PCA: USA Education Institute feature engineering Case study

Dataset: usa_education_inst_data.csv

Context

In this case study, we will use the Education dataset which contains information on educational institutes in the USA. The data has various attributes about the number of applications received, enrolments, faculty education, financial aspects, and the graduation rate of each institute.

Objective

The objective of this problem is to reduce the number of features by using dimensionality reduction techniques such as high correlation filter and PCA, and extract insights about the data.

Dataset

The Education dataset contains information on various colleges in the USA. It contains the following features:

- Names: Names of various universities and colleges
- Apps: The number of applications received
- Accept: The number of applications accepted
- Enroll: The number of new students enrolled
- Top10perc: The percentage of new students from the top 10% of Higher Secondary class
- Top25perc: The percentage of new students from the top 25% of Higher Secondary class
- F_Undergrad: The number of full-time undergraduate students
- P_Undergrad: The number of part-time undergraduate students
- Outstate: The number of students for whom the particular college or university is out-of-state tuition
- Room_Board: The cost of room and board
- Books: The estimated cost of books for a student
- Personal: The estimated personal spending for a student
- PhD: The percentage of faculties with a Ph.D.
- Terminal: The percentage of faculties with a terminal degree
- S_F_Ratio: Student/faculty ratio
- perc_alumni: Percentage of alumni who donate
- Expend: The instructional expenditure per student
- Grad_Rate: Graduation rate