Home Assignment Topics for Data Science

Home Assignment Instructions

Consider the following dataset. Apply the following: **Use “R Only”**

1. Use minimum two different appropriate visualization tools for visualizing/comparing the data. **Use ggplot2 command.**
2. Presentation of Meta Data.
3. Demonstrate the Data pre-processing carried out (if any).
4. Perform Hypothesis testing by setting an appropriate hypothesis.
5. Find the principal components. How many principal components are required to describe 90% of total variance? (Hint: Use prcomp command).
6. Check whether any two or more variables are correlated. Find the degree of correlation. Also plot the scatter plot for the same.
7. Apply regression to predict the future value of one/more variables if applicable. (If not applicable, justify).
8. Apply a suitable classifier to classify the given data into one/two classes if applicable. (If not applicable, justify).

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B3-3 | Gross domestic expenditure on R & D | <https://data.un.org/> |