

```

import pandas as pd
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
import numpy as np

data = {
    'Name': ['Alice', 'Bob', 'Charlie', np.nan],
    'Age': [25, 30, np.nan, 22],
    'Salary': [50000, 60000, 55000, None],
    'Department': ['HR', 'IT', 'Finance', 'IT']
}
df = pd.DataFrame(data)
print("Original Data:\n", df)

numeric_features = ['Age', 'Salary']
categorical_features = ['Department']

numeric_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='mean')),
    ('scaler', StandardScaler())
])

categorical_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='most_frequent')),
    ('encoder', OneHotEncoder(handle_unknown='ignore'))
])

preprocessor = ColumnTransformer(
    transformers=[
        ('num', numeric_transformer, numeric_features),
        ('cat', categorical_transformer, categorical_features)
    ]
)

X_processed = preprocessor.fit_transform(df)

cat_feature_names = preprocessor.named_transformers_['cat']['encoder'].get_feature_names_out(categorical_features)
all_feature_names = numeric_features + list(cat_feature_names)
processed_df = pd.DataFrame(X_processed.toarray() if hasattr(X_processed, "toarray") else X_processed,
                           columns=all_feature_names)

print("\nProcessed Data:\n", processed_df)

processed_df.to_csv('processed_data.csv', index=False)
print("\n✅ ETL Process Completed. File saved as 'processed_data.csv'")

```



Original Data:

	Name	Age	Salary	Department
0	Alice	25.0	50000.0	HR
1	Bob	30.0	60000.0	IT
2	Charlie	NaN	55000.0	Finance
3	NaN	22.0	NaN	IT

Processed Data:

	Age	Salary	Department_Finance	Department_HR	Department_IT
0	-0.233285	-1.414214	0.0	1.0	0.0
1	1.516351	1.414214	0.0	0.0	1.0
2	0.000000	0.000000	1.0	0.0	0.0
3	-1.283066	0.000000	0.0	0.0	1.0

✅ ETL Process Completed. File saved as 'processed\_data.csv'

