1. Data Preprocessing

Load a dataset (CSV/Excel/JSON) using Pandas.
Handle missing values (mean, median, or mode imputation).
Perform feature scaling (min-max normalization or standardization).

Example to **load a dataset**, **handle missing values**, and **perform feature scaling** (both Min-Max normalization and standardization) using **Pandas** on **Google Colab**.

1. Load a Dataset (CSV/Excel/JSON) using Pandas

Let's start by loading a dataset from a CSV file. You can upload your dataset directly to Google Colab or read it from a URL.

```
import pandas as pd

# Load CSV from URL or Google Drive
url = 'https://raw.githubusercontent.com/openai/data/master/titanic.csv'  #
Example dataset
data = pd.read_csv(url)

# Display the first few rows of the dataset
data.head()
```

Alternatively, if you want to upload a dataset from your local machine, you can use:

```
from google.colab import files

# Upload a file manually
uploaded = files.upload()

# Load the dataset (after uploading)
data = pd.read_csv('your_file_name.csv')
```

2. Handle Missing Values (Mean, Median, or Mode Imputation)

Next, we can handle missing values by imputing them using the **mean**, **median**, or **mode** of the column.

```
# Check for missing values
print(data.isnull().sum())

# Impute missing values with mean (for numerical columns)
data['Age'].fillna(data['Age'].mean(), inplace=True)

# Alternatively, you can impute with the median
data['Age'].fillna(data['Age'].median(), inplace=True)

# Or, impute with the mode (for categorical columns)
```

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

# Verify that missing values are handled
print(data.isnull().sum())
```

3. Feature Scaling (Min-Max Normalization or Standardization)

Now we perform **feature scaling**. We'll show both **Min-Max normalization** and **Standardization**.

Min-Max Normalization

This scales the data to a specific range, typically [0, 1].

```
from sklearn.preprocessing import MinMaxScaler

# Initialize MinMaxScaler
scaler = MinMaxScaler()

# Select columns to normalize (numerical columns)
columns_to_normalize = ['Age', 'Fare']

# Apply Min-Max normalization
data[columns_to_normalize] =
scaler.fit_transform(data[columns_to_normalize])

# Display the first few rows after normalization
data[columns_to_normalize].head()
```

Standardization (Z-Score Normalization)

Standardization transforms the data such that it has a mean of 0 and a standard deviation of 1.

```
from sklearn.preprocessing import StandardScaler

# Initialize StandardScaler
scaler = StandardScaler()

# Apply standardization to numerical columns
data[columns_to_normalize] =
scaler.fit_transform(data[columns_to_normalize])

# Display the first few rows after standardization
data[columns_to_normalize].head()
```

Full Example for Google Colab:

```
# Import necessary libraries
import pandas as pd
from sklearn.preprocessing import MinMaxScaler, StandardScaler
from google.colab import files
```

```
# Upload a dataset (replace 'your_file.csv' with your actual file name if
uploading manually)
uploaded = files.upload()

# Load dataset
data = pd.read_csv('titanic.csv')  # Use the correct file name after upload

# Check for missing values
print("Missing values before imputation:")
print(data.isnull().sum())

# Handle missing values (impute with mean, median, or mode)
data['Age'].fillna(data['Age'].mean(), inplace=True)  # Impute with mean
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)  # Impute
with mode

# Check for missing values after imputation
print("\nMissing values after imputation:")
print(data.isnull().sum())

# Feature scaling (Min-Max Normalization)
scaler_minmax = MinMaxScaler()
columns_to_normalize = ['Age', 'Fare']
data[columns_to_normalize] =
scaler_minmax.fit_transform(data[columns_to_normalize])

# Feature scaling (Standardization)
scaler_standard = StandardScaler()
data[columns_to_normalize] =
scaler_standard.fit_transform(data[columns_to_normalize])

# Display the first few rows of the dataset
data.head()
```

Conclusion:

This code demonstrates how to:

1. **Load a dataset** (from a URL or by uploading).
2. **Handle missing values** by imputing with the **mean**, **median**, or **mode**.
3. **Scale features** using **Min-Max normalization** or **standardization**.

This setup can be run directly in Google Colab to preprocess datasets for further analysis or modeling.