

```

40     ('scaler', StandardScaler())
41 ])
42 categorical_pipeline = Pipeline([
43     ('imputer', SimpleImputer(strategy='most_frequent')),
44     ('onehot', OneHotEncoder(drop='first'))
45 ])
46 preprocessor = ColumnTransformer([
47     ('num', numeric_pipeline, numeric_features),
48     ('cat', categorical_pipeline, categorical_features)
49 ])
50
51 # Full model
52 model = Pipeline([
53     ('preprocessor', preprocessor),
54     ('regressor', LinearRegression())
55 ])
56
57 # Train
58 model.fit(X_train, y_train)
59
60 # Predict
61 y_pred = model.predict(X_test)
62
63 # Results
64 print("\n" + "="*50)
65 print("REGRESSION ON IRIS (with missing + categorical)")
66 print("="*50)
67 print(f" $R^2$  Score: {r2_score(y_test, y_pred):.4f}")
68 print(f"MAE: {mean_absolute_error(y_test, y_pred):.4f}")
69
70 # Plot
71 plt.figure(figsize=(8,6))
72 plt.scatter(y_test, y_pred, color='purple', alpha=0.7)
73 plt.plot([y.min(), y.max()], [y.min(), y.max()], 'r--')
74 plt.xlabel('Actual Petal Width (cm)')
75 plt.ylabel('Predicted Petal Width (cm)')
76 plt.title('Linear Regression on Iris\n(Missing Values + Species
    ↪ Included)')
77 plt.grid(True, alpha=0.3)
78 plt.show()
79
80 # Predict on new flower
81 new_flower = pd.DataFrame({
82     'sepal length (cm)': [5.9],
83     'sepal width (cm)': [3.0],
84     'petal length (cm)': [np.nan],
85     'species':
86         ['virginica']

```