**Data Science Final Project. 2024**

Group Task (4-5 persons max)

**Objective:**

Analyze the given data set from a large international store. Based on your analysis provide recommendations to the store owner so that profit can be increased.

However, bear in mind that this dataset is not straightforward to analyze. You should not blindly accept the results from your software.

**Subtasks:**

1. Perform a preliminary data analysis of the dataset and discuss the findings. For example, and not limited to:
   1. The frequency of each items purchased
   2. The number of transactions per month, on weekdays, weekends, time, etc (depending on dataset)
   3. The number of customers
   4. The number of items purchase
   5. Any other

Discuss if any of these analysis yield any understanding that will help the store owner.

1. Perform a market basket analysis using association rules to find:
   1. Support
   2. Confidence
   3. Lift
   4. Any other rules
2. Discuss the findings from the association rules.
3. Use of any other techniques to enhance understanding of the dataset (for example clustering, textual analysis, etc). Include only techniques that are useful and can explain the data. Explain why you use them. Hint: Do not use Recency, Frequency, Monetary clustering – they serve no purpose.
4. What recommendations can you give the store owner to **increase** the sales of the products? Your recommendations should be based on the results of your analysis. Be **specific** about your recommendations. Your recommendations should be based on your analysis. Prepare a list of recommendations, for example: 1. Item 123 should be …, 2. Product abc should be etc etc.

It will be even better if you could predict the increase in sales according to your recommendations. However, you need to show evidence for it (such as performing a regression on it), you just can’t claim without evidence.

**Presentation**

You will be required to give a short presentation to the class. Presentation is limited to 5 minutes about 5-8 slides max. Each team member must be presenting.

**Important**: Present your significant findings and recommendation first. Do not start with the usual preliminary analysis, what commands you run, etc.

Rough guide of slides

Slide 1-2: Recommendations to Owner and Why

Slide 2-3: Justification and assumptions for recommendations

Slide 3-5: Technique and results used that are useful and not useful

Slide 6-8: Difficulties faced and how can be improved

**Submission**

Only the report and code is required for submission. PPT not required.

**Report**

The report should be about 6-10 pages in length. Keep the introduction short. You can also omit the chapter explaining AR and other techniques. Please include a list of references.

Use a format from <https://www.ieee.org/conferences/publishing/templates.html>

In addition, put the contributions of each group member clearly in a separate page.

**Due date**: Check Moodle

**Marking Scheme**

**Report** **– group** (100)

1. Preliminary data analysis (Evidence: 10, Explanation: Quality & Relevance: 10)
2. Market basket analysis with AR (Evidence of ruleset: 10, Explanation: Quality & Relevance: 15)
3. Other applied techniques (Evidence of other techniques: 10, Explanation: Quality & Relevance: 15)
4. Recommendations (Quality and Relevance: 15)
5. Conclusion (10)
6. References (5)

**Presentation – individual** (Tentative 100)

1. Quality of Delivery:
   1. Tone and Loudness (20)
   2. Eye contact (20)
   3. Preparedness and fluency (20)
2. Quality of Content:
   1. Professional slide design (20)
   2. Sufficiency of content (20)