

MY JOURNEY

as a **Data Analyst Intern**

Jeevallucas Gautama

PENNYU ANAK INTERN BATCH 1
PENNYU GROUP



Hello, I'm
Lucas

Data Analyst Intern at Pennyu Group
Aug 2023 - Nov 2023

I am passionate about transforming raw data into actionable insights to drive strategic decisions. During my internship at Pennyu Group, I have honed my skills in data analysis, visualization, and **customer behavior research**, contributing to impactful projects that **enhance marketing strategies and operational efficiency.**





Education



Universitas Kristen Duta Wacana

Aug 2021 - 2025 (expected)

Bachelor Degree in Information Systems
GPA: 3.83/4.00



Zenius

Aug 2023 - Dec 2023

Data Analytics Program:
Completed the program with grade of “A”
across all modules and managed to be
candidates for Group Final Project
Competition Winner.

Experience



Data Analyst Intern

Pennyu Group

2024

Contributed to customer segmentation and sales trend analysis project, focusing on optimizing marketing strategies and ad effectiveness.

Skills & Abilities

Hard Skills:



Microsoft Excel

- Power Query
- Basic data cleaning



SQL

- Extracting data
- Join tables
- Common table expressions



Python

- Library: Numpy, Pandas, sklearn
- Advanced data cleaning
- EDA



Microsoft Power BI

- DAX
- Visualization
- Dashboard



ClickUp & Notion

- Project Collaboration
- Project Management

Soft Skills:

- Problem Solving: Defining Problem, MECE, 5 Whys Framework
- Project Management: Collaboration, Agile Framework
- Data Storytelling: Creating a story with data-narrative-visuals

NOTABLE TASKS

Excel Skill Set

- **Basic SQL:**
 - Data entry, formatting, and basic calculations using formulas (SUM, AVERAGE, IF).
 - Removing duplicates and handling blank cells.
 - Sorting and filtering data for quick analysis.
- **Intermediate SQL:**
 - Using Power Query for importing, transforming, and combining data from various sources.
 - Applying advanced functions like VLOOKUP, HLOOKUP, INDEX-MATCH, and TEXT.
 - Basic pivot table creation for summarizing and analyzing data.
- **Advanced SQL:**
 - Automating repetitive tasks using macros and VBA.
 - Advanced data cleaning techniques like splitting text, combining columns, and handling inconsistent formats using Power Query.

Excel Skill Set

- Utilized Power Query in Microsoft Excel on 2024 Pennyu Marketplace Sales Data.
- Performed basic data cleaning tasks, including:
 - Removing unnecessary columns.
 - Changing column data types to match analytical needs.
 - Replacing values in currency columns (e.g., removing decimal points to correct formatting for whole numbers).
- Handling blank cells to ensure data consistency.

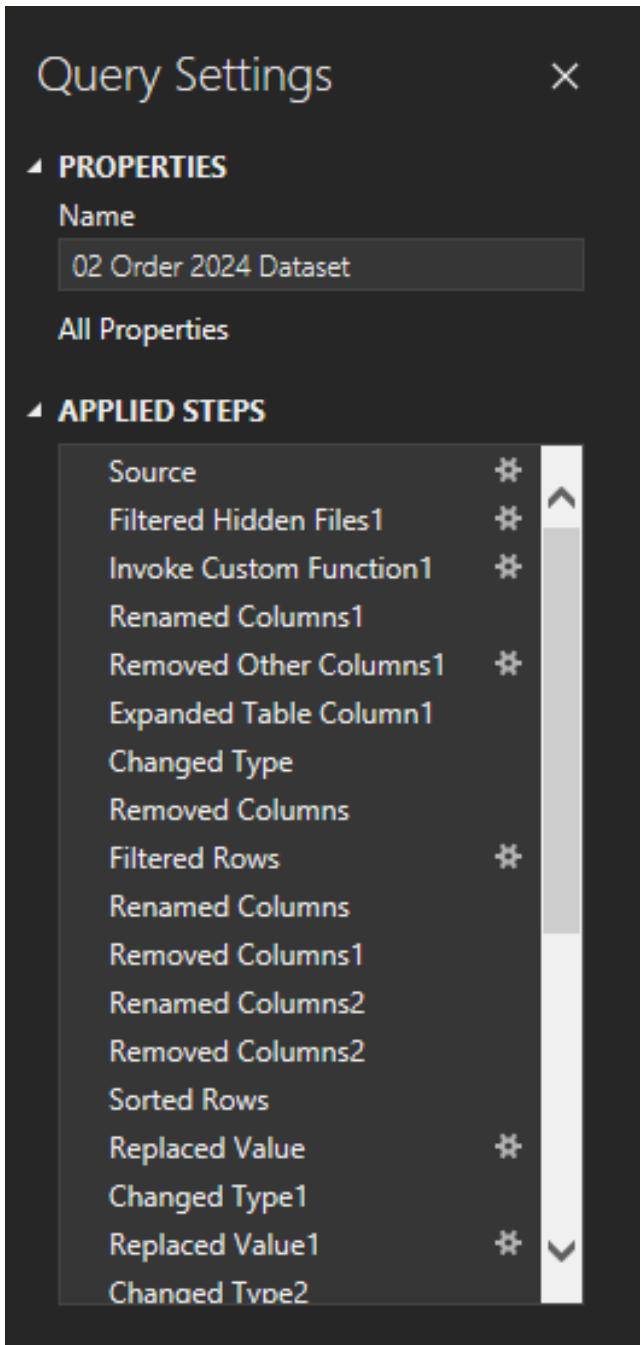
Power Query Advanced Editor

```
1 let
2     Source = Folder.Files("D:\Pennyu Group\Week 3\Project 1\01 Raw Datasets\02 Order 2024 Dataset"),
3     "#Filtered Hidden Files1" = Table.SelectRows(Source, each [Attributes]?[Hidden] <> true),
4     "#Invoke Custom Function1" = Table.AddColumn("#Filtered Hidden Files1", "Transform File", each "#Transform File"([Content])),
5     "#Renamed Columns1" = Table.RenameColumns(#"Invoke Custom Function1", {"Name", "Source.Name"}),
6     "#Removed Other Columns1" = Table.SelectColumns(#"Renamed Columns1", {"Source.Name", "Transform File"}),
7     "#Expanded Table Column1" = Table.ExpandTableColumn(#"Removed Other Columns1", "Transform File", Table.ColumnNames("#Transform File"("Sample File"))),
8     "#Changed Type" = Table.TransformColumnTypes(#"Expanded Table Column1", {"Source.Name", type text}, {"No. Pesanan", type text}, {"Alasan Pembatalan", type text}, {"Status Pembatalan/ Pengembalian", type text}, {"No. Resi", type text}, {"Opsi Pengiriman", type text}, {"Antar ke counter/ pick-up", type text}, {"Pesanan Harus Dikirimkan Sebelum (Menghindari keterlambatan)", type datetime}, {"Waktu Pengiriman Diatur", type datetime}, {"Waktu Pesanan Dibuat", type datetime}, {"Waktu Pembayaran Dilakukan", type text}, {"Metode Pembayaran", type text}, {"SKU Induk", type text}, {"Nama Produk", type text}, {"Nomor Referensi SKU", type text}, {"Nama Variasi", type text}, {"Harga Awal", type text}, {"Harga Setelah Diskon", type text}, {"Jumlah", Int64.Type}, {"Returned quantity", type any}, {"Total Harga Produk", type text}, {"Total Diskon", type text}, {"Diskon Dari Penjual", type text}, {"Diskon Dari Shopee", Int64.Type}, {"Berat Produk", type text}, {"Jumlah Produk di Pesan", Int64.Type}, {"Total Berat", type text}, {"Voucher Ditanggung Penjual", Int64.Type}, {"Cashback Koin", Int64.Type}, {"Voucher Ditanggung Shopee", type number}, {"Paket Diskon (Diskon dari Shopee)", Int64.Type}, {"Paket Diskon (Diskon dari Penjual)", Int64.Type}, {"Potongan Koin Shopee", Int64.Type}, {"Diskon Kartu Kredit", Int64.Type}, {"Ongkos Kirim Dibayar oleh Pembeli", type number}, {"Estimasi Potongan Biaya Pengiriman", type text}, {"Ongkos Kirim Pengembalian Barang", Int64.Type}, {"Total Pembayaran", type text}, {"Perkiraaan Ongkos Kirim", type text}, {"Catatan dari Pembeli", type text}, {"Catatan", type any}, {"Username (Pembeli)", type text}, {"Nama Penerima", type text}, {"No. Telepon", Int64.Type}, {"Alamat Pengiriman", type text}, {"Kota/Kabupaten", type text}, {"Provinsi", type text}, {"Waktu Pesanan Selesai", type datetime}),
9     "#Removed Columns" = Table.RemoveColumns(#"Changed Type", {"Source.Name}),
10    "#Filtered Rows" = Table.SelectRows(#"Removed Columns", each ([Status Pesanan] <> "Batal")),
11    "#Renamed Columns" = Table.RenameColumns(#"Filtered Rows", {"Alasan Pembatalan", "Dibatalkan Oleh"}, {"Status Pembatalan/ Pengembalian", "Alasan Pembatalan"}),
12    "#Removed Columns1" = Table.RemoveColumns(#"Renamed Columns", {"Returned quantity", "Diskon Dari Shopee"}),
13    "#Renamed Columns2" = Table.RenameColumns(#"Removed Columns1", {"Berat Produk", "Berat Produk (gr)"}, {"Total Berat", "Total Berat (gr)"}, {"Total Berat (gr)", "Total Berat (gr)"}, {"Paket Diskon (Diskon dari Shopee)", "Paket Diskon (Diskon dari Penjual)"}, {"Potongan Koin Shopee", "Diskon Kartu Kredit"}, {"Ongkos Kirim Dibayar oleh Pembeli", "Ongkos Kirim Pengembalian Barang"}, {"Catatan", "Catatan"}),
14    "#Removed Columns2" = Table.RemoveColumns(#"Renamed Columns2", {"Voucher Ditanggung Penjual", "Cashback Koin"}, {"Voucher Ditanggung Shopee", "Paket Diskon"}, {"Paket Diskon (Diskon dari Shopee)", "Paket Diskon (Diskon dari Penjual)"}, {"Potongan Koin Shopee", "Diskon Kartu Kredit"}, {"Ongkos Kirim Dibayar oleh Pembeli", "Ongkos Kirim Pengembalian Barang"}, {"Catatan"}),
15    "#Sorted Rows" = Table.Sort(#"Removed Columns2", {"Waktu Pesanan Dibuat", Order.Ascending}),
16    "#Replaced Value" = Table.ReplaceValue(#"Sorted Rows", "", "", Replacer.ReplaceText, {"Harga Awal"}),
17    "#Changed Type1" = Table.TransformColumnTypes(#"Replaced Value", {"Harga Awal", Int64.Type}),
18    "#Replaced Value1" = Table.ReplaceValue(#"Changed Type1", "", "", Replacer.ReplaceText, {"Harga Setelah Diskon"}),
19    "#Changed Type2" = Table.TransformColumnTypes(#"Replaced Value1", {"Harga Setelah Diskon", Int64.Type}),
20    "#Replaced Value2" = Table.ReplaceValue(#"Changed Type2", "", "", Replacer.ReplaceText, {"Total Harga Produk"}),
21    "#Changed Type3" = Table.TransformColumnTypes(#"Replaced Value2", {"Total Harga Produk", Int64.Type})
```

✓ No syntax errors have been detected.

Done Cancel

Power Query Applied Steps



SQL Skill Set

- **Basic SQL:**
 - Writing simple SELECT statements for data extraction.
 - Using basic WHERE clauses for filtering (e.g., date ranges, string matching).
 - Applying grouping with GROUP BY and aggregations (SUM, COUNT).
 - Filtering and basic selection of columns.
- **Intermediate SQL:**
 - Performing multiple JOIN operations to combine related tables (e.g., INNER JOIN with stock_quant, product_product, etc.).
 - Using GROUP BY on multiple columns and integrating aggregations.
 - Joining and grouping across product and category data.
- **Advanced SQL:**
 - Constructing complex queries with multiple JOINs across different hierarchies and filtering conditions.
 - Calculating derived fields (e.g., amount_of_invoiced as a computed value).
 - Handling advanced grouping and conditions (e.g., combining OR and AND operators with careful precedence).

SQL Skill Set

Query 1

```
SELECT
    pc.category_name AS category_name,
    pt.product_name AS product_name,
    SUM(sq.quantity) AS total_quantity
FROM
    stock_data sq
JOIN
    product_details pt ON sq.template_id = pt.template_id
JOIN
    product_categories pc ON pt.category_id = pc.category_id
WHERE
    pc.category_name IN ('Category A', 'Category B', 'Category C', 'Category D')
    AND sq.quantity >= 0
GROUP BY
    pc.category_name,
    pt.product_name;
```

Result 1

category_name	product_name	total_quantity
Category A	Product Alpha	50
Category A	Product Beta	30
Category B	Product Gamma	20

Query 2

```
SELECT
    rp.customer_name AS customer_name,
    rk.region_name AS region_name,
    rka.district_name AS district_name,
    rcs.state_name AS state_name,
    ai.invoice_date AS invoice_date,
    ai.invoice_number AS invoice_number,
    pc.category_name AS product_category,
    pt.product_name AS product_name,
    ail.quantity,
    (ail.quantity * ail.unit_price) AS total_invoiced,
    ai.source_reference AS source_reference
FROM invoice_data ai
JOIN customer_data rp ON ai.customer_id = rp.customer_id
JOIN regions rk ON ai.region_id = rk.region_id
JOIN districts rka ON ai.district_id = rka.district_id
JOIN states rcs ON ai.state_id = rcs.state_id
JOIN invoice_line_data ail ON ai.invoice_id = ail.invoice_id
JOIN stock_data sq ON ail.product_id = sq.product_id
JOIN product_details pt ON sq.template_id = pt.template_id
JOIN product_categories pc ON pt.category_id = pc.category_id
WHERE rp.company_id = 202 OR ai.company_id = 202
    AND ai.invoice_date BETWEEN '2024-01-01' AND '2024-12-31'
GROUP BY
    rp.customer_name,
    rk.region_name,
    rka.district_name,
    rcs.state_name,
    ai.invoice_date,
    ai.invoice_number,
    pc.category_name,
    pt.product_name,
    ail.quantity,
    ail.unit_price,
    ai.source_reference;
```

Result 2

customer_name	region_name	district_name	state_name	invoice_date	invoice_number	product_category	product_name	quantity	total_invoiced	source_reference
Gautama	Region W	District V	State U	2024-05-15	INV002	Category A	Product Beta	5	1000.00	Order002
Jeevallucas	Region X	District Y	State Z	2024-02-10	INV001	Category A	Product Alpha	10	1500.00	Order001

Queried product categories (Category 1, Category 2, Category 3, Category 4) to calculate total quantities, extracted customer details (e.g., name, address, phone) for Company Client 1 in 2023-2024, and retrieved 2024 invoice data for Company Client 2 to analyze sales performance using SQL.

Python Skill Set

- **Basic:**
 - Writing scripts to read, clean, and process data files (e.g., CSV, Excel).
 - Using Python's built-in functions for basic data handling and exploration.
- **Intermediate:**
 - Leveraging libraries like Pandas for data manipulation (e.g., filtering, grouping, aggregation).
 - Summarizing data statistics (.describe(), .info()).
- **Advanced:**
 - Handling complex issues such as newline characters in CSV fields, inconsistent formats, and missing data.
 - Automating repetitive cleaning tasks for large datasets, as demonstrated in the script provided.

Python Skill Set

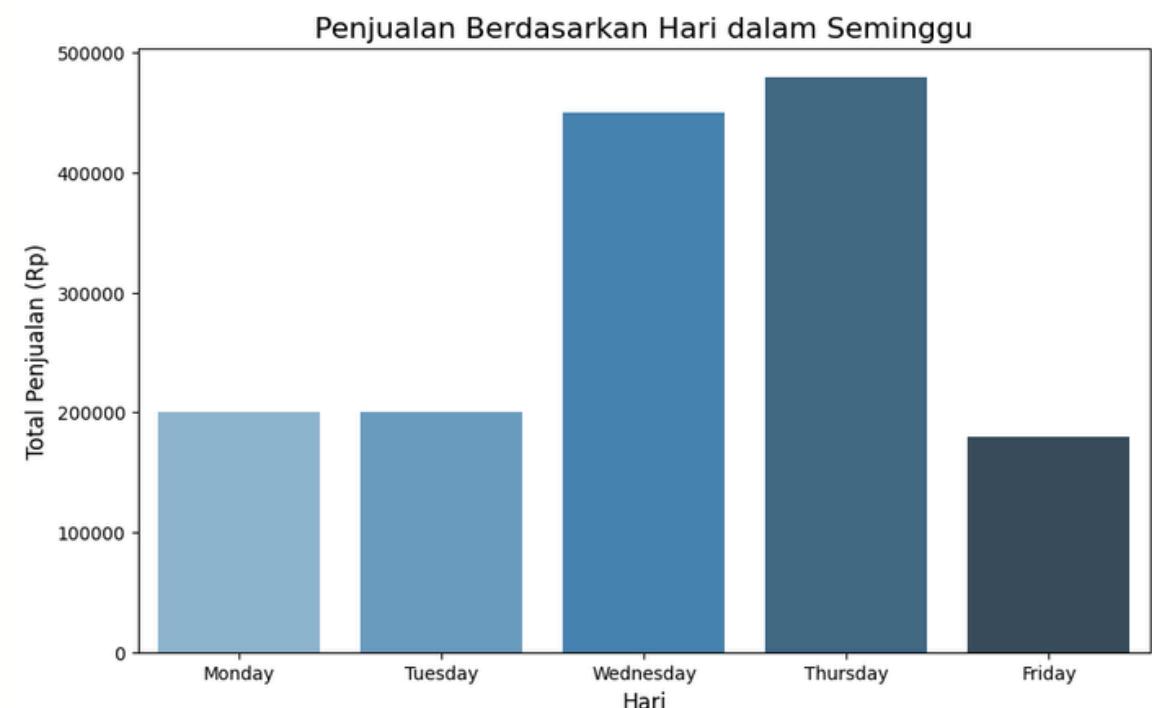
Python Code

```
[14] # Total penjualan per hari
daily_sales = order_2024.groupby('Hari')['Total Harga'].sum().reset_index()
order_2024['Hari'] = pd.Categorical(order_2024['Hari'],
                                      categories=['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'],
                                      ordered=True)

plt.figure(figsize=(10, 6))
sns.barplot(data=daily_sales, x='Hari', y='Total Harga', palette='Blues_d')
plt.title("Penjualan Berdasarkan Hari dalam Seminggu", fontsize=16)
plt.xlabel("Hari", fontsize=12)
plt.ylabel("Total Penjualan (Rp)", fontsize=12)
plt.show()
```

Calculated total daily sales for 2024 by grouping transactions by weekday and summing the sales amounts. Days were ordered from Monday to Sunday, and results were visualized in a bar chart to identify sales trends across the week.

Python Result



Recommendations:

Focus on increasing sales on days with **low performance**, such as **Monday, Tuesday, and Friday**, through special promotions, discounts, or loyalty programs to attract customers, while maximizing the potential of **high-sales** days like **Wednesday and Thursday** with strategic campaigns.

Power BI Skill Set

- **Basic:**

- Creating simple visualizations (e.g., bar charts, line charts, pie charts).
- Importing and transforming data using Power BI's interface.
- Building basic dashboards with interactive filters and slicers.

- **Intermediate:**

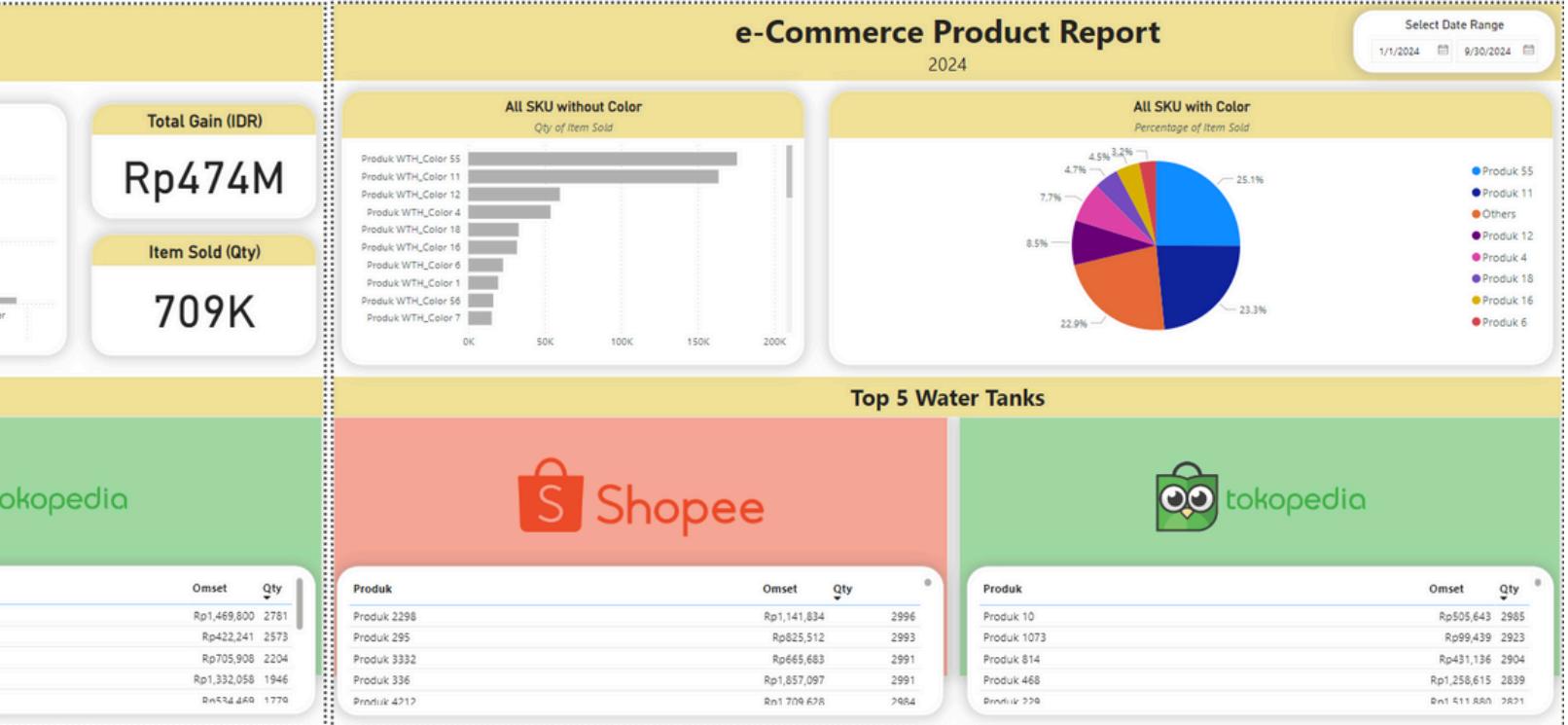
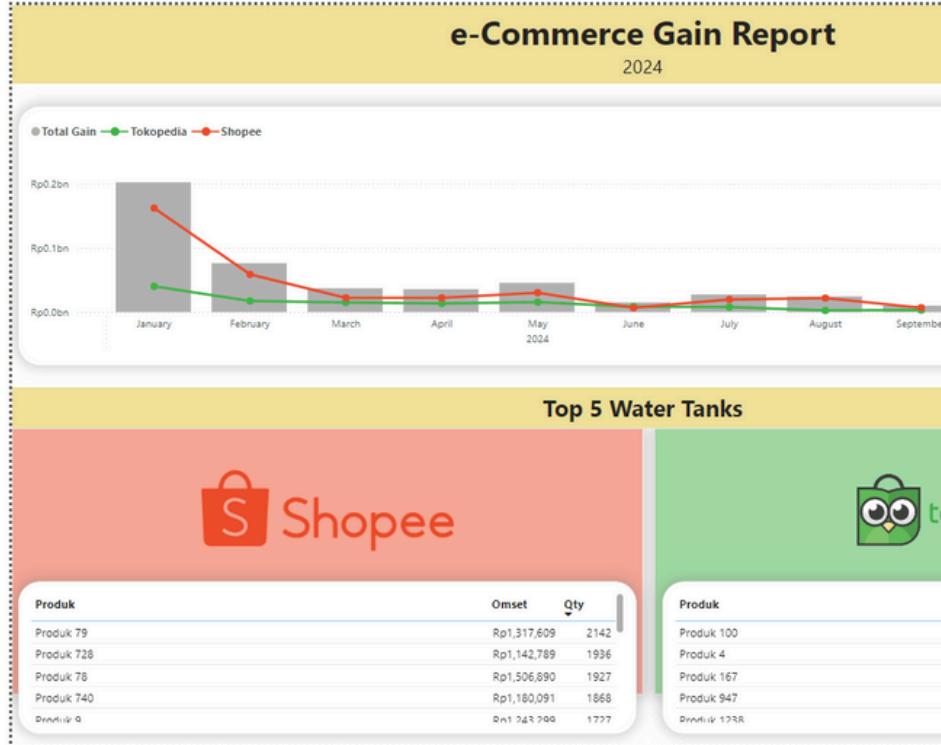
- Writing DAX (Data Analysis Expressions) formulas for calculated columns and measures.
- Creating relationships between tables for seamless data modeling.
- Designing dashboards with multiple visualizations to highlight key insights.

- **Advanced:**

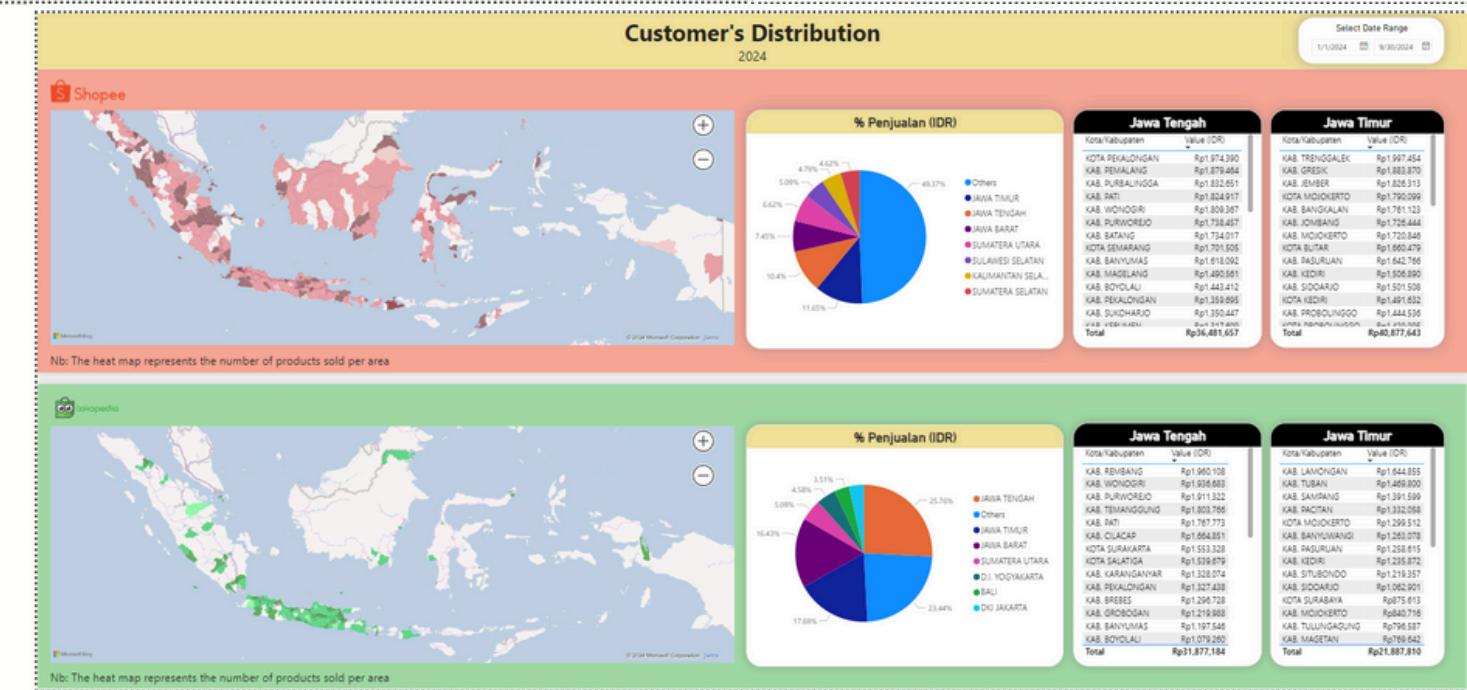
- Developing complex DAX measures for advanced calculations (e.g., year-over-year growth, rolling averages).
- Building dynamic dashboards with drill-through capabilities and tooltips.
- Optimizing data models for performance and scalability.
- Maps and geographic data analysis

Power BI Skill Set

Page 1



Page 2



Page 3

At Pennyu, I was entrusted with the task of transitioning the company's dashboards from Google Looker Studio to Microsoft Power BI, marking the first time Power BI was implemented in the organization. While this dashboard is a direct functional conversion of the existing Looker Studio dashboard, the visuals and user interface have been significantly enhanced and developed further in Power BI to ensure data accuracy, maintain original insights, and provide a more polished and effective presentation. This project reflects my role in pioneering Pennyu's adoption of Power BI during my days as an intern.

PROJECT GIVEN TO ME

Background & Project Goals

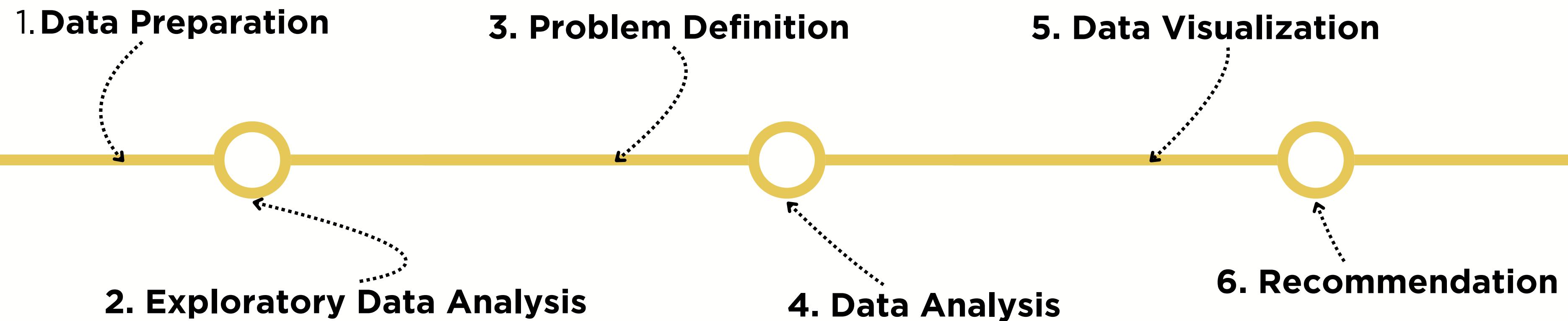
- Pennyu Group, established in 1991, is one of Indonesia's prominent manufacturing companies, serving over **4,000 retailers** and **40 distributors** nationwide.
- The company specializes in **essential products** such as water tanks, bio-septic systems, and plumbing fixtures. It operates across **multiple sectors**, including construction, biotech, and energy.
- Pennyu has three production centers **certified by SNI standards** and **strategically located** across Java, supporting its extensive distribution network.
- With **growing operational data** from platforms like **Shopee**, Pennyu aims to optimize **customer retention** and **boost sales** through data analytics.



OBJECTIVE

Ultimately, the main goal is to **drive sales growth** and **customer satisfaction** for Pennyu's Shopee operations. However, as a Data Analyst Intern, my objective is to **provide actionable recommendations** to key stakeholders on how to **optimize marketing strategies** and **improve customer retention** based on data-driven insights and patterns derived from the analysis.

Methodology

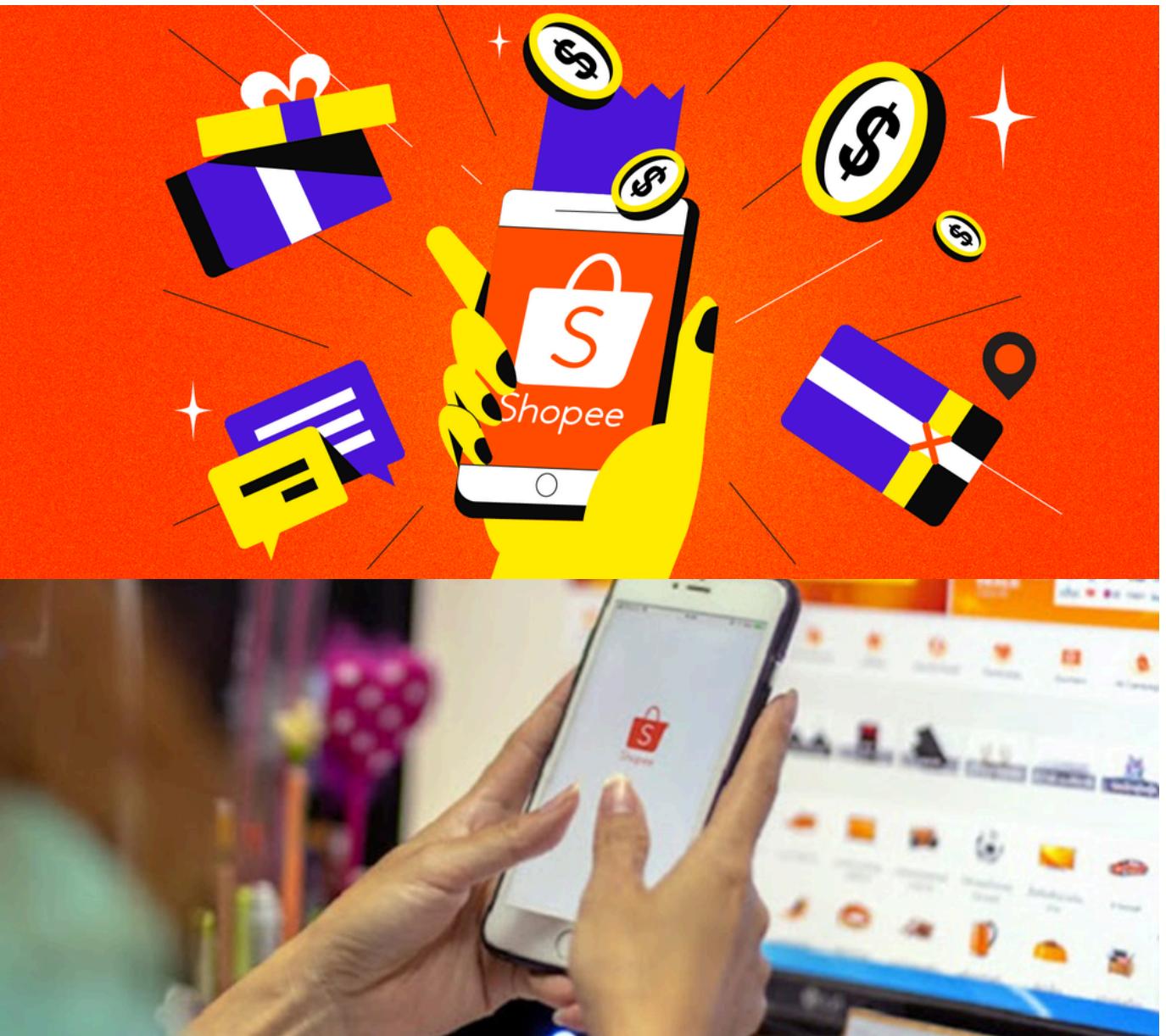


Data Source & Data Overview

The dataset provides a comprehensive overview of the **operational activities** of Pennyu Group's Shopee official store, covering the period from **January to September 2024**. It comprises of **89 columns** and **14,000+ rows**.

This dataset includes detailed customer transaction records, covering order IDs, dates, product quantities, total prices, shipping costs, and other relevant metrics. It incorporates advertising data, such as impressions, clicks, conversions, and spending details.

The data reflects customer purchasing behavior, segmented trends, and the effectiveness of targeted marketing campaigns, enabling insights into seasonal demand, ad performance, and customer retention strategies during this period.

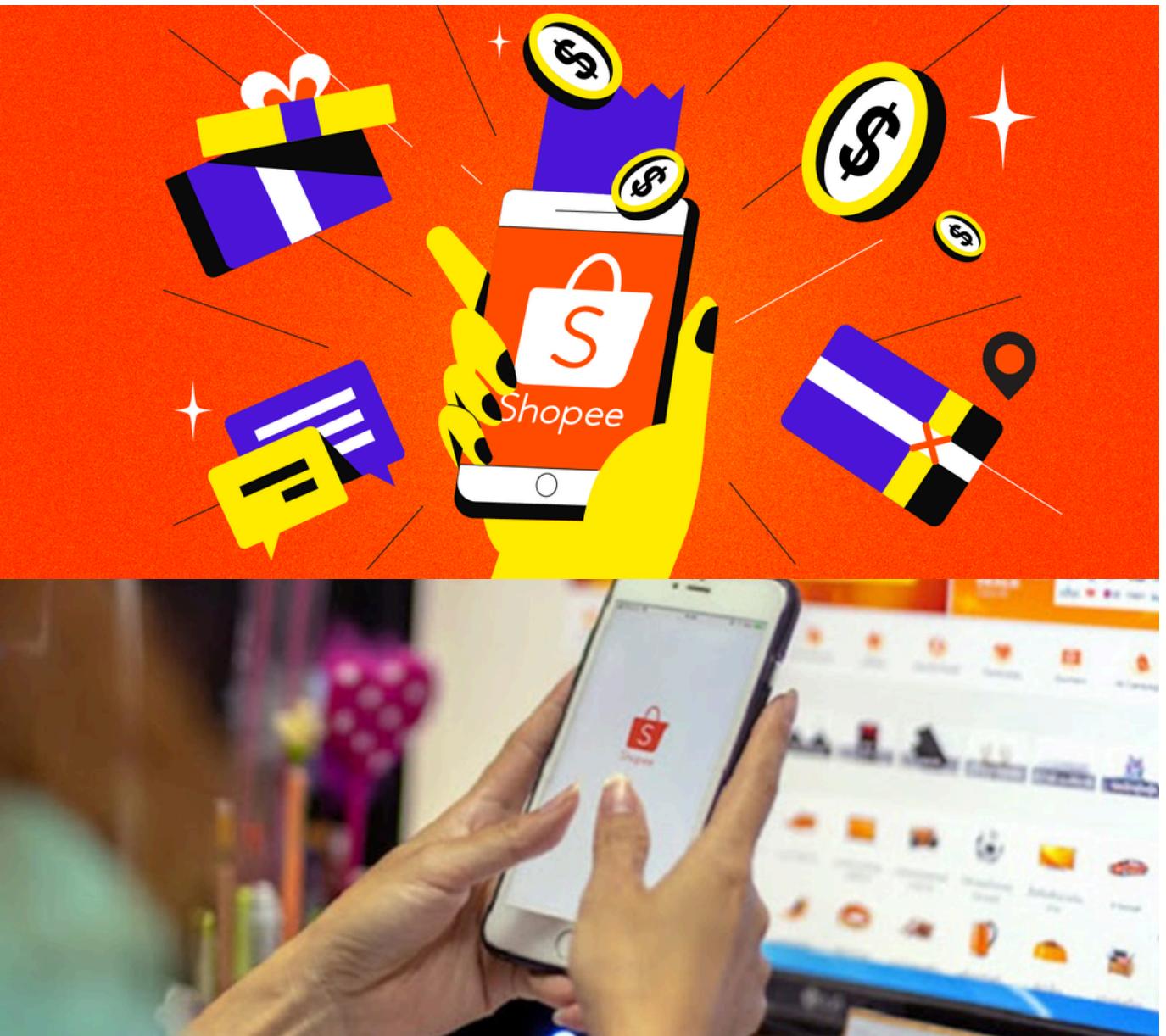


Data Source & Data Overview

The dataset provides a comprehensive overview of the **operational activities** of Pennyu Group's Shopee official store, covering the period from **January to September 2024**. It comprises of **89 columns** and **14,000+ rows**.

This dataset includes detailed customer transaction records, covering order IDs, dates, product quantities, total prices, shipping costs, and other relevant metrics. It incorporates advertising data, such as impressions, clicks, conversions, and spending details.

The data reflects customer purchasing behavior, segmented trends, and the effectiveness of targeted marketing campaigns, enabling insights into seasonal demand, ad performance, and customer retention strategies during this period.



...

Thank you!



Feel free to reach me out!



wa.me/6281393405010



linkedin.com/in/jeevallucas-gautama



jnanamaitriya@gmail.com