- Tính tương quan và đa cộng tuyến:  
import pandas as pd

corr\_matrix = df.corr(method='pearson')

print(corr\_matrix)

- pip install joblib matplotlib numpy pandas seaborn scikit-learn

- Cài xlrd: pip install xlrd openpyxl

- Xóa cột trong df: df.drop(“iris” axis= 1) ( nếu xóa cột)

df.drop(“3” axis= 0) ( nếu xóa dòng chỉ số 3)

- Ghép theo cột: print(pd.concat([X, y], axis = 1).head())

- Dự đoán:

for i in range(5):

    X\_new = X\_test[i:i+1]

    y\_pred = knn.predict(X\_new)

    print(f"Mẫu {i} dự đoán là {y\_pred[0]}")

- Chuyển male, female: sex = {'Male':1, 'Female':0}

- Kiểm tra tổng giá trị thiếu theo cột: print(data.isnull().sum())

- Kiểm tra nhanh xem thiếu hay không: print(data.isnull().values.any())

- Xử lý thủ công từng cột:

data['Age'] = data['Age'].fillna(data['Age'].mean())

- Xử lý tất cả:

for col in data.select\_dtypes(include=['float64', 'int64']).columns:

data[col].fillna(data[col].mean(), inplace=True)

- Thuật toán nào cần chuẩn hóa ???

- Đổi tên: X.rename(columns={'Age\_bin': 'Age\_Group'}, inplace=True)

- Cài: pip install imageio

Hoặc !pip install imageio

- pip uninstall imageio xòn rồi cài lại

- Update: pip install --upgrade imageio

- Xóa một lần: X = data.drop(['User ID', 'EstimatedSalary\_K', 'Purchased'], axis = 1)

- from sklearn.metrics import accuracy\_score, confusion\_matrix, classification\_report