In [1]:

```python
import pandas as pd
```

In [2]:

```python
# 1. Load Dataset

df = pd.read_csv("Mall_Customers.csv")
df.head()
```

Out[2]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

In [3]:

```python
# 2. Standardize Column Names

df.columns = (
    df.columns
    .str.strip()
    .str.lower()
    .str.replace(" ", "_")
)
```

In [4]:

```python
# 3. Check Missing Values

print("\nMissing Values:\n", df.isnull().sum())
```

```

Missing Values:
 customerid                0
gender                    0
age                       0
annual_income_(k$)        0
spending_score_(1-100)    0
dtype: int64
```

In [5]:

```python
# Fill missing values
for col in df.columns:
    if df[col].dtype == "object":
        df[col] = df[col].fillna(df[col].mode()[0])   # categorical → mode
    else:
        df[col] = df[col].fillna(df[col].mean())      # numeric → mean
```

In [6]:

```
# 4. Remove Duplicates

before = df.shape[0]
df = df.drop_duplicates()
after = df.shape[0]
print(f"\nDuplicates Removed: {before - after}")
```

```

Duplicates Removed: 0
```

In [7]:

```
# 5. Standardize Text Values

for col in df.select_dtypes(include="object").columns:
    df[col] = df[col].str.strip().str.lower()

if "gender" in df.columns:
    df["gender"] = df["gender"].replace({
        "m": "male",
        "f": "female"
    })
```

In [8]:

```
# 6. Convert Date Columns

for col in df.columns:
    if "date" in col:
        df[col] = pd.to_datetime(df[col], errors="coerce")
```

In [9]:

```
# 7. Fix Data Types

# Example conversions (if columns exist)
if "age" in df.columns:
    df["age"] = df["age"].astype(int)
```

In [10]:

```
# 8. Final Check

print("\nFinal Info:")
print(df.info())
```

```

Final Info:
<class 'pandas.core.frame.DataFrame'>
Int64Index: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype 
---  ------                  --------------  ----- 
 0   customerid              200 non-null    int64 
 1   gender                  200 non-null    object
 2   age                     200 non-null    int32 
 3   annual_income_(k$)      200 non-null    int64 
 4   spending_score_(1-100)  200 non-null    int64 
dtypes: int32(1), int64(3), object(1)
memory usage: 8.6+ KB
None
```

In [11]:

```
# 9. Save Cleaned Dataset

cleaned_file = "cleaned_mall_customers.csv"
df.to_csv(cleaned_file, index=False)
```

In [12]:

```
print("\nCleaned file saved as:", cleaned_file)
```

```

Cleaned file saved as: cleaned_mall_customers.csv
```