|  |  |  |
| --- | --- | --- |
| **Exploratory Data Analysis** | | |
| Total Marks: 30 | | |
| Please use the suicides Dataset provided | | |
| **Section A** | | |
| 1. | How many categorical variables does the data contain? Please state the number of categories for each such variable. | 2 marks |
| 2. | How strong is the correlation between HDI and suicides\_no? | 2 marks |
| 3. | Which generation has the highest number of suicides/100k pop? | 2 marks |
| 4. | Which country has the least number of suicides between 1990-1995? | 2 marks |
| **Section B** | | |
| 5. | Are there any countries with no suicides recorded? Create a new data frame which ranks these countries by HDI. | 3 marks |
| 6. | Generate suitable graphs for comparing suicides between men and women for the top 5 countries with the highest suicide rate per 100,000. | 3 marks |
| 7. | Are there any redundant columns in the dataset? Which coulmn is it? Can that column be dropped? State your reasons. | 3 marks |
| 8. | Please obtain the distribution of suicides for each age group for Argentina. Plot these as graphs | 3 marks |
| 9. | Generate a correlation heatmap for the dataset. Which pairs of variables are highly correlated. | 3 marks |
| **Section C** | | |
| 10. | Generate the following tables:   1. A table containing the columns ‘Country’, ‘Year’, ‘Total suicides’. Total Suicides has to be calculated from the existing table.(3 marks) 2. A Table containing the columns ‘Country’, ‘Year’, ‘per capita gdp’.(2 marks) 3. Merge the above two tables using ‘Country’ as the merge column.(2 marks) | 7 marks |