

**The Hong Kong Polytechnic University**  
**COMP3134 Business Intelligence and Customer Relationship**  
**Management**

**Group Project (15%)**

**Due date: 20<sup>th</sup> Nov., 2025, 23:59**

**Rules and Guidelines**

All project work must be original. If any part of your work uses AI tools such as ChatGPT or Bing AI, you must clearly state where and how AI was used; otherwise, it will be treated as plagiarism or academic dishonesty and **result in a zero** for this project. If you quote or refer to text, research, or findings from other sources, you must provide proper citations. For more information, please read:

<https://www.polyu.edu.hk/elc/-/media/department/elc/content/language-resources/referencing/apa.pdf?rev=1de722ffdaf44774a8dd05aa52fbab9d&hash=6EE51403C0DB8BCFC1CB6E2964251DA8>

Late submission penalty: deduct -1.5 pts per day

**Project Overview: Descriptive Analysis and Business Insights**

You are members of the Data Science Team at RetailX, a company operating several shopping malls. Management has provided you with recent sales and customer data and tasked you with performing a descriptive analysis. Your primary goal is to identify key customer groups and purchasing patterns to inform the next targeted marketing promotion two months later. For example, you can use predictive analytics to analyse customer purchasing data and suggest a strategy to select their product mix in the coming promotion campaign.

The following results are particularly useful for such planning:

- Regression / Classification (discover possible causal relationships among data).
- Clustering (discover similarities and differences among clusters).
- Associations (discover interesting relationships among data).

You are free to use any software or publicly available methods and packages to help with your predictive analytics.

## Project Requirements

In the project, your group is given datasets with the following fields

Sales data description:

Field	Description	Format or Possible values
Invoice no	Invoice identification number	INV{XXXXXX} where X are digits, e.g. INV00001
Customer id	Customer identification number	C{XXXXXX} where X are digits, e.g. C00001
Product id list	List of product identification number (separated by comma)	P{XXXXXX} where X are digits, e.g. P00001
Invoice date	Date of purchase	DD/MM/YYYY, e.g. 01/12/2025
Shopping mall	Shopping mall location (location code)	MK / TKO/ ST /CYB corresponding to “Mong Kok”, “Tsuen Kwun O”, “Shatin”, “Causeway Bay”

Customer data description:

Field	Description	
Customer id	Customer identification number	C{XXXXXX} where X are digits, e.g. C00001
Gender	Gender	M (Male) or F (Female)
Age	Customer age	Positive Integer
Payment method	Payment used by customer	'Credit Card', 'Cash' or 'Mobile Payment'

Product data description:

Field	Description	Format or Possible values
Product id	Product identification number	P{XXXXXX} where X are digits, e.g. P00001
Category	General item categorization groups	'Electronics', 'Clothing', 'Groceries', 'Books', or 'Toys'
Price	Price of the product	Floating point with 2 decimal digits (in HKD)

You will have to perform the following steps for your project:

**1. Downloading your Dataset**

- Go to Blackboard to download your group's specific dataset file package
- The package should contain 3 csv files :  
1. sales\_{grp id}.csv, 2. customers\_{grp id}.csv, 3. products\_{grp id}.csv  
where "grp id" is your group ID.

**2. Discovery of Data Characteristics and Analysis**

- Perform a preliminary understanding of the data characteristics with descriptive statistics & visualization and identify key trends.

**3. Prediction and Visualization:**

- Select one of the data mining techniques that helps to obtain critical insight for the coming promotion campaign (e.g., Regression, classification, clustering or association analysis). You are free to choose any specific algorithm which had not been discussed in the lectures.
- Implement the algorithm, then process and validate it using the provided data.
- Predict future trends or insight on previous data.
- Visualize the results.

**4. Determine a Marketing Strategy**

- Use 1–2 PPT slides to present the recommended market strategy, grounded in the team's own data analysis and visuals:
  - Target segment, positioning statement/map, and core value proposition (reference chart/metrics from the team's analysis).
  - One acquisition tactic and one retention tactic, each tied to a specific insight from the team's visualization.

**5. Summarization and Presentation:**

- Prepare a comprehensive summary with detailed methodology, analysis, and findings.
- Present the project outcomes, highlighting key insights and challenges.

## **Deliverables and Evaluation Criteria:**

### **1. Submission Package:**

- **Source files:**
  - Include a copy of your source file(s) of your analysis tool, e.g. Python code (either in .py source files or .ipynb Jupyter Notebook) along with the accompanying dataset (if the original dataset is transformed, modified or augmented). The source files **MUST** reproduce all prediction results presented in the presentation.
  - Include a README file with instructions on how to run your source files. **NO marks** will be awarded if your code **cannot be executed properly**.
- **Presentation slide:**
  - Include a copy of your presentation slide file (e.g. pptx file)
  - Write your group number and members' student IDs on the slide

### **2. Presentation:**

- **Requirements:**
  - Deliver a concise summary of the project, emphasizing key insights and suggestion for the coming promotion campaign in a **4-minute** presentation.
  - Utilize roughly **10-20 slides**, with slide numbers optionally placed in the lower right corner.
- **Content Composition:**

You are recommended to present contents with a similar structure as below:

  - **Key Highlights:** Begin with a summary of the most important findings and insights from your analysis.
  - **Data Processing and Analysis Steps:** Describe the steps you took to prepare and analyse the data, including any ETL (Extract, Transform, Load) processes, key data cleaning or transformation actions, and the main approaches or trials you used during analysis and model validation.
  - **Strategy and Suggestions:** Summarize your main results and insights, and provide practical recommendations or strategies for the upcoming promotion campaign based on your analysis.
- **Oral Presentation:**
  - One or two members will represent the entire group in a **4-minute** presentation conducted in **English**. Exceeding the time limit will result in a penalty for the marks.
  - **Evaluation Criteria:**
    - The presentation will be assessed based on how clearly you describe and present the data, the suitability and correct use of your chosen data mining technique, and how well you draw useful insights for planning future promotions.
    - Key aspects include the quality of your data analysis and visualizations, the soundness of your chosen methods, and how clearly you explain your findings and suggestions. Your presentation should show a logical approach, thoughtful analysis, and practical recommendations, with an emphasis on clear communication and teamwork. Additionally, the presenters' attitude and time management will be considered.