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

Tools Used:

Google Colab

Libraries: pandas, numpy

1️⃣ Import Required Libraries

Copy code

Python

```
import pandas as pd

import numpy as np
```

2️⃣ Load the Dataset (Titanic)

Copy code

Python

```
# Load Titanic dataset

df = pd.read_csv("titanic.csv")

# Display first 5 rows

df.head()
```

3️⃣ Basic Dataset Information

Copy code

Python

```
# Dataset shape

print("Shape:", df.shape)

# Column names and data types

df.info()
```

4️⃣ Check Missing Values

Copy code

Python

Check missing values in each column

df.isnull().sum()

5 Simple Data Cleaning

◆ Fill Missing Values

Copy code

Python

Fill missing Age with mean

df['Age'].fillna(df['Age'].mean(), inplace=True)

Fill missing Embarked with mode

df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)

◆ Drop Unnecessary Column

Copy code

Python

Drop Cabin column (too many missing values)

df.drop(columns=['Cabin'], inplace=True)

6 Remove Duplicate Rows

Copy code

Python

df.drop_duplicates(inplace=True)

7 Verify Cleaning

Copy code

Python

Check missing values again

df.isnull().sum()

Final dataset info

df.info()

8️⃣ Basic Statistics

Copy code

Python

df.describe()

✅ Final Output

Dataset successfully loaded

Missing values handled

Unnecessary column removed

Duplicates removed

Clean dataset ready for analysis

📌 Conclusion (for submission):

In this task, the Titanic dataset was loaded using Pandas. Basic data exploration was performed. Missing values were handled using mean and mode, unnecessary columns were removed, and duplicate records were dropped. The dataset is now clean and ready for further analysis.