Form 2 (Task 2)

While solving the task, create a report in a **Jupyter Notebook** following the sections of this form.

1. Student information	
a. Name	
b. Student number	
c. Date	
2. Question 1. Task identification. Given the assigned data set, is this an association rule mining problem or a text mining problem? Explain your reasoning. (Analyse with Python, compare, explain)	
3. Question 2. Analyse the dataset potential. Does the dataset you have been assigned have a specific class associated to the instances of the dataset? Can this be considered a classification problem, regression, or would clustering be a better choice? Are there problems in the dataset? What other analysis can you think about? What about the features of the dataset, anything important to report? Explain your reasoning and use charting where appropriate. (Analyse with Python, compare, explain.)	
4. Question 3. Data analysis Part 1. If you have an association rule mining problem: a. Explain all the steps needed in the analysis. b. Explain the meaning of the found itemsets. c. Select one of the categorical variables in the dataset and apply a classification algorithm of your choice selected from those studied in the course. Report the confusion matrix obtained and explain the result. Alternatively, take a subset of the continuous attributes in your data set and apply a clustering algorithm, explaining the results. Provide the Python Code. If you have a text mining problem: a. Add a label to the documents groups you have been provided and explain why you selected those labels. b. Apply the word to vector necessary transformations (stop word removal, tokenization, TF-IDF, word frequencies…) . Explain your selection. Provide the Python code. c. Apply a classification or a clustering algorithm, then highlight the results in your own words. Provide the Python code.	
5. Data analysis Part 2. If in answering Question 3c. you applied a clustering (classification or regression) algorithm, now apply a classification (clustering) algorithm and explain the results. Provide the Python code. Provide the evaluation metrics. Explain the evaluation metrics.	