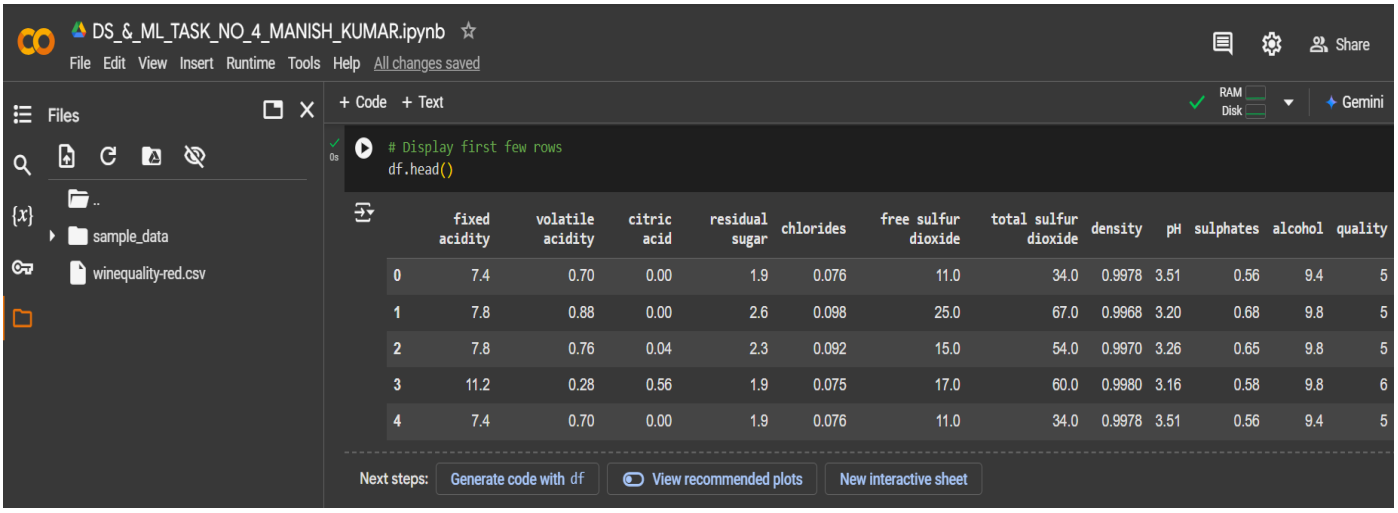
Task 4 Report
Implement a Custom Colormap for a Matplotlib Plot

1. Task Description

In this task, the objective was to implement a **custom colormap** for visualizing high-dimensional data from the **Wine Quality** dataset. The goal was to enhance the interpretability of visualizations by using distinct colors to represent different wine quality scores. This involved creating a custom color palette and applying it to a scatter plot of two features (fixed acidity vs. volatile acidity) as well as a heatmap of feature correlations. This approach helps to better distinguish and understand data patterns, especially in datasets with categorical or discrete target variables, like wine quality.

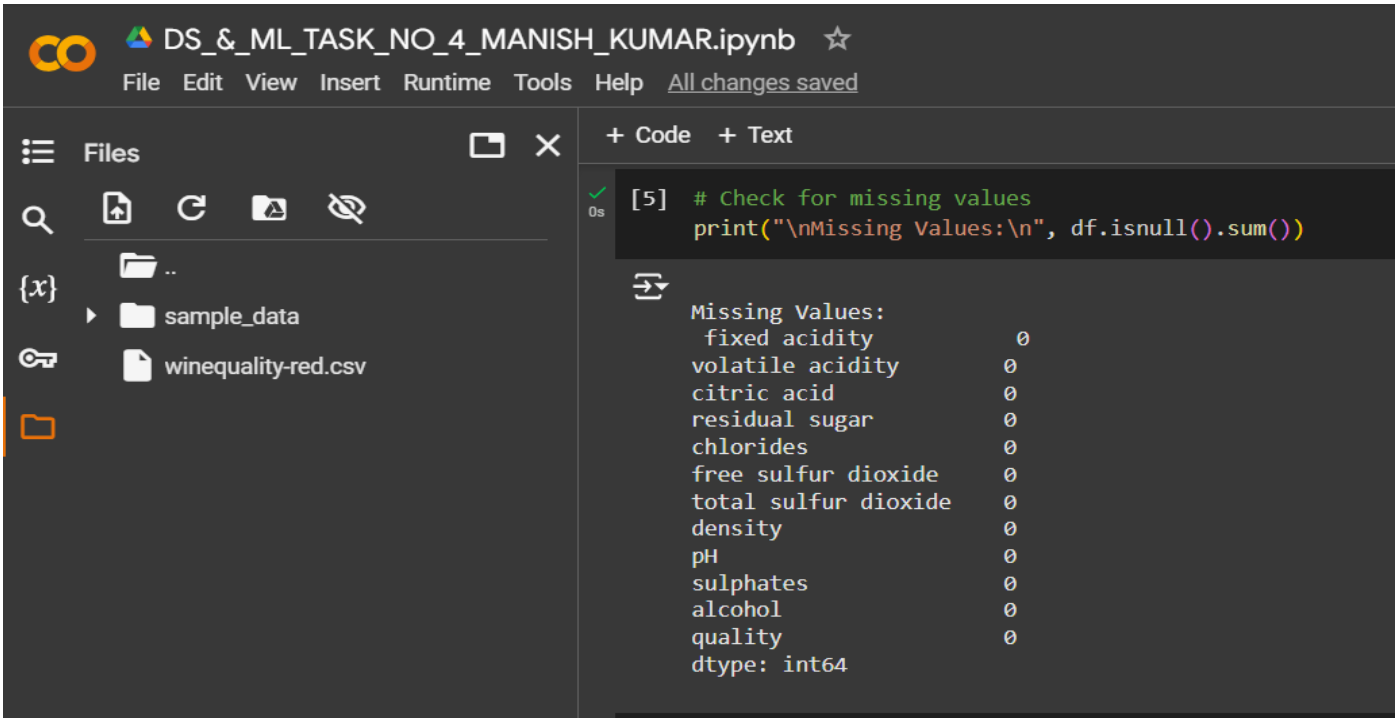
2. Attach Screenshot of Output

❖ **Dataset Preview:**



	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

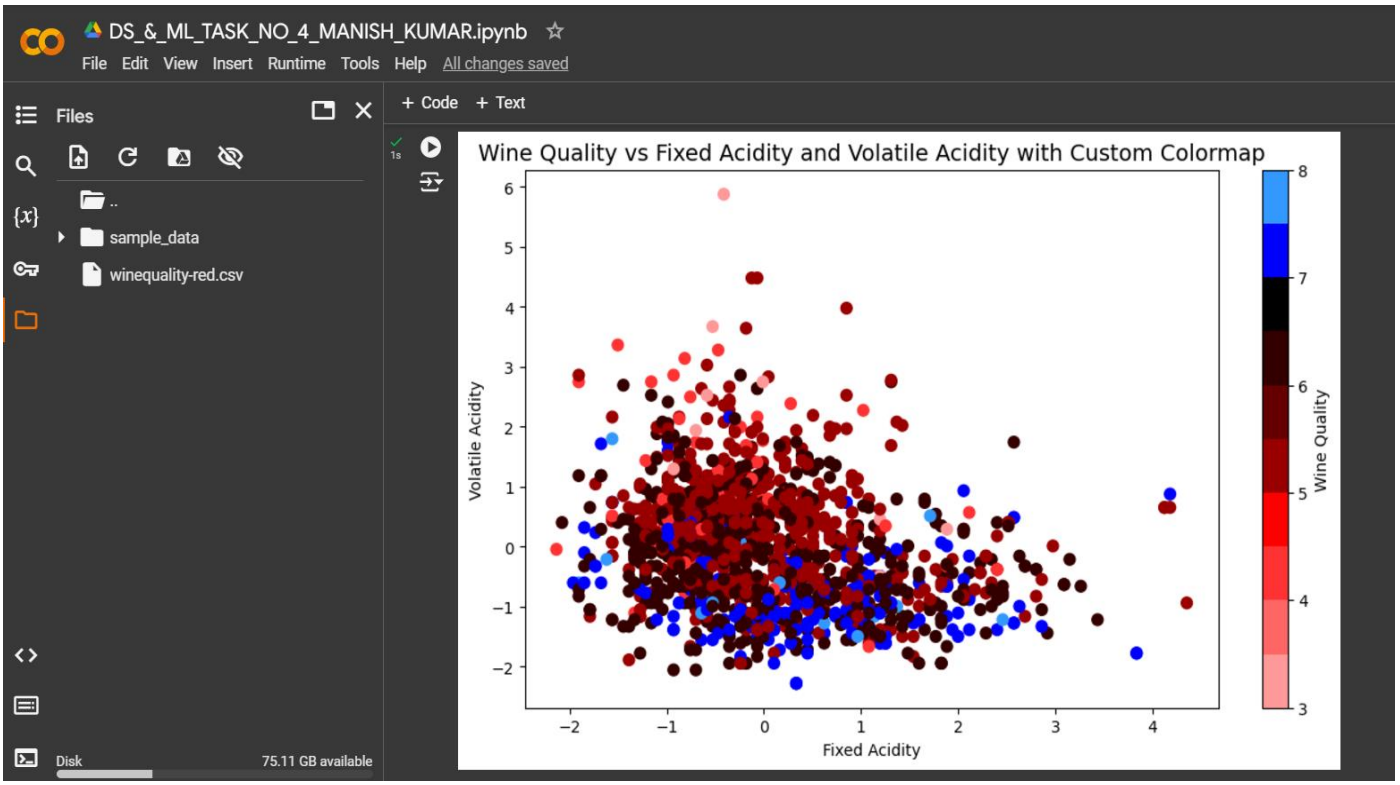
❖ **Check for missing values:**



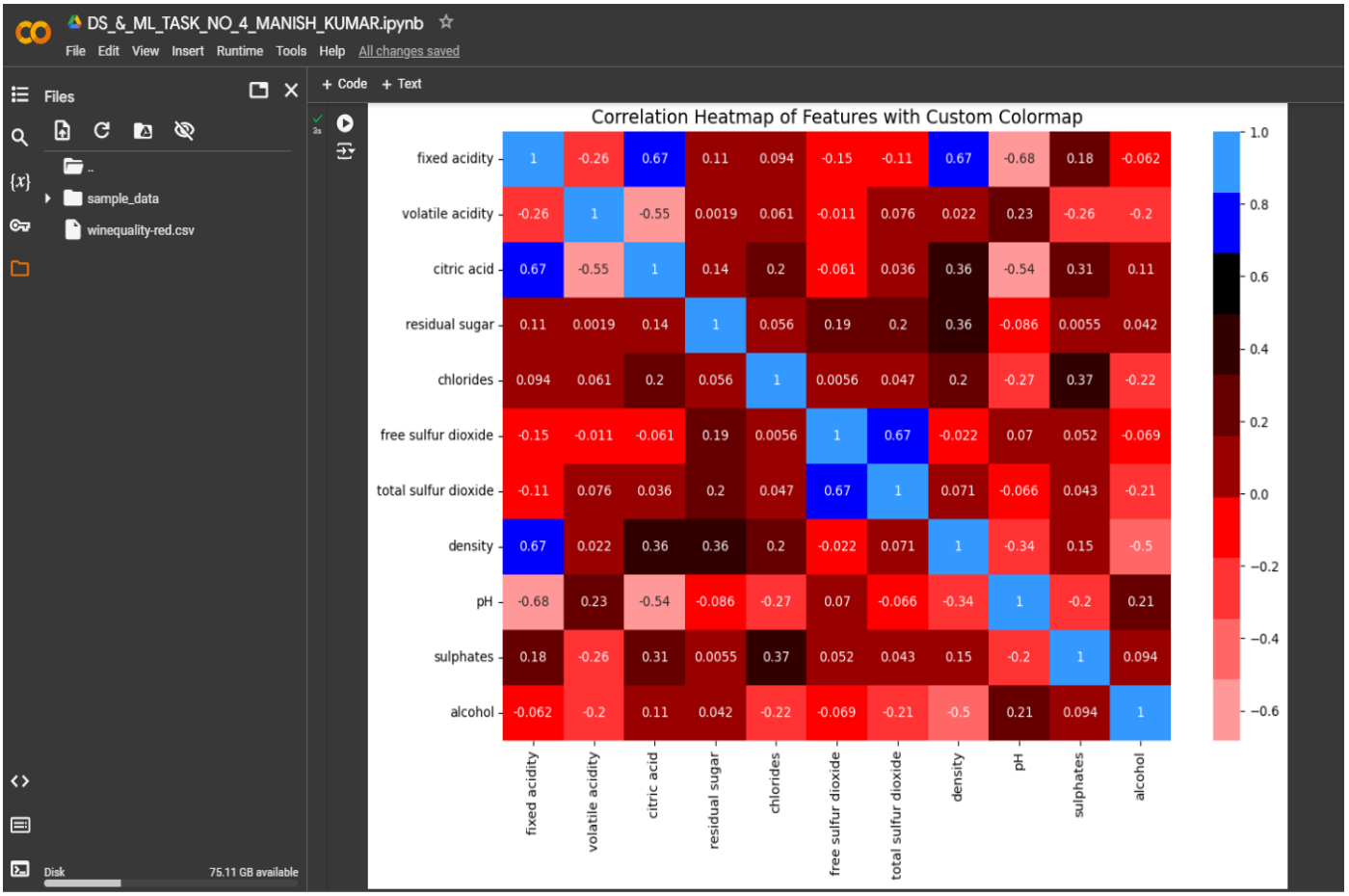
Task 4 Report

Implement a Custom Colormap for a Matplotlib Plot

❖ **Scatter Plot with custom colormap:**



❖ **Correlation heatmap with custom colormap:**



Task 4 Report
Implement a Custom Colormap for a Matplotlib Plot

3. Describe Widget/Algorithm Used in Task

Algorithm Used:

The task primarily utilizes several algorithms and techniques to facilitate the visualization of high-dimensional data using a custom colormap:

- **Color Mapping (Custom Colormap)**:
  - **ListedColormap** from Matplotlib is used to create a custom colormap for visual differentiation between distinct categories. This function allows us to define a specific set of colors and associate them with various levels of wine quality. By assigning unique colors to each category of the target variable (wine quality), the visualization becomes clearer, helping viewers easily identify and distinguish between different classes.
- **Data Preprocessing**:
  - **StandardScaler** from **Scikit-learn** is applied to standardize the dataset. This preprocessing step ensures that each feature has a mean of 0 and a standard deviation of 1. Standardizing the data is important because it removes any bias that may arise from different scales of the features (e.g., features like acidity may have larger values than volatile acidity). This helps in ensuring that the visualizations are accurate and that the color mapping reflects meaningful patterns in the data.
- **Data Visualization**:
  - **Scatter Plot**: The custom colormap is applied to a scatter plot of two features: fixed acidity and volatile acidity. Each point in the scatter plot represents an observation (a wine), with the color of the point representing the wine's quality. This type of plot helps to reveal potential clusters or trends between these two features while also giving insight into how wine quality is distributed across different feature values.
  - **Heatmap**: A heatmap is created using **Seaborn** to visualize the correlation matrix between numerical features in the dataset. The color intensity in the heatmap represents the strength of the correlation, allowing for an intuitive understanding of how different features interact with one another. Strong correlations can suggest that certain features may be combined or used as predictor in further analysis.

Tools Used:

- **Pandas**: **Pandas** is a powerful library for data manipulation and analysis. In this task, it was used to load the dataset and perform essential data cleaning and transformation tasks. For example, missing values are handled efficiently using Pandas methods, and the dataset is prepared for the next steps. The library also provides easy-to-use functions for inspecting the data, making it straightforward to perform exploratory data analysis (EDA) to understand the structure and content of the dataset.

- **Matplotlib**: **Matplotlib** is a widely-used library for creating a wide range of static, animated, and interactive visualizations in Python. In this task, **Matplotlib** was primarily used to generate the scatter plot and apply the custom colormap. The ListedColormap function from Matplotlib enables the creation of a

Task 4 Report
Implement a Custom Colormap for a Matplotlib Plot

color palette that can be used to color-code different categories in the scatter plot. This is a key feature that enhances the effectiveness of the visualization, making it easy to distinguish different wine quality categories.

- **Seaborn**: **Seaborn** is a Python data visualization library based on Matplotlib, designed to make it easier to create aesthetically pleasing and informative statistical plots. It was used in this task to generate a heatmap which visualizes the correlation matrix of various numerical features. The heatmap uses color gradients to represent correlation strength, providing a clearer understanding of how different features relate to each other, which can be useful for further feature engineering or analysis.
- **Scikit-learn**: **Scikit-learn** is a comprehensive library for machine learning and data preprocessing. In this task, **Scikit-learn's StandardScaler** was used to standardize the dataset. Standardizing ensures that all features contribute equally to the visualizations, especially when there are features with different ranges. By standardizing the features, we make sure that the visual representation is not skewed by one feature dominating the others due to its scale.

***** The End *****