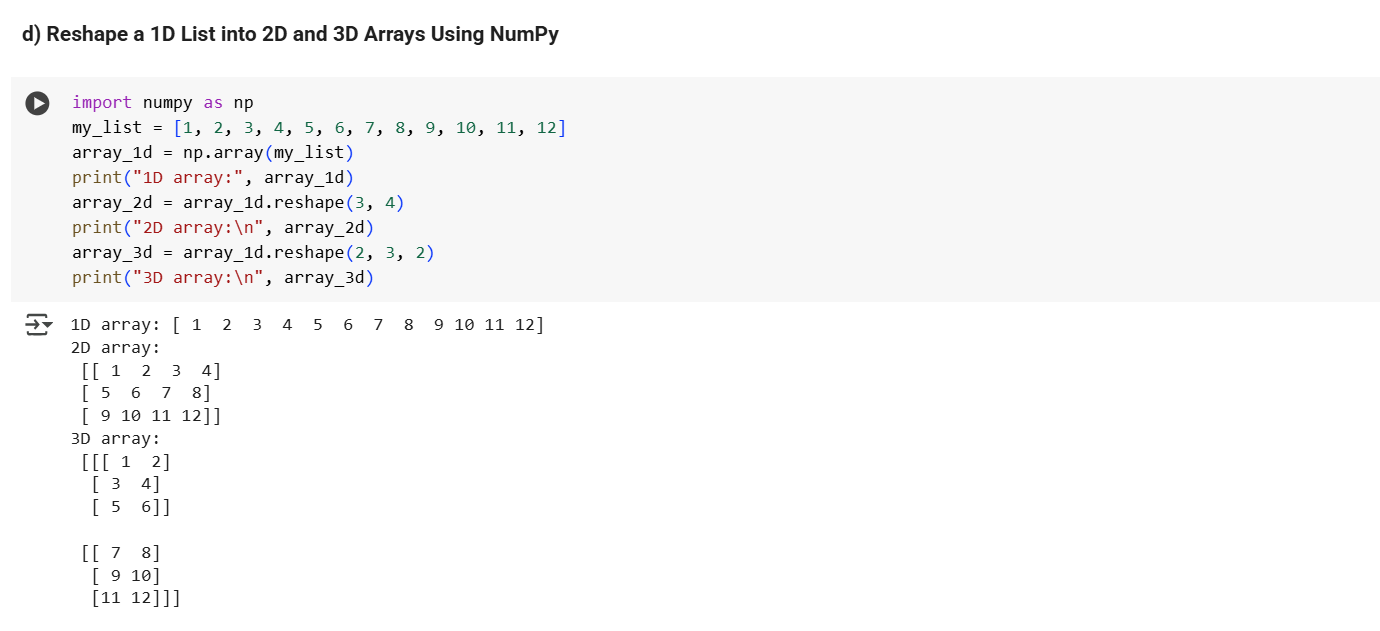
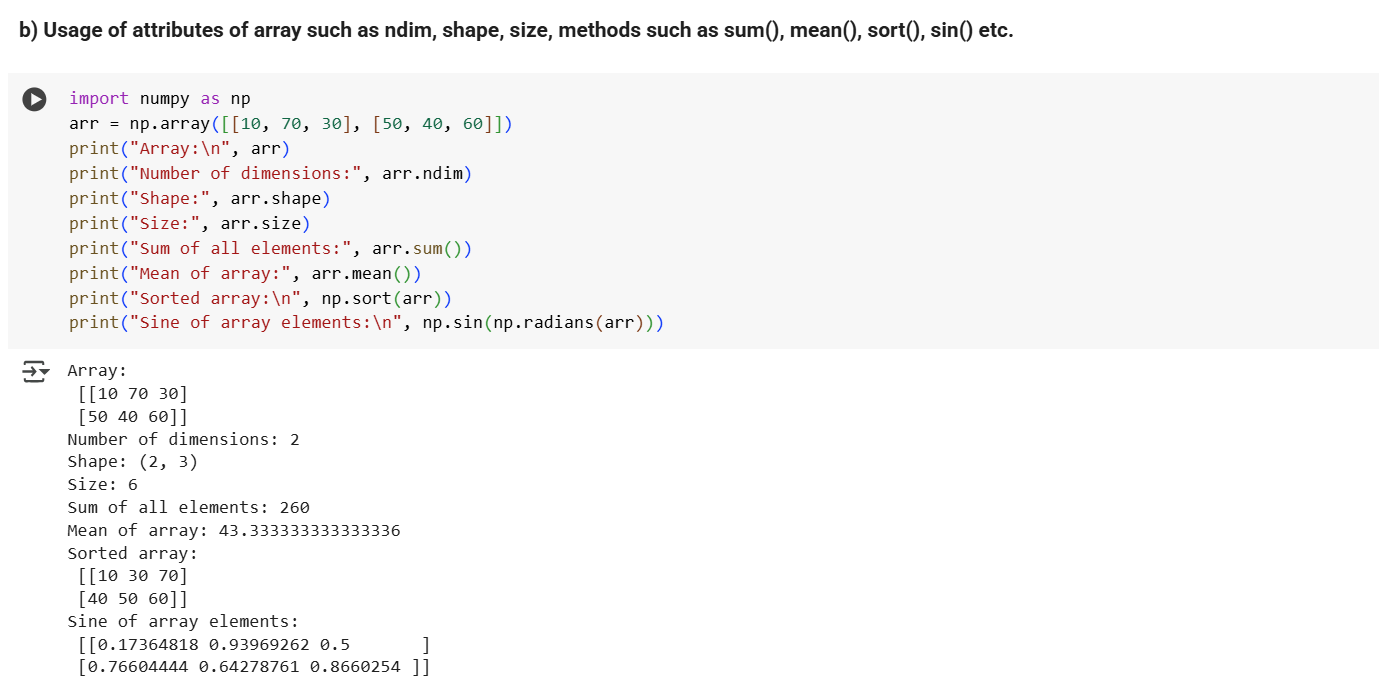
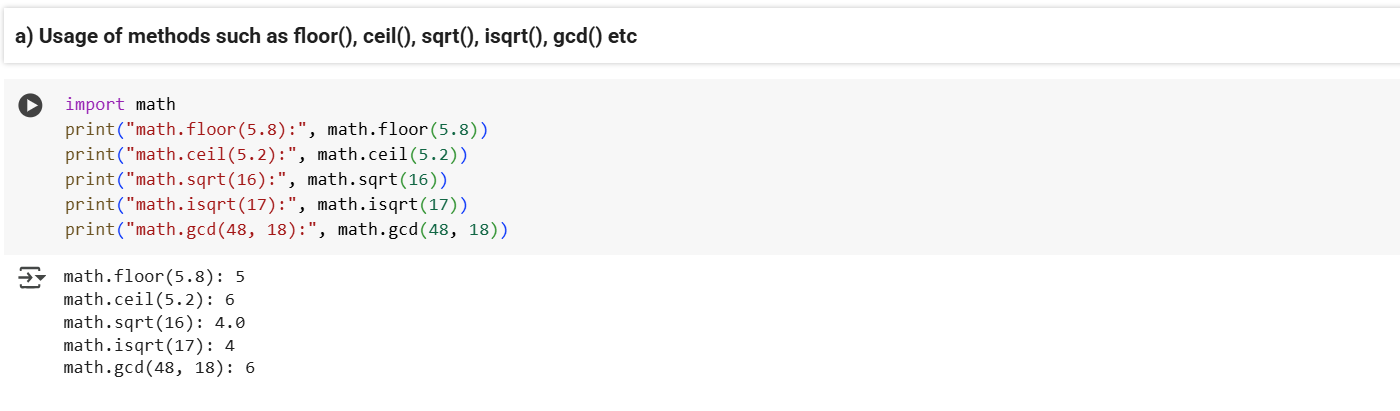
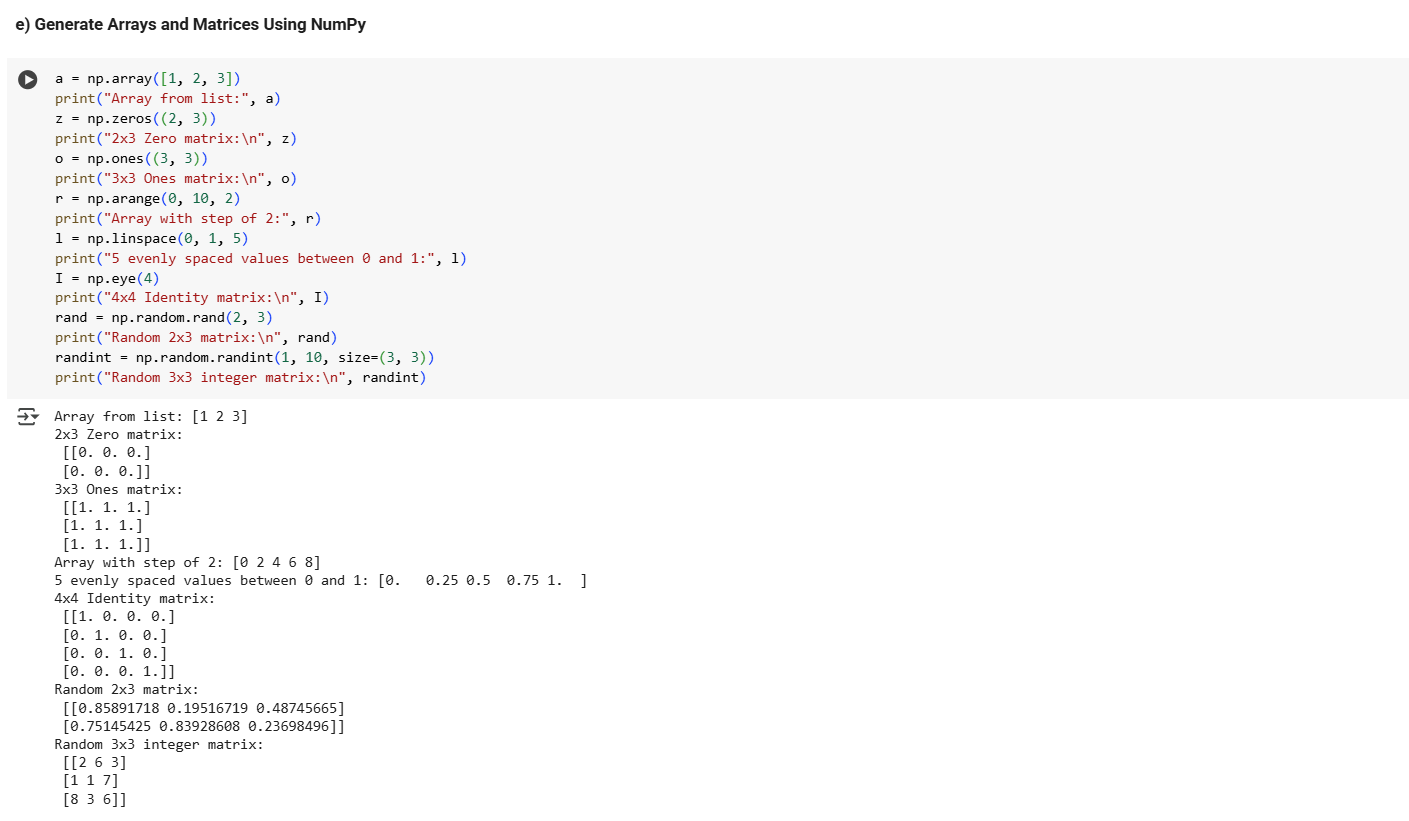
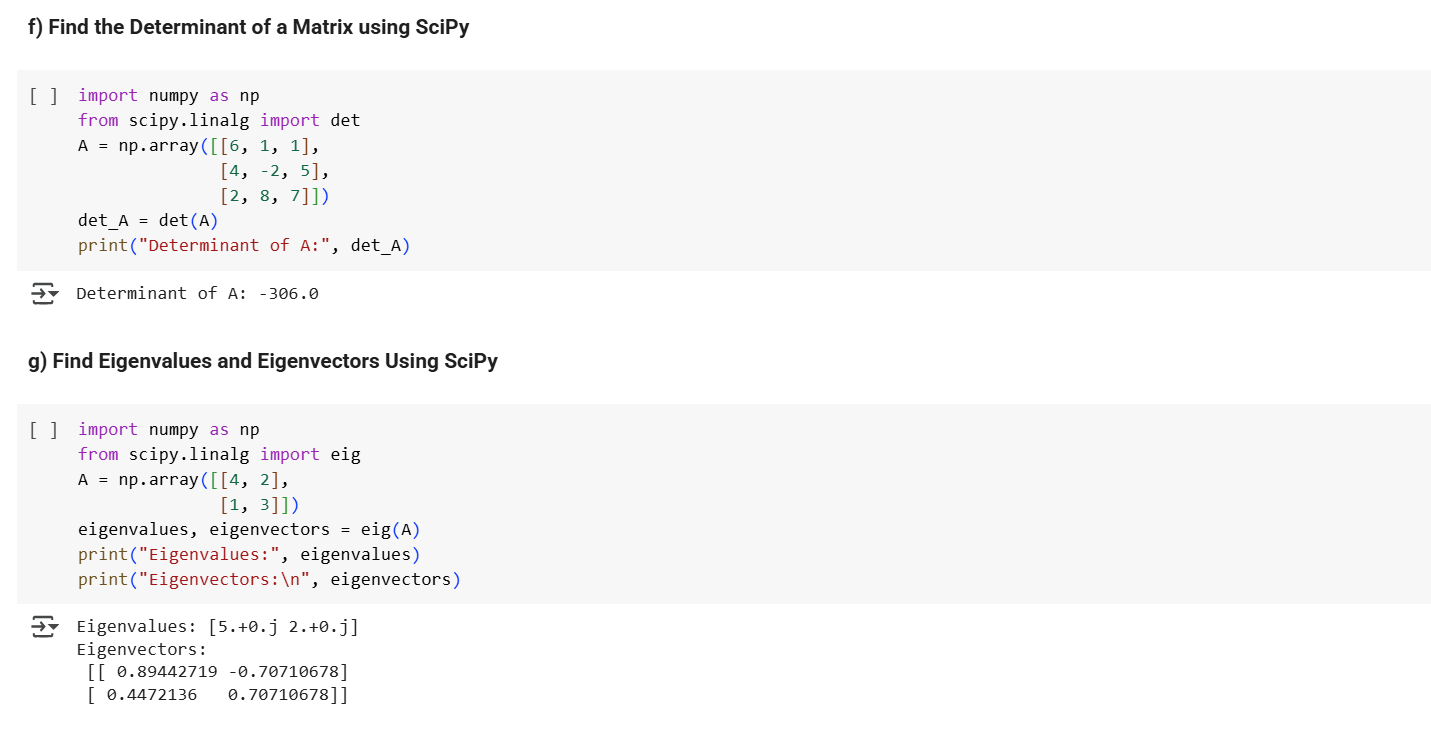
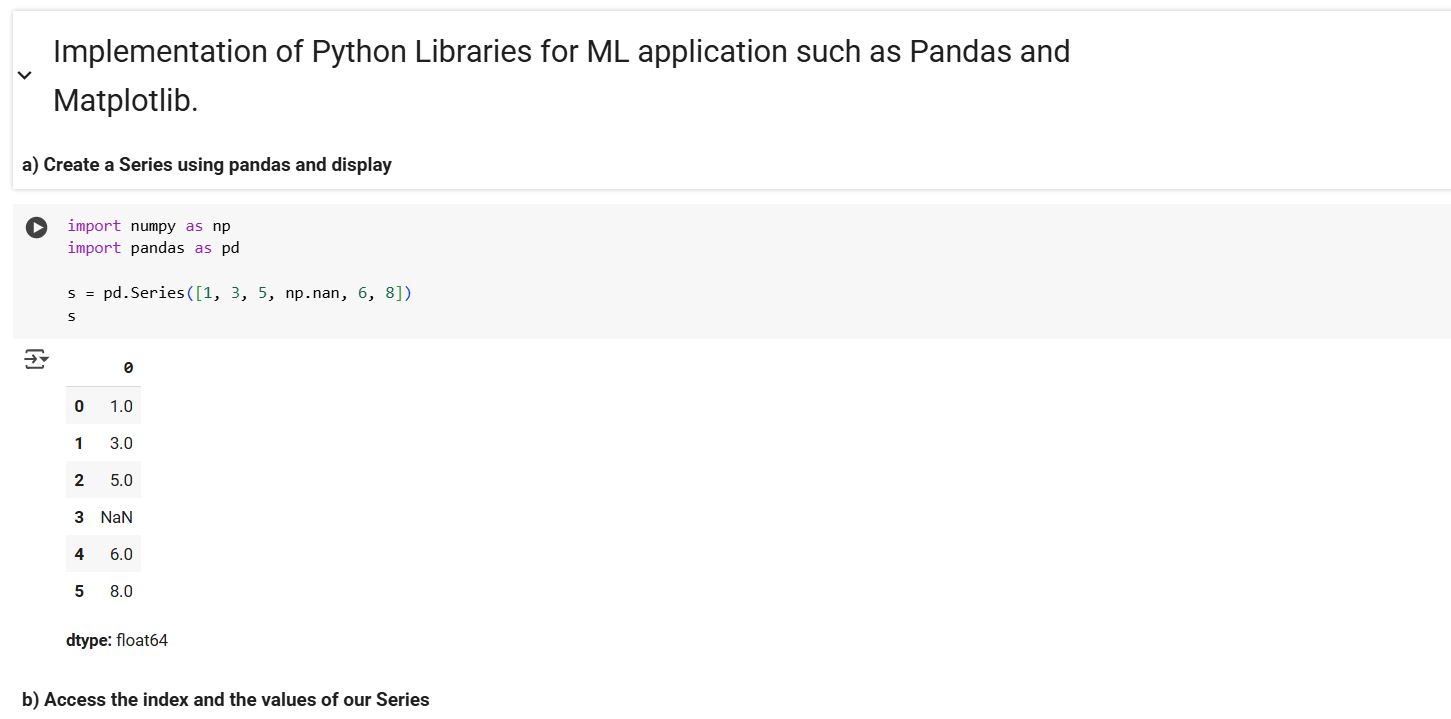
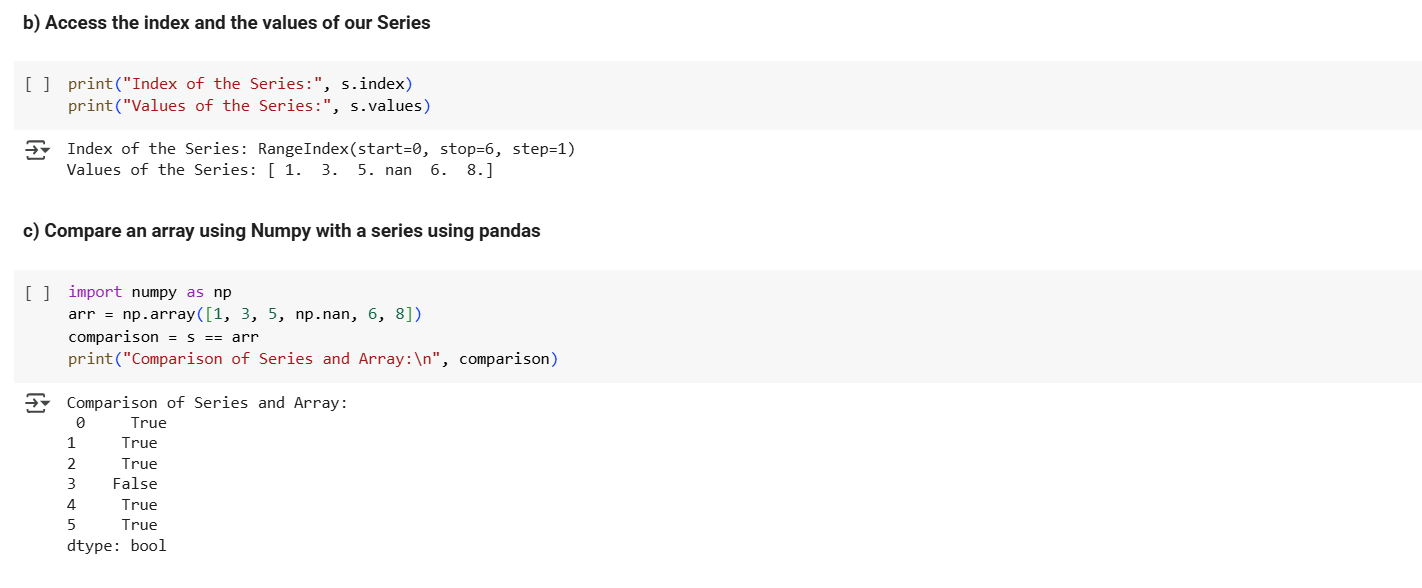
Sankara Krishnan P – 2023115074 – Lab Exercise – 2[21-07-2025]



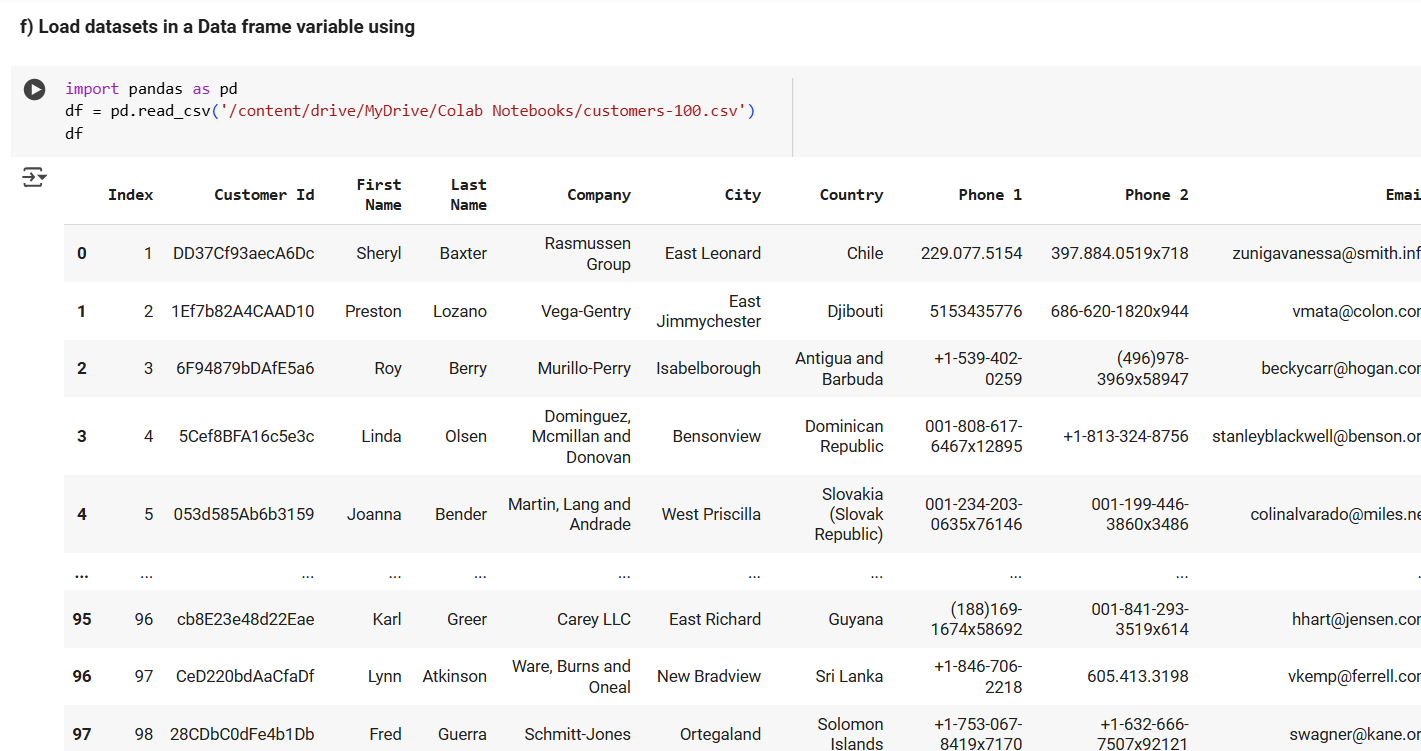


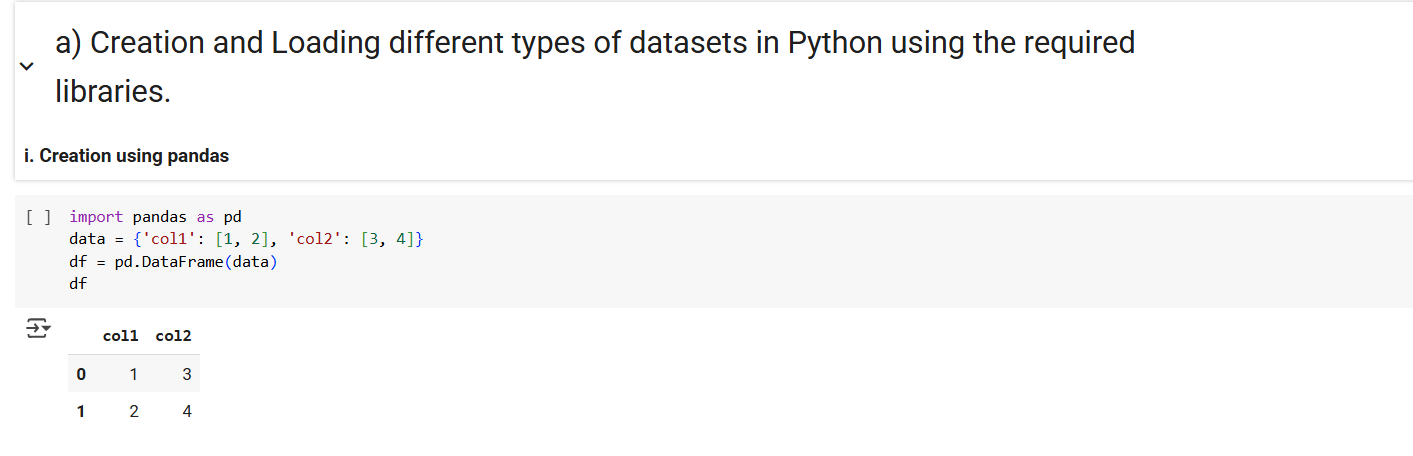


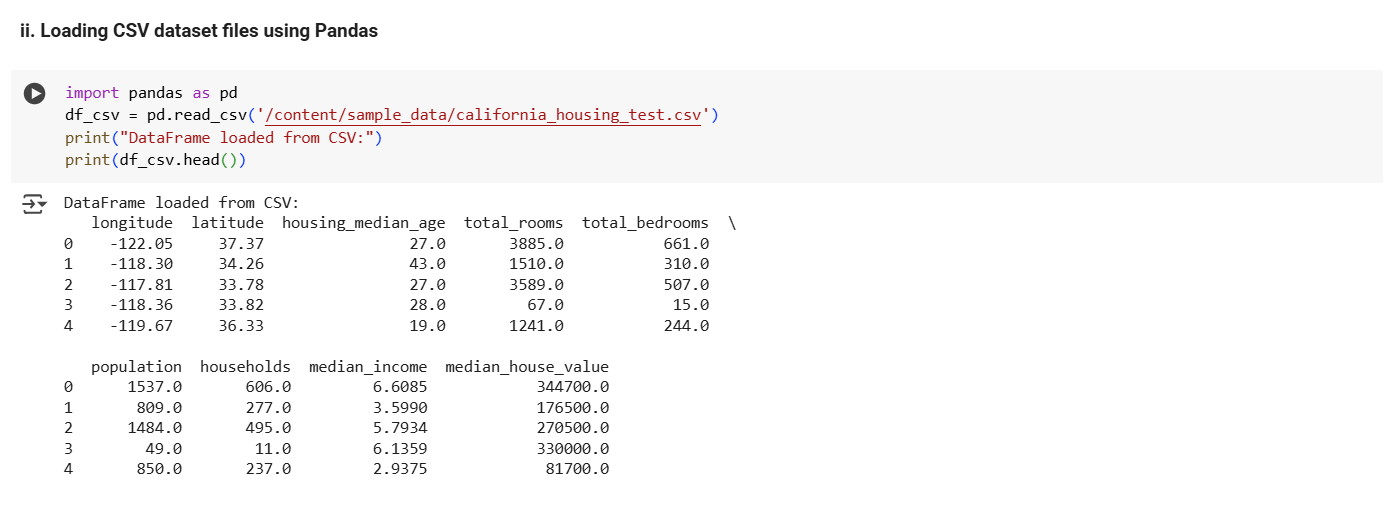




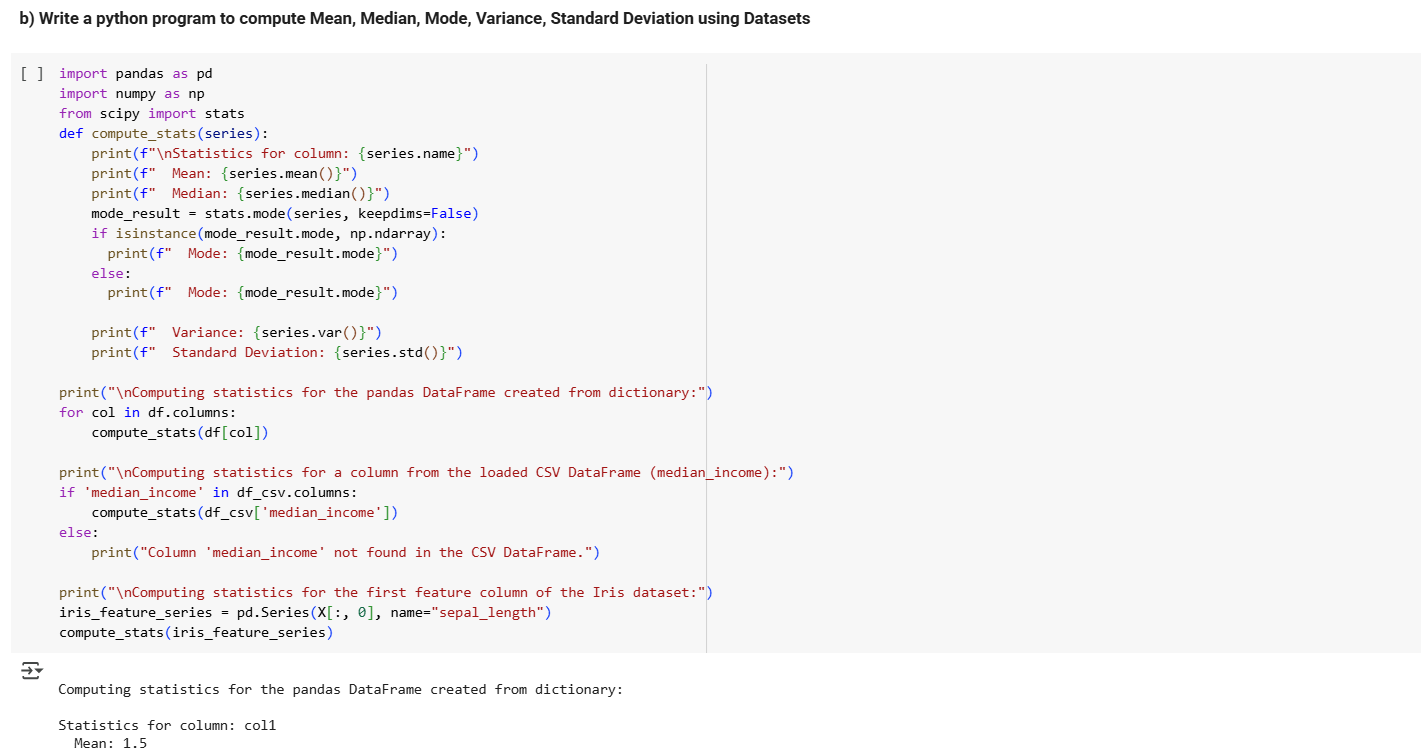


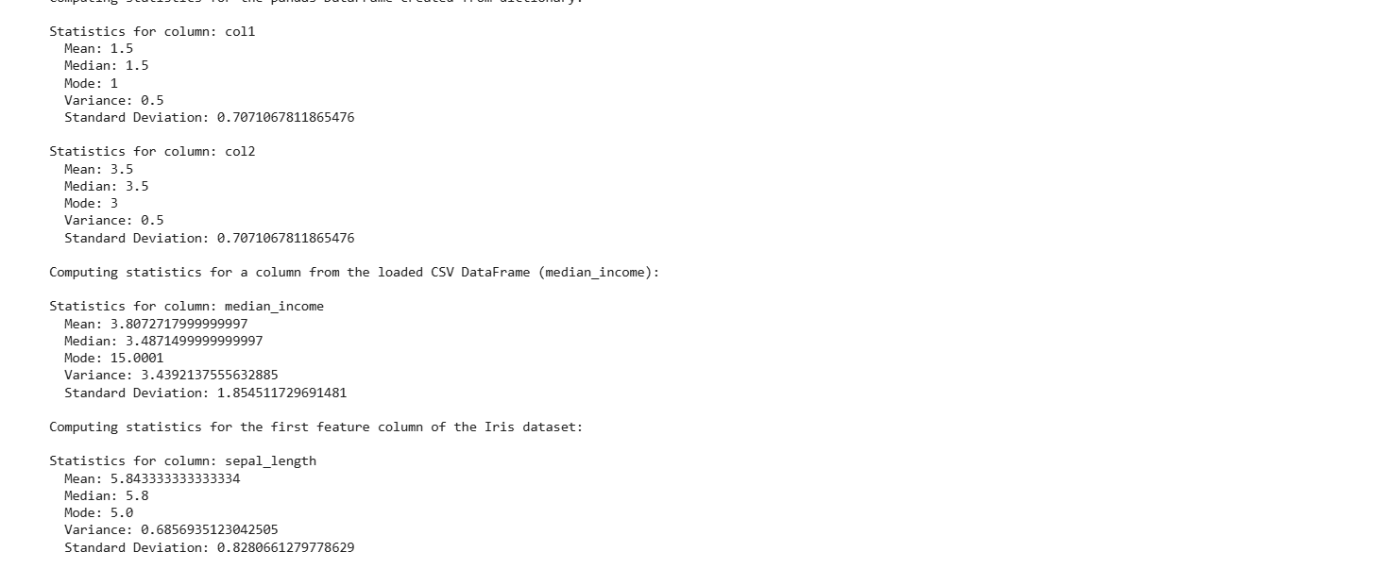




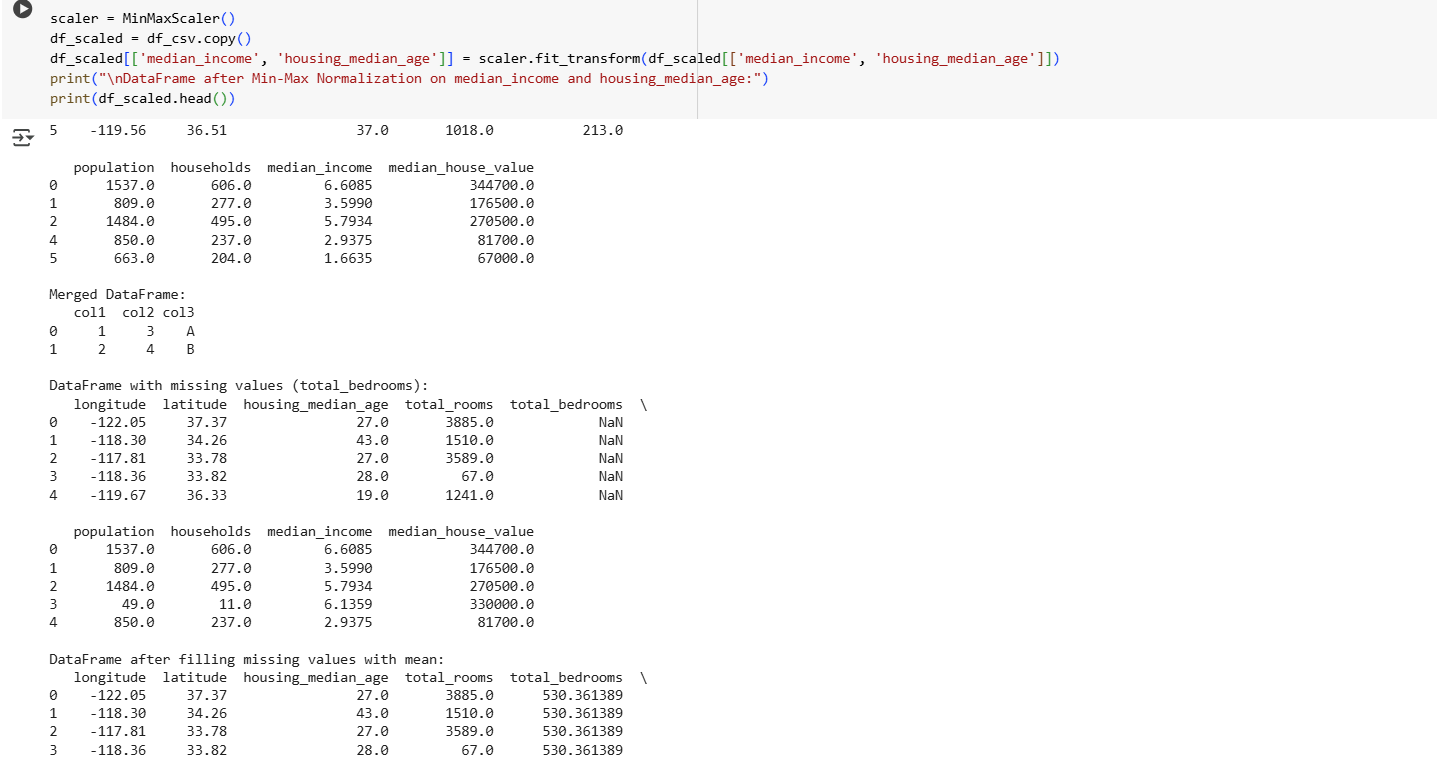


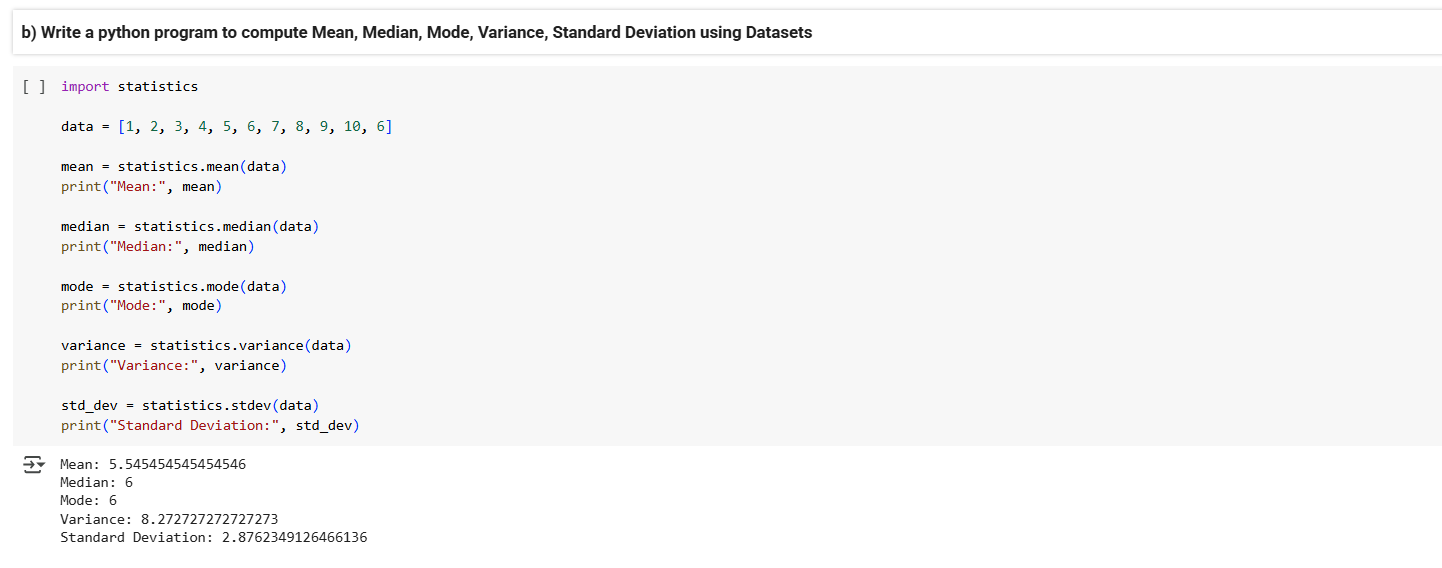




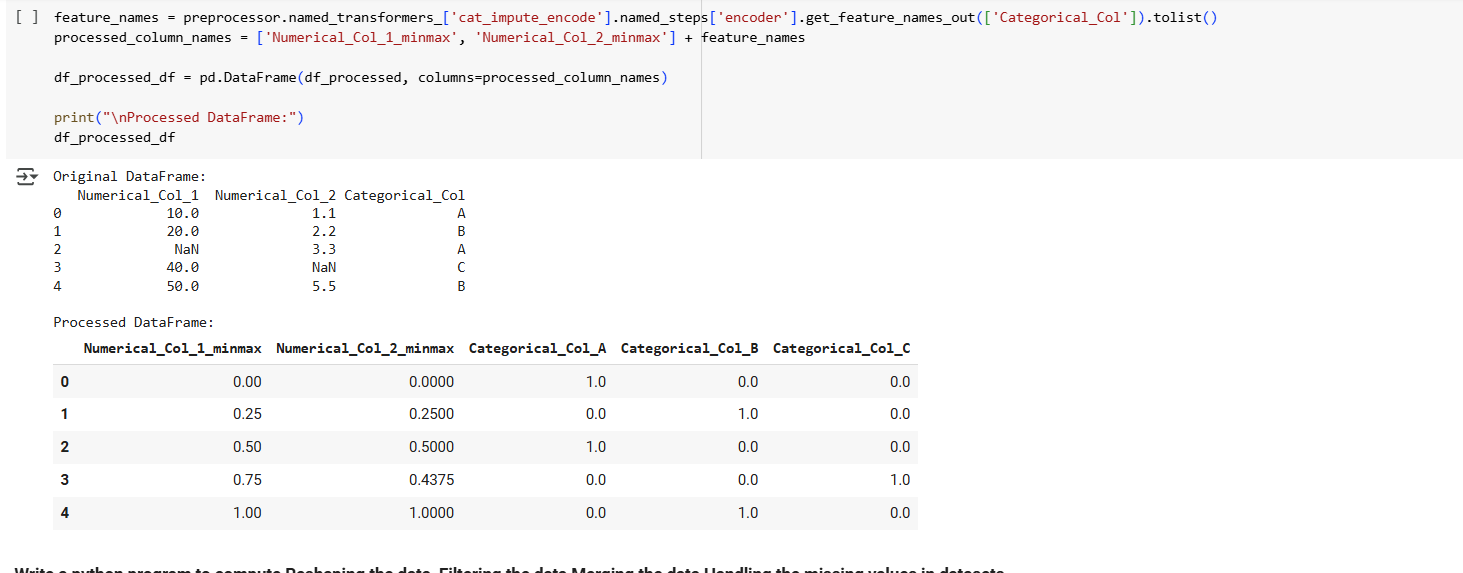


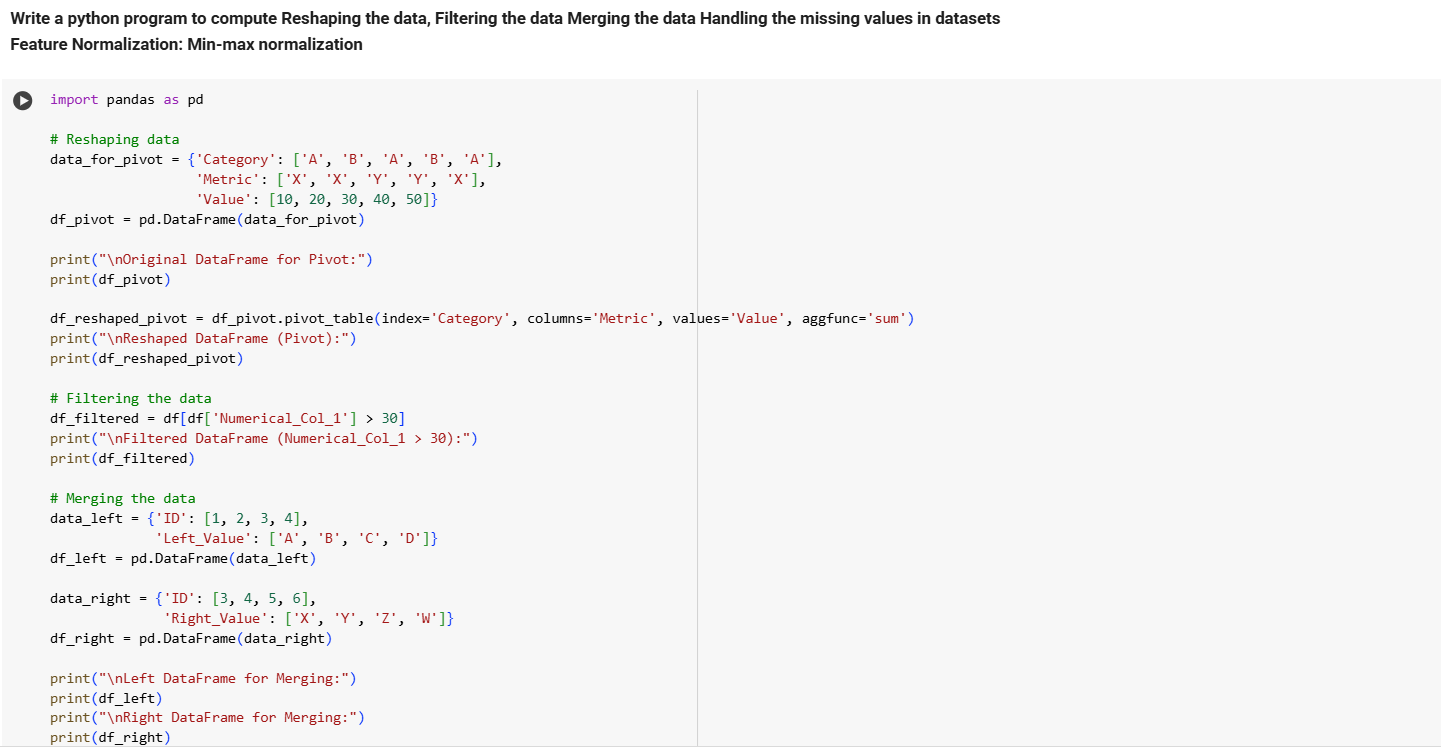


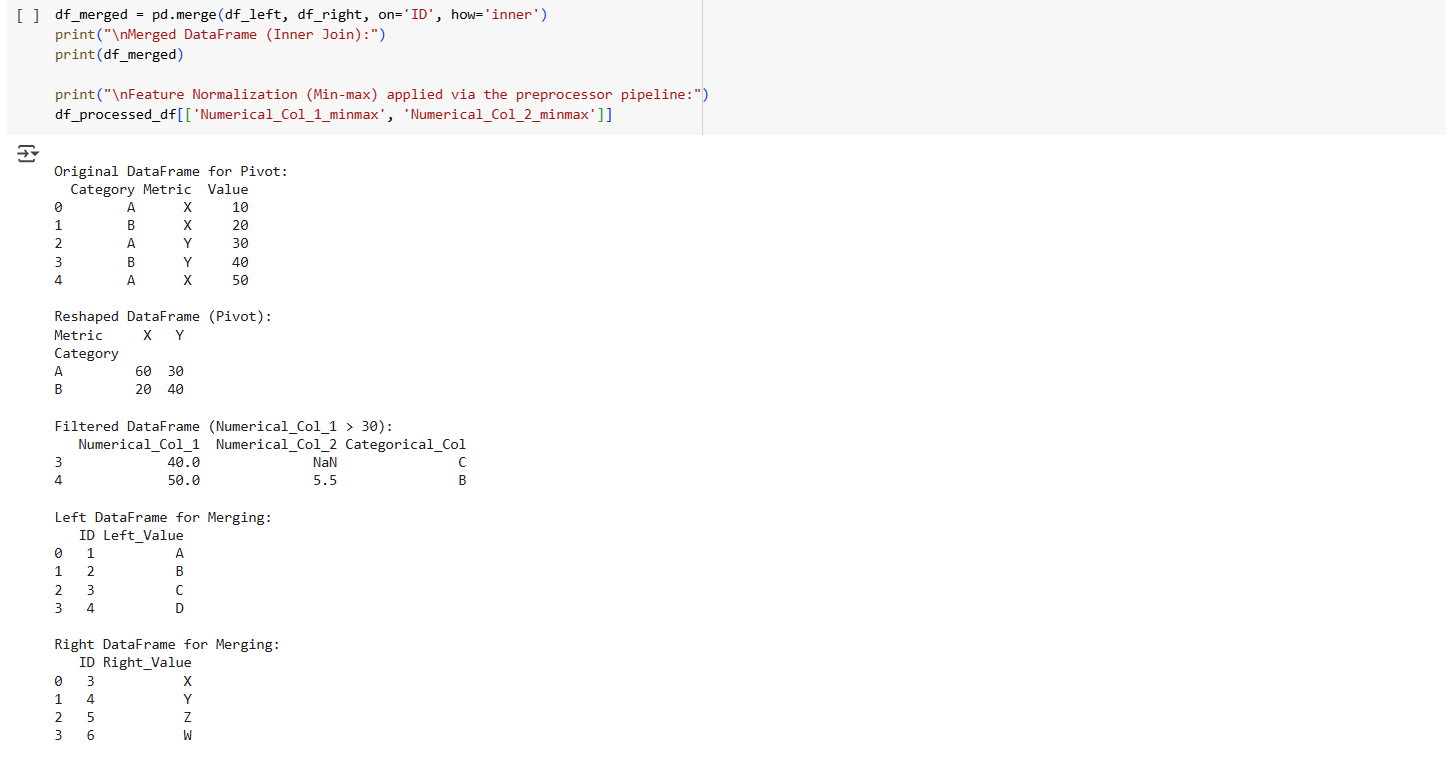












**Descriptive Statistics**

**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**

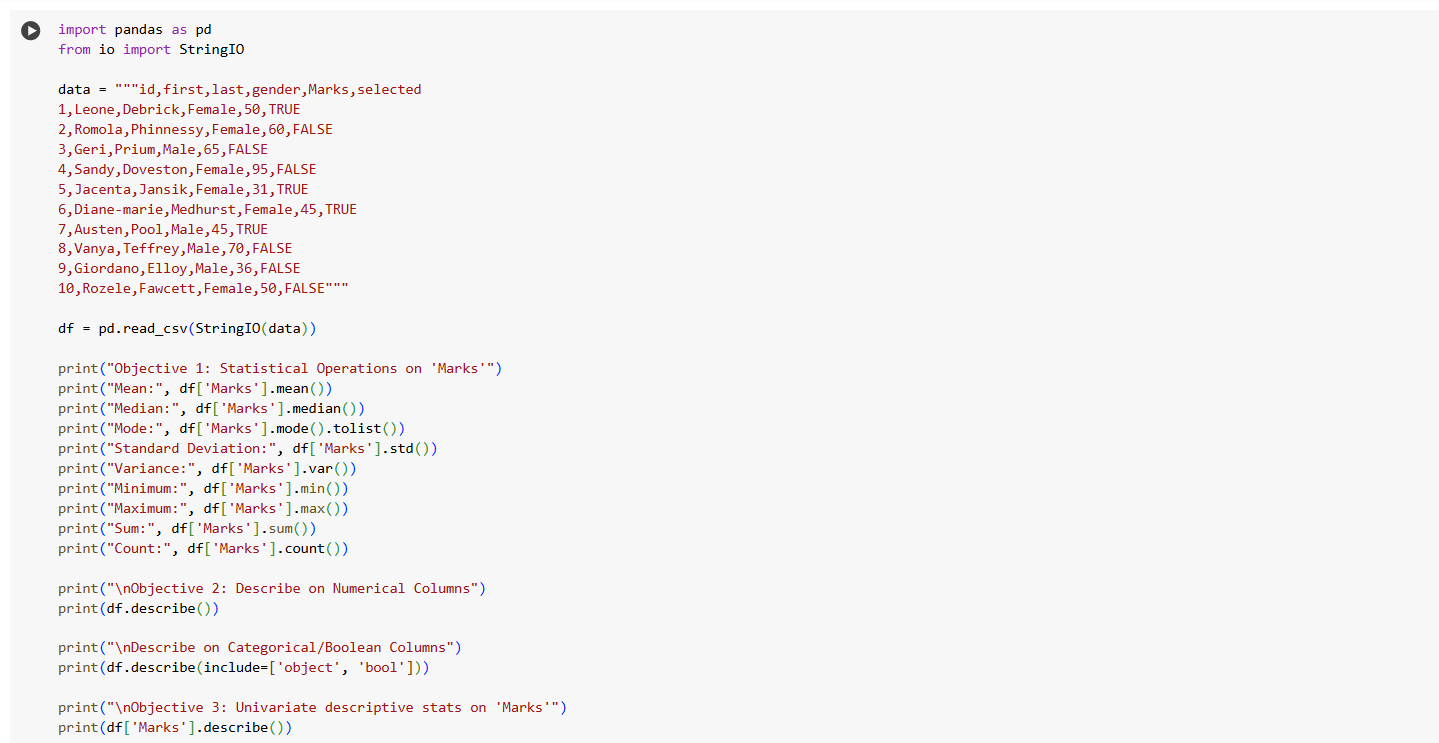
| **id** | **first** | **last** | **gender** | **Marks** | **selected** |
| --- | --- | --- | --- | --- | --- |
| **1** | **Leone** | **Debrick** | **Female** | **50** | **TRUE** |
| **2** | **Romola** | **Phinnessy** | **Female** | **60** | **FALSE** |
| **3** | **Geri** | **Prium** | **Male** | **65** | **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** |

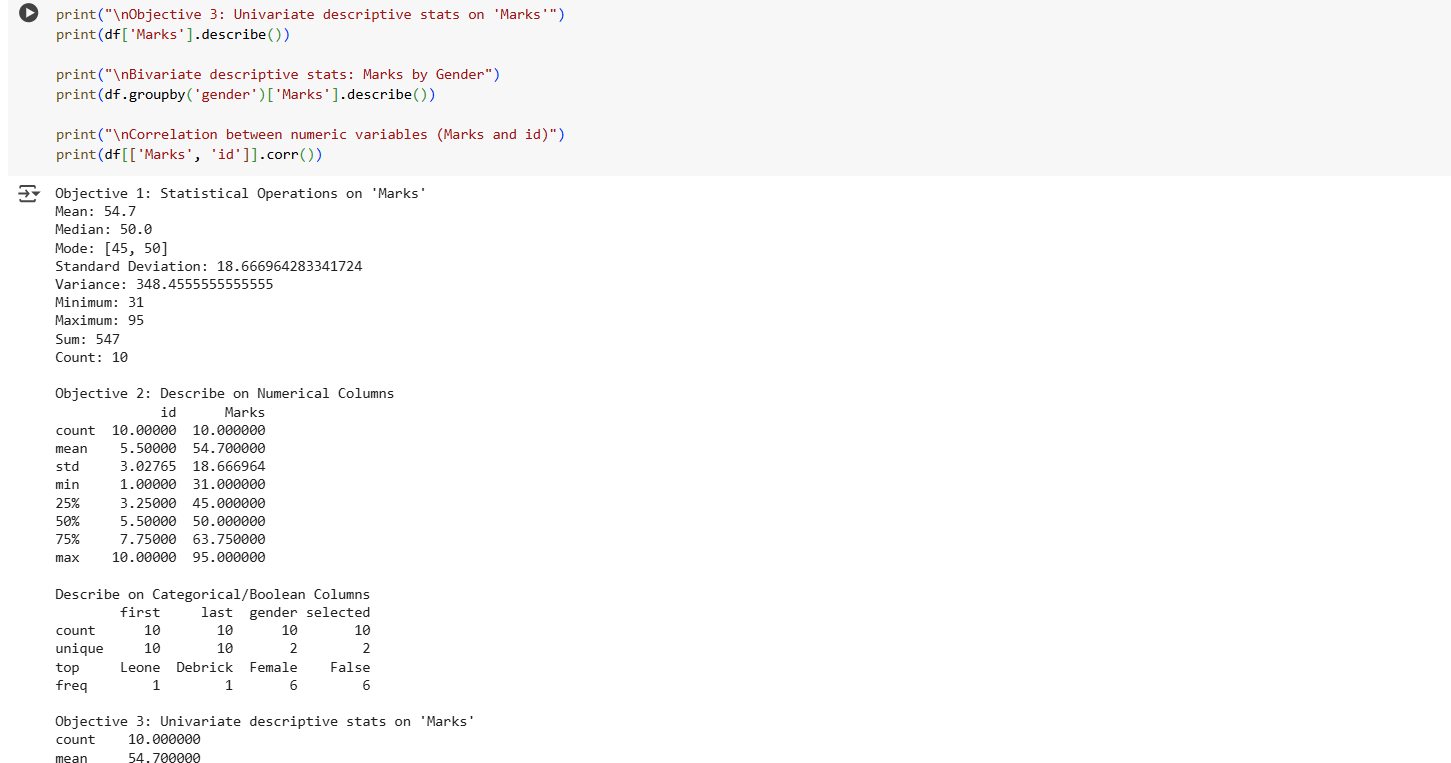
**The objectives of this experiment are:**

**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

**Reference to the Textbook and Explanation**

All the fundamentals are given in Chapter 2 and Appendix 2.

The variable in the dataset Female and Male can be changed to 0 or 1 using Label Encoder. It is done as given below:

df\_gender\_encode=LabelEncoder()

df.gender=df\_gender\_encode.fit\_transform(df.gender)

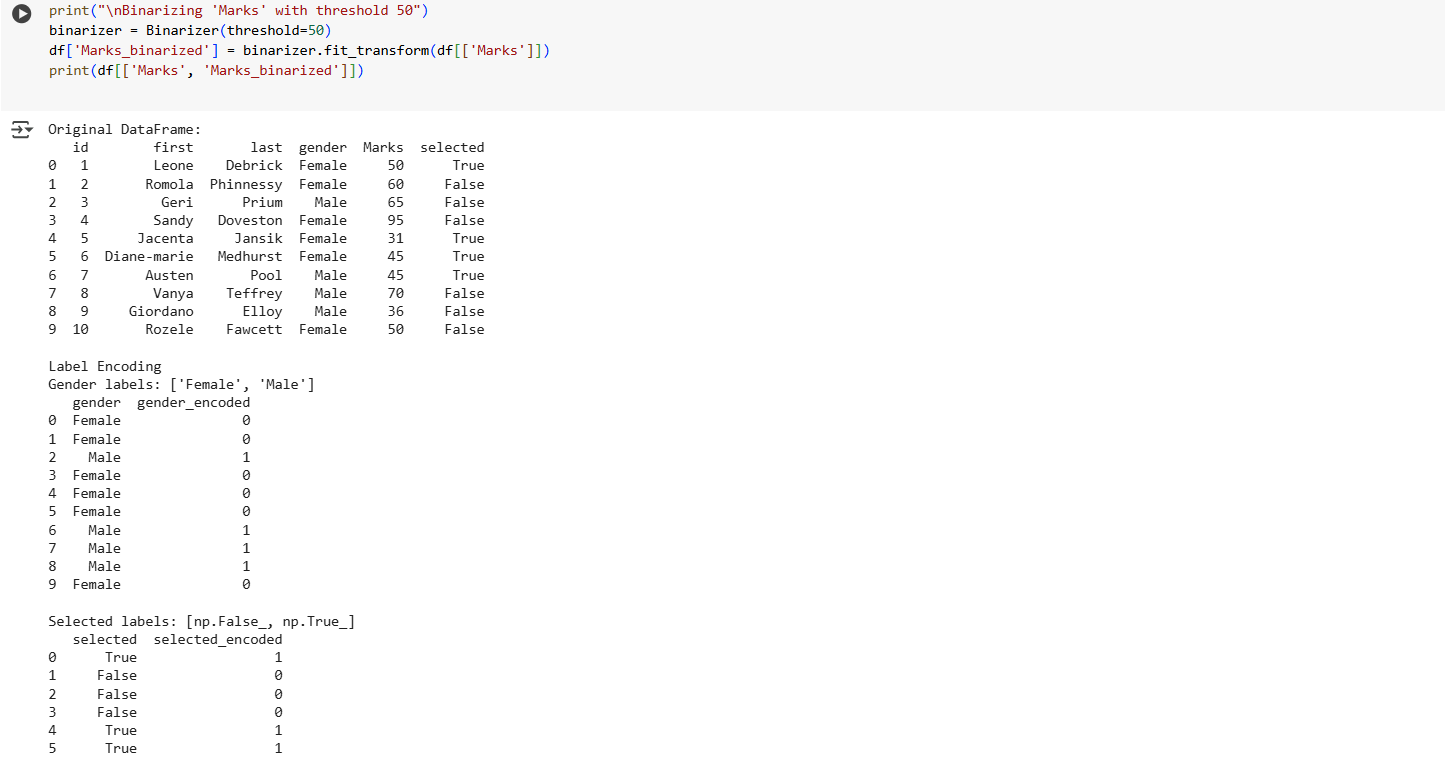
Scaling can be done as follows:

df.Marks = preprocessing.scale(df.Marks)

scaled\_df= preprocessing.scale(df.Marks)

Scaling removes the mean





Implement Dimensionality reduction using Principle component Analysis method

