Task 5: Python Basics – Reading Data + Simple Cleaning (Pandas)

Tools Used:
- Google Colab
- Python Libraries: pandas, numpy

Dataset Used:
- Titanic Dataset

Steps Performed

1. Imported pandas and numpy libraries.

2. Loaded the dataset using pd.read_csv().

3. Viewed dataset structure using head() and info().

4. Identified missing values using isnull().sum().

5. Filled numeric missing values with median and categorical values with mode.

6. Dropped irrelevant columns with too many missing values.

7. Removed duplicate records using drop_duplicates().

8. Converted data types where necessary using astype().

9. Created a new column (Age_Group) using logical conditions.

10. Saved the cleaned dataset using to_csv().

Conclusion

The dataset was successfully cleaned and prepared for further analysis or modeling. All missing values, duplicates, and data inconsistencies were handled using pandas.