```
# ziad aburas group A
# https://github.com/ziadaburas/study/tree/main/3B/DM
import pandas as pd
path="/content/sample_data/__Mall_Customers_toCleanAssignment (1).csv"
dataset = pd.read_csv(path,encoding_errors='ignore')
dataset.rename(columns={"Spending Score (1-100)":"Spending_Score_1_100", "Annual

```

```
df1 = dataset.drop('CustomerID',axis=1)
df1.head()
```

	Gender	Age	Annual_Income_k_Dollar	Spending_Score_1_100
0	Male	19.0	15.0	39.0
1	1	19.0	NaN	39.0
2	Female	20.0	16.0	6.0
3	Male	NaN	15.0	39.0
4	Female	NaN	16.0	77.0

Next steps: Generate code with df1 | View recommended plots | New interactive sheet

```
df2=df1.copy()
df2.Gender=df2.Gender.astype(str)
df2.Gender.unique()

df2.loc[df2.Gender=='1','Gender']='Male'
df2.loc[df2.Gender=='0','Gender']='Female'
df2.Gender.unique()

df2.where((df2.Gender=='Male')| (df2.Gender=='Female'),other=None,inplace=True
df2.Gender.unique()

df2.drop(index=df2[df2.Gender.isnull()].index,inplace=True)
df2.Gender.unique()

```

```
array(['Male', 'Female'], dtype=object)
```

```
df3=df2.copy()
df3.describe()

# df3.Spending_Score_1_100.fillna(df3.Spending_Score_1_100.mean(),inplace=True
df3.Spending_Score_1_100=df3.Spending_Score_1_100.fillna(df3.Spending_Score_1_
df3.isnull().sum()
```

```

df3.Spending_Score_1_100[df3.Spending_Score_1_100 <1]
df3.Spending_Score_1_100[df3.Spending_Score_1_100 >100]
#
# df3.Spending_Score_1_100[df3.Spending_Score_1_100 <1]=abs(df3.Spending_Score
# df3.Spending_Score_1_100[df3.Spending_Score_1_100 >100]=df3.Spending_Score_1
df3.loc[df3.Spending_Score_1_100 <1,'Spending_Score_1_100']=abs(df3.Spending_S
df3.loc[df3.Spending_Score_1_100 >100,'Spending_Score_1_100']=df3.Spending_Sco
df3.describe()


```

	Age	Annual_Income_k_Dollar	Spending_Score_1_100
count	841.000000	835.000000	989.000000
mean	37.353151	68.415569	49.691770
std	18.468252	34.212392	23.042389
min	-53.000000	-43.000000	3.000000
25%	23.000000	40.000000	36.000000
50%	35.000000	65.000000	49.798949
75%	54.000000	97.000000	65.000000
max	70.000000	137.000000	99.000000

```
df4=df3.copy()
df4.describe()
# df4.Annual_Income_k_Dollar.fillna(df4.Annual_Income_k_Dollar.mean(),inplace=
df4.Annual_Income_k_Dollar=df4.Annual_Income_k_Dollar.fillna(df4.Annual_Income
df4.isnull().sum()

df4.Annual_Income_k_Dollar[df4.Annual_Income_k_Dollar <1]
# df4.Annual_Income_k_Dollar[df4.Annual_Income_k_Dollar <1]=abs(df4.Annual_Inc
df4.loc[df4.Annual_Income_k_Dollar <1,'Annual_Income_k_Dollar']=abs(df4.Annual
df4.describe()
```

	Age	Annual_Income_k_Dollar	Spending_Score_1_100
count	841.000000	989.000000	989.000000
mean	37.353151	68.587460	49.691770
std	18.468252	31.055916	23.042389
min	-53.000000	15.000000	3.000000
25%	23.000000	45.000000	36.000000
50%	35.000000	68.415569	49.798949

50%	55.000000	88.415966	49.798945
75%	54.000000	89.000000	65.000000
max	70.000000	137.000000	99.000000

```
df5=df4.copy()
df5.isnull().sum()
df5 = df5.dropna()
df5.isnull().sum()

df5.Age[df5.Age <1]
df5.loc[df5.Age <1,'Age']=abs(df5.Age[df5.Age <1])

df5.describe()
df5[df5.Age < 15].count()

# Drop if age <15 because they are childern not working
df5.drop(index=df5[df5.Age < 15].index,inplace=True)
df5[df5.Age < 15].count()

df5.describe()

df5.Age=df5.Age.astype(int)
print(df5.Age.dtype)
df5.info()
```

```
int64
<class 'pandas.core.frame.DataFrame'>
Index: 752 entries, 0 to 997
Data columns (total 4 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Gender                  752 non-null    object
 1   Age                     752 non-null    int64
 2   Annual_Income_k_Dollar  752 non-null    float64
 3   Spending_Score_1_100    752 non-null    float64
dtypes: float64(2), int64(1), object(1)
memory usage: 29.4+ KB
```

```
df5.head(20)
df5.to_csv("/content/drive/MyDrive/Colab Notebooks/Mall_Customers_toCleanAssignm
```