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 |
| B1-1 | SBA Disaster Loan Data FY2017 | <https://catalog.data.gov/dataset/sba-disaster-loan-data-fy2017> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B1-2 | Population, surface area and density | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B1-3 | County and State Medicare spend | <https://data.world/adamhelsinger/county-state-medicare-spend> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B1-4 | Population growth, fertility, life expectancy and mortality | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B1-5 | Population, Population Change, and Estimated Components of Population Change: April 1, 2010 to July 1, 2019 (CO-EST2019-alldata) | https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B1-6 | Intentional homicides and other crimes | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B2-1 | Treasury Recovery Act Data | <https://catalog.data.gov/dataset/treasury-recovery-act-data> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B2-2 | GVA by kind of economic activity | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B2-3 | Ingestion of swimming pool water by recreational | https://catalog.data.gov/dataset/ingestion-of-swimming-pool-water-by-recreational |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B2-4 | Education at the primary, secondary and tertiary levels | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B2-5 | Index of industrial production | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B2-6 | Teaching staff at the primary, secondary and tertiary levels | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B3-1 | Public expenditure on education | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B3-2 | Inventory of Owned and Leased Properties (IOLP) | https://catalog.data.gov/dataset/inventory-of-owned-and-leased-properties-iolp |

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

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B3-4 | Employment by economic activity | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B3-5 | Energy production, trade and consumption | <https://data.un.org/> |

|  |  |  |
| --- | --- | --- |
| Grp No | Name of Data Set | Link |
| B3-6 | Finances of Selected State | https://data.world/finance/finances-of-selected-state/workspace |