

Exploratory Data Analysis-1

Assignment Questions



Q1. What are the key features of the wine quality data set? Discuss the importance of each feature in predicting the quality of wine.

Q2. How did you handle missing data in the wine quality data set during the feature engineering process? Discuss the advantages and disadvantages of different imputation techniques.

Q3. What are the key factors that affect students' performance in exams? How would you go about analyzing these factors using statistical techniques?

Q4. Describe the process of feature engineering in the context of the student performance data set. How did you select and transform the variables for your model?

Q5. Load the wine quality data set and perform exploratory data analysis (EDA) to identify the distribution of each feature. Which feature(s) exhibit non-normality, and what transformations could be applied to these features to improve normality?

Q6. Using the wine quality data set, perform principal component analysis (PCA) to reduce the number of features. What is the minimum number of principal components required to explain 90% of the variance in the data?

Note:- We can use Wine quality dataset and Student Performance Data set as per the discussion in lecture.

Note: Create your assignment in Jupyter notebook and upload it to GitHub & share that github repository link through your dashboard. Make sure the repository is public.