmailAddress		null
#	EmailPromotion		null
Abc	PhoneNumber		null
#	TerritoryID		1, 4
Abc	TerritoryName		Northwest, Southwest
Abc	CountryRegionCode		US
Abc	CountryRegionName		United States
T|F	Transaction_flag		False



Once the data is obtained, we create an output component and configure the storage location and format for exporting the data and write options (in this case, I have chosen to export it as a Tableau extract file with the .hyper extension). If you want to add the Incremental Load option, you can select the Incremental Refresh section as an additional step. However, please note that you need to set up incremental load by ticking the **"Incremental Load"** option in **"Settings"** tab and selecting a fixed data field.



Continue following the same process for each remaining dimension table until you have completed the task



After completing the flow, you will receive Tableau Extract files as shown in the figure.

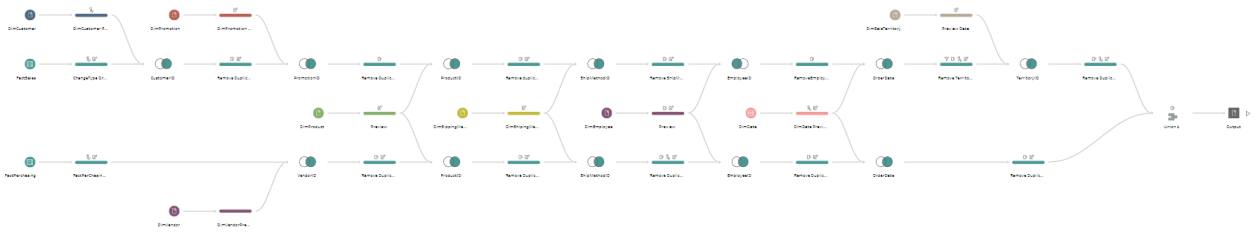
DimCustomer.hyper	8/6/2023 11:10 PM	Tableau Extract	1,280 KB
DimEmployee.hyper	8/6/2023 11:12 PM	Tableau Extract	128 KB
DimProduct.hyper	8/6/2023 11:12 PM	Tableau Extract	64 KB
DimPromotion.hyper	8/6/2023 11:10 PM	Tableau Extract	64 KB
DimReseller.hyper	8/6/2023 11:12 PM	Tableau Extract	128 KB
DimSaleTerritory.hyper	8/6/2023 11:11 PM	Tableau Extract	64 KB
DimShippingMethod.hyper	8/6/2023 11:11 PM	Tableau Extract	64 KB
DimVendor.hyper	8/6/2023 11:12 PM	Tableau Extract	64 KB

Now, we move to the next part.

4.2.2 Fact Table's ETL Process

At this point, we have obtained data from the Extract files (referred to as Dimension tables). We then proceed with the Main flow. Essentially, this process is similar to the previous section, but with some additional elements. Firstly, we continue to extract data into the Fact table using the 'Custom SQL' option. Next, we establish connections between the Dimension tables and join their data with the Fact table using the primary keys of the Dimension tables. This is repeated until all Dimension tables are connected. We then proceed similarly with the remaining Fact tables.

Finally, after making adjustments and data refinements, we consolidate the two Fact tables using the Union component and export the result as a new Extract file.



At this point, the ETL process is complete. Now, we can move to visualize with Tableau Desktop.

CHAPTER 5: DASHBOARD SYSTEMS

5.1. Report and dashboard systems (Report tree)

5.1.1. Sales module

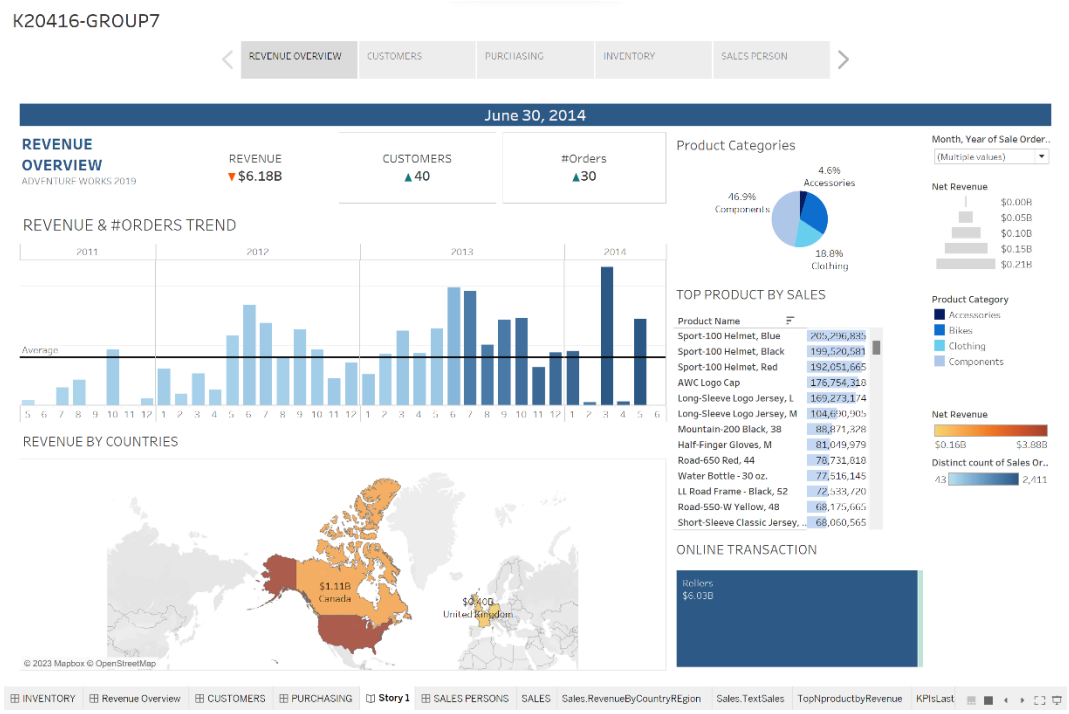


Figure 20. Revenue Overview Dashboard

Overall, The Revenue and #Orders Sales Trend chart shows the monthly revenue and number of orders for Adventure Works from 2011 to 2014. The chart reveals that the revenue and the number of orders have a positive correlation, meaning that they tend to increase or decrease together. The chart also shows that the peak sales season for Adventure Works is from May to July, while the lowest sales season is from November to February. This could be related to the weather and the demand for outdoor activities.

Furthermore, the Day over Day Changes table shows the percentage change in revenue, number of customers, and number of orders for each day compared to the previous day. The table helps to identify the daily fluctuations in sales performance and the factors that might affect them. For example, the table shows that on July 4th, 2013, there was a significant drop in all three metrics, which could be explained by the fact that it was a national holiday in the US and people were less likely to shop online. On the other hand, on August 31st, 2013, there was a significant increase in all three metrics, which could be attributed to the end-of-summer sales and promotions offered by Adventure Works.

The Product Category pie chart shows the percentage of revenue contributed by each product category for Adventure Works. The chart shows that the most profitable product category is Bikes, which accounts for 66% of the total revenue, followed by Components (18%), Clothing (11%), and Accessories (5%). This indicates that Adventure Works has a strong competitive advantage in selling high-quality bicycles and related parts, but it could also explore other opportunities to diversify its product portfolio and increase its market share in other categories.

The Top Product by Sales table shows the top 10 products sold by Adventure Works based on their total revenue. The table shows that the most popular product is Road-150 Red, 62, which is a high-end road bike that costs $3,578. The table also shows that eight out of the top 10 products are bikes, which confirms that Adventure Works is mainly focused on selling bicycles. However, the table also shows that two non-bike products made it to the top 10 list: HL Mountain Pedal and HL Road Pedal, which are both components that cost $80 each. This suggests that Adventure Works could also leverage its expertise in selling bike components and accessories to increase its sales and customer loyalty.

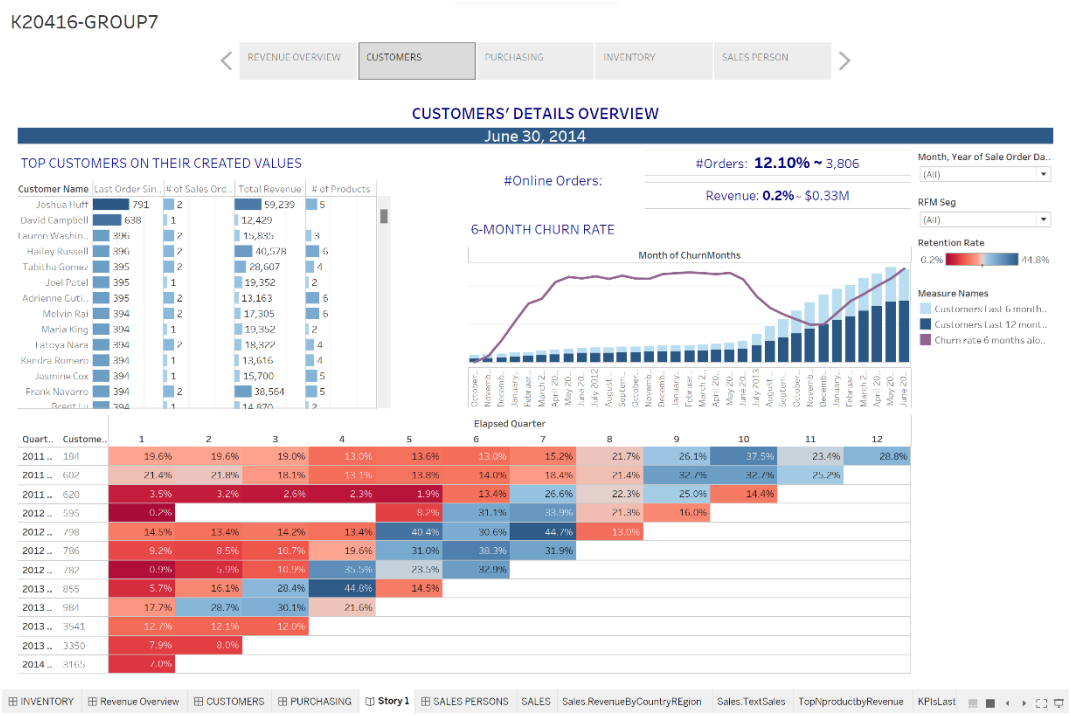


Figure. Customer's Details Overview

The Cohort Analysis chart shows the percentage of customers who returned to make another purchase within a given time period. The chart is divided into cohorts based on the month of their first purchase, and the columns show the percentage of customers who made a second, third, fourth, etc. purchase in the subsequent months.

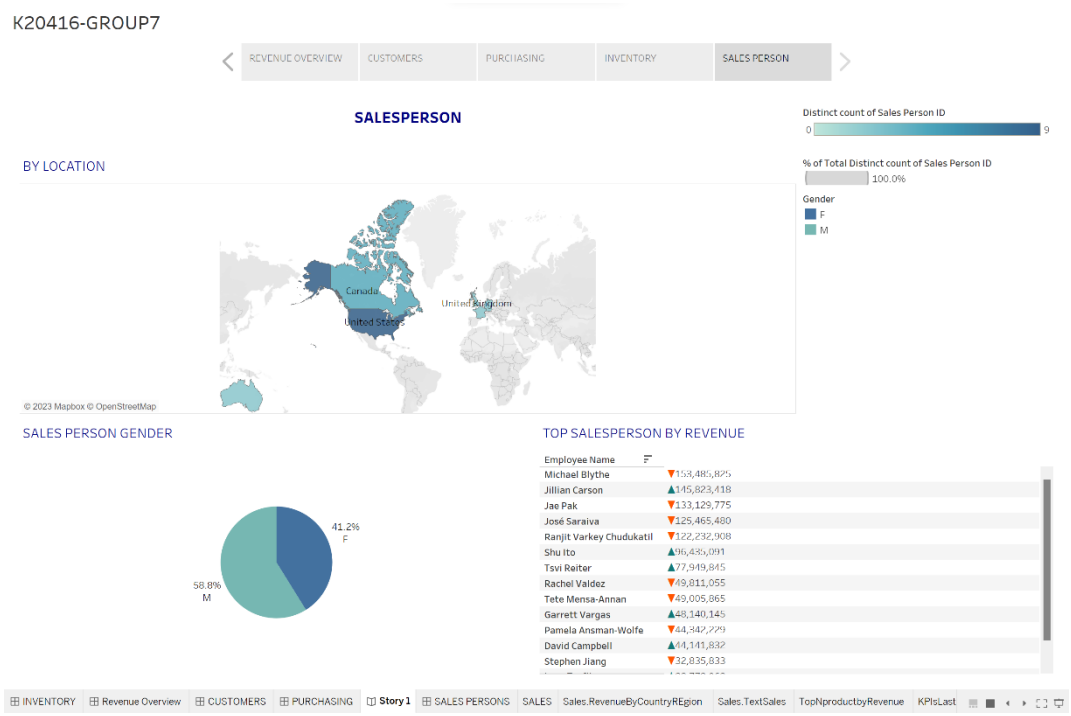
The chart helps to measure the customer retention rate and the customer lifetime value of your online store. You can use this information to evaluate the effectiveness of your customer acquisition and retention efforts, and identify the factors that influence customer loyalty. For example, you can compare the retention rates of different cohorts and see how they are affected by seasonality, promotions, product quality, or customer service.

The 6-Month Churn Rate chart shows the percentage of customers who stopped making purchases after six months from their first purchase. The chart combines a line chart and a bar chart to show the trend and the distribution of the churn rate over time. The chart helps to monitor the customer attrition rate and the customer churn risk of your online store. You can use this information to prevent customer churn and increase customer loyalty. For example, you can identify the customers who are at risk of churning based on their purchase history, behavior, or feedback, and reach out to them with targeted offers, incentives, or reminders to encourage them to come back and make another purchase.

The Online Orders and Revenue KPIs show the number of online orders and the total revenue generated by your online store in the last 30 days. The KPIs also show the percentage change compared to the previous 30 days. The KPIs help to track the sales performance and growth of your online store. You can use this information to set goals and benchmarks for your online store, and adjust your sales and marketing strategies accordingly. For example, you can analyze the factors that affect your online orders and revenue, such as traffic sources, conversion rates, average order value, or product mix.

Customer RFM Segmentation.

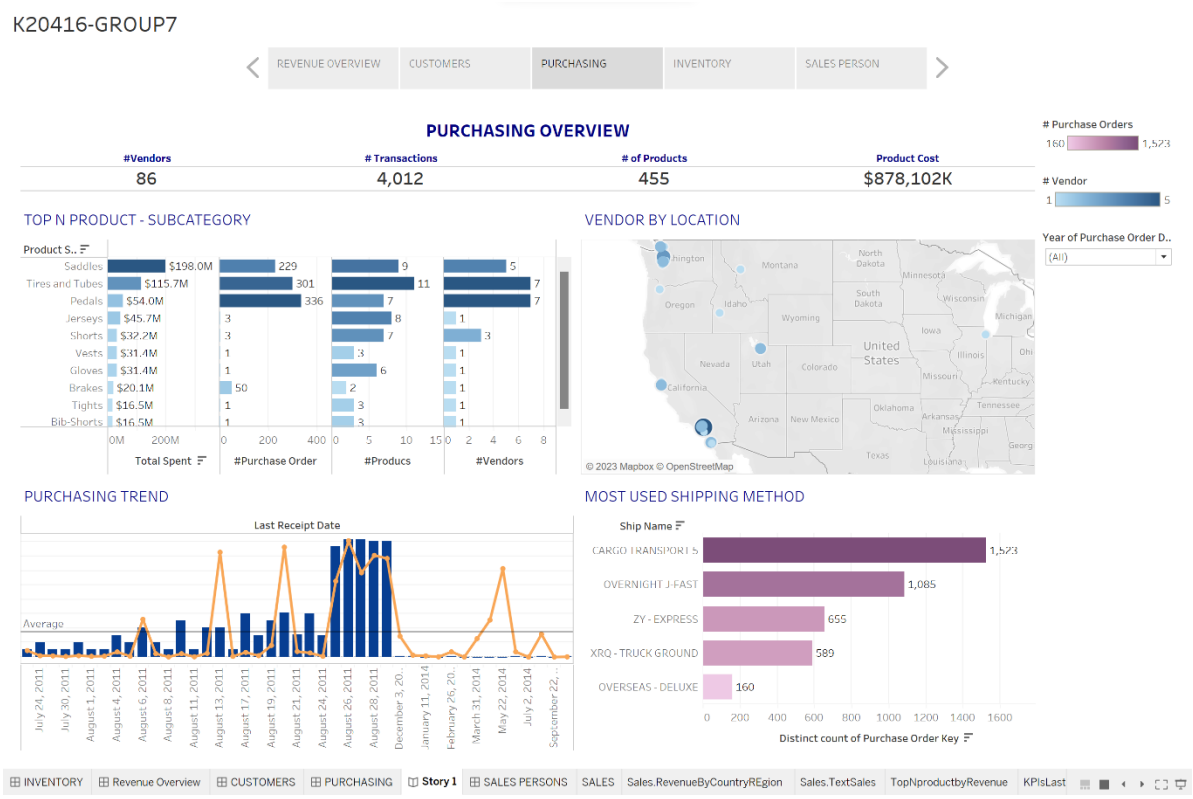
As mentioned, we have used RFM model to segment customers following four quartiles. Then, the results include 7 segments, who are occasional buyers, Big Spenders, Loyal Customer, Best Customer, Lost Cheap Customer and Potential to become Best.



Sales person overview

For sales staff, most of the staff can be found residing in the US and a few others in neighboring countries. Of the 17 salespeople, Michael Blythe was the best performer on the last day of data, with sales of more than $153 million but lower than the day before.

5.1.2. Purchase module



Purchasing overview

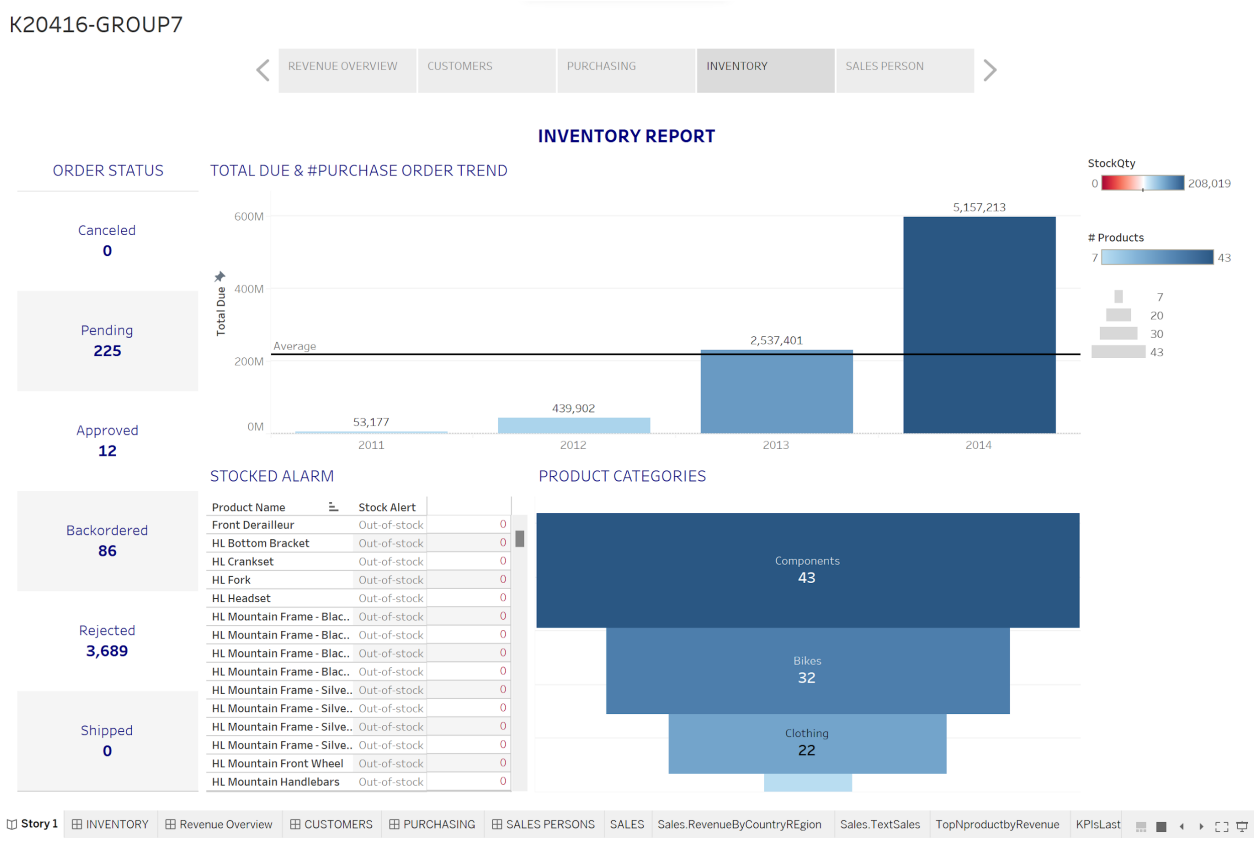
At first, the Vendors, Transactions, Products, and Total Spent KPIs show the number and amount of purchases made by your company in the last year. The KPIs also show the percentage change compared to the previous year. The KPIs help to monitor the purchasing volume and cost of your company, and evaluate the efficiency and profitability of your purchasing decisions. You can use this information to optimize your purchasing budget and strategy, and negotiate better deals with your vendors.

The Top Product-Subcategory by Total Spent table shows the details of the top 10 product-subcategories purchased by your company based on their total spent, purchase order, and number of vendors. The table helps to identify the most expensive and frequent purchases made by your company, and the diversity and reliability of your vendors. You can use this information to prioritize your purchasing needs and preferences, and diversify or consolidate your vendor relationships.

The Purchasing Trend chart shows the monthly trend of the total spent and the number of transactions for your company's purchases. The chart combines a line chart and a bar chart to show the correlation and variation of the two metrics over time. The chart helps to analyze the seasonality and volatility of your purchasing behavior and expenses, and identify the factors that influence them. For example, you can see how your purchases are affected by demand, supply, price, quality, or availability of products.

The Most Used Shipping Method chart shows the percentage of purchases made by your company using different shipping methods, such as standard, express, or overnight. The chart helps to compare the cost and speed of different shipping options, and evaluate their suitability and convenience for your company's needs. You can use this information to choose the best shipping method for your purchases, and balance between saving money and time.

The Vendors Location map shows the geographic distribution of your company's vendors by country. The map helps to visualize the global reach and diversity of your company's supply chain, and assess the risks and opportunities of sourcing products from different regions. You can use this information to expand or reduce your vendor network, and explore new or existing markets for your products.



Inventory overview

The Number of Orders by Status chart shows the distribution of orders placed by your customers based on their current status, such as pending, canceled, approved, backordered, rejected, or shipped. The chart helps to track the progress and performance of your order fulfillment process, and identify any issues or delays that might affect your customer satisfaction and revenue. You can use this information to optimize your order management system, and improve your communication and coordination with your customers and suppliers.

The Stocked Alarm table shows the details of the products that have low or high stock levels in your inventory. The table includes the product name, category, subcategory, current stock, minimum stock, maximum stock, and stocked alarm. The stocked alarm is a color-coded indicator that shows whether the product is understocked (red), overstocked (yellow), or optimally stocked (green). The table helps to alert you of any potential stock shortages or surpluses that might affect your sales and costs. You can use this information to adjust your inventory levels and reorder points, and balance between meeting customer demand and minimizing inventory holding costs.

The Product Category Funnel chart shows the percentage of products in each category that are sold out, low in stock, or in stock in your inventory. The chart helps to compare the sales and inventory performance of different product categories, and evaluate their profitability and popularity. You can use this information to prioritize your inventory replenishment and allocation, and align your product mix with customer preferences and market trends.

The Total Due and Purchase Order Trend chart shows the yearly trend of the total amount due and the number of purchase orders made by your company to replenish your inventory. The chart combines a line chart and a bar chart to show the correlation and variation of the two metrics over time. The chart helps to analyze the cash flow and purchasing behavior of your company, and identify the factors that influence them. For example, you can see how your purchases are affected by seasonality, demand, supply, price, quality, or availability of products.

5.3. Evaluation and Discussion

1. 5.3.1. Evaluation

The evaluation of the data visualization project reveals its effectiveness in achieving its intended goals and offers insights into its performance and value to AdventureWorks. This section provides a comprehensive assessment of the project's outcomes and its alignment with the initial objectives.

The data visualization project has successfully met its primary objectives of providing actionable insights into the Sales and Purchase modules of AdventureWorks. The interactive dashboards have enabled users to gain a deeper understanding of sales trends, customer behaviors, and vendor performance. The integration of key performance indicators (KPIs) within the visualizations has facilitated efficient monitoring of critical business metrics, aiding decision-making processes.

The project's strengths lie in its user-friendly interface, which allows users across different departments to explore and analyze data effortlessly. The intuitive design of the dashboards promotes widespread adoption and utilization. By transforming complex datasets into visually engaging representations, the project has enhanced the accessibility of data-driven insights to a broader audience within the organization.

Moreover, the data visualization project's impact on informed decision-making is evident through the identification of growth opportunities and operational improvements. Real-time access to visualized data has expedited the identification of sales patterns and customer preferences, enabling timely adjustments to strategies. This, in turn, has contributed to increased sales efficiency, optimized resource allocation, and heightened customer satisfaction.

2. 5.3.2 Discussion

The discussion surrounding the data visualization project delves into its implications, significance, and potential for future enhancements. It highlights both the immediate benefits realized and the broader organizational changes prompted by the project's implementation.

The data visualization project's impact on AdventureWorks has been profound. By transforming raw data into meaningful insights, the project has elevated data-driven decision-making to new heights. The ability to visualize sales and purchase trends has enabled the company to make agile and well-informed choices, resulting in improved operational efficiency and revenue growth.

Beyond its immediate outcomes, the project has cultivated a culture of data-driven thinking within AdventureWorks. Employees from diverse departments are now more inclined to engage with data and derive insights from it. This cultural shift has led to enhanced cross-functional collaboration, as teams leverage the visualizations to align their efforts and strategies based on shared insights.

Looking ahead, there are opportunities for the project's expansion and refinement. Incorporating additional modules, such as Inventory Management or Marketing Analytics, could provide a comprehensive view of the company's operations. Furthermore, continuous data quality management and regular updates to the visualizations will ensure that the insights remain accurate, relevant, and reliable.

CHAPTER 6: CONCLUSION AND FUTURE WORKS

6.1. Results

The data visualization project utilizing Tableau has yielded insightful outcomes that contribute significantly to AdventureWorks' decision-making processes and operational efficiency. Through the implementation of interactive dashboards and visualizations, key performance indicators (KPIs) related to the Sales and Purchase modules have been effectively presented and analyzed. The dashboards provide a comprehensive overview of the company's sales revenue, customer segmentation, purchasing trends, and vendor performance. With these visualizations, AdventureWorks is now equipped to identify customer preferences, monitor sales territories, and track purchasing patterns in real-time. Additionally, the project has enabled the identification of high-performing products and vendors, as well as areas for potential cost-saving and process improvement. Overall, the results showcase the power of data visualization in transforming raw data into actionable insights for driving business growth.

6.2. Limitations

While the data visualization project has provided valuable insights, it is essential to acknowledge certain limitations. One of the primary limitations is the quality and completeness of the underlying data. The accuracy of the visualizations heavily relies on the accuracy and consistency of the data sources. Incomplete or inaccurate data could lead to misleading conclusions and insights. Moreover, the project's scope is focused on the Sales and Purchase modules, and thus, insights into other aspects of the business may be limited. Additionally, the visualizations are based on historical data, and external factors or unforeseen events that occurred after the data collection period are not accounted for in the analysis. Lastly, the project's success is contingent upon user engagement and adoption of the dashboards; any lack of user interaction could impact the effectiveness of the insights presented.

6.3. Future works

To build upon the current achievements and address the identified limitations, several avenues for future works are recommended. Firstly, enhancing data quality through continuous data cleansing and validation processes will contribute to more accurate and reliable visualizations. Incorporating real-time data feeds and automating data extraction processes will ensure that insights remain up-to-date and relevant.

Expanding the scope of the project to include additional business modules, such as Inventory Management or Customer Service, will provide a more comprehensive view of the company's operations. Integration with advanced analytics and machine learning algorithms can enable predictive insights, aiding in forecasting sales trends and demand fluctuations. Furthermore, fostering user engagement through training and workshops will encourage wider adoption of the dashboards and increase their impact on decision-making. Finally, exploring opportunities to decentralize data storage while ensuring data security and compliance will enhance data accessibility and utilization. In this way, the data visualization project can continue to evolve, delivering deeper insights and greater value to AdventureWorks' business operations.

REFERENCES

[1] Microsoft. Adventure Works Cycles Business Scenarios. Retrieved 4/5/2023 from: https://learn.microsoft.com/en-us/previous-versions/sql/sql-server 2008/ms124825(v=sql.100)

[2] Gartner. Analytics and Business Intelligence (ABI). Retrieved 4/5/2023 from: https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi

[3] Microsoft. What is business intelligence?. Retrieved 4/5/2023 from:https://powerbi.microsoft.com/en-cy/what-is-business-intelligence/

[4] Techtarget. business intelligence (BI). Retrieved 4/5/2023 from: https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-BI

[5] Wu, L., Barash, G., & Bartolini, C. (2007, June). A service-oriented architecture for business intelligence. In IEEE international conference on service-oriented computing and applications (SOCA'07) (pp. 279-285). IEEE.

[6] Ong, I. L., Siew, P. H., & Wong, S. F. (2011). A five-layered business intelligence architecture. Communications of the IBIMA, 2011, 1-11.

[7] IBM. What is ETL?. Retrieved 5/5/2023 from: https://www.ibm.com/topics/etl

APPENDIX

PROJECT MEMBER EVALUATION (Signed by all members)