



GROUP ASSIGNMENT

TECHNOLOGY PARK MALAYSIA

MODULE CODE: AICT020-4-2-EDAV

**MODULE NAME: EXPLORATORY DATA ANALYTICS &
VISUALIZATION**

INTAKE CODE: UCDF2407ICT(DI)

HAND OUT WEEK: WEEK 2

HAND IN WEEK: WEEK 13 @ 16.1.2026

GROUP NAME: TEAM NISSAN

DATASET NAME: RetailSports

Student Name	TP No.
HENG EE SERN	TP081786
LAEU ZI-LI	TP083604
LOW ZE XUAN	TP082620
MUHAMMAD FARRIS BIN RAZMAN	TP082730
TAN HAO SHUAN	TP080852

Table of Contents

1.0 Summary of the Cleaned Dataset.....	3
1.1 Explanation of Each Visualizations & Statistical Findings	5
1.1.1 Executive Summary	5
1.1.2 Product Strategy	9
1.1.3 Customer and Operations.....	13
1.1.4 Analysis Sandbox.....	17
2.0 Actionable Recommendations	20
2.1 Logistics Optimization by Fixing the Nationwide 5-Day Delay of Processing Orders.....	20
2.2 Margin Recovery by Assortment Rationalization of the Golf Category.....	22
2.3 Restructuring Express Stores	24
2.4 Reallocate Inefficient Discounts from Swimming to Footwear	26
2.5 Pivot Marketing Strategy to the Underserved Female Demographic	27
2.6 Implement Dynamic Pricing to Boost Margin.....	28
3.0 Conclusion	29
4.0 References.....	30
5.0 Appendix	31
5.1 Dashboard Link.....	31
5.1 Work Breakdown Structure.....	31

1.0 Summary of the Cleaned Dataset

Data cleaning is a very important process to ensure the integrity of the Retail Sports dataset. Firstly, data validation and type conversion was performed to verify the data type across the dataset, as shown in Figure 1. By performing data validation and type conversion, accurate calculation can be ensured during the modelling phase. Raw tables were then duplicated and labelled as “Clean” to allow fallback. Secondly, text cleaning and standardization. A new column called CustomerName was created by merging First Name and Last Name for readability and simplicity. At the same time, all text columns were trimmed and cleaned to remove irregular whitespace and non-printable characters.

Thirdly, handling missing values. A comprehensive null values assessment was conducted and confirmed several missing values columns. For example, Email column contained 5% missing values, Phone column contained 10% missing values and BirthDate column contained 3% missing values. The Model column in Products table which contained 100% missing values was excluded from analysis.

Feature extraction and engineering was performed to create new analytical dimensions. The “@” delimiter was used in Email column to isolate email domains for marketing channel purposes. A new column named HasEmail was also created to show customers who have an email account. A conditional Region column was also engineered to classify customers into "Peninsular Malaysia" or "East Malaysia" based on their state. For Products table, PriceBand was created to categorise prices into Low, Mid or High and BrandStatus was created to check the availability. In the Sales table, ShipDelayDays was calculated to check the days delayed for each shipment and NegativeFlag was check for non-delayed shipments. Lastly, LineAmount was to check the total amount needed to pay after discount.

Lastly, data modelling. Relationships were established between cleaned dimension tables and fact tables.

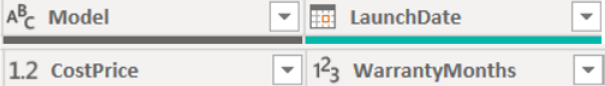
Action taken	Reason	Comments / Example
Check the data type for all columns and tables	Every data type is correct and suitable	 <p>Texts are used for Model, date is used for LaunchDate, decimal is used for CostPrice and whole numbers are used for WarrantyMonths</p>

Figure 1: Data Validation and Type Conversion

1.1 Explanation of Each Visualizations & Statistical Findings

1.1.1 Executive Summary



Figure 2: Executive Summary Report Page and all of its visuals

1.1.1.1 Line Chart

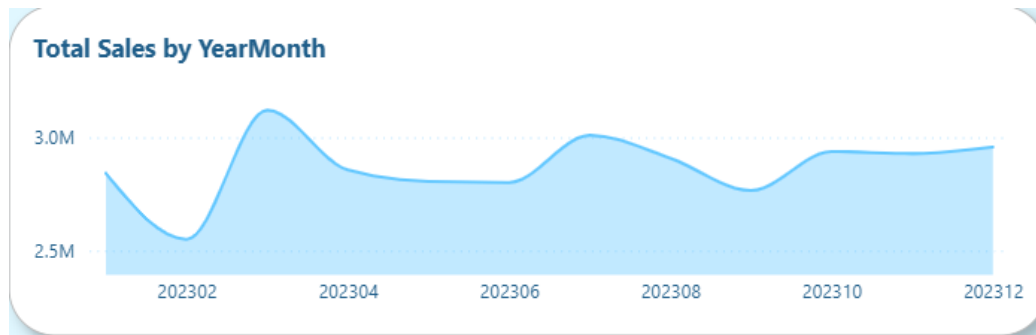


Figure 3: Line Chart indicating the Total Sales by YearMonth

This line chart shows the total sales by time in YYYYMM format. The highlighted area below the graph tracks the sales performance over the period of time. X-axis shows the time covering the entire year of 2023. Y-axis shows the total sales in millions (M), ranging from roughly 2.5M to 3.0M.

The graph indicates the highest peak sale around March 2023, with the sales exceeding 3 million, while the lowest drop in sales in February 2023 around 2.5 million. After April 2023, the sales stabilized around 2.8 million. Other increment in sales include 2023 July reaching close to 3 million before dropping back to its initial 2.8 million, and October 2023 with an increase to around 2.9 million with a continuous increment until the end of 2023.

1.1.1.2 Clustered Bar Chart

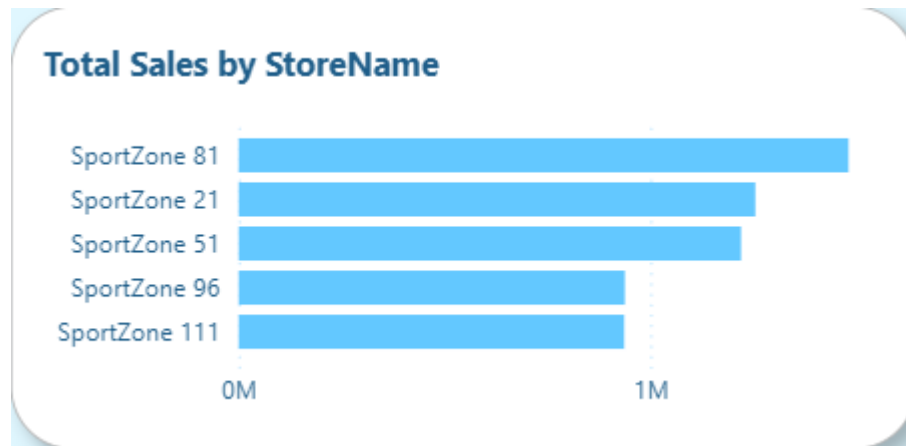


Figure 4: Clustered bar chart

The clustered bar chart the total sales by stores. The Y-axis shows the names of different stores such as SportZone 81 and SportZone 111, while the X-axis shows the total sales in millions (M) of that store.

The graph shows SportZone 81 as the top-performing store with a significantly higher sales of almost 1.5 million, whereas SportZone 96 and SportZone 111 having the lowest performance of less than 1 million. SportZone 21 and SportZone 51 having crossed 1 million sales, following close to the leading store.

1.1.1.3 Map



Figure 5: A Map that indicates the Total Sales and Average Processing Day by State

This map shows the total sales and average processing day by state. The bubble size of a certain area represents the total sales. Larger circle indicates higher sales volume within the region.

The map highlights the higher volume of sales around Kuala Lumpur area. Other areas having lower sales volume such as Ipoh, Melaka and Seremban.

1.1.2 Product Strategy

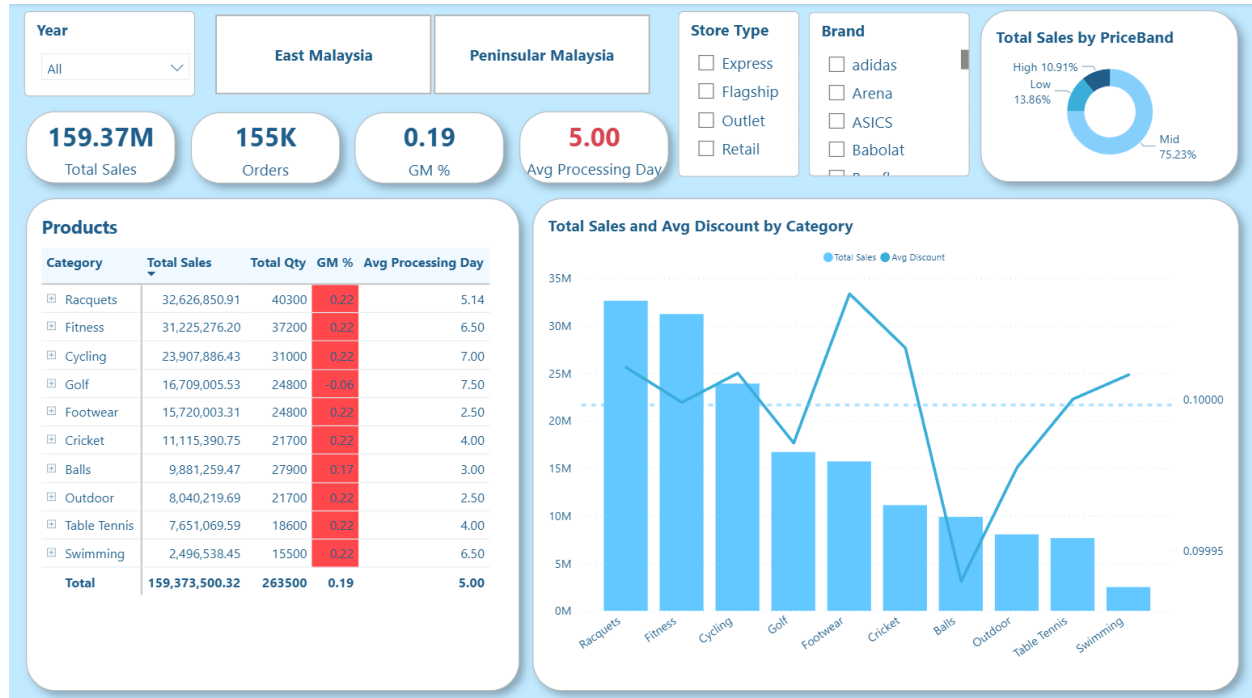


Figure 6: Product Strategy Report Page with all of its visuals

1.1.2.1 Donut Chart

Figure 7: Donut Chart indicating the Total Sales by PriceBand

This donut chart shows the three distinct pricing categories, high, mid and low. The chart indicates medium price band as the overwhelming driver of revenue, with a total of 75.27% from the total sales. The minority segments such as high price band and low price band occupy 10.89% and 13.84% of total sales.

1.1.2.2 Matrix

Products Strategy				
Category	Total Sales	Total Qty	GM %	Avg Processing Day
⊕ Racquets	32,626,850.91	40300	0.22	5.14
⊕ Fitness	31,225,276.20	37200	0.22	6.50
⊕ Cycling	23,907,886.43	31000	0.22	7.00
⊕ Golf	16,709,005.53	24800	-0.06	7.50
⊕ Footwear	15,720,003.31	24800	0.22	2.50
⊕ Cricket	11,115,390.75	21700	0.22	4.00
⊕ Balls	9,881,259.47	27900	0.17	3.00
⊕ Outdoor	8,040,219.69	21700	0.22	2.50
⊕ Table Tennis	7,651,069.59	18600	0.22	4.00
⊕ Swimming	2,496,538.45	15500	0.22	6.50
Total	159,373,500.32	263500	0.19	5.00

Figure 8: Matrix for Product Strategy that indicates the Gross Margin % of each category

This matrix shows the detailed data table for the performance of different product categories. The data are broken down to categories including total sales, total quantity, gross margin percentage and average processing day.

Data shows racquets as the top-selling product with a profit of over 32 million while swimming equipment being around 2 million. Many products such as racquets, fitness and table tennis have the same gross margin percentage with a total of 22% whereas golf having a low percentage of -6%.

1.1.2.3 Line and Clustered Column Chart

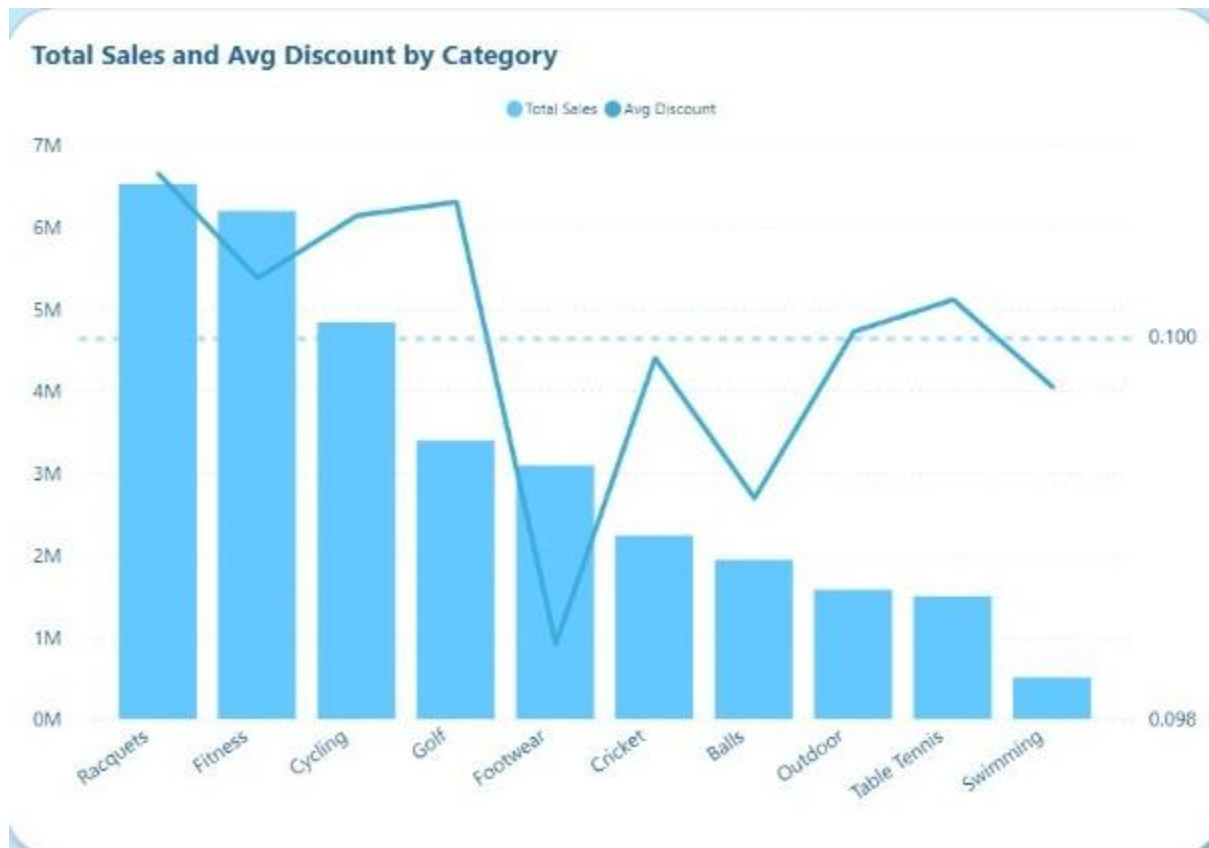


Figure 9: Line and Clustered Column Chart indicating the Total Sales and Average Discount by Category

This combo chart plots the total sales against the average discount. The clustered column chart indicates the total sales based on category. The line indicates the average discount for the total profit margin.

The chart shows a very low discount rate for footwear and a high discount rate for racquets. Unlike racquets, golf have high discount rates but low sales, whereas footwear having a very low discount rate produce an average amount of sales.

1.1.3 Customer and Operations



Figure 10: Customer and Operations Report Page with all of its visuals

1.1.3.1 Clustered Bar Chart

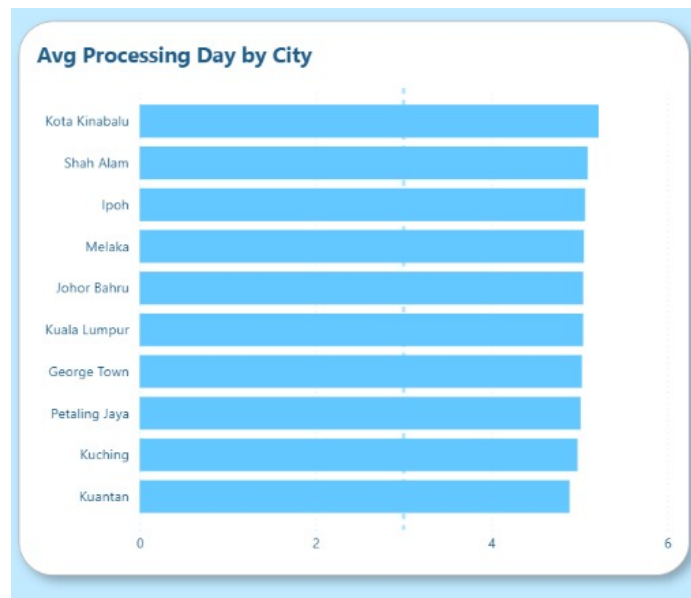


Figure 11: Clustered Bar Chart indicating the Average Processing Day by City

This clustered bar chart ranks different Cities by their Average Processing Day. This visual identifying which specific geographical areas are contributing most the high 5-day average. Kota Kinabalu has the longest bar so that its show that it is a worst performer in terms of processing times.

1.1.3.2 Clustered Column Chart

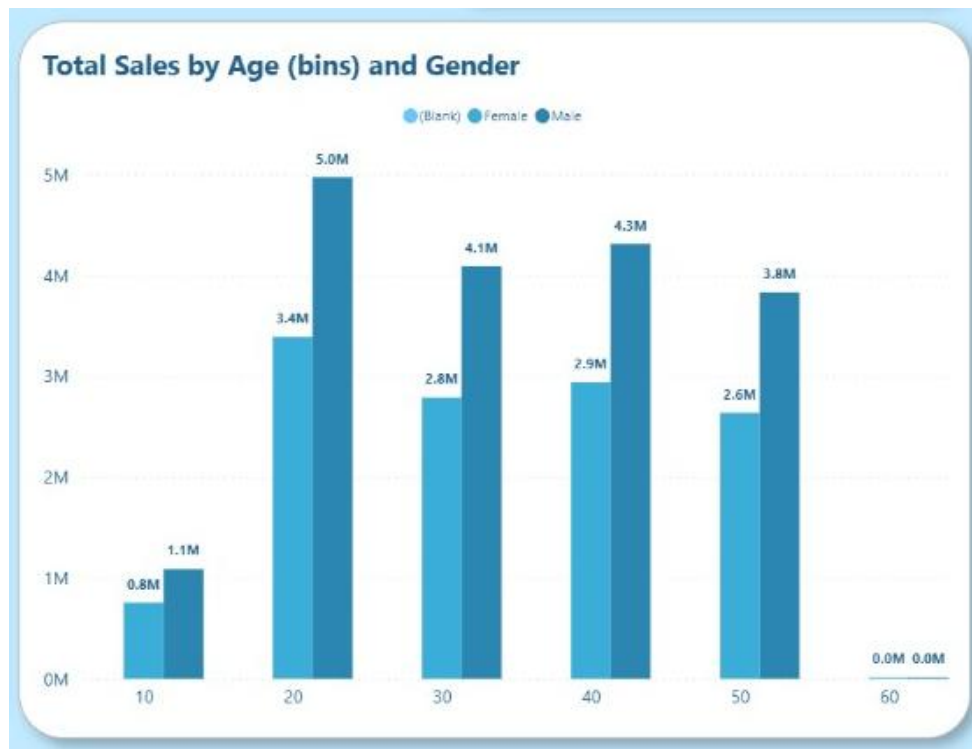


Figure 12: Clustered Column Chart indicating the Total Sales by Age and Gender

This chart segments Total Sales across various Age Bins (e.g., 20s, 30s) and clusters them by Gender (Male/Female). The visual confirms that the 20-29 age bin is the most revenue-critical demographic. Within this bin, the Male cluster drives significantly higher sales (24.9M) compared to the Female cluster (17.1M)

1.1.3.3 Gauge



Figure 13: Gauge indicating the Percentage of Late Order Rate

The gauge in Figure x indicated a catastrophic service failure. A 60.00% late order rate is unacceptable and signals a profound crisis in the company's operational and logistical capabilities. This failure is the single greatest threat to customer loyalty and brand reputation.

1.1.4 Analysis Sandbox

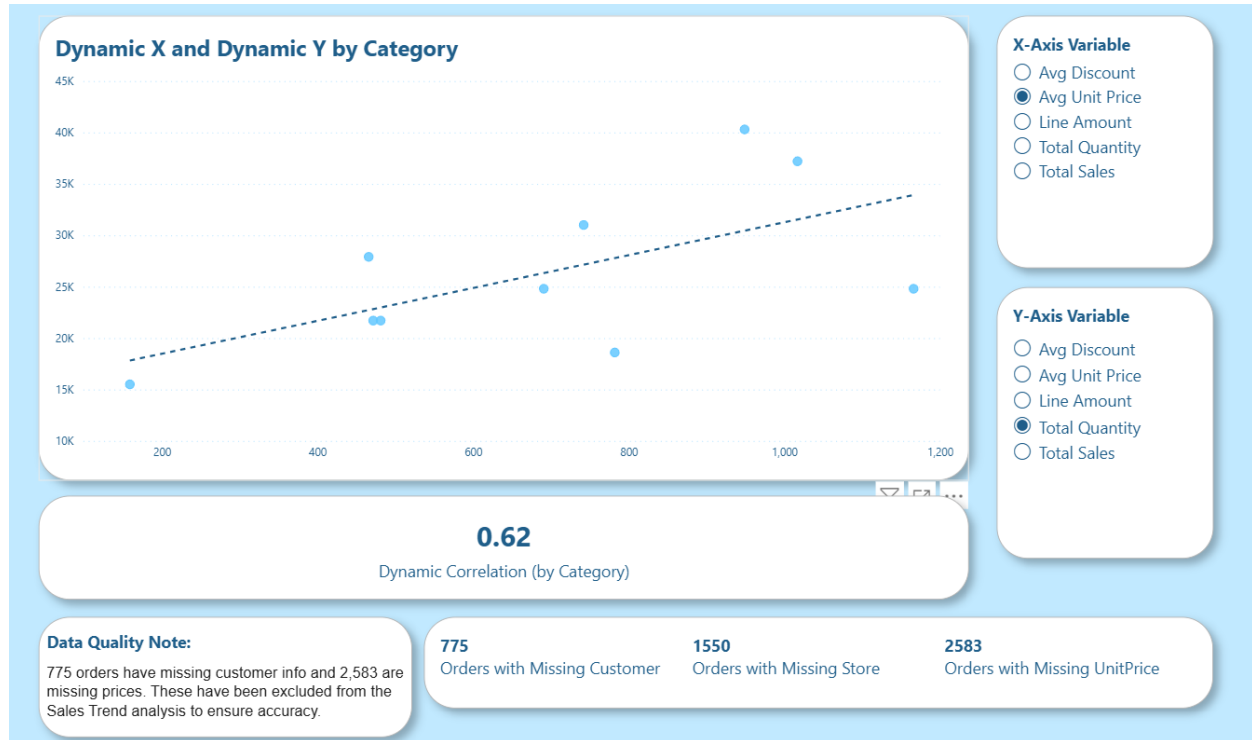


Figure 14: Analyst Sandbox Report Page and all its visuals

1.1.4.1 Scatter Chart

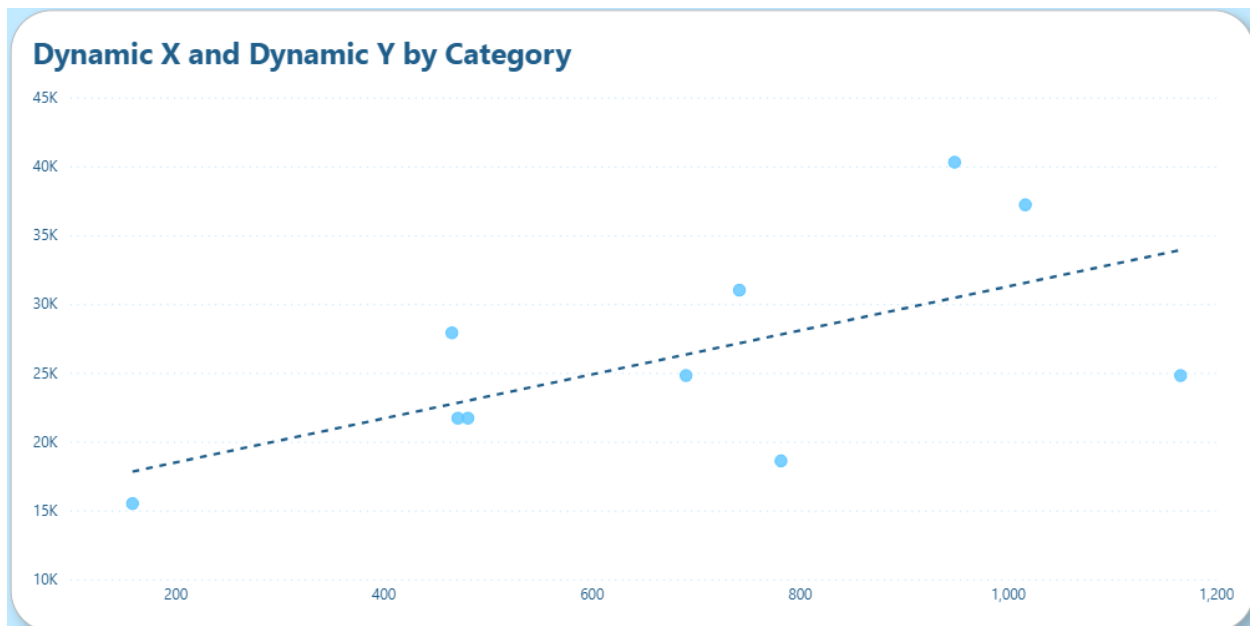


Figure 15: Scatter Chart indicating the Dynamic X and Dynamic Y by Category

Figure 15 shows the scatter chart of Analyst Sandbox. This interactive visualization allows users to dynamically select different variables for X and Y axes using Measure Selectors as List Slicer.

1.1.4.2 Dynamic Correlation by Category Card

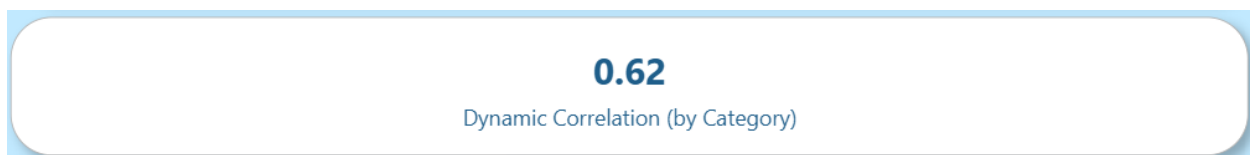


Figure 16: Card indicating the Dynamic Correlation by Category

Figure 16 shows the Card of dynamic correlation by Category. It calculates and displays Pearson Correlation Coefficient based on the variables selected in the scatter chart. The purpose of this visualization is for quantification. Scatter chart provides a visual representation of the trend, while this card provides the precise mathematical value that shows the strength of the specific relationship. A value of 0.62 shows a moderate to strong positive relationship between Avg Unit Price and Total Quantity, which indicates when Avg Unit Price increases, Total Quantity also tends to increase.

1.1.4.3 Statistical Findings of Orders Card

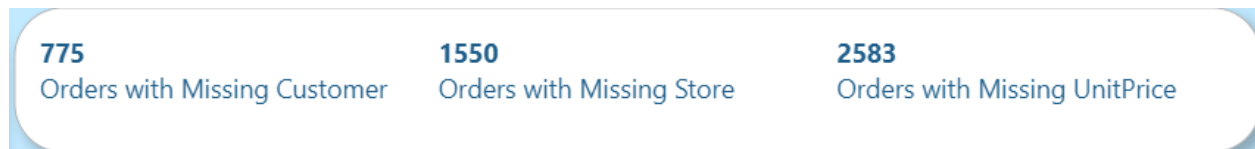


Figure 17: Card indicating the Statistical Findings of Orders

In Figure 17, the statistical findings card act as an audit log. It displays the count of records that has failed the validation check during the cleaning process.

2.0 Actionable Recommendations

2.1 Logistics Optimization by Fixing the Nationwide 5-Day Delay of Processing Orders

Based on data analysis, the report has indicated an underwhelmingly inefficient process of package fulfillment. As shown in Figure 18, in 2025, the reported average time for delivery across Peninsular Malaysia was 5.02 days while East Malaysia was 5.11 days, making both delivery areas very similar in terms of performance averaging 5.04 days. This concludes that the geographical distance is not solely the factor that causes the time difference, it is instead proven that this problem occurs during the "picking-and-packing" process at the warehouse, way before the courier would eventually arrive to pick-up the goods. By resolving this issue, it is expected to immediately reduce the value of the "Late Order Rate" KPI.

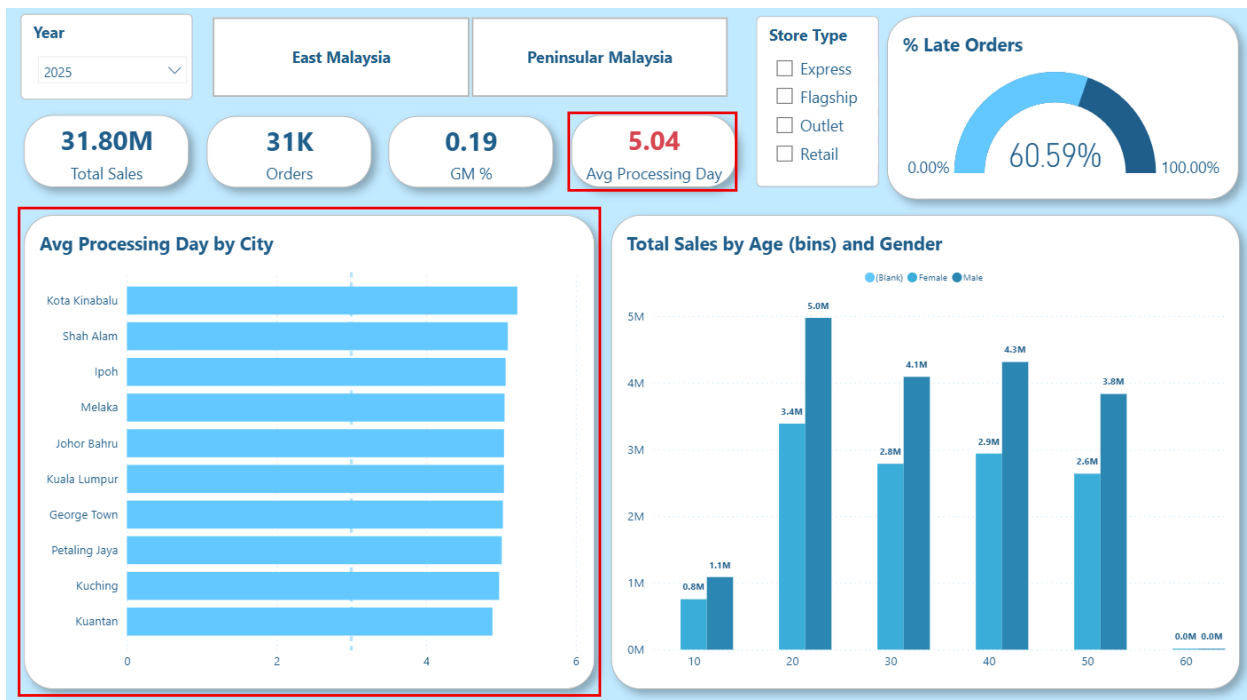


Figure 18: Customer and Operations Report Page highlighting the poor average processing days

The actionable steps to take to resolve this issue are by implementing a warehouse audit that enforces a "24-hour shipping guarantee" KPI, it makes strict policy to abide by. Derived from

the data, it has implied that orders are completely stagnant for an average of 2-3 days at the warehouse before the hand-over to the logistics partner. Other than that, a review on the Vendor Service Level Agreement (SLA) would determine whether the factor of poor performance is on the logistics partner itself, specifically not being able to deliver within 3 working days, prompting to switch to a more reliable courier.

The actions mentioned are expected to reduce the value in "ShipDelayDays", targeting a 3-day delivery guarantee within Peninsular Malaysia. On the other hand, an improved customer retention is expected as an improved overall delivery speed would result in improving customer satisfaction and return to purchase more. Lastly, a significant reduction of the value in "Late Order Rate" percentage is expected.

2.2 Margin Recovery by Assortment Rationalization of the Golf Category

It is found that the "Golf" category aims for high revenue but causes a negative profit in return. As shown in Figure 19, although this specific category has been raking in RM 16.7 million in accumulated sales, the negative gross margin of -6%, it causes generating no profit at all, resulting in an all-time net loss of RM 1 million from those sales. If this issue is not resolved immediately, the company would lose a significant portion of money on every sale within the "Golf" category.

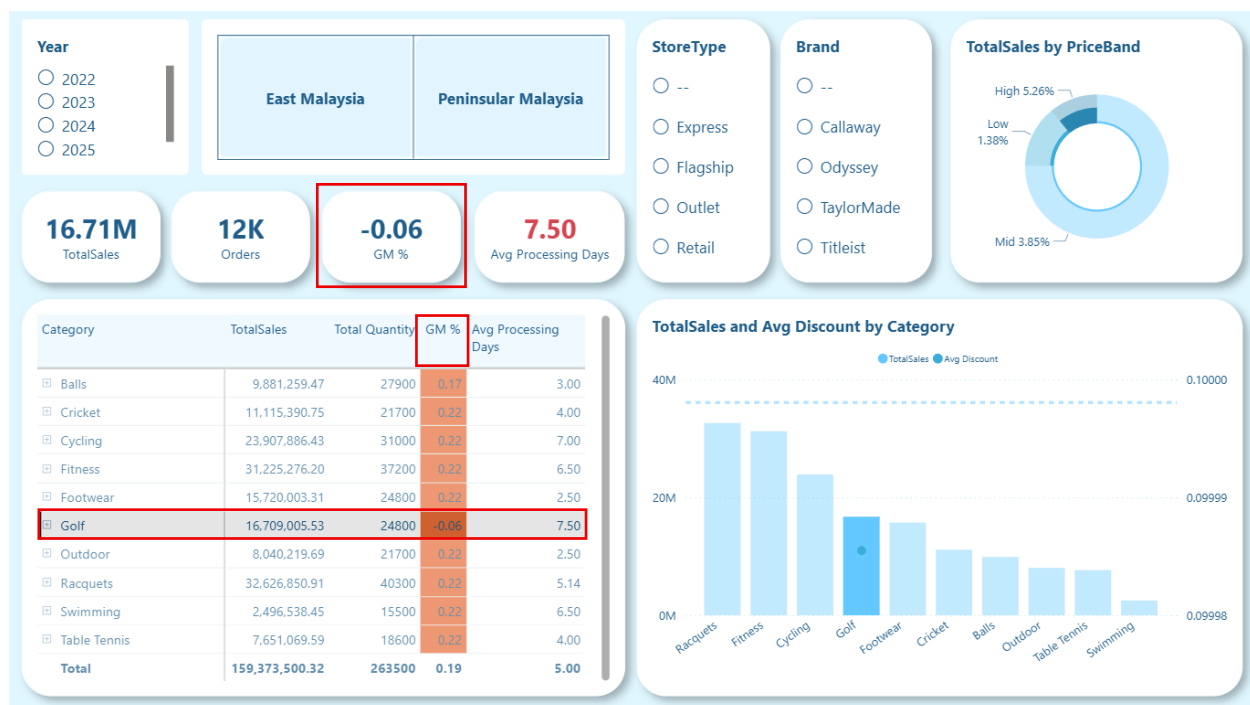


Figure 19: Product Strategy Report Page highlighting the profit loss under the Golf category

To immediately resolve this issue, a "surgical" removal of specific SKUs is conducted by un-listing/discontinuing the products that are causing a loss. By implementing the "Bottom 20%" Rule, it helps to identify exactly which SKUs have a negative margin, where the cost of acquiring the goods is more than its selling price (Tardi, 2025). Other than that, negotiating with the supplier by presenting the sales volume data of RM 16.7 million in revenue as leverage, convincing them to lower the prices to acquire the goods from them, as the movement of volume is there but unable to sustain the cost of the goods.

It is expected that there would be an immediate profit turnaround by converting the loss into an even or a more positive margin by one-quarter of a year. On the other hand, it should optimize the cash flow by eliminating the capital held by "dead stock" to focus on investing in categories of higher margin like Fitness, which has an approximate margin of 22%.

2.3 Restructuring Express Stores

Highlighting the issue specifically, as shown in Figure 20, it indicates that there are significant performance deviations between the profit margin depending on different store types. It is reported that ‘Outlet’ stores have the best profit margin of approximately 23%, while the ‘Express’ stores are way behind, generating only a 15% profit margin. Analysing this issue further, it is found that the ‘Express’ stores either suffered hefty operational costs that are unmatched to the sales generated or a poor product assortment that is inefficient and less appealing compared to other store types.

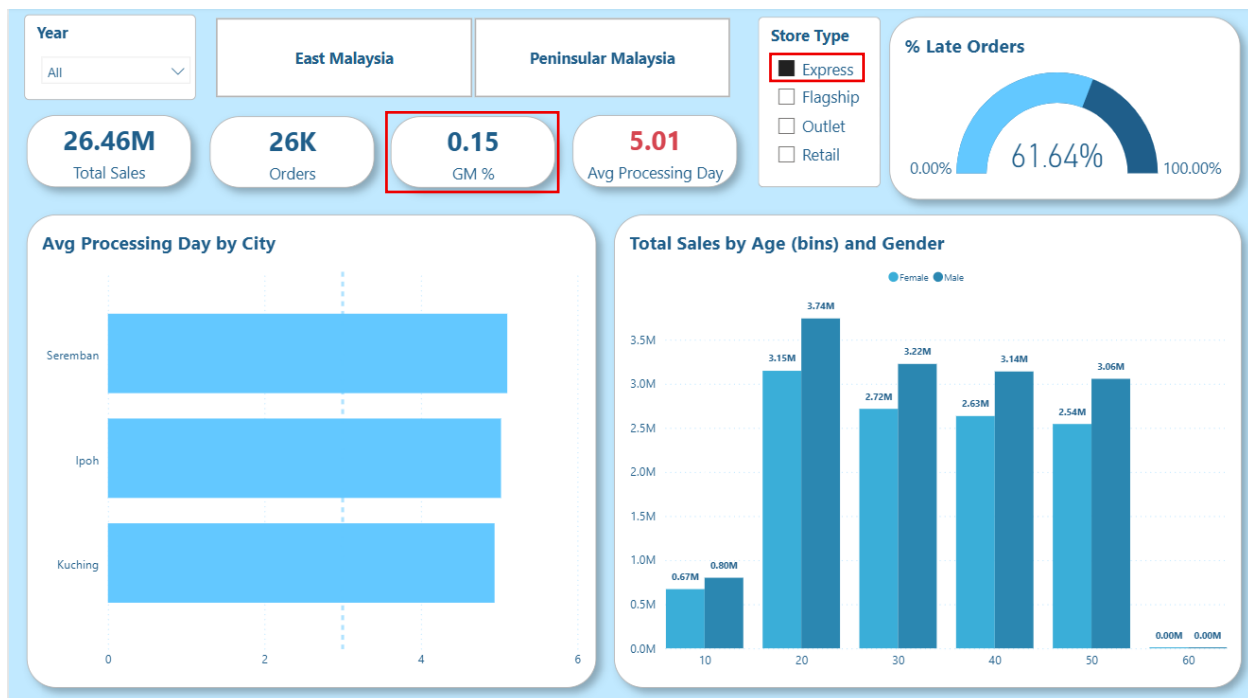


Figure 20: Customer and Operations Report Page highlighting the underwhelming performance of Express stores

The immediate actions to be implemented are inventory audits, specifically by re-planning a more strategic assortment of products to be sold within the ‘Express’ stores. Assuming that there are a majority of low-margin items being sold, especially the consumable products, while the stores are missing stock of high-margin products. Other than that, a format review can be implemented. To be more specific, a condition is made where if the store’s performance could not reach at least an 18% margin in 6 months, the following store can be reconsidered to either rebrand into a higher-

margin into a 'Retail' store type, or else, it is best to identify the bottom 10% of the stores and permanently close them down.

It is expected that a significant profit margin increases from a low 15% to a high 20% could potentially generate improved net profits, without relying on putting down more funds to open new stores that would lead to more operational costs into picture. The stores' operations should improve in efficiency as it will close the performance gap between the 'Express' stores and other store types.

2.4 Reallocate Inefficient Discounts from Swimming to Footwear

In the Total Sales and Average Discount by Category chart shown in Figure 21, we can observe that Swimming has the lowest sales, yet it receives a higher average discount from Footwear. This trend indicates inefficient spending to drive swimming purchases, but the volume does not respond proportionally. Conversely, Footwear shows healthy sales volume despite having the lowest discount rate on the chart.

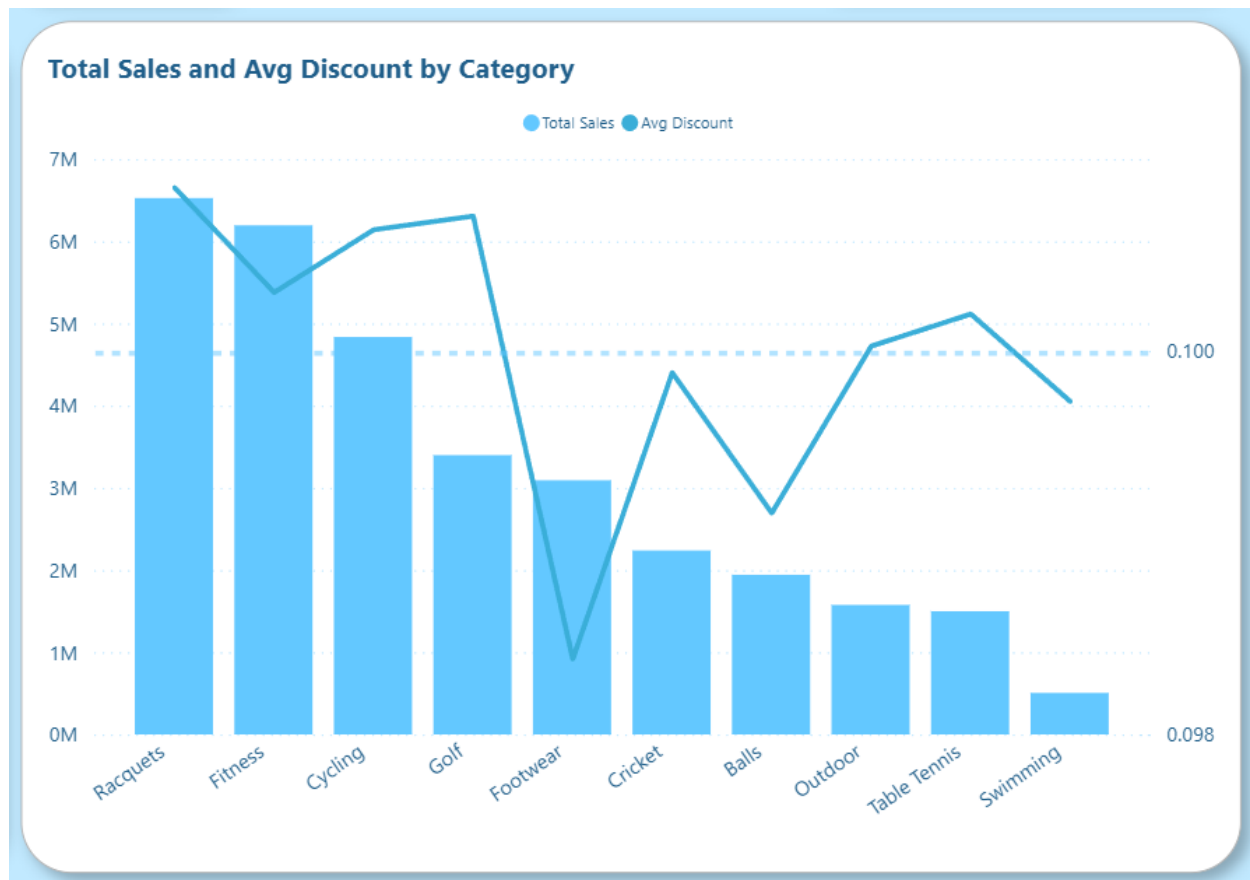


Figure 21: Line and Clustered Column Chart indicating the Total Sales and Average Discount by Category

Therefore, a shift of promotional budget from Swimming to a high-potential category such as Footwear is recommended. Discounts for Swimming products can be drastically reduced. The data suggest that low demand for Swimming products was not price related. Marketing funds can be reallocated to the Footwear category to subsidise a strategic price drop. Since Footwear already has organic traction at full price, adding a discount is highly likely to boost the sales volume into the top 3 categories.

2.5 Pivot Marketing Strategy to the Underserved Female Demographic

The Total Sales by Age and Gender chart in Figure 22 shows an obvious trend. The sales of the company are predominantly driven by Males across all age groups. The Female demographic is significantly lower. This represents a massive opportunity for growth in the female market.

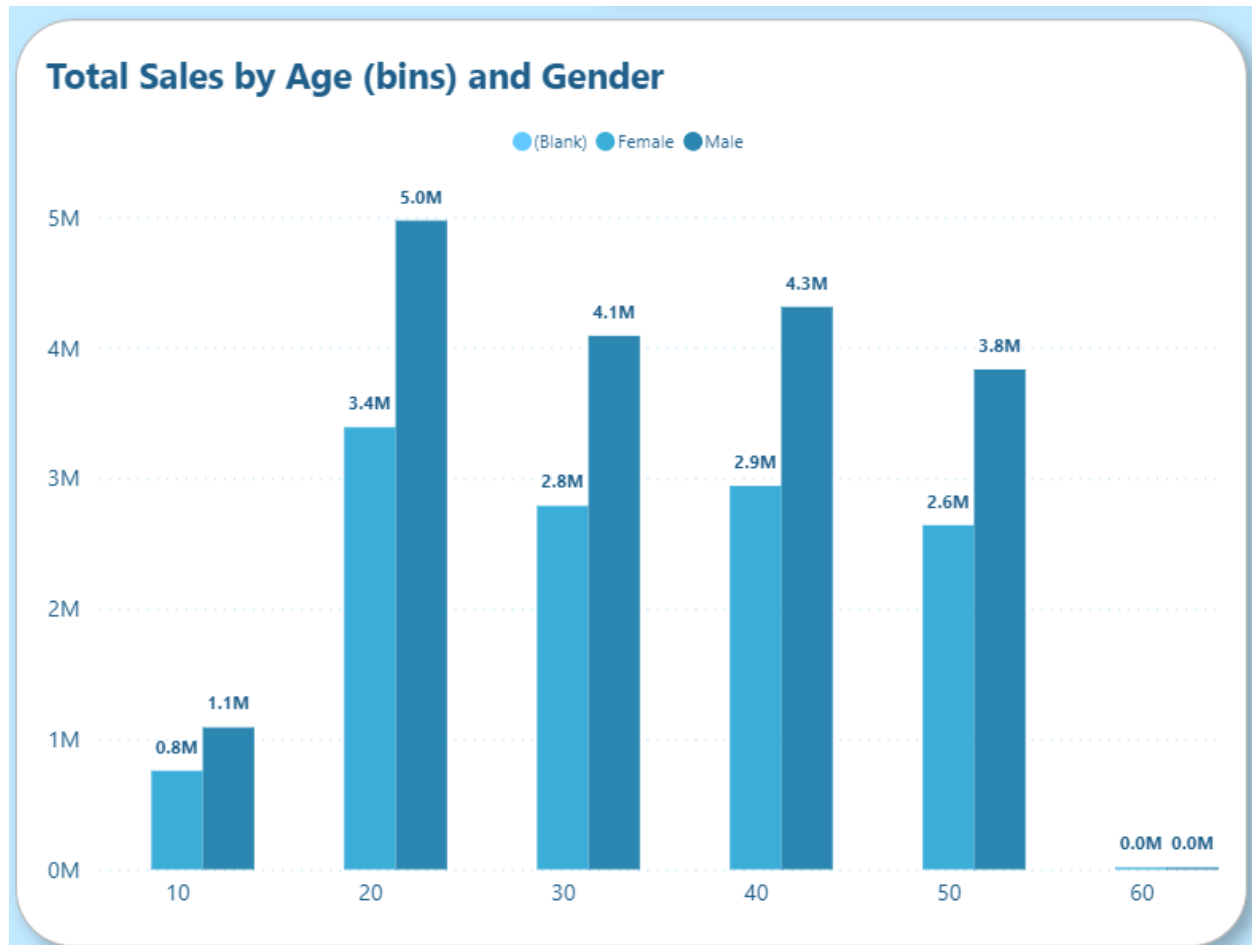


Figure 22: Clustered Column Chart indicating the Total Sales by Age and Gender

To drive the sales of the female market, a “Women’s Activewear” expansion campaign can be launched. First, an inventory audit can be done to analyse whether low female sales are due to the lack of stock or variety. If the stock is mostly unisex or male-focused gear, it will naturally alienate female buyers. Then, new categories can be added dedicated to females, such as Yoga, Pilates, and Women’s Running. Lastly, a partnership with female fitness influencers can drive sales to these new categories.

2.6 Implement Dynamic Pricing to Boost Margin

As shown in Figure 23, the correlation between Average Unit Price and Total Sales shows a strong correlation of 0.73. This indicates the strong pricing power of products. However, the business operates at a low gross margin, only 0.19, suggesting that there is missed revenue.

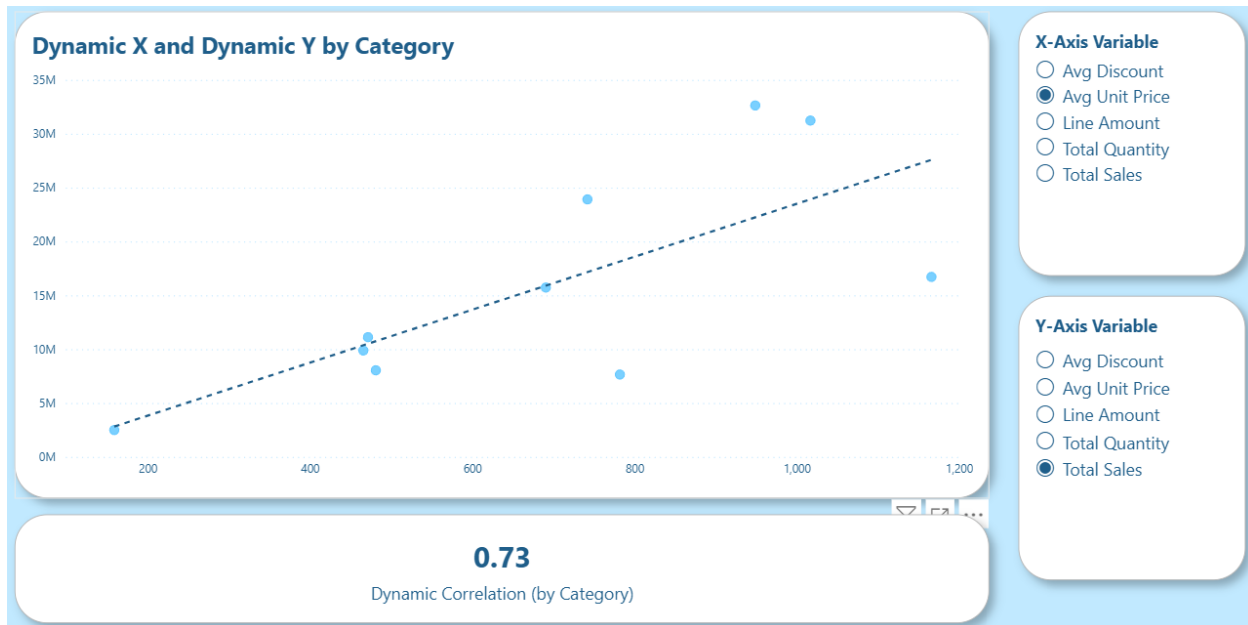


Figure 23: Scatter Chart indicating the Average Unit Price and Total Sales by Category

To combat this issue, a dynamic pricing system can be implemented to automatically adjust prices based on the inventory levels, competitor prices, and demand patterns. This can start with a pilot test on Racquets and Fitness categories that make up 40% of the company's sales, then roll out company wide. An increase in gross margin and revenue is expected.

3.0 Conclusion

Overall, the project achieved the implementation of a full-fledged solution for Exploratory Data Analytics and Visualization (EDAV) that provided the retail firm with valuable insights superior to what can normally be achieved through reporting. This solution stands out specifically due to the creation of a sustainable data model that has been carefully compiled and built with heavy data cleansing and data transformation and measure creation efforts that are explicitly mentioned in the Log Sheet. Also, the creation of a star data structure with specific data tables for the dimensions (Customers, Products, and Store) and a Fact Sales table helps maintain data validity. Additionally, the creation of a variety of measures like time intelligence measures 'Sales YTD' and data monitors like 'Orders with Missing Customers' helps make the data model future-proof.

The analytical value is seen in the dynamic, executable results that are created. Crucially, the dashboard goes beyond reporting, with the inclusion of sophisticated segmentation techniques, such as Sales by Age and Gender, and sophisticated analysis techniques, such as Dynamic Correlation. This enables readers to understand not just where they are failing (for instance, with regard to their 60% late order rate, as indicated in the Gauge chart) but how they actually fail, with regard to, for instance, discount levels and sales, allowing for a determination of the value of the project by turning information into a predictive, value-generative asset, where decisions are driven not just by intuition or gut feel, but by hard, measurable evidence, to change, for instance, the way they distribute marketing budgets or to rectify operational inefficiencies with regard to regional logistics distribution channels.

4.0 References





- Gallo, A. (2014, October 29). *The value of keeping the right customers*. Harvard Business Review. <https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>
- Tardi, C. (2025, May 30). *The 80-20 Rule (aka Pareto Principle): What It Is and How It Works*. Investopedia. <https://www.investopedia.com/terms/1/80-20-rule.asp>
- Tuovila, A. (2022, August 24). *Sales per square foot: Overview, measurements and FAQ*. Investopedia. <https://www.investopedia.com/terms/s/sales-per-square-foot.asp>

5.0 Appendix

5.1 Dashboard Link

<https://app.powerbi.com/groups/me/dashboards/104238e1-9156-4488-a8a3-1041a0743ba4?experience=power-bi>

5.1 Work Breakdown Structure

Name	Task / Description / Responsibility	Signature
Heng Ee Sern (TP081786)	<ul style="list-style-type: none"> - 2.4 Reallocate Inefficient Discounts from Swimming to Footwear - 2.5 Pivot Marketing Strategy to the Underserved Female Demographic - 2.6 Implement Dynamic Pricing to Boost Margin 	
Laeu Zi-Li (TP083604)	<ul style="list-style-type: none"> - Executive Summary - Product Strategy 	
Low Ze Xuan (TP082620)	<ul style="list-style-type: none"> - Summary of the Cleaned Dataset - Analysis Sandbox 	
Muhammad Farris Bin Razman (TP082730)	<ul style="list-style-type: none"> - Actionable Recommendations - 2.1 Logistics Optimization by Fixing the Nationwide 5-Day Delay of Processing Orders - 2.2 Margin Recovery by Assortment Rationalization of the Golf Category 	

	<ul style="list-style-type: none">- 2.3 Restructuring Express Stores	
Tan Hao Shuan (TP080852)	<ul style="list-style-type: none">- Customer and Operation- Conclusion	