UNIVERSITI UTARA MALAYSIA
SCHOOL OF COMPUTING
COLLEGE ARTS AND SCIENCE

STTHK2113 DATA ANALYTICS
GROUP B
SEMESTER A242

GROUP ASSIGNMENT 1 & 2

PREPARED BY:

NO.	NAME	MATRIC NUMBER
1.	TAN HOU REN	301235
2.	LIM WEI	300827
3.	DANIEL JAKSON	301153
4.	CHONNIKARN CHIRA A/P SOM PHONG	301206
5.	NUR FATINI BINTI MAHAMAD RAZALI	301193

PREPARED FOR:
TS. DR. MOHAMED ALI SAIP

DATE OF SUBMISSION:
15 JUNE 2025

Table of Contents

Introduction to the Chosen Business

Business Name: Yummy Fresh Rice Combo King
Location: Varsity Mall, Universiti Utara Malaysia
Type: Chinese food store specializing in rice combo meals with various sauces.

Menu Highlights with Code:

I. Rice:
- Salt Egg Butter Milk Chicken [1A]
- Salt Egg Butter Milk Fish [1B]
- Salt Egg Mala Chicken [1MA]
- Salt Egg Mala Fish [1MB]
- Sweet & Sour Chicken [2A]
- Sweet & Sour Fish [2B]
- Jin Xiang Chicken [3A]
- Jin Xiang Fish [3B]
- Gong Bao Chicken [4A]
- Gong Bao Fish [4B]
- Thai Patpet Chicken [5A]
- Thai Patpet Fish [5B]
- Si Chuan Mala Chicken [6A]
- Si Chuan Mala Fish [6B]
- Honey Chicken [7A]
- Honey Fish [7B]
- Black Pepper Chicken [8A]
- Black Pepper Fish [8B]
- Teriyaki Sauce Smoked Duck [9]
- Salt Egg Mala Sauce Smoked Duck [9M]
- Black Pepper Sauce Smoked Duck [9B]

The average price for the chicken menu ranges from RM7 to RM8.50, while the fish menu is priced between RM8 and RM9.50. Smoked duck is available at RM9 each.

II. Noodles:
- Cantonese Yuen Yang [13] and Cantonese Yee Mee [14]
- RM6.5 each

III. Snacks:
- Fried Chicken Wantan [11], Chicken Roll [11A], Fish Ball [11B] and Golden Fish Roll [11C]
- RM5 each

Why we choose this shop:
We selected Yummy Fresh Rice Combo King for our business and data analysis task because of its location within the UUM campus, the wide selection of budget menus, and its popularity among students from various residential colleges (INASIS). The business has regular day-to-day operations and a high volume of student customers, which is a good starting point in observing actual customer behaviour, sales trends, and business performance. Also, the cooperative nature of the owner ensured it was an ideal choice to follow up with quantitative and qualitative research.

Introduction to the Data Collected

i) Sales Data

The dataset contains detailed sales transactions for YUMMY FRESH RICE COMBO KING VMALL on December 2024, with a total of 6,798 records.

Important columns included in this dataset:

- CODE and DESCRIPTION: Identifying each menu item.
- QTY and UNIT PRICE: Quantifying and pricing individual sales.
- DATE and TIME: Timestamp of each transaction.
- TYPE: Classify the order as dine in or take out.
- PAYMENT METHOD: Indicating how the customer paid (Cash or QR Pay).
- TOTAL AMOUNT: Calculated revenue per transaction.

This dataset provides a holistic picture of customer purchasing behaviours and associated activity. Based on the dataset, we can analyse sales trends, menu items selling popularity, revenue performance, and payment methods used. The data collected offers some insight into the hours of peak business and consumer buying options at different times of day. This allows for the development of dashboard visualization, identification of key findings and assessment of the business overall performance.

Both the original dataset (YummyFresh_OriData.csv) and the processed dataset (YummyFresh_ProcessedData.xlsx) are included in the attached submission files.

ii) Customer Feedback Google Form

In addition, a short customer feedback survey was carried out to gather supplementary information on user background, customer satisfaction, and suggestions for improvement.

The data was collected through a Google Form targeting customers of Yummy Fresh Rice Combo King. Although the responses are limited, it offers some initial insights and should be considered a helpful guide toward improving the business.

The collected responses are provided in the attached CSV file (YummyFresh_FeedbackResponses.xlsx), and the original Google Form link is included below:
Google Form Link

Data Collection Activities

Purpose of Data Collection:
To analyse customer behaviour, sales trends, and tastes to facilitate fact-based decision-making as well as maximize marketing efforts.

First visit: 29 April 2025

- We directly approached the owner of Yummy Fresh Rice Combo King for permission to gather business data for our study. The goal was made explicit, with a focus on how all collected data would be used exclusively for academic research and politely shared with the company for both parties' benefit.


Figures 1 and 2: Our team engaging in conversation with the owner of Yummy Fresh Rice Combo King during the data collection request.

Second visit: 11 May 2025

- We went inside during this session to watch how customers behaved, when the restaurant was busiest, and how popular the menu was. The objective was to examine activities in real time and gather information that would subsequently be utilized to offer data-driven suggestions for enhancing the business's performance.

Figures 3, 4, and 5: Menu and customer queuing scenes at Yummy Fresh Rice Combo King.

Third visit: 18 May 2025

- We conducted a customer feedback survey and collected 20 responses to gain additional insights. In addition to the survey, we also held a structured interview with the owner, focusing on several key topics:
    - Customer demographics (ethnicity, gender)
    - Peak hours each day and each week
    - Top-selling items
    - Trends in sales performance
    - Primary sources of clients
- We gained deeper qualitative insights from this session that bolstered our data analysis and provided a more comprehensive understanding of customer behaviour and business concerns.

Figure 6: Group photo with the owner of Yummy Fresh Rice Combo King after the data collection session.

Customer Observation Summary:

- **Gender:** Observations indicate that higher percentage of female customers
- **Ethnicity:** Around 70 percent of the customers are Chinese, with the remaining 30 percent being people from other ethnic groups such as Indian and Malay.
- **Peak Hours of Operation:**
  Day: Saturdays are the busiest days, mostly because of on-campus co-curricular activities (KOK).
  Time: From 5:00 PM to 8:30 PM, the business has its highest volume of customers.
- **Trends in Demographics by Time:**
  Around 6:00 PM, the largest percentage of customers are of Chinese ethnicity. However, after 8:00 p.m., the customer base diversifies, with an increase in the number of Indian and other ethnic group customers.

Sales Trends:

- Sales typically increase during holidays and weekends when more students are free and available to visit.
- It was also observed that sales were better in the last semester compared to the current one.

Popular Items:

- Rice with Salt Egg Butter Milk Sauces
- Rice with Salt Egg Mala Sauces

Customer Sources:

The highest number of customers are students residing in INASIS SME Bank and INASIS Bank Muamalat, also the majority of the customers are student residents within UUM (on-campus student

Data Preprocessing Steps

Load the dataset from the excel file.

```
7    df= pd.read_csv('sales_data.csv')
```

Check the structure of data by using first 5 rows of data.

```
11    print(df.head(5))
```

Check column data types and basic info.

```
15    print(df.info())
```

Data Cleaning with imputation

The columns TYPE and CODE are found with the missing values.
Fills missing values in the TYPE column with the most frequent value (mode).
Fill in missing CODE values by referencing the DESCRIPTION field.

```
19  print('Missing values: ')
20  print(df.isnull().sum())
21
22  df['TYPE']= df['TYPE'].fillna(df['TYPE'].mode()[0])
23
24  # 1. Create a mapping from DESCRIPTION to CODE NUMBER (only from non-missing rows)
25  code_map = df.dropna(subset=['CODE']).drop_duplicates(subset=['DESCRIPTION'])[['DESCRIPTION', 'CODE']]
26  code_map = dict(zip(code_map['DESCRIPTION'], code_map['CODE']))
27
28  # 2. Use the map to fill missing CODE NUMBERs
29  df['CODE'] = df.apply(
30      lambda row: code_map[row['DESCRIPTION']] if pd.isnull(row['CODE']) and row['DESCRIPTION'] in code_map else row['CODE'],
31      axis=1
32  )
```

Data Reduction (Feature Selection)

Drop the DESCRIPTION column because it is irrelevant to analysis to reduce the size of data needed to be processed and improve analysis efficiency.

```
29    df=df.drop(columns=['DESCRIPTION'])
```

Data Transformation

Convert DATE to datetime format.

```
23    df['DATE'] = pd.to_datetime(df['DATE'], dayfirst=True)
```

Standardize TIME format to HH:MM:SS.

```
27  df['TIME'] = df['TIME'].apply(lambda x: x if len(x.strip().split(':')) == 3 else x + ':00')
```

Extract day of the week.

```
33    df['DAY'] = df['DATE'].dt.day_name()
```

Calculate the total sales amount per day AND Plot boxplot of daily total sales.

```
39  daily_sales = df.groupby('DATE')['TOTAL AMOUNT'].sum()
40
41  # Plot a single boxplot of 31 daily totals
42  plt.figure(figsize=(6, 6))
43  sns.boxplot(y=daily_sales)
44  plt.title('Boxplot of TOTAL DAILY AMOUNT (31 Days)')
45  plt.ylabel('Total Daily Sales Amount')
46  plt.tight_layout()
47  plt.show()
```

Presentation and Interpretation of Descriptive Analysis Findings

```
Daily Sales Statistics:
Mean: 2126.83
Median: 2011.00
Mode: RM1000–RM2000
Standard Deviation: 581.69
Range: 2183.00
```

The store earns around RM2127 a day, and RM1000 to RM2000 is the sales range experienced on most days. Daily sales can be very different from each other, with some days performing much better or worse. The difference between the highest and lowest day in sales is RM2183. Overall, the store's daily revenues are not consistent and tend to change depending on weekend or peak hour.

```
Summary Statistics for Transaction Totals by Customers
Mean: 15.84
Median: 16.00
Mode: 8.00
Standard Deviation: 6.79
Variance: 46.17
Range: 41.00
```

Customers spend an average of RM15.84 per order and most orders are approximately RM16, so spending is fairly consistent. The most common amount spent is RM8, likely from a top-selling product. Customer expenditure habits are fairly consistent overall, with most transactions being in an expected price range.

```
Frequency of TYPE (per transaction):
 TYPE
Dine In     1834
Delivery    1791
Name: count, dtype: int64
```

The business receives a comparable number of dine-in and delivery orders, which means that the two services are as popular among customers. While the number of dine-in orders is slightly more frequent, the variation is minimal, which means that the business is efficiently satisfying on-site and off-site customers.

```
Frequency of PAYMENT METHOD (per transaction):
 PAYMENT METHOD
QR Pay    2747
Cash       878
```

Most orders use QR Pay, indicating a strong customer acceptance of cashless payment over cash. This may be due to the fact that the store has a technology-savvy or a young customer base who are comfortable with cashless options.

```
Cross-tabulation: TYPE vs PAYMENT METHOD (per transaction)
 PAYMENT METHOD  Cash  QR Pay
TYPE
Delivery           0    1791
Dine In          878     956
```

From the output all delivery transactions were paid through QR Pay, implying that the system only accommodates electronic payments for deliveries. For orders that are dine-in, both Cash and QR Pay were used by customers, although QR Pay was used a bit more frequently. Because Cash was never applied to deliveries, there is no doubt that Cash payments are dine-in only, and this supports the assumption that delivery orders must be paid with QR-based payment.

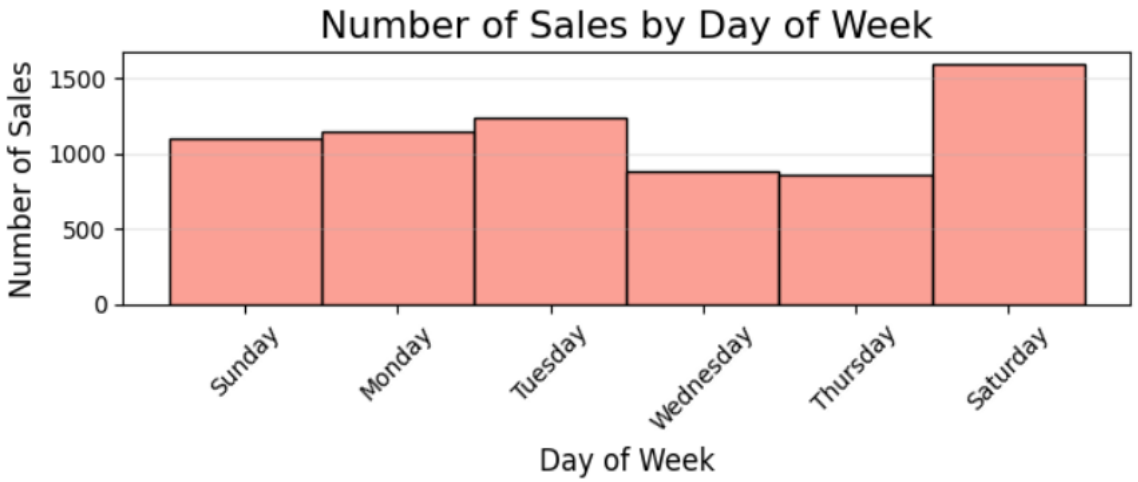

Diagram 1: A histogram for number of sales by day of week

From the histogram:

- **Saturday** recorded the highest number of sales, with over 1500 sales transactions-peak sale days.
- Moderate sales from **Sunday to Tuesday** with around 1000-1200 sales.
- **Wednesday and Thursday** tend to show minimum sales that fall below 1000.

Insight: Saturday is important for business operations, while promotional strategies are needed for midweek days to drive traffic.

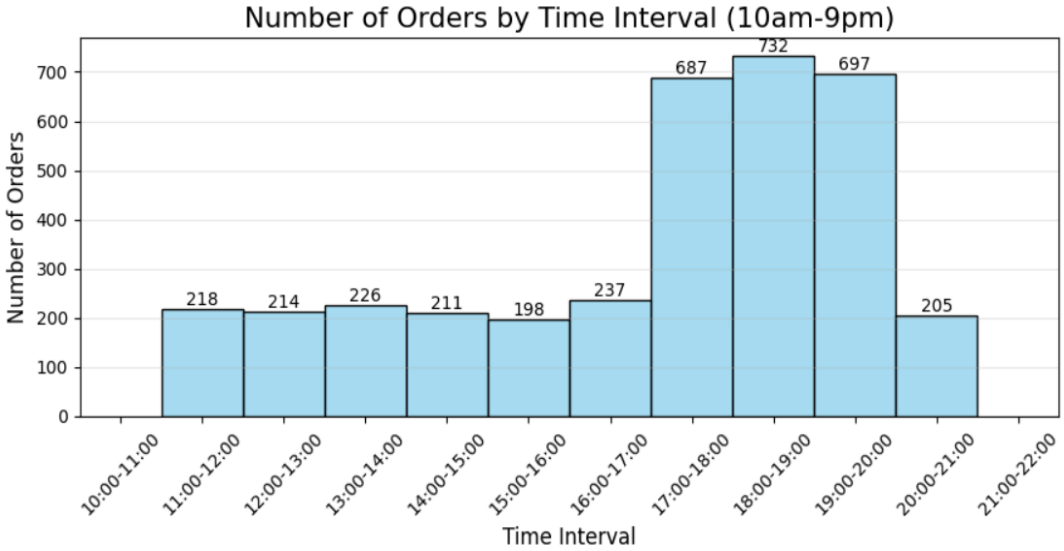

Diagram 2: A histogram of number of sales between the time interval (10am to 9pm only)

From the histogram:

- The histogram clearly shows a surge in orders **after 5 PM**, peaking at 6–7 PM.
- The highest number of orders occurred from **18:00–19:00**, with 732 orders.
- Closely followed by **19:00–20:00** (697 orders) and **17:00–18:00** (687 orders).
- This means that dinner time is the busiest time of the day for the business.

Insight: More staff and resources should be allocated in the evening hours to be able to accommodate peak demand.

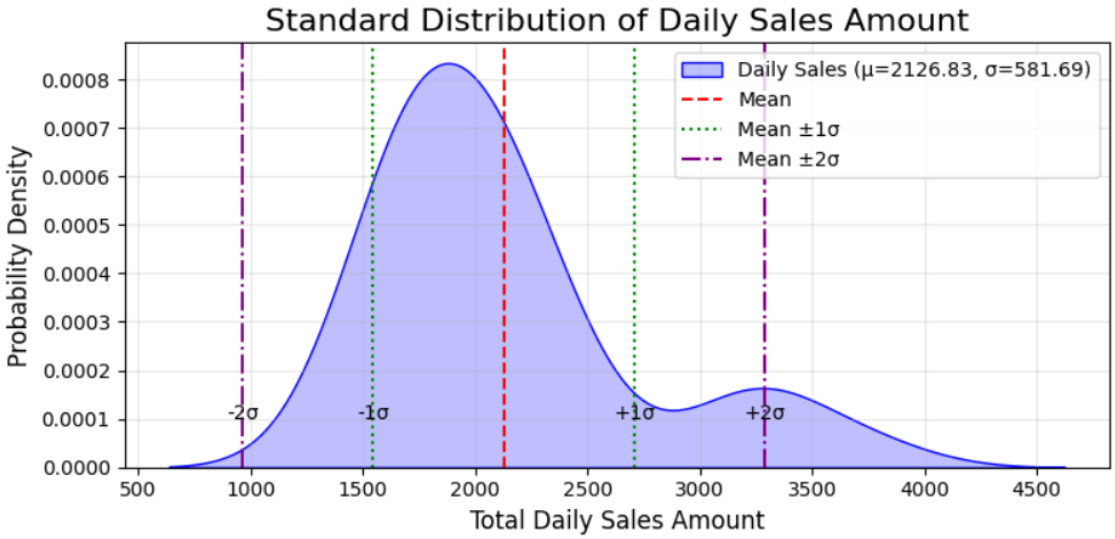

Diagram 3: A standard normal distribution graph for the total sales in day

The normal distribution graph clearly illustrates:

- The central tendency (mean) which is 2126.83 and spread (standard deviation) of daily sales which is 582.69.
- Symmetry in sales distribution, with most data points clustered near the mean.
- Vertical lines marking ±1σ and ±2σ, highlighting the range of typical and outlier sales days.

Insight: The business operates with consistent daily sales centered around 2126.83, but variability exists. However, there is some volatility in sales at certain times. Low sale outliers can be reduced while, maximum sales outliers can be used to make profits for the revenue optimization of the sales pattern. For example, investigates outlier days concerning external factors like weather and holidays to further plan for them in the future.

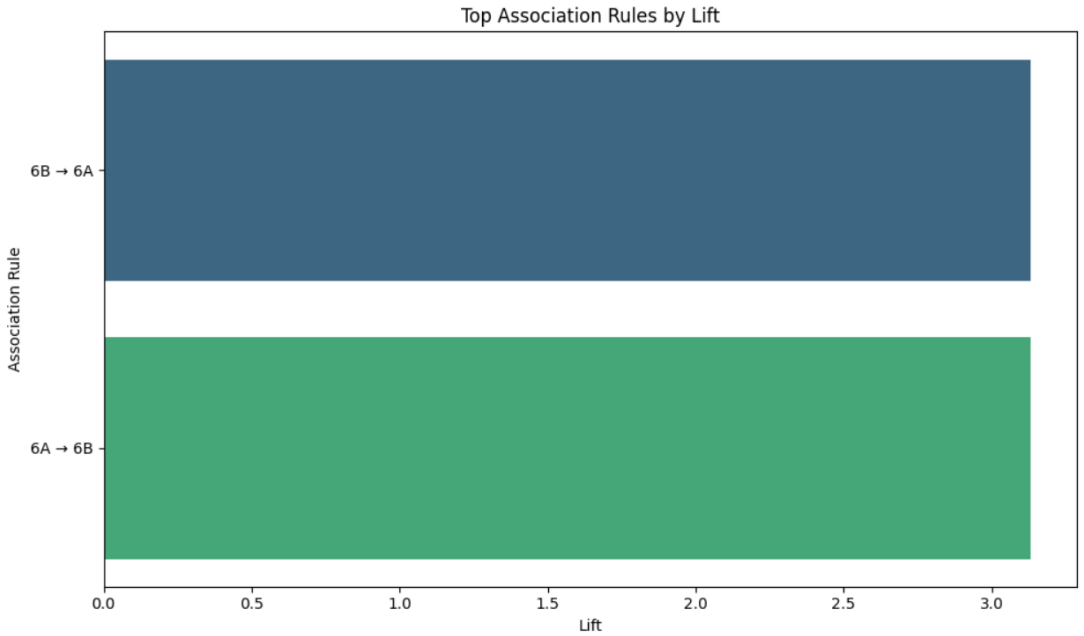

Diagram 4: An association rule analysis for the item

From the diagram:

- The association rule analysis reveals that 6A (Sichuan Mala Chicken) and 6B (Sichuan Mala Fish) are frequently purchased together with a lift value > 3, indicating a strong association between these two spicy dishes
- The high lift indicates that customers look for this spicy flavor pairing actively, thereby making it a possible signature combo.

Data Dashboard

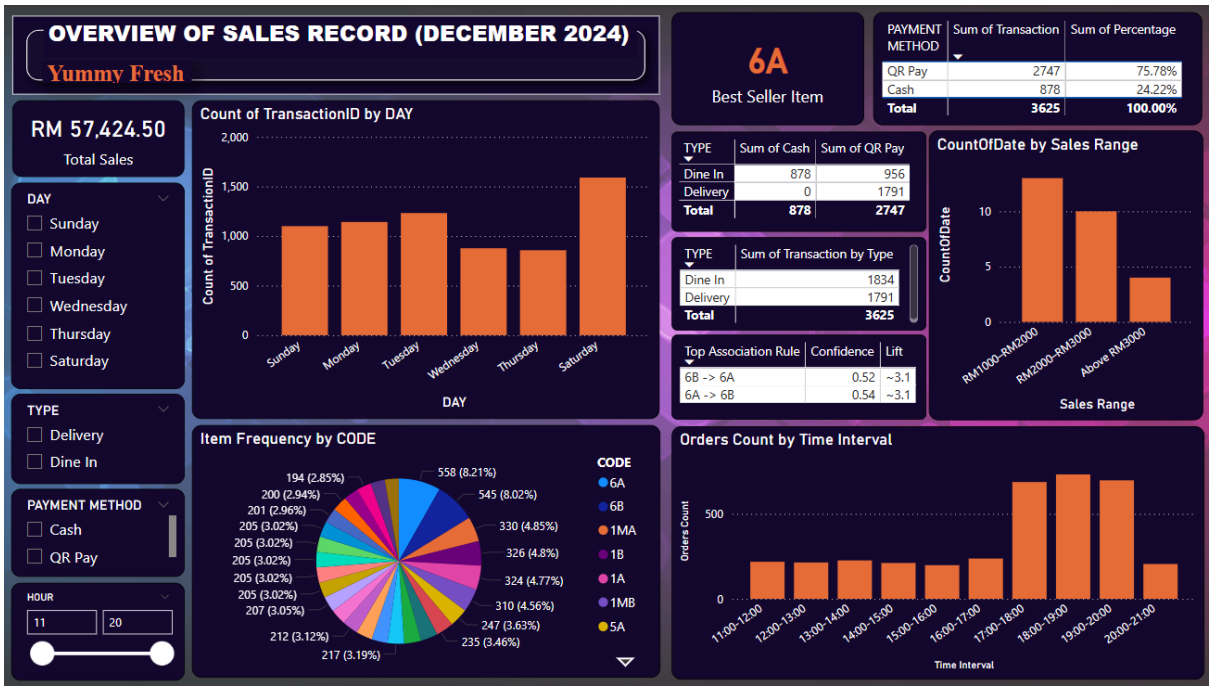

Figure 7: Data Dashboard

Figure 7 displays a dashboard visualizing Yummy Fresh's sales performance for December 2024. The interactive version of this dashboard can be accessed via the attached Power BI file, YummyFresh_Dashboard.pbix, submitted together with this report.

Insights Gained and Potential Implications

1. **Observation:**
   Based on the horizontal bar chart, the association rules show that 6A(Si Chuan Mala Sauce - Chicken) and 6B(Si Chuan Mala Sauce - Fish) are frequently bought together, with lift values greater than 3. This indicates a strong relationship between the two items which customers often purchase as a pair.

   Business Insight:
   We recommend combining items 6A and 6B into a "Best Seller Combo" or offering them as a discounted set. This can encourage faster ordering during peak hours and increase average order value through upselling. It can also improve customer experience by reducing decision time.

2. **Observation:**
   Based on the histogram of Number of Orders by Time Interval, it shows that peak periods are
   - 17:00–18:00 (687 orders)
   - 18:00–19:00 (732 orders)
   - 19:00–20:00 (697 orders)

   These three consecutive time slots account for the majority of daily sales activity.

   - 6A (448 orders)
   - 6B (426 orders)
   - 1MA (220 orders)
   - 1A (214 orders)
   - 1B (213 orders)
   - 1MB (205 orders)

   These six dishes represent the highest sales between 17:00 and 20:00.

   Business Insight:
   To enhance operations during peak hours of 5:00 PM to 8:00 PM, increasing kitchen and counter staff would be an effective way to enhance service speed and quality. Additionally, if the six highest selling dishes were prepared in advance, there would be less wait time and improve speed of fulfilling orders to customers. Including quick combo sets with popular dishes may enhance the ordering process and good service. The owner may need to evaluate when she orders and prepares certain items in her inventory to ensure all inventory needed for high demand during busy hours is available.

3. **Observation:**
Based on the boxplot and normal distribution graph,
Mean daily sales = RM 2126.83
Standard deviation = RM 581.69
Most daily sales fall within RM 1000 - RM 2000

Outliers occur primarily on Saturdays, with sales exceeding RM 3500, suggesting weekend peak demand.

Business Insight:
Sales are stable across most days, but Saturdays consistently outperform the rest.
We recommend capitalizing on this trend by implementing weekend-only promotions or loyalty rewards to encourage even higher spending. Furthermore, consider extending operating hours slightly on Saturdays, and increasing inventory and staff coverage to fully accommodate the higher customer volume and maximize revenue.

4. **Observation:**
Based on the collected data as below:

Frequency of SERVICE TYPE:
Dine In =1834
Delivery =1791

Frequency of PAYMENT METHOD:
QR Pay = 2747
Cash = 878

SERVICE TYPE vs PAYMENT METHOD

SERVICE TYPE	Cash	QR Pay
Delivery	0	1791
Dine In	878	956

Business Insight:
This data reveals an interesting picture of customer payment preference. Customers clearly prefer QR payments (2747 out of 3625 transactions were QR). Dine-in customers preferred QR payments to cash (956 QR and 878 cash) and QR payments were used for 100% of delivery. It seems that QR payments have been adopted wholesale, especially in delivery. Businesses will need to continue to support QR payments, potentially with incentives to support the behaviour we see customers are exhibiting. Nevertheless, owners still want to offer both QR and cash payment options for dine-in transactions for the benefit and convenience of the customers. With only 878 transactions with cash, owners do not need to have a large amount of cash on hand, just enough change to provide when requested.

5. **Observation:**
   Based on the histogram of Number of Sales by Day of Week, it shows that Saturday consistently records the highest sales volume, followed by Tuesday, Monday, and Sunday. In contrast, Wednesday and Thursday show a noticeable drop in transactions, indicating a midweek slowdown in customer activity.

   Business Insight:
   To address the sales gap midweek, we recommend launching midweek-specific promotions such as "Combo Wednesday" or "Student Deals Thursday" to drive traffic on slower days. This can help balance the weekly revenue and make better use of operational resources on low-traffic days.

6. **Observation:**
   Based on the pie chart in the dashboard, it shows the lowest selling dishes in a week as below:
   - 3A (185 orders)
   - 5B (191 orders)
   - 9 (194 orders)
   - 7B (196 orders)
   - 11A (200 orders)

   Business Insight:
   Owners may modify the recipes of these lowest selling dishes, or change the items themselves to perhaps better appeal to customers. Second, promote these items by offering limited-time introductory specials, discounts, or bundle them in conjunction with other items. Other ideas to boost popularity would be to put them in a higher visibility area on the menu or print them as a daily special. Owners also should seek customer feedback which may help to explain the poor sales results of these dishes.

7. **Observation:**
   Based on the Google Form responses, all respondents are UUM students from a variety of ethnic backgrounds, with a balanced gender distribution, indicating a broad and diverse student customer base.

   Business Insight:
   The business should maintain a universal menu that appeals to all customer groups, while also introducing limited-time items or campaigns tailored to specific student segments to boost engagement. For example, offering ethnic-inspired seasonal specials like a "Rendang Special" during festive months can effectively attract more attention and drive sales.

8. **Observation:**
   Based on the data from the questionnaire (Google Form), the percentage of students come from different INASIS is shows as below.

   Laluan D : 40%
   Laluan B : 40%
   Laluan C : 15%
   Laluan A : 5%

Business Insight:
Laluan D has only two INASIS and is far from the V MALL, but it is taking a significant portion of the sales for the Yummy Fresh. This may be due to the lack of similar food options in the Laluan D. Additionally, owners could try to understand what drives their purchasing behaviour and perhaps it could help replicate the success in other routes like Laluan C and A, which currently contribute less to sales.

9. **Observation:**
Based on the Google Form responses, the majority of customers rated the food and service quality as “Excellent,” with very few ratings falling into the “Good” or lower categories.

Business Insight:
This indicates strong operational performance overall. However, it is essential to maintain this high standard during peak hours. To support ongoing quality monitoring and enable real-time issue tracking, consider implementing a simple feedback system at the counter, such as a QR code survey.

10. **Observation:**
Most suggestions (from open text of Google Form responses) focus on:
    - Portion sizes
    - Long queue times
    - The amount of sauce

Business Insight:
Although only mentioned by a few respondents, some useful suggestions emerged regarding sauce quantity, queue times, and portion size. One respondent noted that the rice could feel dry when too little sauce is given, highlighting an opportunity to slightly increase sauce portions or offer an optional "extra sauce" add-on. A comment about long queues during peak hours suggests the need to reduce waiting time. This could be addressed by adding an extra service counter or developing a simple mobile ordering app, allowing customers to pre-order their meals before arriving and view estimated wait times. A request for larger portion sizes suggests that offering a size upgrade for a small fee could add perceived value and increase customer satisfaction.

Limitation of Analysis

1. No Demographic Information (e.g., Gender, Race)
    - The dataset does not include any personal or demographic identifiers.
    - This limits the ability to analyse a customer group for preferences based on gender or culture.
    - Insights like “Which group prefers spicy options?” or “Do males spend more on combos?” cannot be explored.

2. No Accommodation or Inasis Data
    - Since the data do not store any information on student accommodation (inasis), it is not possible to:
      i) Examine purchasing habits by dorm or block
      ii) Analyse walk-in patterns vs. distance

3. No Inventory or Stock Data
    - Inventory levels and item availability are not tracked in the dataset.
    - Sales data can be used to estimate high-demand items. However, the absence of real-time inventory data limits the capabilities of:
      i) Identify items which have frequently gone out of stock
      ii) Keeping track of waste levels for unsold perishable items
      iii) Identify reorder points based on actual depletion of stock, not just based on simple sales velocity

4. No Data about Cost of Goods Sold (COGS)
    - There is no cost data related to the items in this data set (e.g. price of raw ingredients or preparation of each item) limits:
      i) Analysis of profit margins by menu item
      ii) Recognition of low margin or high waste items
      iii) Making informed decisions relating pricing strategy or menu improvement

5. Time Frame is Limited
    - The dataset only contains sales data from one month.
    - This limits:
      i) Understanding of seasonal trends
      ii) Planning for exam periods, holidays, and semester breaks
      iii) Evaluation of long-term growth or drop-offs

Presentation to the Business Owner

In the final stage, the key insights and data dashboards were presented to the owner of Yummy Fresh Rice Combo King. It was to translate analytics into executable business strategies. We pointed out peak hours, hot-selling products, payment preferences, and areas for improvement in operations, such as staff planning, portion ratios, and potential promotions for the middle of the week.

Figure 8: Dashboard presentation and key findings shared with the owner of Yummy Fresh Rice Combo King.

The dashboard served as a visual interface to allow the owner to track customer behaviour trends and opportunities for generating revenue, like bundling popular dishes or offering discounts on less popular ones. The owner responded positively and expressed interest in applying some of the findings for fast service, sales planning, and menu visibility. This collaborative approach helped to reinforce the value of data-driven decision-making by small businesses and showcased in a practical fashion the direct positive impact analytics can have on day-to-day business operations.

Figure 9: Dashboard interface shown to the owner during the presentation.

Group Member Contributions

Name	Contribution
Lim Wei	Led the overall project coordination. Monitored customers and recorded peak hours and demographic trends. Worked on data preprocessing and insights generation.
Tan Hou Ren	Initiated contact with the business owner and managed all communications for site visits. Worked on data preprocessing and generating business insights.
Daniel Jakson	Developed the data dashboard and contributed to the presentation and interpretation of descriptive analysis findings. Also assisted in preparing the data for analysis and validating results for accuracy.
Chonnikarn Chira A/P Som Phong	Compiled the report documentation. Developed the data dashboard and assisted in the presentation of descriptive analysis findings.
Nur Fatini binti Mahamad Razali	Created and disseminated the Google Form for customer feedback. Completed the Introduction and documented the Data Collection Activities section of the report.

Conclusion

In summary, this project has provided useful information about the operations, and customers behaviours at Yummy Fresh Rice Combo King. By sourcing, cleaning, and analysing the data, patterns were identified that can improve the quality of services offered in the business, and customer satisfaction. Notable results include QR-based payment preferences, peak demand during evening hours, frequently ordered mala or salted egg rice combos, and Saturday was a profitable day and might be where marketing and promotion efforts may be spent.

The observed patterns also identified operational aspects like long wait times for queuing customers, and inconsistent portioning of sauces. However, these aspects can be mitigated with changes to human resources, planning before delivery starts in peak times, and other better service experiences. Though the analysis offered quite reasonable recommendations, the mentioned limitations including the lack of demographic data, inventory tracking, and profit margins must be acknowledged. These diminish expansion and suggest areas for data collection efforts in the future.

This project highlights not just the value of data analytics, but how data analytics can lead to better decision-making. By taking the changes into consideration, and tracking data over time, Yummy Fresh Rice Combo King can improve efficiency, customer experience, and increase profitability.