UNIVERSITY OF ECONOMICS AND LAW

FACULTY OF INFORMATION AND SYSTEM

ASSIGNMENT REPORT

Project*: *ANALYSIS THE ROOT CAUSE OF AN E-COMMERCE COMPANY'S TRAFFIC DECLINE

Course: **Data Analytics in Business**

Lecturer: **Mr. Nguyen Van Ho**

Group: **Eight**

Ho Chi Minh City, January 28th 2023

ACKNOWLEDGMENTS

Firstly, we would like to thank our academic supervisor **Mr. Nguyen Van Ho** for providing us all the necessary help for the completion of this report. Thank you very much Sir for guiding us to successfully start and complete this report. We are also thankful for the patience that you have shown during the project.

Secondly, with the passage of time, we are students of the University of Economics and Law standing on the other entity of our course completion, hence are finalized with our report named

"**ANALYSIS THE ROOT CAUSE OF AN E-COMMERCE COMPANY'S TRAFFIC DECLINE**". Vividly enough, our project comprises good endeavors. But no doubt, our contribution will be best evaluated on your sharp scale of acceptance & analytical remarks. Consequently, we are submitting our report on your very concern. Hopefully, you will discover our well-researched, informative approach as a hallmark of hard work.

Instead, in case of any further clarification or elaboration as to our report, we would welcome the opportunity to consult with you to explore how our findings could best meet your needs.

Thank you again and all the best to you!

Group EIGHT.

COMMITMENT

This report "**ANALYSIS THE ROOT CAUSE OF AN E-COMMERCE COMPANY'S TRAFFIC DECLINE**". is carried out by our team throughout and with the support of lecturer Mr. Nguyen Van Ho. We completely create the parameters, tables and images shown in the article based on the theory and research papers related to this topic that we have mentioned in the references at the end of this report. We certify that the entire content of the report is truthful, unique, and free of fraud. We take all responsibility for ensuring the transparency of the work.

Ho Chi Minh City, 2023

Group Eight.

TEAM MEMBER

No.	Full name	Student ID
1	Nguyễn Thành Công	K194111526
2	Lê Khánh Hân	K194111535
3	Nguyễn Quốc Thắng	K194111577
4	Nguyễn Phạm Anh Thư	K194111580
5	Bùi Thị Thu Trang	K194111581
6	Trịnh Ngọc Bảo Trân	K194111584
7	Nguyễn Thanh Trúc	K194111587
8	Đồng Thị Mỹ Uyên	K194111591

Contents

LIST OF ABBREVIATIONS

Abbreviation	Definition
BI	Business Intelligence
ETL	Extract, Transform and Load
KPI	Key Performance Indicator
MDX	Multidimensional Expressions
RLS	Row-level Security
DBA	Database Business Analyst

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: PROJECT OVERVIEW

1. Reason

Today, e-commerce is a crucial part of the worldwide retail industry. Global E-commerce sales growth will be fueled by consumer demand and aided by technical advancements. It has grown to accommodate people's evolving preferences and to improve the convenience of online shopping for today's consumer.

After the Covid-19 epidemic, e-commerce is giving many companies new opportunities. The public formerly found e-commerce to be rather weird, but these activities are now very common. Consumer behavior and e-commerce activities have changed quickly as a result of the pandemic, which has helped to advance the growth of the e-commerce industry and the digital economy.

However, in addition to the benefits, e-commerce also confronts a number of difficulties that must be overcome in light of the Covid-19 epidemic. With the purpose to have a good insight about e-commerce data and explore the e-commerce market, my team decided to evaluate the reasons for rising/falling client access to e-commerce exchanges and present a number of suggestions to help improve the effectiveness of these activities.

2. Project Objective

2.1 Overview objective

The objective of this project is to conduct an overview study on data analysis methods, decision support models and business intelligence solutions (Business Intelligence - BI), business performance assessment and management. That way, use these studies to analyze data and build financial analysis reports that support management decision-making.

2.2 Detail objective

- Research overview on decision support models and BI business intelligence solutions, data storage models and data processing methods.
- Clarifying the role of BI in business, especially in e-commerce.
- Research and use Microsoft Power BI.
- Improve e-commerce business efficiency, increase revenue for businesses, thereby seeing the potential benefits of BI.

3. Objects & Scopes:

3.1 Objects

Real commercial data of S e-commerce company from 2016 to 2018

3.2 Scopes

Company S

4. Research method

Diagnostic analysis is one of the four primary types of data analysis that answers the question "Why is this happening?". Diagnostic analytics helps companies make more informed decisions about how to solve problems and drive continued success. Companies can better understand the cause of the patterns they have observed in their data using diagnostic analysis. Diagnostic analysis can include many techniques, including data drilling and data mining. Companies may need to examine additional data sources, potentially including external data so as to investigate the root causes of trends.

Importance of Diagnostic Analytics

Every company can benefit from better understanding its business performance to drive success and fix problems. Diagnostic analytics helps companies better understand the internal and external factors affecting its results. It presents a more comprehensive picture of each situation, which help businesses make better decisions.

How Does Diagnostic Analytics Work?

Diagnostic analytics uses a variety of techniques to provide insights into the causes of trends. These include:

- Data drilling: Drilling down into a dataset can reveal more detailed information about which aspects of the data are driving the observed trends.
- Data mining hunts through large volumes of data to find patterns and associations within the data.
- Correlation analysis examines how strongly different variables are linked to each other.

Benefits and Limitations of Diagnostic Analysis

Understanding the causes of business results is critical to a company's ability to grow and learn from its mistakes. Diagnostic analysis allows companies to focus on factors that lead to success or cause failure, including contributing factors that may not be obvious at first glance. Diagnostic analytics can help instill a culture of data-driven analytics across the enterprise. When business leaders understand that a company has the tools to investigate the cause of a problem, they are more likely to use diagnostic analysis in their decision making.

One limitation of diagnostic analysis is that it focuses on historical data; it can only help the business understand why events happened in the past. In addition, further investigation may be needed to determine whether the correlations revealed by the diagnostic analyzes actually suggest cause and effect. To look to the future, businesses need to use other analytical techniques, such as predictive analytics, looking at the potential future impact of trends and events, and prescriptive analytics. , suggests actions a business can take to influence the outcome of future trends.

5. Tool & Programing Language

Power BI: converts data from different data sources to interactive dashboards and BI reports, calculate and visualize the information generated from the dataset.

6. Structure of project

- Chapter 1. Project overview

  Beginning with the first chapter, the project is summarized in light of current market conditions, business problems and set remedial goals.

- Chapter 2. Theoretical basis

  This chapter mainly defines BI, ETL Process, Data Warehouse and Data Mart, KPI and MDX language for multidimensional data analysis and OLAP. In addition, we also identify the benefits, advantages and disadvantages of the definitions above .

- Chapter 3. Requirements Analytics and BI Solution Cycle

  In chapter 3 we focus on analyzing requirements and proposing BI solutions. We work on a sales module, so we will learn about the sales department and sales revenue. Business requirements analysis is very important in this chapter, if any income or business requirements analysis is missing this will greatly affect data warehouse construction. The proposed BI solution is also presented in this chapter through the analysis of business requirements.

- Chapter 4. Visualization and Report Analysis

  We present the results after analyzing the data through Power BI. Visualization methods have long been used to support exploratory data analysis tasks.

- Chapter 5. Discussion and Management Implications

  In general, the project has analyzed and understood the sales process. Analyze requirements and identify problems to be solved, make business decisions, design and build a Sales data warehouse. Analyze data through Power BI tools. However, there are still unresolved issues such as the data contains many null values that make it difficult to analyze and the data is not updated automatically.

CHAPTER II: THEORETICAL BASIS

In this chapter 2, the research team will conduct a review of previous studies related to the research topic in order to review the concepts and form the basic theoretical basis to lay the foundation for the research. this. Then, the study will propose a research model.

---

1. Overview about BI

1.1 What is business intelligence?

To assist businesses in making more data-driven decisions, business intelligence integrates business analytics, data mining, data visualization, data tools and infrastructure, and best practices. In actuality, you can determine if your organization has modern business intelligence when you have a thorough understanding of its data and use it to drive change, get rid of inefficiencies, and react swiftly to supply or market changes. Flexible self-service analysis, controlled data on reliable platforms, empowered business users, and speed to insight are prioritized by modern BI solutions.

It's crucial to remember that this is a relatively contemporary understanding of BI, and the term has a troubled past as a catchphrase. Traditional business intelligence, complete with capital letters, first appeared as a method of information sharing between firms in the 1960s. In 1989, the phrase "business intelligence" was introduced along with computer decision-making models. These programs continued to evolve, transforming data into insights before becoming a specific service solution from BI teams with IT-dependent support. This essay is just the tip of the iceberg and will serve as an introduction to BI.

1.2 How business intelligence works

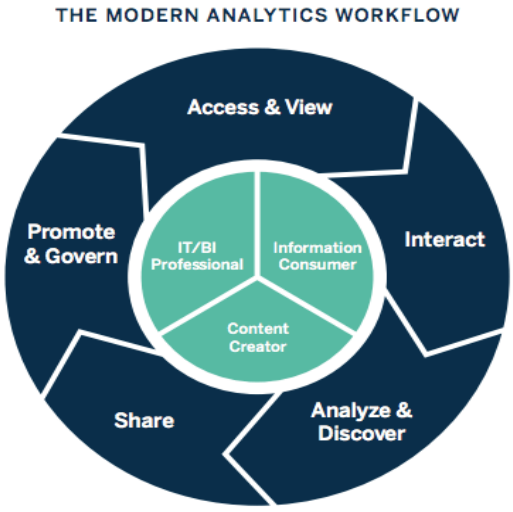


Figure 1. *A circular flow chart of the modern BI workflow*

Organizations and businesses have objectives and questions. They collect the relevant data, analyze it, and decide which steps to take to achieve their goals in order to respond to these questions and monitor performance against these objectives.

On the technological side, systems used by businesses are mined for raw data. Data is processed, then it is kept in files, applications, data warehouses, and the cloud. Users can access the data once it has been stored and begin the analysis process to provide answers to business problems.

Data visualization tools, which turn data into charts or graphs and present them to any important stakeholders or decision-makers, are another feature of BI platforms.

1.3 BI methods

Corporate intelligence is a broad word that refers to the procedures and techniques for gathering, archiving, and analyzing data from business operations or activities in order to improve performance, rather than just one particular "thing." All of these factors combine to produce a holistic picture of a firm that enables decision-makers to take better, more effective actions. Business intelligence has developed over the last few years to now incorporate additional procedures and activities to aid in performance. These procedures consist of:

- Databases, statistics, and machine learning (ML) are all used in data mining to find patterns in massive datasets.
- Reporting: Providing stakeholders with data analysis so they may reach conclusions and take action.
- Performance metrics and benchmarking: Monitoring performance versus goals by comparing current performance data to previous performance data, generally utilizing customized dashboards
- Descriptive analytics: Investigating events through basic data analysis
- Querying is the process of asking queries of data sets and BI extracting the replies.
- Statistical analysis: Using the findings from descriptive analytics, further analyzing the data to determine how and why a pattern occurred.
- Data visualization is the process of visualizing data analysis through the use of graphs, charts, and histograms to make the information easier to understand.
- Visual analysis: Investigating data with visual storytelling to share findings immediately and maintain analysis flow
- Data collection from various sources, identification of the dimensions and measurements, and preparation for analysis

1.4 Benefits of business intelligence (BI)

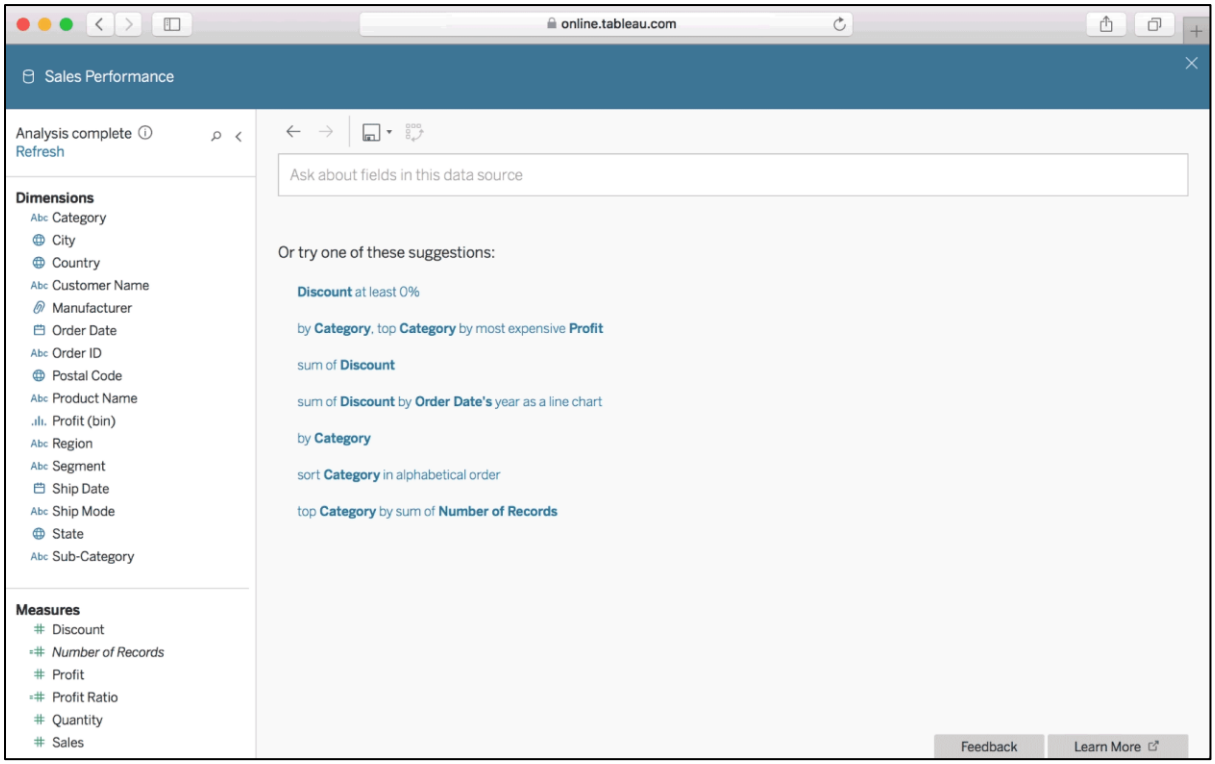


Figure 2. *Great BI helps businesses and organizations ask and answer questions of their data.*

Faster analysis, intuitive dashboards

BI solutions are made to do intensive data processing in the cloud or on the servers of your business. The data is then analyzed using user queries, drag-and-drop reports, and dashboards by BI tools after being collected from various sources and placed in a data warehouse. Lenovo improved reporting efficiency across many divisions by 95% thanks to business intelligence. Multiple monthly reports were consolidated by their HR department into a single snapshot dashboard. Additionally, PepsiCo used the potential of BI to reduce analysis time by up to 90%. Business intelligence dashboards have the advantage of simplifying and simplifying data analysis, enabling non-technical users to tell stories with data without having to learn code.

Increased organizational efficiency

BI gives executives the tools to access data, get a complete picture of their business, and compare their performance to that of the entire company. Leaders might find opportunities by taking an overall view of the organization. Pfizer uses BI systems to facilitate departmental communication and has created models to improve patient diagnosis and create quicker, more effective clinical trial processes. PEMCO, an insurance firm, uses Tableau to handle and quickly resolve claims. BI allows firms to leverage data to develop on new programs and products for the business because they spend less time on data analysis and report compilation.

Data-driven business decisions

Better business decisions are made possible by accurate data and quick reporting capabilities. Before meeting with potential customers, MillerCoors' sales team could access real-time data and sales estimates thanks to specialized mobile dashboards. Sales can now respond to clients' or prospects' demands with assurance because they are aware of the most recent facts. Leaders are no longer

required to wait days or even weeks for reports and cope with the possibility of obsolete information.

Improved customer experience

Customer pleasure and experience can be directly impacted by business intelligence. More than 1,500 staff dashboards were produced by Verizon after BI tools were implemented across several divisions. These dashboards gathered operational data as well as text information from customer assistance chat sessions. Verizon was able to discover areas to enhance customer service and cut the number of support calls by 43% using this data.

Improved employee satisfaction

IT organizations and analysts can respond to business user requirements more quickly thanks to BI solutions. With little training, departments that previously couldn't access their own data without contacting analysts or IT may now begin data analysis. BI is made to be scalable, offering data solutions to departments that require them and to people who are insatiably curious about data. When BrownForman expanded its BI platform to 1,000 users worldwide, it discovered that it complemented its current data architecture. Non-technical consumers should have a seamless and simple viewing experience using BI software.

Trusted and governed data

BI systems improve data analysis and organizing. When using traditional data analysis, users must visit many databases to find the answers to their reporting queries because separate departments' data is compartmentalized. A single data warehouse may now be created by combining all of these internal databases with other data sources including customer, social, and even historical weather data. The same data can be accessed simultaneously by all departments within a company. Utilizing business intelligence technologies, the marketing firm Tinuiti centralized more than 100 data sources, saving its clients hundreds of hours of analysis time.

Increased competitive advantage

When businesses are aware of their performance in the market, they can be more competitive. To determine when to enter and depart the market and how to best position itself, Rosenblatt Securities examined data from thousands of sources. Businesses can use BI to stay abreast of sectoral changes, track seasonal market shifts, and foresee client wants.

1.5 BI strategy in business

Your BI strategy consists of three key components:

- Vision. What are your goals in developing a BI practice within your organization?
- Processes and people. Who is going to develop and oversee the BI strategy? How then?
- Instruments and architecture What dashboards and products are we looking to build? for what geographical areas? And what effect will they have there?

Without a BI strategy	With a BI strategy
Multiple versions of the truth; people refer to different data when making decisions	The single version of the truth that leads to effective business decisions
Unclear names and definitions that get everyone guessing	Consistent definitions
Personnel overhead with different departments having their own BI	BI specialists and analysts maintaining the BI ecosystem
Data quality is an afterthought	Data quality is a priority

Figure 3. *Company comparison between having BI strategy and having no BI strategy*

5.1.1 Create a BI vision

To create a concise vision of your BI strategy, you first need to define your current state. Once you know where you are now, you will be able to define what is possible to achieve. First, you need to analyze where you are now by connecting data from different sources.

Then, set up your objectives and priorities, which will help you understand how BI can contribute to your business performance. Afterward, you will be on the path toward establishing clear and realistic expectations. At this stage, it is important to determine the following:

- What data will be collected?
- Who will be involved in BI processes?
- How will BI strategy be integrated with the corporate business operations?
- How to deliver BI solutions?
- What BI tool to select?
- Which KPIs do you need to monitor?
- What will BI lifecycle management look like?

5.1.2 Assemble a BI team

The business intelligence team is responsible for data discovery, analyzing, and connecting it with end-users. For large organizations, there are a variety of BI roles and responsibilities. However, they can be combined and narrowed to one position if you have a limited budget for human resources. If you are planning to create your in-house team, consider the following main roles:

- BI project manager – bridges the gaps between business and technology stakeholders.
- BI architect – established BI infrastructure by translating business requirements into the data warehouse.
- BI analyst – a person who is responsible for data mining and analysis to achieve meaningful insights.
- ETL developer – manages ETL processes for the data warehouse.
- Data visualization analyst – a person who delivers informative clean visuals from the analyzed

data to end-users.

- System administrator – installs and monitors the hardware.

5.1.3 Establish your BI architecture

There are different architectural approaches available depending on required configurations. According to Chartio's eBook Cloud Data Management, there are four stages of data sophistication:

- Source data (siloed data from datasets like Excel spreadsheets, Salesforce, Hubspot, etc.)
- Data lake (a single repository for storing unstructured data from data sources. Data in the lake is unstructured and is not ready for analysis).
- Data warehouse (it is a single source of truth where data is structured, cleaned, and ready for consumption).
- Datamart (it is a data set designed for specific use cases).

Different architectural approaches have specific capabilities. Depending on the growing organizational needs within your company, you can select one-tier, two-tier, and three-tier architecture:

- One-tier architecture is designed without domain-specific data mart layers (see above).
- The two-tier architecture is focused on adding data mart levels between the data warehouse and user interface.
- The three-tier architecture allows multidimensional data analysis with the help of OLAP cubes (online analytical processing) which means that data is represented from multiple dimensions. To put it simply, data is divided hierarchically into categories and subcategories.

5.1.4 Choose the best software vendor for your business needs

The supplier needs to meet the requirements for:

- Can the software/tool be integrated with your company's existing systems?
- What deployment options does the solution offer: on-premises or cloud?
- Are there flexible user permissions for data fields?
- Is the software easy to use and navigate?
- Do your business requirements correspond to the software you select?
- Is the software scalable enough if your company's demands will grow?
- Are the dashboards customizable?
- Does a vendor offer training to all the users, not only the tech team?

Besides, the choice of supplier depends on the detailed goals of the business, and the cost that your business is willing to pay.

5.1.5 Select BI platform and environment

There are such deployment options available: on-premise, cloud, hybrid. To define which options will best suit your business goal, let's compare them.

- On-premise is a model where infrastructure is installed locally so that it is under your total control. So, if you have a particular use case and are concerned about high privacy and security – this type can be perfect for you.
- The cloud environment is a model where data processing and management are performed by

third-party providers in the cloud which eliminates the expenditures on the hardware and infrastructure purchases. Data security threats are the main ones constraining businesses from adopting cloud computing.

- Hybrid cloud connects on-premises data centers and public clouds that are integrated. This solution suits best businesses that want to store their critical workloads on on-premises (much more secure) and less-sensitive resources on a cloud provider.

5.1.6 Build a BI roadmap

A business intelligence roadmap is a visual document representing planned activities by dates, KPIs, deliverables, and milestones within the timeline. The roadmap should contain technical details and be attractive to end-users as well.

2. Data Warehouse & Data Mart

Data warehouse

Data warehouse is a sort of data management system intended to assist and enable business intelligence (BI) activities, particularly analytics. Data warehouses frequently have a lot of historical data and are only meant to be used for queries and analysis. A data warehouse typically uses a variety of sources, including transaction programs and application log files.

Large volumes of data from many sources are centralized and consolidated in a data warehouse. Organizations can gain useful business insights from their data using its analytical skills to enhance decision-making. It creates a historical record over time that data scientists and business analysts can use to their advantage. A data warehouse can be thought of as the "single source of truth" for an organization because of these features.

Frequently found in a typical data warehouse are the following components:

+ A relational database for data management and storing
+ A method for data preparation through extraction, loading, and transformation (ELT)
+ Data mining, reporting, and statistical analysis capabilities
+ Tools for client analysis that visualize and present data to business users
+ Other, more advanced analytical applications that use data science and artificial intelligence (AI) techniques, as well as graph and spatial aspects to enable more types of large-scale data analysis

Data Mart

A Data Mart is a basic type of data warehouse that is concentrated on one topic or area of business. People may access data and discover insights more quickly when using a data mart since they don't have to spend time manually combining data from several sources or searching inside a more complicated data warehouse.

A data mart makes it simpler to retrieve the data needed by a particular team or line of business inside the company. For instance, combing through and compiling data from disparate systems could be time-consuming, inaccurate, and ultimately expensive if your marketing team is seeking for data to aid improve campaign performance over the holiday season.

Spreadsheets are the most common tool used to communicate and collaborate on data that they are required to locate from multiple sources. The typical outcome of is what is known as the "spreadsheet nightmare": human errors, misunderstanding, difficult reconciliations, and many sources of truth. As a central location where the necessary data is gathered and arranged before reports, dashboards, and visualizations are prepared, data marts have grown in popularity.

3. OLAP & OLTP

On-Line Analytical Processing (OLAP) is an essential element of decision support, which has increasingly become a focus of the database industry. Rollup (raising the amount of aggregation), drill-down (decreasing the degree of aggregation or increasing detail), slice-and-dice (selection and projection), and pivot are examples of OLAP processes (re-orienting the multidimensional view of data). Slice, dice, and other OLAP procedures are also available.

Online transaction processing, or OLTP, is a sort of data processing that involves carrying out several transactions that are happening simultaneously. These transactions are typically referred to as economic or financial transactions, and they are secured, documented, and available at any moment for an enterprise's accounting or reporting needs. Businesses depend on the transaction data kept in the database, which is utilized for reporting or analyzed to support data-driven decision making.

Large-scale real-time transaction processing (OLTP) facilitates the execution of numerous transactions by numerous users, whereas online analytical processing (OLAP) typically entails querying these transactions (also known as records) in a database for analytical reasons. OLAP assists businesses in gaining insights from transaction data so they may utilize it to make better decisions.

OLTP systems	OLAP systems
Enable the real-time execution of large numbers of database transactions by large numbers of people	Usually involve querying many records (even all records) in a database for analytical purposes
Require lightning-fast response times	Require response times that are orders of magnitude slower than those required by OLTP
Modify small amounts of data frequently and usually involve a balance of reads and writes	Do not modify data at all; workloads are usually read-intensive
To improve response times, use indexed data t	To make it simple to access a large number of records, store data in a columnar structure.
Require frequent or concurrent database backups	Require far less frequent database backup
Require relatively little storage space	Typically have significant storage space requirements, because they store large amounts of historical data
Usually run simple queries involving just one or a few records	Run complex queries involving large numbers of records

Table 1. *OLTP and OLAP comparison*

4. KPI in Business

Key performance indicators (KPIs) are checkpoints along the path to success in online retail. Ecommerce business owners can track their success toward sales, marketing, and customer service objectives by doing so.

KPIs are a crucial tool for ensuring that teams are contributing to the general objectives of the company. Here are a few of the main justifications for why key performance indicators are necessary: KPIs keep teams moving in the same direction whether measuring project success or employee performance; from risk concerns to financial indicators, key performance indicators give a realistic view of the state of firm, KPIs also enable you to see successes and shortcomings clearly, so that you may increase what works and decrease what doesn't, hold the teams accountable by ensuring that each member contributes value through the use of key performance indicators (KPIs) that both employees and managers can use to monitor progress.

There are numerous types of KPIs:

- Strategic: These broad-based KPIs keep track of company objectives. Typically, executives use one or two strategic KPIs to measure the performance of the company at any particular time. Examples include market share, revenue, and return on investment.

- Operational: These KPIs are usually more time-sensitive and concentrate on the effectiveness and efficiency of organizational procedures. Examples include regional sales, monthly transportation expenditures on average, and acquisition costs (CPA).

- Functional unit: Numerous KPIs are connected to particular functions, like finance or IT. Finance KPIs track gross profit margin or return on assets, whereas IT may track time to resolution or average uptime.

- Leading vs Lagging: Lagging KPIs track what has actually occurred, but leading KPIs can aid in outcome prediction. To make sure they're tracking the most crucial information, organizations utilize a combination of both.

5. Row-level security (RLS) with Power BI

Row-level security (RLS) is a data governance capability of Power BI that restricts data based on the authorization context of the logged-in user. RLS forms an integral part of an organization's data protection strategy as it implements appropriate data visibility for end-users. Power BI Row-level security ensures that end users have visibility only into the data they are supposed to see. Inappropriate access management can lead to chaos and unforeseen circumstances in any organization.

Row-level security is a way to protect sensitive data by limiting visibility access to Power BI data and reports. Row-level security Power BI is a horizontal limitation applied to rows within a table. Power BI applies filters on the data for the users with limited visibility based on the instructions

outlined by the administrator. The filters apply data access limitations at the row level, and these filters can be defined within roles.

The static method in Row-level security uses a fixed value in the DAX filter, while the dynamic method uses a DAX function. RLS involves several configuration steps, which should be completed in the following order:

- Create a report in Microsoft Power BI Desktop.
    - Import the data.
    - Confirm the data model between both tables.
    - Create the report visuals.
- Create Row-level security roles in Power BI Desktop by using DAX.
- Test the roles in Power BI Desktop.
- Deploy the report to Microsoft Power BI service.
- Add members to the role in Power BI service.
- Test the roles in Power BI service.

CHAPTER III: REQUIREMENTS ANALYTICS AND BI SOLUTION CYCLE

In this chapter 3, the team will perform the analysis of the requirements from the stakeholders to understand their needs and take steps according to the standard BI process to get useful insights and implement analytic results communication.

The picture below is the process that the team will apply to provide BI solutions for business.

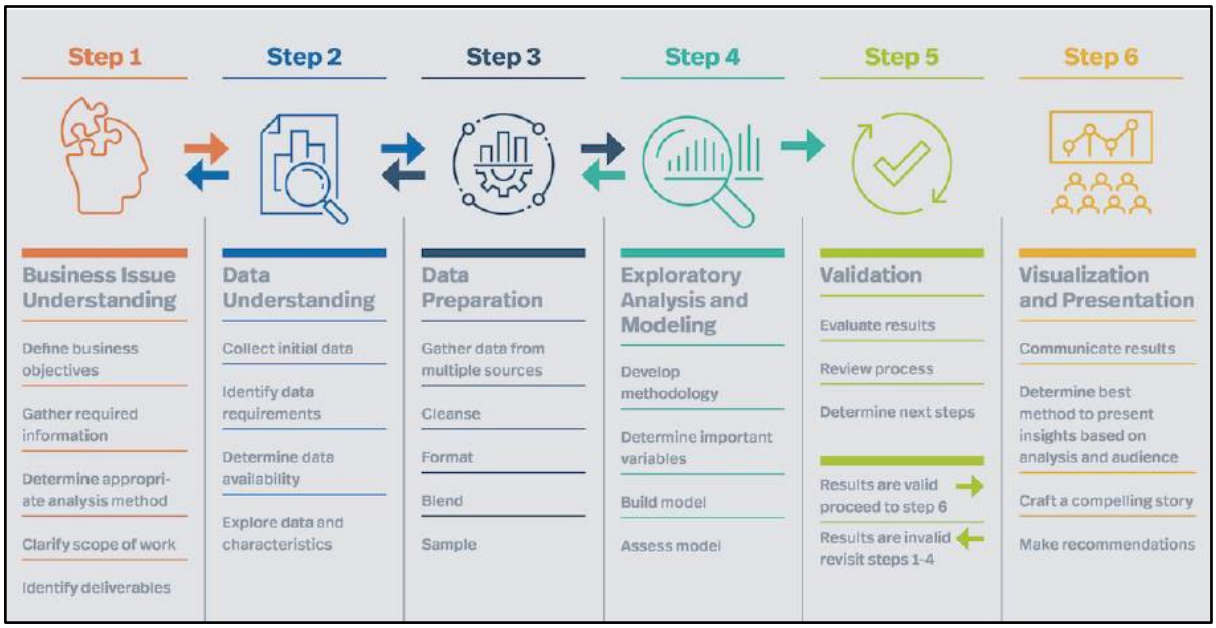


Figure 4. *Data analytics project lifecycle*

1. Business Issue Understanding

1.1 Define business objectives

Today, e-commerce is growing strongly and attracting a lot of new customers. Therefore, company S is trying to find ways to attract customers compared to its competitors in the market. Company S has identified the right things to do including expanding customer reach, reducing service costs, and providing an exceptional customer experience.

Company S uses data analysis so it can make data-driven decisions. Business analytics provides a business with an overview and insight into how companies can become more efficient, and these insights will allow that business to optimize and automate its processes.

The value of data analysis is that it increases productivity while helping to eliminate doubts. The most successful businesses have both short-term and long-term plans and projects. Substantial field data collection and analysis ensures that a company can still locate its valuable resources where they are needed most. It is essential to understand which areas need to be prioritized to help grow and move forward.

The ability to gauge customer needs and customer satisfaction through data analytics allows businesses to give customers what they want when they want it. With data analytics, many companies

have the opportunity to develop innovative new products to meet the ever-changing needs of their customers.

1.2 Gather required information

The covid-19 pandemic raged over 6 months last year which makes the E-commerce market have outstanding developments. E-commerce's role in promoting post-pandemic economic development is recognized. However, in the middle of this year, not only company S but also all E-commerce companies recorded a slowdown because of the government's reopening policy and the rapid growth of the new platform - Tiktok, especially Tik Tok Shop. It significantly impacts the current business. Now, The Board of Directors (BOD) wants to know the root of these problems, why there is such a change, and any remedy for them for the upcoming downtrend in the economy. They have listed all metrics which are corresponding to each objective and they need the support of the Data team for tracking and analysis. Hence, provide them advice based on their view of data to help them have deeper insights and can make them actionable.

No.	Objectives	Metrics
1	Expanding customer reach	Traffics: average session time, number of sessions, number of screens of 1 session
		Active customers, new customers, retention customers, churn rate.
		The number of categories, sellers, number products each category.
		Promotion Performance on E-commerce platform (Flash sale, Ads package, online store campaign)
		Marketing Performance
		Search through filters (pricing, sku, transfer, promotion)
2	Reducing cost-to-serve	Auto-reply messages cost saving.
		Less calls to customer service centers.

		Cost of advertising, logistics, customer services, operation
		Low-value high-demand items /categories
		Staffs' performance
		Customers's time spent on every transaction, by every employee, for each customer, product
		Conversion rate
3	Creating differentiated customer experiences	Features that customers used the most.
		Type of promotion that customers used the most.
		Loyalty program enrollment
		Social Media channel often used by customers
		Age, Gender, Interests, Habits of potential customers
		New trends
		Shopping cart abandonment rate
		Bounce rate (Number of people who go to the website and then go out without taking any action)
		Device types that customers use most to access the platform

Table 2. *Objectives and related Metrics*

1.3 Determine appropriate method:

The Data Team decided to use diagnostic analytic method for helping the business team understand clearly the pain points they meet and the root cause of them.

Why do we use this method?

Diagnostic analytics is an advanced form of analytics that examines data or content, measuring historical data against other data to answer the question "Why did that happen?" It is characterized by techniques such as drill-down, data exploration, data mining, and correlations. The diagnostic analysis focuses less on what happened and why something happened. Collectively, these analyzes are looking at processes and causes, rather than results.

The diagnostic analysis also provides insight into a particular problem. Every company can benefit from a better understanding of its business performance to replicate success and fix any problems. Diagnostic analytics helps companies better understand the internal and external factors affecting their results. It paints a more comprehensive picture of each situation, helping businesses make better decisions.

As the context above, "The Board of Directors (BOD) wants to know **the root of these problems**, **why** there is such a change and any remedy for them for the upcoming downtrend in the economy". The goal of diagnostic analytics is to help enterprises to answer the question of **why** this condition occurs, so this is the most suitable method to use in this case. The diagnostic analysis includes basic descriptive analysis, using the results of the descriptive analysis to find **the root cause of events**. From there, the performance indicators are analyzed more closely. The diagnostic analysis uses a variety of techniques to provide insights into the causes of trends. These include:

- Data Mining: Drilling into the data set can reveal more detailed information about which aspects of the data are driving the observed trends.
- Data mining searches through large volumes of data to find patterns and associations in the data.
- Correlation analysis examines how strongly different variables are associated with each other.

1.4 Clarify scope of work:

Customer Management Analyst includes:

- Master Data Management (Customer)
- Sales activities
- Social media engagement

1.5 Identify deliverables:

Because it would take too long to assess and present the metrics within three months. To make metrics easier to grasp initially and to maintain, DBA separates them into two parts. The KPIs of traffic, client retention, and category management will be the primary emphasis of phase 1. The team will explain why our business's client reach has reduced using these figures. Phase 2 will begin as soon as this project is completed. To reduce modification requests while the data team is working on the project, this strategy was accepted by the business team and includes the signatures of both teams.

No.	Metrics	Definition	Description
1	Number of sessions		Comparison with previous period (week, month)
2	Average session time	Average time that users browse the app.	Comparison with previous period (week, month)
3	Hourly/Weekly traffic		Shown the most / less visited times of the day/week
4	Avg. screens of 1 session		Comparison with previous period (week, month)
5	Active customers	Active: open app in month	Shown by Month
5	New customers	New: created (view by monthly)	Shown by Month
6	Retention customer	Retention: Placed an order last month and got back to place a new order this month.	Shown by Month and emphasize the time retention customer rate is high/low to optimize or expand campaigns.
7	Churn rate.	Churn rate: churn customers/(total customers - deleted customer) Churn Customer: un-open app in a month	Shown by Month and emphasize the time churn rate is low to alert.
8	Number of categories		Show the no of category parallel with no of active.
9	Top most viewed products/ top selling products		

Table 3. *Deliverable Metrics Sign-off*

2. Data understanding

2.1 Collect initial data:

The Data team must first gather data in order to examine these indicators. Following the DBA's clarification of the business needs, the Data Engineer team and they discussed how to include in the database the information they require on customers, customer events, product categories, sales orders, dates, sellers, and promotions.

2.2 Identify data requirements:

Based on business requirements and DBA transfer, Data Engineers have found and built datasets including the tables below for having enough information to serve business needs.

No.	Table	Description
1	Fact_Events	Detail event information about the time they browsed the app, type of events, type of platforms,...
2	Fact_Orders	Detail order information about order date, customer, seller, productid, and quantity.
3	Dim_Customer	Detailed customer information about customer: name, created date, deleted date
4	Dim_Items	Detail item information about item name, category, price,..
5	DimDate	Date rollup: 7 days ago, 28 days ago.
6	Dim Promotion	Detail promotion information about promotion name, start date, end date, promotion type,...

Table 4. *Datasets requested from DBAs.*

Besides internal data provided by Data Engineers, BI also needs some external data sources such as Customer Behavior, Category Information on other platforms.

2.3 Determine data availability

After cross-checking the database, Data Engineers have clarified tables that are available or unavailable. Given tables are what BI needs and are available in the database.

Table Name	Attribute
Fact Event	Event_ID
	Event Time
	User_ID
	Event Name
	PlatformType
	Parameter Name
	Parameter Value
Fact Orders	InvoiceID
	Line_item_id
	User_ID
	Item_ID
	CreatedAt
	Quantity
Dim_Customer	CreatedDate
	DeletedDate
	UserID
	FirstName
	Email
	Last Name

Dim Items	Item ID
	Item Name
	Created At
	Adjective
	Category
	Price
Dim Date	Date
	7 Day ago
	28 Day ago

Table 5. *Datasets provided by the Data Engineer team*

2.4 Explore data and characterize

The Face Event Table records 428364 unique events in 3 months on many platforms and many types of different events.

The Face Orders Table records about 400k unique orders in 3 months, and 1M products were sold out.

The Dim Customer Table records about 112k unique users.

The Dim Item Table records about 2k unique items in 30 categories.

3. Data preparation

3.1 Gather data from multi-sources

The first requirement from DBA is that they need data from external sources related to customer behaviors, and category information on other platforms. However, the Data Engineer team cannot find enough data information as required and the collected data is also not qualified. In light of this, the data team decides that the BI team will evaluate the internal source first until the Data Engineer team has gathered sufficient data and has thoroughly examined the data quality, after which the BI team will be able to exploit it.

3.2 Cleanse

After loading all data files into PowerBI, the BI team will check data quality first in the Power Query window.

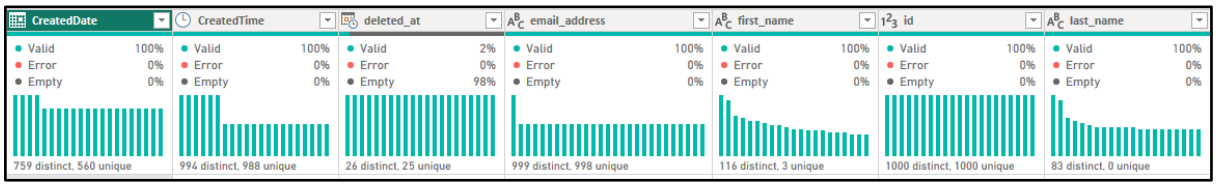


Figure 5. *User Data Quality*

All columns in the Users table are qualified at the maximum rate. The percentage of valid value is totally 100%. Empty and Error Value are 0%.

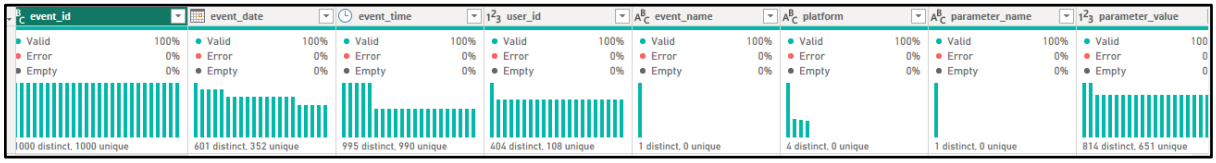


Figure 6. *Events Data Quality*

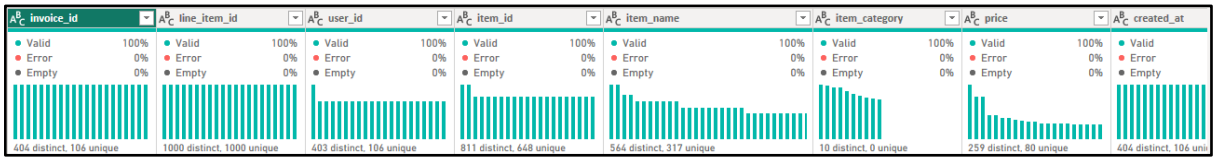


Figure 7. *Order Data Quality*

Is the same as the Users table, all columns in the 2 tables: Events and Orders have the maximum quality rate. However, only the Item table has 2 columns: Adjective and Modifier columns aren't entirely accurate (88% for adjectives and 64% for modifiers).

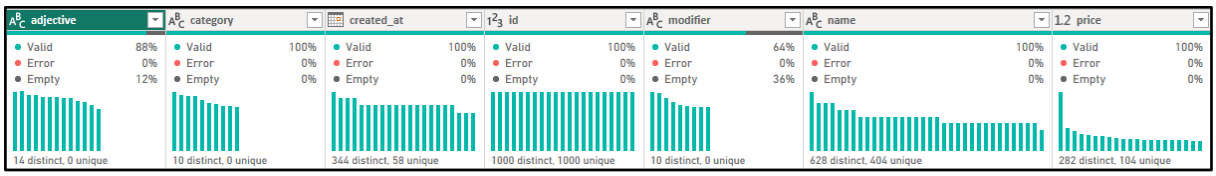


Figure 8. *Items Data Quality*

The team came to the conclusion that certain values in both columns are simply empty and do not represent mistakes. Thankfully, the team has established a pattern for these columns that places the adjective as the first word and the modifier as the final word from left to right. As a result, using this rule using Power Mcode in the Power Query window, we may quickly fill up blank data.

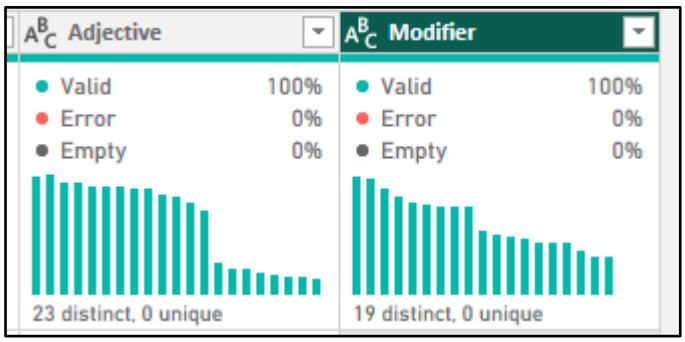


Figure 9. *Adjective and Modifier after refill*

Values in these columns after filling in the blank value and then, the team has the data reaching at the highest quality rate.

In addition, we removed unnecessary columns to reduce data capacity.

3.3 Format

In order to analyze more easily, the team has conducted some processes formatting the data for a better data model and more convenience when choosing attributes to visualize.

All columns have data type in DateTime, we will separate them into 2 columns: date and time for the purposes below:

- Decrease the uniqueness of value, and improve performance when data refresh.
- Is in the same format with date columns in the Date table.
- Can be easily analyzed by date or time depending on the analysis direction.

Besides, the BI team also changed the Type of Data of columns into the correct type based on the required information tables.

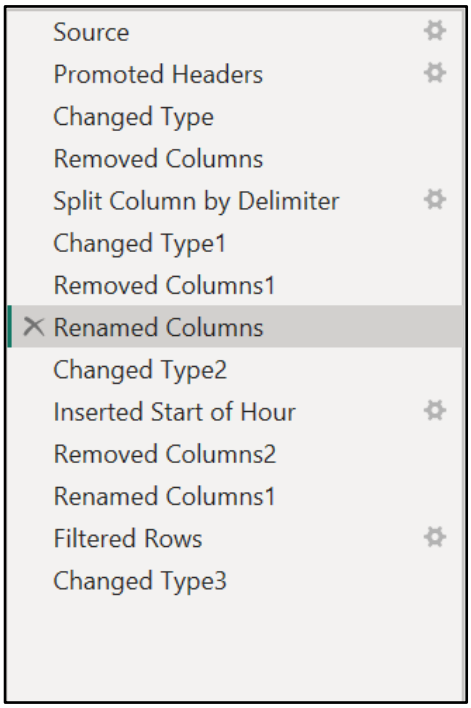


Figure 10. *Example for Fact Order Formating Process*

This picture above is an example for the formatting process. The process goes through many steps and has been done carefully by the team before proceeding to the next step - Modelling Data.

4. Exploratory Analysis and Modeling

4.1 Propose Data Model.

The team decided to adopt Galaxy Schema as the data model for this project since the datasets contain two fact tables and three dimension tables.

The data warehouse schema is being replaced by the Galaxy Data Schema, commonly referred to as a Fact Constellation Schema. The Galaxy Schema makes use of numerous fact tables joined by shared normalized dimension tables, unlike the Star and Snowflake Schemas. Galaxy Schema may be compared to an interconnected, fully normalized star schema that does not contain any redundant or inconsistent data.

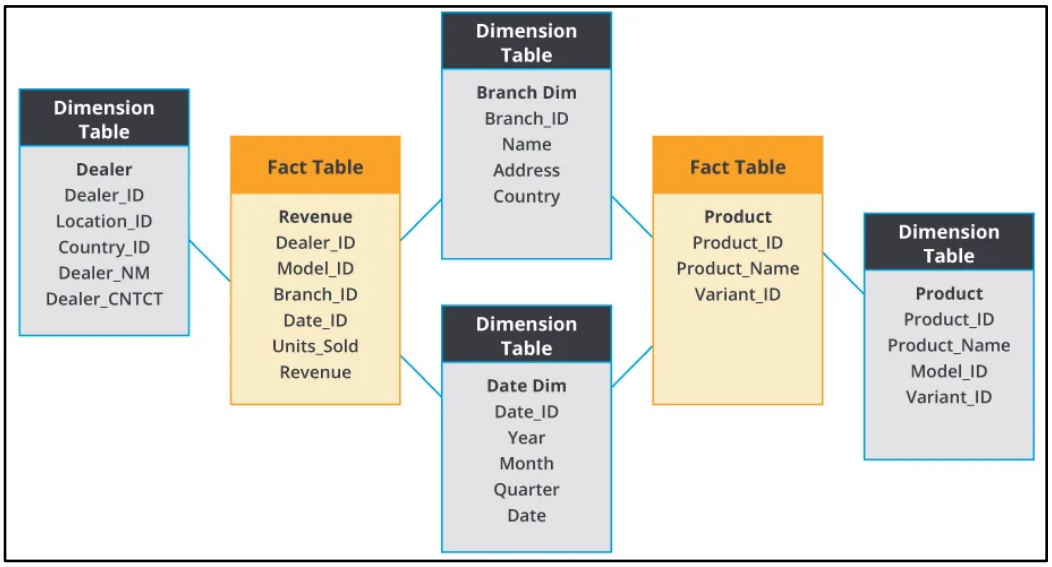


Figure 11. *Example of Galaxy Schema*

Characteristics of the Galaxy Schema:

- Galaxy Schema is multidimensional acting as a strong design consideration for complex database systems
- Galaxy Schema reduces redundancy to near zero redundancy as a result of normalization
- Galaxy Schema is known for high data quality and accuracy and lends to effective reporting and analytics

4.2 Design data model

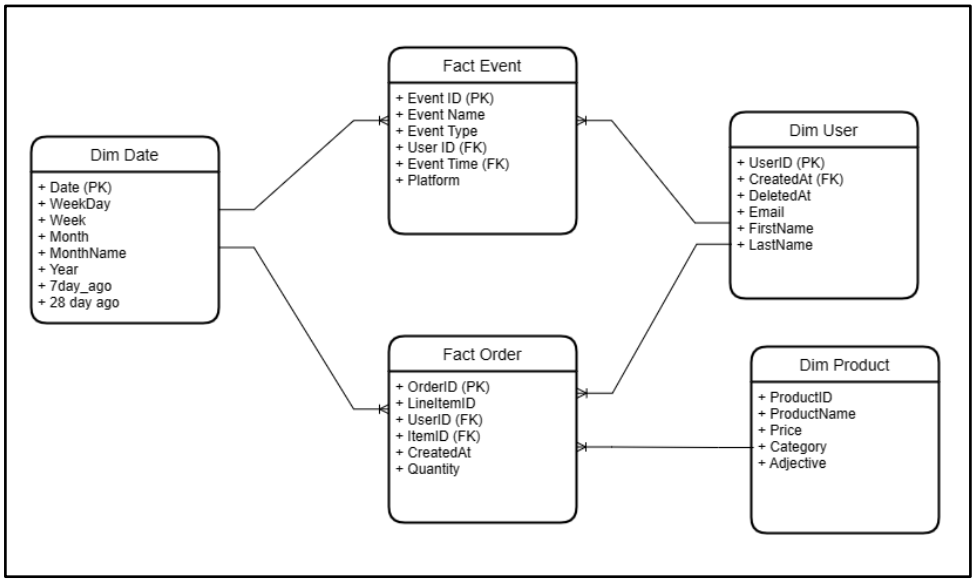


Figure 12. *Data Model*

Fact Event table have 2 relationships:

- Many-to-one with Dim Date. (EventTime - Date)

- Many-to-one with Dim User. (User ID - UserID)

Fact Order table has 3 relationships:

- Many-to-one with Dim Date. (CreatedAt - Date)
- Many-to-one with Dim User. (UserID - UserID)
- Many-to-one with Dim Product. (ItemID - ProductID)

4.3 Exploratory Analysis

Before applying diagnostic analytics, we use the descriptive analytics method to visualize all the metrics first to have a glance at the data and find the factors to analyze then.

* Metric 1: Number of Sessions:

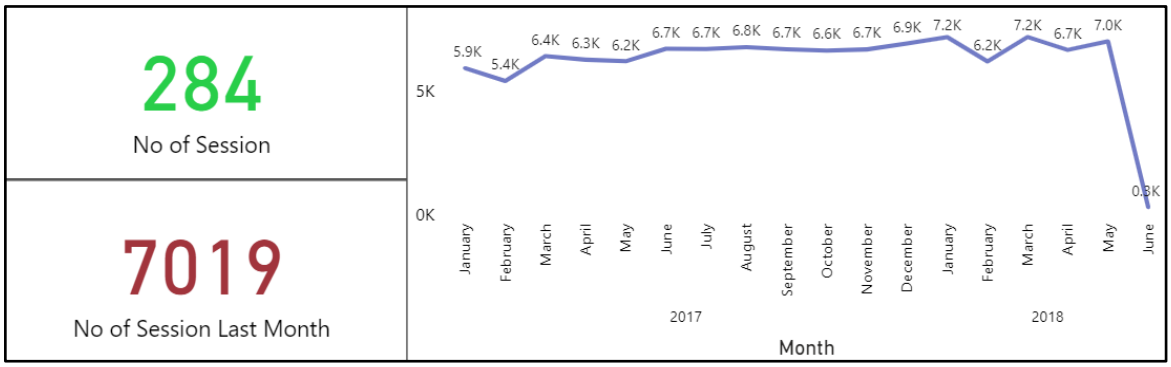


Figure 13. *Number of Sessions this Month compare vs Last Month & Number of Sessions Monthly.*

In general, traffic in 2017 and 2018 barely increased (skip June 2018 because the data is not available for the full month of June). Despite the chart's minor rise, it appears as though the growth is progressively slowing down.

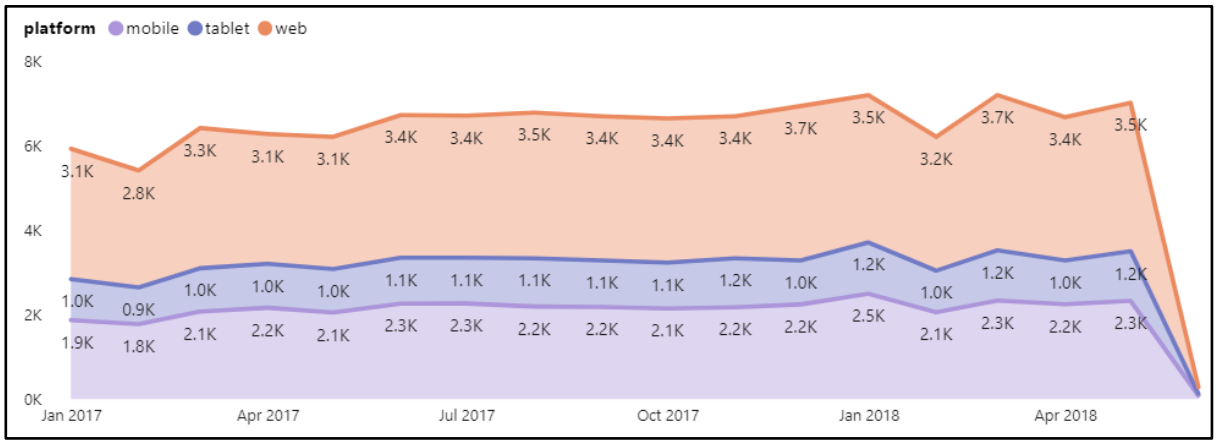


Figure 14. *Number of Sessions by platform Monthly in 2018*

Web is the platform with the largest traffic when looking at the number of sessions by platforms, more than the combination of mobile and tablet. The percentage of traffic on the platform has not changed much. The team may thus draw the conclusion that platform is not what influences consumer reach.

* Metric 2: Hourly/ Weekly Traffic:

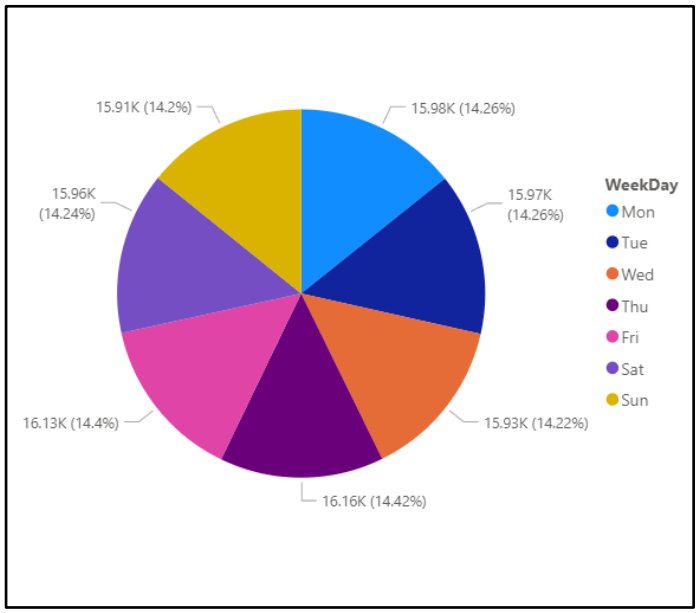


Figure 15. *Weekly Traffic*

In this figure, there is no difference in traffic percentage among days of the week. The difference between the percentages of the highest and lowest traffic is only 0.02%. It's really insignificant to conclude that day of week is the deciding factor on traffic.

Event Time	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Total
12:00:00 AM	660	694	629	636	642	665	705	**4631**
1:00:00 AM	720	637	692	709	696	664	668	**4786**
2:00:00 AM	679	675	715	712	709	658	699	**4847**
3:00:00 AM	651	680	645	679	679	671	689	**4694**
4:00:00 AM	668	643	671	694	663	635	640	**4614**
5:00:00 AM	658	673	663	701	652	660	659	**4666**
6:00:00 AM	655	669	677	672	658	688	617	**4636**
7:00:00 AM	679	651	677	690	663	680	659	**4699**
8:00:00 AM	685	644	644	675	668	705	643	**4664**
9:00:00 AM	659	670	680	708	682	669	623	**4691**
10:00:00 AM	660	650	651	698	686	663	636	**4644**
11:00:00 AM	673	633	657	646	705	701	713	**4728**
12:00:00 PM	680	670	688	678	679	661	697	**4753**
1:00:00 PM	639	703	647	620	691	670	682	**4652**
2:00:00 PM	666	718	695	654	658	704	668	**4763**
3:00:00 PM	655	641	619	646	679	666	667	**4573**
4:00:00 PM	672	701	705	667	654	662	683	**4744**
5:00:00 PM	668	670	671	639	674	606	651	**4579**
6:00:00 PM	666	670	600	673	642	677	632	**4560**
7:00:00 PM	654	683	691	655	691	599	655	**4628**
8:00:00 PM	652	646	628	701	638	655	699	**4619**
9:00:00 PM	648	624	672	686	693	734	648	**4705**
10:00:00 PM	682	669	660	659	677	662	620	**4629**
11:00:00 PM	650	660	650	662	650	601	655	**4528**
Total	**15979**	**15974**	**15927**	**16160**	**16129**	**15956**	**15908**	**112033**

Figure 16. *Hourly Traffic*

The team identifies the two times when users use the app most frequently as 1AM:3AM and 11AM:12PM when looking at hourly views in further detail. However, there isn't much of a difference between the times. The day with the most equally spread traffic is Monday.

* Metrics 3: Avg. screens per 1 session

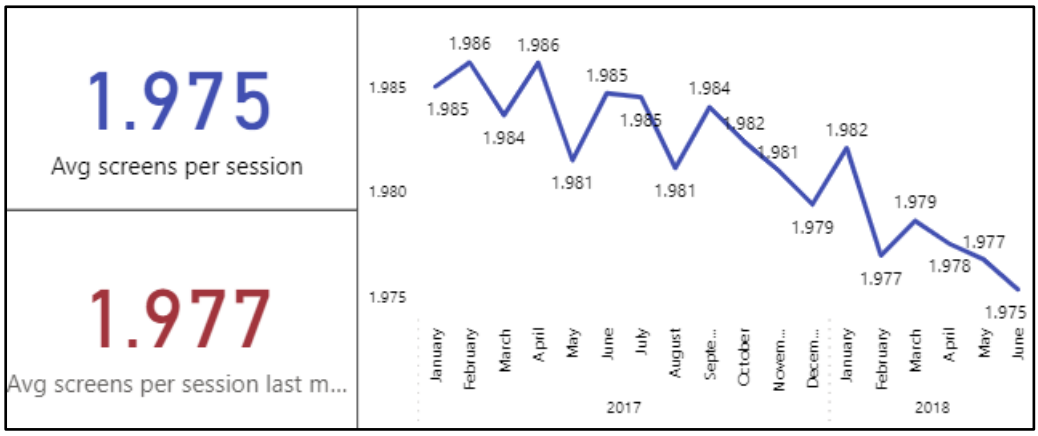


Figure 17. *Average Screens per 1 Session.*

The average number of screens per session monthly has complex fluctuations, but it usually comes approximately to 2. In addition, by the figure, the decrease in this metric is evidentable over time and the slope of the metric line is very high. 2 screens per session is not a high number and thus the company will be seriously warned if it starts to decline.

* Metrics 4: Active customers

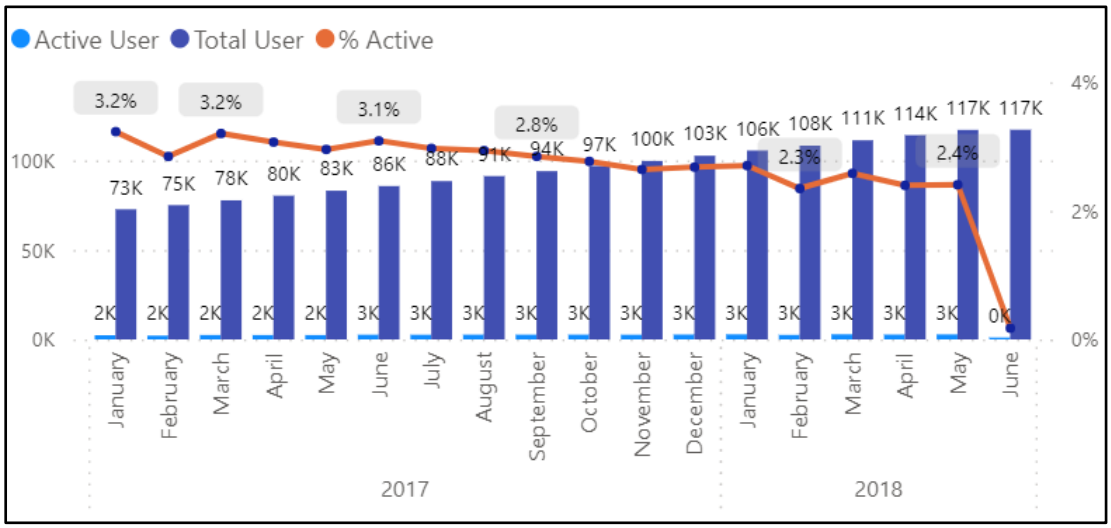


Figure 18. *Active User & Total User & % Active Monthly.*

This figure illustrates that the number of active users accounts for a tiny percentage comparison with the total user. Even if the total number of users is rising silently and uniformly, the percentage of active users, which is typically at 3%, is also beginning to decline.

* Metrics 5: New customers

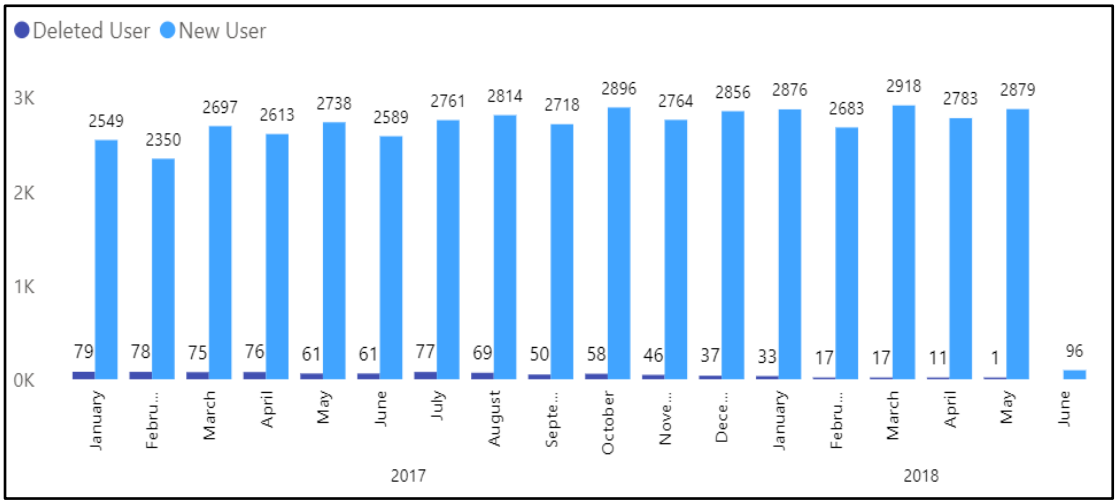


Figure 19. *New User & Deleted User Monthly*

While the number of deleted users has considerably dropped after August 2017, the number of new users regularly increases but not equally. Although the firm is showing signs of growth, the marketing team has a hurdle in making use of these users.

* Metrics 6: Retention customer

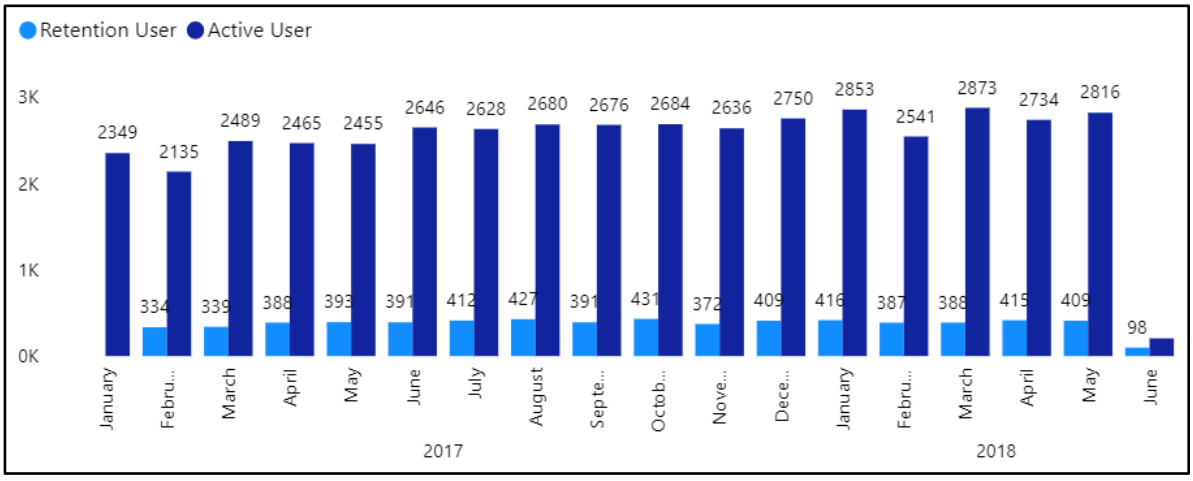


Figure 20. *Retention User & Active User Monthly*

Excluding January 2017, Retention Users maintains a steady monthly count but percentage of user retention with active users is quite low.

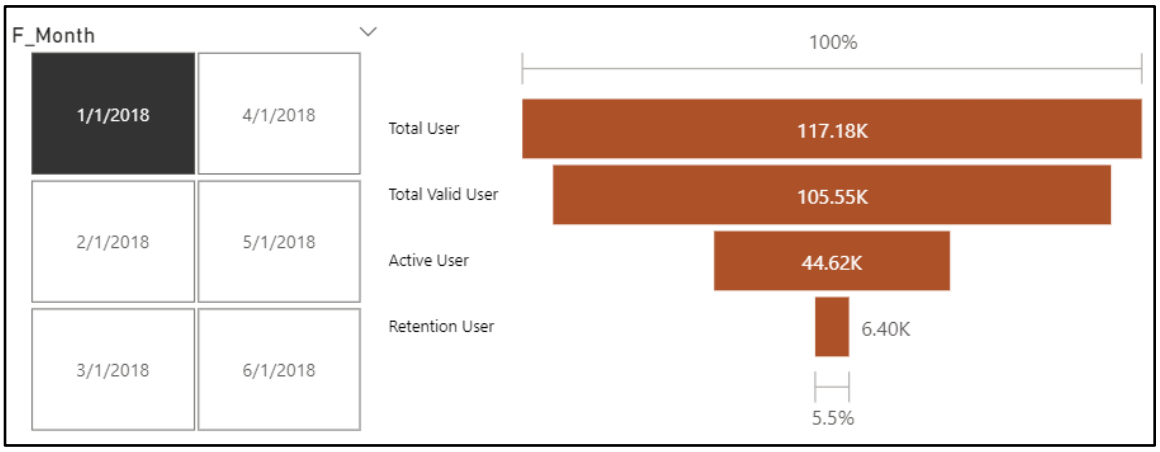


Figure 21. *User Funnel Monthly*

User Funnel Chart shows the user's conservation rate of the company. Total Valid User is the user who has not deleted their accounts. Active user is the user opening app in the selected month. Retention user is the user opening app in the selected month and also opened app in the previous month of the selected month. Finally, the conversion rate is just over 5%.

* Metrics 7: Churn rate

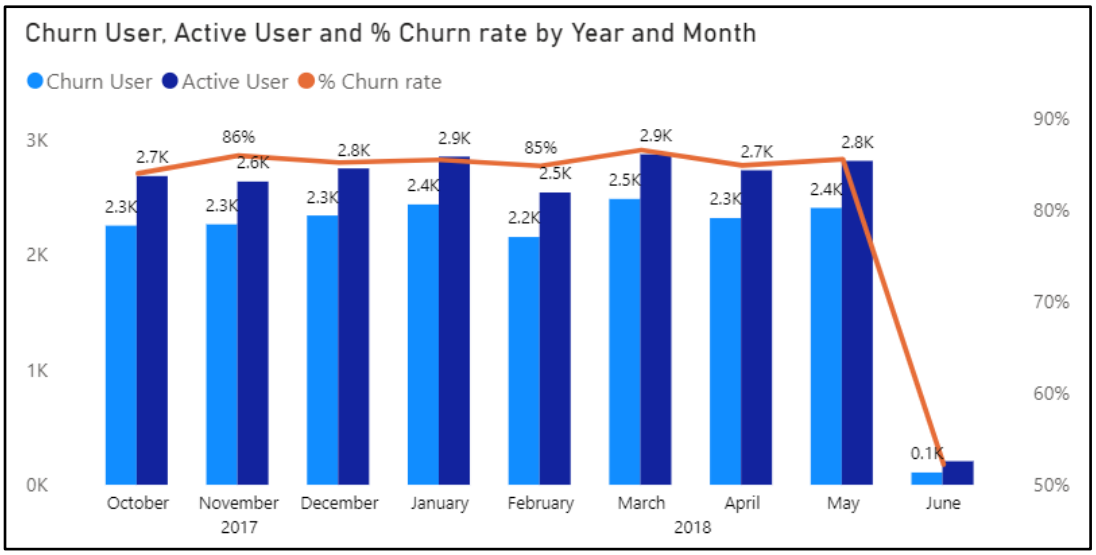


Figure 22. *Churn User, Active User & % Churn Rate Monthly*

A significant portion of all monthly active users are accounted for by churn users, and this number has not decreased. That is followed by the bounce rate, which continues to be quite high (over 80% each month).

* Metrics 8: Number of Category

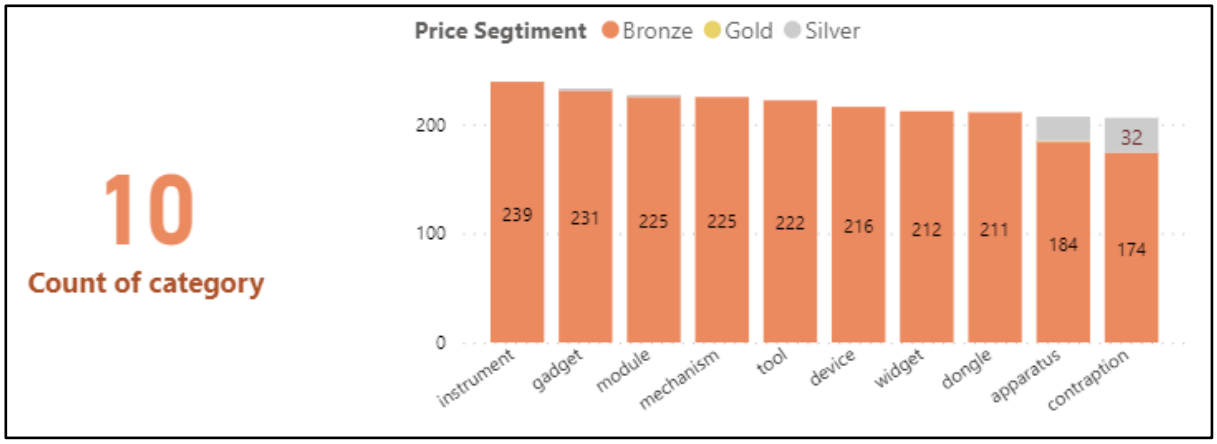


Figure 23. *Number of category and number of products by Category, Price Segment*

There are 10 categories and 3 price segments in the product list. In general, product quantity in each category has no big difference. Bronze is the segment with the most products. Silver also have a middle quantiy at apparatus and contraption category. Gold products account for a very small amount.

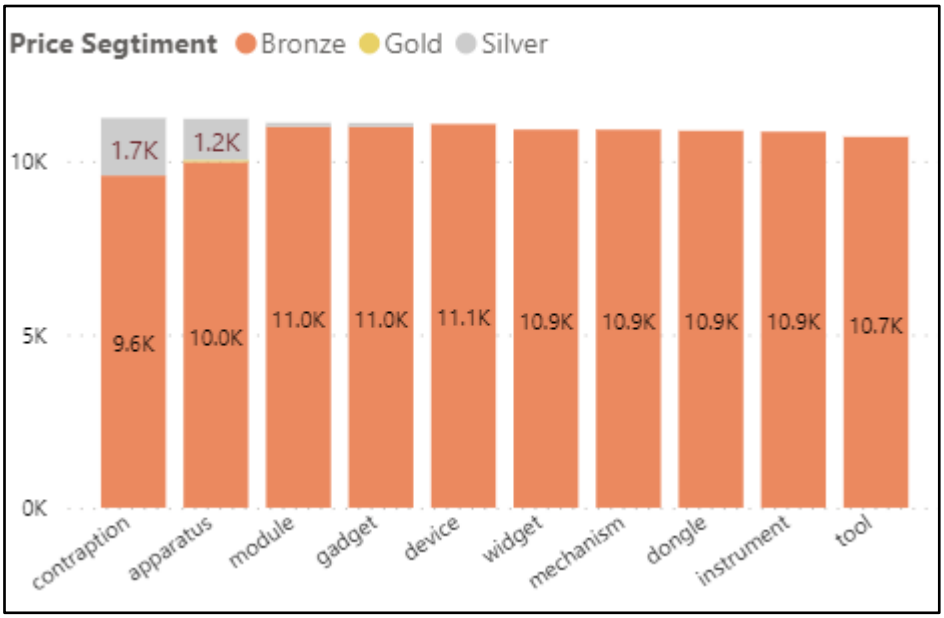


Figure 24. *Number of sessions by Category, Price Segment*

Following the figure above, the number of sessions is also the same with the product quantity. The Bronze Segment performs outstandingly with the others segment.

Metrics 9: Top most viewed products/ top selling products

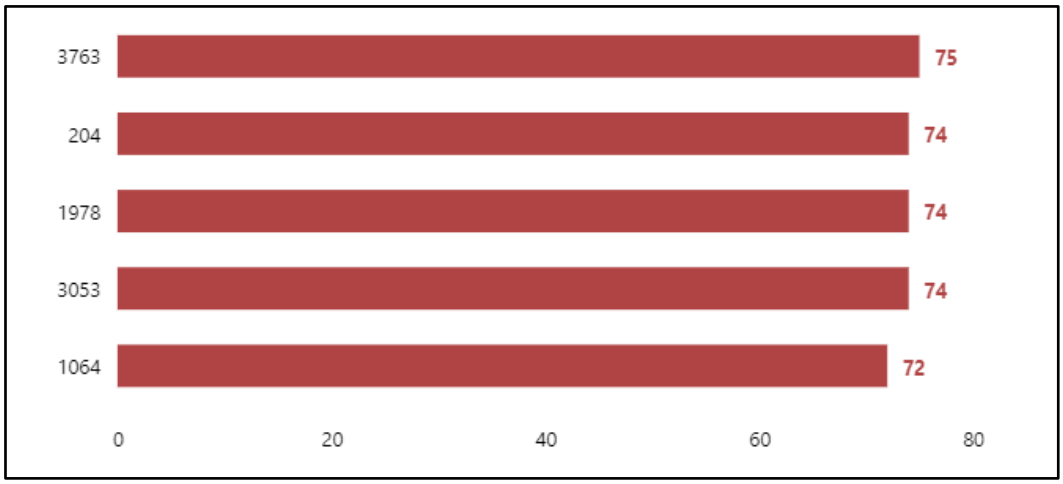


Figure 25. *Top Viewed Products*

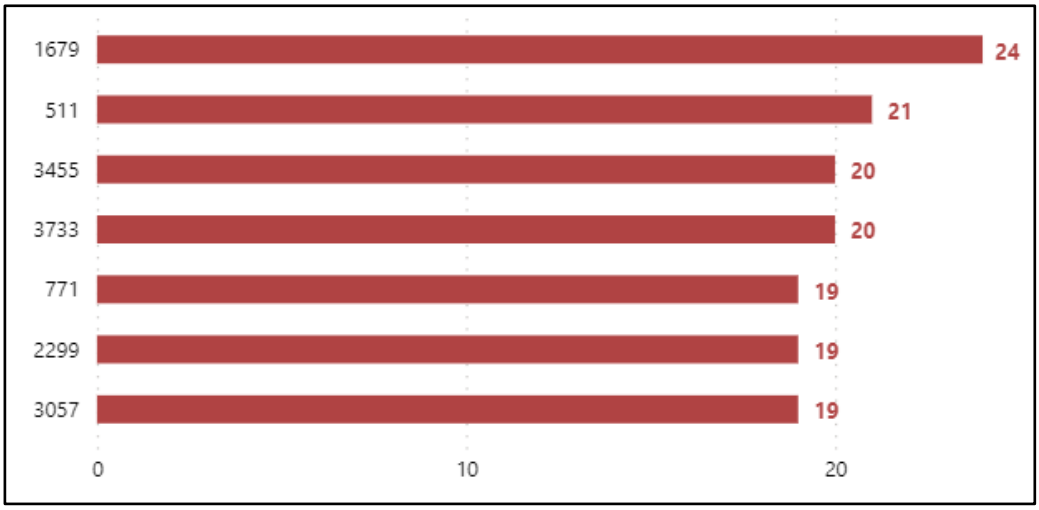


Figure 26. *Top Selling Products*

The two figures above illustrate top products having the most views and the most sales. Products in both top lists do not overlap, indicating that although they are receiving plenty of views, sales may not be as high as expected.

5. Validation

To ensure the number shown is correct, the team has checked in Excel files to validate data and logic calculations.

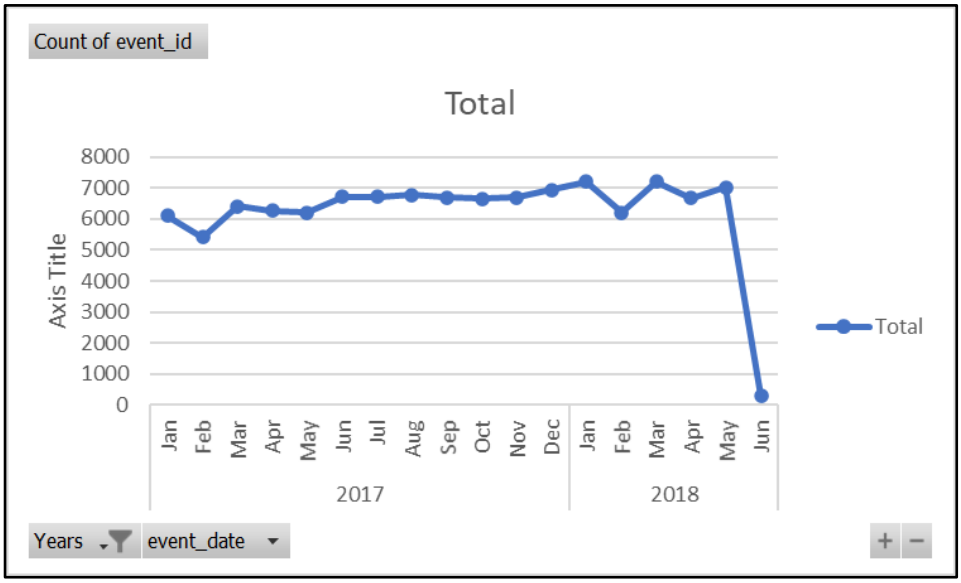


Figure 27. *Number of Sessions Monthly by Excel*

The line chart made by Excel is the same number as the chart in PowerBI. These are the validation images showing the data have matched with the number of visuals made by PowerBI.

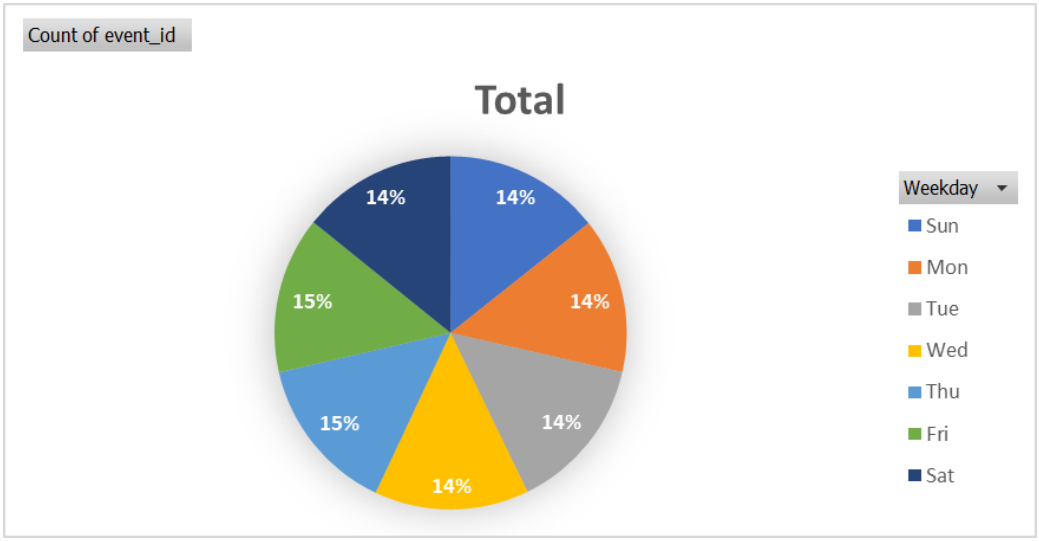


Figure 28. *Number of Session by Weekday in Excel*

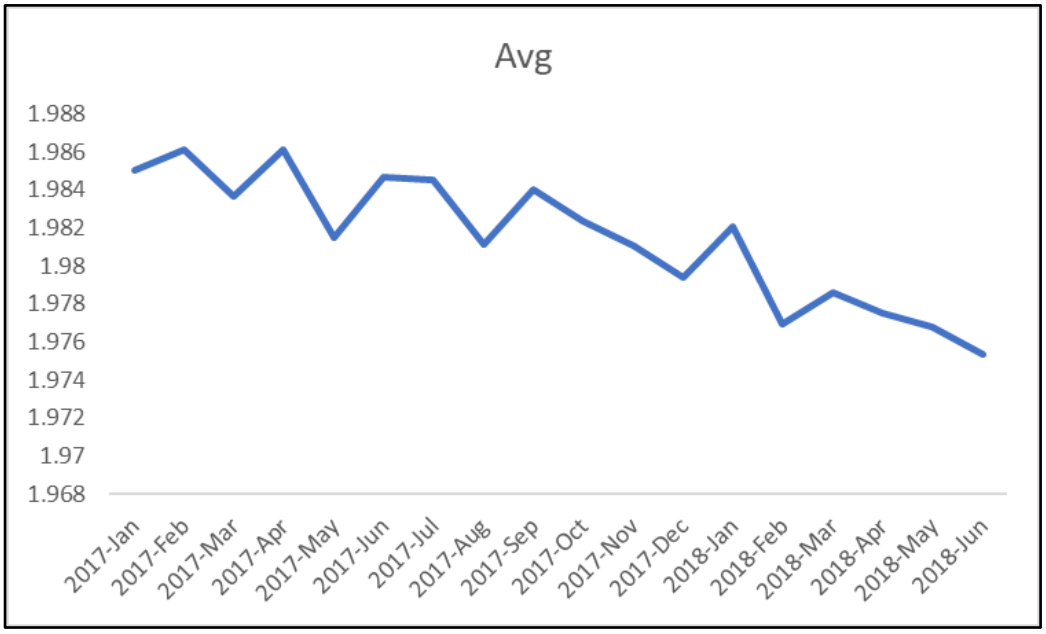


Figure 29. *Average screens per session monthly in Excel.*

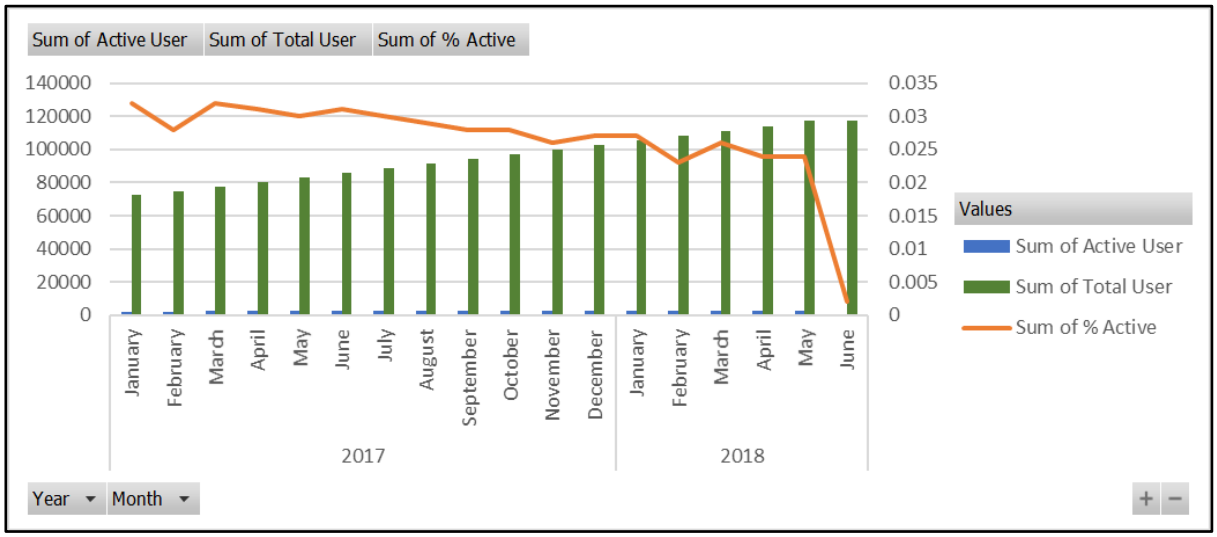


Figure 30. *Active user/ Total user & % Active User Monthly in Excel.*

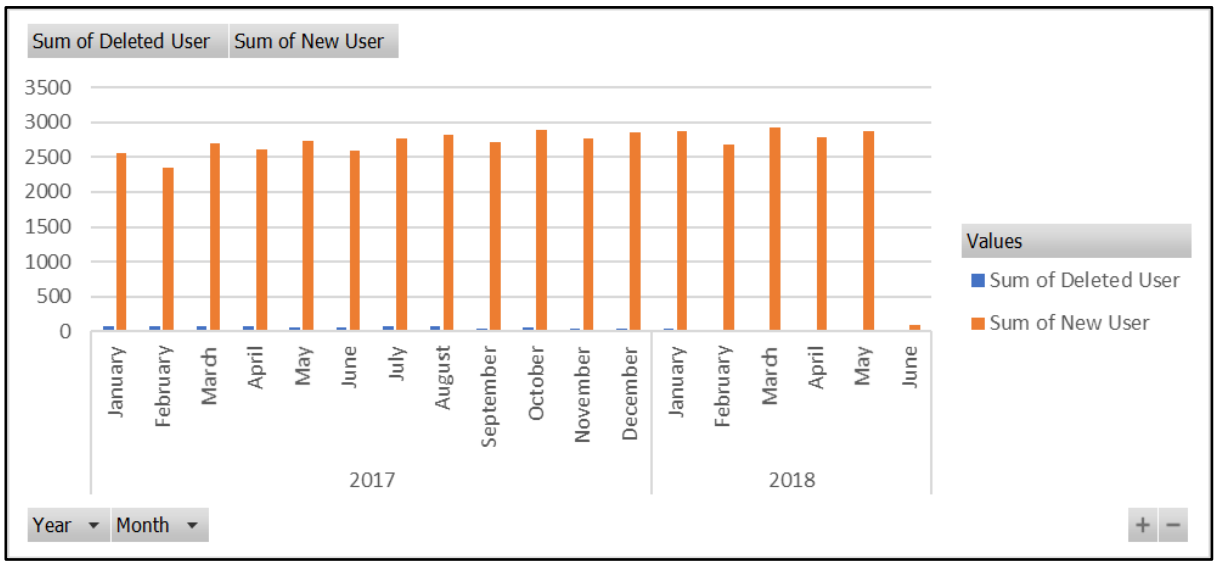


Figure 31. *New Users & Deleted User Monthly in Excel.*

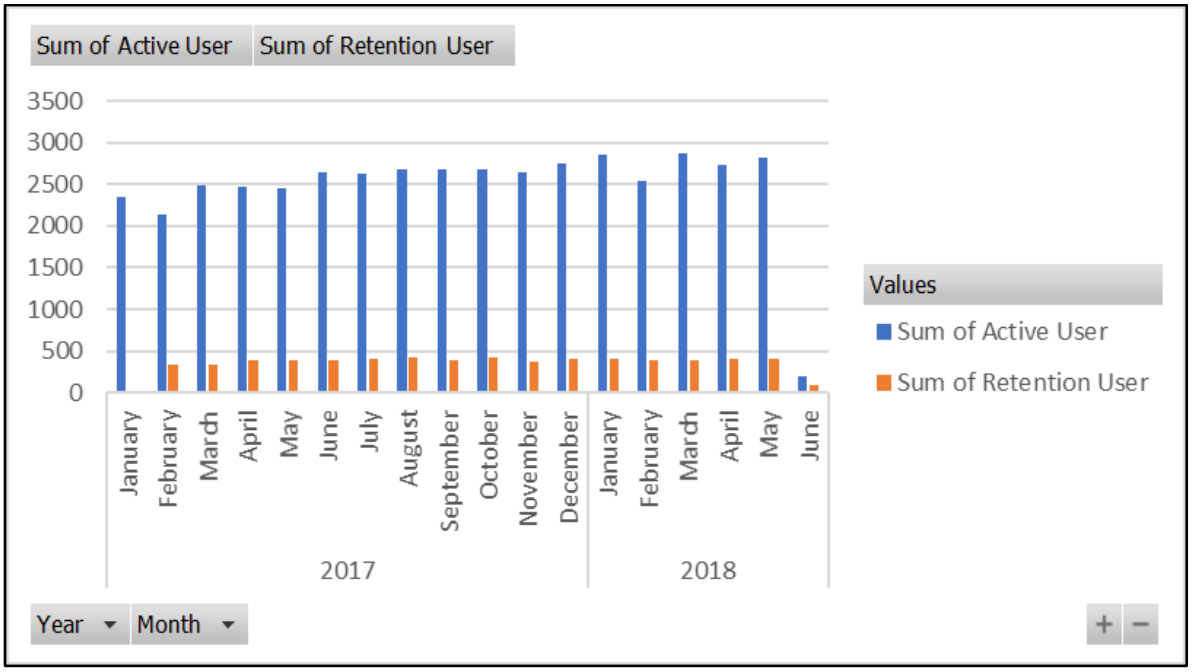


Figure 32. *Retention User & Active User Monthly in Excel.*

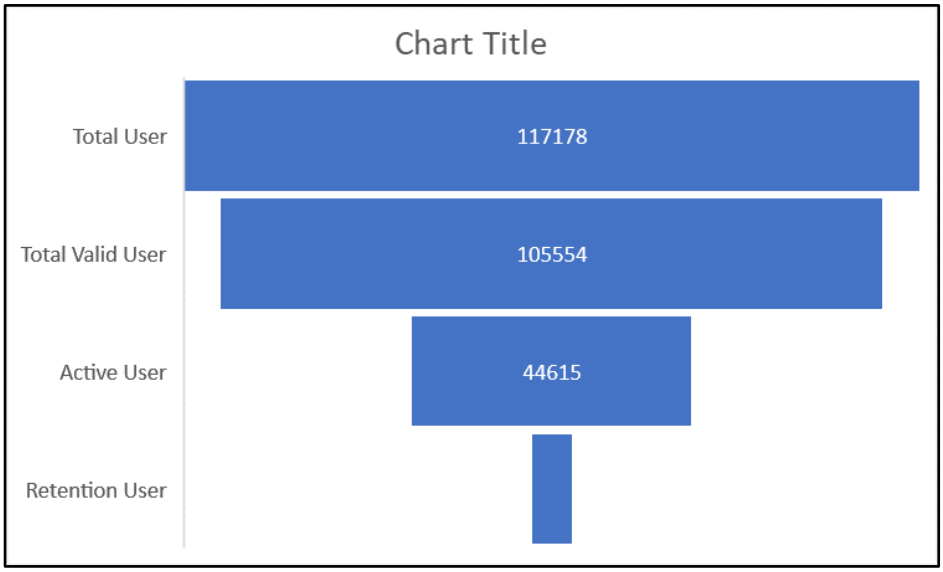


Figure 33. *User Funnel in Excel.*

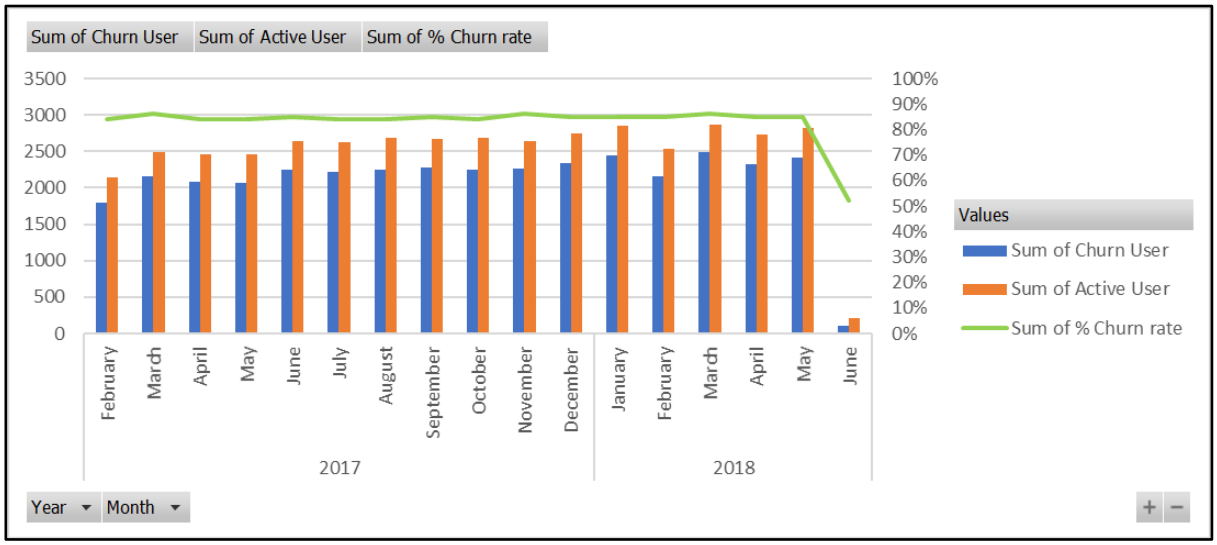


Figure 34. *Churn User & Active User & % Churn Rate Monthly in Excel.*

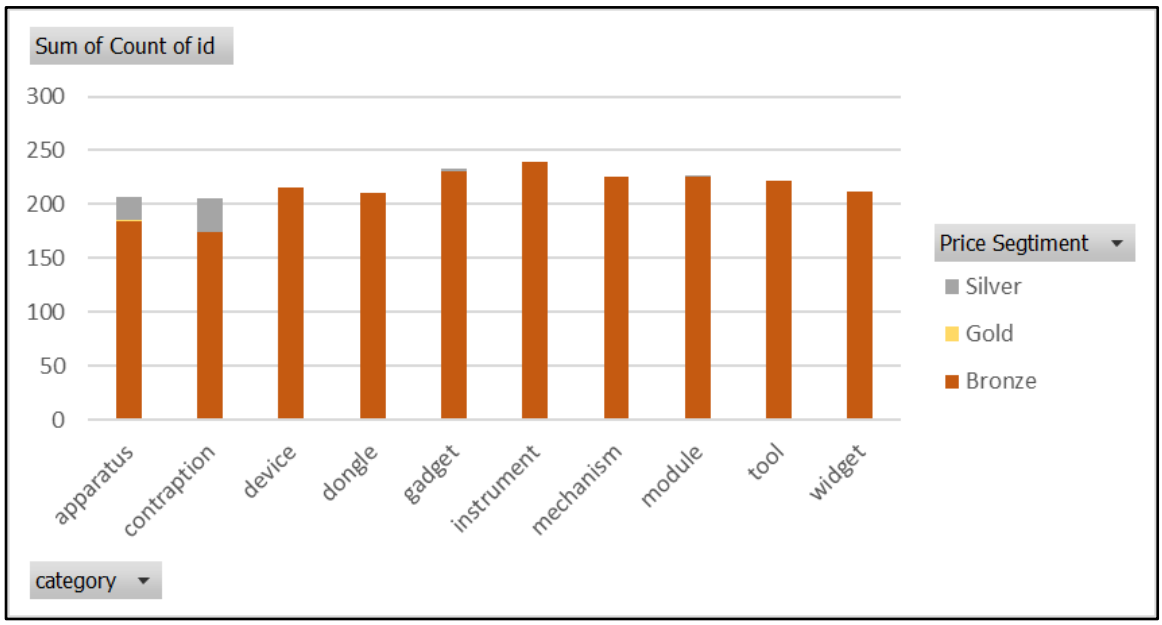


Figure 35. *Number of products by Category, Price Segment in Excel*

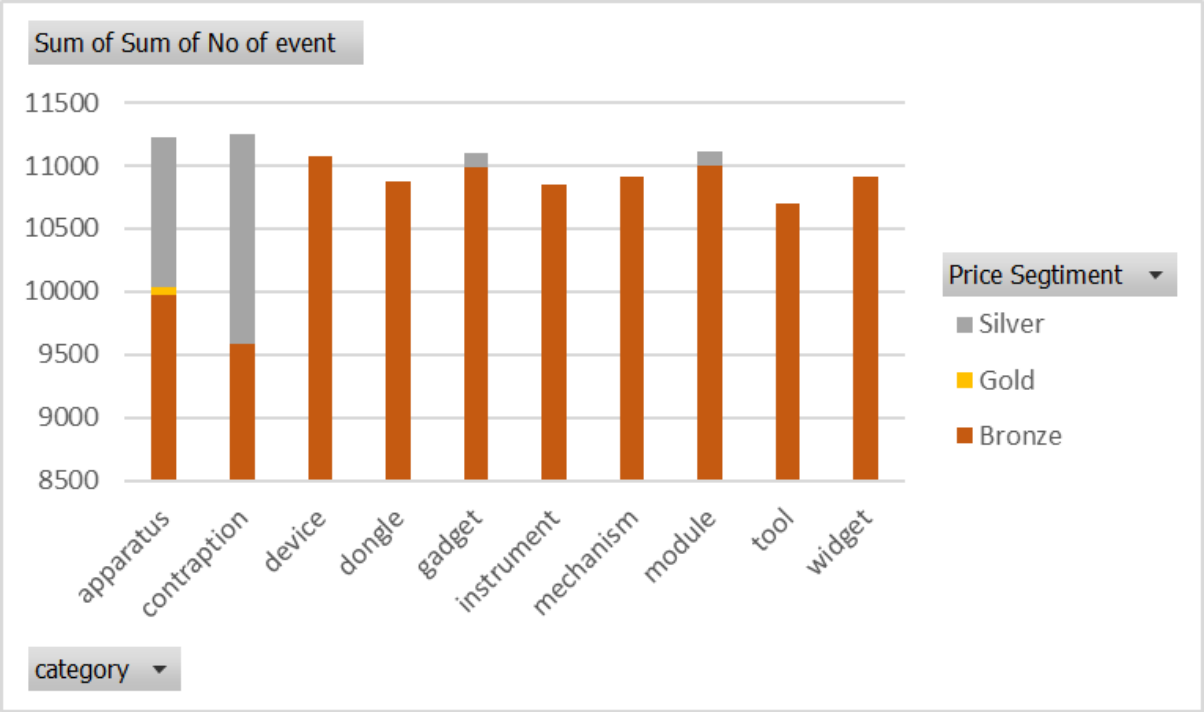


Figure 36. *Number of sessions by Category, Price Segment in Excel.*

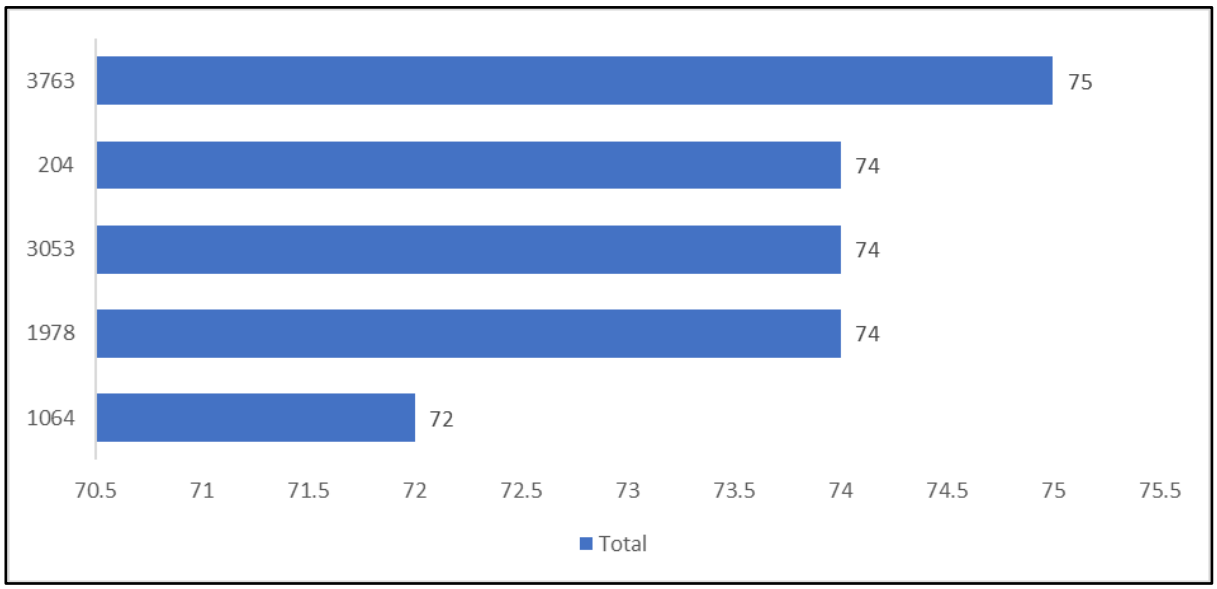


Figure 37. *Top Viewed Products*

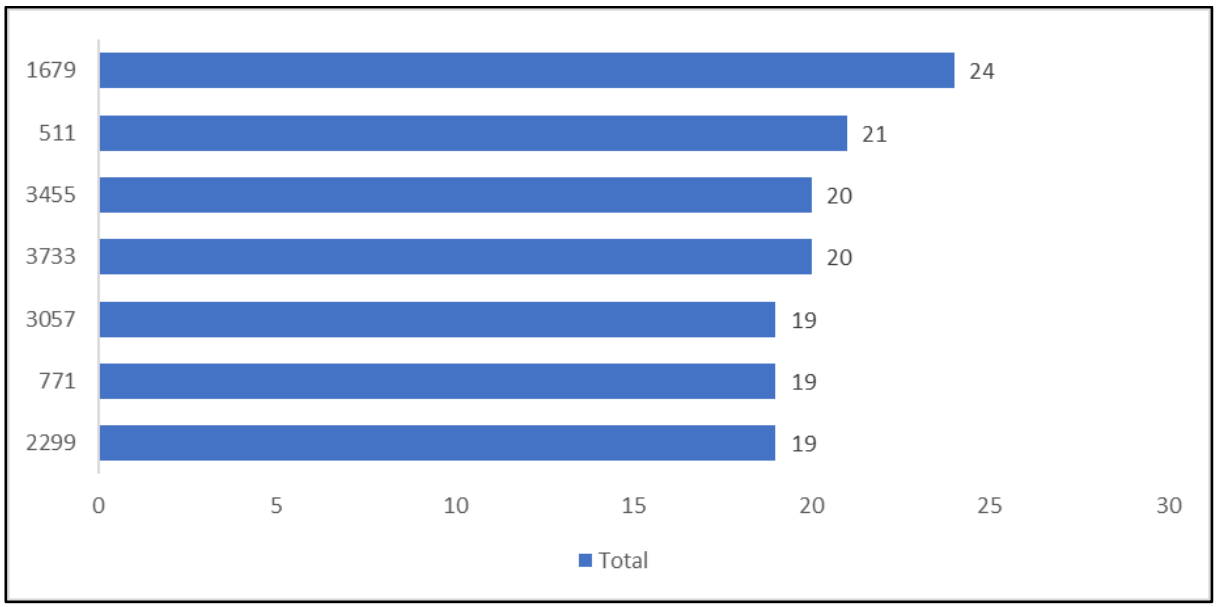


Figure 38. *Top Selling Products*

CHAPTER IV: VISUALIZATION AND REPORT ANALYSIS

In this chapter 4, the team will utilize the exploratory analysis in chapter III and add-in some deep-dive analysis for crafting the data story-telling more informative and attractive.

---

1. Craft a compelling story

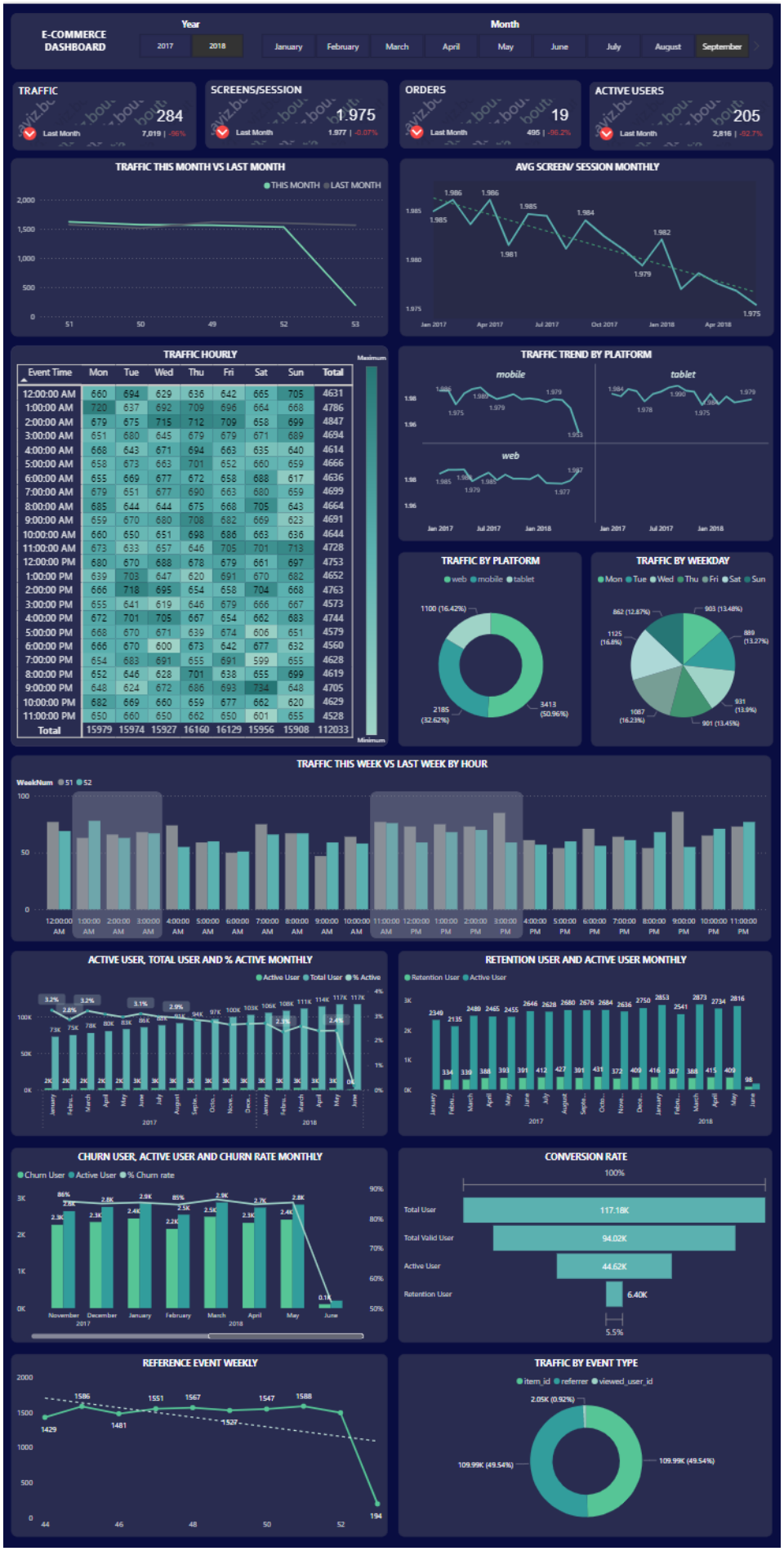

E-COMMERCE DASHBOARD
Year
2017
2018
Month
January
February
March
April
May
June
July
August
September
TRAFFIC
284
Last Month
7,019 | -96%
SCREENS/SESSION
1.975
Last Month
1.977 | -0.07%
ORDERS
19
Last Month
495 | -96.2%
ACTIVE USERS
205
Last Month
2,816 | -92.7%
TRAFFIC THIS MONTH VS LAST MONTH
THIS MONTH
LAST MONTH
AVG SCREEN/ SESSION MONTHLY
TRAFFIC HOURLY
Maximum
Minimum
Event Time | Mon | Tue | Wed | Thu | Fri | Sat | Sun | Total
12:00:00 AM | 660 | 694 | 629 | 636 | 642 | 665 | 705 | 4631
1:00:00 AM | 720 | 637 | 692 | 709 | 696 | 664 | 668 | 4786
2:00:00 AM | 679 | 675 | 715 | 712 | 709 | 658 | 699 | 4847
3:00:00 AM | 651 | 680 | 645 | 679 | 679 | 671 | 689 | 4694
4:00:00 AM | 668 | 643 | 671 | 694 | 663 | 635 | 640 | 4614
5:00:00 AM | 658 | 673 | 663 | 701 | 652 | 660 | 659 | 4666
6:00:00 AM | 655 | 669 | 677 | 672 | 658 | 688 | 617 | 4636
7:00:00 AM | 679 | 651 | 677 | 690 | 663 | 680 | 659 | 4699
8:00:00 AM | 685 | 644 | 644 | 675 | 668 | 705 | 643 | 4664
9:00:00 AM | 659 | 670 | 680 | 708 | 682 | 669 | 623 | 4691
10:00:00 AM | 660 | 650 | 651 | 698 | 686 | 663 | 636 | 4644
11:00:00 AM | 673 | 633 | 657 | 646 | 705 | 701 | 713 | 4728
12:00:00 PM | 680 | 670 | 688 | 678 | 679 | 661 | 697 | 4753
1:00:00 PM | 639 | 703 | 647 | 620 | 691 | 670 | 682 | 4652
2:00:00 PM | 666 | 718 | 695 | 654 | 658 | 704 | 668 | 4763
3:00:00 PM | 655 | 641 | 619 | 646 | 679 | 666 | 667 | 4573
4:00:00 PM | 672 | 701 | 705 | 667 | 654 | 662 | 683 | 4744
5:00:00 PM | 668 | 670 | 671 | 639 | 674 | 606 | 651 | 4579
6:00:00 PM | 666 | 670 | 600 | 673 | 642 | 677 | 632 | 4560
7:00:00 PM | 654 | 683 | 691 | 655 | 691 | 599 | 655 | 4628
8:00:00 PM | 652 | 646 | 628 | 701 | 638 | 655 | 699 | 4619
9:00:00 PM | 648 | 624 | 672 | 686 | 693 | 734 | 648 | 4705
10:00:00 PM | 682 | 669 | 660 | 659 | 677 | 662 | 620 | 4629
11:00:00 PM | 650 | 660 | 650 | 662 | 650 | 601 | 655 | 4528
Total | 15979 | 15974 | 15927 | 16160 | 16129 | 15956 | 15908 | 112033
TRAFFIC TREND BY PLATFORM
mobile
tablet
web
TRAFFIC BY PLATFORM
web
mobile
tablet
1100 (16.42%)
2185 (32.62%)
3413 (50.96%)
TRAFFIC BY WEEKDAY
Mon
Tue
Wed
Thu
Fri
Sat
Sun
862 (12.87%)
903 (13.48%)
1125 (16.8%)
889 (13.27%)
931 (13.9%)
1087 (16.23%)
901 (13.45%)
TRAFFIC THIS WEEK VS LAST WEEK BY HOUR
WeekNum
51
52
ACTIVE USER, TOTAL USER AND % ACTIVE MONTHLY
Active User
Total User
% Active
RETENTION USER AND ACTIVE USER MONTHLY
Retention User
Active User
CHURN USER, ACTIVE USER AND CHURN RATE MONTHLY
Churn User
Active User
% Churn rate
CONVERSION RATE
100%
Total User
117.18K
Total Valid User
94.02K
Active User
44.62K
Retention User
6.40K
5.5%
REFERENCE EVENT WEEKLY
TRAFFIC BY EVENT TYPE
item_id
referrer
viewed_user_id
2.05K (0.92%)
109.99K (49.54%)
109.99K (49.54%)

Figure 39. *Dashboard Appearance*

Dashboard will show the key metrics by month and year that the user will select to view. Main metrics which the users want to track will be shown first and shown by trendline charts. Next, the team will show the factors we have analyzed to find the cause of the trend or some factors that really impact the main metric. The team will next conduct a thorough examination of these aspects.

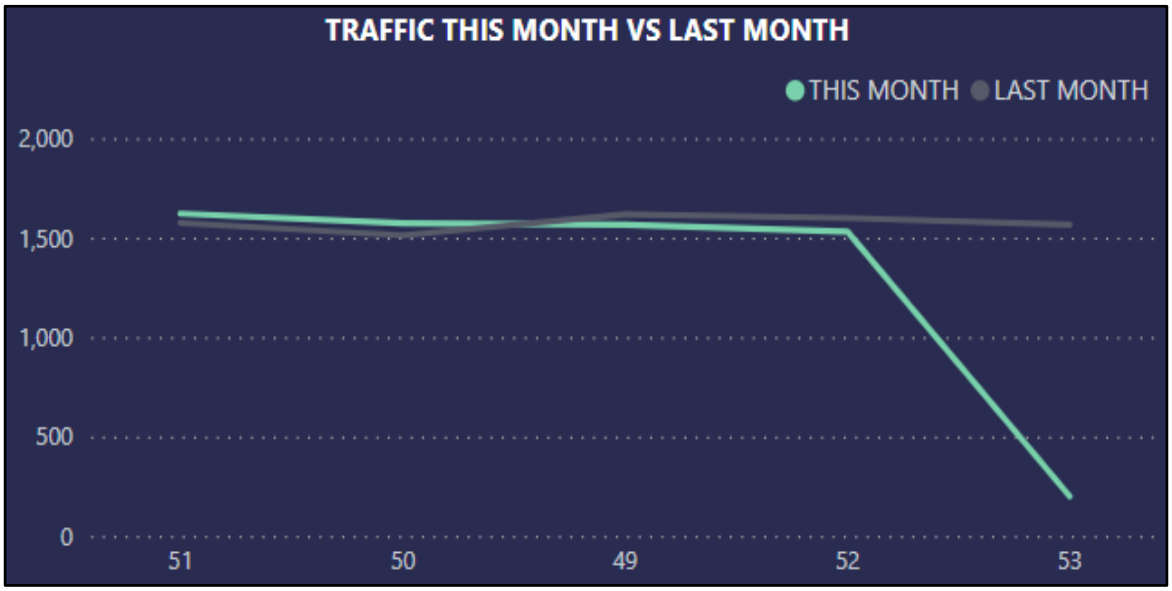


Figure 40. *Number of Sessions of Latest 4 Weeks Comparing with 4-week backwards.*

The most recent month did not experience a rise in the number of sessions, and it is clear that this trend will continue over the next months based on traffic statistics from the last four weeks compared to the four weeks prior.

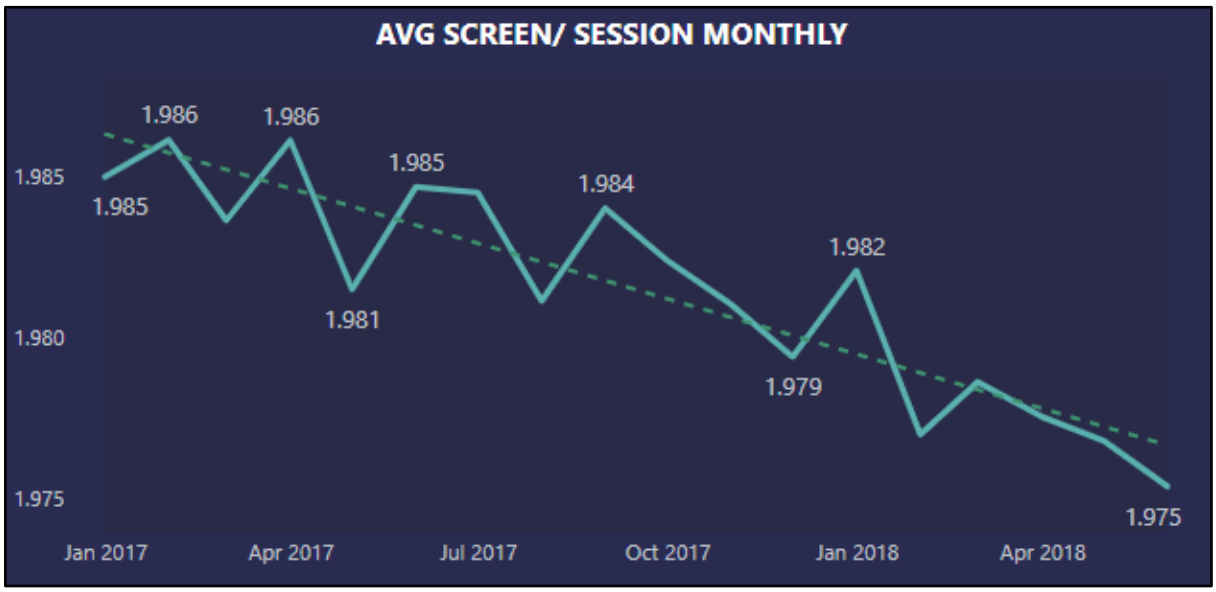


Figure 41. *Average Screens per Session trend line.*

It can be seen that avg screen is trending up but decreasing more immediately after that from 2017 to 2018, starting from April 2018 there was a marked decrease and no sign of increasing again.

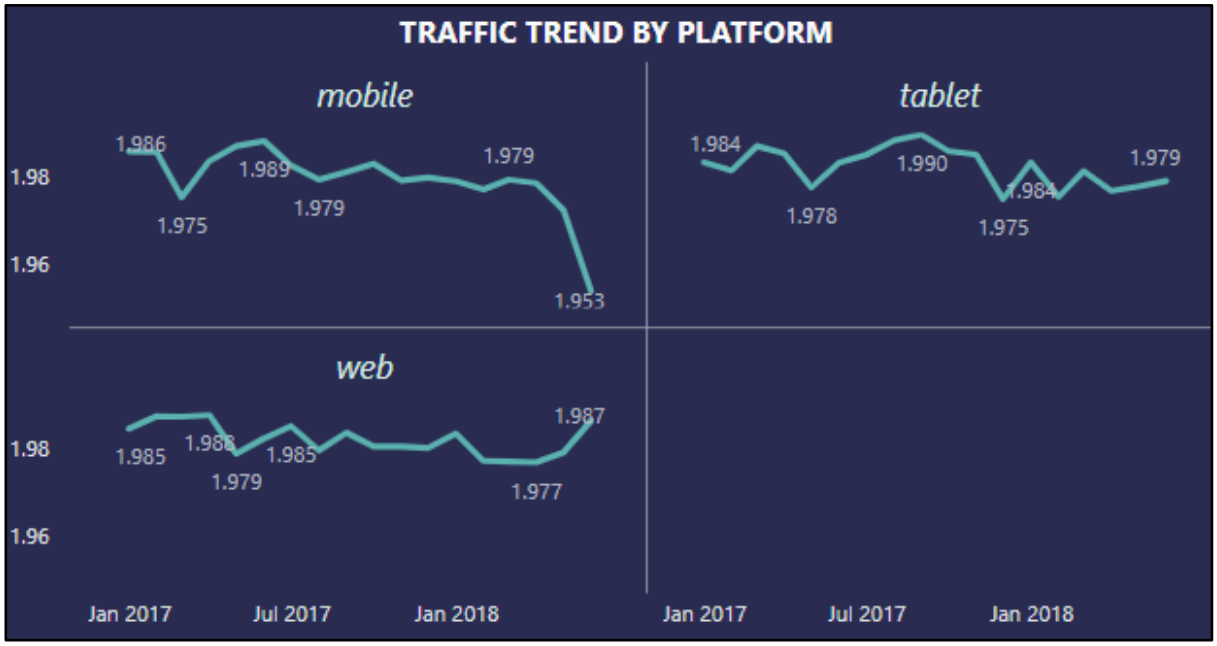
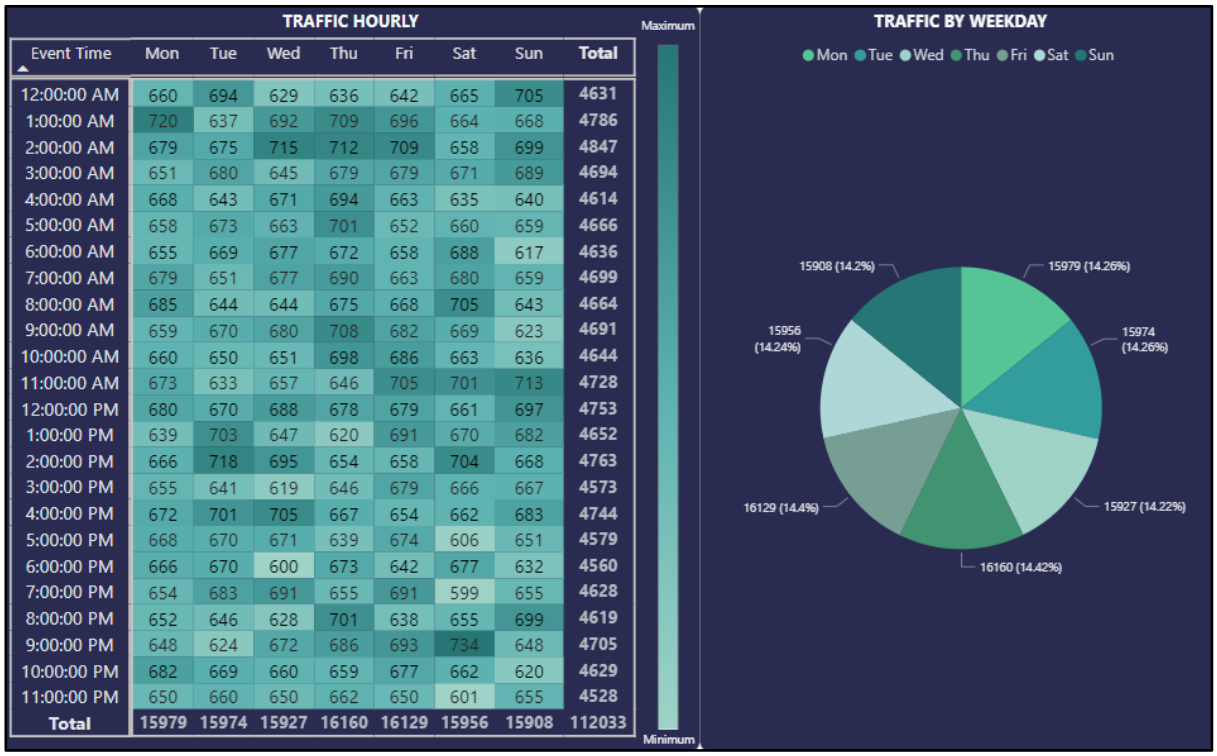


Figure 42. *Average Screens per Session by platform trend line.*

We can observe the number of visitors on 3 platforms: mobile, tablet and web. On the mobile platform, the number of visits decreased markedly and rarely saw an increasing trend. For tablet platforms, the chart still shows a larger decrease, and finally for web platforms, the increase is also negligible, only from 1977 to 1987.



TRAFFIC HOURLY

Event Time	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Total
12:00:00 AM	660	694	629	636	642	665	705	4631
1:00:00 AM	720	637	692	709	696	664	668	4786
2:00:00 AM	679	675	715	712	709	658	699	4847
3:00:00 AM	651	680	645	679	679	671	689	4694
4:00:00 AM	668	643	671	694	663	635	640	4614
5:00:00 AM	658	673	663	701	652	660	659	4666
6:00:00 AM	655	669	677	672	658	688	617	4636
7:00:00 AM	679	651	677	690	663	680	659	4699
8:00:00 AM	685	644	644	675	668	705	643	4664
9:00:00 AM	659	670	680	708	682	669	623	4691
10:00:00 AM	660	650	651	698	686	663	636	4644
11:00:00 AM	673	633	657	646	705	701	713	4728
12:00:00 PM	680	670	688	678	679	661	697	4753
1:00:00 PM	639	703	647	620	691	670	682	4652
2:00:00 PM	666	718	695	654	658	704	668	4763
3:00:00 PM	655	641	619	646	679	666	667	4573
4:00:00 PM	672	701	705	667	654	662	683	4744
5:00:00 PM	668	670	671	639	674	606	651	4579
6:00:00 PM	666	670	600	673	642	677	632	4560
7:00:00 PM	654	683	691	655	691	599	655	4628
8:00:00 PM	652	646	628	701	638	655	699	4619
9:00:00 PM	648	624	672	686	693	734	648	4705
10:00:00 PM	682	669	660	659	677	662	620	4629
11:00:00 PM	650	660	650	662	650	601	655	4528
Total	15979	15974	15927	16160	16129	15956	15908	112033

Figure 43. *Traffic Hourly and by days in week.*

Based on traffic data by time frame of weekdays, we can only see that the traffic have a great similarity between weekdays and are almost equal. But analyzing more deeply into each specific time

frame, the data team realized that the 1AM:3AM and 11AM:12PM time frames mark a larger amount of traffic than other time frames because these are the hours when users have a high demand for workspace or entertainment.

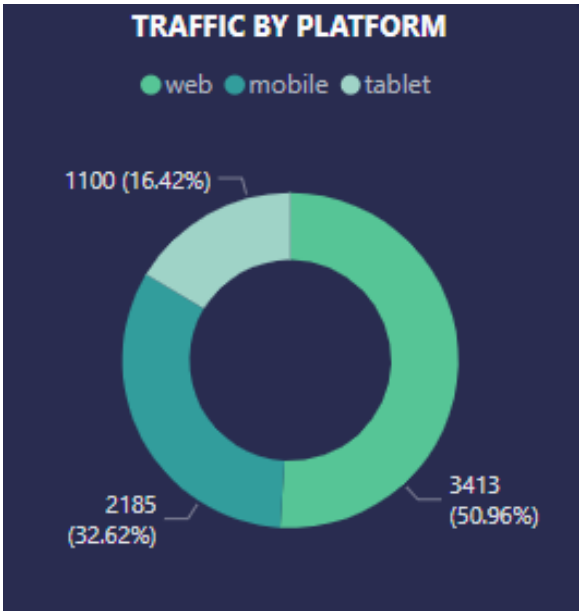


Figure 44. *Traffic by Platform Type.*

Because it is the morning shift's busiest hour and many individuals use laptops to access the internet, the 11 AM to 12 PM time period provides a significant volume of online traffic. In contrast, people typically use their mobile devices for pleasure and convenience in bed after a long day at work, therefore the 1AM–3AM time period saw a spike in the amount of people utilizing their mobile platform to access our app.

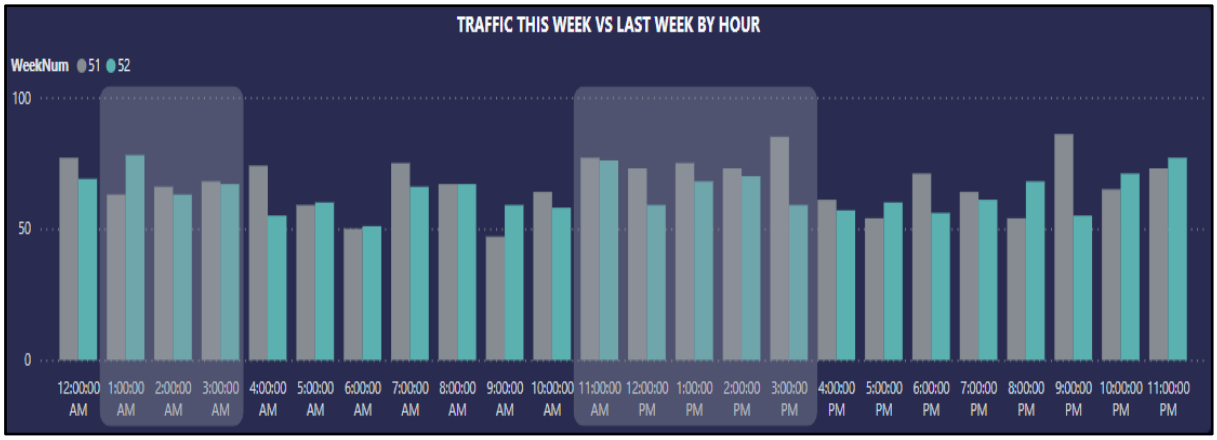


Figure 45. *Traffic of the last 2 weeks by hour.*

` Based on the traffic data of the current week compared to the previous week, we can see that the 1AM-3AM time frame recorded a slight decrease and the 12PM-3PM time frame has a sudden decline and needs attention for business strategic actions to recover.

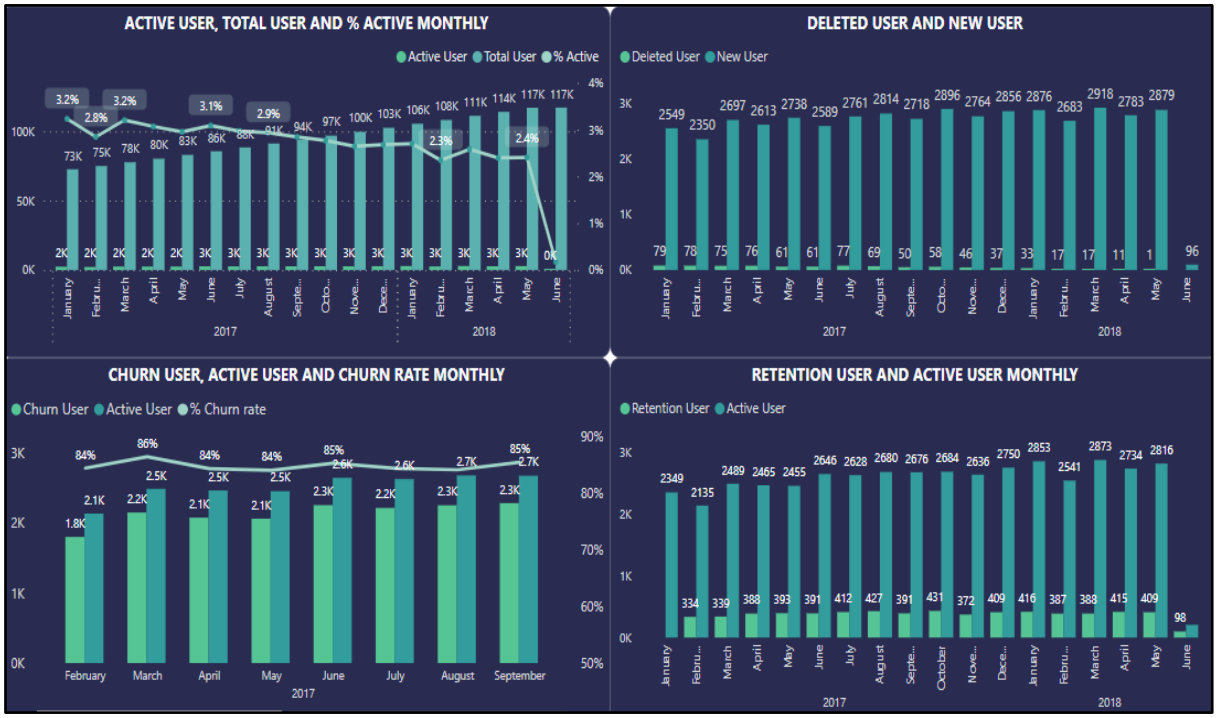


Figure 46. *Active, Deleted, New, Retention and Churn User Monthly in 1-page.*

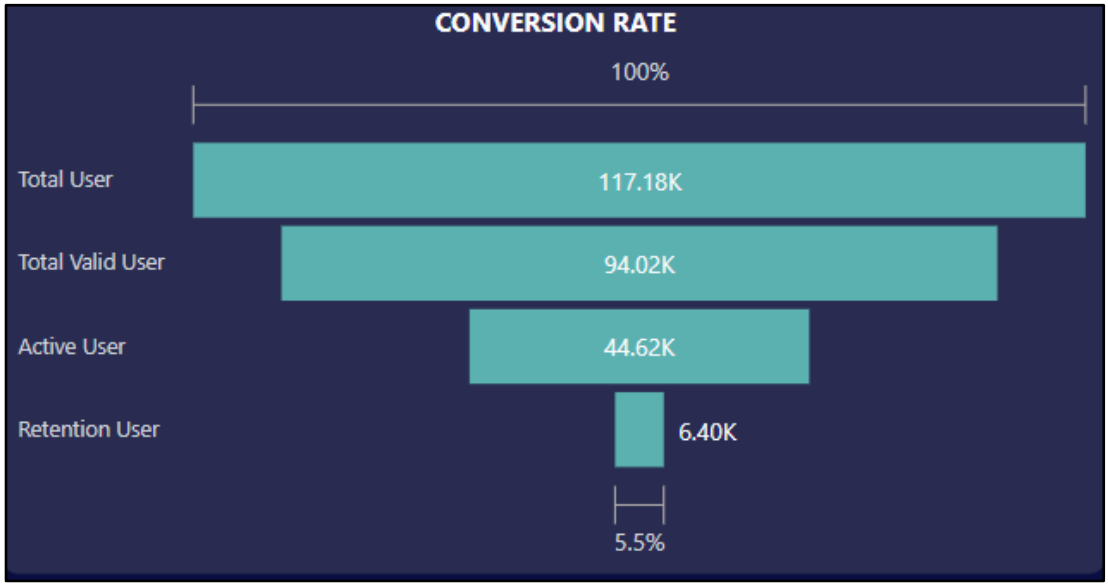


Figure 47. *User Conversion Funnel.*

According to the charts from the dashboard above, we can summarize the causes of the decrease in traffic including:

+ Figure 1 - Active user, total user and % active monthly: It can be seen that a lot of new users are generated, however, the percentage of real active users is very few of them, while active users are customers that will bring traffic.
+ Figure 2 - Deleted user and new user: New users are created a lot but the amount of deleted users do not increase. Although this is a positive point, the marketing team has not taken advantage of this effectively yet.

+ Figure 3 - Churn user, active user and churn rate monthly: the percentage of churn user over all active user is quite alarming, about 85%, which means that most users tend to leave rather than stay on the website .
+ Figure 4 - Retention user and active user monthly: Low customer return rate. According to statistics, 80% of traffic comes from 20% of loyal customers → this is the main reason why traffic drops.
+ Figure 5 - Conversion rate: Compared to the total number of users (about 120K), the number of valid users is only about 94K, of which just over 44K are active users. Retained users are counted as over 6K, which is 5.5% of total users.

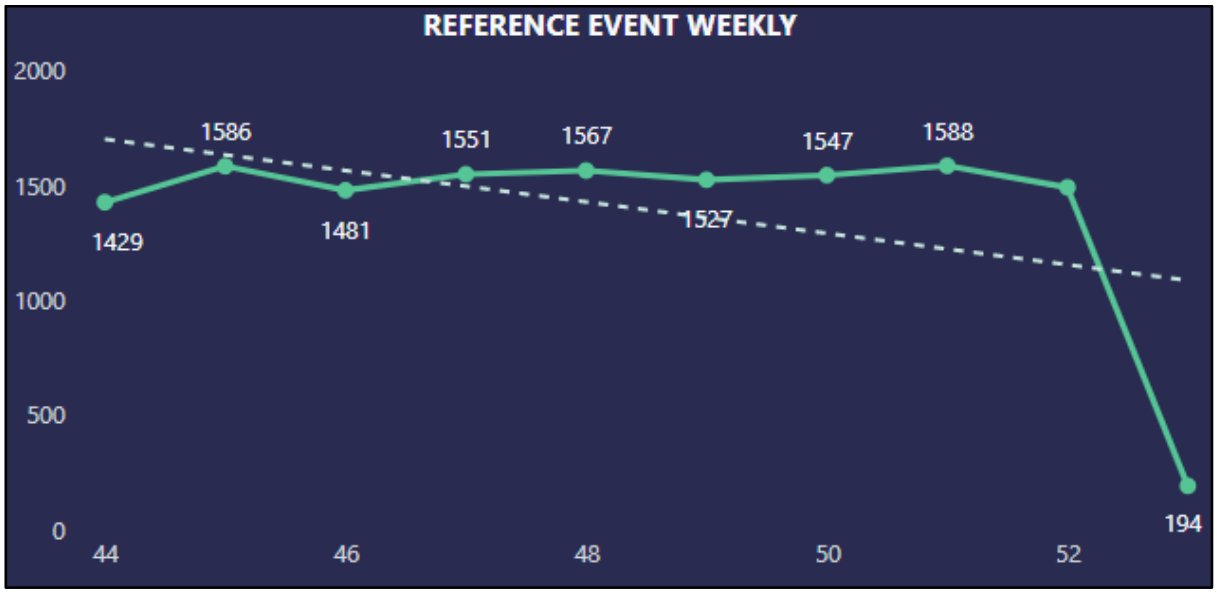


Figure 48. *Referrer Events Last 10 Weeks.*

The first thing you should realize is that how active users heavily depends on Referrer Events of your company. One of the reasons is these events are not attractive enough to attract new users and retain customers. Businesses should need to increase their marketing efforts in order to grow the customer base.

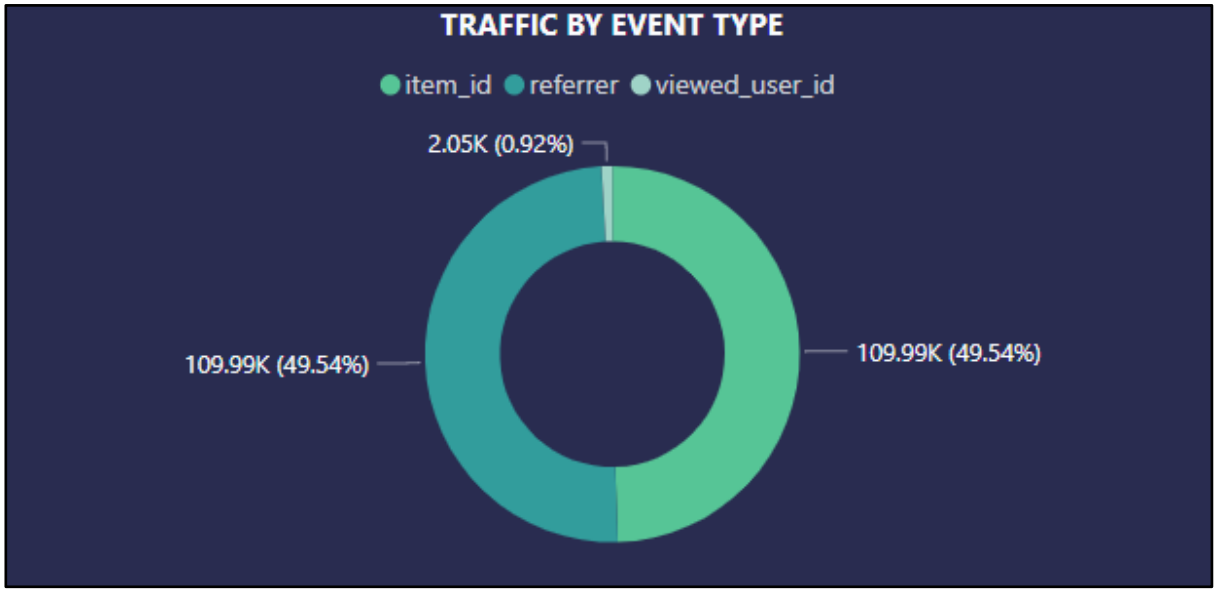


Figure 49. *Traffic by Event Type.*

Traffic mostly comes from watching products and introducing new user, the rate of introducing new people is equal to the rate of watching products, which proves that users only run programs to earn points and products. They would check whether there was an external and internal factor that could have impacted users. An illustration of the external factor is a new competitor's debut or an existing competitor's release of a potent feature.

In terms of category and price segment, there is no difference found when the number of views of the product corresponds to the number of products, that means, the more products the segment has, the easier it is to have more views. The team have tested some analysis that drill each segment into fields such as modifier and adjective to have a quick view and draw some conclusions.

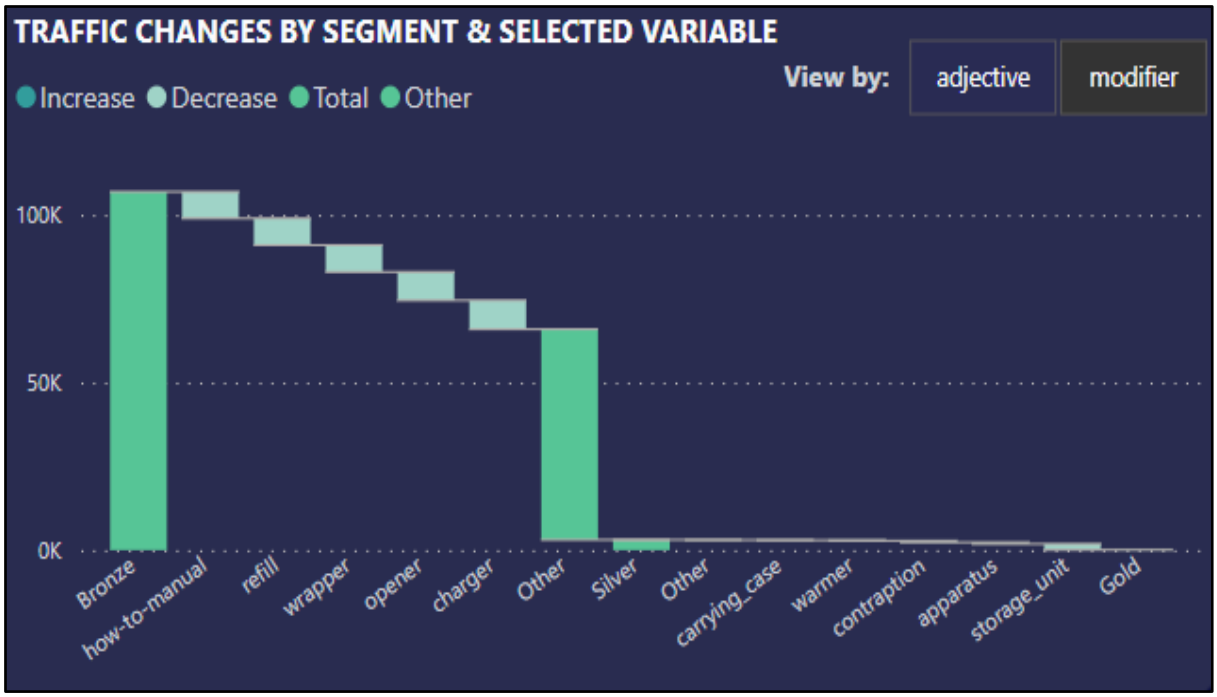


Figure 50. *Waterfall chart of traffic by price segment and modifier.*

In general, all price segments have a decrease in traffic for modifiers. The Bronze segment has the strongest decline, followed by the Silver segment and finally the Gold segment. Products with modifiers in a cheaper segment have the largest amount of traffic but also have the most drastic decrease, focusing on some modifiers including how-to-manual, refill, wrapper, opener, charger.

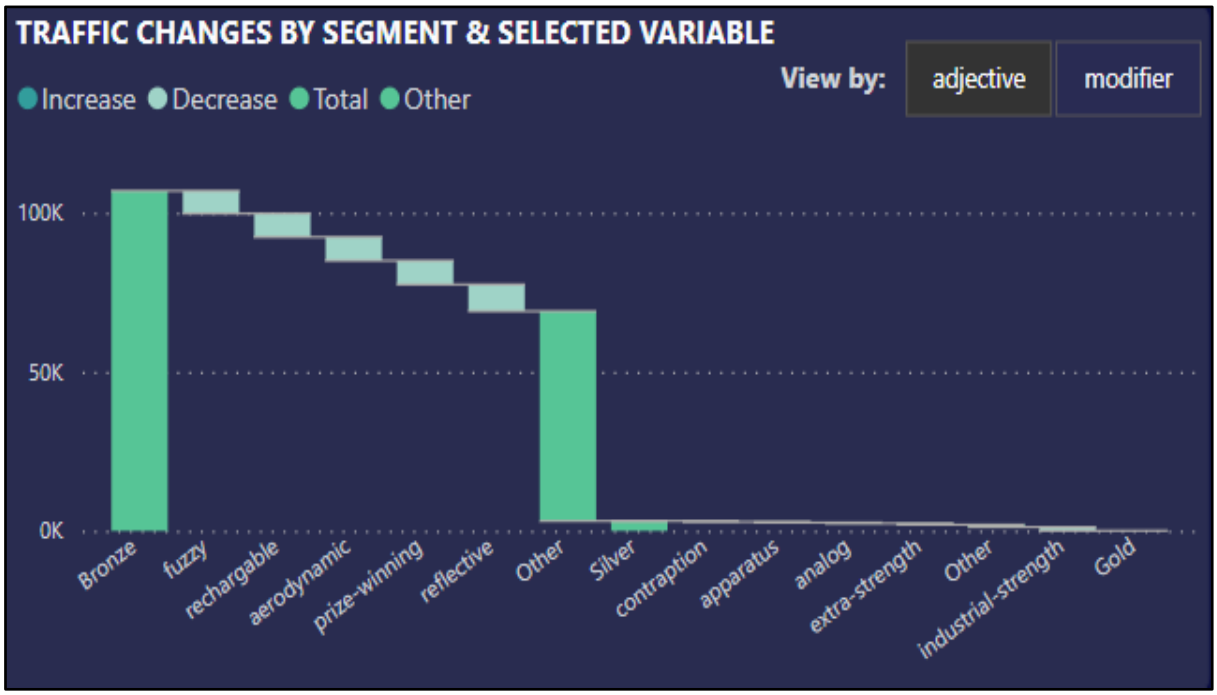


Figure 51. *Waterfall chart of traffic by price segment and adjective.*

Most of the products with modifiers above have a decrease in traffic in many adjectives. The traffic changed in a negative way, The Bronze segment continued to have the biggest drop, focusing on some adjectives including fuzzy, rechargeable, aerodynamic, prize-winning, reflective.

TOP 5 VIEWED PRODUCT

Product ID	Sessions
1521	11
2072	10
3676	10
208	9
963	9
1590	9
3071	9
3333	9
3701	9
3878	9

TOP 5 SELLING PRODUCT

Product ID	Orders
3601	4
1443	3
1676	3
3733	3
3461	2

Figure 52. *Top 5 viewed products and top 5 selling products.*

Items that attract a lot of traffic but receive little purchases, as seen by the graph of the top 5 most viewed products and the top 5 best sellers. Bounce rates - the proportion of visitors who briefly browse a product page before leaving - have been rising over time. The reason is that even if there are a lot of visits due to the poor conversion rate, there are less buyers for this product than for others even though there are fewer views.

Conclusion: Traffic in this month declined because the impact of these factors below:

- Platform: Tablet and Mobile.

- Timeframe: 1AM:3AM and 11AM:12PM are the timeframes brings users and traffic the most in the past but they have decreased recently.

- Churn rate is high and retention rate is low as well.

- Reference Program is not interesting enough to bring more user.

- Modifier and Adjective: a list of modifiers and adjectives above.

2. Security Management

The team will intervew the business team to get the list of users which they want to share the report and dashboard with. To ensure less risky and high data governance, the data team also wants to know the role of each person and their detailed permission to grant access to the dashboard and set rules for them. After working closely with the business team, we finally have the users' list.

Name	Email	Role	Permission
Nguyen Van A	ngvana@abc.company	E-commerce & Strategy Planning Manager	Full
Nguyen Van B	ngvanb@abc.company	E-commerce Assistant Manager	Full
Nguyen Thi C	ngvanb@abc.company	Consumer Experience Manager	Full
Nguyen Thi D	ngvanD@abc.company	Brand Manager (Bronze Group)	Bronze Segment
Nguyen Thi E	ngvanE@abc.company	Brand Manager (Silver Group)	Silver Segment
Nguyen Thi F	ngvanF@abc.company	Brand Manager (Gold Group)	Bronze Segment

Table 6. *List of users and their permission.*

Based on the table above, the data team will set-up all charts on the dashboard to show correspondingly to each user. The user when using dashboards will be aware that they can see that

data that are in their permissions. Therefore, some error charts are shown because they are not for their usage and not in their permissions as well.

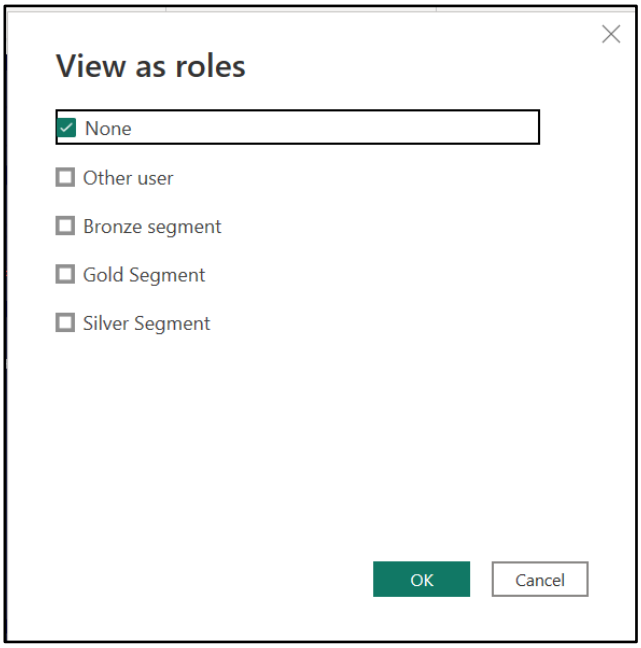


Figure 53. *List of view as user role*

The team has added the list that the business team giving and defined 4 groups will access the dashboard. The Bronze group only sees the numeric related to the product belonging to the bronze segment.There are the similar to the Silver and Gold group. The Group name Other is a group of managers who have full permissions and make a decision based on the overall picture the data team has visualized.

CHAPTER V: DISCUSSION AND IMPLICATIONS MANAGEMENT

The project has applied diagnostic analysis to assist the company S find out the reason why their traffic decreased. The in-depth investigation also pointed out the factor impacting the decline, the root cause of the problem and showed the way how they impact the traffic. In addition, the team showed some factors that don't have any bearing on the traffic. Then, in order to reinforce the factors that affect traffic, we have broken them down into small factors to monitor more carefully and deeply to get the most specific and accurate diagnoses possible. Multifactorial diagnostic project has shown that there are many factors that interact with each other, they are not independent.

The project's constraints, however, prevent the data team from being able to meet the requirements from the business team. We have not examined how the traffic effect on the sales and some fields in datasets were not unused. We also realized that the interaction between the filter and the charts or among the charts are not running perfectly. It has some errors that make the team has to disable the interaction among them to avoid displaying the incorrect number because of the unstandardized data model.

REFERENCES

1. Microsoft (2022, August 10). Configure row-level security with the static method. Retrieved January 28, 2023, from https://learn.microsoft.com/en-us/training/modules/row-level-security-power-bi/2-static-method

2. Verma, R., Sharma, D., Chhabra, M., & Chola, A. (2022, March 21). Power BI Row Level Security: Made Easy - Learn | Hevo. Hevo Data. Retrieved January 28, 2023, from https://hevodata.com/learn/power-bi-row-level-security/

3. Bahrynovska, T. (2022, January 4). *Business Intelligence Strategy - Everything You Need to Know*. Forbytes. Retrieved December 31, 2022, from https://forbytes.com/blog/business-intelligence-strategy/

4. *Business intelligence: A complete overview*. (n.d.). Tableau. Retrieved December 31, 2022, from https://www.tableau.com/learn/articles/business-intelligence#what-is

5. *Business Intelligence Strategy: Creating Your BI Roadmap*. (2020, November 10). AltexSoft. Retrieved December 31, 2022, from https://www.altexsoft.com/blog/business-intelligence-strategy/

6. *Top 7 benefits of business intelligence*. (n.d.). Tableau. Retrieved December 31, 2022, from https://www.tableau.com/learn/articles/business-intelligence/enterprise-business-intelligence/benefits

7. Tom Robertson (May 18, 2020). What is Diagnostic Analysis; Steps to do this. Retrieved January 15, 2023, from https://www.datascienceacademy.io/blog/what-is-diagnostic-analysis;-steps-to-do-this/

8. Marc Holliday (December 9, 2021). What Is Diagnostic Analytics? How It Works and Examples. Retrieved January 15, 2023, from https://www.netsuite.com/portal/resource/articles/data-warehouse/diagnostic-analytics.shtml

9. Buuck, B. (2022, May 31). Schemas Used in Data Warehouses: Star, Galaxy, and Snowflake. StreamSets. Retrieved January 29, 2023, from https://streamsets.com/blog/schemas-data-warehouses-star-galaxy-snowflake/

10. Oracle. (n.d.). What Is a Data Warehouse? Oracle. Retrieved January 29, 2023, from https://www.oracle.com/database/what-is-a-data-warehouse/

11. Oracle (n.d) What is a Data Mart? Oracle. Retrieved January 29, 2023, from https://www.oracle.com/autonomous-database/what-is-data-mart/

12. Oracle (n.d) What is OLTP? Oracle. Retrieved January 29, 2023, from https://www.oracle.com/database/what-is-oltp/