VIETNAM NATIONAL UNIVERSITY - HCMC

UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION SYSTEMS

BI SOLUTIONS FOR PRODUCTION PROCESS OF ADVENTURE WORKS COMPANY

Course: Business Intelligence and Decision Support System

Group 6

Lecturer: M.A Nguyen Van Ho

M.A Nguyen Van Tuyen

B.A Nguyen Anh Nhat

Class: K19406

Ho Chi Minh City, 24th May 2022

ACKNOWLEDGEMENTS

First of all, our team would like to thank the Faculty of Information Systems for facilitating the opening of the course “Business Intelligence and Decision Support System” to help us achieve brand new knowledge.

We would also like to thank Mr. Nguyen Van Ho, the lecturer of this course who is always enthusiastic about guiding and helping us a lot in the process of implementing the subject project.

We also do not forget to appreciate Mr. Nguyen Van Tuyen's dedication in supporting the team with practical knowledge about SQL, Data Warehouse Building, MDX, KPIs ... This way we can complete our report in the best way possible.

Finally, we sincerely thank Mr. Ho Trung Thanh for the knowledge related to the BI course that he has imparted, Mr. Nguyen Anh Nhat has shared very well about his work experience as well as professional knowledge in the field of data. All this support helps our team a lot in completing the course and having for ourselves useful knowledge.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Acronym	Description
AW/ ADW	Adventure Works
BI	Business Intelligence
ETL	Extract, Transform, and Load
IT	Information Technology
KPI	Key Performance Indicator
MDX	Multidimensional Expressions
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
R&D	Research and development
ROI	Return on Investment
SQL	Structured Query Language
SSAS	SQL Server Analysis Services
SSIS	SQL Server Integration Services

CHAPTER 1: INTRODUCTION

1.1. Business case for the project

Nowadays data and data analytics are increasingly becoming an important part of companies' business. When companies expand their market, the amount of data also increases as well. Moreover, analyzing and processing that huge amount of data is extremely necessary in supporting managers to make important decisions.

That data can provide managers with in-depth insights into sales trends, customer behavior, business revenue, product models and resource allocation, among other things. Managers can then make more effective decisions and plan for future business growth, maximizing revenue and lowering costs.

The technologies, applications, strategies, and practices used to collect, analyze, integrate, and present relevant business information are referred to as Business Intelligence (BI) solutions. By extracting and analyzing data from raw data, BI solutions can aid in the production of accurate reports.

BI exists to support and facilitate better business decisions. BI provides organizations with access to information that is critical to the success of a variety of departments and areas, including sales, finance, and marketing. Effectively leveraging BI will provide your company with more actionable data, great insights into industry trends, and a more strategically geared decision-making model.

In this project, we will build a BI solution through some following stages: Production Process Analysis, Data Source building phase, ETL processes and Project Reports to better understand the effects that BI brings to the business.

1.2. Objectives of the project

1.2.1. General Objective

Learning how to transform raw data into structured data that can be analyzed and visualized as dashboards to produce more visually appealing reports. Understanding more about the AdventureWorks company, particularly the production department, and using SQL Server to create a Data Warehouse then store them for analysis, making reports and decision support.

1.2.2 Specific Objectives

- Investigating data analysis methods and basic theories.
- How Business Intelligence helps at each stage of the business process.
- Designing and constructing a data warehouse with Microsoft SQL Server.
- How companies can make decisions based on data analysis.
- Using ETL processes to figure out the financial reports.

1.3. Research Objects

1.3.1 AdventureWorks's business issues

The Adventure Works database contains a large number of tables that are similar to those found in a bicycle shop. The examples will use various tables from the database to highlight BI reporting from the Adventure Works Company's various segments. Human Resources, Manufacturing, Sales, Purchasing, and Contact Management are among these segments. In each chapter, a different aspect of the data will be highlighted using a specific set of tools.

Adventure Works sells not only bicycles, but also accessories, clothing, and components. Bottles, bike racks, brakes, and other accessories are available. Clothing options include caps, gloves, jerseys, and so on. Adventure Works sells brakes, chains,

derailleurs, and other components. Many of those items are manufactured by vendors, so AdventureWorks acts as a reseller.

Adventure Works employs a total of 290 people in various functions such as sales, production, purchasing, engineering, finance, information services, marketing, shipping and receiving, and R&D. Adventure Works' customers include over 700 stores and over 19000 people worldwide, and its vendors number around 100 companies that supply raw materials, accessories, clothing, and components.

Despite the fact that AdventureWorks is a fictional company, it is designed to be a realistic case similar to a real company in the industry. Adventure Works offers a database and data warehouse that covers sales, material management, production, finance, and human capital management. As a result, the researcher uses this fictitious company as a case study to create a Self-service BI system.

1.3.2 Limits of the data

There is little information about the AdventureWorks data production process on the internet. While the manufacturing process generates a large amount of data, it has many tables, making it more difficult to redesign and build a Data Warehouse to manage and analyze the data.

1.4. Scope of the project

1.4.1 Data Available

- Data from Production Department of AdventureWorks
- AdventureWorks2014.bak

1.4.2 Tool

- Microsoft SQL Server Management Studio
- Visual Studio (SSIS, SSAS)

- Excel, Word
- Power BI
- MDX

1.5 Value and desired outcome of the project

- We look forward to building a BI solution to help increase revenue per product by 10%.
- We hope this BI solution project will help AdventureWorks reduce costs in its production and business activities by 20%.

1.6 Structure of project

We divide the project structure into 7 chapters:

- CHAPTER 1: INTRODUCTION

Chapter 1 focuses mainly on giving the reasons for the implementation of the project, the scope as well as the brief introduction of the BI solution.

- CHAPTER 2: THEORETICAL BASIS

Chapter 2 is about the basic foundations based on the theory related to BI, ETL, OLAP techniques, KPIs ... to get an overview of related concepts.

- CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION

In this chapter, we introduce the business process related to the production department and also figure out the requirements for data analysis.

- CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

We design a data warehouse model and then conduct the ETL process to pour data from the data source into the data warehouse.

- CHAPTER 5: DATA ANALYTICS

Chapter 5 shows how we conduct data analytics with SSAS Technology, how we use MDX to create KPIs and how we learn about the OLAP techniques.

- CHAPTER 6: VISUALIZATION AND FORECASTING OR PREDICTIVE MODEL

We visualize data using Power BI to have further insights into the production department.

- CHAPTER 7: CONCLUSION AND FUTURE WORKS

Finally, Chapter 7 summarizes what we have done, the limitations that need to be overcome as well as future development work for the project.

CHAPTER 2: THEORETICAL BASIS

2.1 BI theory

2.1.1 Concept of BI

BI stands for Business Intelligence, which is a collection of processes, architectures, and technologies that transform raw data into useful information that drives profitable business decisions.

To provide users with detailed intelligence about the state of the business, BI tools access and analyze data sets and show analytical findings in reports, summaries, dashboards, graphs, charts, and maps. The phrase "Business Intelligence" is also used to describe a set of technologies that deliver quick, easy-to-understand information about a company's current state, based on available data.

2.1.1.1 The importance of BI

"A real-time enterprise without real-time business intelligence is a real fast, dumb organization" said Stephen Bobst, CTO Teradata. We can see how important a BI is according to the opinion above. Let's take a look at six of the most important:

- BI turns data into actionable information. A business intelligence (BI) system is an analytical tool that can provide you with the information you need to make successful strategic plans for your company. This is because such a system would be able to recognize key trends and patterns in your organization's data, making it easier for you to make important connections between seemingly unrelated areas of your business. As a result, a business intelligence system can help you better understand the implications of various organizational processes and improve your ability to identify appropriate opportunities for your company, allowing you to plan for the future.

- BI is measurements which creates KPI (Key Performance Indicators) based on historic data.
- BI help us get a better understanding of customer behavior. Once we know what our customers want, we can use that knowledge to create products that are in line with current trends and, as a result, increase our profits.
- BI helps on data visualization, which improves data quality and, as a result, decision-making quality.
- The use of business intelligence (BI) boosts productivity. BI accelerates and streamlines the process of analyzing and interpreting data, allowing us to understand business data as soon as it arrives and generate reports with a single mouse click. This frees up more time for us and our employees to focus on running our company rather than analyzing it.
- BI will increase the visibility of various critical processes and allow for the identification of any areas that require improvement.

2.1.1.2. The process of developing a BI solution

Here are 6 steps:

- Step 1: Understanding business need (collect and analyze requirements)
- Step 2: Learn, analyze source data and the current state of the system
- Step 3: Data modeling (Data Warehouse) and ETL. Build a Data Warehouse model with the required information from the report with an appropriate structure for data retrieval and analysis. Choosing ETL tools and transfer data from data sources into Data Mart
- Step 4: Building multi dimensional cubes and visualizing
- Step 5: Testing and Quality controlling
- Step 6: Go LIVE, deployment, operation and maintenance.

2.2 ETL Theory and practical usage:

2.2.1 Concept of ETL

ETL is the acronym for extract, transform and load, a data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system to perform needed analysis. For our project, SSIS (SQL Server Integration Services) tools are used to implement ETL process.

2.2.2 Practical usage of ETL

ETL tools break down data silos and make it easy for our group to access and analyze data, and turn it into business intelligence by providing clean and filtered data structures.

ETL enables us to gather data from multiple sources and across platforms into one centralized data warehouse. Therefore, we can enhance our analysis, visualization throughout large data sets. And as a result, our ability to generate reports and metrics will be improved for a better strategy.

2.2.3 ETL Process

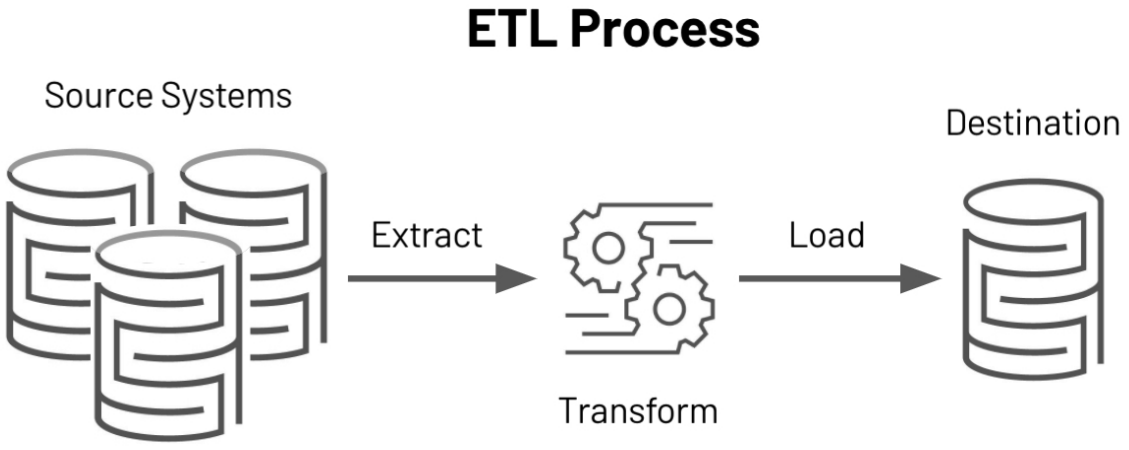


Figure 1. ETL Process

As mentioned above, ETL process contains three following steps:

First is Extract. During data extraction, raw data is copied or exported from source locations to a staging area. Data management teams can extract data from a variety of data sources, which can be structured or unstructured.

The next step is Transform, the most vital step in the ETL process. During this phase, data is transformed and consolidated for its intended analytical use case. The follow-up tasks may involve in this phase can be:

- Filtering, cleansing, de-duplicating, validating, and authenticating the data
- Performing calculations, translations, or summarizations based on the raw data.
- Conducting audits to ensure data quality and compliance
- Removing, encrypting, or protecting data governed by industry or governmental regulators
- Formatting the data into tables or joined tables to match the schema of the target data warehouse.

Last but not least, Loading. In this last step, the transformed data is moved from the staging area into a target data warehouse.

2.3 KPI Theory and practical usage

2.3.1 What is KPI?

KPI stands for key performance indicator, a quantifiable measure of performance over time for a specific objective. KPIs provide targets for teams to shoot for, milestones to gauge progress, and insights that help people across the organization make better decisions

2.3.2 The importance of KPIs

KPIs are important to business objectives because they keep objectives at the forefront of decision making.

It’s essential that business objectives are well communicated across an organization, so when people know and are responsible for their own KPIs, it ensures that the business's overarching goals are top of mind.

KPIs also ensure that performance is measured not blindly in pursuit of the KPI but in relation to the larger business objectives. This means that every part of work is done with intentionality and for the right purpose.

2.3.3 Process of creating KPIs

Developing meaningful KPIs that track, and clearly visualize performance takes some planning. Each KPI needs to address a specific business objective, and provide timely, accurate information to assess progress towards goals.

Here's how to create a KPI:

1. Establish a clear objective

If a goal of our team is to develop a BI solution for A.W company, then a KPI objective maybe to “analyze product data and support manager decision-making for the company's upcoming activities”. State clearly, and in simple terms the purpose of the KPI. This provides guidance for anyone viewing the KPI to interpret the data in the correct context.

2. Outline the criteria for success

What will the target be? Is it attainable? when should it be accomplished? and how will progress be monitored? Targets should be realistic, changes to business processes take time to implement. In the initial stages of KPI monitoring it's best to focus on long-term targets with midterm monitoring.

3. Collect the data

Investigate the availability and accuracy of the data. Data may be available automatically from existing systems or hidden in reports and databases. This data will all need to be pulled together at regular intervals for reporting in one central place.

4. Build the KPI formula

Some KPIs contain but a single metric or measure. However most rely on a combination brought together under a single calculated formula. For example, a KPI that measures productivity in revenue would look like this: Total Revenue divided by the total number of products. Build formulas and create calculations with test data to see if the results are what our team would expect.

5. Present KPIs

To efficiently communicate KPIs we'll need to translate the data into understandable visuals such as graphs and charts. Dashboards for Operational KPIs, or Reports for Strategic KPIs offer a convenient way to create, track and distribute our KPIs.

2.4 MDX language for analyzing multidimensional data and OLAP

2.4.1 What is MDX language?

MDX (Multidimensional Expressions) is a calculation/query language used in database management systems to define queries for online analytical processing (OLAP). The SQL extension multidimensional expression allows you to query data stored in a multidimensional structure. Multidimensional databases are OLAP cubes that are used for reporting and analysis.

In an OLAP application, the MDX query basics help you interact with dimensions, hierarchies, members, and other things. MDX can be used to query data stored in a SQL Server database. As a result, a dataset containing cell and axis data is returned by an expression.

2.4.2 OLAP technique

OLAP stands for Online Analytical Processing, and it is a type of software that allows users to evaluate data from various databases at the same time. It's a tool that allows analysts to obtain and analyze business data from various perspectives.

Grouping, aggregating, and joining data is a common task for analysts. These data mining OLAP processes are resource expensive. Data can be pre-calculated and pre-aggregated using OLAP, which speeds up analysis.

Most corporate data has numerous dimensions, or subcategories, into which it is divided for presentation, tracking, or analysis. Sales numbers, for example, could have multiple dimensions relating to geography (region, nation, state/province, store), time (year, month, week, day), product (clothing, men/women/children, brand, kind), and more.

However, data sets are kept in tables in a data warehouse, and each table can only organize data into two of these dimensions at a time. OLAP collects data from several relational data sets and reorganizes it into a multidimensional structure that allows for extremely quick processing and analysis.

An OLAP Cube is at the heart of the OLAP idea. The OLAP cube is a data structure that is designed to allow for rapid data analysis. One or more cubes are used to separate OLAP databases. The cubes are constructed in such a way that it is simple to create and read reports.

2.4.3 MDX method and structure

2.4.3.1 MDX method

It’s possible to exploit MDX to query data maintained in an SQL Server. As a result, an expression returns a dataset containing cell and axis data. The statement SELECT is the

most common MDX query example. The statement selects a result set to return, specifies its syntax, and constructs the MDX query.

Multidimensional Expressions interact with and return numerous forms of OLAP data through MDX queries:

- Scalar is either a number or a string.
- Dimension is a dimension in a cube and it's the initial part to arrange measure and attribute. Dimensions are organized independently. They include members arranged in hierarchies whilst hierarchies contain levels.
- Hierarchy is primarily a dimension MDX hierarchy and is organized inside dimensions and can't be determined by unique names.
- Level belongs to a dimension hierarchy and in its turn is organized in a dimension hierarchy.
- Members belong to a dimension hierarchy and are arranged in an OLAP dimension hierarchy. They can be determined by unique names like WITH member in MDX query. All levels are specified depending on a hierarchy hence there can be 2 various members to coordinate in sets if one member belongs to 2 various hierarchies.
- Tuple is a collection of members from various dimensions and can be determined by several members.
- Set in its turn is a collection of tuples from identical dimensions/hierarchies.
- Other: there are also member properties as an alternative to data attributes.

SELECT Statement:

The SELECT statement is the most fundamental query in MDX, as previously stated. The SELECT statement selects a selection of multidimensional data from a data set. The Choose keyword points to the beginning of the query and describes what you wish to select. Both simple and complicated queries are possible.

A simple MDX query example can be as following:

```
SELECT
  { [Measures].[Sales Amount],
```

```
    [Measures].[Tax Amount] } ON COLUMNS,
  { [Date].[Fiscal].[Fiscal Year].&[2002],
    [Date].[Fiscal].[Fiscal Year].&[2003] } ON ROWS
FROM [Adventure Works]
WHERE ( [Sales Territory].[Southwest] )
```

In this example, the query defines the following result set information:

- The SELECT clause sets the query axes as the Sales Amount and Tax Amount members of the Measures dimension, and the 2002 and 2003 members of the Date dimension.
- The FROM clause indicates that the data source is the Adventure Works cube.
- The WHERE clause defines the slicer axis as the Southwest member of the Sales Territory dimension.

2.4.3.2 MDX Query Structure

The SELECT statement in MDX specifies a result set including a subset of multidimensional data retrieved from the cube, an MDX query must include the following clauses to define this set:

WITHclause: permits the named sets to be calculated while the SELECT and WHERE clauses are being processed;

SELECTclause: specifies which dimension elements will be used in each axis of the MDX query structure.

FROMclause: identifies the searched cube and specifies which multidimensional data source should be used to fill the result set of the SELECT MDX query;

WHEREclause: specifies which member or dimension will be utilized as a slicer dimension (the slicer usually refers to the axis formed by the WHERE clause).

2.5 Data warehouse and Data mart theory and practical usage

2.5.1 Data Warehouses and Data Marts theory

Data warehouse is a centralized repository for data from various sources such as ERP, legacy, or external systems and applications. It collects data from across the enterprise. Then makes it available for analysis, reporting, and other business intelligence (BI) functions to aid manufacturers in making data-driven decisions.

Data mart is a subset of a data warehouse that comprises structured data on a single subject or business line. It divides enormous datasets into smaller, more manageable bits. Such as inventory or production data, to increase data aggregation and agility without the need to retrieve data extracts from a centralized corporate database.

The difference between a data warehouse and a data mart is the size and or scope of information in the database.

We need them because the information in their source databases is not organized to make it readily accessible. They are too complicated, difficult to access, or resource-intensive to access.

The influx and amount of data are what contributed to the problem. Sifting through the "noise" in data to provide quick and accurate insights is a key business value to BI.

2.5.2 How does the data get into the data warehouse in practical usage?

A data staging area is where data is stored when brought over from the data sources. Staged data is typically a 1-1 of the source system. A data lake is a type of staging area. The data is in its natural/ raw format and is usually object blobs or files.

A data integration area is where data is transformed before being added to the data warehouse. The transformation includes combining data with multiple sources, cleaning,

de-duplication, adding business rules, adding calculations, and optimizing the data for retrieval.

2.5.3 Using the 5C's to prepare data

Clean—dirty data can really muddy up a company's attempt at real-time disclosure and puts the CFO at high risk when signing off on financial reports and even press releases based on incorrect information. Dirty data has missing items, invalid entries, and other problems that wreak havoc with automated data integration and data analysis. Customer and prospect data, for example, is notorious for being dirty. Most source data is dirty to some degree, which is why data profiling and cleansing are critical steps in data warehousing.

Consistent—there should be no arguments about whose version of the data is the correct one. Management meetings should never have to break down into arguments about whose number is correct when they really need to focus on how to improve customer satisfaction, increase sales, or improve profits. Business people using different hierarchies or calculations for metrics will argue regardless of how clean the transactional data is.

Conformed—the business needs to analyze the data across common, shareable dimensions if business people across the enterprise are to use the same information for their decision-making.

Current—the business needs to base decisions on whatever currency is necessary for that type of decision. In some cases, such as detecting credit card fraud, the data needs to be up to the minute.

Comprehensive—business people should have all the data they need to do their jobs—regardless of where the data came from and its level of granularity.

CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION

3.1 Production Process

3.1.1 Production Department

A production department is a group of functions that is responsible for the manufacture of goods, or a fully functioning department that converts raw materials, assembles components into finished goods, and packages them.

Adventure Works (AW) is a fictional company that specializes in manufacturing bicycles. The company's product line includes 97 different brands of bikes, grouped into three categories: mountain bikes, road bikes, and touring bikes. In addition to manufacturing bicycles, AW also manufactures some of its own components. Other components are purchased from outside vendors, as well as all accessories and clothing.

3.1.2 Production Process

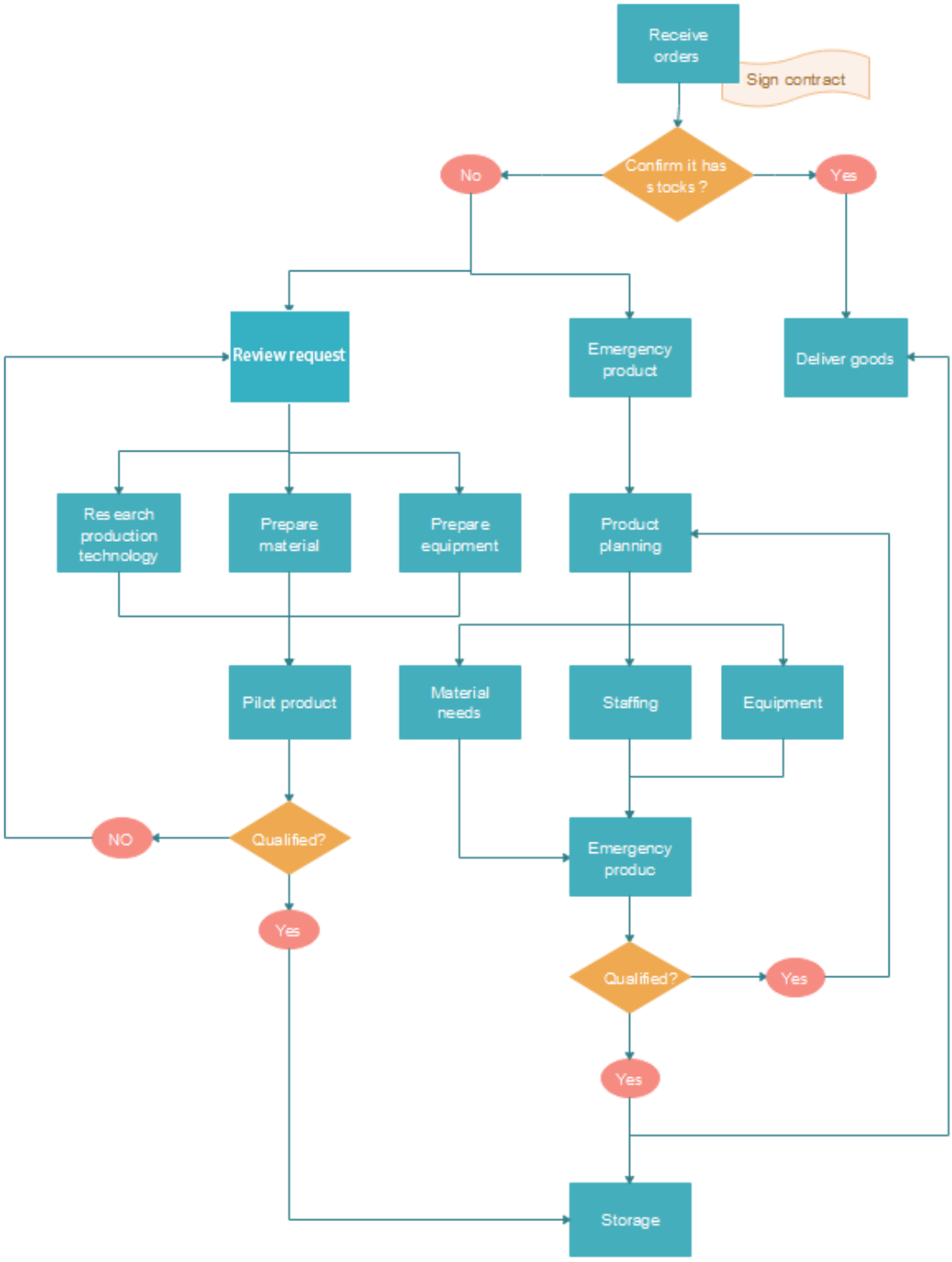


Figure 2. The BPMN of the production process

When receiving orders, the production department will confirm and review the equipment and machinery situation at the factory again to start the production process.

When the equipment has been confirmed to work stably, the list of materials needed to create a finished product will be updated, then the information will be sent to the warehouse for checking again and make sure that the quantity of materials needed for production is still sufficient.

After being qualified, the product is confirmed in the system. Then, materials are moved to storage, and system reports that they are now available for consumption by other processes. All information will be saved on the ERP system.

3.2 Data source and challenges

3.2.1 Data source

AdventureWorks Database is a Microsoft product sample for an online transaction processing (OLTP) database. The AdventureWorks Database supports a fictitious, multinational manufacturing company called Adventure Works Cycles

In our project, we focus on the Production process with different requirements and tables from the original data source of Adventure Works. After searching and analyzing the data sources, we finally have our own database. We mainly worked on the Production process with 11 tables and lots of requirements. In the original database, the Production process of Adventure Work has 25 different tables

3.2.2 Challenges

During time of executing, we got a lot of difficulties, especially are those:

- Challenges in using these databases is their small size which is geared towards ease of distribution, more than demonstrating all of the functionality of SQL Server

- We had problems when data was duplicated a lot. It makes our cleaning data process exceed the deadline
- Adventure Works's data is not constantly updated, the current analysis is only based on data from many years ago

3.3 Business Requirements Analysis

3.3.1 Case's Setting - Customer Demand:

According to a British article, there was a significant growth in the demand for entertainment and health originated from the success of the bicycle race Tour De France as well as the Olympics London 2012. This has led to the resurrection of the cycling industry over a period of 4 years (2011 - 2014), especially 2012 experienced a peak in bicycles sales. In order to prove this statement, statistics from the table Sales.SalesOrderHeader will compare total sales and its growth rate from 2011 to 2014.

3.3.2 Customer Segment:

However, to deeply understand the customer segment, we have to answer the following question: "Whether our customer is new or old, what is their buying capacity?", rather than growth rate. Therefore, the calculation here will be made from the table Sales.Customer, Sales.SalesPerson and so on.

Once we acknowledged halfly what was happening in this period, we noticed that there is an issue coming from AW's category including: Components, Bicycle, Clothing and Accessories. Mainly from Bicycle and its real profit from customer segment's need which was analyzed above. To clarify this statement, we need data from Production.ProductCategory, Production.ProductSubcategory, Production.ProductModel and Sales.SalesOrderDetail.

After indicating what the relationship between each product line and its sales, we start to analyze the correlation among the product line' factors to its sales to define the causation and whether AW have to create a new product line:

- Cost from ProductCostHistory
- Size, Weight, Color, etc,... of each model, category from Product , productSubcategory, ProductCategory, ProductModel
- The number of customers chose to buy that model/category
- Each model/category profit (KPI: ROI)
- Routing effectiveness by comparing the duration between planned date and actual date; and the portion of scrapped product.

As a result, we will find out what we have to do with the present product category/model and what enhancement we can apply to product's routing?

3.3.3 IT requirements Analysis (IT & Infrastructure)

The IT requirements for implementing a BI Solution project are incredibly vital and necessary. IT requirements must be meticulously prepared throughout the project implementation process, from the Initiation phase to the Closure phase.

IT requirements for project implementation may include the following:

3.3.3.1. Hardware

BI Solution project may have to perform heavy tasks, so the team needs to be equipped with strongly configured computers and enough capacity to be able to implement the project as smoothly as possible.

3.3.3.2. Software

For the Planning phase, the project team needs to have access to Microsoft Project software to be able to create a schedule that organizes the work in phases and monitors the progress of the project.

The team will need to adopt the draw.io tool in order to draw a sample data warehouse containing fact tables and dimensions tables suitable for the production subsystem. In addition, the Microsoft SQL Management Studio 18 tool helps the project team query the database and create a database for our production subsystem based on the AdventureWorks database.

Next, based on the construction of the data warehouse, the project team needs to understand and familiarize ourselves with software such as Visual Studio to pour data from the data source into the data warehouse that the team has built.

3.3.4. Comparative Analysis of BI and Data Visualization Tools

3.3.4.1 Surveying and evaluation

Businesses employ Business Intelligence (BI) tools to improve their decision-making processes. This is frequently accomplished by gathering, integrating, and analyzing massive amounts of data before producing meaningful findings. As a result, BI has a direct impact on an organization's strategic and business operations.

BI makes data discovery accessible to everyone, not just skilled analytics experts, by utilizing a variety of tools and apps. However, most organizations find it difficult to select the correct technology that corresponds with their objectives. There are three BI tools on the market that have acquired greater traction in recent years than the others. *Microsoft Power BI*, *Google Data Studio*, and *Tableau* are the three tools.

The first allows you to quickly generate customized charts and reports, the second provides dynamic interactive charts with advanced BI capabilities, and Tableau has a highly intuitive interface with customisable dashboards.

For the past five years, Gartner has positioned Microsoft Power BI and Tableau in the Leader's Quadrant for Business Intelligence and Analytics Platform. This is determined by analyzing and rating various categories such as infrastructure, product capabilities, architecture, sales and marketing strategies, and customer feedback, among others. This report is up to date as of January 2022.

Figure 3. Magic quadrant for Analytics and Business Intelligence Platforms

Here's a quick comparison of the top three most popular business intelligence (BI) tools on the market today

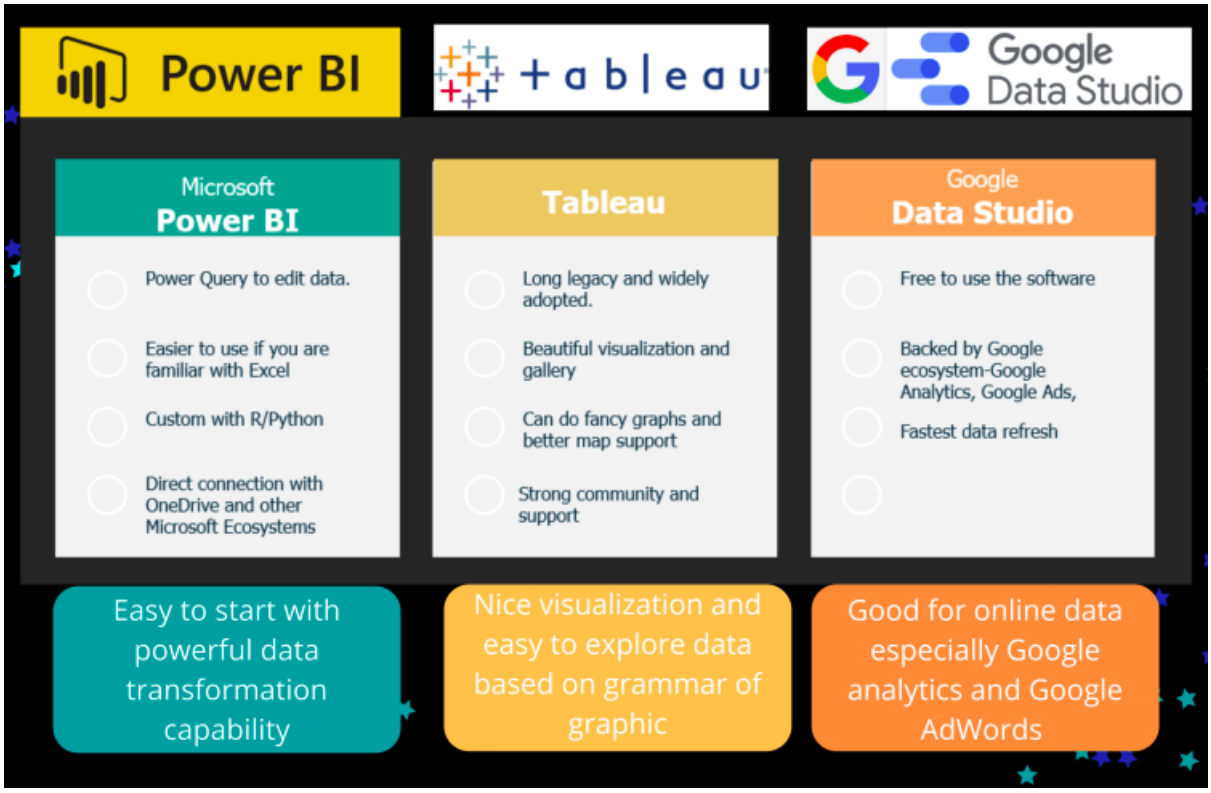


Figure 4. Comparison of the top 3 most popular BI tools

3.3.4.2 Proposing BI solution for the project

Choosing the best Business Intelligence Software is critical to successful data interpretation and analysis, and many business issues can be traced back to ineffective BI tool selection processes.

If money is an issue, Google Data Studio would be the best option. On the other hand, if a company needs to perform complex data transformations, is willing to pay, and wants dynamic report generation, Microsoft Power BI or Tableau are the best options.

With the BI Solution for AdventureWorks project, the proposed team should combine multiple BI tools together to get the most intuitive comparison and evaluation possible. Power BI has an easy-to-use interface and exceptional data visualization capabilities while Tableau enables the connection of data from virtually any source and in virtually any format. A simple drag-and-drop interface could be used to create an interactive visual

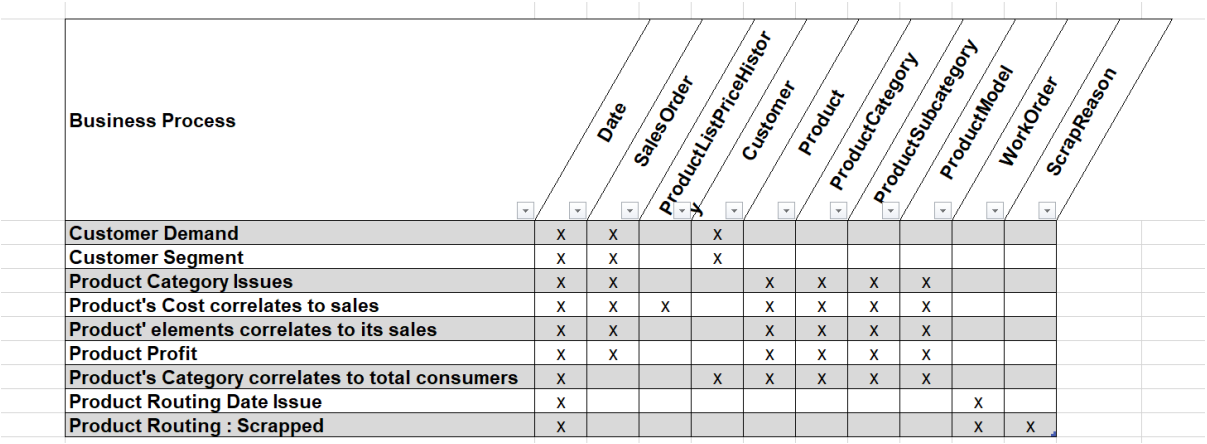
analytics dashboard. Data can be converted into charts, graphs, maps, and other visualizations using this method.

CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

In the chapter 4, the team discuss designing data warehouse, a data warehouse model, and ETL process to integrate data from the data source into the data warehouse

4.1 Designing data warehouse

4.1.1. Bus Matrix



Business Process	Date	SalesOrder	ProductListPriceHistor	Customer	Product	ProductCategory	ProductSubcategory	ProductModel	WorkOrder	ScrapReason
Customer Demand	x	x		x						
Customer Segment	x	x		x						
Product Category Issues	x	x			x	x	x	x		
Product's Cost correlates to sales	x	x	x		x	x	x	x		
Product' elements correlates to its sales	x	x			x	x	x	x		
Product Profit	x	x			x	x	x	x		
Product's Category correlates to total consumers	x			x	x	x	x	x		
Product Routing Date Issue	x								x	
Product Routing : Scrapped	x								x	x

Figure 5. Bus matrix

Developed by Kimball, the Bus Matrix Architecture provides us with a manageable design of a data warehouse. It shows details of which core business process comes along with the associated dimensions. The business process here includes what we have stated in the requirements of the project and theirs 10 needed dimensions related to the production process and its correlation to sales.

4.1.2. Master Data

Object	**Description**
Product	All information such as size, weight, class, color,... of a product

Customer	All information of a customer
Product Category, Subcategory, Model	Name of Category and Subcategory

Table 1. Master data

4.1.3. Transaction Data

Object	**Description**
Sales Order	Data, document of sales process
Product Cost	Standard cost of an product at a period of time (Auctual/ Planned cost)
Work Order Routing	Detail of an scheduled/ actual manufacturing date (Start and End date)
List Price	Price of an product at a period of time

Table 2. Transaction data

4.1.4. Fact and dimension tables (Mapping)

Attribute Name	Data Type	IsNull	Table(source)	Attribute Name	Data Type	IsNull	ETL Rules
DimCustomer							
CustomerKey (PK)	Int						Auto Generate
CustomerID	Int		Sales.Customer	CustomerID	Int		
Title	Nvarchar(5)	x		Title	Nvarchar(5)		
Suffix	Nvarchar(5)	x		Suffix	Nvarchar(5)		
First Name	Nvarchar(50)	x		First Name	Nvarchar(50)		
Middle Name	Nvarchar(50)	x		Middle Name	Nvarchar(50)		
Last Name	Nvarchar(50)	x	Person.Person	Last Name	Nvarchar(50)		From Source
DimProduct							
ProductKey (PK)	Int						Auto Generate
ProductID	Int			ProductID	Int		
ProductModelID	Int			ProductModelID	Int		
ProductName	Nvarchar(50)			Name	Nvarchar(50)		
ProductNumber	Nvarchar(25)			ProductNumber	Nvarchar(25)		
ProductLine	nchar(10)	x		ProductLine	Nvarchar(2)	x	
Size	nvarchar(5)			Size	Nvarchar(5)	x	
Weight	decimal(8, 2)			Weight	Decimal(8,2)	x	
Class	nchar(2)			Class	nchar(2)	x	
Color	nvarchar(15)	x		Color	Nvarchar(15)	x	
Style	nchar(10)	x		Style	nchar(2)	x	
ProductModelID	int	x		ProductModelID	int		
ProductSubcategoryID	int	x	Production.Product	ProductSubcategoryID	int		
ProductCategoryID	int	x	Production.ProductCategory	ProductCategoryID	int		
Model	Nvarchar(50)		Production.ProductModel	Name	Nvarchar(50)		
Category	Nvarchar(50)		Production.ProductCategory	Name	Nvarchar(50)		
SubCategory	Nvarchar(50)		Production.ProductSubCategory	Name	Nvarchar(50)		From Source
DimPrice							
PriceKey							Auto Generate
ProductID	int			ProductID	int		
ListPrice	money		Production.ProductListPriceHistory	ListPrice	money		
StandardCost	money			StandardCost	money		
StartDate	datetime			StartDate	datetime		
EndDate	datetime	x	Production.ProductCostHistory	EndDate	datetime		From Source
FactProductSale							
FactProductSaleKey (PK)	Int						Auto Generate

Figure 6. Fact and Dimension tables

CustomerKey	Int						
ProductKey	Int						
PriceKey	Int						
OrderDate	Int						
ShipDate	Int						
DueDate							
ProdcutID	Int		Production.Product				
SalesOrderID	Datetime			ModifiedDate	Datetime		
Freight	Money			Freight	Money		
SalesAmount	Money			SubTotal	Money		
SalesOrderNumber			Sales.SaleOrderHeader				
UnitPrice	Money			UnitPrice	Money		
OrderQty	int			OrderQty	Money		
UnitPriceDiscount			Sales.SalesOrderDetail	UnitPriceDiscount	Money		
StandardCost	Money		Production.Product	StandardCost	Money		From Source
ProductCost	Money						
FactWorkOrderRouting							
WorkOrderKey (PK)	Int						Auto Generate
ProductKey							
WorkOrderID	Int			WorkOrderID	Int		
OperationSequence	Int						
ScheduledStartDate	Datetime	x		ScheduledStartDate	Datetime		
ScheduledEndDate	Datetime	x		ScheduledEndDate	Datetime		
ActualStartDate	Datetime	x		ActualStartDate	Datetime	x	
ActualEndDate	Datetime	x		ActualEndDate	Datetime	x	
PlannedCost	Money			PlannedCost	Money		
ActualCost	Money		Production.WorkOrderRouting	ActualCost	Money	x	
ScrappedQuantity	int	x	Production.WorkOrder	ScrappedQuantity	int		
ProductID	int		Production.Product	ProductID	int		From Source

Figure 7. Fact and Dimension tables (cont)

4.1.5 Data Warehouse Model

Designing Data Warehouse with Galaxy schema:

Following the requirements, we designed a data warehouse including 6 tables with 2 fact tables and 4 dimension tables. With data from 2 business processes are sales and production.

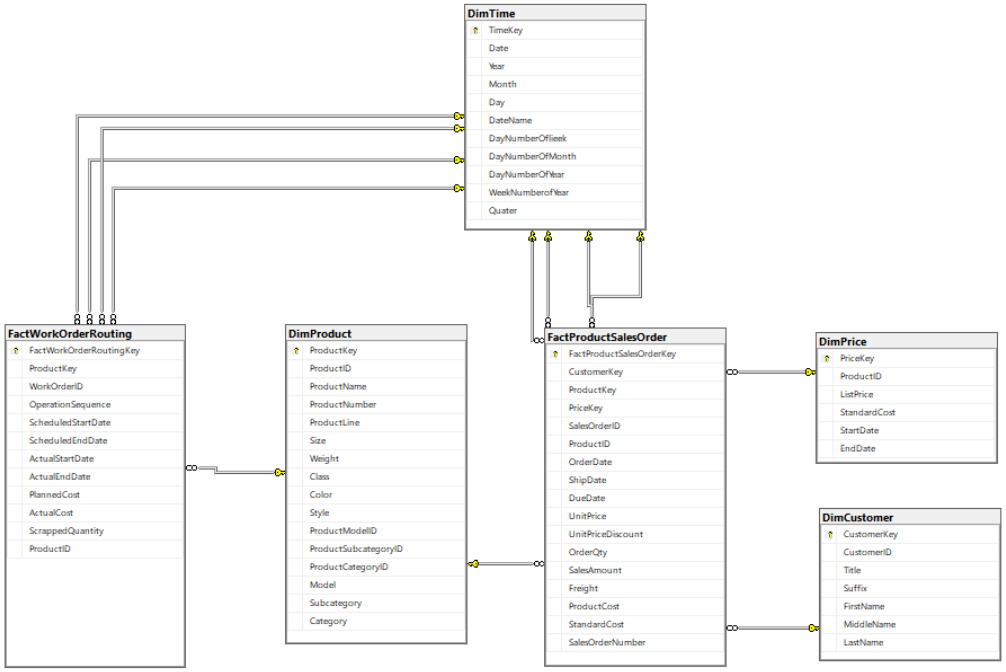


Figure 8. Data warehouse model

4.2 ETL processes

4.2.1. Dimension Table's ETL Process:

These is the steps for the ETL process of Dim tables, the process to ETL for other Dim tables is similar:

Step 1: Create a new Integration Services Project in Visual Studio

Step 2: New SSIS Package -> Data Flow Task -> Double click on Data Flow Task -> OLE DB Source from SSIS Toolbox > Double click on OLE DB Source -> Select source database -> Data access mode select SQL command or Table or view

Step 3: Insert SQL command or Table or view

Step 4: Drag Slowly Changing Dimension from SSIS toolbox, select Dimension table and keys then select a change type for slowly changing dimension columns and finish step

Step 5: Run

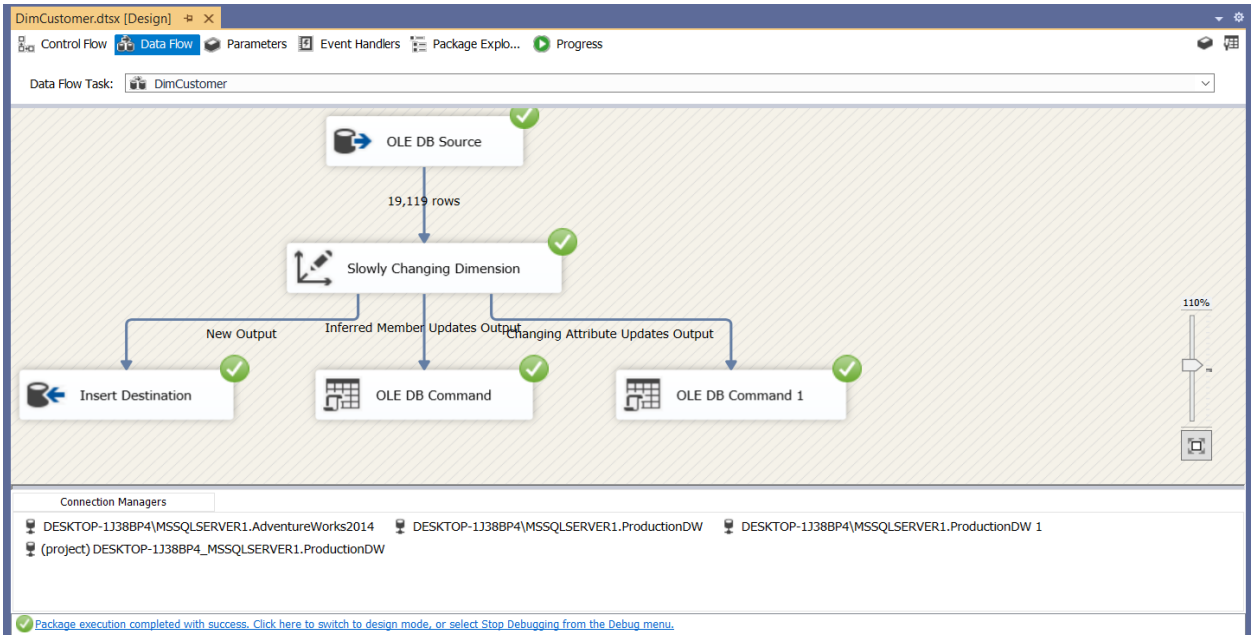


Figure 9. ETL for dimCustomer

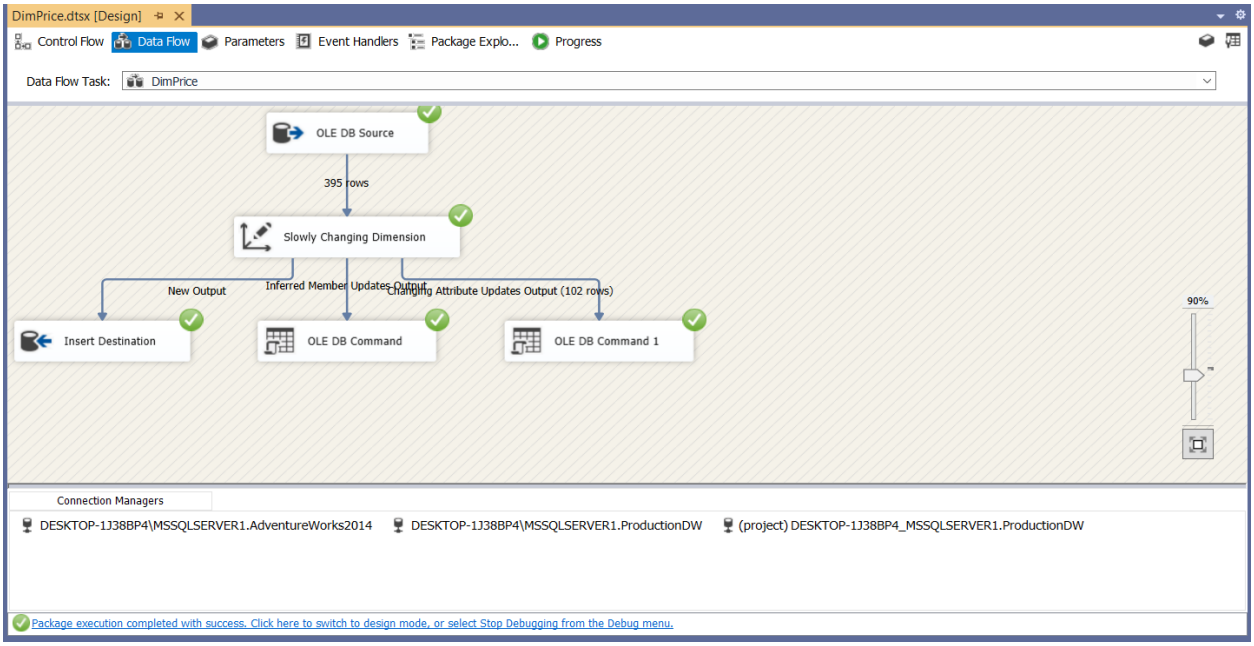


Figure 10. ETL for dimPrice

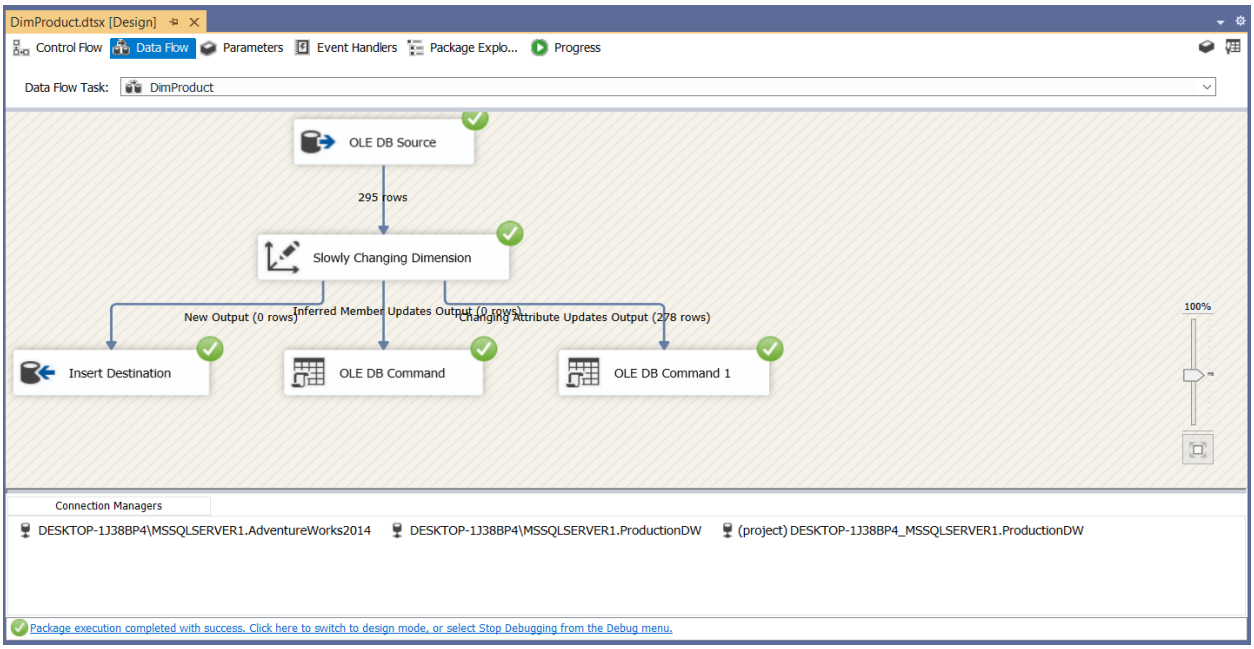


Figure 11. ETL for dimProduct that has sales record (mean that we only loading the product that are sold)

4.2.2. Fact Table's ETL Process

Step 1: Create a new Integration Services Project in Visual Studio

Step 2: New SSIS Package -> Data Flow Task -> Double click on Data Flow Task -> OLE DB Source from SSIS Toolbox > Double click on OLE DB Source -> Select source database -> Data access mode select SQL command.

Step 3: Insert SQL command.

Step 4: Insert lookup and mapping then set input column and lookup column.

Step 5: Create "Merge" from SSIS toolbox to merge data between data source and data warehouse.

Step 6: Set up condition to update and insert.

Step 7: Choose OLE DB Destination from SSIS toolbox to insert and OLE Command to update data.

Step 8: Run.

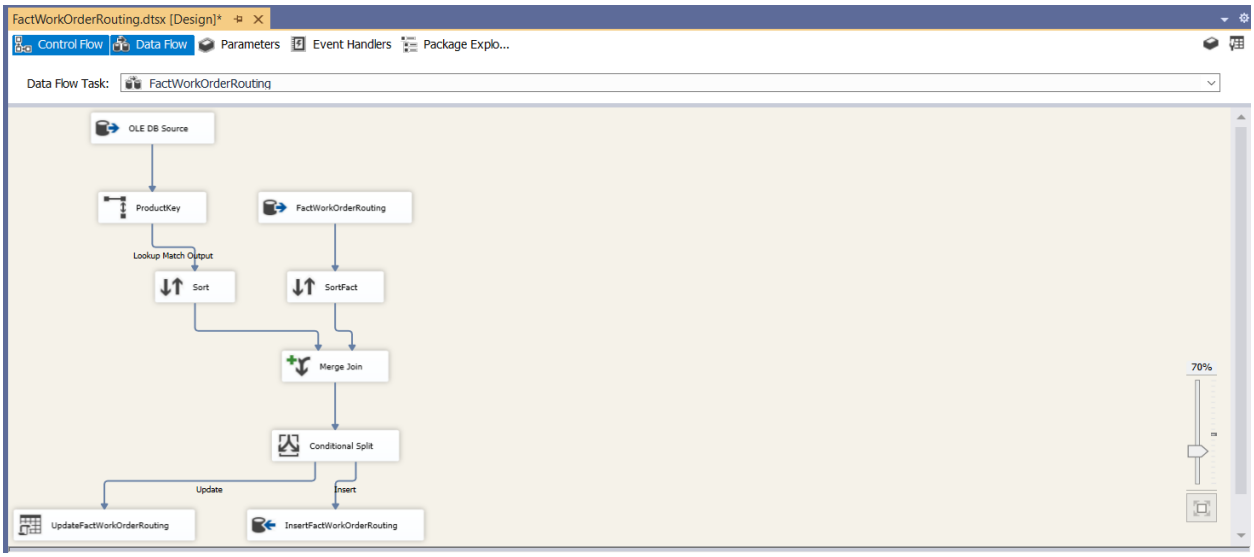


Figure 12. ETL for FactWorkOrderRouting

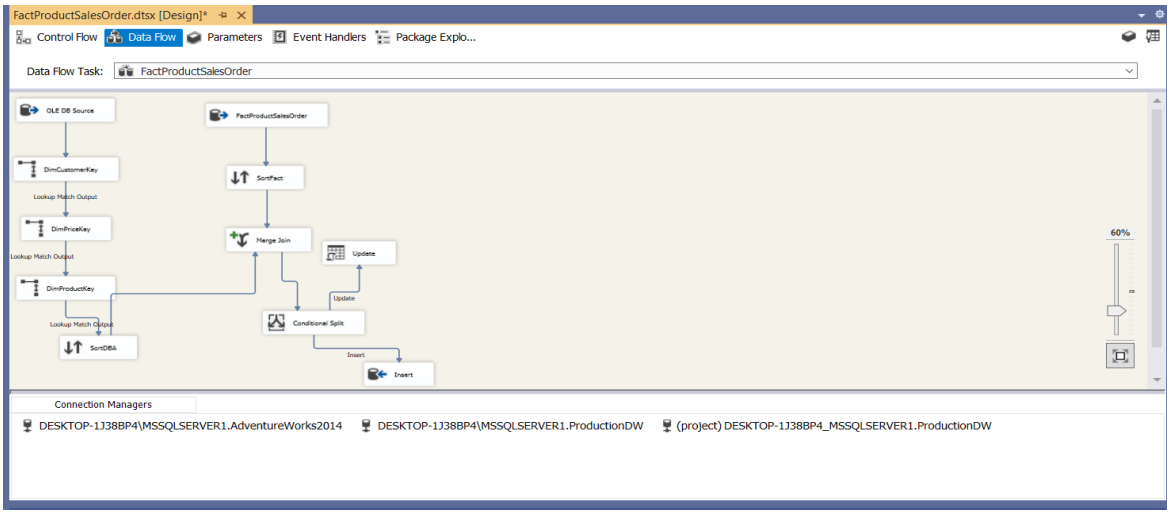


Figure 13. ETL for FactProductOrderSales

CHAPTER 5: DATA ANALYTICS

5.1. Data analytics with SSAS technology

5.1.1 Building the cube:

Step 1: In Data Sources tab -> New data source -> Tick at Create a data source based on an existing or new connection then choose data connection -> Tick at Use the service account -> Finish

Step 2: In Data Source View -> New data source view -> Select a data source -> choose all Dim and Fact table and switch them from Available objects to Included objects -> Finish

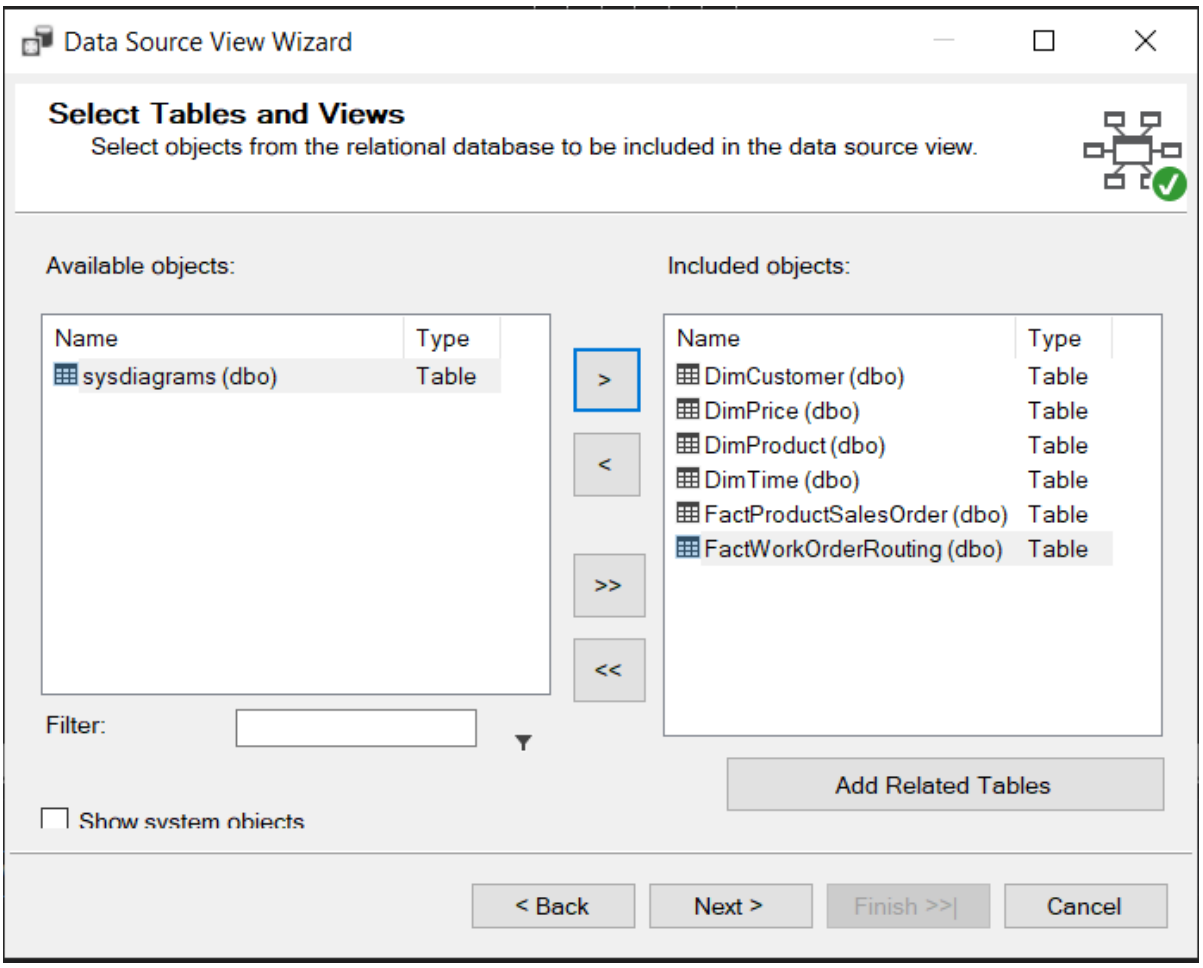


Figure 14. Step 1+2 Building the cube

Step 3: In Cubes tab -> New cube -> Tick Use existing table -> Tick all Fact table -> Then click “Next” to end.

Step 4: After create cube, the team transferred attributes from Data source view column to Attribute column

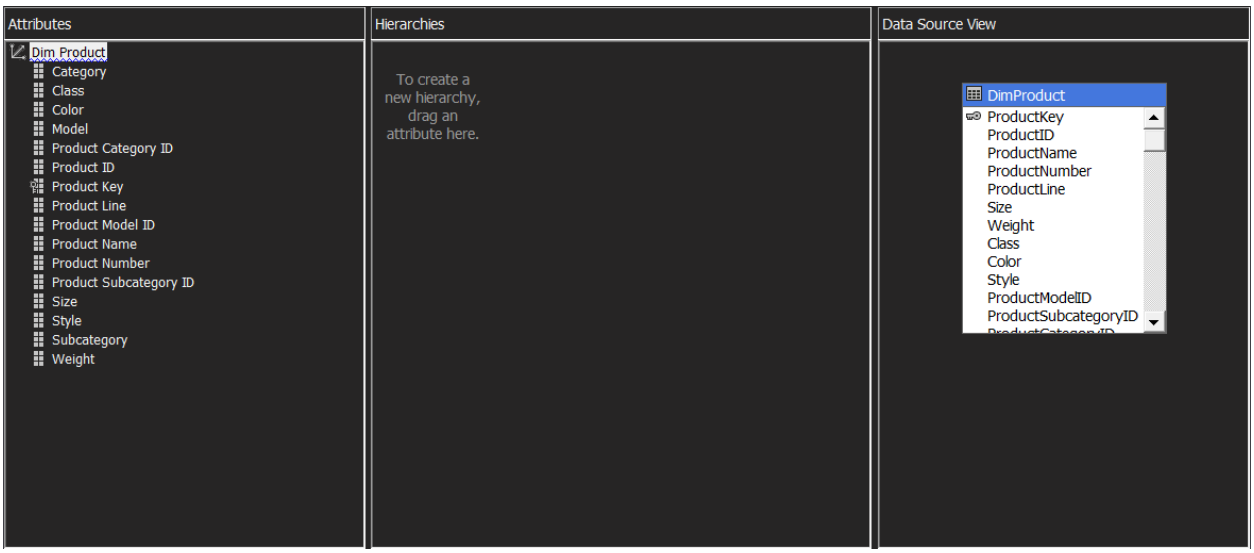


Figure 15. Step 4 Building the cube

Step 5: After the setup step 4 are successful, the screen will show dimension and fact tables and relationship among the tables

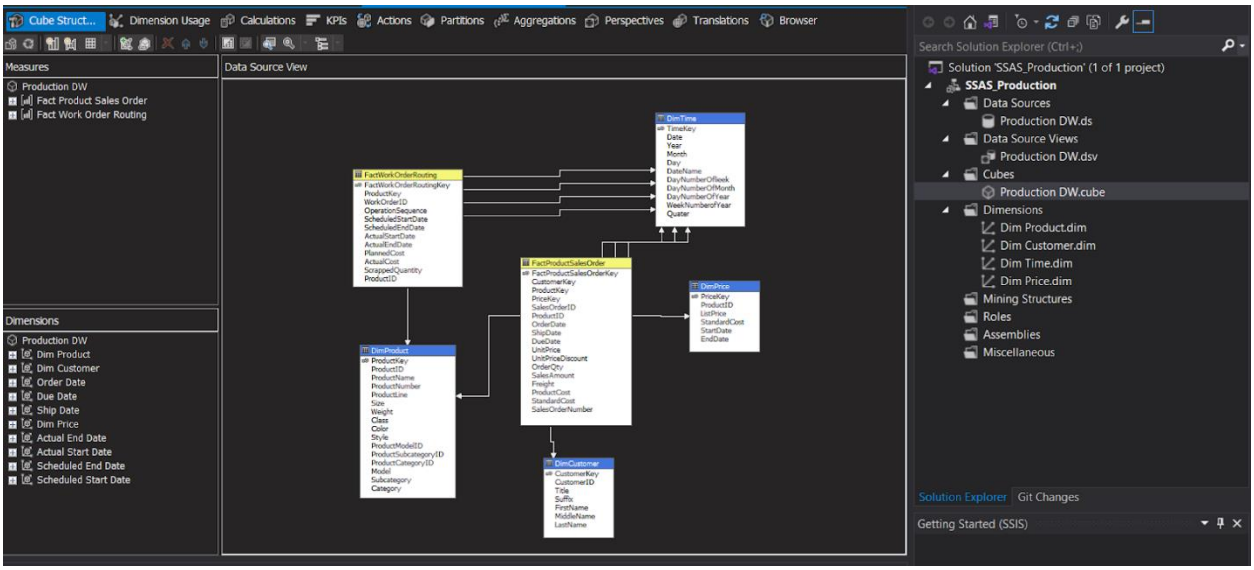


Figure 16. Step 5 Building the cube

5.1.2 Analysis with SSAS:

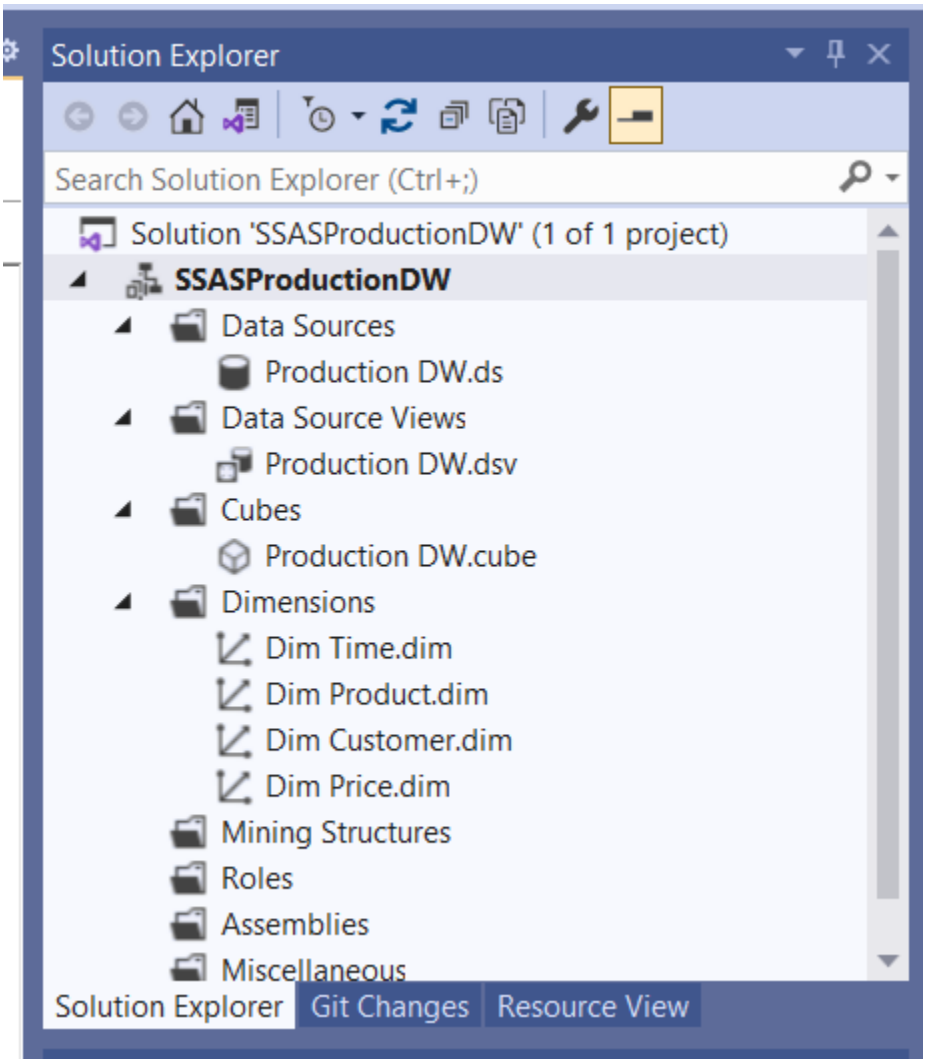


Figure 17. Step to Analysis with SSAS

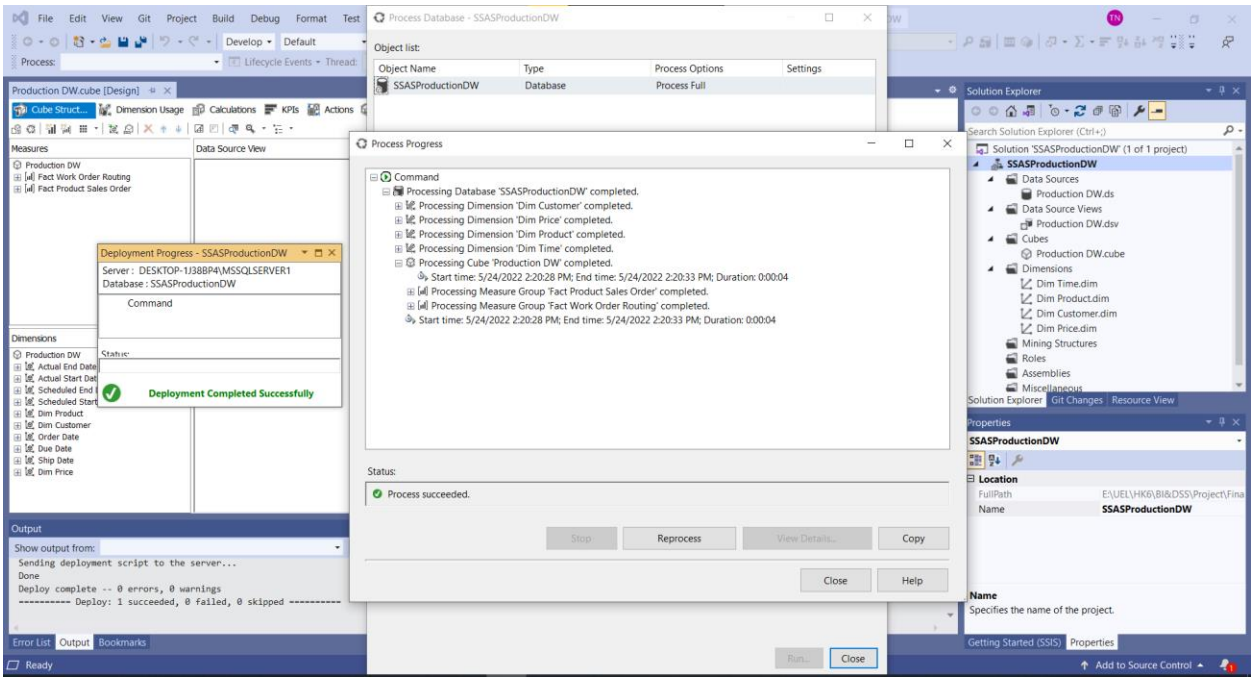


Figure 18. Step to Analysis with SSAS (cont)

5.1.3 Building KPIs system

The team built two KPI to analyze profit rates over the years, consider whether the actual profit rate is achieved as the target or not and the trend of profit compared to the previous year is growing or decreasing.

5.1.3.1 KPI Profit for category

Coding for KPI:

- Value Expression: [Measures].[profit]
- Goal Expression:

case

when [Dim Product].[Category] is [Dim Product].[Category].&[Bikes] then 0.1

when [Dim Product].[Category] is [Dim Product].[Category].&[Clothing] then 0.1

when [Dim Product].[Category] is [Dim Product].[Category].&[Accessories] then 0.3

else 0.05

end

●Status indicator:

case

when KPIVALUE("KPIProduct")/KPIGOAL("KPIProduct") > 0.9 then 1

when KPIVALUE("KPIProduct")/KPIGOAL("KPIProduct") <=0.9

and KPIVALUE("KPIProduct")/KPIGOAL("KPIProduct") > 0.8 then 0

else -1

end

● Trend:

case

when ISEMPTY(PARALLELPERIOD([Order Date].[Year].[Year] ,1,[Order Date].[Year])) then 0

when [Measures].[profit] > PARALLELPERIOD([Order Date].[Year].[Year] ,1,[Order Date].[Year]) then 1

when [Measures].[profit] = PARALLELPERIOD([Order Date].[Year].[Year] ,1,[Order Date].[Year]) then 0

else -1

end

Row Labels	KPIProfit	KPIProduct Goal	KPIProduct Status	KPIProduct Trend
2011				
Accessories	35.14%	0.3		
Bikes	13.65%	0.1		
Clothing	-26.54%	0.1		
Components	-0.14%	0.05		
2012				
Accessories	32.39%	0.3		
Bikes	2.57%	0.1		
Clothing	14.28%	0.1		
Components	2.34%	0.05		
2013				
Accessories	49.21%	0.3		
Bikes	7.13%	0.1		
Clothing	13.66%	0.1		
Components	5.18%	0.05		
2014				
Accessories	55.67%	0.3		
Bikes	17.05%	0.1		
Clothing	20.22%	0.1		
Components	6.54%	0.05		
Grand Total	**8.53%**	**0.05**		

Figure 19. Result of KPI Profit for Category

5.1.3.2 KPI Profit for Product Line

Coding for KPI:

● Value Expression: [Measures].[Profit]

● Goal Expression:

case

when [Dim Product].[Product Line] is [Dim Product].[Product Line].&[M] then 0.15

when [Dim Product].[Product Line] is [Dim Product].[Product Line].&[R] then 0.1

when [Dim Product].[Product Line] is [Dim Product].[Product Line].&[T] then 0.1

else 0.2

end

●Status indicator:

case

when KPIVALUE("KPIProfit")/KPIGOAL("KPIProfit") > 0.9 then 1

when KPIVALUE("KPIProfit")/KPIGOAL("KPIProfit") <= 0.9

and KPIVALUE("KPIProfit")/KPIGOAL("KPIProfit") > 0.8 then 0

else -1

end

● Trend:

case

when ISEMPTY(PARALLELPERIOD([Order Date].[Year].[Year] , 1, [Order Date].[Year])) then 0

when [Measures].[profit] > (PARALLELPERIOD([Order Date].[Year].[Year] , 1, [Order Date].[Year])) then 1

when [Measures].[profit] = (PARALLELPERIOD([Order Date].[Year].[Year] , 1, [Order Date].[Year])) then 0

else -1

end

Row Labels	profit	KPIProfit Goal	KPIProfit Status	KPIProfit Trend
Unknown				
2011		0.2		
2012	25.57%	0.2		
2013	25.85%	0.2		
2014	25.98%	0.2		
M				
2011	10.34%	0.15		
2012	2.91%	0.15		
2013	17.72%	0.15		
2014	26.10%	0.15		
R				
2011	15.32%	0.1		
2012	2.27%	0.1		
2013	3.74%	0.1		
2014	12.56%	0.1		
S				
2011	-6.86%	0.2		
2012	15.56%	0.2		
2013	20.57%	0.2		
2014	30.52%	0.2		
T				
2011		0.1		
2012		0.1		
2013	-4.24%	0.1		
2014	10.19%	0.1		
Grand Total	**8.53%**	**0.2**		

Figure 20. KPI table for customer trend

5.2. Data analytics with MDX and OLAP technique

5.2.1 Slice & Dice:

Briefly, slice and dice is the process of breaking down a large body of information into smaller pieces or examining it from different perspectives in order to better understand it.

Slicing refers to selecting a subset of the cube by choosing a single value for one of its dimension and creating a smaller cube with one less dimension.

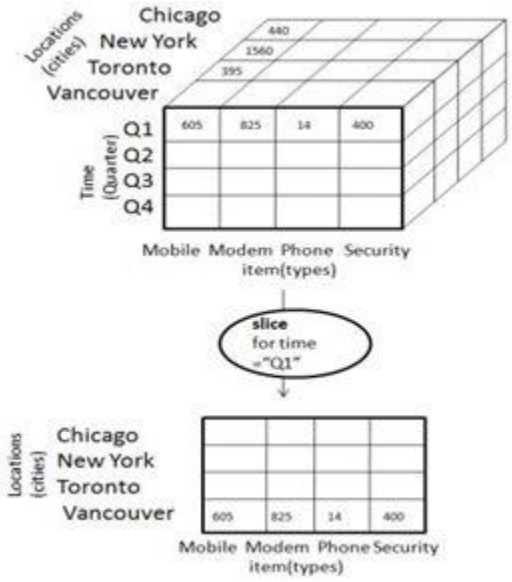


Figure 21. Example of Slicing function

Dicing on the other hand generates a subcube by picking two or more values from multiple dimensions of the cube. The cube is rotated independent of its dimension.

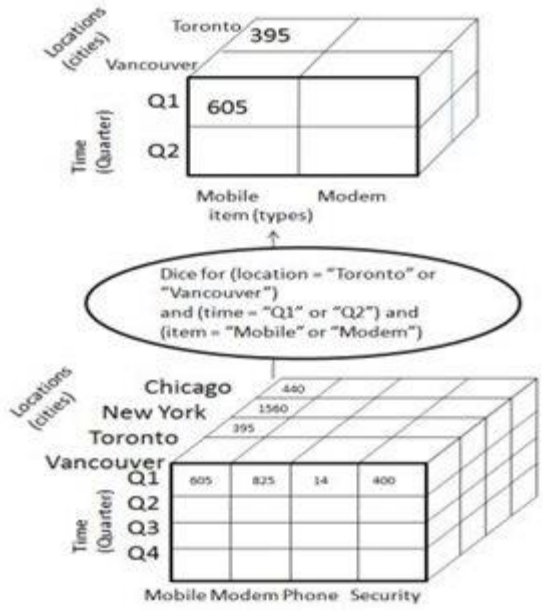


Figure 22. Example of Dicing function

Thus dicing and slicing allows detailed analysis of data from different angles.

5.2.2 Drill down:

Drill Down is the process of extracting data by subdividing the data size. It is performed by either of the following ways:

- By stepping down a concept hierarchy for a dimension
- By introducing a new dimension.

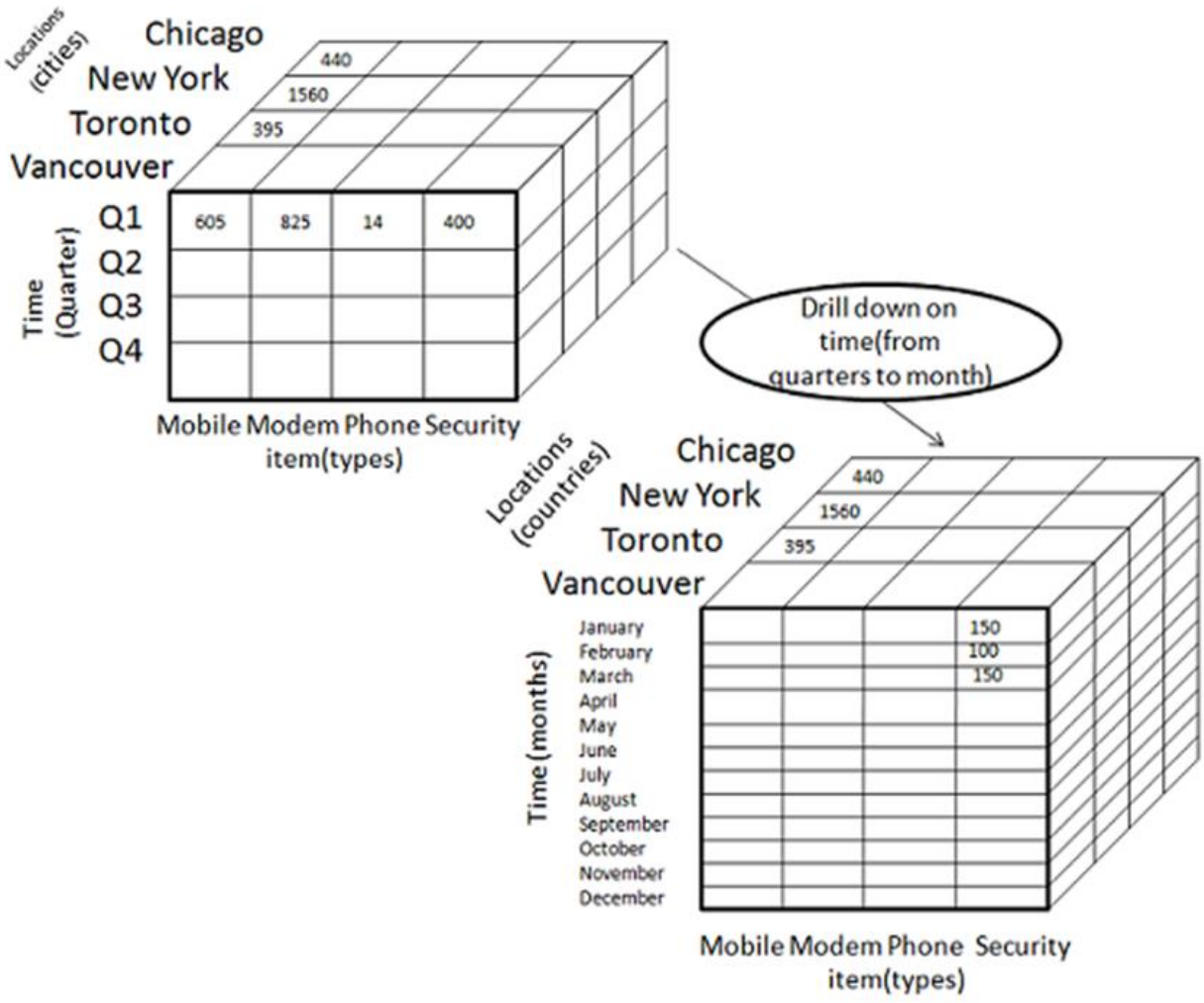


Figure 23. Example of Drill-down function

- Drill-down is performed by stepping down a concept hierarchy for the dimension time.
- Initially the concept hierarchy was "day < month < quarter < year."
- On drilling down, the time dimension is descended from the level of quarter to the level of month.
- When drill-down is performed, one or more dimensions from the data cube are added.
- It navigates the data from less detailed data to highly detailed data.

5.2.3 Roll up:

Going up a level in the hierarchy is called rolling up the data. Roll up is the opposite of the drill down function - it performs aggregation on a data cube by moving up in the concept hierarchy or by reducing the number of dimensions.

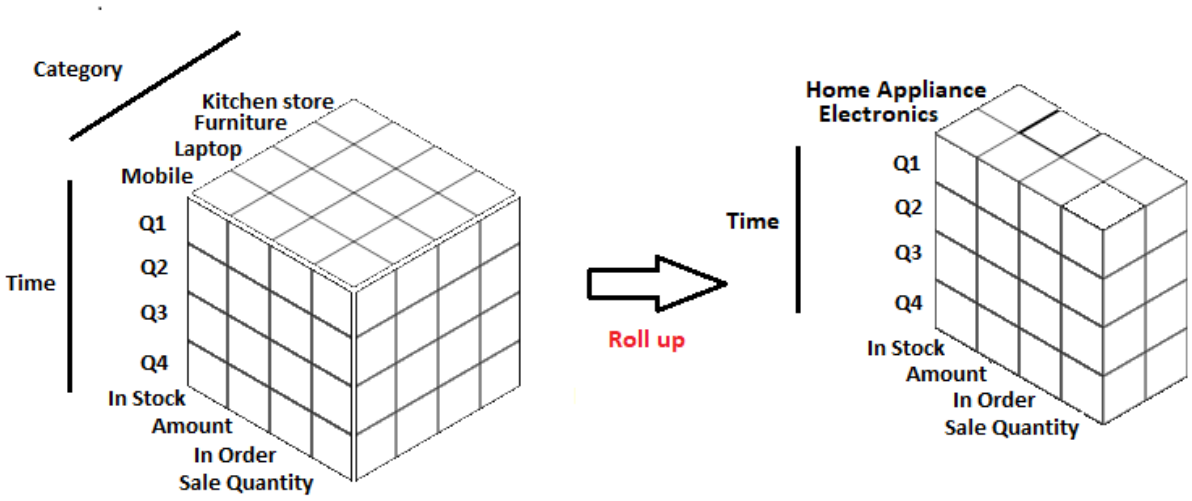


Figure 24. Example of Roll-up function

5.2.4 Rotate or Pivot:

Pivoting, also called rotating, is similar to creating a spreadsheet Pivot Table. It roughly "pivots" a table to rotate some row values into columns to show summary information. Pivoting allows you to summarize a data column by primary and secondary columns.

For example, if your data looks like this, with Sales Bicycles in the column headings and and Quarters along the left side:

Sales by Bicycle Type	Road	Touring	MTB
Qtr 1	$54,600	$67,000	$60,000
Qtr 2	$43,200	$92,000	$43,000
Qtr 3	$22,000	$37,000	$57,000
Qtr 4	$32,300	$51,000	$67,000

Table 3. Example of Rotate or Pivot function

This feature will rearrange the table such that the Quarters are showing in the column headings and the Sales Bicycles can be seen on the left, like this:

Sales by Bicycle Type	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Road	$54,600	$43,200	$22,000	$32,300
Touring	$67,000	$92,000	$37,000	$51,000
MTB	$60,000	$43,000	$57,000	$67,000

Table 4. Example of Rotate or Pivot function (cont)

5.2.5 OLAP use SSAS

5.2.5.1 Average value of orders

	CustomerKey	CustomerName	Average sales order($)
1	37327	Roger Harui	877107.19
2	37224	Andrew Dixon	853849.18
3	37231	Reuben D'sa	841908.77
4	37712	Robert Vessa	816755.58
5	37123	Ryan Calafato	799277.90
6	37148	Joseph Castellucio	787773.04
7	37210	Kirk DeGrasse	746317.53
8	37126	Lindsey Camacho	740985.83
9	37503	Robin McGuigan	730798.71
10	37155	Stacey Cereghino	727272.65
11	37089	Richard Bready	724299.64
12	37336	Valerie Hendricks	711864.76
13	37006	François Ferrier	700803.79
14	37225	Blaine Dockter	693502.49
15	37422	Anton Kirilov	671618.03
16	37698	Mandy Vance	643745.90
17	37466	Kevin Liu	636226.47
18	37032	John Arthur	618616.13
19	37125	Barbara Calone	617340.46
20	37643	Marcia Sultan	602559.89

Figure 25. Customers whose total orders value is greater than the average value

The table above shows information about customers whose total orders value is greater than the average value of each order, and these customers have 2 or more orders. From there, we can consider the rate of returning customers using the formula total number of customers who buy 2 or more times / total number of customers, thereby making appropriate policies to increase the rate of returning customers, and exploit the potential of existing customers.

5.2.5.2 Total sales return

	Total sales return	Sales return Percent(%)
1	8301	43.42

Figure 26. Number and the percentage of returning customers

The table above provides information about the number of returning customers and the percentage of returning customers compared to the total number of AdventureWork customers.

5.2.6 OLAP use Pivot Table from SSAS:

5.2.6.1 Products Correlations

With cost

Category	Accessories	
Product Name	**Sales Amount**	**Product Cost**
Hitch Rack - 4-Bike	237,096.2	142,090.1
Sport-100 Helmet, Blue	165,406.6	88,240.9
Sport-100 Helmet, Black	160,869.5	85,479.7
Sport-100 Helmet, Red	157,772.4	81,998.8
Hydration Pack - 70 oz.	105,826.4	56,783.6
HL Mountain Tire	48,860.0	18,273.6
Fender Set - Mountain	46,619.6	17,435.7
All-Purpose Bike Stand	39,591.0	14,807.0
ML Mountain Tire	34,818.4	13,022.1
HL Road Tire	27,970.8	10,461.1
Grand Total	**1,024,830.9**	**528,592.6**

Figure 27. Sales amount and product cost by product in Accessories

Category	Bikes	
Product Name	**Sales Amount**	**Product Cost**
Mountain-200 Black, 38	4,400,592.8	3,727,148.3
Mountain-200 Black, 42	4,009,494.8	3,335,278.2
Mountain-200 Silver, 38	3,693,678.0	3,029,893.1
Mountain-200 Silver, 42	3,438,478.9	2,827,394.0
Mountain-200 Silver, 46	3,434,256.9	2,804,612.8
Mountain-200 Black, 46	3,309,673.2	2,642,932.5
Road-150 Red, 56	1,847,818.6	1,441,739.3
Road-150 Red, 62	1,769,096.7	1,302,776.5
Road-150 Red, 48	1,540,803.1	1,070,448.0
Road-150 Red, 52	1,415,563.6	994,452.7
Grand Total	**28,859,456.6**	**23,176,675.5**

Figure 28. Sales amount and product cost by product in Bikes

Category	Clothing	
Product Name	**Sales Amount**	**Product Cost**
Classic Vest, S	156,398.1	100,862.0
Women's Mountain Shorts, S	137,164.1	86,277.1
Women's Mountain Shorts, L	136,774.0	84,915.9
Women's Tights, L	94,090.6	65,671.6
Women's Tights, S	91,330.8	64,094.0
Classic Vest, M	90,250.6	54,242.7
Men's Bib-Shorts, M	86,166.0	59,987.4
Full-Finger Gloves, L	69,943.2	52,936.3
Women's Mountain Shorts, M	57,685.8	29,814.8
Half-Finger Gloves, M	54,545.5	31,727.8
Grand Total	**974,348.8**	**630,529.6**

Figure 29. Sales amount and product cost by product in Clothing

Category	Components	
Product Name	**Sales Amount**	**Product Cost**
HL Mountain Frame - Silver, 38	930,780.7	901,123.4
HL Mountain Frame - Black, 42	901,590.2	872,807.4
ML Mountain Frame - Black, 48	200,284.5	178,200.7
ML Mountain Frame-W - Silver, 40	195,826.4	179,238.8
HL Mountain Rear Wheel	166,013.3	123,491.0
HL Crankset	148,622.6	110,227.0
HL Mountain Frame - Silver, 48	141,635.1	122,278.3
ML Mountain Rear Wheel	119,892.2	89,180.6
HL Road Front Wheel	112,286.4	83,091.9
ML Road Front Wheel	79,354.0	58,670.5
Grand Total	**2,996,285.4**	**2,718,309.6**

Figure 30. Sales amount and product cost by product in Components

The 4 tables above show the top 10 highest-grossing products by category. We can see that for Category Bikes, Accessories and Components, Product Cost is proportional to Sales Amount, meaning that the higher the product cost, the higher the revenue while in the remaining category - Clothing, the product cost has no effect on product sales.

Color

	Year	Product Color	Order Quantity
1	2011	Red	5251
2	2011	Black	3491
3	2011	Multi	1572
4	2011	Silver	1499
5	2011	White	674
6	2011	Blue	360
7	2012	Black	28797
8	2012	Red	13562
9	2012	Multi	7940
10	2012	Silver	3994
11	2012	Yellow	3282
12	2012	Blue	1425
13	2012	White	523
14	2013	Black	36011
15	2013	Yellow	18399
16	2013	Blue	13251
17	2013	Silver	12635
18	2013	Multi	10823
19	2013	Red	8302
20	2013	Silver/Black	2606
21	2013	White	2481
22	2014	Black	13638
23	2014	Yellow	10875
24	2014	Blue	8623
25	2014	Silver	6895
26	2014	Multi	4738
27	2014	Red	2114
28	2014	White	1539
29	2014	Silver/Black	1325

Figure 31. Product color with order quantity by year

The table above mentions the top Product color which has a large Order quantity each year. It also shows us that the red color drew customers' attention more than other colors in 2011. However, in the year that followed, the black color was crowned with Order quantity seems very higher than others (in 2014, order quantity is just a bit higher instead).

Category SubCategory Model:

Order Qty	Year				
Product Name	Bikes	Clothing	Accessories	Components	Grand Total
2011	7889	2246	1032	1647	12814
2012	25832	17797	5446	14327	63402
2013	37541	36484	30545	24303	128873
2014	19006	17143	24909	8767	69825
Grand Total	**90268**	**73670**	**61932**	**49044**	**274914**

Figure 32. Order quantity of each Category by year

The table above shows the Order quantity of each Category from 2011 to 2014. Bikes always have the most order quantity, the next are clothing, the third are accessories but its demand was the most in 2014.

Row Labels	Order Qty
Road Bikes	47196
Mountain Bikes	28321
Jerseys	22711
Helmets	19541
Tires and Tubes	18006
Touring Bikes	14751
Gloves	13012
Road Frames	11753
Mountain Frames	11621
Bottles and Cages	10552
Grand Total	**197464**

Figure 33. Top 10 Order quantity by Subcategory

The table above indicates the top 10 Order quantity by Subcategory and Road Bikes is on top 1.

Row Labels	Order Qty
Sport-100	19541
Road-650	17609
Mountain-200	14596
Long-Sleeve Logo Jersey	13637
Short-Sleeve Classic Jersey	9074
Road-250	8587
Cycling Cap	8311
Women's Mountain Shorts	7679
Road-550-W	7183
Half-Finger Gloves	6928
Grand Total	**113145**

Figure 34. Top 10 Order quantity by Model

The table above indicates the top 10 Order quantity by Model and Sport-100 is on top 1.

5.2.6.2 Product Routing

Efficiency

```
SELECT CAST(
        COUNT(*) * 100. /
        (SELECT COUNT(*)
        FROM FACTWORKORDERROUTING) AS DECIMAL(9,2)
    ) AS '(%) ACTUAL COST / PLANNED COST'
FROM FACTWORKORDERROUTING
WHERE PLANNEDCOST = ACTUALCOST
```

(%) Actual Cost / Planned Cost
100.00

Figure 35. 100% Actual cost equal to Planned cost of each product

The Portion of scrapped product

```
SELECT
    MODEL,
    SUM(SCRAPPEDQUANTITY) AS 'TOTAL SCRAPPED QUANTITY'
FROM DIMPRODUCT A
JOIN FACTWORKORDERROUTING B ON A.PRODUCTID = B.PRODUCTID
GROUP BY MODEL

    ORDER BY SUM(SCRAPPEDQUANTITY) DESC
```

Scrapped products quantity categorized into models

	model	Total Scrapped Quantity
1	Front Derailleur	348
2	LL Fork	280
3	ML Crankset	225
4	LL Road Handlebars	189
5	ML Headset	154
6	HL Mountain Handlebars	138
7	LL Bottom Bracket	107
8	Rear Derailleur	105
9	ML Mountain Front Wheel	100
10	Touring Rear Wheel	87
11	HL Fork	76
12	HL Bottom Bracket	75
13	ML Road Handlebars	69
14	HL Mountain Frame	60
15	ML Road Rear Wheel	58
16	ML Road Frame-W	54

Figure 36. Top 16 Models by Scrapped Quantity

The chart above indicates top 16 Models which have large Scrapped Quantity. It helps us figure out the model need to be improve in production process and address thoroughly the issue in a timely manner

Scrapped products quantity categorized into category and subcategory

```
SELECT
        CATEGORY,
        SUM(SCRAPPEDQUANTITY) AS 'TOTAL SCRAPPED QUANTITY'
FROM DIMPRODUCT A
JOIN FACTWORKORDERROUTING B ON A.PRODUCTID = B.PRODUCTID
GROUP BY CATEGORY
        ORDER BY SUM(SCRAPPEDQUANTITY) DESC
```

	Category	Total Scrapped Quantity
1	Components	2648
2	Bikes	148

Figure 37. Scrapped products quantity categorized into Category

```
SELECT
        SUBCATEGORY,
        SUM(SCRAPPEDQUANTITY) AS 'TOTAL SCRAPPED QUANTITY'
FROM DIMPRODUCT A
JOIN FACTWORKORDERROUTING B ON A.PRODUCTID=B.PRODUCTID
GROUP BY SUBCATEGORY
ORDER BY SUM(SCRAPPEDQUANTITY) DESC
```

	Subcategory	Total Scrapped Quantity
1	Handlebars	522
2	Wheels	474
3	Derailleurs	453
4	Forks	398
5	Cranksets	255
6	Headsets	230
7	Bottom Brackets	196
8	Road Bikes	89
9	Mountain Frames	66
10	Road Frames	54
11	Mountain Bikes	43
12	Touring Bikes	16

Figure 38. Scrapped products quantity categorized into Subcategory

The chart shows the number of product categories and subcategories damaged. It is mostly focused on components which require meticulousness and of course the machine is not about doing that stuff well.

Routing effectiveness

```
SELECT CAST(
        COUNT(*) * 100. /
            (SELECT COUNT(*)
            FROM FACTWORKORDERROUTING) AS
DECIMAL(9,2)
    ) AS '(%) TOTAL WORK ORDERS DELIVERED LATELY / TOTAL
WORKORDER'
FROM FACTWORKORDERROUTING
WHERE SCHEDULEDENDDATE != ACTUALENDDATE
```

(%) total work orders delivered lately / total WorkOrder
58.69

Figure 39. Percentage of Delayed work order

The number above shows that we got numerous delayed WorkOrder, 58% is not eligible for delivery standards in a large company like Adventure Work Co.

Factory	Number of delayed Work Order	(%) Delayed WorkOrder in each Factory / Total Delayed WorkOrder
5	420	1.23
4	515	1.51
3	935	2.74
2	3492	10.25
1	5558	16.32
7	8494	24.94
6	14648	43.00

Figure 40. Factory, number and the rate between delayed Work order of each Factory and total delayed Work Order

The chart above shows the number of delayed Work Order, which has actual end date were later than the scheduled end date. It indicates the factory and the rate between delayed Work order of each Factory and total delayed Work Order

We got fortunate when not having any delayed Start date (100% Actual start date = Scheduled Start date). However, we don't have any Work Order to be completed before Scheduled End date (0% Actual End Date < Scheduled End Date)

CHAPTER 6: VISUALIZATION AND FORECASTING OR PREDICTIVE MODEL

A data visualization increases the project's diversity and effectiveness. Below are some worksheets and dashboards that our team created using various BI tools to have a better understanding of the production department.

6.1. Report and dashboard systems (structure):

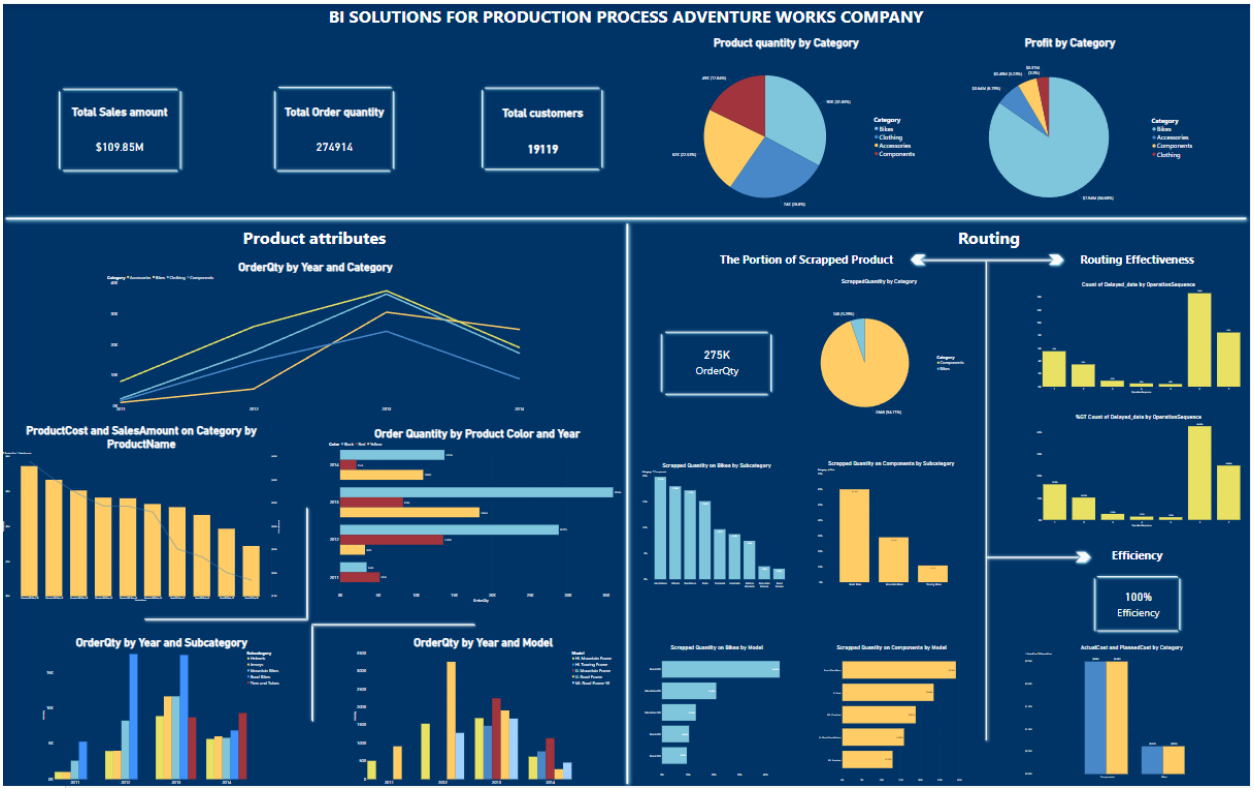


Figure 41. Reports and dashboard systems from PowerBI

6.2. Data analysis with the Pivot Table tool in Microsoft Excel

Order Qty	Year				
Product Name	2011	2012	2013	2014	Grand Total
AWC Logo Cap					
Clothing	545	1913	3745	2108	8311
Water Bottle - 30 oz.					
Accessories			3705	3110	6815
Sport-100 Helmet, Blue					
Accessories	360	1425	3041	1917	6743
Long-Sleeve Logo Jersey, L					
Clothing	544	1970	2921	1157	6592
Sport-100 Helmet, Black					
Accessories	341	1319	3010	1862	6532
Sport-100 Helmet, Red					
Accessories	331	1224	2894	1817	6266
Classic Vest, S					
Clothing			2615	1632	4247
Patch Kit/8 Patches					
Accessories			2036	1829	3865
Short-Sleeve Classic Jersey, XL					
Clothing			2416	1448	3864
Long-Sleeve Logo Jersey, M					
Clothing	228	1038	1649	721	3636
Grand Total	**2349**	**8889**	**28032**	**17601**	**56871**

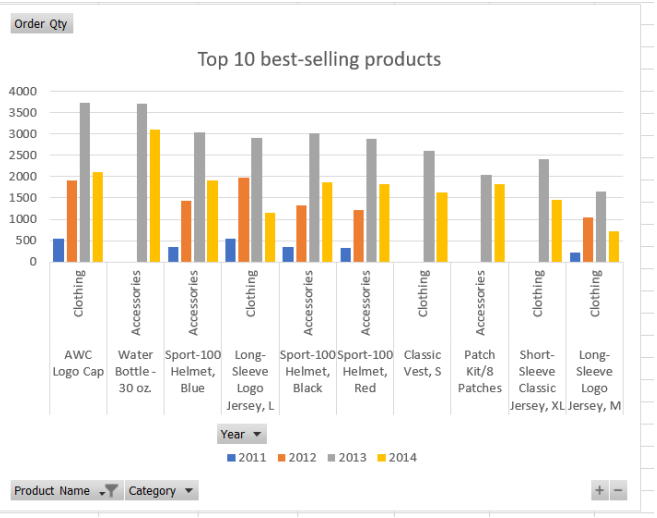


Figure 42. Data analysis with Pivot table tool

The pivot table above informs us of the top 10 best-selling products in 4 years which all belong to two categories Clothing and Accessories. This is an understandable trend which we will explain later on.

Order Qty	Year				
Product Name	Accessories	Bikes	Clothing	Components	Grand Total
2011	1032	7922	2246	1647	12847
2012	5446	25841	17797	14327	63411
2013	30986	37623	36602	24303	129514
2014	24468	18882	17025	8767	69142
Grand Total	**61932**	**90268**	**73670**	**49044**	**274914**

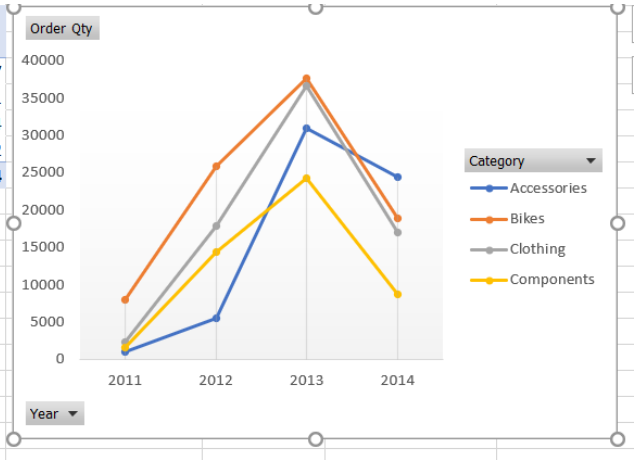


Figure 43. Data analysis with Pivot table tool (cont)

This line chart from the pivot table describes the sales trend for each category, apparently, all the 4 reach their highest point of sales in 2013 and spares slight differences to each other, let alone accessories as it surged more quickly than other three.

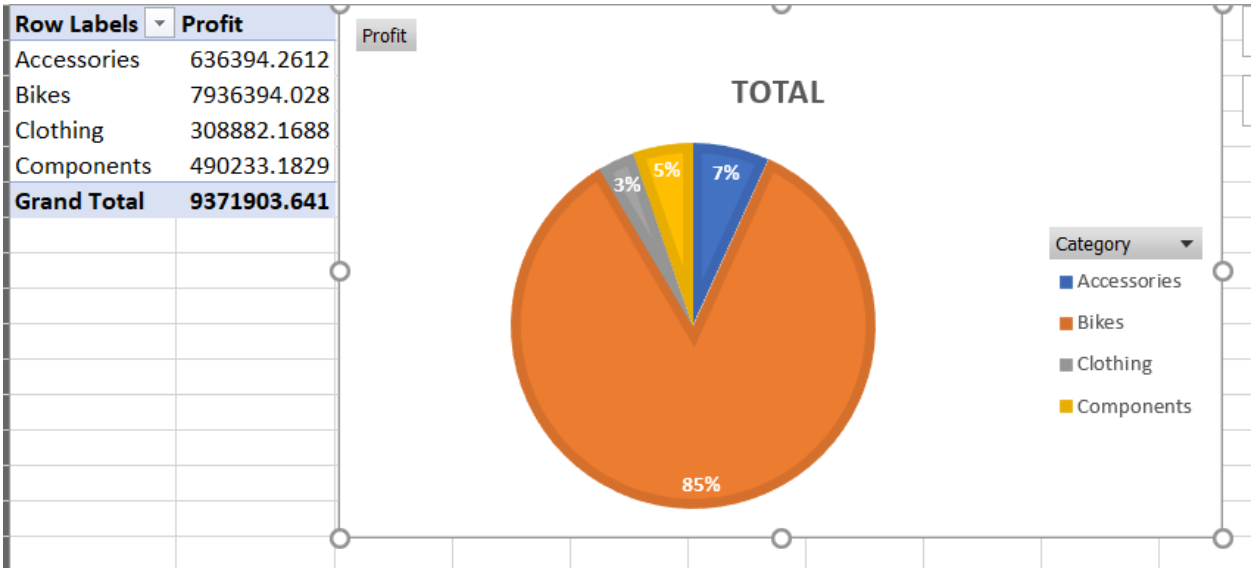


Figure 44. Data analysis with Pivot table tool (cont)

6.3. Data analysis with Power BI:

As the dashboard from the top revealed some astonishing findings supporting our project in wrapping the whole background of Adventures Work's Product categories issues.

It is shown evidently that in 4 years on record, the company reached total sales of 109.85 Million dollars came from 275k orders sold to 19119 customers. The two next pie charts which illustrated each category sold units and profit they brought to the company.

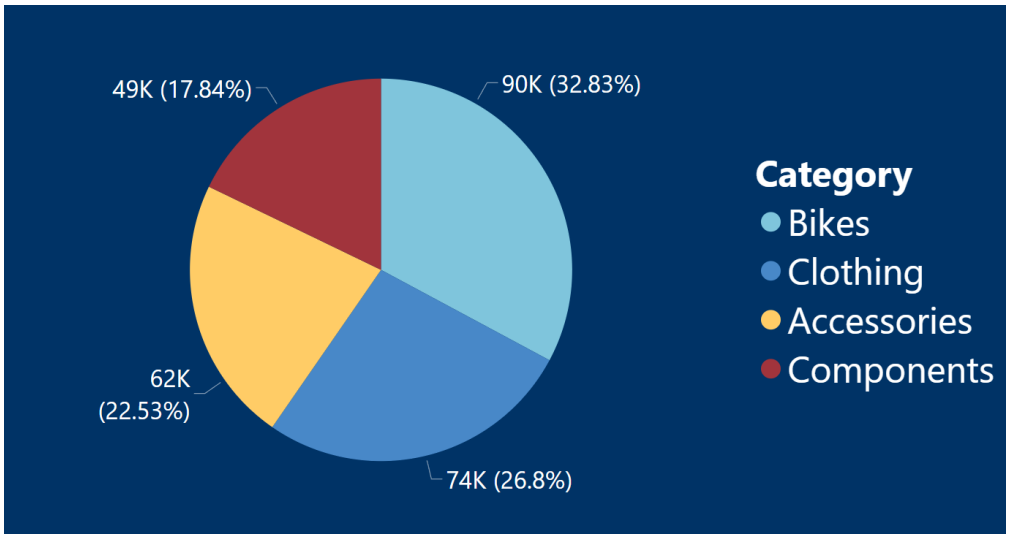


Figure 45. Product quantity by Category

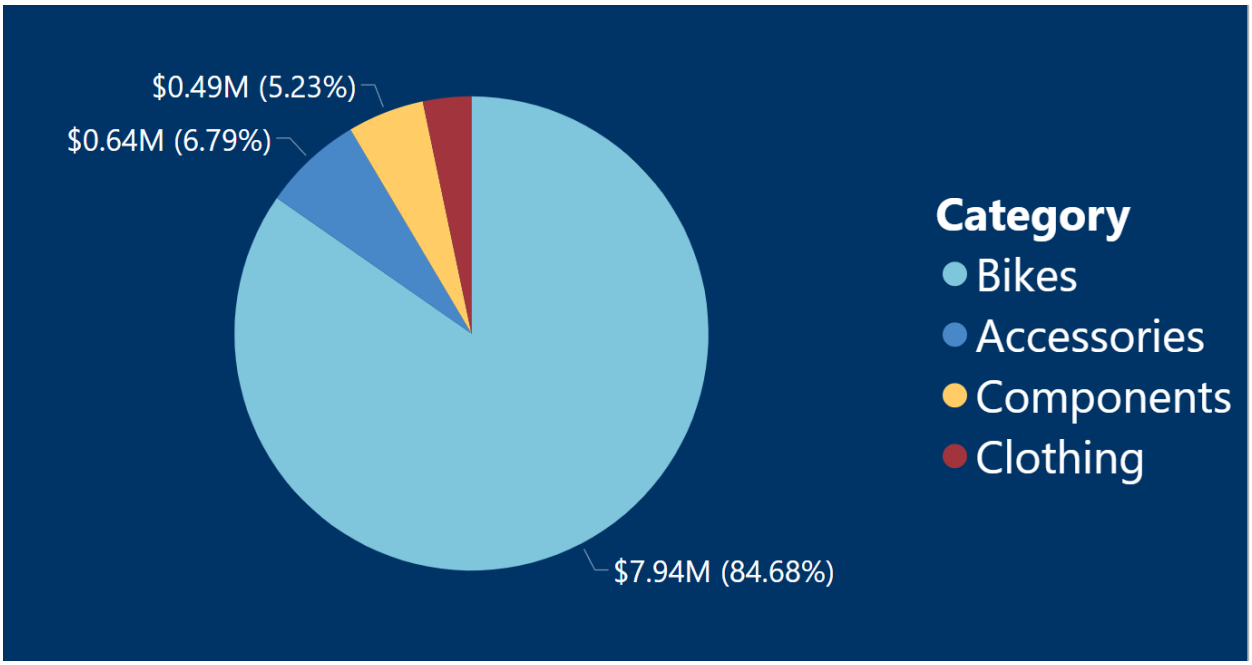


Figure 46. Profit by Category

As observed from the first chart, we can see that each category sold units was not comparatively different, though bikes still took the largest proportion of total quantities sold (32,83%). In stark contrast, the second chart highlighted the dominant portion of profit brought by bikes which stated that the category Bikes contributed 84,68% of the company total profit (before tax). Needless to say that bikes are what Adventure Work lives for.

Another thing to mention is that all category sold quantities experience the same trend in a 4-years period.

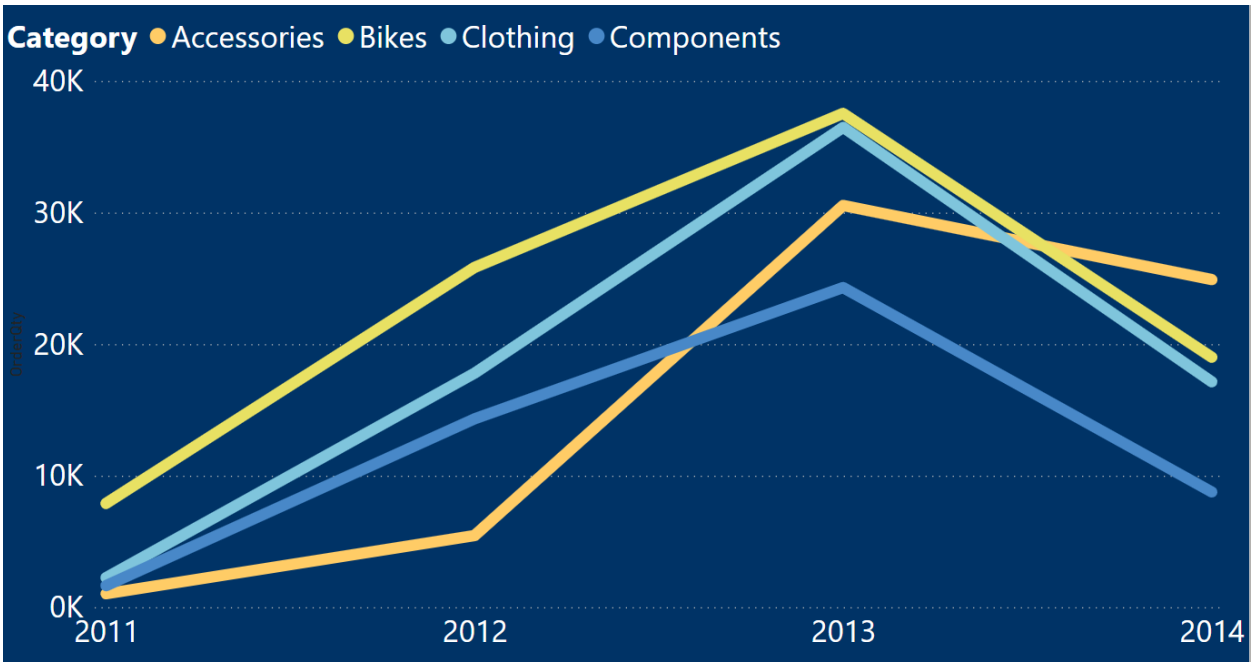


Figure 47. Order quantity by Year and Category

As the data has started to be recorded since 2011, the 4 categories increased dramatically in the first two years and reached the highest point of sales in 2013. In the next year, 2014, brought a steady decline in this figure for all categories. This has led our group to suspect what was happening with the 4 categories and what might pose threats to the company' survival if this trend continued.

6.3.1. Products Components correlates with theirs sold quantities:

Taking everything into account, we have noticed that there are some possible elements of a product that may affect its profitability.

6.3.1.1. Standard Cost

From the visualization dashboard we can see that Product Cost has been related to the sales of that product. For all three categories including Bikes,

Accessories and Components, Product Cost changed in a descending direction, which also meant that revenue declined proportionally.

We saw that the cost of the product determines the level of revenue that the product brings, for example, if the total cost of the "Road-150 Red, 56" product is $1.4 million, the revenue is about $1.8 million while the product "Mountain-200 Black, 38" has a total product cost of $3.7 million, sales amount is $4.4 million which meant that the higher the cost, the higher the Sales Amount is.

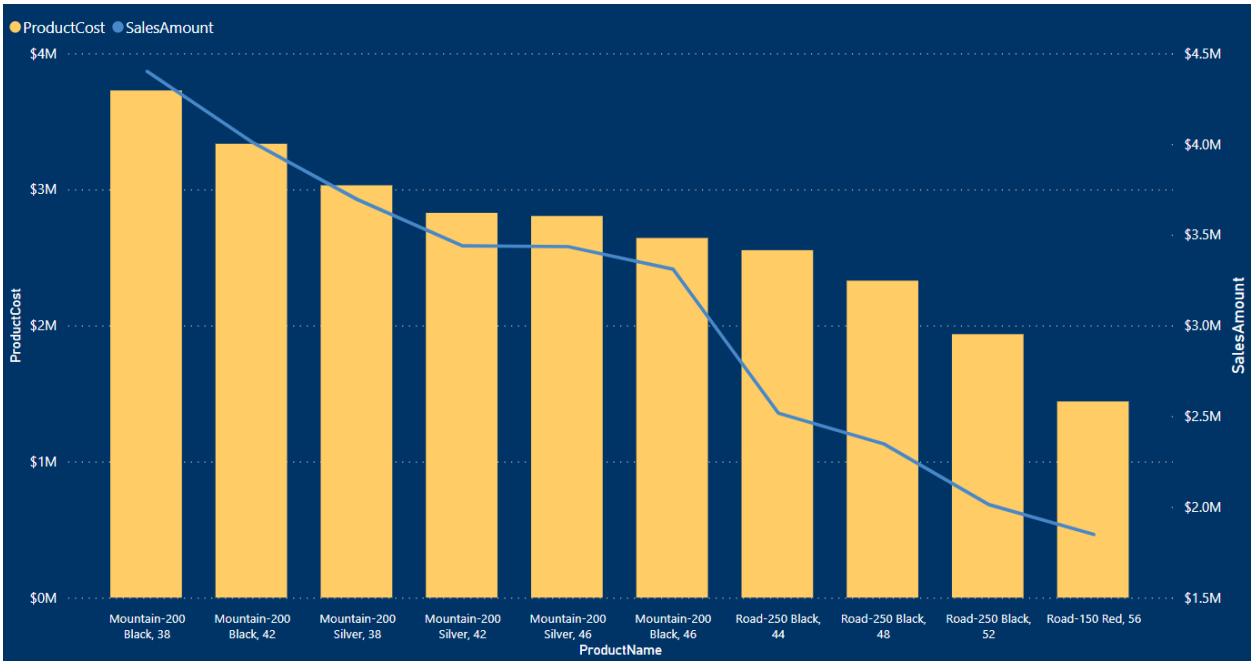


Figure 48. ProductCost and SalesAmount on Bikes by ProductName

As for the Clothing category, Product Cost did not seem to have an impact on product sales. This is reflected in the fact that the Sales Amount line rises and falls continuously rather than a downward or upward line. For example, the product column "Classic Vest, S", with a total product cost of about $100k while revenue is about $156k, obviously if ranked in the top 10 products by Sales Amount, "Long-Sleeve Logo Jersey, XL" had a higher product cost ($114k) but lower revenue (about $95.6k) than "Classic Vest, S".

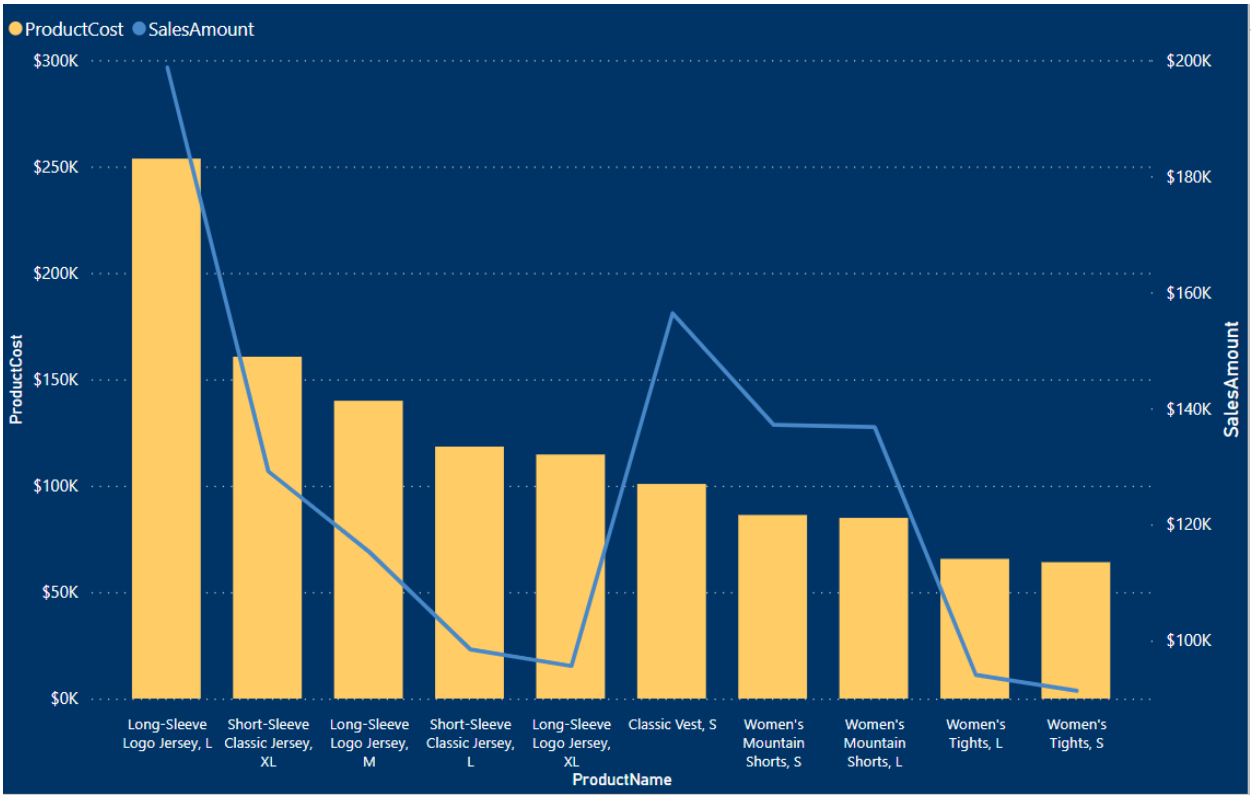


Figure 49. ProductCost and SalesAmount on Clothing by ProductName

This reflected the customer's demand for that product, not the cost, which determines most of the revenue. Managers can rely on this insight to make appropriate Standard Cost decisions depending on the type of product.

6.3.1.2. Color

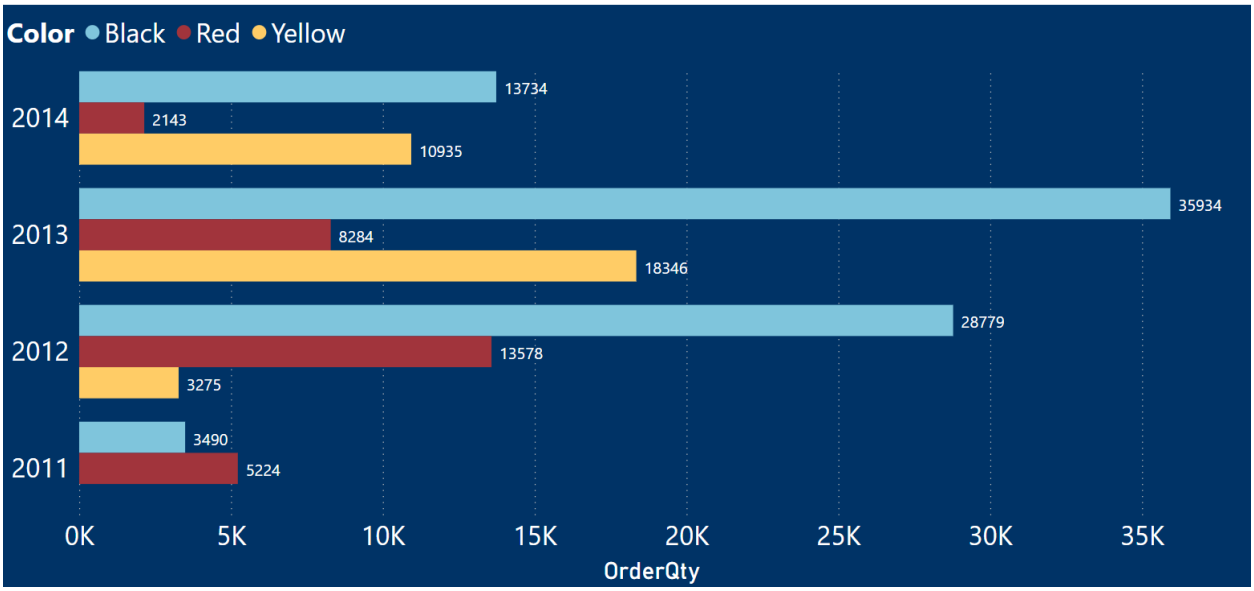


Figure 50. Order quantity by Product color and Year

The stacked chart above shows top 3 color product bestsellers from 2011 to 2014. The black products have been the best sold since 2012. On the contrary, the red products sold decrease year by year since 2012. From there, we should consider whether to promote the development of yellow products or not when the number of products sold increases each year and have trend to replace the proportion of red products.

6.3.1.3. Subcategory:

Accessories

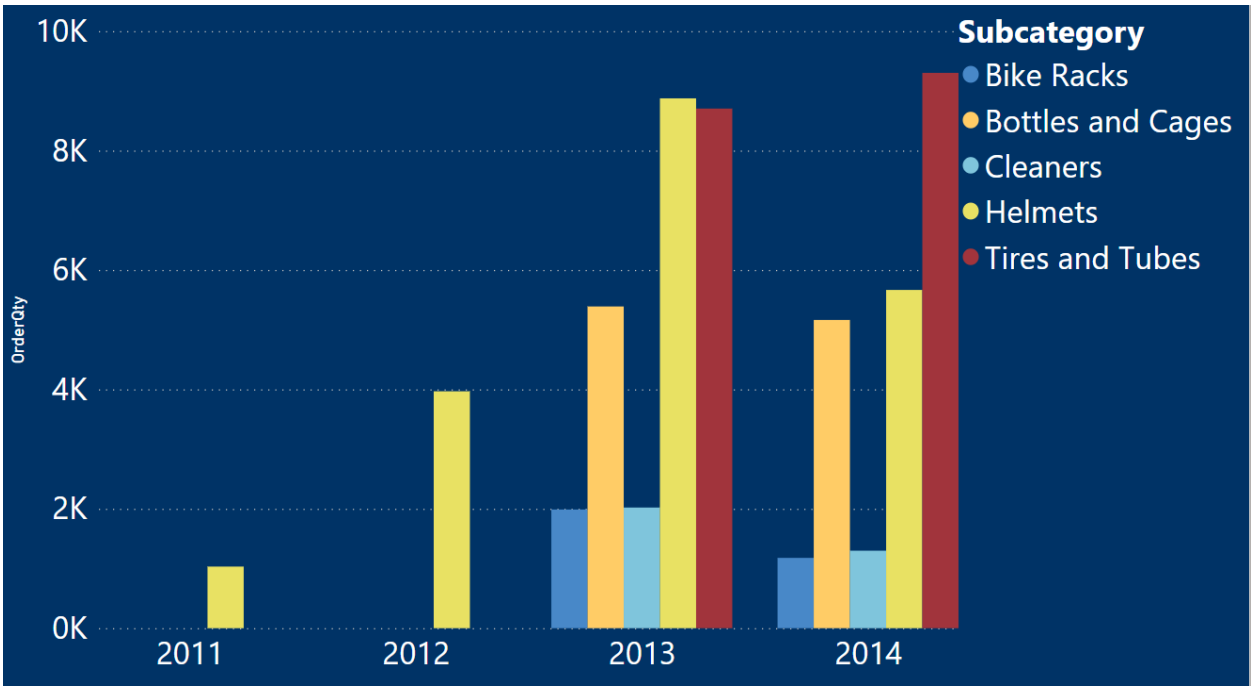


Figure 51. Order quantity by Year and Subcategory (Accessories)

The clustered column chart above shows that accessories amounts were sold from 2011 to 2014. The accessories of AdventureWorks have been diversified when it has only Helmet in 2011 and 2012 but since 2014, more accessories have been added to the product list. Not only that, the number of purchased accessories also increased sharply over the years. This can provide a strategic direction for the development of the sales of accessories for the company

Components

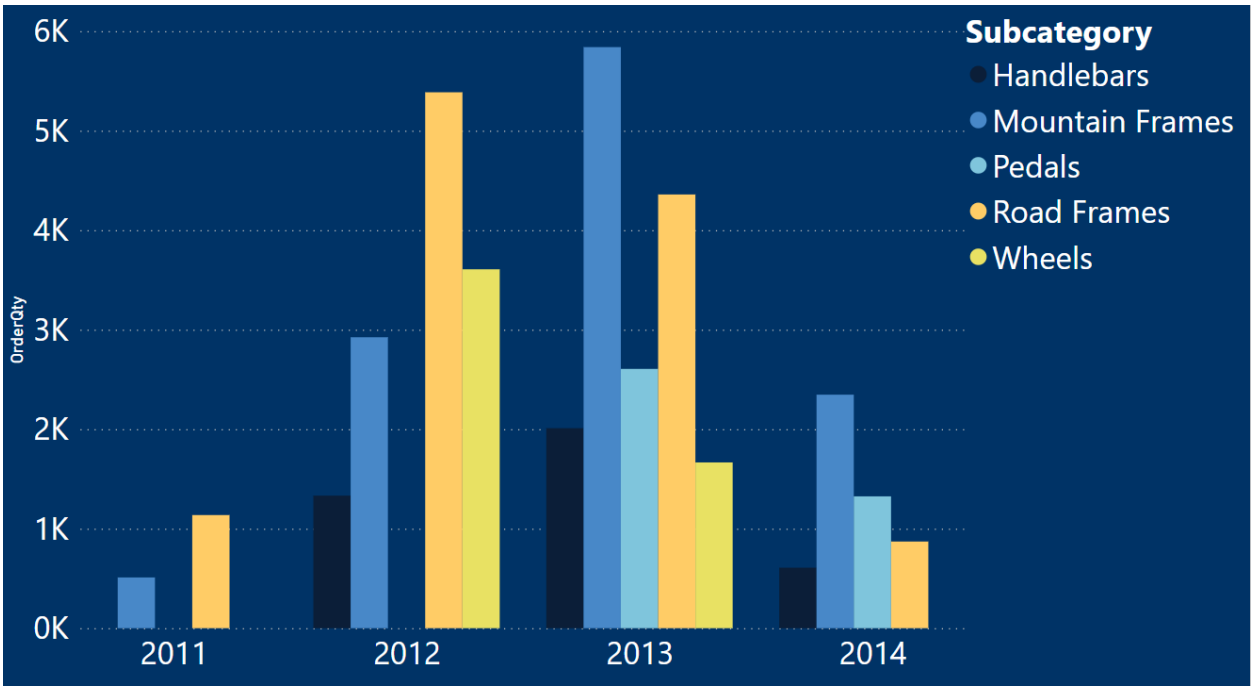


Figure 52. Order quantity by Year and Subcategory (Components)

The chart above provides information for the best-selling retail bicycle components from 2011 to 2014. From the chart above it is easy to see the best-selling are Mountain and Road bike frames. The objective reason may be that the amount of frames purchased is due to the customer's need to redesign their own bicycles. For cyclists, when buying high-priced bikes and to satisfy themselves, they often tend to buy parts such as bicycle frames to redesign themselves according to their own style. For example, redesigning the bicycle based on ordinary steel or aluminum monolithic materials to ensure lightness and durability; choose frames that have a different color from the original or have an angular design;…

Clothing

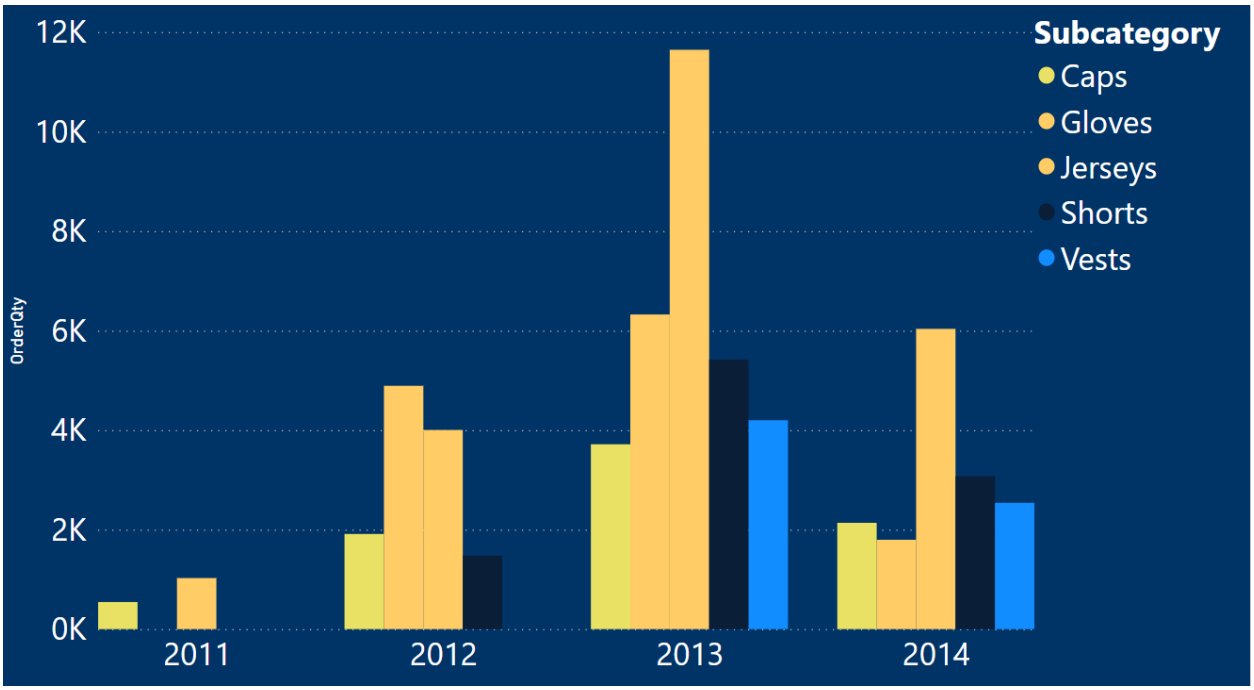


Figure 53. Order quantity by Year and Subcategory (Clothing)

The products of clothing for cyclists are also diversified over the years. The large volume of Jerseys sold also proves that customers tend to buy it when they buy bicycles, so the company can further diversify in terms of clothing designs to serve user demand and at the same time promoting the increase in revenue from the sale of clothing.

Bikes

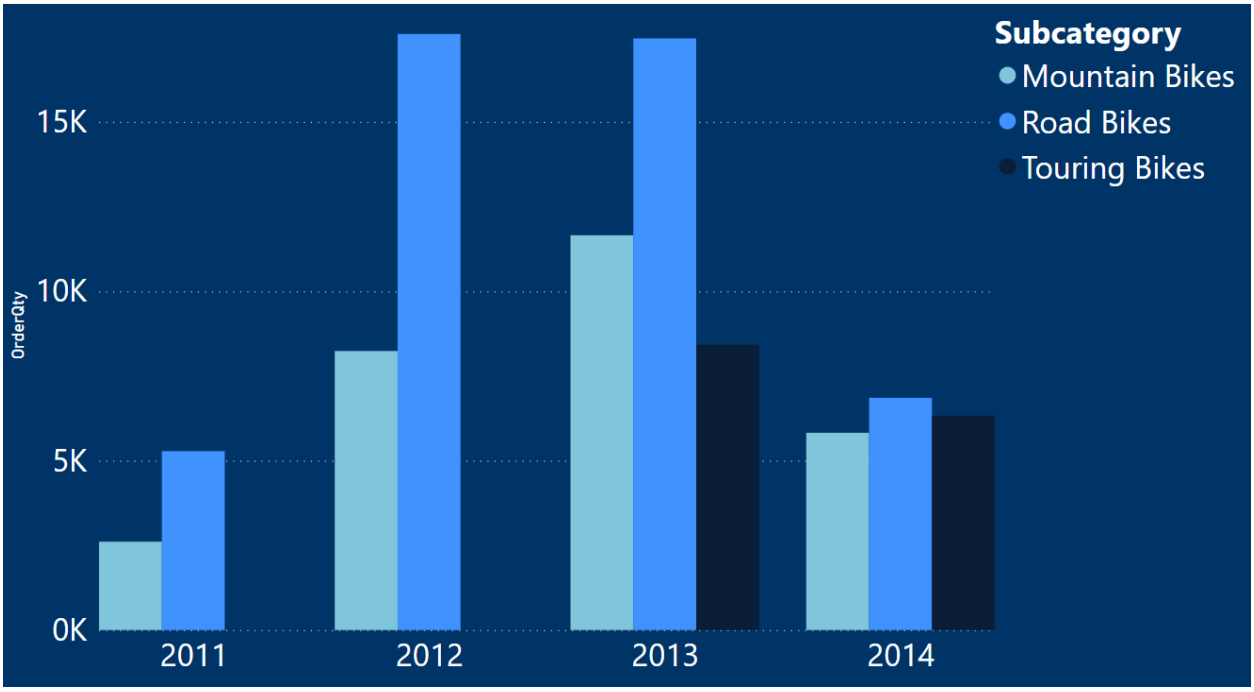


Figure 54. Order quantity by Year and Subcategory (Bikes)

The number of sales in 2012 and 2013 increased and there was a new product called Touring bike from 2013. For the sales amount, there was a gradual decrease in 2014, but the rate of decrease of the product line Touting is not too high compared to the previous year, while the purchased rate of 2 product lines Road and Mountain has decreased markedly.

6.3.1.4 Model:

Accessories

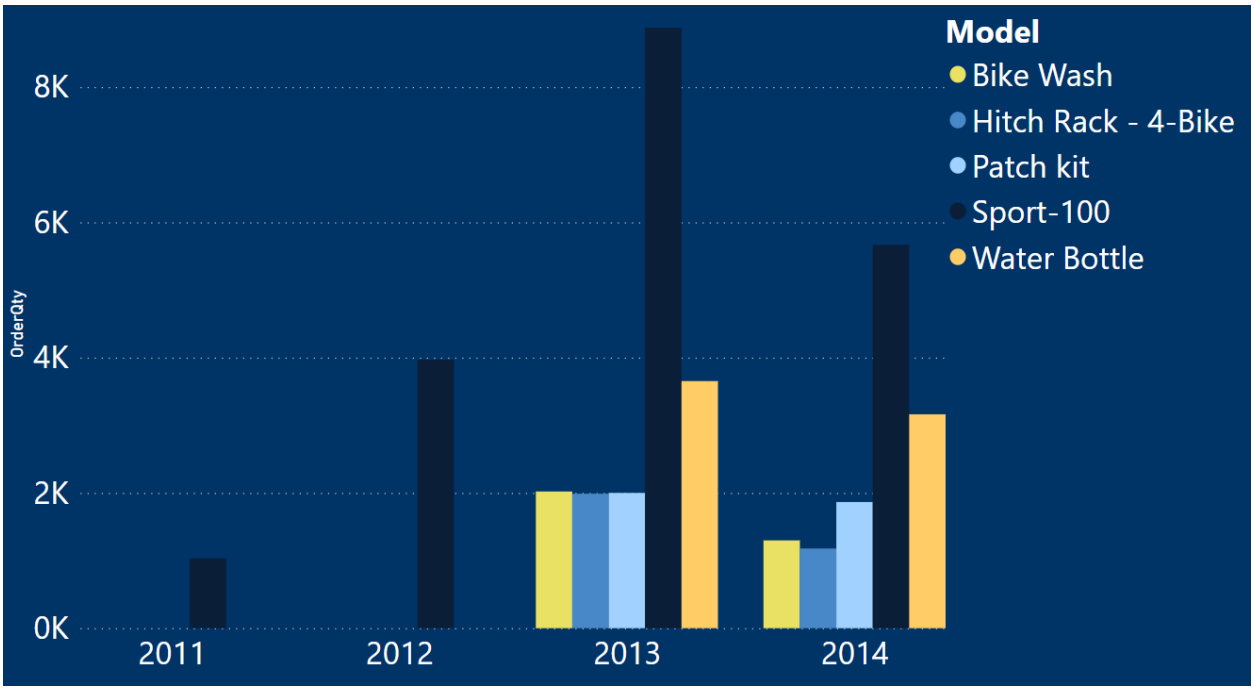


Figure 55. Order quantity by Year and Model (Accessories)

From the clustered bar chart above, we easily identify the most sold model from accessories is the Sport - 100, a type of helmet, which is really necessary for cycling as now people would pay more cautious to their safety. This helmet places the first rank every year despite the establishment of new impressive accessories models such as water bottle, patch kit, hitch rack and bike wash.

Components

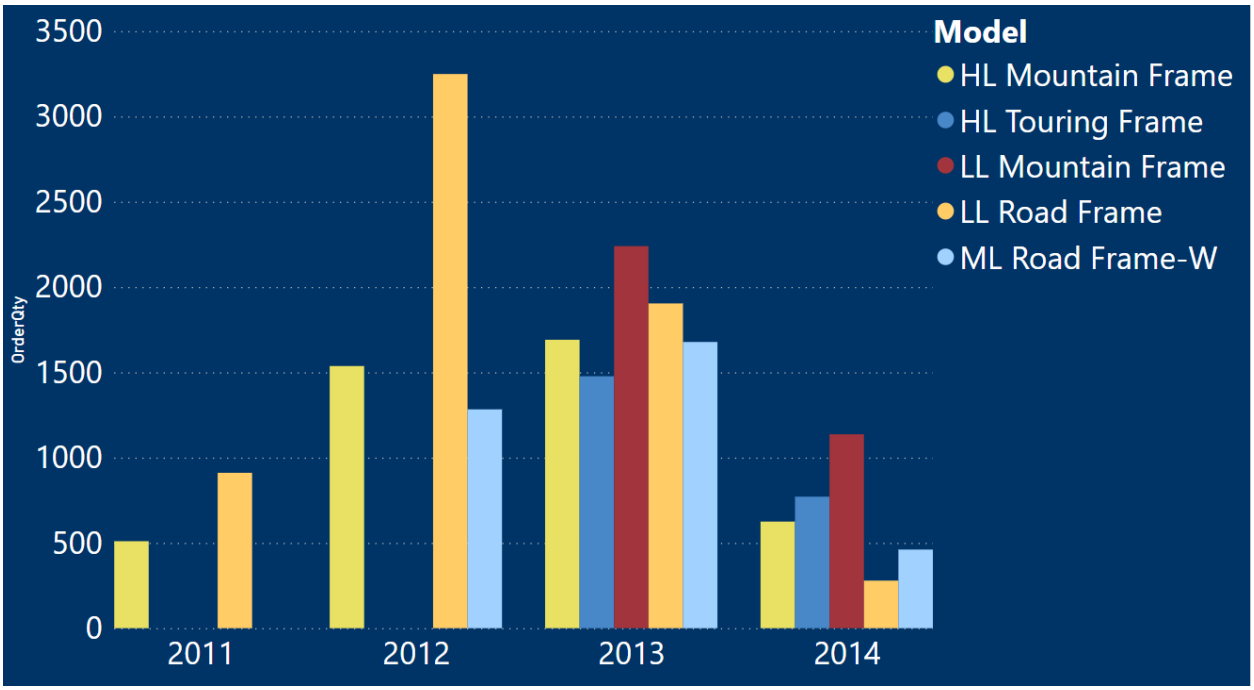


Figure 56. Order quantity by Year and Model (Components)

It is undeniable that the need for frames still ascends continually throughout the 4 year. Being the best-selling in both 2011 and 2012, LL Road Frame just now was the second-best-selling in 2013 and then fell slightly behind in 2014. An explanation for this could be new models of frames appearance in 2013 (especially the LL Mountain Frame) being helpful in meeting the customer demand in upgrading their bikes

Clothing

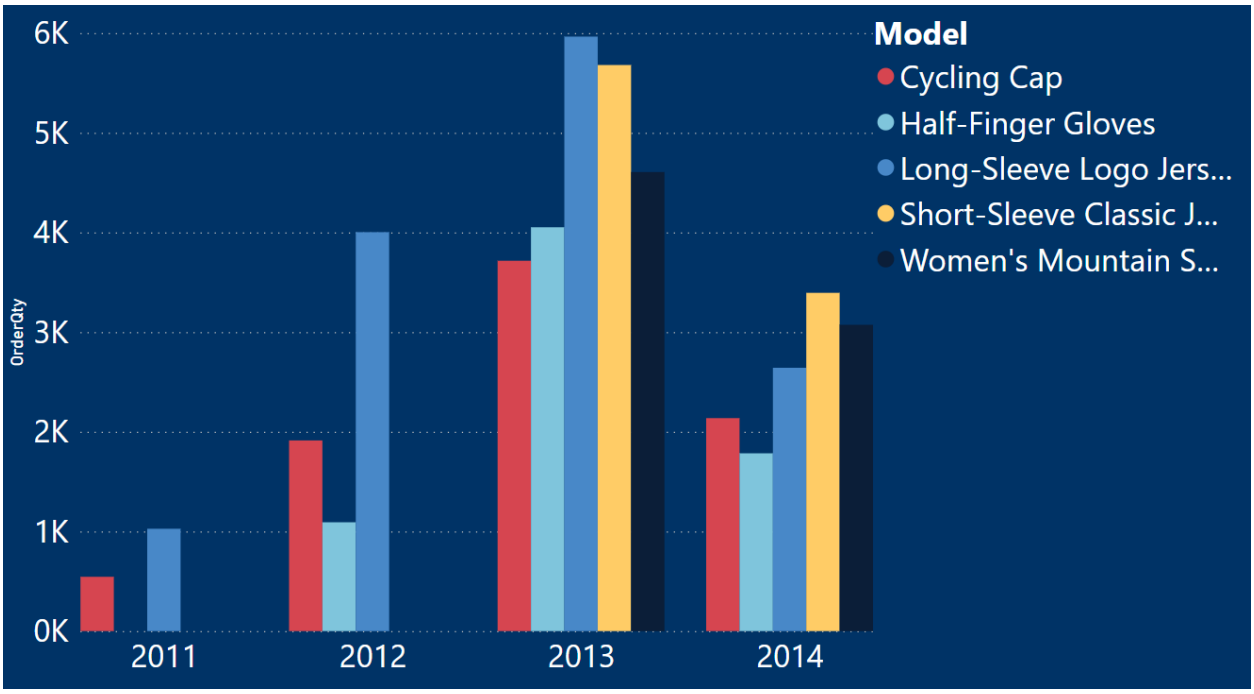


Figure 57. Order quantity by Year and Model (Clothing)

Apparently from the chart above, the clothing experience total sold quantities experience the same trend with the two above. As the new product began to launch, the old ones gradually lost their sales.

Bikes

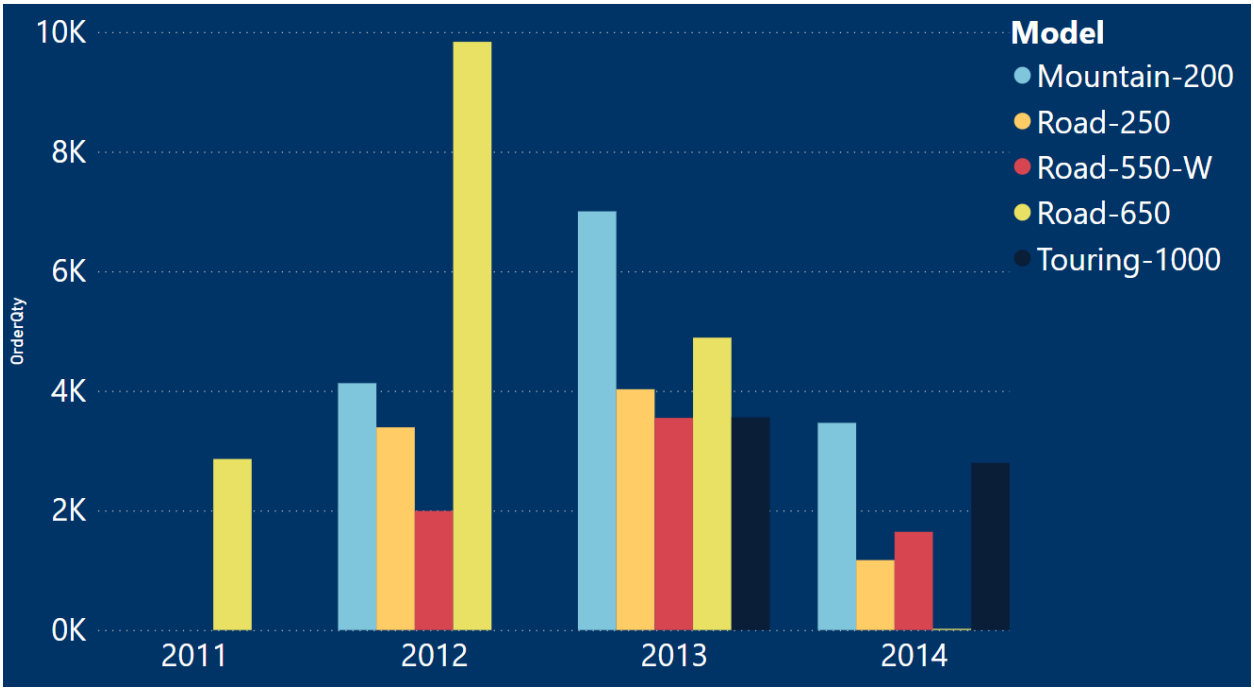


Figure 58. Order quantity by Year and Model (Bikes)

The chart above also contributed to the conclusion that new products pose a real threat to the former ones. While being the best-selling in 2012 with a leading portion, the model Road-650 sales went down sharply and then hit a low in 2014.

6.3.2. Product Routing Effectiveness

6.3.2.1 Portion of scrapped products

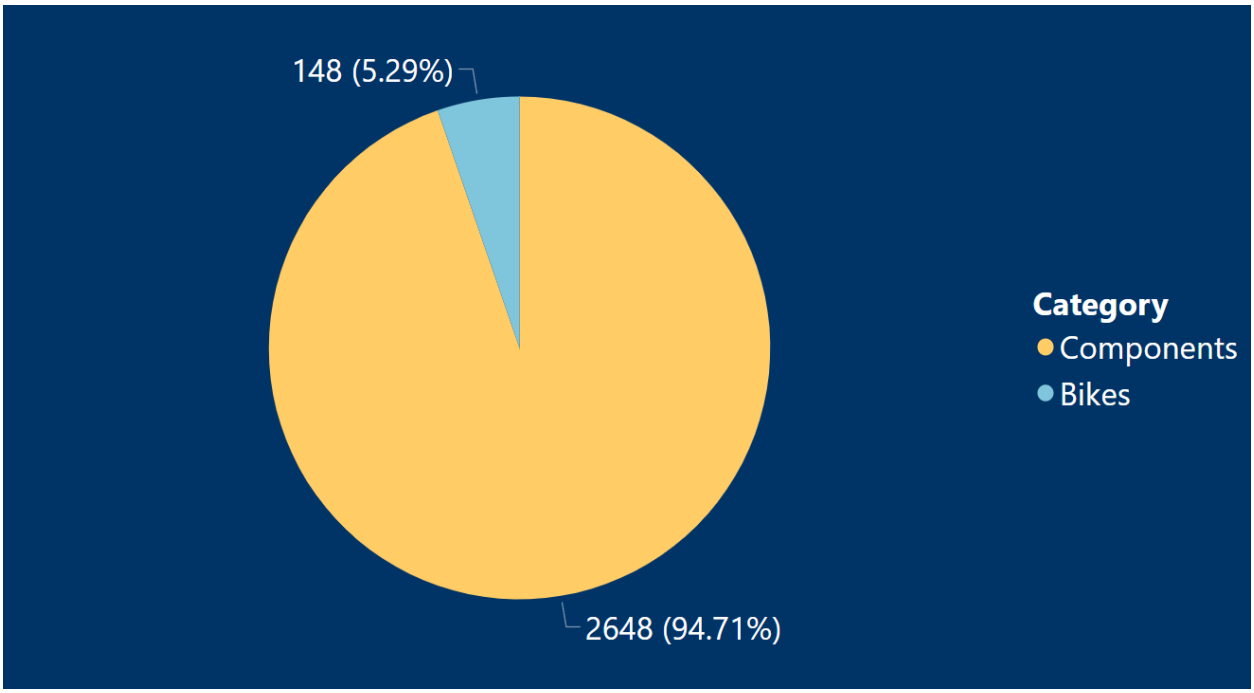


Figure 59. Scrapped quantity by Category

Pie chart above indicates the percentage of Scrapped products (Components and Bikes). The components were badly damaged. Above 94 percent of Scrapped products were components. This can be explained by a critical reason, components required tough technique skills to be completed. Meanwhile, the bikes, the main products taking up to nearly 85% profit, were created by a fine-tuning process.

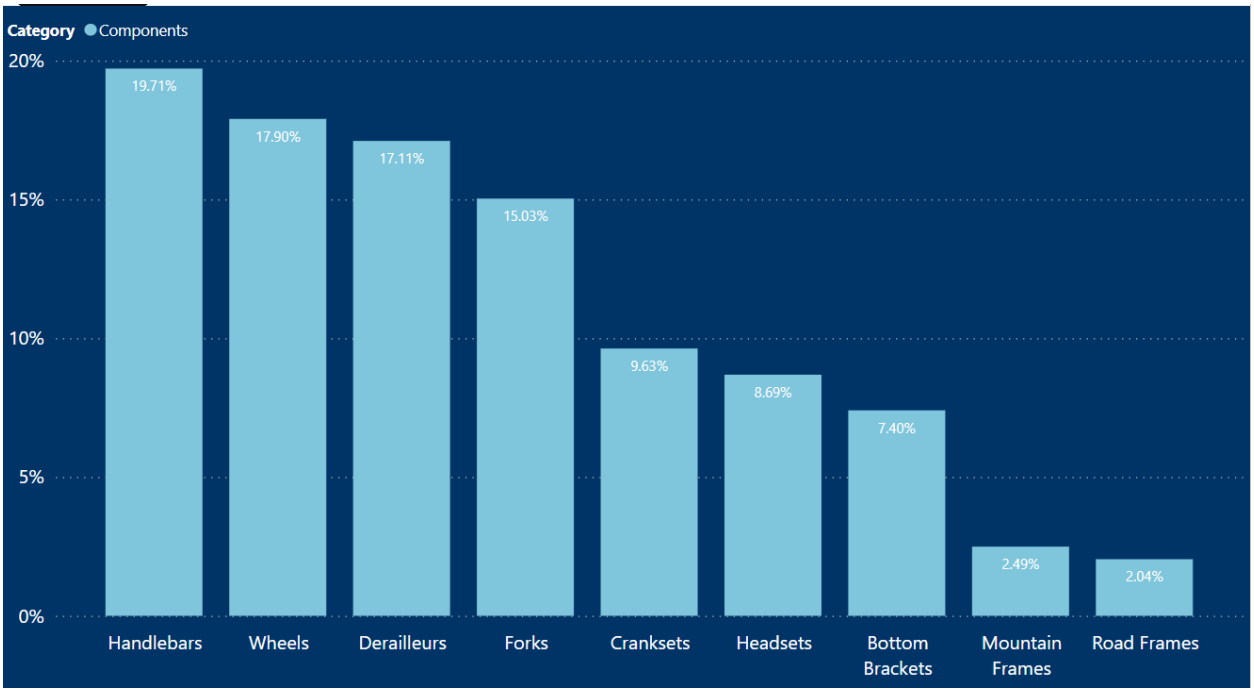


Figure 60. Scrapped quantity on Bikes by Subcategory

Diving into Scrapped components with this column chart. Top 5 components mostly damaged take up to 75.2% of total scrapped products. As explained above, top 5 components really need high-level skills to be created.

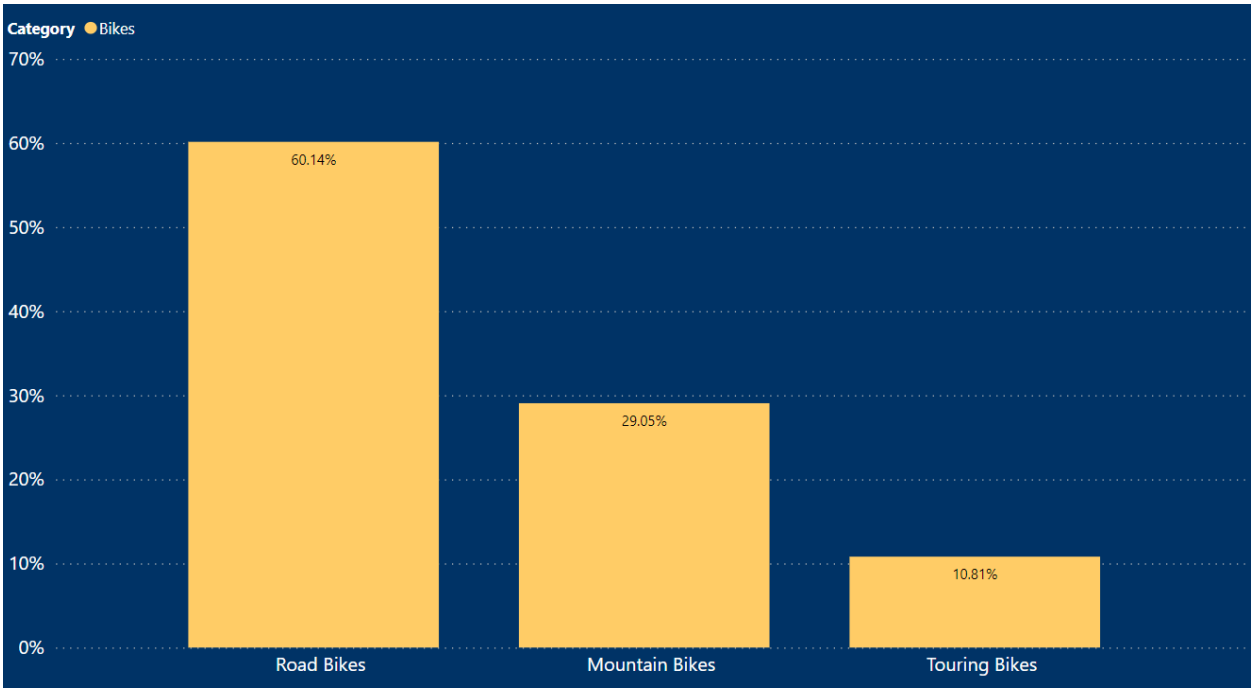


Figure 61. Scrapped quantity on Components by Subcategory

The chart above was only the quantity of scrapped Bikes. RoadBike is one of the bikes that is very difficult to be perfect, it took up to 60% total scrapped bikes, trailing behind Mountain bikes which took 29%. European people tended to buy road bikes instead of touring bikes, because of the huge impact of the 100th Tour De France and ⅔ Europe area is plain. That created road bicycle habit people there. Order quantities were high, so the number of scrapped bikes also increased.

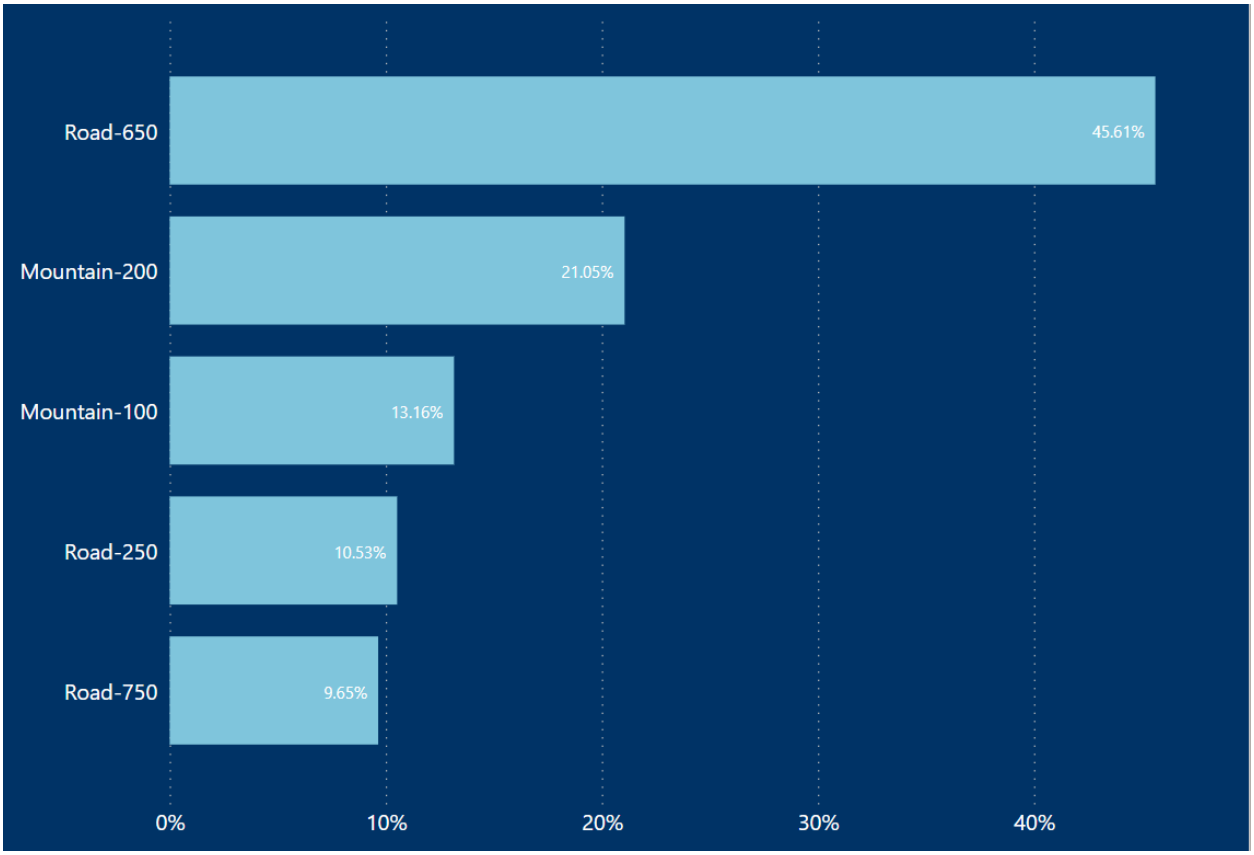


Figure 62. Scrapped quantity on Bikes by Model

The Chart above shows the percentage of scrapped bikes divided into models. Road-650, the 2nd best selling bike, but took up to 46% of total scrapped quantity and more than twice of the 2nd scrapped quantity (Mountain-200). Trailing behind, the second and the third are Mountain bikes (Model: 200 and 100). There were no touring bike models in top 5 scrapped bikes.

6.3.2.2. Routing effectiveness

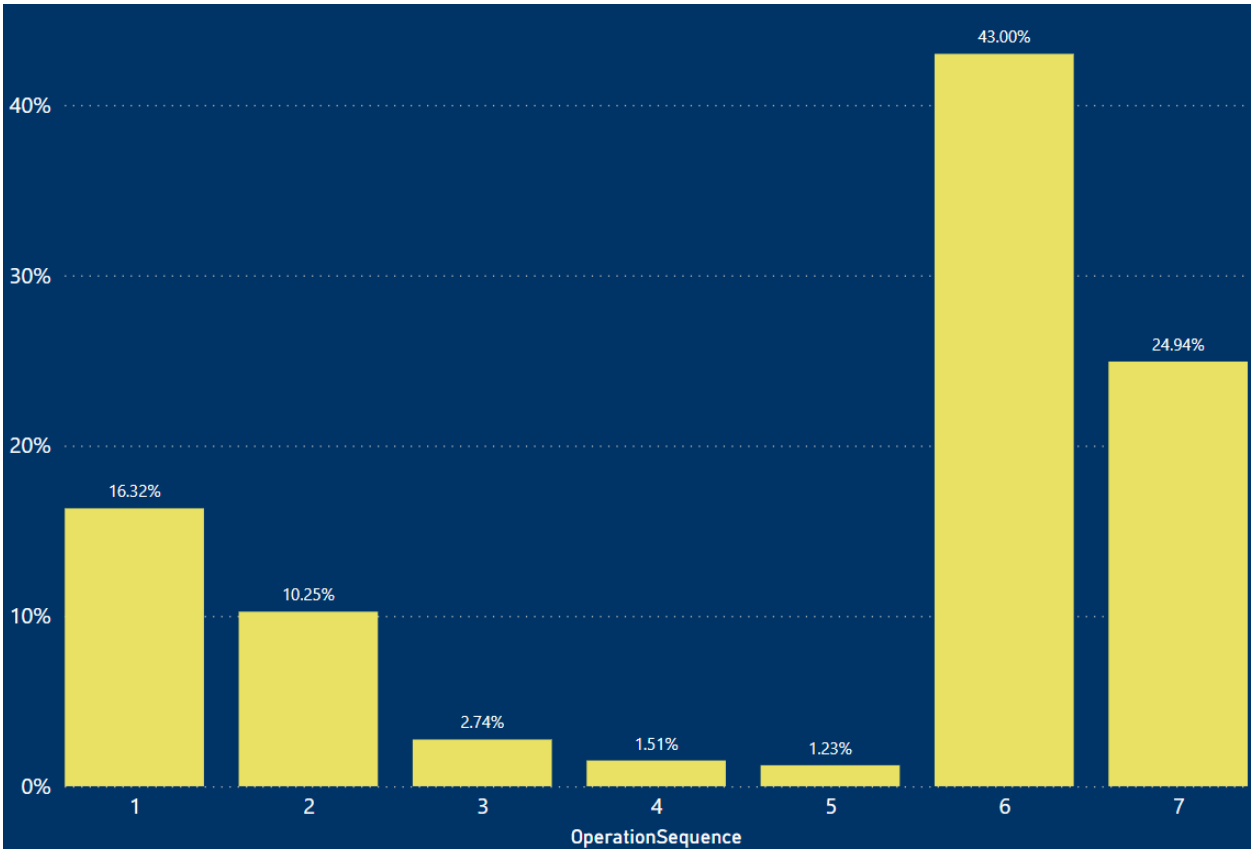


Figure 63. Percentage of Delayed date by Operation sequence

A column chart above indicates the percent of scrapped products in each Operation. The sixth Operation with the highest number of scrapped product (approximately 43%)

Figure 64. Efficiency between scheduled cost and actual cost

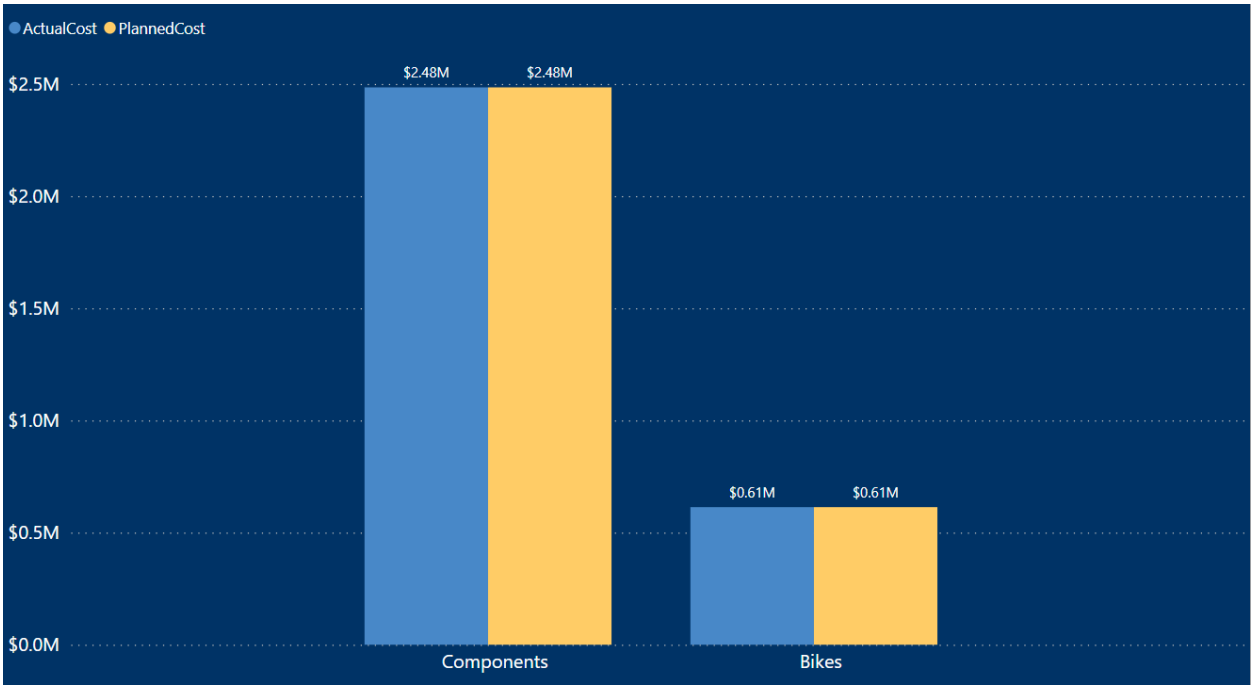


Figure 65. Actual cost and Planned cost by Category

Both charts were about indicating the Efficiency between scheduled cost and actual cost. 100% Efficiency Incredible number. All expenses are suitable to the company's budget. This was actually a good point. Well-organized cost management helps business generate profit as much as possible. This kind of profit was reinvested after that, to improve product quality as well. This really was a perfect cycle.

6.4 Evaluation and Discussion:

In the final analysis, our group will summarize the reports in order to conduct the ultimate goals from the beginning.

First of all, with the KPI Goals of each category, Accessories and Bikes both fluctuated in a 4-year period, once meeting the goal and next year not. Besides, Clothing and Components highlighted the same trend when both had slight failures in goals in the first two years but then succeeded at the end of the time recorded.

Though Adventure Work seems to have quality revenue (..) but their sales return rate did not reach the basic standard (..), more than half of the customers are not willing to buy the second time, let alone their high buying capacity.

Those two problems resulted in us progressing the analysis with the fact that all 4 categories had a similar trend in sales (increased heavily till the highest point in 2013 and then fell significantly in 2014). We have set up some theories to understand this sophisticated movement.

First to mention is the cost, apparently, sales only correlates with cost when it comes to the 3 three categories Bikes, Accessories and Components. Next thing to notice that black has been the customer's favorite color since it was first introduced. Other factors that need deeper analysis are Subcategory and Model. Our group found out that the two ways of division experienced opposite trends.

For Subcategory, customers tend to keep their buying habit with specific subcategories stable. It was hard for a new subcategories started to be our customer's interest

On the other hand, they had a tendency to easily shift from old models to new ones. This may lead to the extinction of the former models in near future.

Moreover, we also took product routing into consideration. Despite a successful rate of efficiency in routing, there has been a big gap between scrapped products from finished

goods and semi-finished goods (materials included). This may be caused by the 6th Operation phase.

For the final observation, our group has suggested some enhancements for the product department to take for better future

Change in mindset, change in the system and change in operation. ADW has an alarming issue despite the outstanding growth in years 2012 -> 2014.

There are needs to recreate both the design and formula to improve the sales performance and to enhance customer satisfaction. The sales of accessories and components are having great performance, the company may consider expanding the market on these two factors to hit strongly on the psychology of the loyal customer group and passion for bicycles.

The campaigns recently don’t make things go well like those had been done before. The company have to reduce operating expenses with strategic cuts and automation, reshape the company's brand identity and reputation

Optimizing all things should be applied right now, increasing convenience for customers is one of the permanent goals

The analysis above is a premise, a tool to help ADW make decisions in developing new product lines in the production aspect.

CHAPTER 7: CONCLUSION AND FUTURE WORKS

7.1. Results

After implementing the project of this subject, the team summarized the results achieved as below:

- Understand the role of Business Intelligence (BI) in helping managers make important business decisions.
- Understand the basic concepts and applications of ETL, MDX, KPI and OLAP techniques…
- Learn about AdventureWorks' production process and identify the data issues the company is experiencing from which to come up with the necessary business requirements.
- Build a data warehouse based on the knowledge learned and the use of SQL Server tools, Visual Studio…
- Build a report dashboard using Power BI to visualize data to make reviews and insights.
- Acquire teamworks skills, problem-solving skills, skills to build requirements.
- Learn to work seriously and aggressively to meet the deadlines.
- Get the most overview of data-related work and motivate future major-oriented choices (Data Analyst,...).

7.2. Limitations

Besides the positive side achieved, there are still limitations the team needs to overcome in order to be able to complete future projects better:

- In the process of implementing the subject project, the team does not understand the data, so the implementation of pouring data into the data warehouse is difficult.
- There isn't much time to learn and complete the project, a lot of knowledge and applications haven't been deepened.

- We have built data warehouses and visualized using power BI however adventureworks data is virtual so we do not know that the actual types of data in the future will be difficult or easy to process.

7.3. Future works

- Learn more about the data to better implement future data projects if possible.
- Conduct the surveys about the connection between BI and the customer's needs in practice to get important insights.
- Use the knowledge learned to pursue the work of a data analyst if possible.

REFERENCES

1. Education, I. C. (2020). *www.ibm.com*. Retrieved from https://www.ibm.com/cloud/learn/etl
2. Education, I. C. (2020). *www.ibm.com*. Retrieved from https://www.ibm.com/cloud/learn/olap#:~:text=Roll%20up%20Roll%20up%20is%20the%20opposite%20of,viewing%20each%20country%27s%20data%2C%20rather%20than%20each%20city.
3. Microsoft. (n.d.). *support.microsoft.com*. Retrieved from https://support.microsoft.com/en-us/office/overview-of-online-analytical-processing-olap-15d2cdde-f70b-4277-b009-ed732b75fdd6
4. Microsoft. (n.d.). *support.microsoft.com*. Retrieved from https://support.microsoft.com/en-gb/office/transpose-rotate-data-from-rows-to-columns-or-vice-versa-3419f2e3-beab-4318-aae5-d0f862209744#OfficeVersion=Windows
5. Qlik. (n.d.). *www.qlik.com*. Retrieved from https://www.qlik.com/us/kpi/
6. Taylor, D. (2022). *www.guru99.com*. Retrieved from https://www.guru99.com/business-intelligence-definition-example.html#:~:text=BI%20%28Business%20Intelligence%29%20is%20a%20set%20of%20processes%2C,to%20transform%20data%20into%20actionable%20intelligence%20and%20knowledge.
7. Team, M. S. (2015). *Slicing and Dicing Meaning & Definition*. Retrieved from www.mbaskool.com: https://www.mbaskool.com/business-concepts/it-and-systems/13503-slicing-and-dicing.html
8. Tutorialspoint. (n.d.). *www.tutorialspoint.com*. Retrieved from https://www.tutorialspoint.com/dwh/dwh_olap.htm