SULIT

FACULTY OF COMPUTING
UTM Johor Bahru

UNIVERSITI TEKNOLOGI MALAYSIA
FINAL EXAMINATION SEMESTER 1 2022/2023

SUBJECT CODE	**:**	**MCSD1013**
SUBJECT NAME	**:**	**BUSINESS INTELLIGENCE AND ANALYTICS**
SECTION	**:**	**01**
TIME	**:**	0900
DATE/DAY	**:**	**19 JUNE – 3 JULY 2023**
VENUES	**:**	**ONLINE**

INSTRUCTIONS:

This alternative assessment is an INDIVIDUAL assessment consists of **THREE (3)** questions/phases:

1) Data preparation
2) Dashboard development
3) Reporting of analysis
   Each student will be evaluated based on ALL the questions/phases given (Please attach this page as the front page of your report)
4) Submission to be done through E-learning

(Please Write Your Lecturer Name And Section In Your Answer Booklet)

Name	**FAN CHIN WEI**
I/C No.	**970324-01-6883**
Year / Course	**1/MCSD**
Section	**01**
Lecturer Name	**DR. NOR ERNE NAZIRA BINTI BAZIN**

This question paper consists of **____ (__)** printed pages excluding this page.

Table of Contents

Business Problem Formulation

The supermarket chain is experiencing stagnant or declining sales across its stores. The company lacks a centralized and comprehensive analytics solution to identify sales trends, customer preferences, and opportunities for improvement. The aim of this project is to analyze and optimize supermarket sales performance by examining customer flowrate and devising effective business strategies. The problem entails understanding the factors that influence sales performance such as customer behaviour, product placement and product pricing. By analyzing customer flow rate, we can identify peak hours, bottlenecks, customer preferences and customer gender that enable us to improve the overall shopping experience. In addition, this analysis will help in the development of targeted business strategies that encompass personalized marketing campaigns, optimal inventory management and efficient staff allocation. To address these issues, companies need to carefully analyze their sales performance and customers detail in order to develop continuously optimized strategies to increase sales. This project emphasizes on increasing sales performance of supermarket with the help of Power BI tools to develop business strategies. The objective of this project is analyzing sales data to identify trends, patterns, factors influencing sales performance. Another objective is gaining insights into customer preference, behaviour, and shopping habits. In summary, this project is empowering the supermarket store management team with the necessary tools to leverage analytics effectively.

Stakeholder

The key stakeholders involved in this project are:

a) Senior management: Responsible for providing support, resources, and guidance.

b) Store managers: Responsible for implementing strategies and ensuring successful adoption.

c) Sales and marketing teams: Responsible for executing sales and marketing initiatives based on data-driven insights.

d) IT department: Responsible for providing technical support and ensuring data integrity.

e) Customers: Beneficiaries of the project through enhanced shopping experiences.

Data Source and Type

In this project, the data obtained from Kaggle website: (https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales?resource=download) to develop data-driven strategies to enhance sales, improve customer satisfaction and increase customer retention. Besides that, another dataset obtained from online: (https://www.citypopulation.de/en/myanmar/cities/) represents as abbreviation city of Myanmar. This dataset is in .csv format which consisting of 1000 rows and 17 columns. In this project, we are used two datasets to conduct this project. Thus, data ingestion is needed in this case. Table 1 shown the information of the dataset used in this project.

No.	Column Name	Data Type	Description
1	Invoice id	String	Computer generated slip invoice identification number
2	Branch	String	Branch of supermarket 1. Branch A = Yangon city 2. Branch B = Mandalaycity 3. Branch C = Naypyitw
3	City	String	Location of supermarket
4	Customer type	String	Type of customers, recorded by

			members for customers using member card and normal for without member card
5	Gender	String	Gender type of customer
6	Product line	String	General item categorization groups
7	Unit Price	Float	Price of each product in $
8	Quantity	Integer	Number of products purchased by customer
9	Tax	Float	5% tax fee for customer buying
10	Total	Float	Total price including tax
11	Date	String	Date of purchase
12	Time	String	Purchase time
13	Payment	String	Payment used by customer for purchase
14	COGS	Float	Cost of goods sold
15	Gross margin percentage	Float	Gross margin percentage
16	Gross income	Float	Gross income
17	Rating	Float	Customer stratification rating

Implementation of Project Planning

In this section, the data cleaning and preparation process is executed in this step. Figure 1 shows supermarket sales data as below. Figure 1 only shows 17 columns of attributes and 28 rows data. Figure 2 shows 2 columns of attributes and 16 rows data. This dataset renames as "Supermarket Data" and "State Abbreviation" are used and imports into PowerBI for data preparation and build dashboard visualization.

1	Invoice ID	Branch	City	Custome	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating
2	750-67-8428	A	Yangon	Member	Female	Health and be	74.69	7	26.1415	548.972	1/5/2019	13:08	Ewallet	522.83	4.761904762	26.1415	9.1
3	226-31-3081	C	Naypyita	Normal	Female	Electronic acc	15.28	5	3.82	80.22	3/8/2019	10:29	Cash	76.4	4.761904762	3.82	9.6
4	631-41-3108	A	Yangon	Normal	Male	Home and life	46.33	7	16.2155	340.526	3/3/2019	13:23	Credit ca	324.31	4.761904762	16.2155	7.4
5	123-19-1176	A	Yangon	Member	Male	Health and be	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.761904762	23.288	8.4
6	373-73-7910	A	Yangon	Normal	Male	Sports and tra	86.31	7	30.2085	634.379	2/8/2019	10:37	Ewallet	604.17	4.761904762	30.2085	5.3
7	699-14-3026	C	Naypyita	Normal	Male	Electronic acc	85.39	7	29.8865	627.617	3/25/2019	18:30	Ewallet	597.73	4.761904762	29.8865	4.1
8	355-53-5943	A	Yangon	Member	Female	Electronic acc	68.84	6	20.652	433.692	2/25/2019	14:36	Ewallet	413.04	4.761904762	20.652	5.8
9	315-22-5665	C	Naypyita	Normal	Female	Home and life	73.56	10	36.78	772.38	2/24/2019	11:38	Ewallet	735.6	4.761904762	36.78	8
10	665-32-9167	A	Yangon	Member	Female	Health and be	36.26	2	3.626	76.146	1/10/2019	17:15	Credit ca	72.52	4.761904762	3.626	7.2
11	692-92-5582	B	Mandala	Member	Female	Food and beve	54.84	3	8.226	172.746	2/20/2019	13:27	Credit ca	164.52	4.761904762	8.226	5.9
12	351-62-0822	B	Mandala	Member	Female	Fashion acces	14.48	4	2.896	60.816	2/6/2019	18:07	Ewallet	57.92	4.761904762	2.896	4.5
13	529-56-3974	B	Mandala	Member	Male	Electronic acc	25.51	4	5.102	107.142	3/9/2019	17:03	Cash	102.04	4.761904762	5.102	6.8
14	365-64-0515	A	Yangon	Normal	Female	Electronic acc	46.95	5	11.7375	246.488	2/12/2019	10:25	Ewallet	234.75	4.761904762	11.7375	7.1
15	252-56-2699	A	Yangon	Normal	Male	Food and beve	43.19	10	21.595	453.495	2/7/2019	16:48	Ewallet	431.9	4.761904762	21.595	8.2
16	829-34-3910	A	Yangon	Normal	Female	Health and be	71.38	10	35.69	749.49	3/29/2019	19:21	Cash	713.8	4.761904762	35.69	5.7
17	299-46-1805	B	Mandala	Member	Female	Sports and tra	93.72	6	28.116	590.436	1/15/2019	16:19	Cash	562.32	4.761904762	28.116	4.5
18	656-95-9349	A	Yangon	Member	Female	Health and be	68.93	7	24.1255	506.636	3/11/2019	11:03	Credit ca	482.51	4.761904762	24.1255	4.6
19	765-26-6951	A	Yangon	Normal	Male	Sports and tra	72.61	6	21.783	457.443	1/1/2019	10:39	Credit ca	435.66	4.761904762	21.783	6.9
20	329-62-1586	A	Yangon	Normal	Male	Food and beve	54.67	3	8.2005	172.211	1/21/2019	18:00	Credit ca	164.01	4.761904762	8.2005	8.6
21	319-50-3348	B	Mandala	Normal	Female	Home and life	40.3	2	4.03	84.63	3/11/2019	15:30	Ewallet	80.6	4.761904762	4.03	4.4
22	300-71-4605	C	Naypyita	Member	Male	Electronic acc	86.04	5	21.51	451.71	2/25/2019	11:24	Ewallet	430.2	4.761904762	21.51	4.8
23	371-85-5789	B	Mandala	Normal	Male	Health and be	87.98	3	13.197	277.137	3/5/2019	10:40	Ewallet	263.94	4.761904762	13.197	5.1
24	273-16-6619	B	Mandala	Normal	Male	Home and life	33.2	2	3.32	69.72	3/15/2019	12:20	Credit ca	66.4	4.761904762	3.32	4.4
25	636-48-8204	A	Yangon	Normal	Male	Electronic acc	34.56	5	8.64	181.44	2/17/2019	11:15	Ewallet	172.8	4.761904762	8.64	9.9
26	549-59-1358	A	Yangon	Member	Male	Sports and tra	88.63	3	13.2945	279.185	3/2/2019	17:36	Ewallet	265.89	4.761904762	13.2945	6
27	227-03-5010	A	Yangon	Member	Female	Home and life	52.59	8	21.036	441.756	3/22/2019	19:20	Credit ca	420.72	4.761904762	21.036	8.5
28	649-29-6775	B	Mandala	Normal	Male	Fashion acces	33.52	1	1.676	35.196	2/8/2019	15:31	Cash	33.52	4.761904762	1.676	6.7

Figure 1 Supermarket sales data

City name	Abbreviation
Myanmar	MMR
Ayeyarwady [Irrawaddy]	AYE
Bago	BAG
Chin	CHI
Kachin	KAC
Kayah	KYH
Kayin (Karen)	KYN
Magway (Magwe)	MAG
Mandalay	MAN
Mon	MON
Naypyitaw	NAY
Rakhine (Arakan)	RAK
Sagaing	SAG
Shan	SHA
Tanintharyi (Tenasserim)	TAN
Yangon	YAN

Figure 2 Abbreviation city of Myanmar

Data cleaning and preparation is an important phase in the data analysis process since it helps to ensure the accuracy and quality of the final results. The data preparation stages in this project are carried out using Python, Power Query Editor and PowerBI Desktop. Python is used to extensively investigate the dataset to understand the patterns, trends and identifying any potential patterns. Based on results from figure 3 and figure 4, there are no null values and duplicated data in the dataset.

```
import numpy as np
sales_data.isnull().sum()
```

```
Invoice ID                0
Branch                    0
City                      0
Customer type             0
Gender                    0
Product line              0
Unit price                0
Quantity                  0
Tax 5%                    0
Total                     0
Date                      0
Time                      0
Payment                   0
cogs                      0
gross margin percentage   0
gross income              0
Rating                    0
dtype: int64
```

Figure 3 Checking null values

```
sales_data[sales_data.duplicated()]
```

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating

Figure 4 Checking duplicated data

In addition, each column are checked using Power Query Editor function such as column quality, column distribution and column profile as shown in figure 5 and figure 6.

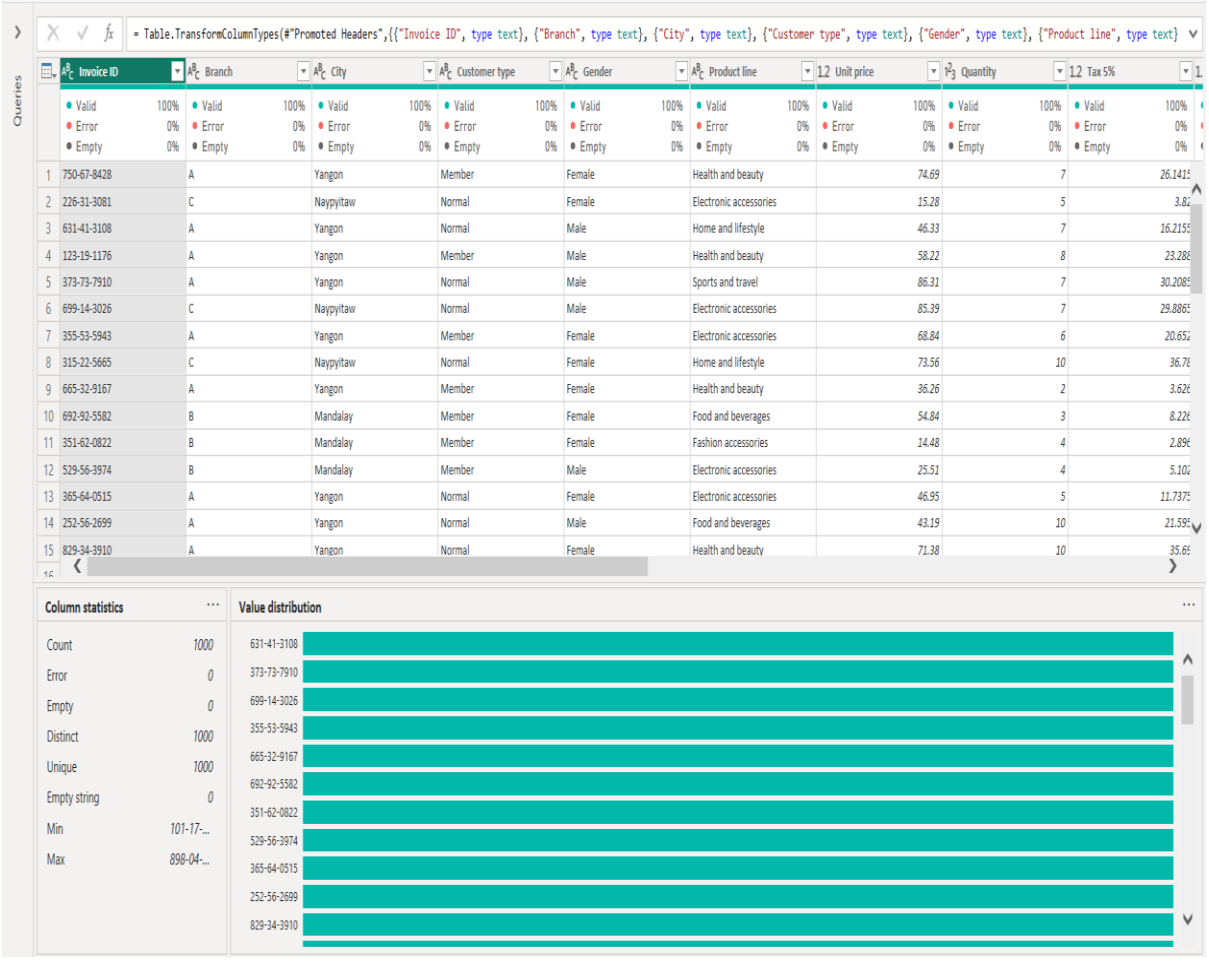

Figure 5 Data Checking

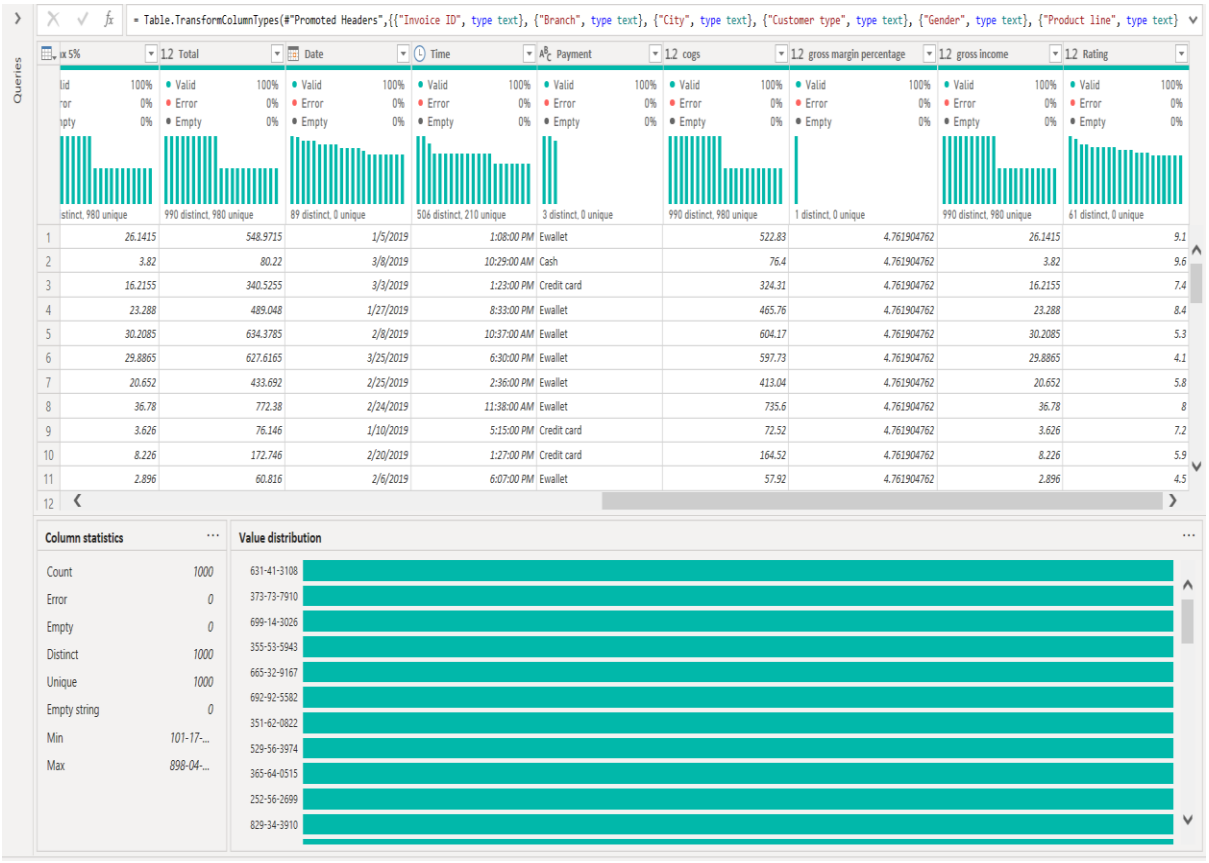

Figure 6 Data Checking

Data Visualization and Analytics

In this phase, the supermarket sales data will be fully utilized and visualized by using dashboards to explore the ales performance and factors to affect sales performance which align with the objectives of the project. There are 4 interactive dashboards are created such as net sales performance dashboard, customer information dashboard, sales influencers dashboard and predictive dashboard. The goal and objectives of this project shows as below:

1. Analyzing sales data to identify trends, patterns, and factors influencing sales performance.
2. Gaining insights into customer preferences, behaviour, and shopping habits.
3. Developing data-driven strategies to enhance sales, improve customer satisfaction, and increase customer retention.
4. Providing stakeholders with interactive dashboards and reports for real-time monitoring and decision-making.
5. Empowering the supermarket store management team with the necessary tools and knowledge to leverage analytics effectively.

Dashboard 1: Net sales performance dashboard

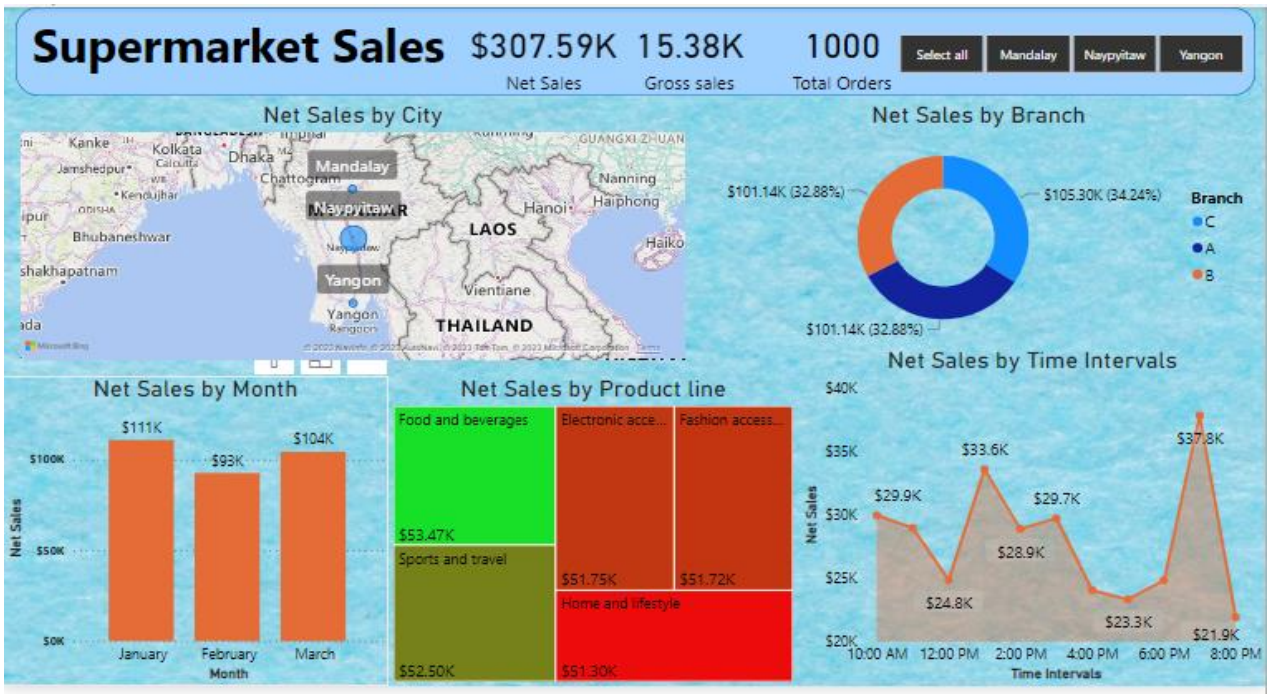

Figure 7 Net sales performance dashboard

The net sales performance dashboard is used to visualize the net sales of the supermarket as shown in figure 7. Based on dashboard, the net sales, gross sales and total orders are calculated using measures in PowerBI. These informations are important in understanding the sales performance of the supermarket. At the right corner of dashboard, there is a city filter for the audience to filter the visualization by different cities. Several informations like net sales, gross sales and total sales are using cards to present in figure 7. Based on figure 7, these are the total net sales of $307.59k, the total of $15.3k of gross sales and 1000 total orders in supermarket.

In addition, a map is used to visualize the supermarket branch at different cities such as Yangon, Naypyitaw and Mandalay. Next, a pie chart is used to visualize the net sales by different branches. Based on pie chart, branch C at Naypyitaw city as highest net sales with 34.24%, while branch A and B have same 32.88% of net sales at Yangon and Mandalay city. Besides that, clustered column chart is used to visualize net sales by month. From this column chart, January has highest net sale of $111k, while February has the lowest net sales of $93k. Furthermore, a treemap is utilized to show top 5 of net sales by different products sold in supermarket According to the treemap analysis, the product category of food and beverages has generated the highest net sales of $53.47k. Furthermore, this product category is known to be popular among supermarket shoppers. On the other hand, an area chart is used to visualize the net sales by different time intervals. Based on the area chart analysis, there has been a significant increase in net sales during the time period between 6pm and 7pm. Specifically, there has been a remarkable growth of 52.52% in net sales during this hour. The total net sales generated during this time period amounts to $37.8k.

Dashboard 2: Customer Information dashboard

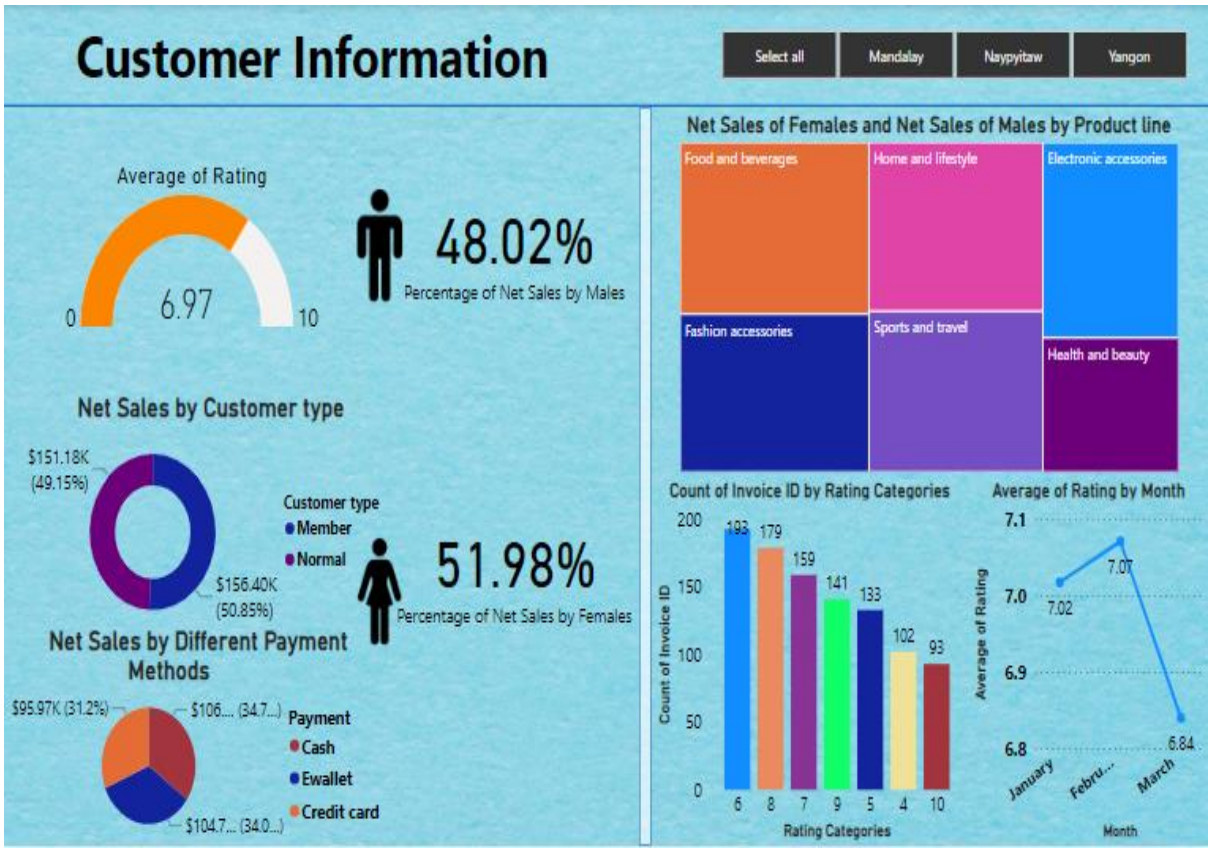

Figure 8 Customer information dashboard

The customer information dashboard is used to visualize the insights into customer preferences, behaviour and shopping habits in figure 8. The audience can filter the data by city name they wish to know. According to figure 8, the gauge used to represent the average rating of the supermarket from customers shows a value of 6.97. Additionally, the percentage of net sales attributed to females is higher than the percentage of net sales attributed to males. Specifically, the net sales by females for 51.98% of the total, while net sales by males represent 48.02% of the total. Based on the donut chart analysis, the net sales generated by members surpass those from non-members, amounting to $156.40k. Consequently, members accounted for 50.85% of the total net sales. The donut chart provides a visualization of net sales by various payment methods. It reveals that cash transactions recorded the highest sales value, amounting to $106.86k. Ewallet transactions ranked second with $104.76k, while credit card transactions generated $95.97k in sales. Consequently, cash transactions accounted for 34.74% of the total net sales. Moreover, a treemap is employed to illustrate the net sales of females and males across different product categories in the supermarket. According to the treemap analysis, the

product category of food and beverages has generated the highest net sales for both females and males. Specifically, females contributed a total net sales value of $31,591.35, while males contributed $21,879.93 to the net sales of food and beverages. Additionally, a stacked column chart is utilized to depict the count of invoice IDs across different rating categories. The chart reveals that the rating category of 6 has the highest count of invoice IDs, surpassing even the rating category of 10, which has the lowest count of invoice IDs at 93. Lastly, a line chart is employed to represent the average rating by month. The chart indicates that February has the highest average rating, reaching 7.07 out of 10. On the other hand, March has the lowest average rating, standing at 6.84.

Dashboard 3: Sales Influencers dashboard

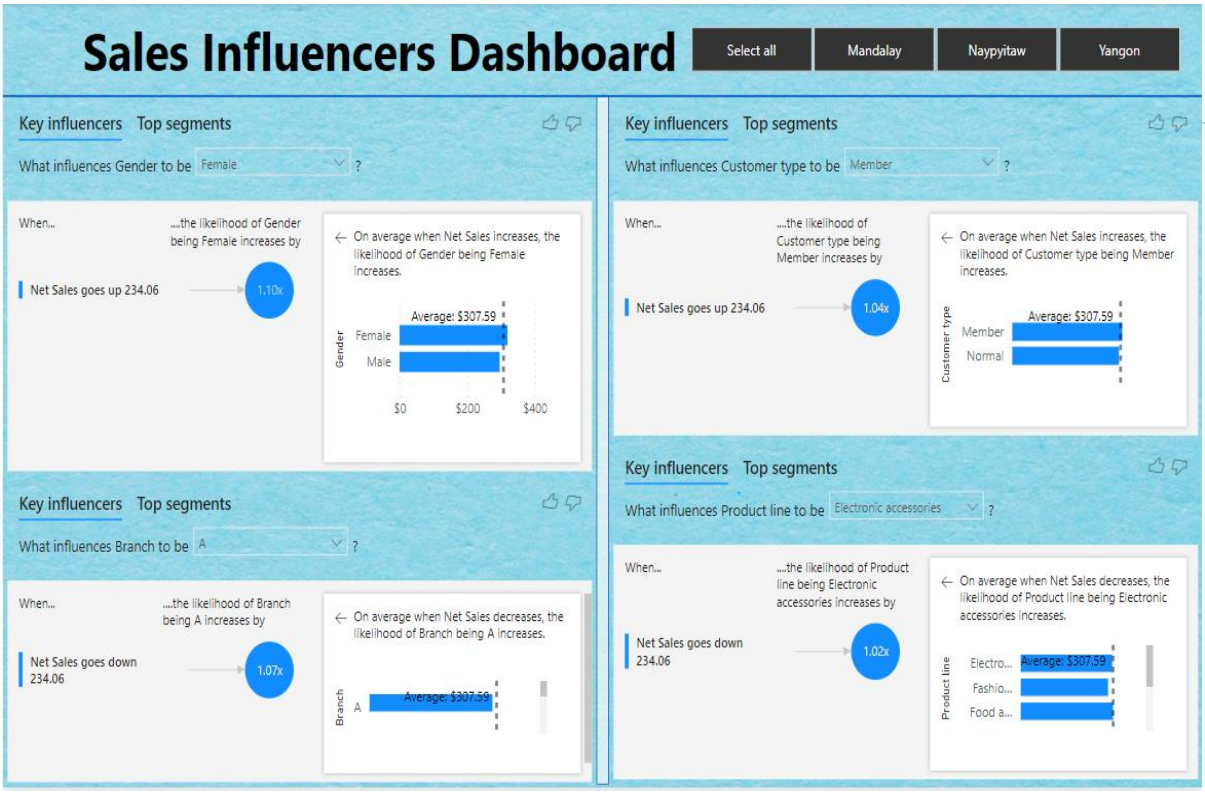

Figure 9 Sales influencers dashboard

In the context of supermarket sales, the Key Influencers tool, depicted in Figure 9, is utilized to identify the factors that exert the most significant influence on a specific metric. The primary goal of employing the Key Influencers tool in Power BI is to provide insights into the variables that have a substantial impact on sales. By using this tool, the audience is able to comprehend the relationship between different variables and how they contribute to overall

sales. In this project, the focus is on investigating various variables illustrated in Figure 9 to determine their impact on sales. The objective is to assist stakeholders in gaining a better understanding of the factors that influence sales in the supermarket setting. Based on the Key Influencers visual, the analysis reveals several insights regarding the impact of net sales on different variables. Firstly, the likelihood of the gender being female increases by 1.1 times when net sales rise, with an average net sales value of $307.59k for both genders. Secondly, the customer type being a member shows a 1.04 times increase in likelihood with rising net sales, and the average net sales for both member and non-member customers is $307.59k. Thirdly, the branch being A demonstrates a 1.07 times higher likelihood when net sales increase, and the average net sales across all branches is $307.59k. Lastly, the product line being electronic accessories exhibits a 1.02 times higher likelihood when net sales increase, and the average net sales for all product lines is $307.59k. These findings highlight the relationships between net sales and these variables, shedding light on their potential impact on sales.

Dashboard 4: Predictive dashboard

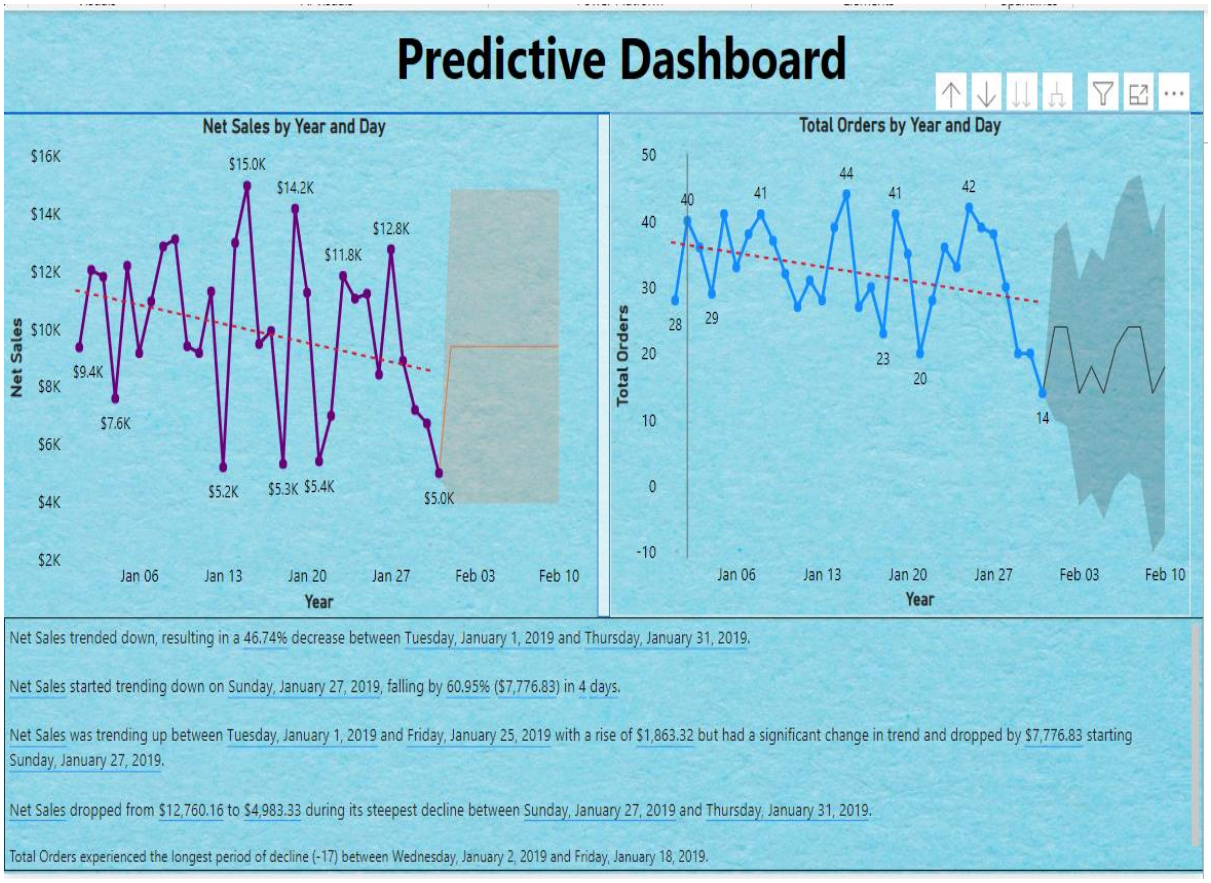

Figure 10 Predictive dashboard

The predictive dashboard for the supermarket focuses on analyzing net sales and total orders over time, both by year and day. The first line chart in the dashboard utilizes the forecast visualization in Power BI, presenting a predictive time series chart. The historical data is represented by a purple line, exhibiting some seasonal patterns. The line appears jagged, with fluctuations from day to day. At the latest historical data point, the line transitions to an orange colour, indicating future forecasted values. The light gray area surrounding the line represents the 95% confidence range. A wider confidence interval suggests a broader range of predicted values. Based on this analysis, the forecasted value for the period between February 1st and February 10th is $9389.41. Lastly, the trend line demonstrates a decreasing net sales trend from January 1st to January 30th.

In addition to that, the second line chart in the dashboard also employs the forecast visualization in Power BI to present a predictive time series chart. The chart represents the historical data with a purple line, showing some seasonal patterns. The line appears jagged, reflecting day-to-day fluctuations. At the latest historical data point, the line transitions to black color to indicate future forecasted values. The light gray area surrounding the line represents the 95% confidence range, with a wider interval implying a broader range of predicted values. Based on this analysis, the forecasted average value for total orders is 24 orders. In conclusion, the trend line in this chart reveals a decreasing trend in total orders from January 1st to January 30th.

Summary of Dashboards

No.	Dashboard	Description
1	**Net Sales Performance**	This dashboard illustrates the net sales by city, branch, month, product line and time intervals. This dashboard provides a visual representation of the net sales, gross sales and total orders of a supermarket.
2	**Customer Information**	This dashboard indicates the relationship between net sales with several attributes such as customer type, payment methods, gender, product categories and rating.
3	**Sales Influencers**	This dashboard utilized the key influencers tool in power BI to investigate the factors that causing the increase of sales.
4	**Predictive**	This dashboard is used to visualize the prediction of net sales and total orders by year and month. The forecast function is applied to predict the future values of net sales and total orders.

Table 1 Summary of dashboard

Suggestions for Stakeholders

According to this study, there are a few suggestions that can be made to relevant stakeholder like sales and marketing team, operations team, finance team and customer service team. First of all, the decreasing trend in net sales from January 1st to January 30th indicates a need for targeted marketing efforts to boost sales. It would be beneficial for the sales and marketing team to analyze the factors contributing to this decline and devise strategies to reverse the trend. Therefore, they could consider running promotions, introducing new products, or conducting customer surveys to identify areas for improvement.

Furthermore, operations team should utilize the valuable insights like forecasted values for net sales and total orders. They can use this information to optimize inventory management, staffing levels, and resource allocation. For example, if the forecast indicates a surge in net sales in February, the operations team can ensure sufficient stock is available and schedule additional staff during peak periods.

Moreover, finance team are used the forecasted net sales and total orders can assist the finance team in budgeting and financial planning. They can align their projections and allocate resources accordingly to support the expected sales volumes. Additionally, the confidence intervals around the forecasts provide an understanding of the range of potential outcomes, allowing the finance team to make informed decisions regarding investments, expenditures, and revenue targets.

In addition, customer service team can utilize the analysis of total orders can be valuable for the customer service team. They can use the forecasted average value of 24 orders to anticipate workload and ensure appropriate staffing levels to handle customer inquiries, resolve issues, and provide a positive customer experience.

Conclusion

In conclusion, the analysis of the data provides valuable insights for relevant stakeholders in the supermarket. The dashboards reveal important trends and patterns related to net sales, rating, customer type, payment methods, branch, product categories, and total orders over time. Stakeholders such as the sales and marketing team, operations team, finance team, and customer service team can benefit from these insights. Suggestions include implementing targeted marketing strategies to reverse declining net sales, optimizing operations based on forecasted values, aligning financial planning with sales projections, and ensuring excellent customer service by anticipating workload. By leveraging these suggestions and acting upon the findings, stakeholders can make informed decisions and take proactive measures to drive sales growth, improve operational efficiency, and enhance customer satisfaction in the supermarket.