```
# ✅ STEP 1: Install dependencies (Colab has them, but just in case)
!pip install pandas scikit-learn seaborn --quiet

# ✅ STEP 2: Import libraries
import pandas as pd
import numpy as np
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.model_selection import train_test_split
import seaborn as sns

# ✅ STEP 3: Load sample dataset (Iris from seaborn)
df = sns.load_dataset("iris")

# Display raw data
print("✅ Raw Data Sample:")
print(df.head())

# ✅ STEP 4: Simulate missing data
df.loc[0, 'sepal_length'] = np.nan
df.loc[5, 'species'] = np.nan

# ✅ STEP 5: Separate features and target
X = df.drop(columns=['species'])
y = df['species']

# ✅ STEP 6: Define numerical and categorical columns
numeric_features = X.select_dtypes(include=['float64', 'int64']).columns.tolist()
categorical_features = X.select_dtypes(include=['object', 'category']).columns.tolist()

# ✅ STEP 7: Define preprocessing pipeline
numeric_pipeline = Pipeline([
    ('imputer', SimpleImputer(strategy='mean')),
    ('scaler', StandardScaler())
])

categorical_pipeline = Pipeline([
    ('imputer', SimpleImputer(strategy='most_frequent')),
    ('encoder', OneHotEncoder(handle_unknown='ignore', sparse_output=False))
])

# Combine transformers
preprocessor = ColumnTransformer([
    ('num', numeric_pipeline, numeric_features),
    ('cat', categorical_pipeline, categorical_features)
])

# ✅ STEP 8: Fit and transform
X_processed = preprocessor.fit_transform(X)

# ✅ STEP 9: Combine processed features with target
processed_df = pd.DataFrame(X_processed)
processed_df['target'] = y.fillna(y.mode()[0]).values

# ✅ STEP 10: Save and display result
processed_df.to_csv("processed_iris.csv", index=False)
print("\n✅ Processed Data Sample:")
print(processed_df.head())
```

```
print("\n✅ Preprocessing complete. File saved as: processed_iris.csv")
```

```
✅ Raw Data Sample:
   sepal_length  sepal_width  petal_length  petal_width species
0           5.1          3.5           1.4          0.2  setosa
1           4.9          3.0           1.4          0.2  setosa
2           4.7          3.2           1.3          0.2  setosa
3           4.6          3.1           1.5          0.2  setosa
4           5.0          3.6           1.4          0.2  setosa

✅ Processed Data Sample:
          0         1         2         3  target
0  0.000000  1.019004 -1.340227 -1.315444  setosa
1 -1.152203 -0.131979 -1.340227 -1.315444  setosa
2 -1.395201  0.328414 -1.397064 -1.315444  setosa
3 -1.516700  0.098217 -1.283389 -1.315444  setosa
4 -1.030704  1.249201 -1.340227 -1.315444  setosa

✅ Preprocessing complete. File saved as: processed_iris.csv
```