

APRIL

2025

TECHHUB CUSTOMER SEGMENTATION

Sales & Customer Insight
Using Python, SQL and Power BI

Prepared by:

Cao Thuc Ta (Erik)

Applying data transformation, RFM segmentation, and visual analytics to uncover customer behavior and optimize business strategy for TechHub, a growing electronics retailer.



<https://github.com/TaCaoThuc>

Table of Contents

| | | |
|------|----------------------------------------------------------------|----|
| I. | Introduction and Project Background..... | 1 |
| II. | Overview of Key Business Frameworks and Techniques..... | 2 |
| 1. | Customer 360..... | 2 |
| 2. | RFM analysis..... | 3 |
| 3. | BCG Matrix..... | 5 |
| 4. | Combing BCG Matrix with RFM Segmentation..... | 6 |
| III. | Project Implementation..... | 8 |
| 1. | Database design and ETL..... | 8 |
| 2. | Sales Overview and Customer Segmentation..... | 8 |
| 3. | Visualization..... | 19 |
| IV. | Conclusion & Future Work..... | 20 |

I. Introduction and Project Background

In today's customer-centric retail environment, understanding purchasing behavior is essential for building loyalty, increasing sales, and staying ahead of competitors. TechHub, a growing electronics retailer, recognized the need to move beyond basic sales reporting and toward a more strategic, data-driven approach — particularly in how it understands and engages its customers.

Scenario:

TechHub had accumulated thousands of sales records, but they existed in the form of a single flat spreadsheet, lacking structure and offering limited analytical value. The management team wanted to answer key questions such as:

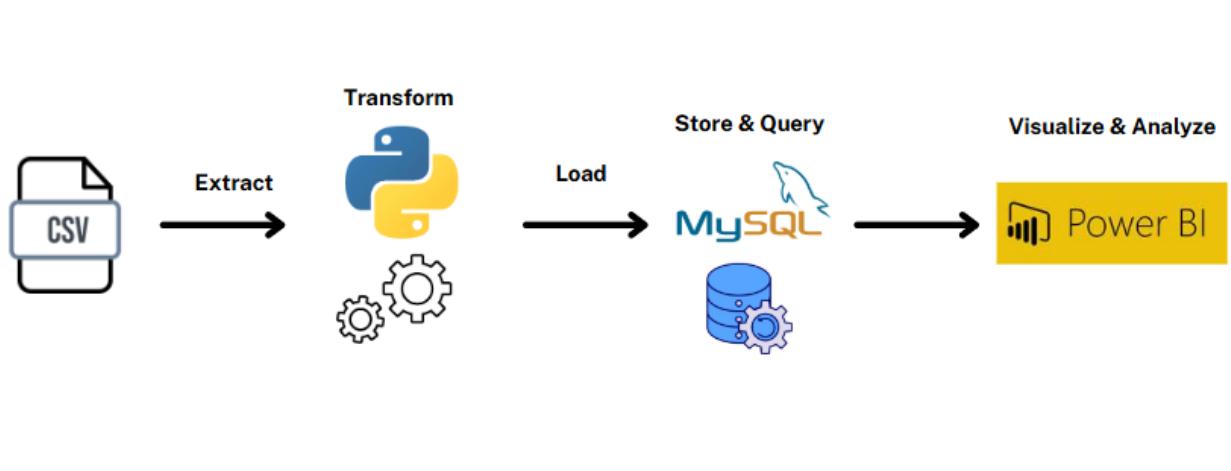
- Who are our most valuable customers?
- Which customers are worth re-engaging?
- How can we personalize marketing and improve retention?

With that in mind, this project was launched with a primary focus on customer segmentation using RFM analysis, supported by complementary sales performance insights.

To support more strategic decision-making, TechHub initiated a data analytics project with the goal of turning raw transactional data into actionable insights. The project was designed to:

- Build a centralized, normalized database structure to store and analyze customer, product, and sales data.
- Uncover sales performance trends across time, geography, and product categories.
- Apply RFM segmentation to categorize customers by their value and engagement levels.
- Provide actionable business recommendations based on insights gained from the data.
- Develop an interactive Power BI dashboard that can be used by business leaders to support marketing, operations, and customer relationship strategies.

The dataset was cleaned, enriched, and normalized using Python, split into seven structured tables, and loaded into a MySQL database. From there, SQL queries were used to perform both sales analysis and customer segmentation, and the results were visualized in Power BI to support clear communication and data-driven decision-making.



Technical workflow of the project

This project not only helps TechHub identify and understand its customer base but also establishes a foundational data infrastructure for more personalized, targeted, and profitable marketing strategies in the future.

II. Overview of Key Business Framework and Techniques used

1. Customer 360

Customer 360 is a strategic approach that integrates customer data from multiple sources — such as purchase history, online platforms, support interactions, and demographic details — to create a comprehensive, real-time view of each customer. This holistic perspective enables businesses to analyze customer behavior, preferences, and value, ultimately allowing for:

- Personalized marketing campaigns
- Targeted product recommendations
- Enhanced customer support and engagement

By unifying demographic, behavioral, transactional, and interactional data, Customer 360 empowers more informed decision-making and helps drive customer-centric growth.

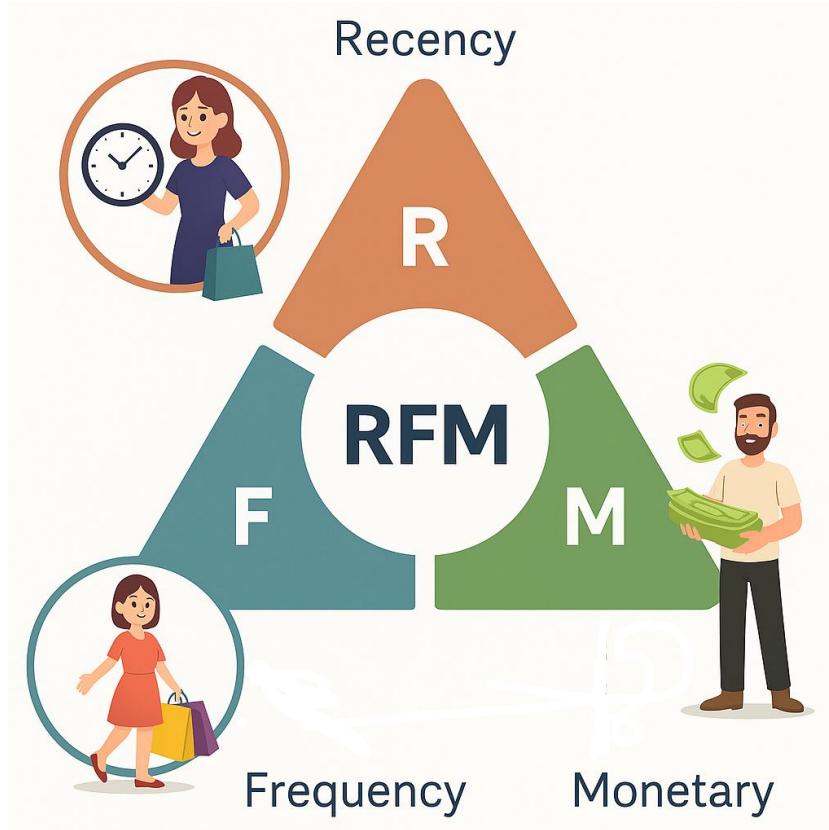


Customer 360 concept

2. RFM analysis

RFM (Recency, Frequency, Monetary) analysis is a key method within the Customer 360 framework, used to segment customers based on their purchasing behavior.

- **Recency (R):** How recently a customer made a purchase
- **Frequency (F):** How often a customer makes purchases
- **Monetary (M):** How much a customer spends in total

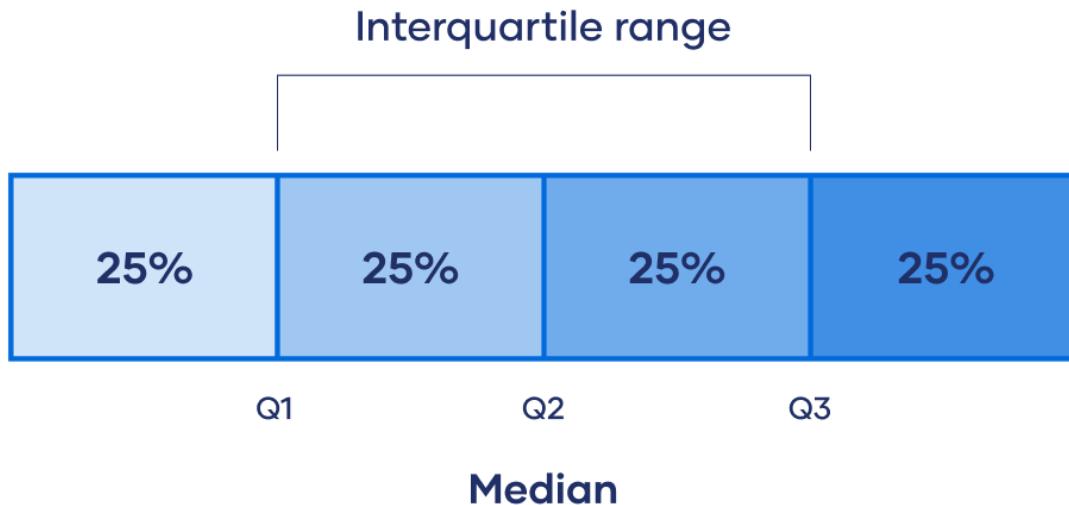


Visual representation of the RFM model

Implementation:

Customer transaction data is first collected, and each customer is assigned R, F, and M values based on their behavior. The next step involves segmenting the data using **interquartile range (IQR)**:

- The dataset is divided into quartiles (Q1, Q2, Q3), each representing 25% of customers.
- Based on which quartile a customer falls into for each metric, they receive a score from 1 (low) to 4 (high).
- For example, a customer in the top 25% of frequency and monetary value would receive F=4 and M=4.



Interquartile range

The three scores are concatenated into a single RFM score (e.g., R=2, F=4, M=3 → RFM Score = 243). These scores are then used to assign customers into actionable segments — such as high-value, at-risk, or one-time buyers — for more targeted business strategies.

3. BCG Matrix

The BCG (Boston Consulting Group) Matrix is a strategic tool used by companies to evaluate their product portfolio and make investment decisions. It helps identify which products or business units to grow, maintain, or divest based on market growth rate and relative market share.



The BCG matrix

The matrix is divided into **four quadrants**:

| BCG Quadrant | Description |
|------------------|-------------------------------------------------------------------------------------------------|
| ⭐ Stars | High market growth, high market share. These are leaders and need investment to sustain growth. |
| 🐮 Cash Cows | Low market growth, high market share. These generate steady cash and fund other areas. |
| ❓ Question Marks | High market growth, low market share. Risky – need heavy investment to grow or may fail. |
| 🐶 Dogs | Low market growth, low market share. Typically candidates for divestment. |

4. Combing BCG Matrix with RFM Segmentation

Although traditionally used for analyzing products or business units, the BCG Matrix can also be applied to customer segmentation by aligning it with RFM scores. This approach enables businesses to evaluate customers not just by their current value, but also by their future potential and growth behavior.

| BCG Quadrant | Customer Type | RFM Behavior | Example Micro-Segment |
|------------------|-------------------------|--------------------------|-----------------------------------------|
| ⭐ Stars | High value, high growth | High R, High F, High M | VIP, Champion, Power Shopper |
| 🐄 Cash Cows | High value, low growth | Low R, High F, High M | Loyal, Dormant Big Spender |
| ❓ Question Marks | Low value, high growth | High R, Low F, Low/Med M | Potential, Newcomer, Curious Visitor |
| 🐶 Dogs | Low value, low growth | Low R, Low F, Low M | One-Time Buyer, At Risk, Bargain Hunter |

Based on the combination of RFM scoring and BCG Matrix principles, TechHub's customer base has been segmented as follows:

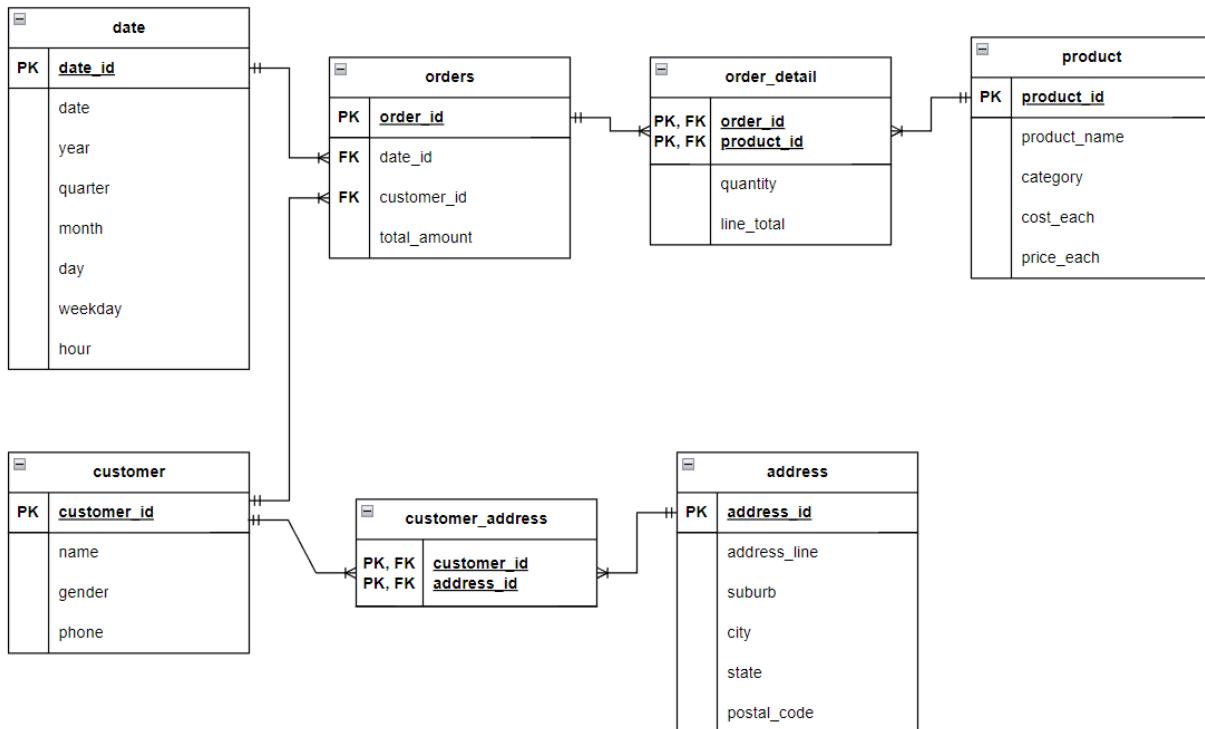
| BCG Quadrant | Segment Name | RFM Criteria | Typical Customer Behavior |
|----------------|-------------------------------------------|-------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Stars | VIP Customers (Tech Enthusiasts) | R ≥ 3, F ≥ 3, M ≥ 3 | Frequently buys high-end electronics, upgrades devices often (e.g., new phone/laptop every year). |
| Cash cows | Loyal Customers (Regular Buyers) | R ≤ 2, F ≥ 3, M ≥ 3 | Loyal customers who buy consistently but not as frequently as VIPs (e.g., buys a laptop one year, accessories the next). |
| | Occasional Spender | R < 3, F < 3, M ≥ 3 | Customers who buy infrequently but spend on high-value items (e.g., bought a high-end laptop but won't return for a long time). |
| Question marks | Potential Big Spenders | R ≥ 3, (F < 3 OR M < 3) | New customers who recently bought but haven't spent much yet (e.g., bought a phone but no accessories). |
| Dogs | One-Time Buyers (Transactional Customers) | R < 3, F < 3, M < 3 | Bought only once and have not returned (e.g., purchased a charger or a low-cost accessory). |

This hybrid framework enables TechHub to develop targeted engagement strategies for each segment — such as retention plans for VIPs, nurturing programs for first-time buyers, and reactivation campaigns for one-time customers — ensuring resources are allocated effectively to maximize customer lifetime value.

III. Project implementation

1. Database design and ETL

Based on the original dataset, a relational schema with seven tables was designed to represent TechHub's data on orders, products, dates, customers, and addresses. This structure ensures efficient organization, normalization, and accessibility, allowing for clear relationships between entities and supporting accurate data analysis and reporting.



The designed data schema

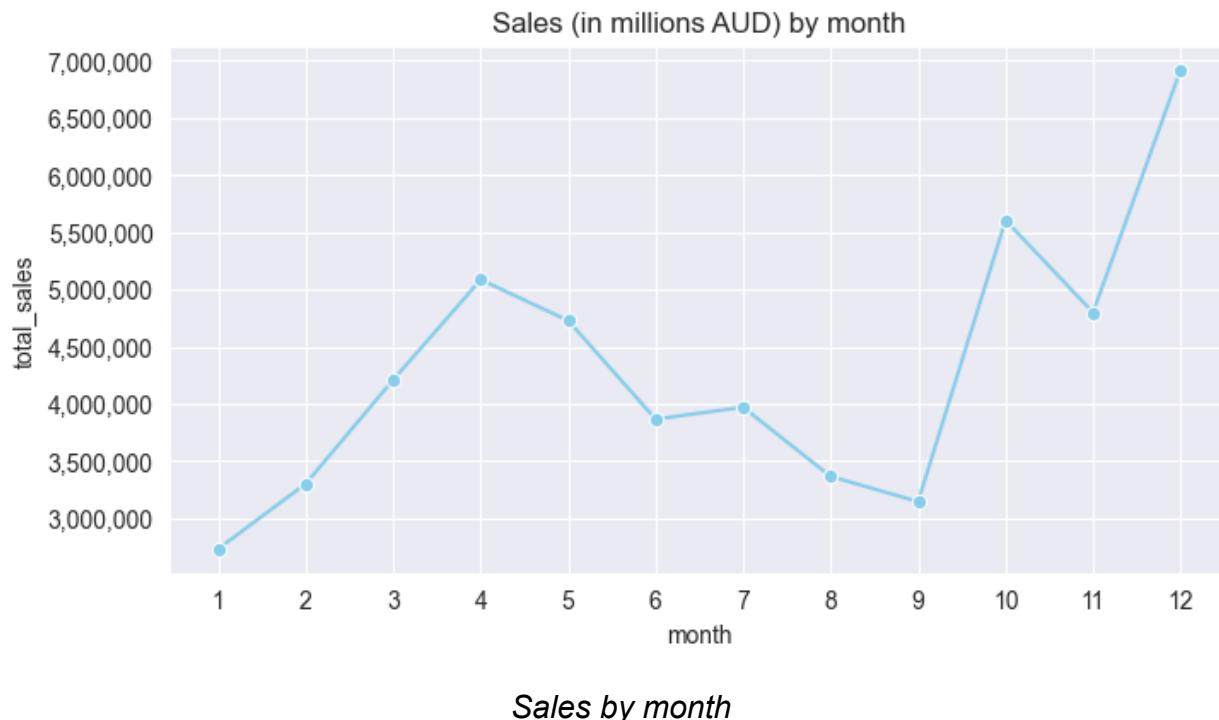
The data was then cleaned, transformed, and split into the respective tables using Python, and subsequently loaded into a MySQL database, preparing it for further analysis and visualization tasks.

[Link to the ETL script](#)

2. Sales Overview and Customer Segmentation

2.1. Sales Overview

a. Sales by Month and Quarter



Insights:

- Revenue is lowest in January, then steadily increases, peaking in December, likely due to holiday shopping.
- At the quarterly level, Q4 generates the highest revenue, while Q1 and Q3 underperform.

Recommendations:

- Prepare holiday stock well in advance, ensuring sufficient inventory for Q4, especially December.
- Extend promotional campaigns into January (e.g., New Year discounts) to maintain post-holiday momentum and address January's low sales.

b. Sales by Weekday



Insights:

- Sunday has the highest sales, likely due to increased leisure time and online activity.
- Surprisingly, Monday also performs well, while Tuesday has the lowest sales volume.

Recommendations:

- Investigate why Tuesday sales are low — this could be due to marketing gaps or customer behavior. Consider running exclusive Tuesday promotions.
- Capitalize on Sunday and Monday with extended weekend campaigns and targeted ad schedules.

c. Sales by Hour



Insights:

- Peak sales hours occur during lunch (11 AM – 1 PM) and evening (7 PM – 8 PM).
- 3 AM to 5 AM records the lowest activity.

Recommendations:

- Schedule ads and promotions around peak hours when customers are most active.
- Introduce late-night deals or flash sales to boost conversions during off-hours, especially on weekends.

d. Sales by Product

| | product_name | total_quantity | total_revenue |
|---|------------------------|----------------|---------------|
| 0 | Macbook Pro Laptop | 4728.0 | 12056400.00 |
| 1 | iPhone | 6849.0 | 7191450.00 |
| 2 | ThinkPad Laptop | 4130.0 | 6194958.64 |
| 3 | Google Phone | 5532.0 | 4978800.00 |
| 4 | 27in 4K Gaming Monitor | 6244.0 | 3652677.36 |

Top 5 product by Revenue

| product_name | total_quantity | total_revenue |
|--------------------------|----------------|---------------|
| AAA Batteries (4-pack) | 31017.0 | 139186.46 |
| AA Batteries (4-pack) | 27635.0 | 159177.61 |
| Wired Headphones | 20557.0 | 369630.19 |
| USB-C Charging Cable | 23975.0 | 429650.99 |
| Lightning Charging Cable | 23217.0 | 520540.66 |

Bottom 5 product by revenue

Insights:

- Laptops are the top-performing category, generating \$18.3 million, with two laptop models appearing in the top 5 best-sellers.
- Home Appliances and Accessories are the worst-performing categories, each contributing only \$1.2 million in revenue.
- MacBook Pro alone generates \$12M, despite relatively lower volume — due to high unit price.
- iPhones and ThinkPads show strong customer demand.
- Google Phones perform moderately well but lag behind iPhones.
- Accessories sell in high volume but generate low revenue, indicating low margins.

Recommendations:

- Focus on high-margin, high-ticket products like MacBooks and iPhones. Introduce financing options and accessory bundles (e.g., "Buy a MacBook, get 20% off AirPods").
- Improve profitability on low-margin accessories by slightly increasing prices or bundling them with premium products.
- Introduce premium accessory lines, such as fast-charging cables, to raise margins without reducing volume.

e. Sales by Category

| category | total_quantity | total_revenue | total_profit |
|-----------------|----------------|---------------|--------------|
| Laptops | 8858.0 | 18251358.64 | 8939447.04 |
| Smartphones | 14449.0 | 13411050.00 | 6568732.19 |
| TV & Monitors | 28941.0 | 11734471.92 | 5747518.64 |
| Headphones | 49675.0 | 5911637.32 | 2895311.61 |
| Accessories | 105844.0 | 1248555.71 | 611306.17 |
| Home Appliances | 1312.0 | 1180800.00 | 578355.85 |

Sales by category

Insights:

- Laptops and Smartphones account for over 60% of total revenue and profit.
- Accessories and Home Appliances have high volume but low profitability.
- Monitors and TVs perform well, indicating interest in work-from-home and gaming setups.

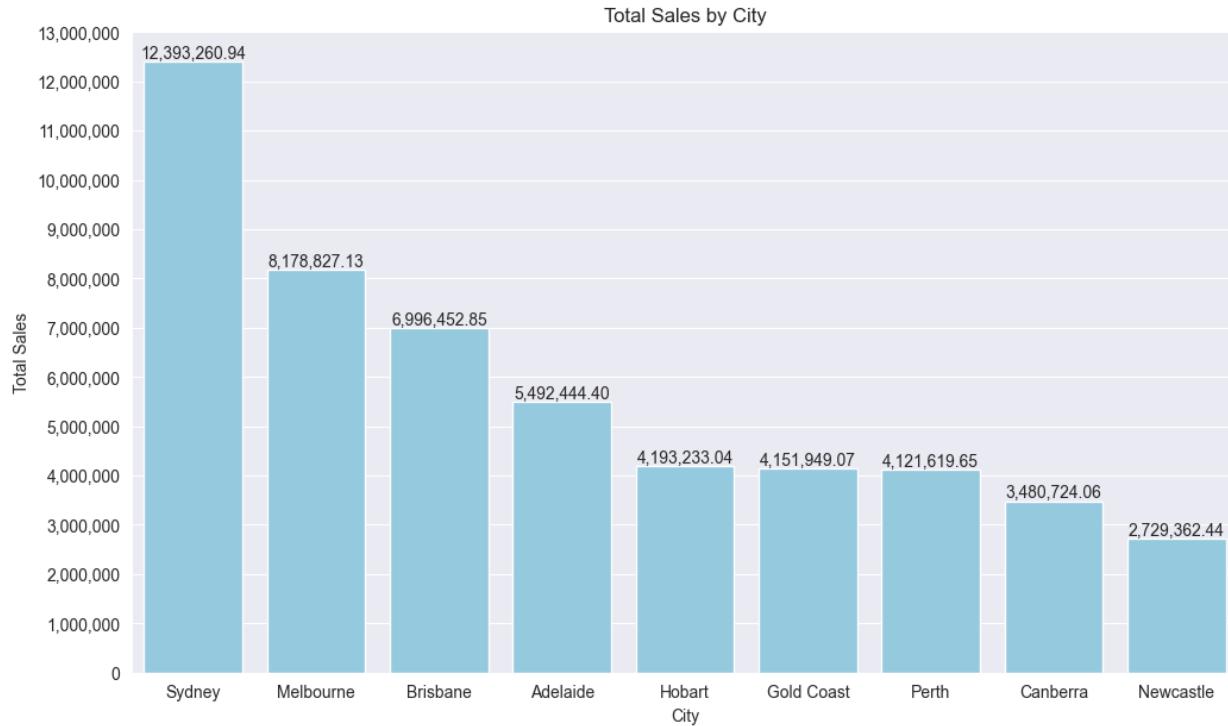
Recommendations:

- Upsell accessories with laptops and smartphones via bundle offers (e.g., "Buy a phone, get 50% off case and charger").
- Expand the monitor and TV product line to meet demand from remote workers and gamers.
- Reevaluate the Home Appliances category — consider focused promotions or phase-out of underperforming items.

f. Sales by State and City

| | city | total_sales |
|---|------------|-------------|
| 0 | Sydney | 12393260.94 |
| 1 | Melbourne | 8178827.13 |
| 2 | Brisbane | 6996452.85 |
| 3 | Adelaide | 5492444.40 |
| 4 | Hobart | 4193233.04 |
| 5 | Gold Coast | 4151949.07 |
| 6 | Perth | 4121619.65 |
| 7 | Canberra | 3480724.06 |
| 8 | Newcastle | 2729362.44 |

Sales by city



Bar chart of sales by city

City-Level Insights:

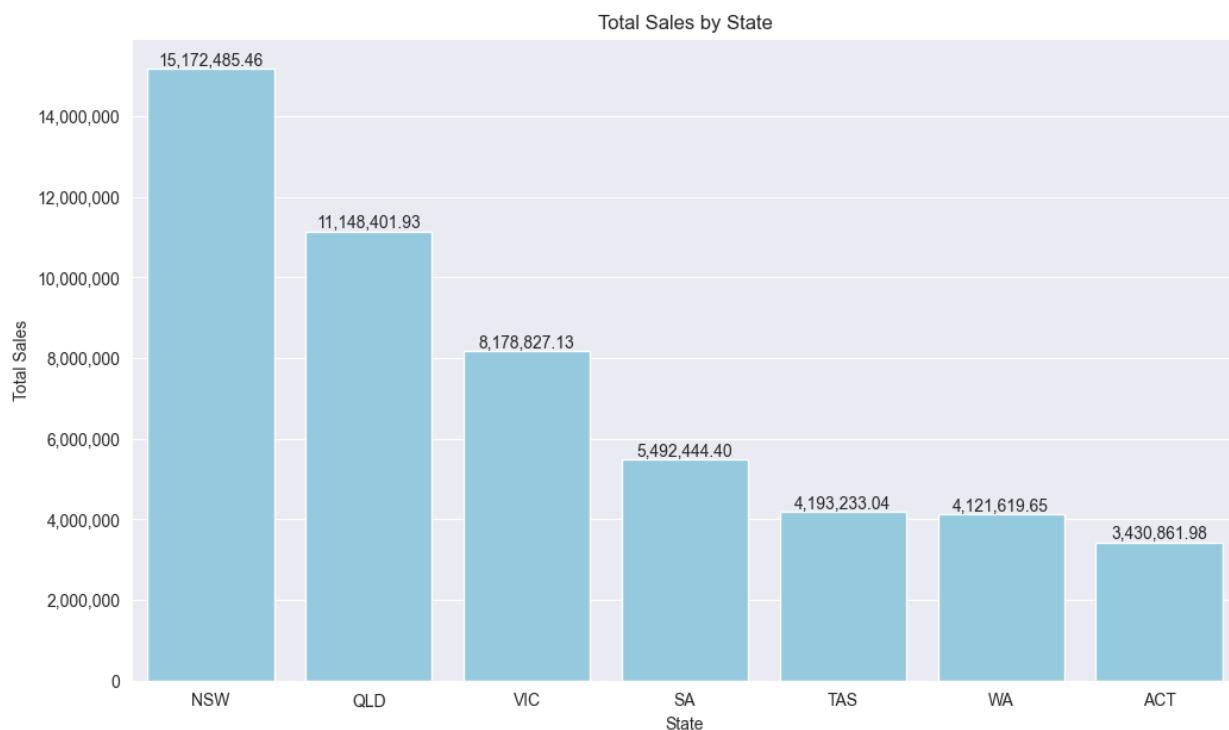
- Sydney is the highest-grossing city, contributing \$12.4 million (24% of total revenue) — making it TechHub's strongest market.
- Melbourne and Brisbane show high potential, though significantly behind Sydney.
- Smaller cities like Newcastle and Canberra underperform, possibly due to lower brand visibility, weaker outreach, or market demand.

City-Level Recommendations:

- Analyze why Sydney outperforms — is it better distribution, brand presence, or customer loyalty? Apply learnings to other key cities.
- Leverage Sydney's dominance by offering VIP programs or loyalty incentives to maintain engagement.
- In underperforming cities, run localized ads and promotions. Consider partnering with local retailers to improve physical and digital presence.

| | state | total_sales |
|---|-------|-------------|
| 0 | NSW | 15172485.46 |
| 1 | QLD | 11148401.93 |
| 2 | VIC | 8178827.13 |
| 3 | SA | 5492444.40 |
| 4 | TAS | 4193233.04 |
| 5 | WA | 4121619.65 |
| 6 | ACT | 3430861.98 |

Sales by state



Bar chart of sales by state

State-Level Insights:

- NSW contributes over 30% of total sales, showing a strong market hold.
- ACT and WA have the lowest sales, pointing to weak demand or ineffective marketing.

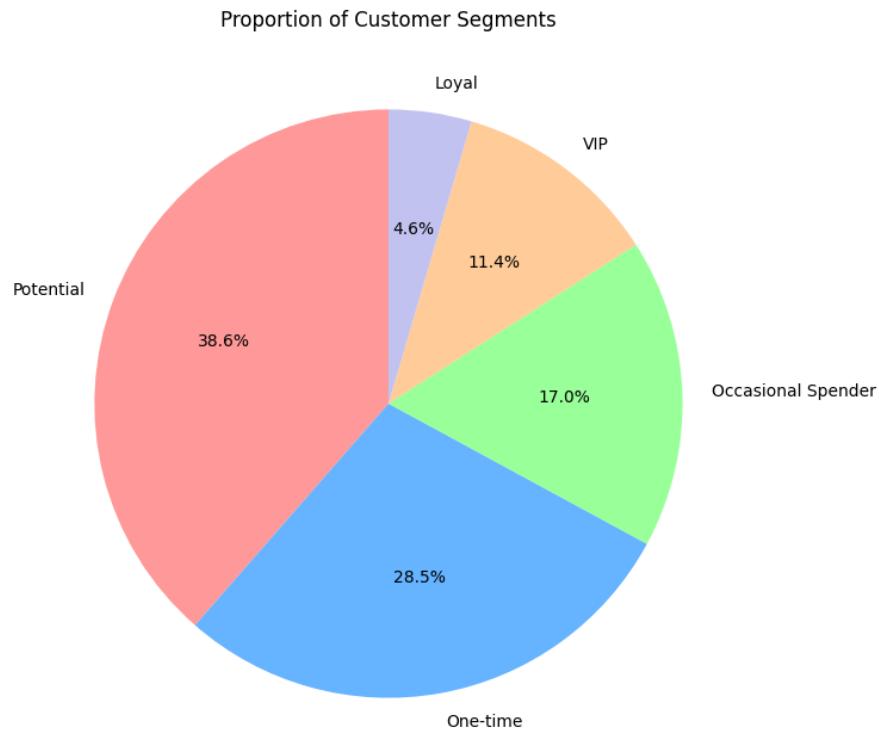
State-Level Recommendations:

- Launch targeted digital ads in WA and ACT using location-based promotions.

- Improve regional logistics — assess whether pricing, shipping time, or product availability affect conversion.
- Partner with local delivery and retail networks to expand TechHub's reach in low-performing states.

2.2. Customer Segmentation

Key insights



Proportion of customer segments

| Segment | Customer Proportion | Revenue Contribution | Observations |
|---------------------|---------------------|----------------------|----------------------------------------------------------|
| VIP | 11.4% | \$14.05M | Small but highly profitable; top-priority group. |
| Loyal | 4.6% | \$5.4M | Reliable, though low in number; potential for nurturing. |
| Potential | 38.6% | \$14.05M | Largest group with untapped value; high opportunity. |
| Occasional Spenders | 17% | \$16.98M | High revenue despite low frequency; great potential. |

| | | | |
|------------------------|-------|---------|------------------------------------------------------------------|
| One-Time Buyers | 28.5% | \$1.24M | Largest group but lowest revenue; urgent need for re-engagement. |
|------------------------|-------|---------|------------------------------------------------------------------|

Business Strategy by Segment

a. VIP (Tech Enthusiasts)

These customers frequently purchase high-end electronics such as the latest phones, laptops, and accessories. They demonstrate strong brand loyalty and a willingness to pay premium prices. Despite making up just 11.4% of the customer base, they contribute a significant \$14.05 million (27%) of total revenue, highlighting the importance of giving them special attention.

Strategy:

- Launch an exclusive VIP program offering early access to product launches, priority service, and personalized deals.
- Upsell premium accessories (e.g., high-end headphones, monitors, peripherals) through curated bundles.
- Use personalized retargeting with AI-powered product recommendations based on their purchase history.

b. Loyal (Regular Buyers)

Though this is the smallest segment (4.6%), they generate a respectable \$5.4 million in revenue. Loyal customers shop consistently but less frequently than VIPs and may not always choose premium items. The goal is to maintain engagement and gradually increase their basket size.

Strategy:

- Implement a points-based loyalty rewards system to incentivize repeat purchases.
- Send seasonal reminders aligned with product lifecycles (e.g., “It’s been 2 years since your last laptop — time for an upgrade?”).
- Offer extended warranties and service plans to build long-term trust and attachment.

c. Potential (First-Time Buyers)

This is the largest segment, accounting for 38.6% of customers. Most have made one big-ticket purchase (e.g., phone or laptop) but haven’t returned. Their spending equals

that of VIPs, showing they have strong potential to become high-value customers if nurtured effectively.

Strategy:

- Send post-purchase follow-up emails with related product offers (e.g., “Bought a MacBook? Get 15% off an SSD or carrying case!”).
- Offer exclusive new-customer incentives, such as \$50 off the next order within 30 days.
- Use retargeting ads on social media and Google to promote accessories or upgrades relevant to their original purchase.
- Provide excellent post-sale support and satisfaction check-ins to build loyalty and trust.

d. Occasional Spenders

These customers buy infrequently but contribute the highest revenue (\$16.98M), despite being just 17% of the base. They typically purchase premium items and are less price-sensitive, but have longer buying cycles. Increasing their purchase frequency even slightly could generate major returns.

Strategy:

- Offer VIP-style perks for large purchases, such as free product servicing, extended warranties, or exclusive consultations.
- Introduce upgrade and trade-in programs to shorten buying cycles (e.g., “Trade your old laptop for \$500 credit toward the latest model”).
- Promote limited-time bundle deals (e.g., “Buy a laptop + monitor, get a premium keyboard free”) to increase cart value.

e. One-Time Buyers (Transactional Customers)

These are customers who purchased only once and never returned. They account for 28.5% of all customers but generate just \$1.24M — the lowest revenue share. Re-engagement is essential for converting them into repeat customers.

Strategy:

- Send post-purchase follow-ups with product guides, thank-you messages, and suggestions for related items.
- Offer incentives for a second purchase, such as free shipping or double loyalty points on the next order.
- Launch win-back campaigns after periods of inactivity (e.g., “We miss you! Here’s 15% off your next purchase.”)

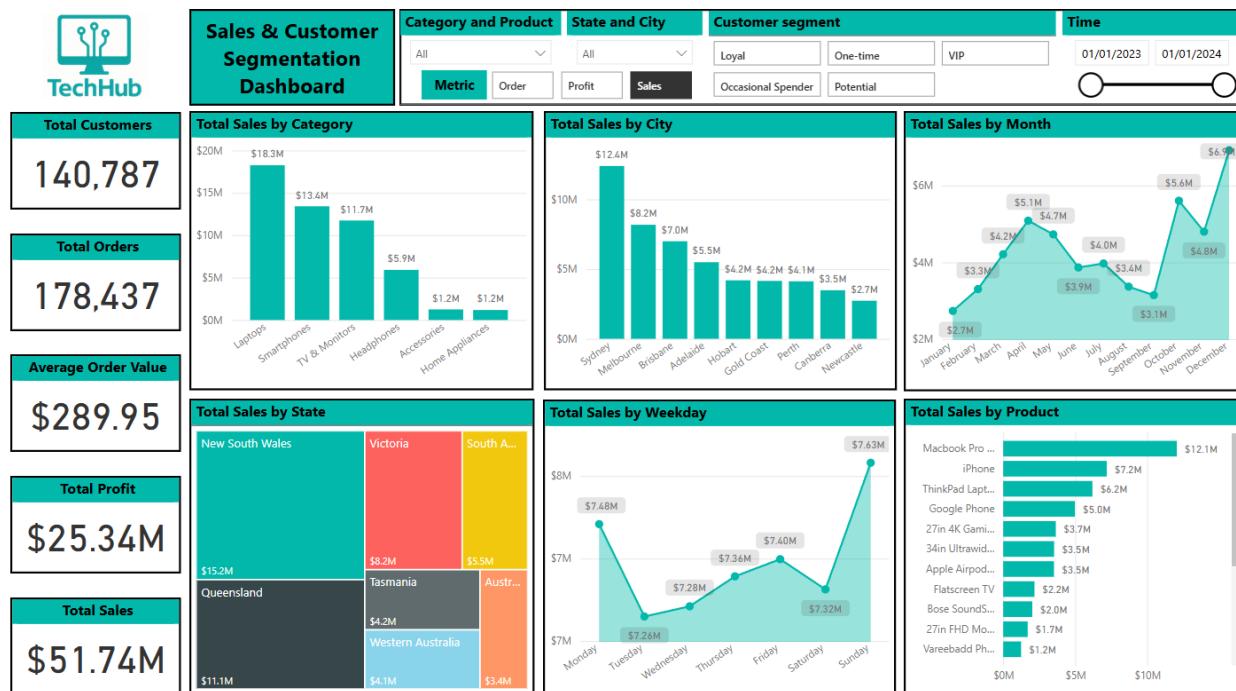
- Create a referral program, offering discounts or small gifts for successful referrals.
- If multiple re-engagement attempts fail, reduce direct marketing but continue sending light-touch updates (e.g., major sales, new releases).

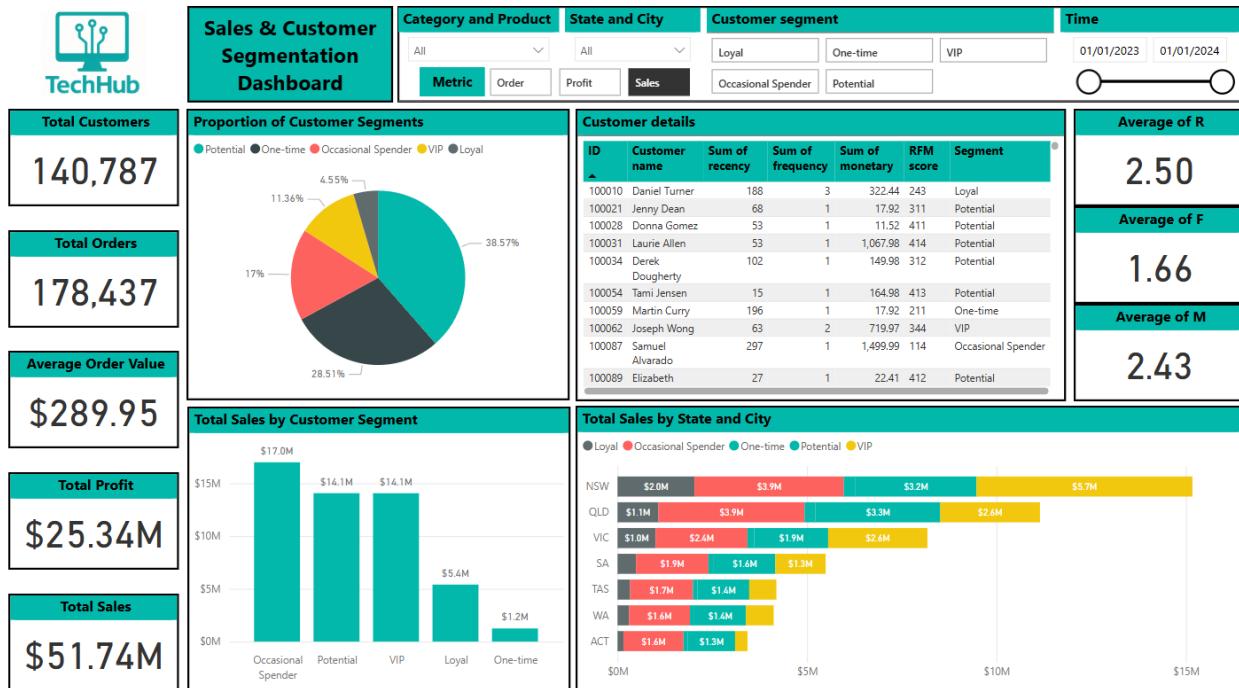
Link to the SQL queries

3. Visualization

Power BI was used to create the visualizations for this project. It was first connected to the MySQL database, enabling direct access to the cleaned and structured data. The resulting dashboard consists of two interactive pages: one presenting a comprehensive sales overview of TechHub, and the other showcasing the results of customer segmentation.

This dashboard aims to support the company in making data-driven decisions by clearly highlighting top-performing cities, products, and time periods. It also identifies key customer segments such as VIPs and Potential buyers, allowing the company to personalize marketing efforts, improve customer retention, and drive revenue growth through targeted strategies.





[Link to the dashboard](#)

IV. Conclusion and Future Work

This project successfully transformed a single raw dataset into a structured, relational database, enabling meaningful insights through both sales analysis and customer segmentation. Using the RFM model alongside the BCG Matrix, TechHub's customers were segmented into five distinct groups, each with tailored business strategies aimed at improving retention, engagement, and long-term profitability. A Power BI dashboard was developed to visualize sales trends, geographic performance, and customer behavior, empowering the business to make informed, data-driven decisions.

To build on the foundation established in this project, several enhancements are recommended:

- Integrate data from other sources such as web traffic, customer support, and marketing platforms to create a more complete Customer 360 view.
- Implement real-time analytics for live monitoring of sales and customer activity, supporting quicker response to demand fluctuations.
- Develop predictive models to estimate Customer Lifetime Value (CLV), identify churn risk, and forecast future sales performance.
- Automate personalized marketing based on RFM segments, using targeted email campaigns and product recommendations.

These future improvements will further enable TechHub to move from descriptive analytics to predictive and prescriptive insights, strengthening its position in a competitive market.