|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics**    **Aim of the Experiment**  **The main aim of this experiment is to explore the given dataset. A sample database is created and is available in the file sample.csv.**  **Sample Dataset**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **id** | **first** | **last** | **gender** | **Marks** | **selected** | | **1** | **Leone** | **Debrick** | **Female** | **50** | **TRUE** | | **2** | **Romola** | **Phinnessy** | **Female** | **60** | **FALSE** | | **3** | **Geri** | **Prium** | **Male** | **65** | **FALSE** | | **4** | **Sandy** | **Doveston** | **Female** | **95** | **FALSE** | | **5** | **Jacenta** | **Jansik** | **Female** | **31** | **TRUE** | | **6** | **Diane-marie** | **Medhurst** | **Female** | **45** | **TRUE** | | **7** | **Austen** | **Pool** | **Male** | **45** | **TRUE** | | **8** | **Vanya** | **Teffrey** | **Male** | **70** | **FALSE** | | **9** | **Giordano** | **Elloy** | **Male** | **36** | **FALSE** | | **10** | **Rozele** | **Fawcett** | **Female** | **50** | **FALSE** |     **The objectives of this experiment are:**  **1. Explore all the statistical operations of Pandas and given in Listing 1**      **2. Use Describe command and explore the dataset as given in Listing 2**    **3. Use Descriptive Statistics for univariate and bivariate data as given in Listing 3** |
| **Data Preprocessing**    **Aim of the Experiment.**  The main aim of this experiment is to preprocess the given dataset. The database is created and is available in the file sample.csv.  Sample Dataset   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | id | first | last | gender | Marks | selected | | 1 | Leone | Debrick | Female | 50 | TRUE | | 2 | Romola | Phinnessy | Female | 60 | FALSE | | 3 | Geri | Prium | Male | 65 | FALSE | | 4 | Sandy | Doveston | Female | 95 | FALSE | | 5 | Jacenta | Jansik | Female | 31 | TRUE | | 6 | Diane-marie | Medhurst | Female | 45 | TRUE | | 7 | Austen | Pool | Male | 45 | TRUE | | 8 | Vanya | Teffrey | Male | 70 | FALSE | | 9 | Giordano | Elloy | Male | 36 | FALSE | | 10 | Rozele | Fawcett | Female | 50 | FALSE |     The objectives of this experiment are  1. Explore Label Encoder    2. Explore Scikit Preprocessing routines like Scaling    3. Explore Scikit Preprocessing routines like Binarizer |