**LAB 3: FILLING OF MISSING VALUE WITH TITANIC DATA SET**

**Theory:**

**Pandas:** A data manipulation and analysis library, providing high-performance, easy-to-use data structures. Primarily used for handling and processing structured data like CSV files.

**NumPy:** Essential for numerical computing in Python, offering support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.

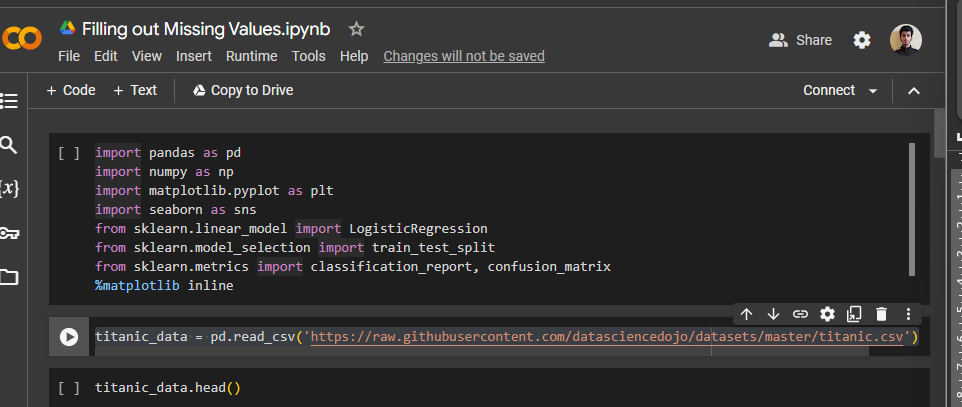
**Matplotlib:** A plotting library for creating static, interactive, and animated visualizations in Python. It is highly customizable and used for drawing various types of graphs and charts.

**Seaborn:** A statistical data visualization library built on top of Matplotlib, facilitating the creation of informative and attractive statistical graphics.

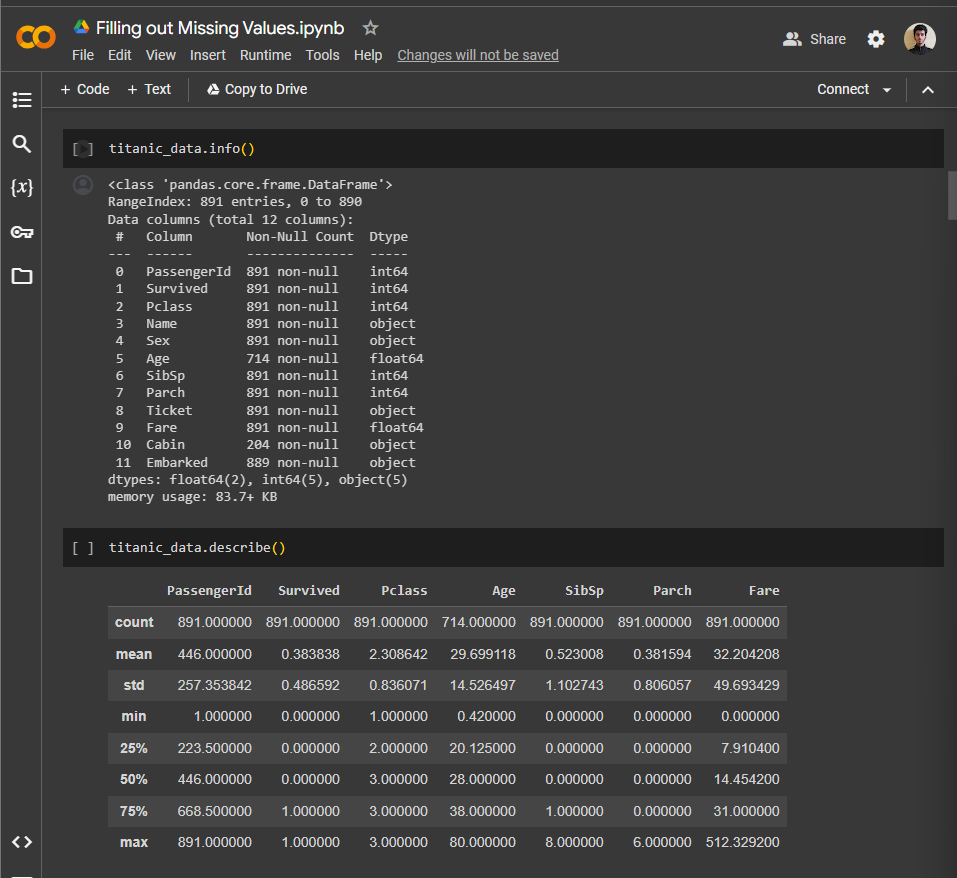
**Scikit-Learn (sklearn):** A tool for data mining and data analysis. It includes various machine learning algorithms for classification, regression, and clustering, alongside utilities for model fitting, data preprocessing, and evaluation.

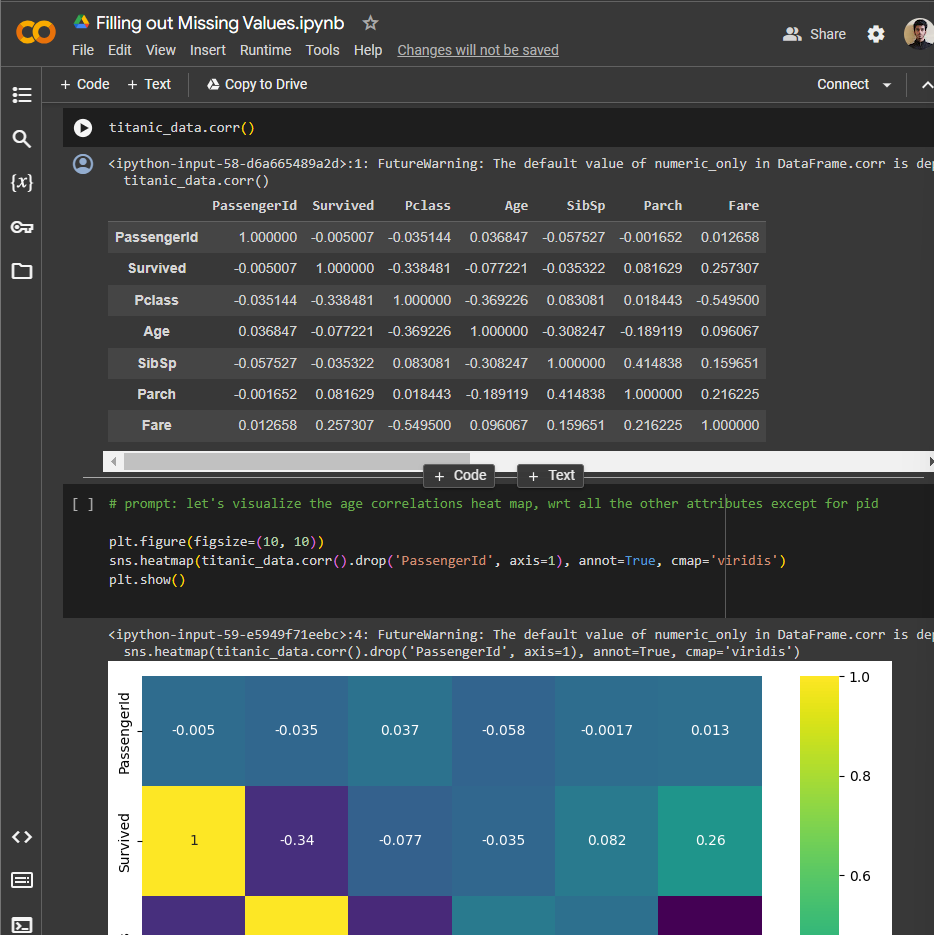
**Code:**

**Importing libraries:**

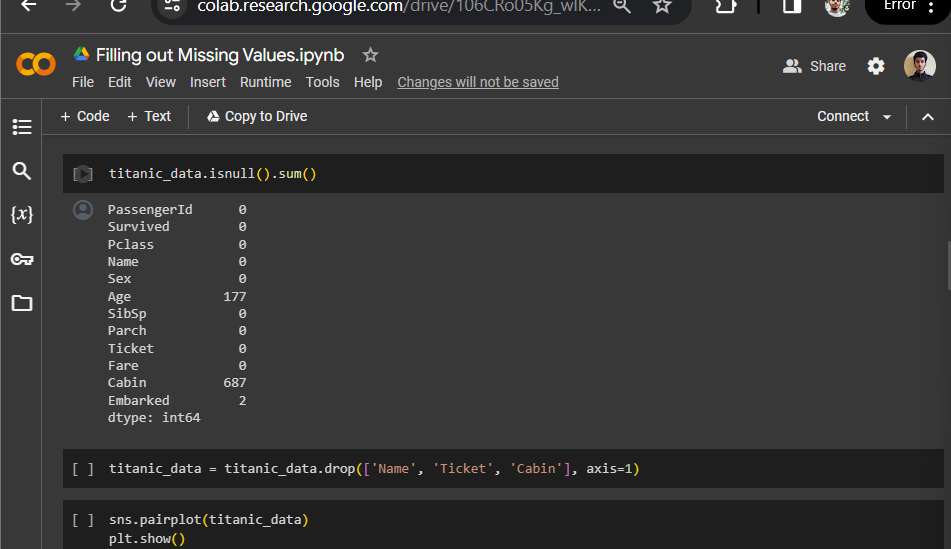
****

**Exploratory data analysis**

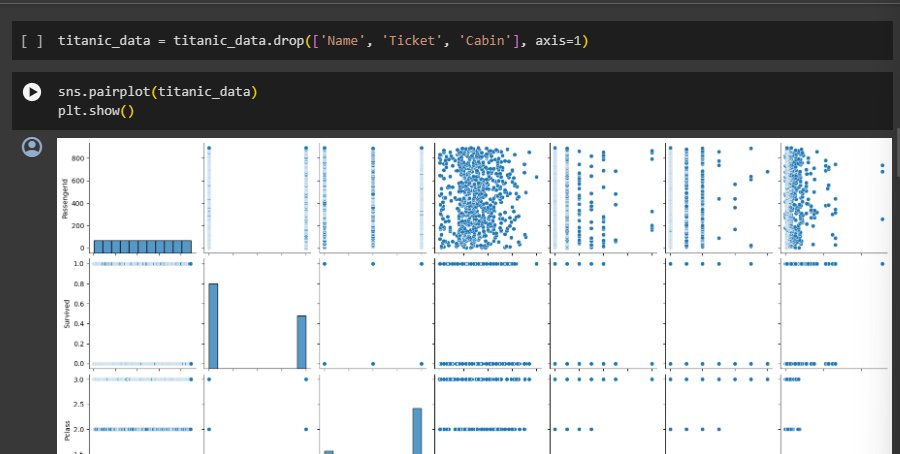
****

****

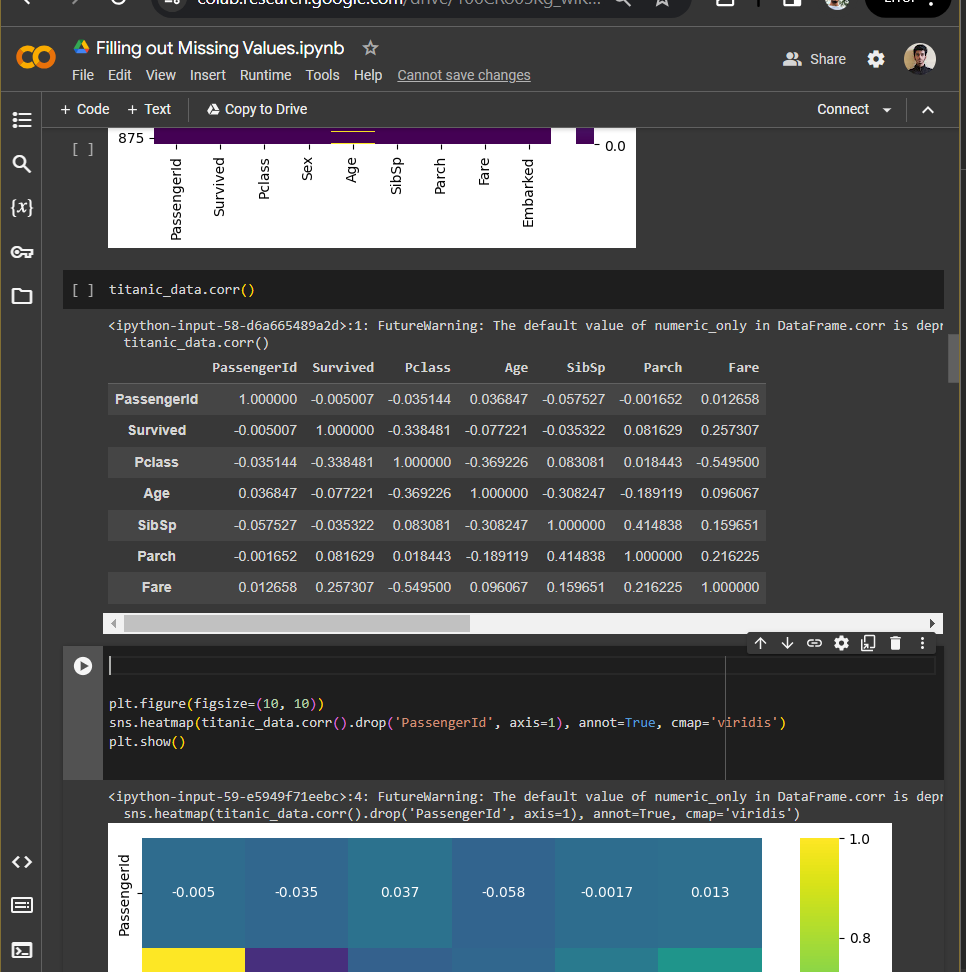
**Checking the null value:**

****

**Dropping unnecessaries columns and visualization:**

****

**Correlation analysis:**

****

It represents the Pearson correlation coefficients between different variables. Each cell in the matrix shows the correlation between two variables. The value of the correlation coefficient ranges from -1 to 1, where:

* 1 indicates a perfect positive correlation.
* -1 indicates a perfect negative correlation.
* 0 indicates no correlation.

Let's interpret the correlations:

* PassengerId: Appears to have very little correlation with other variables, as indicated by values close to 0. This suggests PassengerId is just an identifier.
* Survived:

Negative Correlation with Pclass (-0.338481): Higher class (lower Pclass number) passengers had a higher chance of survival. This could be due to better access to lifeboats in higher classes.

Positive Correlation with Fare (0.257307): Passengers who paid more (higher fares) had a higher survival rate, which might be related to higher-class passengers having a better chance of survival.

* Pclass (Passenger Class):

Negative Correlation with Age (-0.369226): Younger passengers tended to be in lower classes.

Negative Correlation with Fare (-0.549500): Higher classes (lower Pclass number) correspond to higher fares.

* Age:

Negative Correlation with SibSp (-0.308247): Older passengers were likely to have fewer siblings or spouses aboard.

Negative Correlation with Parch (-0.189119): Older passengers were less likely to travel with parents or children.

* SibSp (Siblings/Spouses aboard):

Positive Correlation with Parch (0.414838): Passengers with siblings or spouses aboard were more likely to also have parents or children aboard.

* Parch (Parents/Children aboard):

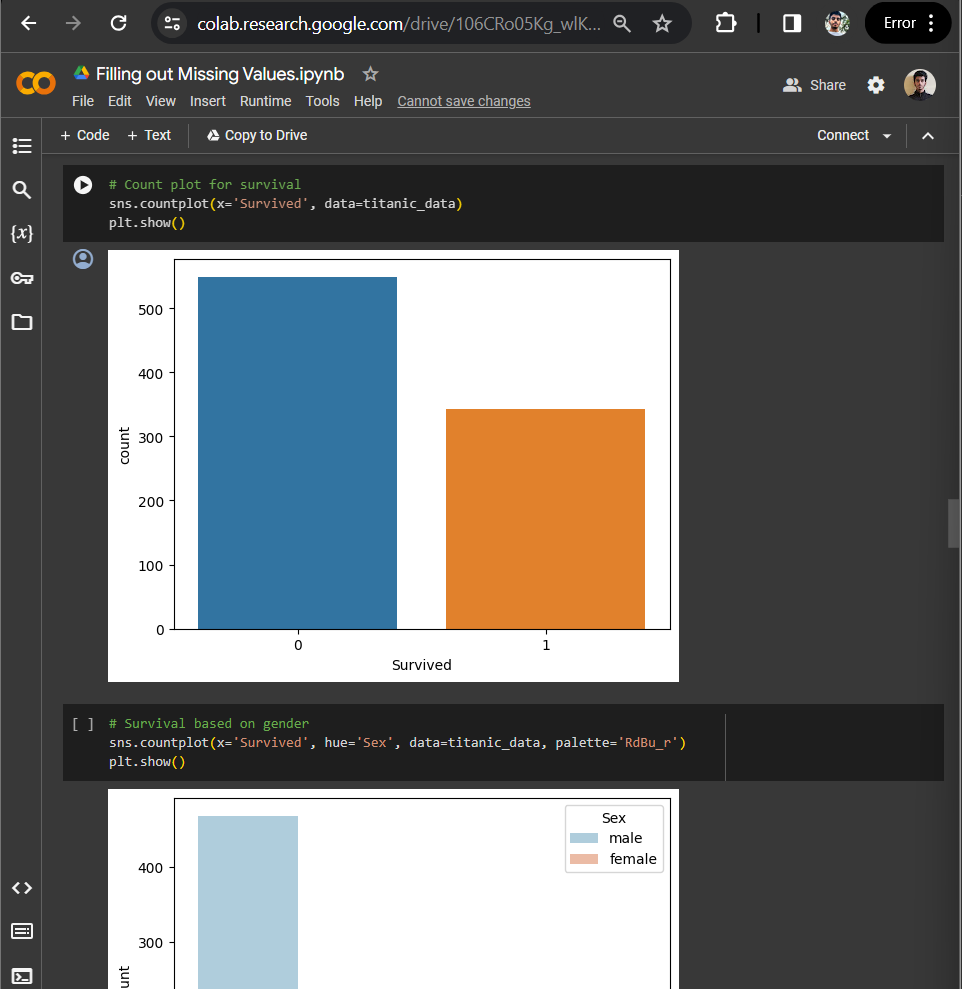
Positive Correlation with Fare (0.216225): Passengers with parents or children aboard tended to pay higher fares, possibly indicating larger cabins or higher classes.

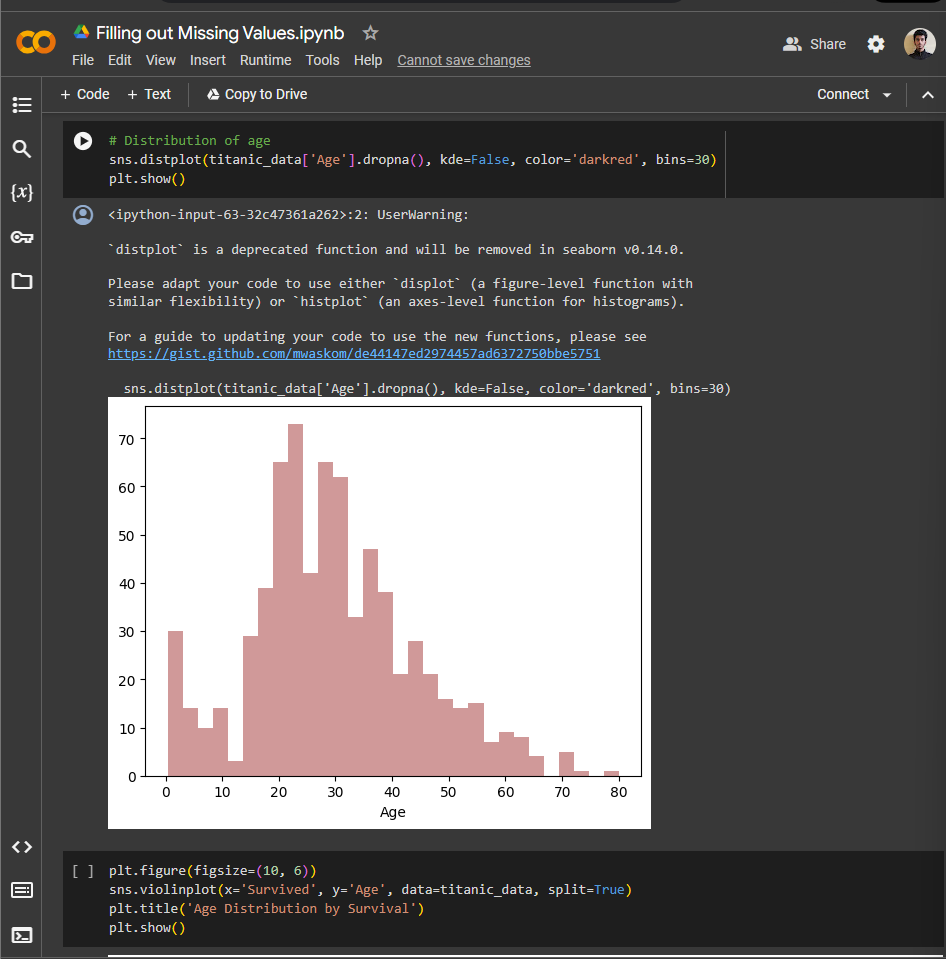
* Fare:

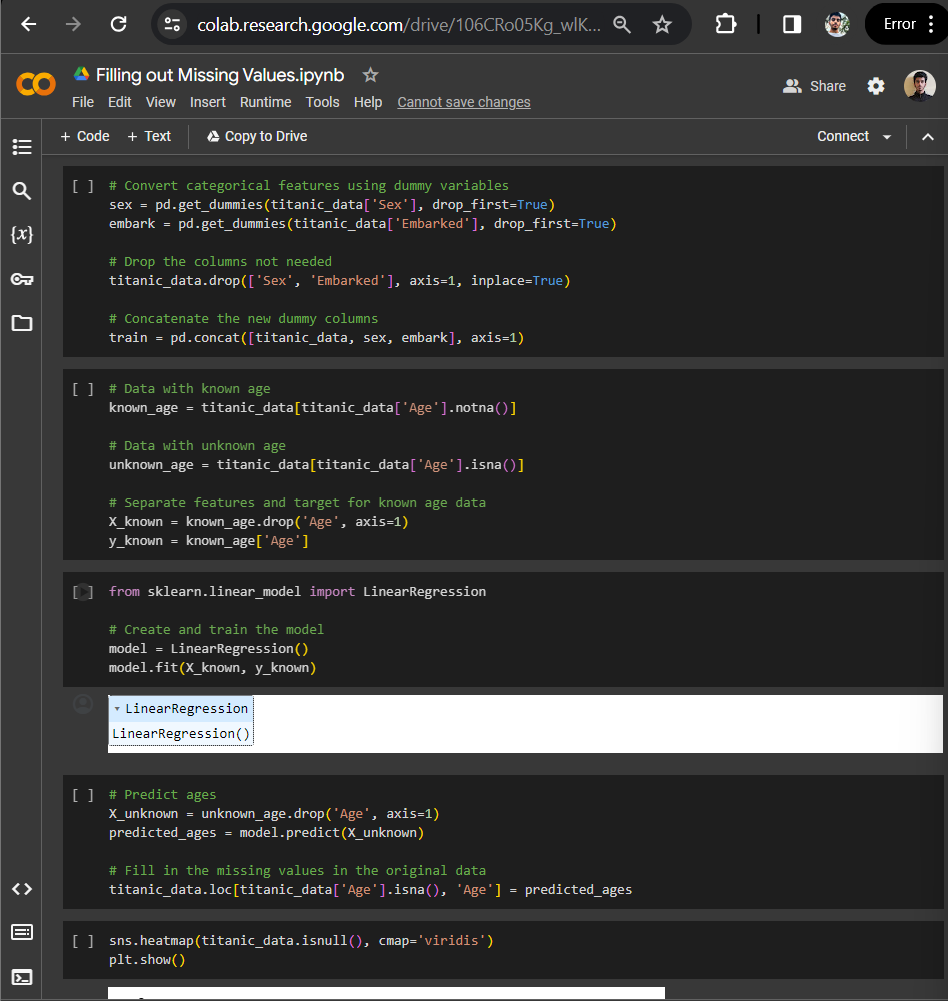
Generally, higher fares correlate with higher chances of survival and lower Passenger class (higher status).

This correlation matrix provides insight into the relationships between different aspects of passenger data on the Titanic, such as socio-economic status (indicated by Pclass and Fare), family size (indicated by SibSp and Parch), and survival rates. Remember, while correlation can indicate a relationship between variables, it does not imply causation.

**Data visualization:**

****

****

****