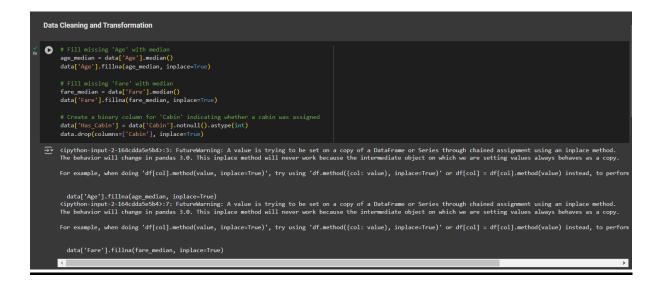
Data Science – Report Task 2

Task - Data Cleaning and Transformation

Use a messy dataset with missing values, duplicate entries, and inconsistent formatting to clean and transform it into a usable form.

Summary - This code cleans and transforms the Titanic dataset from Kaggle for analytical and machine learning purposes. Key tasks include handling missing values (e.g., filling Age and Fare with medians), encoding categorical variables (Sex, Embarked), and simplifying columns like Cabin and Ticket.Advanced features include phonetic clustering for deduplication, title extraction from names, and outlier detection for Age and Fare. New features (FamilySize, Is_Alone, Title) are created, and numerical columns are normalized for consistency. The final cleaned dataset is saved as cleaned_data.csv, ready for predictive modeling or exploratory analysis.





QTechSolutions Dhairya Kalathia

```
| The state of the
```

QTechSolutions Dhairya Kalathia

Output -

The output is a cleaned and transformed version of the Titanic dataset from Kaggle. It includes no missing values, encoded categorical variables (Sex, Embarked), and engineered features like FamilySize, Is_Alone, and Title. Numerical columns (Age, Fare) are normalized, and outliers are flagged. The dataset is deduplicated using phonetic clustering on last names. This cleaned data, saved as cleaned_data.csv, is ready for machine learning, predictive modeling, or detailed analysis.

Skills Learned - Data cleaning, handling missing data.

Tools - Python (Pandas, OpenRefine).

<u>Conclusion</u> - The cleaned and transformed Titanic dataset provides a robust foundation for analytical and predictive tasks. By addressing missing values, simplifying features, and engineering new ones, the dataset is optimized for machine learning models and detailed exploration. The transformations ensure consistency, reduce redundancy, and enhance data usability, making it ideal for extracting meaningful insights or building accurate predictions.

QTechSolutions Dhairya Kalathia