Data cleaning

Import matplotlib

# Load sample data

data = {

'Name': ['Alice', 'Bob', None, 'David', 'Eve', 'Alice'],

'Age': [25, 30, 22, None, 28, 25],

'Email': ['alice@example.com', None, 'charlie@example.com', 'david@example.com', 'eve@example.com', 'alice@example.com']

}

df = pd.DataFrame(data)

print("Original Data:")

print(df)

# 1. Remove duplicates

df = df.drop\_duplicates()

# 2. Fill missing values

df['Name'] = df['Name'].fillna('Unknown')

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

df['Email'] = df['Email'].fillna('noemail@example.com')

# 3. Convert data types (if needed)

df['Age'] = df['Age'].astype(int)

print("\nCleaned Data:")

print(df)

