**Supplementary Information:**

**Supplementary Table 1:** Descriptive Statistics for the given variables

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| **Rank** | 100 | 50.49 | 29.0069 | 1 | 25.75 | 50.5 | 75.25 | 100 |
| **PR** | 100 | 32.8581 | 14.4547 | 1.94 | 24.6225 | 33.59 | 40.88 | 98.56 |
| **Male** | 100 | 50.6234 | 14.1193 | 0 | 42.9825 | 50.535 | 56.72 | 85.67 |
| **Female** | 100 | 49.3766 | 14.1193 | 14.33 | 43.28 | 49.465 | 57.0175 | 100 |
| **OC** | 100 | 1.8539 | 2.52593 | 0 | 0.2175 | 0.995 | 2.13 | 11.7 |
| **EC** | 100 | 13.8844 | 15.4159 | 0 | 3.35 | 8.61 | 19.99 | 77.69 |
| **SC** | 100 | 40.6855 | 23.0645 | 1.08 | 25.1425 | 39.98 | 52.8625 | 93.85 |
| **TFR\_Govt** | 100 | 18.9147 | 23.1918 | 0 | 0.315 | 9.905 | 29.56 | 76.23 |
| **TFR\_Inst** | 100 | 7.0392 | 9.99663 | 0 | 0.5625 | 2.7 | 9.8575 | 55.03 |
| **TFR\_Pvt** | 100 | 1.9842 | 5.12988 | 0 | 0 | 0.12 | 1.26 | 35.21 |
| **TFR\_No** | 100 | 26.6299 | 22.2540 | 0 | 7.9125 | 21.26 | 37.82 | 90.47 |
| **Grad** | 100 | 91.474 | 8.52177 | 64.29 | 88.1925 | 93.47 | 96.785 | 100 |
| **Placed** | 100 | 48.9218 | 24.5842 | 2.7 | 25.85 | 53.27 | 69.5375 | 95.21 |
| **MS** | 100 | 7555435 | 4667798 | 343 | 3781750 | 6769000 | 9969500 | 21610000 |
| **HS** | 100 | 19.3911 | 11.9721 | 2.31 | 10.9725 | 16.71 | 24.415 | 71.71 |
| **HRA** | 100 | 1.93 | 0.78180 | 1 | 1 | 2good | 3 | 3 |
| **Funding** | 100 | 2.25 | 0.72995 | 1 | 2 | 2 | 3 | 3 |
| **NAAC** | 100 | 0.91 | 0.28762 | 0 | 1 | 1 | 1 | 1 |
| **YoE** | 100 | 1968.41 | 36.3857 | 1857 | 1952 | 1980.5 | 1996 | 2012 |

**Supplementary Table 2:** Qualitative statistical tests among features and other state students

|  |  |  |  |
| --- | --- | --- | --- |
| **Features** | **P-Value (0.05)** | **Statistical Test** | **Summary** |
| Type of University | 0.0001 | Kruskal-Wallis ANOVA | There is a significant difference for percentage of other state students in different type of universities |
| HRA | 0.08 | Kruskal-Wallis ANOVA | There is no significant difference for percentage of other state students among different HRA |
| YoE (*ρ=0.33*) | 0.0001 | Spearman Correlation | There is significant difference between YoE and perncentage of Other state students |
| Border(*ρ=0.20*) | 0.004 | There is significant difference between total number of border nearby and percentage of Other State Students |
| Railway (*ρ=0.29*) | 0.0001 | There is significant difference between Railway Station distance from campus and percentage of Other State Students |
| Rank (ρ= -0.27) | 0.0001 | Higher the rank lower the other state students (there is a significant different between both) |

Note: ρ= Spearman Rank Correlation

|  |  |
| --- | --- |
| **(a)** | **(b)** |
| A black background with yellow dots  Description automatically generated | A graph of a bar chart  Description automatically generated with medium confidence |

**Supplementary Figure 1:**

a) Bar chart depicting R2 State Funded Institution’s (SFI) other state student diversity for X, Y, Z, and OA (House Rent Allowance (HRA) Classification of cities) towards Rank of their Institution.

(b) Scatter plot showing the negative correlation between SFI institution Rank and

Percentage of other state students.

|  |  |
| --- | --- |
| **(a)** | **(b)** |
| A chart of a model  Description automatically generated with medium confidence | A graph of a positive tree  Description automatically generated |

**Supplementary Figure 2:**

Classification Learner highlighting the accuracy of coarse tree model using (a) Confusionmatrix and (b) ROC Curve method:

(a) Training Confusion Matrix for Coarse Tree Classification Model with Accuracy (0.8), Precision (0.64), Recall (0.62), and F1 score (0.63)

(b) ROC Curve for Coarse Tree Classification Model in High, Low, and Medium



**Supplementary Figure 3:** Women Students Enrolled in different stream of education (NIRF, 2023)