Dear Participants,

Please find below SMDM Project instructions:

* You have to submit 2 files :
  1. **Answer Report**: In this, you need to submit all the answers to all the questions in a sequential manner. **It should include a detailed explanation of the approach used, insights, inferences, and all outputs of codes like graphs, tables, etc.** Your report should **not** be filled with codes. You will be evaluated based on the business report only. Hence please ensure that your business report is detailed and includes everything apart from the code. **THE REPORT HAS TO BE STRICTLY SUBMITTED IN A PDF/DOC FORMAT. ANY OTHER FORMAT WILL NOT BE CONSIDERED FOR GRADING. 6 Marks are allotted for the "Quality of Business Report".**
  2. **Jupyter Notebook file**: This is a must and will be used for reference while evaluating
* Any assignment found copied/ plagiarized by another person will not be graded and marked as zero.
* Please ensure timely submission as a post-deadline assignment will not be accepted.

Problem 1

**Wholesale Customers Analysis (**[**Download Data**](https://olympus.mygreatlearning.com/courses/80497/files/6317243/download?wrap=1)**[View in a new window](https://olympus.mygreatlearning.com/courses/80497/files/6317243/download?wrap=1))**

**Problem Statement:**

A wholesale distributor operating in different regions of Portugal has information on the annual spending of several items in their stores across different regions and channels. The data consists of 440 large retailers’ annual spending on 6 different varieties of products in 3 different regions (Lisbon, Oporto, Other) and across different sales channels (Hotel, Retail).

**1.1** **Use methods of descriptive statistics to summarize data. Which Region and which Channel spent the most? Which Region and which Channel spent the least?**

**1.2 There are 6 different varieties of items that are considered. Describe and comment/explain all the varieties across Region and Channel? Provide a detailed justification for your answer.**

**1.3 On the basis of a descriptive measure of variability, which item shows the most inconsistent behaviour? Which items show the least inconsistent behaviour?**

**1.4 Are there any outliers in the data? Back up your answer with a suitable plot/technique with the help of detailed comments.**

**1.5 On the basis of your analysis, what are your recommendations for the business? How can your analysis help the business to solve its problem? Answer from the business perspective**

**Problem 2:**

The dataset [Education - Post 12th Standard.csv View in a new window](https://olympus.mygreatlearning.com/courses/80497/files/6317235/download?wrap=1)contains information on various colleges. You are expected to do a Principal Component Analysis for this case study according to the instructions given. The data dictionary of the 'Education - Post 12th Standard.csv' can be found in the following file: [Data Dictionary.xlsx.View in a new window](https://olympus.mygreatlearning.com/courses/80497/files/6317236/download?wrap=1)

* Perform Exploratory Data Analysis [Univariate, Bivariate, and Multivariate analysis to be performed]. What insight do you draw from the EDA?

**Please reflect on all that you have learned while working on this project. This step is critical in cementing all your concepts and closing the loop. Please write down your thoughts in the form.   
  
Link to form :** <https://docs.google.com/forms/d/e/1FAIpQLSfu-8AVwnWSHvWnAEbMyp3nQjwr2bfaToZvymxwi9Xo9UOKaA/viewform>

Top of Form

**Points**

60

Bottom of Form