TASK 5 DA :

Through statistical and visual exploration of the dataset, several key patterns and characteristics were observed. The .describe() and .info() summaries revealed that most numerical features have a wide range and some exhibit skewness, suggesting that transformations (like log scaling) may help normalize the distributions. Categorical columns have some imbalances, indicating the need to account for class imbalance during modeling if relevant.

The heatmap highlighted strong correlations between certain features, while others showed little to no correlation. Highly correlated features may require dimensionality reduction techniques (like PCA) to avoid multicollinearity issues. We can see from the plot that Parch and SibSp are highly correlated as they have darkest colour. From the pairplot, clusters and possible separations among certain classes were visually detectable, suggesting underlying patterns that machine learning models could leverage.

Histograms indicated that many features are right-skewed, and boxplots confirmed the presence of outliers in several columns. These outliers could impact model performance if not handled properly. Scatterplots between selected features showed mostly linear relationships without complex non-linear patterns, making them potentially suitable for models like logistic regression, decision trees, or ensemble methods