BIT 2053

CITY UNIVERSITY MALAYSIA

Est. 1984

Final Project Modern Data

Presented By :

1. Kenneth Chaw Yi Jie
2. Alexandro Elvin Adrian
3. Mohamad Hazrul Ekhwan Bin Moruin @ Abdul Hamid

3 September, 2025

BIT 2053

Introduction

Objective :

-analyzing a real-world dataset using Business Intelligence (BI) tools.

-simulate a data-driven decision-making process

Dataset Chosen :

-"Retail Sales Data Dashboard" from Kaggle.com

-Access Link :

https://www.kaggle.com/datasets/josemiguelterron/retail-sales-data-dashboard

Business Questions Explanation :

Q1) The leadership should understand the seasonality to forecast the demand, manage inventory, and do planning for cash flow during peak periods.

Q2) The executive product team needs to identify the "star" products which means is hot selling to promote and "dog" products which much less buyer bought to discontinue or re-price it in order to maximize the profitability.

Q3) The marketing team needs to take responsibility to know which customer segments and regions that are the most profitable to allocate their budget effectively and tailor campaigns around the area.

BIT 2053

Business Understanding

Business Objective :

-To increase overall annual profit by 20% by optimizing product assortment, marketing spend, and discount strategies.

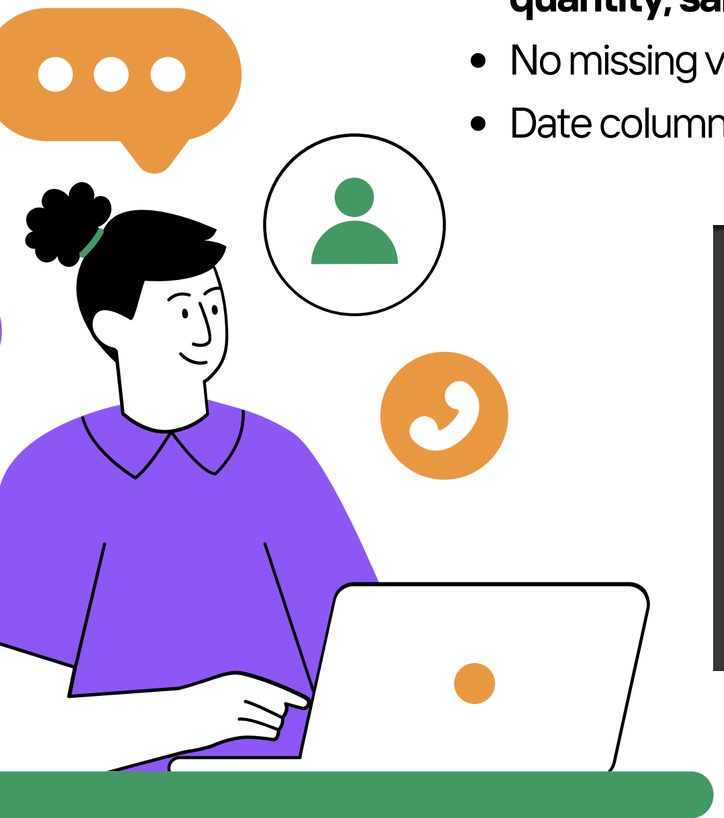
Stakeholders :

-CEO, Marketing Director, Sales Manager, Supply Chain Manager

Data Understanding

BIT 2053

- **Using Google Colab (Write the code)**
- **Dataset:** Retail Sales Data Dashboard (Kaggle)
- **1,000 transactions with:**
  - **i)** Transaction ID, Date, Customer ID, Gender, Age
  - **ii)** Product Category, Quantity, Price per Unit, Total Amount
- Mix of categorical **(e.g., gender, product category)** and numerical **(e.g., quantity, sales)** attributes
- No missing values or duplicates found
- Date column converted from text to datetime for time-series analysis



Choose Files | Copy of Ret... Project.csv

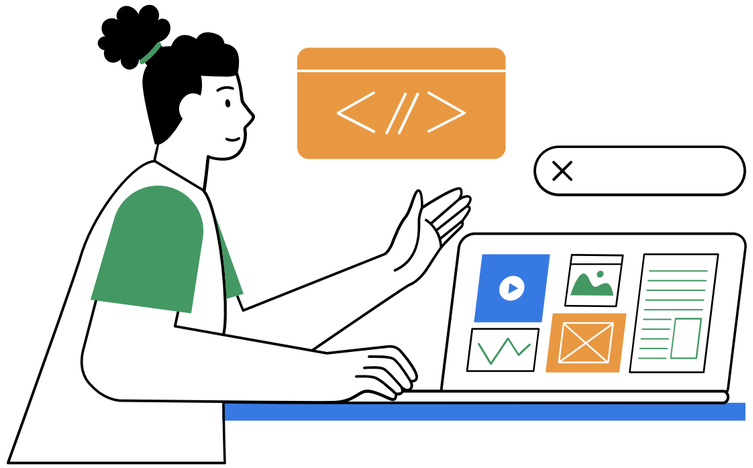
- **Copy of Retail Sales Data Project.csv**(text/csv) - 50605 bytes, last modified: 9/2/2025 - 100% done

```
Saving Copy of Retail Sales Data Project.csv to Copy of Retail Sales Data Project.csv
File uploaded: Copy of Retail Sales Data Project.csv
```

	Transaction ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount
0	1	11/24/2023	CUST001	Male	34	Beauty	3	50	150
1	2	2/27/2023	CUST002	Female	26	Clothing	2	500	1000
2	3	1/13/2023	CUST003	Male	50	Electronics	1	30	30
3	4	5/21/2023	CUST004	Male	37	Clothing	1	500	500
4	5	5/6/2023	CUST005	Male	30	Beauty	2	50	100

BIT 2053

Data Preparation

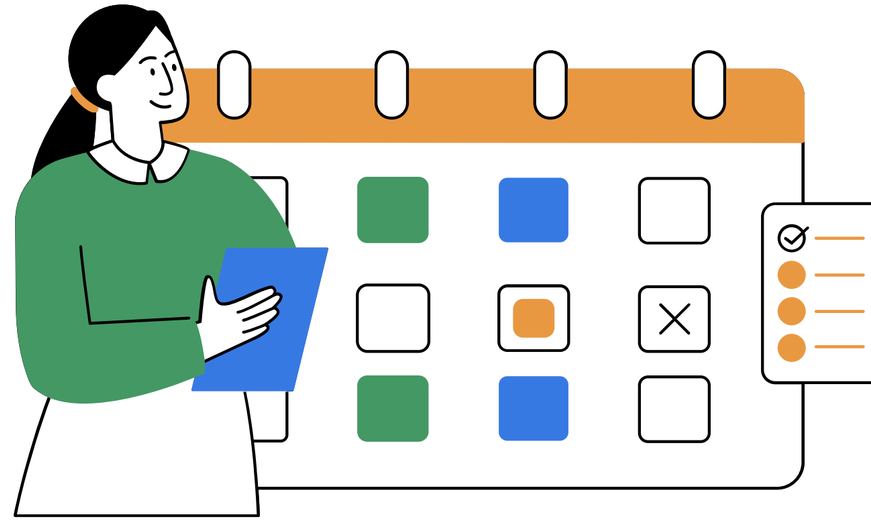


- **Column names standardized** → lowercase, snake_case
- **Date converted to datetime** → extracted: year, month, day, quarter, month_year
- **Numeric validation:** quantity, price_per_unit, total_amount → converted to numeric
- **Feature engineering:**
  I) sales_per_unit = total_amount ÷ quantity
- Final cleaned dataset saved as clean_retail_data.csv
- Ready for Python analysis & Looker Studio dashboards

	transaction_id	date	customer_id	gender	age	product_category	quantity	price_per_unit	total_amount	year	month	day	quarter	month_year	sales_per_unit
0	1	2023-11-24	CUST001	Male	34	Beauty	3	50	150	2023	11	24	2023Q4	2023-11	50.0
1	2	2023-02-27	CUST002	Female	26	Clothing	2	500	1000	2023	2	27	2023Q1	2023-02	500.0
2	3	2023-01-13	CUST003	Male	50	Electronics	1	30	30	2023	1	13	2023Q1	2023-01	30.0
3	4	2023-05-21	CUST004	Male	37	Clothing	1	500	500	2023	5	21	2023Q2	2023-05	500.0
4	5	2023-05-06	CUST005	Male	30	Beauty	2	50	100	2023	5	6	2023Q2	2023-05	50.0

BIT 2053

Data Analysis & Methods



- Applied descriptive analytics to extract business insights
- Created summary tables for later BI dashboards
- Analysis focused on overall, monthly, category, customer, and daily sales

Overall Summary (Figure 6)

- Total Sales
- Total Quantity Sold
- Number of Customers
- Number of Transactions
- Average Sales per Transaction

	total_rows	total_sales	total_quantity	num_customers	num_transactions	avg_sales_per_transaction
0	1000	456000	2514	1000	1000	456.0

Monthly & Category Analysis

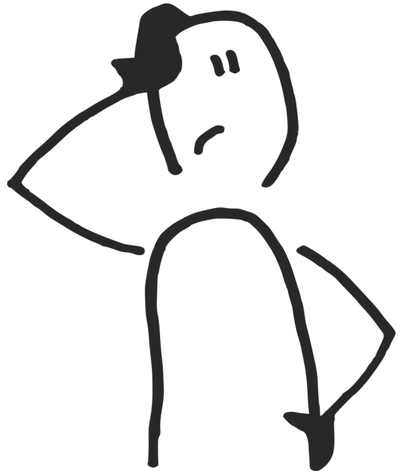
- **Monthly Summary (Figure 7)**
  - Trends in sales, quantities, and customer activity
  - Useful for detecting seasonality & sales growth

	month_year	total_sales	total_quantity	avg_sales	num_customers
0	2023-01	35450	195	466.447368	76
1	2023-02	44060	214	518.352941	85
2	2023-03	28990	194	397.123288	73
3	2023-04	33870	214	393.837209	86
4	2023-05	53150	259	506.190476	105

- **Category Summary (Figure 8)**
  - Top performing product categories
  - Underperforming categories identified for improvement

	product_category	total_sales	total_quantity	avg_price	num_customers
0	Beauty	143515	771	184.055375	307
1	Clothing	155580	894	174.287749	351
2	Electronics	156905	849	181.900585	342

Customer & Daily Analysis



Customer Summary (Figure 9)

- Total spending per customer
- Quantity purchased, avg. order value, transactions per customer
- Helps identify high-value customer

	customer_id	total_sales	total_quantity	avg_order_value	num_transactions
0	CUST001	150	3	150.0	1
1	CUST002	1000	2	1000.0	1
2	CUST003	30	1	30.0	1
3	CUST004	500	1	500.0	1
4	CUST005	100	2	100.0	1

Daily Summary (Figure 10)

- Day to day sales trends
- Detects short term fluctuations & anomalies

	date	total_sales	total_quantity	num_customers	num_transactions
0	2023-01-01	3600	10	3	3
1	2023-01-02	1765	10	4	4
2	2023-01-03	600	2	1	1
3	2023-01-04	1240	8	3	3
4	2023-01-05	1100	5	3	3

BIT 2053

Business Intelligence (BI) Dashboard

Key Performance Indicators (Top Row)

- Total Sales: 456.0K - representing total revenue
- Total Quantity: 2.5K - total number of items sold
- Average Sales per Transaction: 456.0 - indicating the average transaction value

Sales Trend Over Time (Left Chart)

This line graph tracks the total quantity sold from January 2024 back through 2023. Key observations:

- Seasonal Patterns: Shows significant fluctuations with peaks around September 2023 and May 2023
- Recent Performance: Sales appear to have stabilized around 200 units in the most recent period (January 2024)
- Volatility: The trend shows considerable variation, ranging from lows around 175 to peaks over 250 units

Revenue by Gender (Top Right Pie Chart)

- Female customers: 51.1% of total revenue
- Male customers: 48.9% of total revenue
- Gender Balance: Nearly even split with females contributing slightly more to overall revenue

BIT 2053

Business Intelligence (BI) Dashboard

Best-Selling Products (Bottom Left Bar Chart)

- Product performance by total quantity sold:
- Clothing: Leading category with approximately 800+ units
- Electronics: Close second with around 700+ units
- Beauty: Third place with approximately 600+ units

Revenue by Category (Bottom Center Treemap)

Visual representation of revenue distribution:

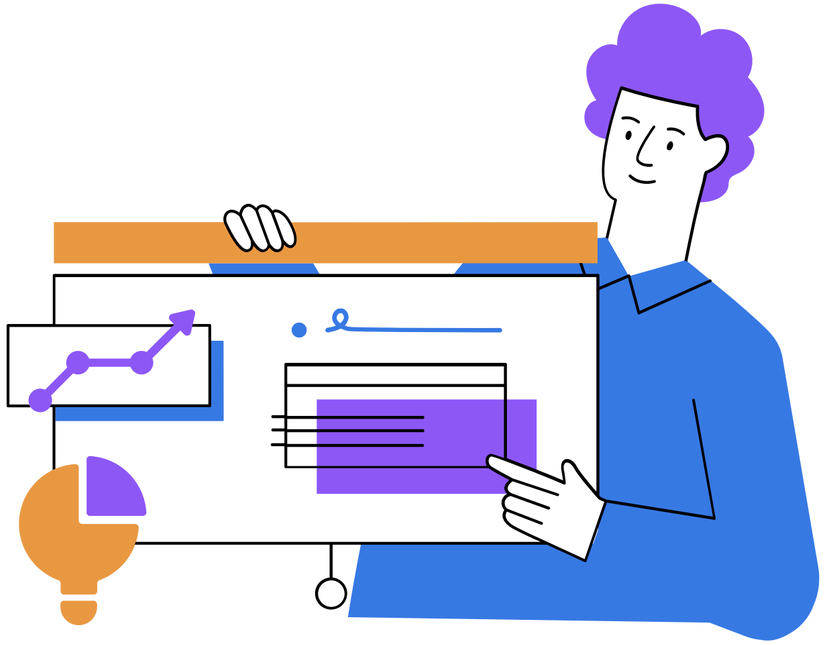
- Electronics: Appears to dominate revenue share despite lower quantity
- Clothing: Significant revenue contributor
- Beauty: Smaller revenue share compared to the other categories

Average Spend by Age Group (Bottom Right Bar Chart)

- Shows spending patterns across age demographics (19-53 years):
- Peak Spending: Age 37 shows the highest average spend at around 750
- Consistent Spending: Most age groups maintain spending levels between 600-700
- Age Range: Covers a broad demographic from young adults to middle-aged customers

BIT 2053

Findings & Insights



Business Performance

1. Strong Transaction Value: With an average of 456.0 per transaction, the business maintains healthy transaction sizes
2. Balanced Customer Base: The nearly 50-50 gender split indicates broad market appeal

Product Strategy Insights

1. Volume vs. Value Discrepancy: While Clothing leads in quantity sold, Electronics appears to generate more revenue per unit, suggesting higher-margin products
2. Category Opportunities: Beauty products show potential for growth, given lower quantities but established customer interest

Customer Demographics

1. Peak Spending Age: The 37-year age group represents the highest-value customer segment
2. Broad Appeal: Consistent spending across age groups from 19-53 suggests the product mix appeals to multiple generations

Recommendations

1.Boost High-Value Electronics

Increase marketing budget by 20% for Electronics; feature prominently online.

2. Target Key Customer Group

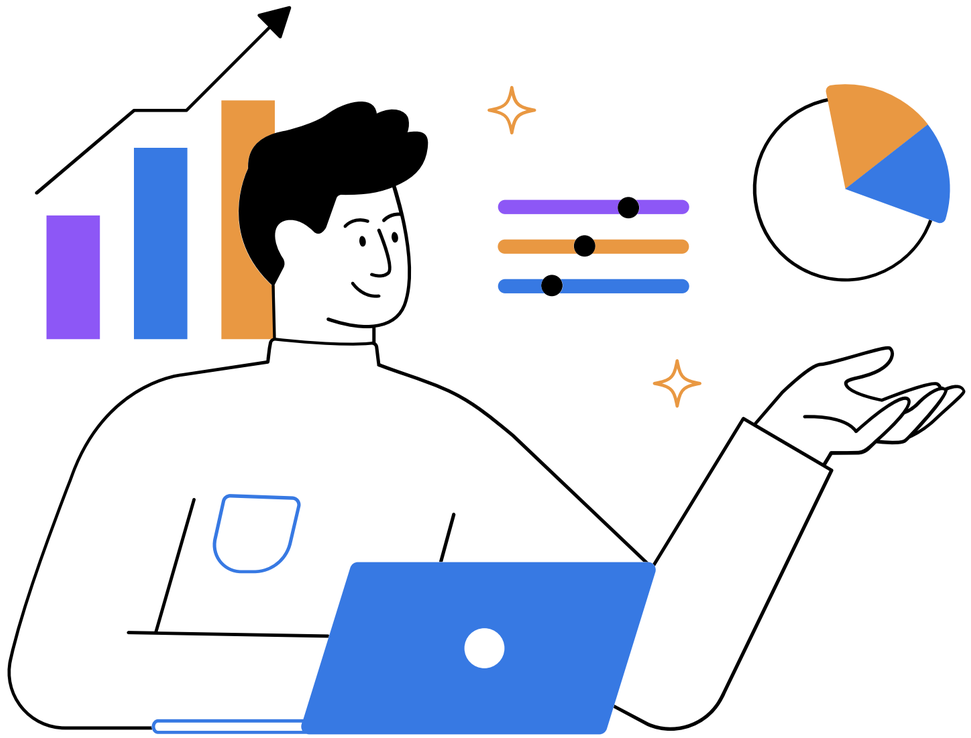
Launch personalized email & loyalty campaigns for customers aged 34-43.

3. Optimize Peak-Season Inventory

Conduct pre-peak inventory review in October; secure stock for best-sellers.

4. Revitalize Beauty Category

Bundle Beauty products with top-selling Electronics; investigate pricing.



Recommendations & Impact

Impacts:

1. Expected 15-20% revenue increase from most profitable segment.
2. Increase customer loyalty and lifetime value from the highest-spending demographic.
3. Prevent stock-outs, reduce emergency costs, and add +10% peak season revenue.
4. Drive 5-10% sales growth in Beauty and improve overall brand appeal.

Conclusion

This project demonstrates:

- BI analytics transform data into actionable strategies
- Electronics drives high-value revenue
- Beauty category offers growth potential
- 37-year-olds are highest-value demographic
- Multi-generational appeal exists across age groups

Recommendations provide clear roadmap for:

- Increased profitability
- Operational efficiency
- Competitive advantage

BIT 2053

Thank You All !