

**Lab Manual**

➢ Submitted by: M ZUBAIR KHAN  
➢ Reg no.: 2023-BS-AI-006. ➢ Submitted to: Mr.Saeed

➢ Subect: Machine Learning

➢ Department: BS-AI(Section-A)

**Table of Content**

|  |  |
| --- | --- |
| **sr** | **Regression project** |
| 1 | Data Preprocessing Steps |
| 2 | Reading Data |
| 3 | Data Cleaning |
| 4 | utlier Detection and Removal |
| 5 | Normalization/Standardization |
| 6 | Standardization |
| 7 | One-Hot Encoding |
| 8 | Data Reduction |
| 9 | Handling Imbalanced Data |
| 10 | Splitting Data |
| 11 | Regression |

**Salary Dataset**

This dataset contains information about employees’ years of experience and their corresponding annual salaries. It is commonly used for simple linear regression tasks to model and predict salary based on experience.

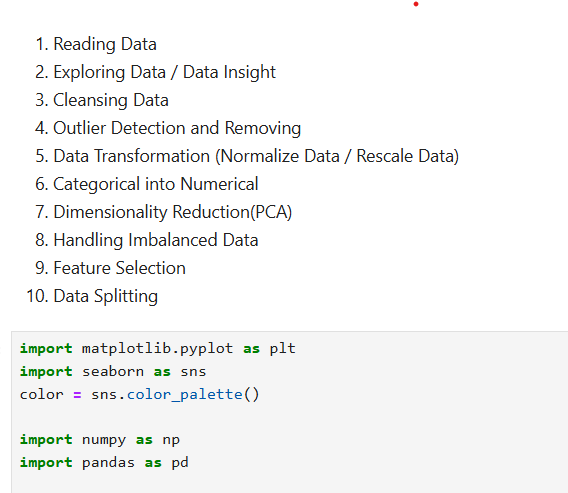
**Columns:**

* YearsExperience: Number of years the employee has worked.
* Salary: Annual salary in U.S. dollars.

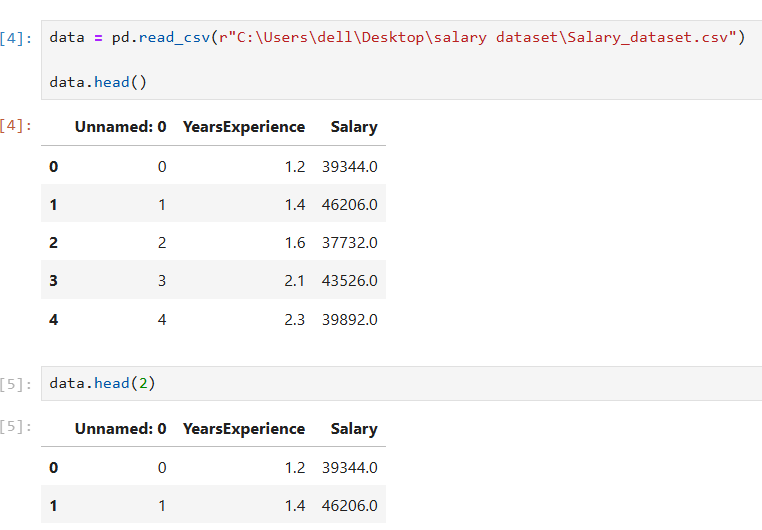
**Use Case:**

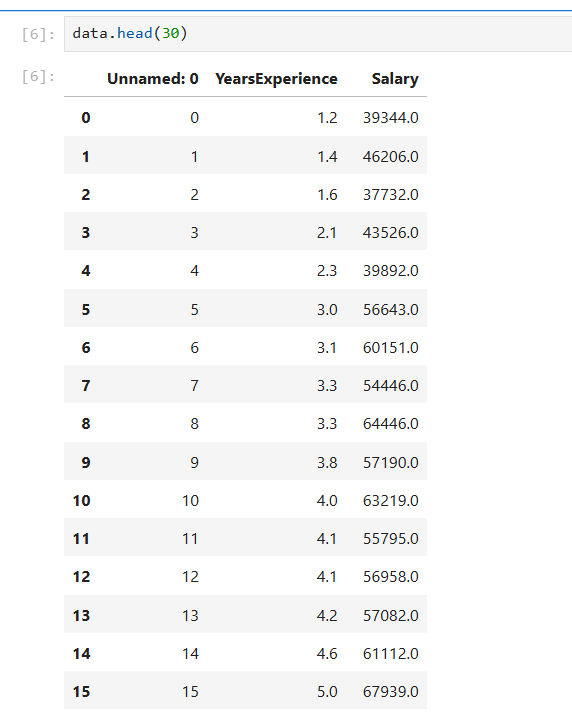
Ideal for learning regression techniques, visualizing linear relationships, and evaluating model performance using metrics like R² and MSE.

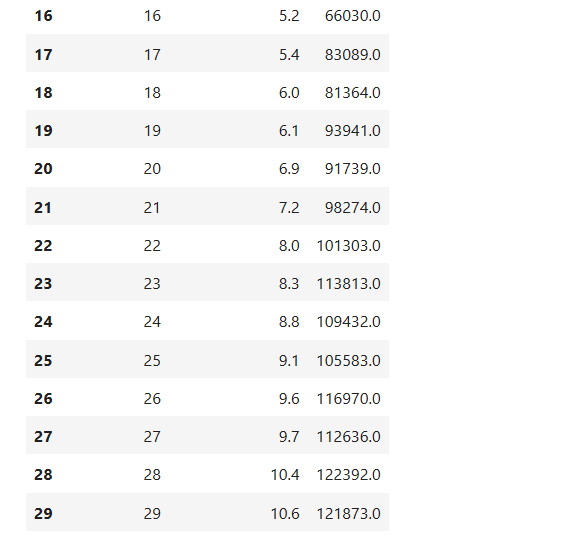
**Data Preprocessing Steps**

****

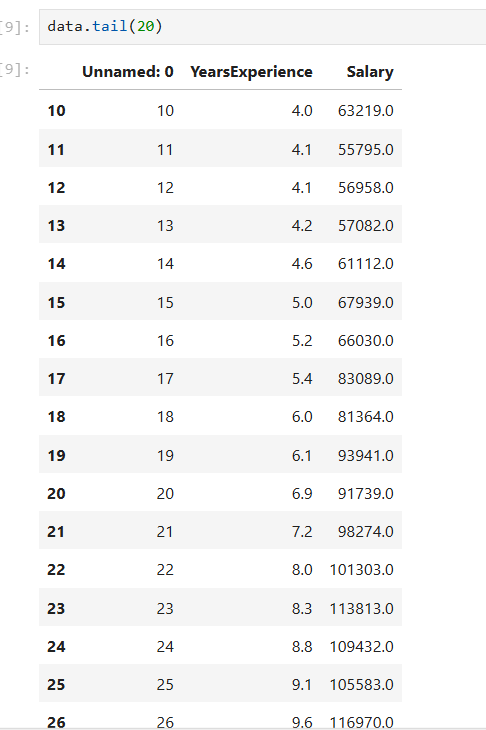
**1: Reading Data:**

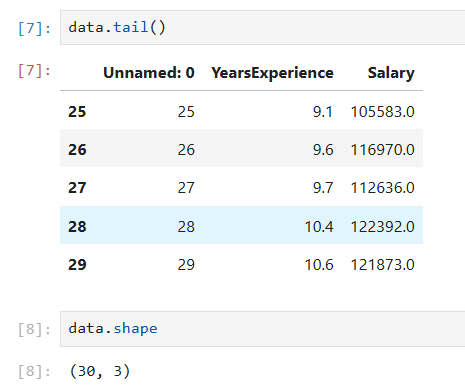
****

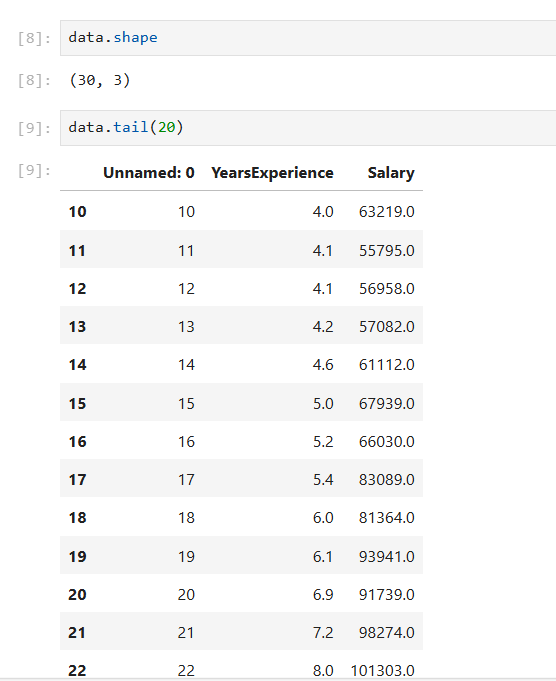
****

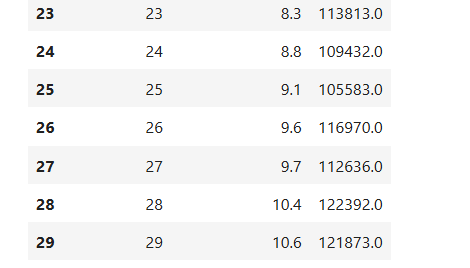
****

****

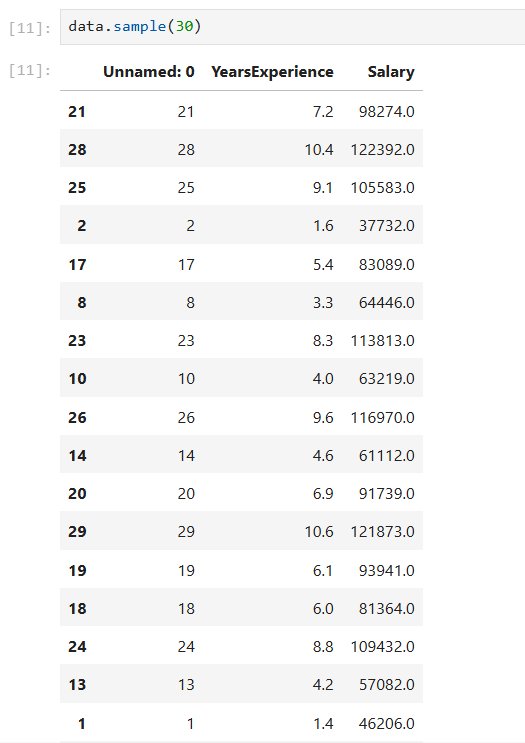
****

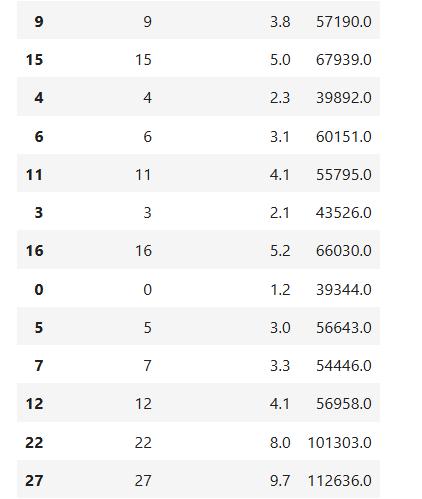


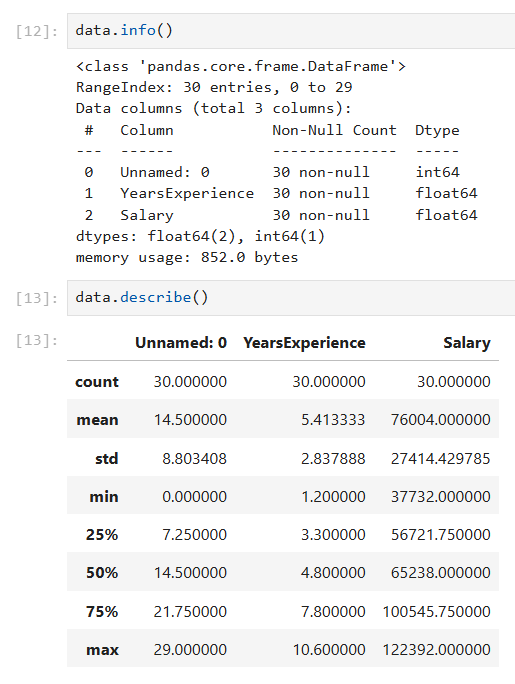










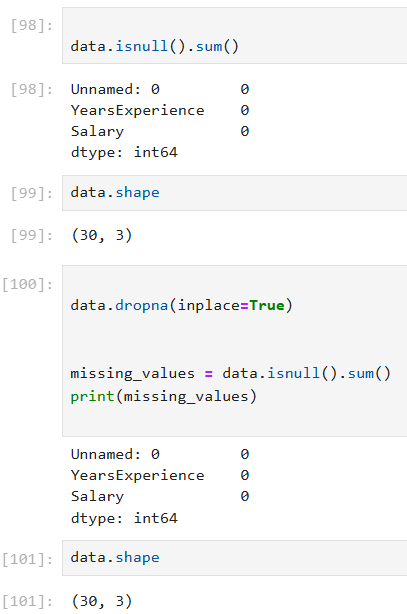


**2: Data Cleaning**

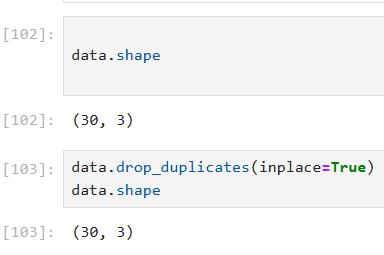
****

****

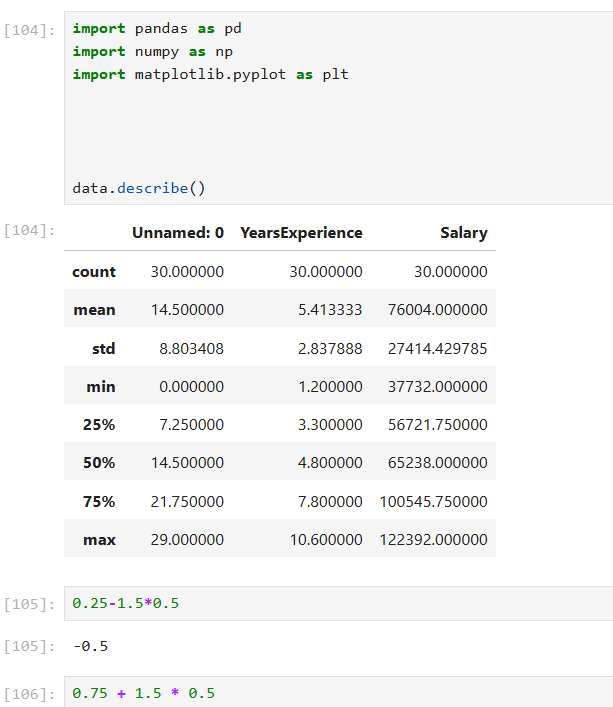
**Removal: Deleting rows with missing values.**

****

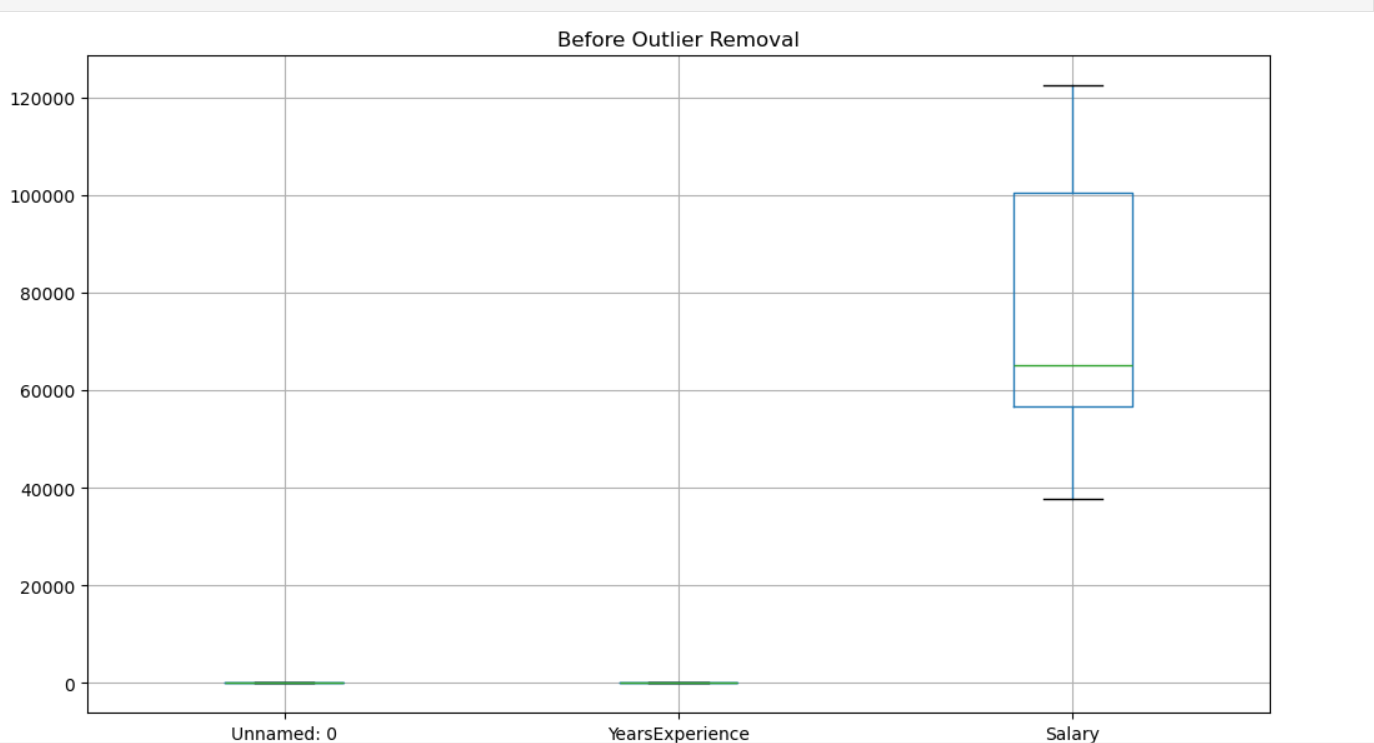
**Removing Duplicates**

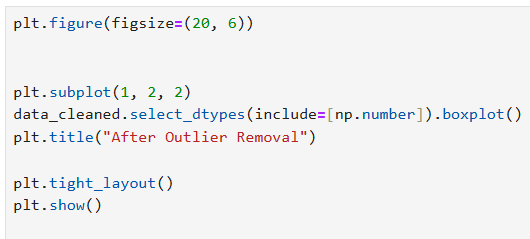
****

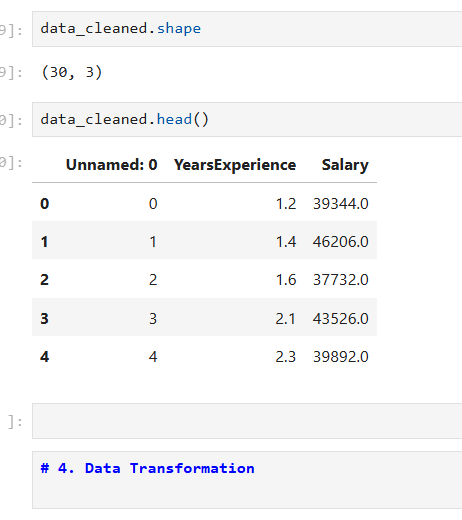
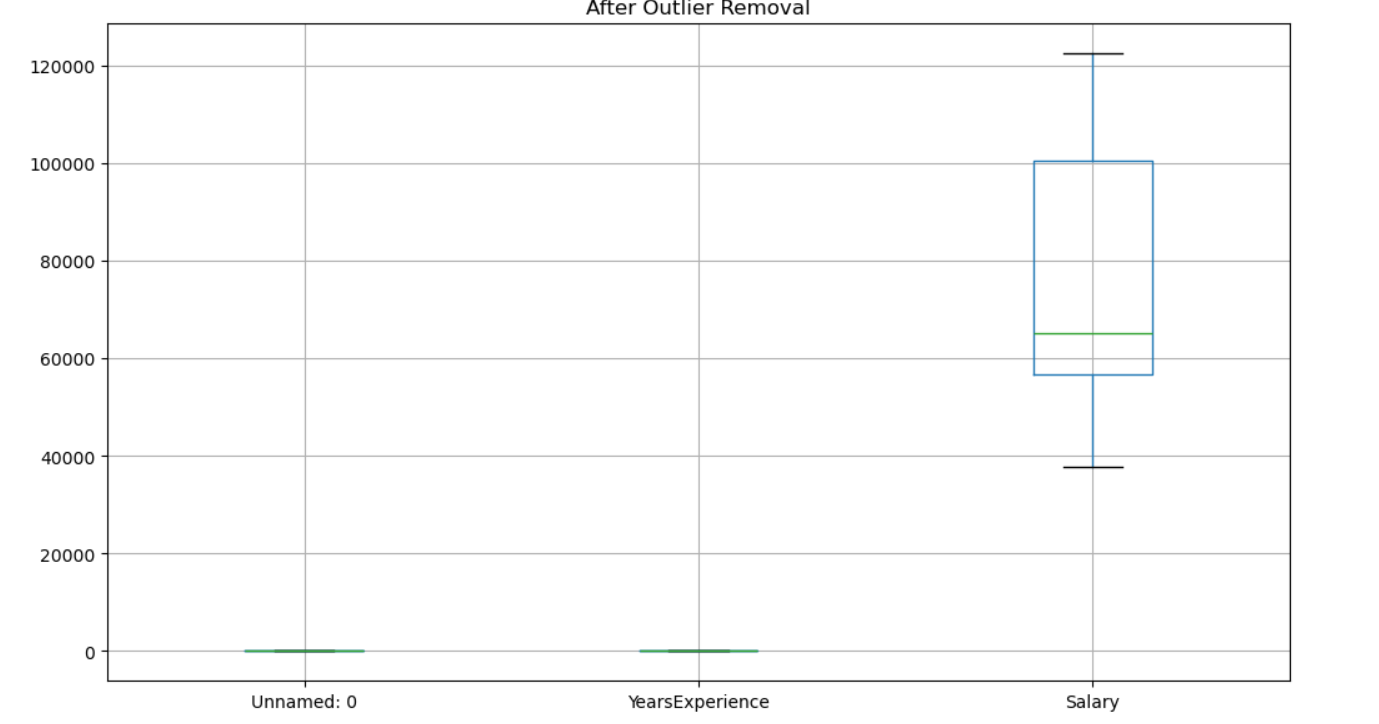
**3: Outlier Detection and Removal**

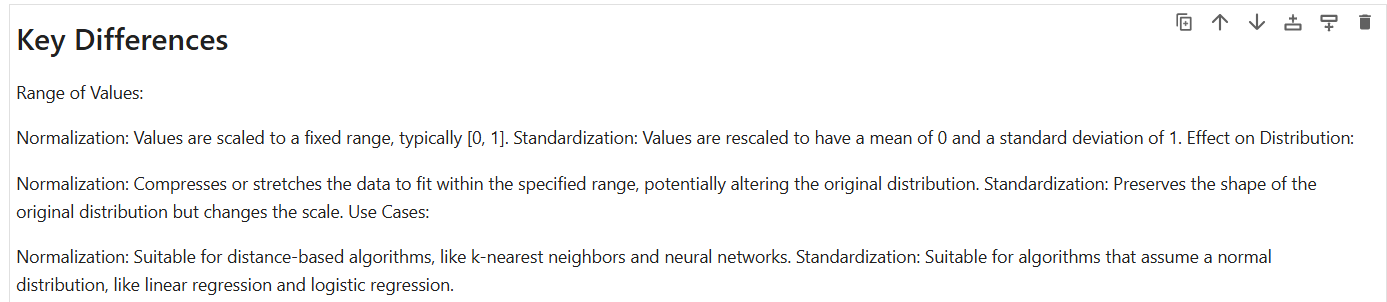
****

****

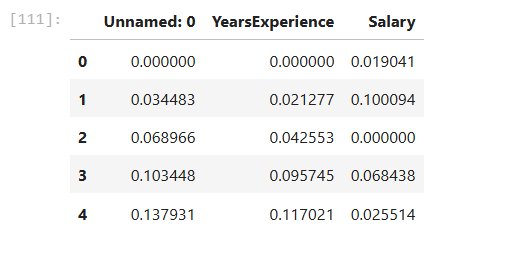
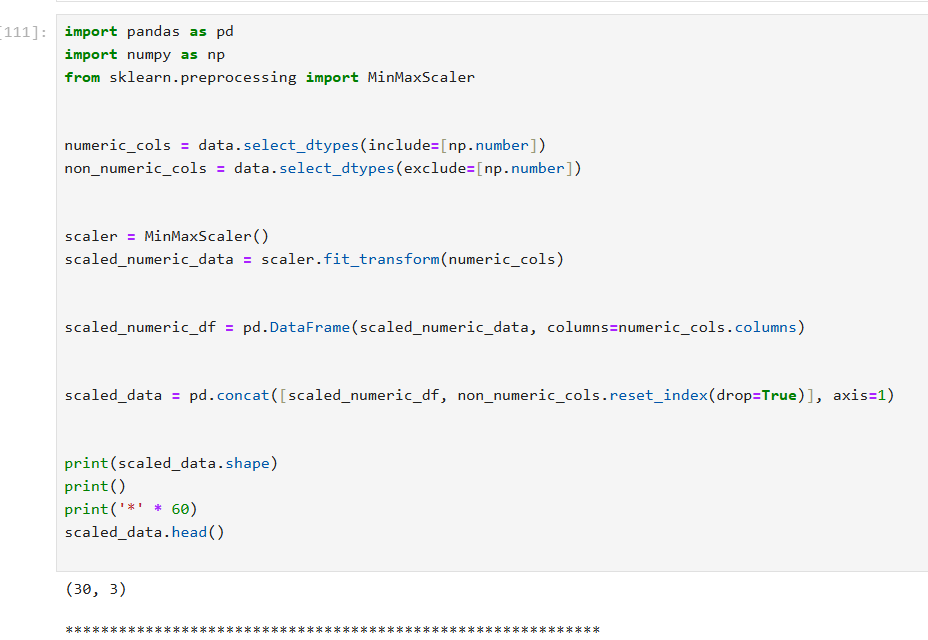
****

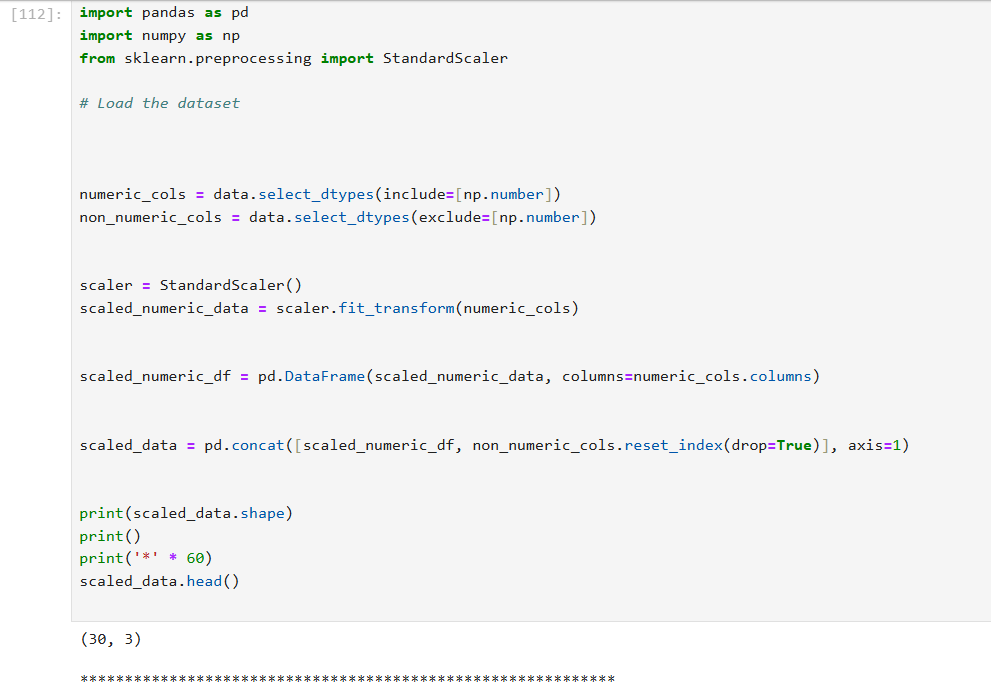
****

****

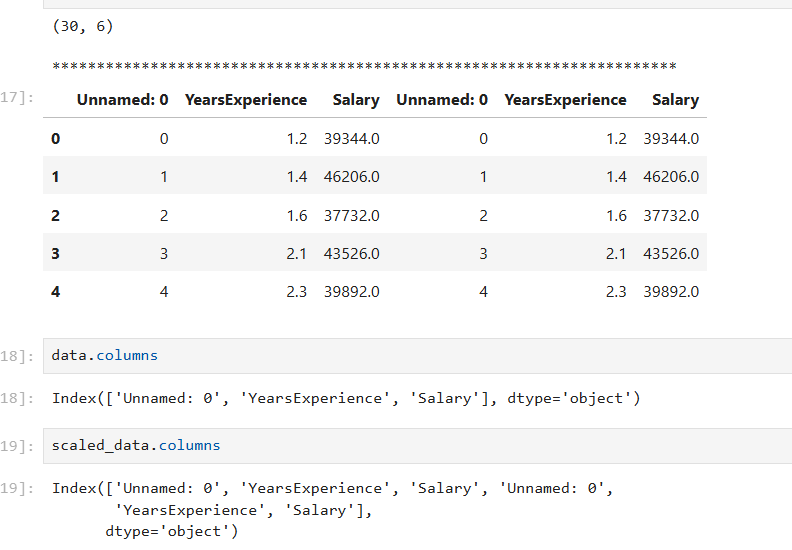
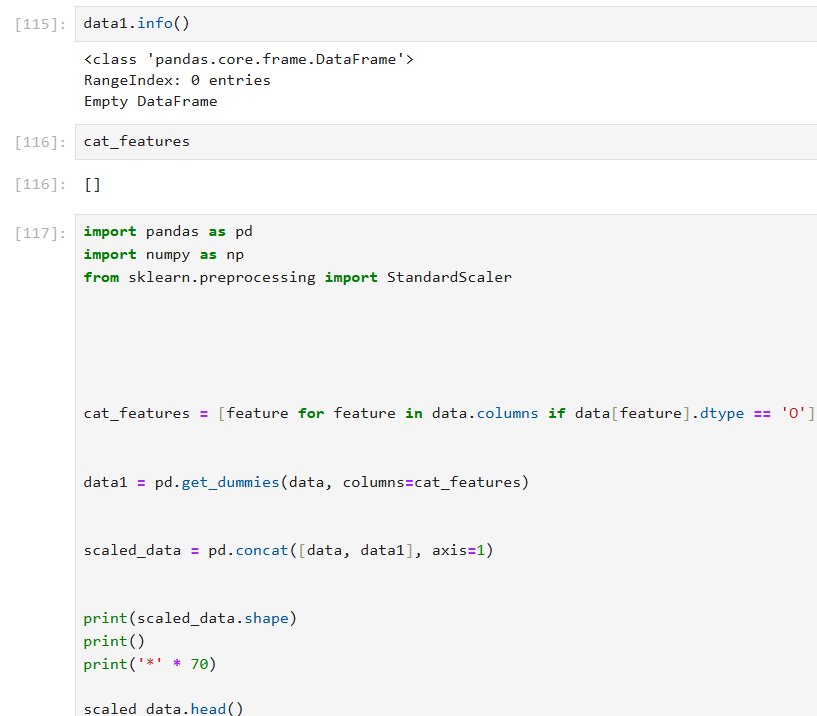
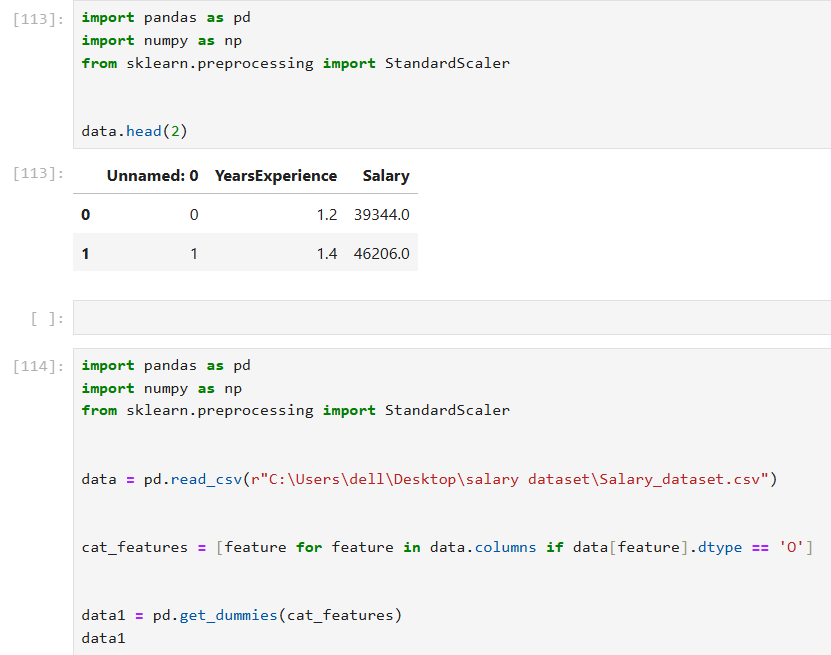
****

**Normalization/Standardization**

**Standardization**

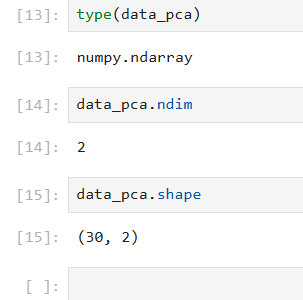
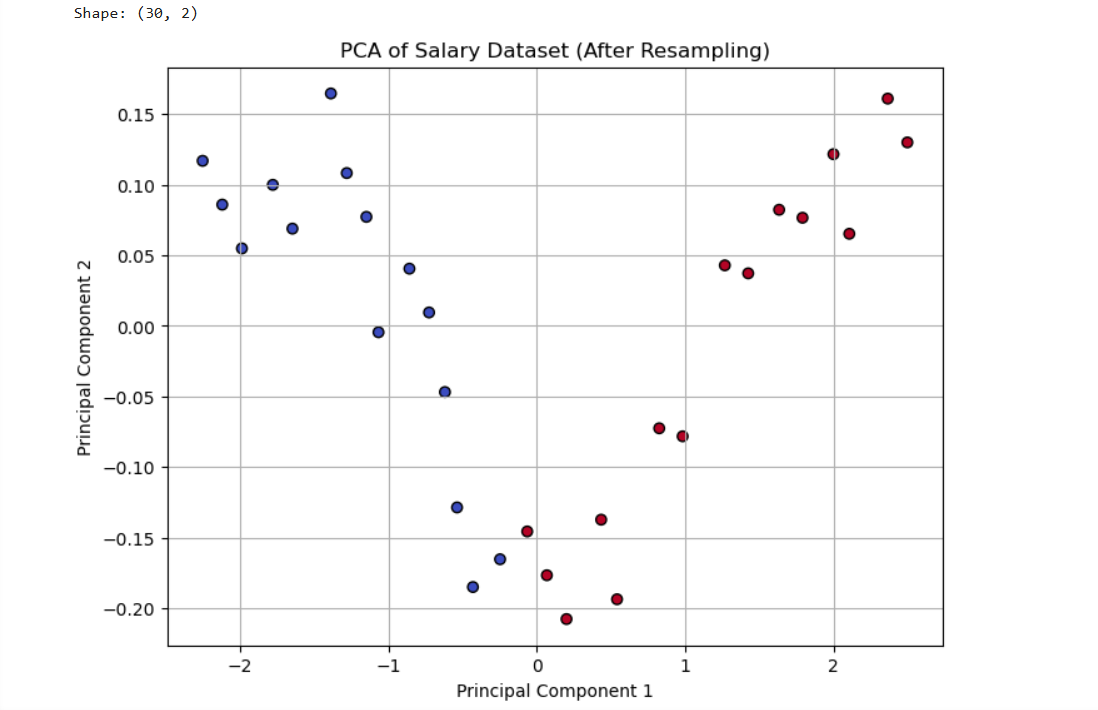
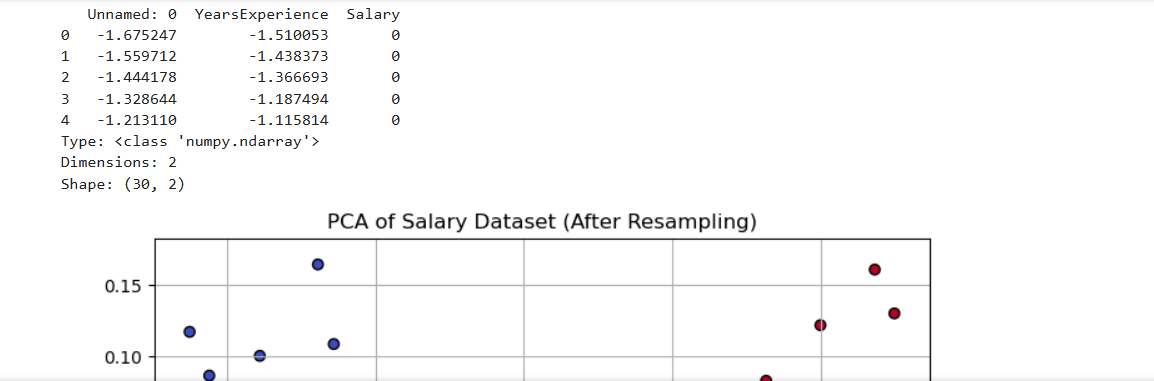
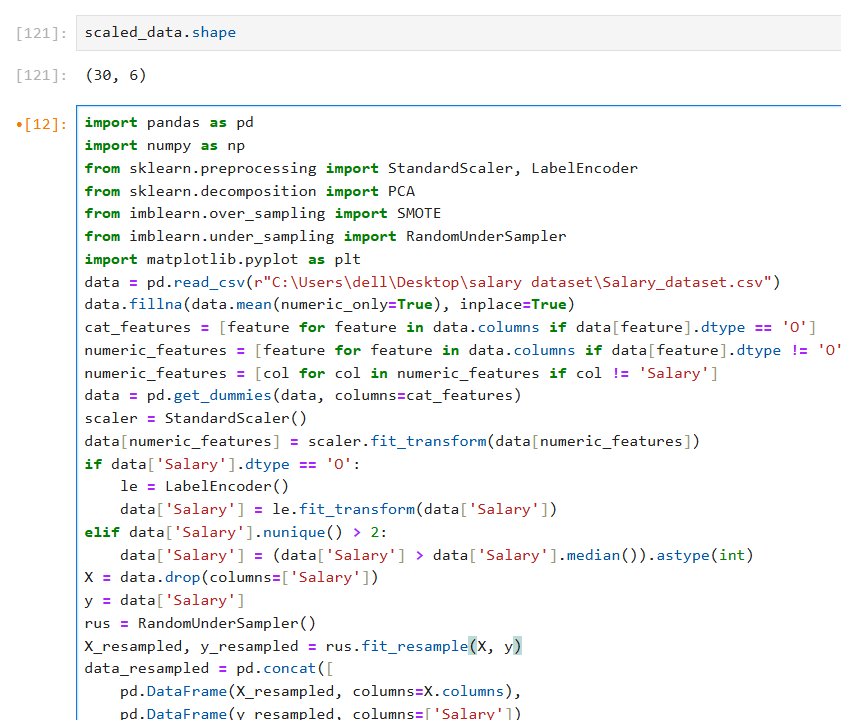
****

**5: One-Hot Encoding**

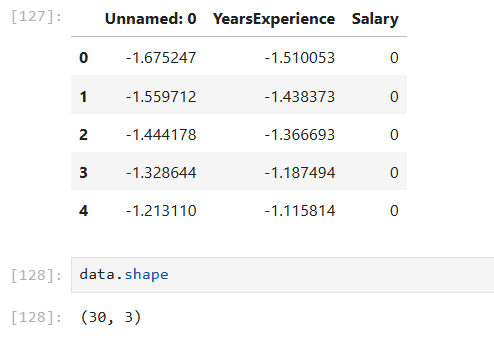
****

**6: Data Reduction**

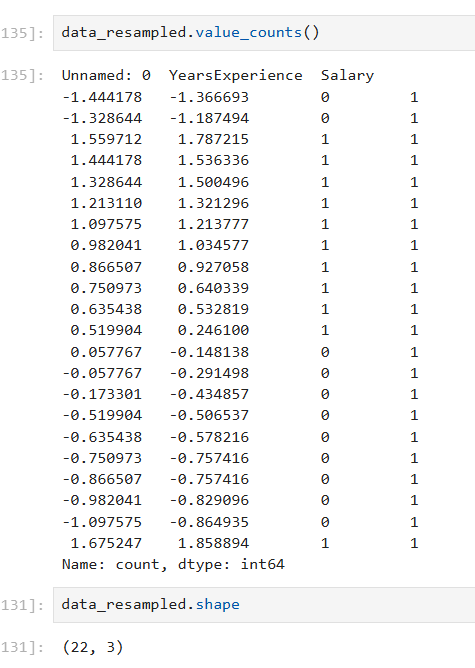
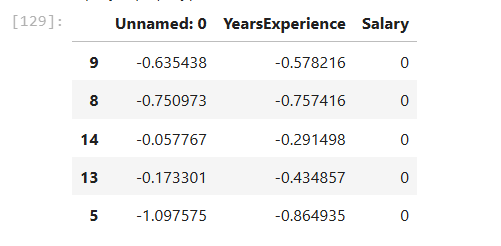
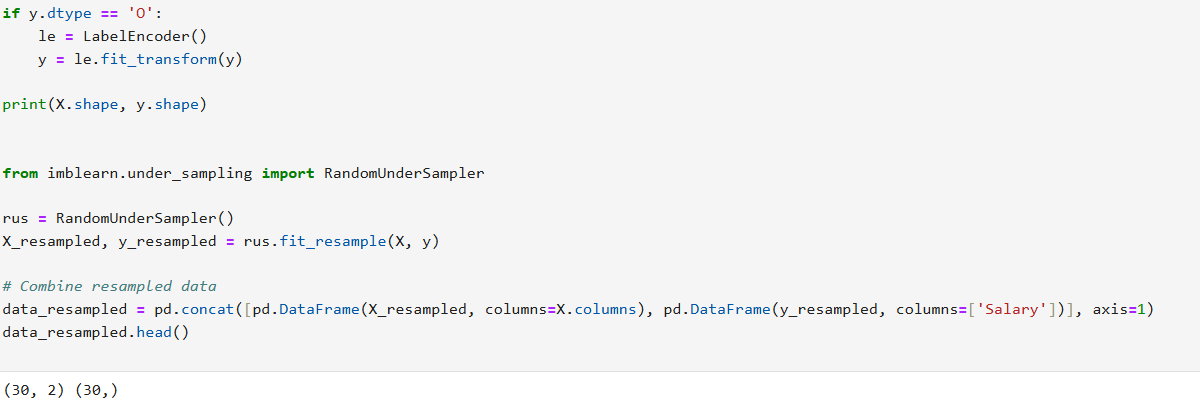
**Dimensionality Reduction**

****

**7: Handling Imbalanced Data**

****

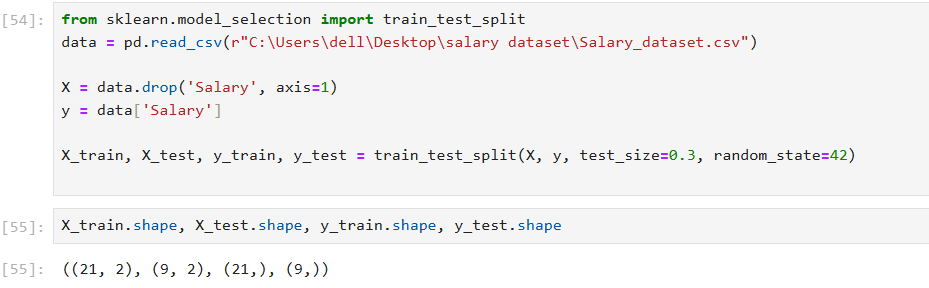
**Undersampling**

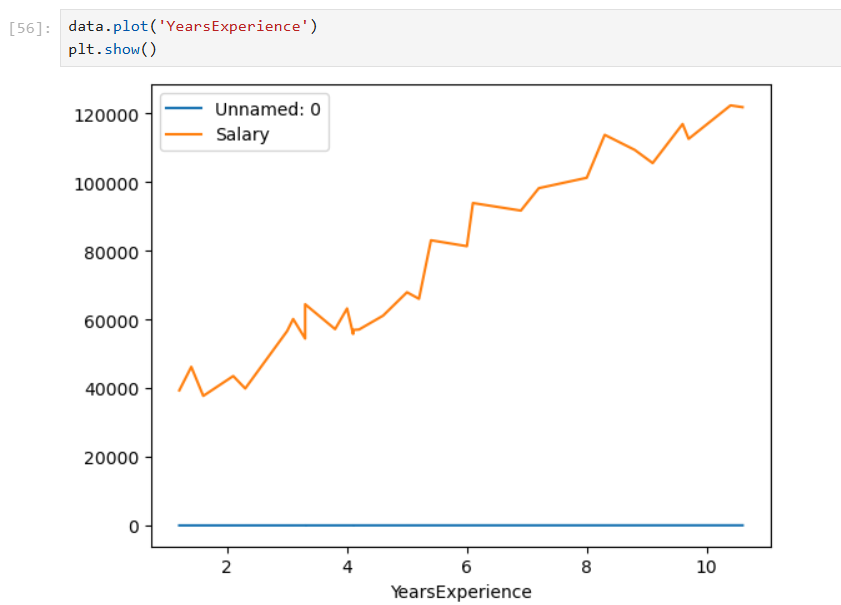
****

**Target Encoder**

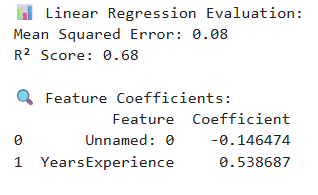
****

**8: Splitting Data**

****

****

**Regression**

****

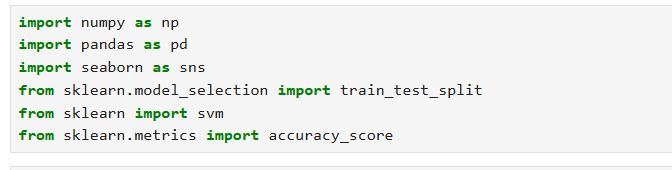
**Prediction**

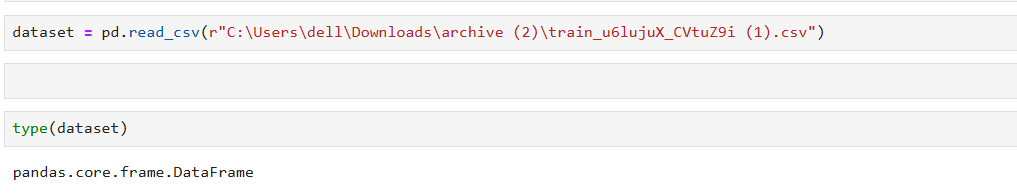
****

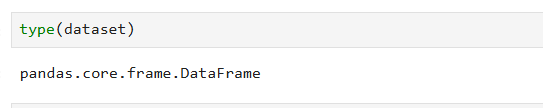
**Project #02**

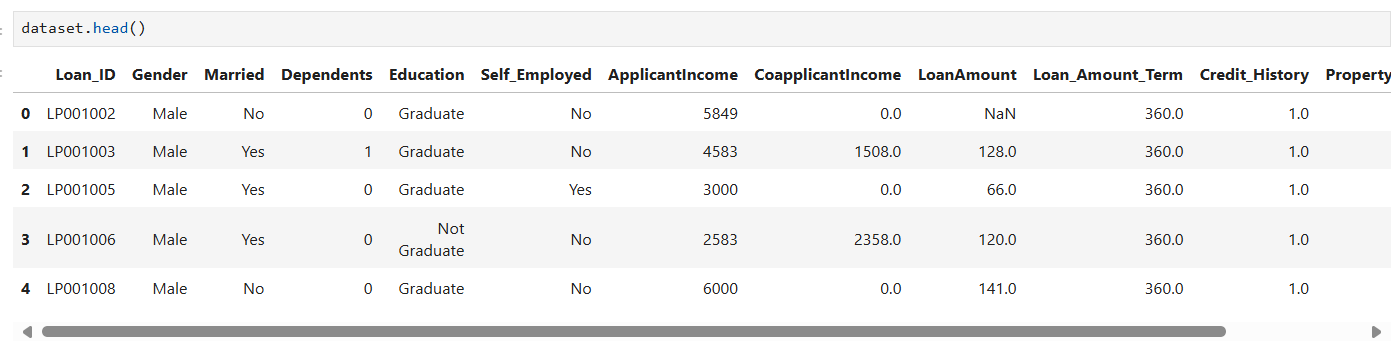
**Classification**

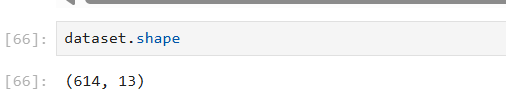
**Import libraries:**

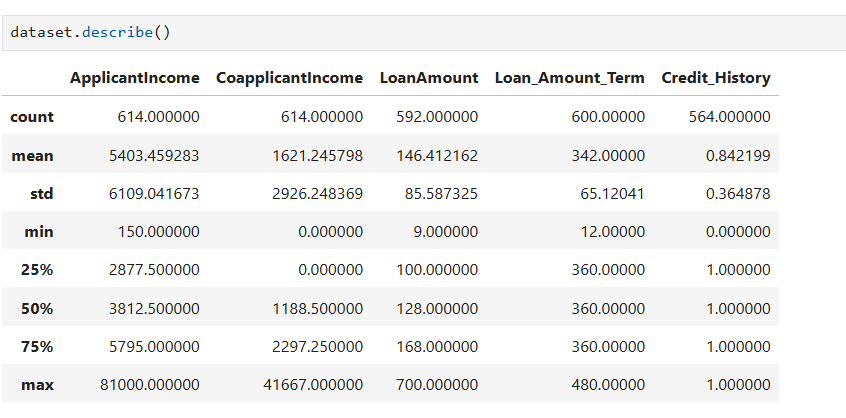
****

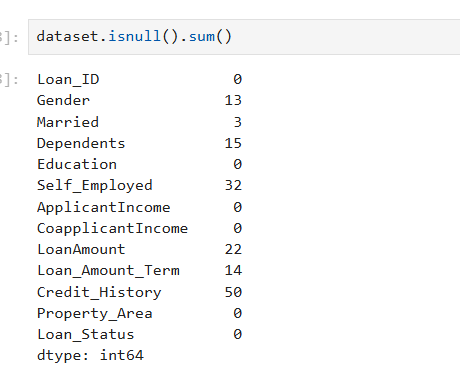
****

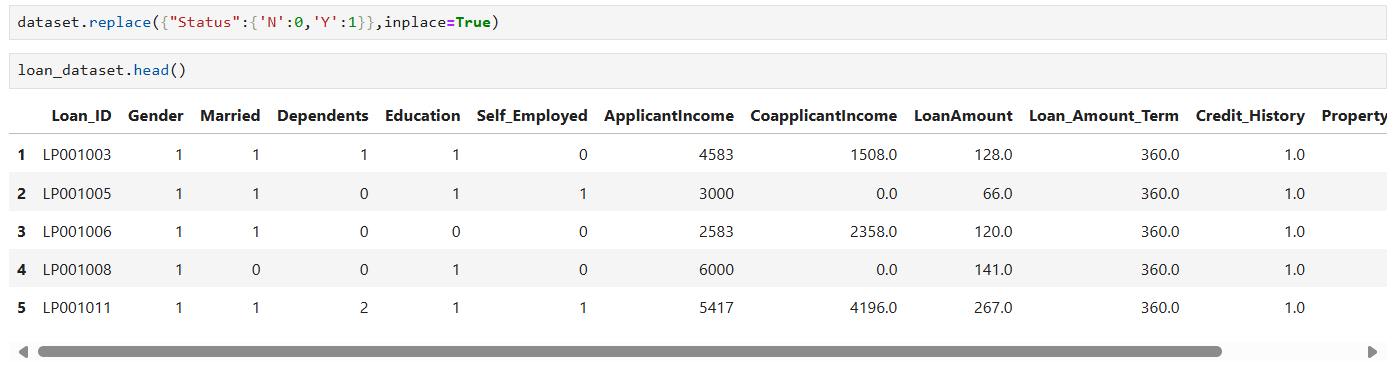
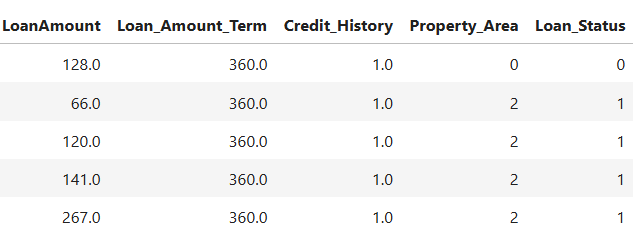
****

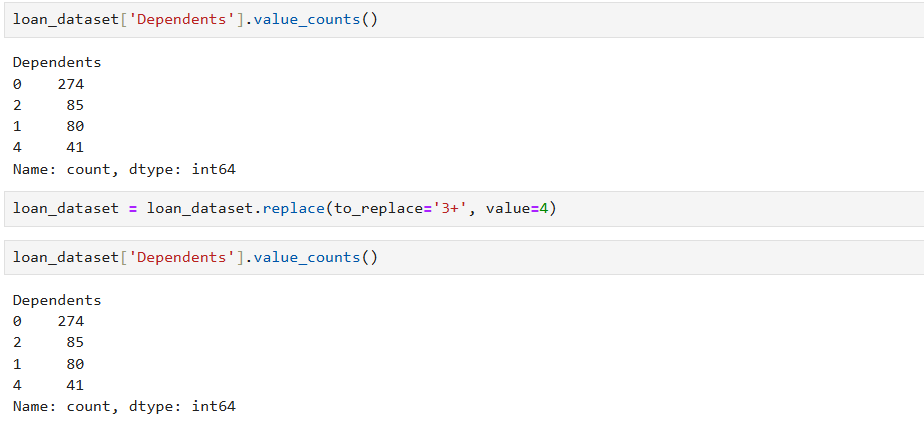
****

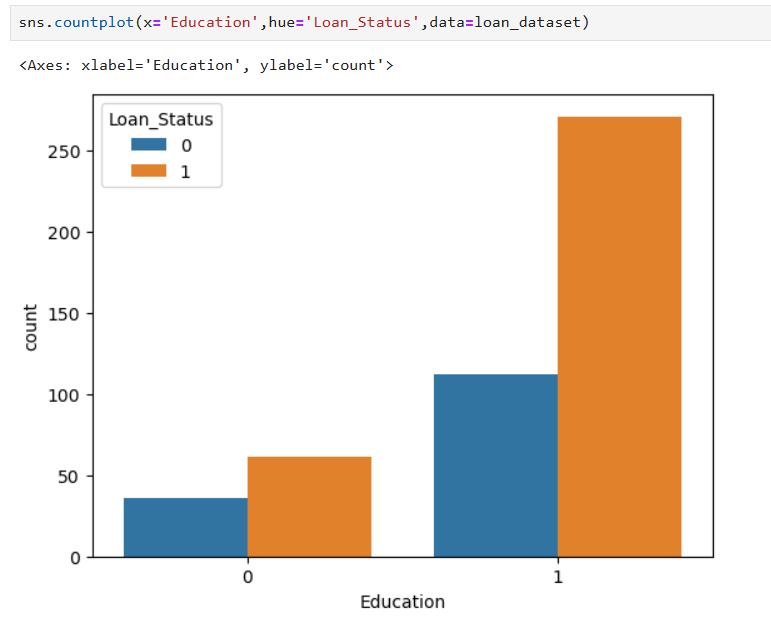
****

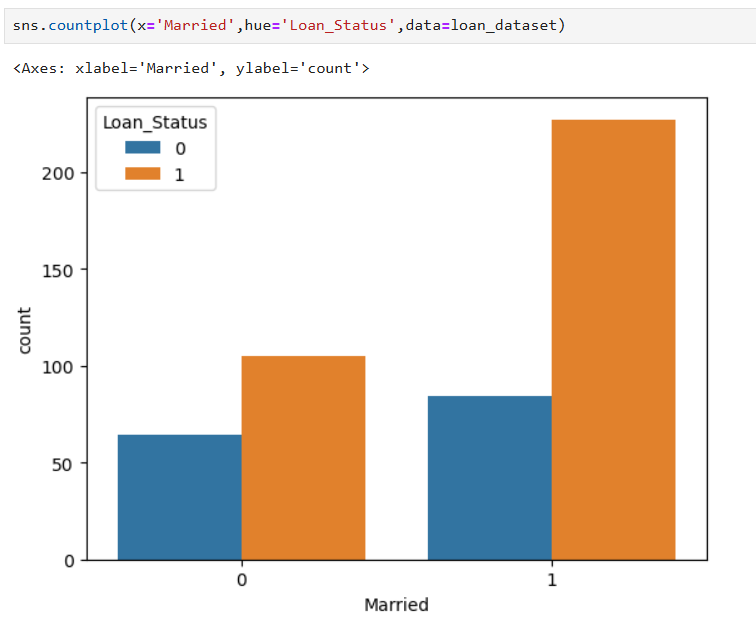
****

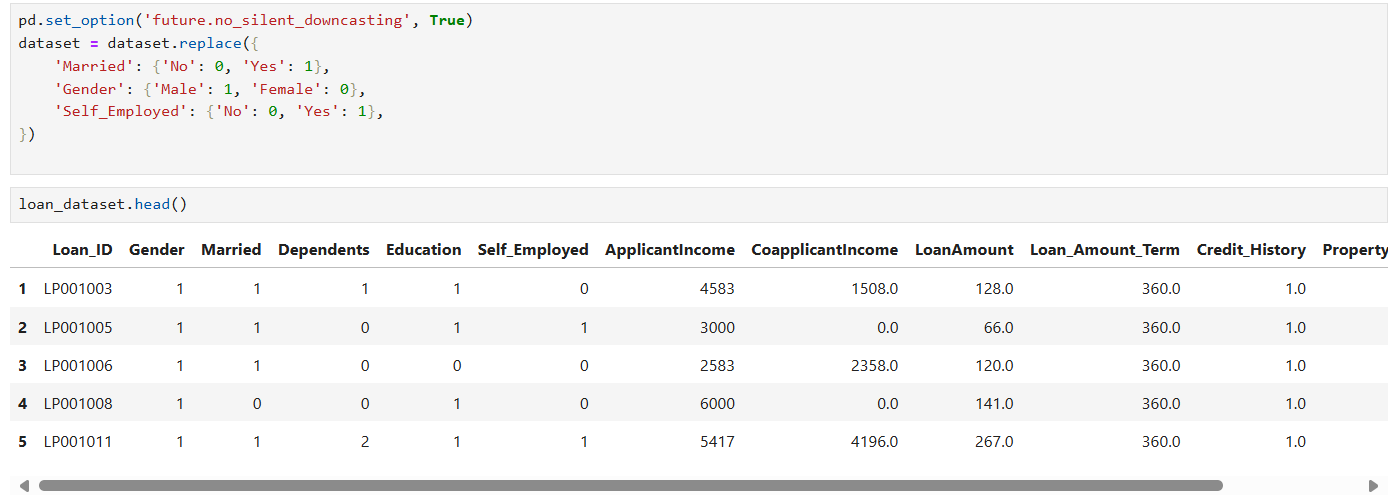
****

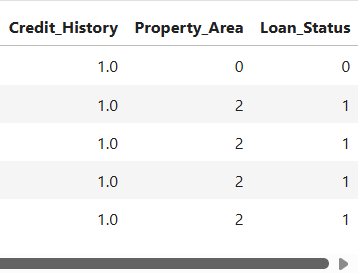
**  
**

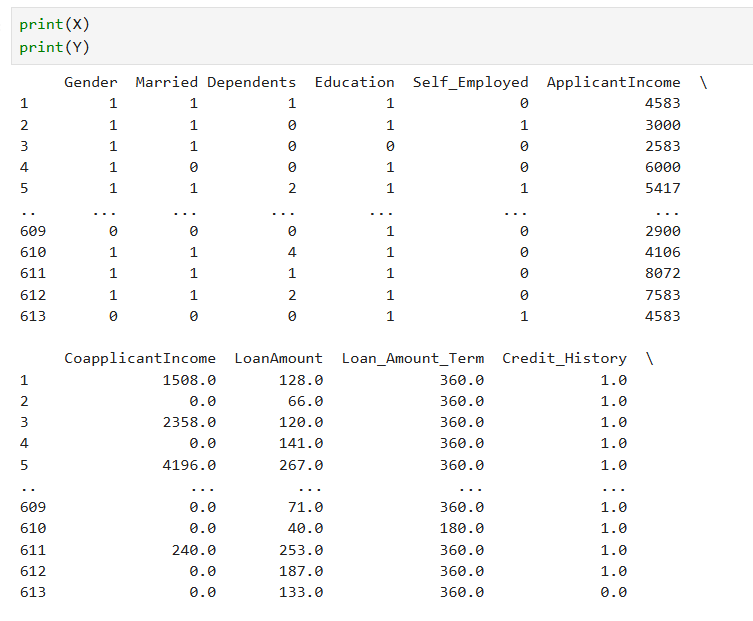
****

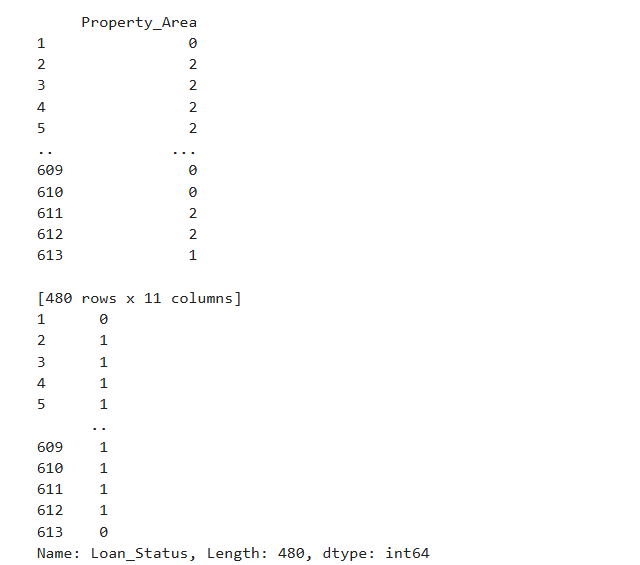
****

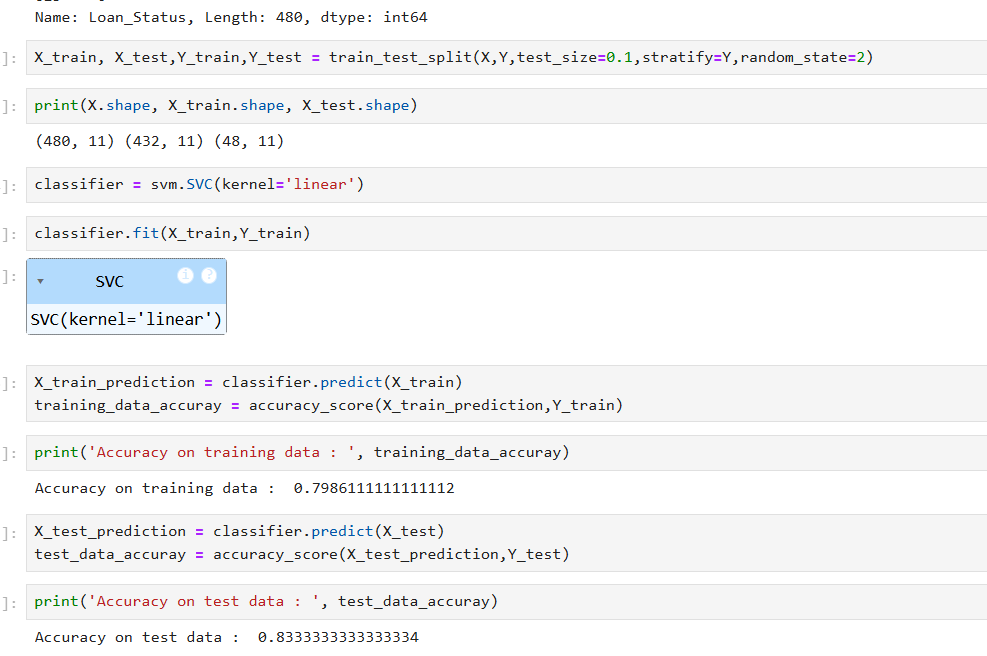
****

****

****

****

****

****