

Problem Statement: The 'Hair Salon.csv' dataset contains various variables used for the context of Market Segmentation. This particular case study is based on various parameters of a salon chain of hair products. You are expected to do Principal Component Analysis for this case study according to the instructions given in the following rubric.

Note: This particular dataset contains the target variable satisfaction as well. Please do drop this variable before doing Principal Component Analysis.

Questions:

- 1) Perform Exploratory Data Analysis [both univariate and multivariate analysis to be performed]. The inferences drawn from this should be properly documented. – **5 points**
- 2) Scale the variables and write the inference for using the type of scaling function for this case study. - **3 points**
- 3) Comment on the comparison between covariance and the correlation matrix after scaling. - **2 points**
- 4) Check the dataset for outliers before and after scaling. Draw your inferences from this exercise. - **3 points**
- 5) Build the covariance matrix, eigenvalues and eigenvector. - **4 points**
- 6) Write the explicit form of the first PC (in terms of Eigen Vectors) – **5 points**
- 7) Discuss the cumulative values of the eigenvalues. How does it help you to decide on the optimum number of principal components? What do the eigenvectors indicate? Perform PCA and export the data of the Principal Component scores into a data frame. – **10 points**
- 8) Mention the business implication of using the Principal Component Analysis for this case study. – **5 points**