```python
# ------------------------------------------------------------
# PROJECT : CUSTOMER SHOPPING BEHAVIOR ANALYSIS USING PANDAS
# ------------------------------------------------------------

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# STEP 1: CREATE DATASET
data = {
    "Customer_ID": range(1, 101),
    "Customer_Name": ["Cust_" + str(i) for i in range(1, 101)],
    "Age": np.random.randint(18, 60, 100),
    "Gender": np.random.choice(["Male", "Female"], 100),
    "Monthly_Income": np.random.randint(15000, 90000, 100),
    "Monthly_Spending": np.random.randint(5000, 60000, 100),
    "Shopping_Frequency": np.random.randint(1, 20, 100)
}

df = pd.DataFrame(data)

# DISPLAY FULL DATASET
print("FULL DATASET:")
print(df)

# ------------------------------------------------------------
# STEP 2: BASIC INSPECTION
# ------------------------------------------------------------

print("\nHEAD:")
print(df.head())

print("\nTAIL:")
print(df.tail())

print("\nINFO:")
print(df.info())

print("\nDESCRIBE:")
print(df.describe())

# ------------------------------------------------------------
# STEP 3: SHAPE, COLUMNS, DTYPES
# ------------------------------------------------------------

print("\nSHAPE:")
print(df.shape)

print("\nCOLUMNS:")
print(df.columns)

print("\nDTYPES:")
print(df.dtypes)

# ------------------------------------------------------------
# STEP 4: DROP, RENAME, SORT, FILLNA, UNIQUE
# ------------------------------------------------------------

print("\nAFTER DROPPING Customer_ID:")
df_dropped = df.drop(columns=["Customer_ID"])
print(df_dropped)

print("\nAFTER RENAMING Customer_Name TO Full_Name:")
df_renamed = df.rename(columns={"Customer_Name": "Full_Name"})
print(df_renamed)

print("\nSORTED BY Monthly_Spending:")
df_sorted = df.sort_values(by="Monthly_Spending")
print(df_sorted)

print("\nFILL NA WITH 0:")
df_filled = df.fillna(0)
print(df_filled)

print("\nUNIQUE ROWS (DROP DUPLICATES):")
df_unique = df.drop_duplicates()
print(df_unique)

# ------------------------------------------------------------
# STEP 5: REPLACE & STRING OPERATIONS
# ------------------------------------------------------------

print("\nREPLACED Male->M, Female->F:")
df_replaced = df.replace({"Male": "M", "Female": "F"})
print(df_replaced)

print("\nADDING COLUMN (Name Contains 'u'):")
df["Contains_u"] = df["Customer_Name"].str.contains("u")
print(df)
```

```python
# ------------------------------------------------------------
# STEP 6: GROUPBY OPERATIONS (MEAN, MAX, MIN, COUNT)
# ------------------------------------------------------------

print("\nMEAN INCOME BY GENDER:")
print(df.groupby("Gender")["Monthly_Income"].mean())

print("\nMAX VALUES GROUPED BY AGE:")
print(df.groupby("Age").max())

print("\nMIN VALUES GROUPED BY AGE:")
print(df.groupby("Age").min())

print("\nCOUNT OF GENDERS:")
print(df["Gender"].value_counts())

# ------------------------------------------------------------
# STEP 7: DATA CLEANING
# ------------------------------------------------------------

print("\nDROPPING NA VALUES:")
df_cleaned = df.dropna()
print(df_cleaned)

print("\nFILL NA WITH 0 AGAIN:")
df_filled_again = df.fillna(0)
print(df_filled_again)

print("\nREPLACE NaN STRING WITH 0:")
df_replaced_nan = df.replace({"NaN": 0})
print(df_replaced_nan)

# ------------------------------------------------------------
# STEP 8: VISUALIZATIONS
# ------------------------------------------------------------

plt.figure(figsize=(15,4))
plt.plot(df["Customer_ID"], df["Monthly_Spending"], marker="o")
plt.title("Monthly Spending Line Plot")
plt.xlabel("Customer ID")
plt.ylabel("Monthly Spending")
plt.show()

df.plot(x="Age", y="Monthly_Income", kind="bar", figsize=(20,5))
plt.title("Age vs Monthly Income (Bar Plot)")
plt.show()

df["Monthly_Spending"].plot(kind="hist", edgecolor="black")
plt.title("Monthly Spending Histogram")
plt.show()

df.plot(x="Monthly_Income", y="Monthly_Spending", kind="scatter")
plt.title("Income vs Spending (Scatter Plot)")
plt.show()
```

```
FULL DATASET:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1        Cust_1   31  Female           86306             31313
1             2        Cust_2   37    Male           28353              5955
2             3        Cust_3   38    Male           51011             53346
3             4        Cust_4   45  Female           80228             25510
4             5        Cust_5   24    Male           46576             32505
..          ...           ...  ...     ...             ...               ...
95           96       Cust_96   48  Female           37034             31002
96           97       Cust_97   28    Male           27696             14714
97           98       Cust_98   37  Female           84130             39055
98           99       Cust_99   24    Male           87063             29934
99          100      Cust_100   53    Male           73577             35450

    Shopping_Frequency
0                    4
1                   18
2                   10
3                   14
4                   19
..                 ...
95                   4
96                  11
97                   7
98                   8
99                   3

[100 rows x 7 columns]

HEAD:
   Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0            1        Cust_1   31  Female           86306             31313
1            2        Cust_2   37    Male           28353              5955
2            3        Cust_3   38    Male           51011             53346
3            4        Cust_4   45  Female           80228             25510
4            5        Cust_5   24    Male           46576             32505

   Shopping_Frequency
0                   4
```

```
   Shopping_Frequency
0                   4
1                  18
2                  10
3                  14
4                  19

TAIL:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
95           96       Cust_96   48  Female           37034             31002
96           97       Cust_97   28    Male           27696             14714
97           98       Cust_98   37  Female           84130             39055
98           99       Cust_99   24    Male           87063             29934
99          100      Cust_100   53    Male           73577             35450

    Shopping_Frequency
95                   4
96                  11
97                   7
98                   8
99                   3

INFO:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Customer_ID         100 non-null    int64
 1   Customer_Name       100 non-null    object
 2   Age                 100 non-null    int64
 3   Gender              100 non-null    object
 4   Monthly_Income      100 non-null    int64
 5   Monthly_Spending    100 non-null    int64
 6   Shopping_Frequency  100 non-null    int64
dtypes: int64(5), object(2)
memory usage: 5.6+ KB
None

DESCRIBE:
       Customer_ID         Age  Monthly_Income  Monthly_Spending  \
count   100.000000  100.000000      100.000000        100.000000
mean     50.500000   37.650000    55238.710000      33064.440000
std      29.011492   12.220512    24212.266542      15649.490557
min       1.000000   18.000000    15523.000000       5101.000000
25%      25.750000   27.000000    31330.750000      19971.750000
50%      50.500000   36.500000    55126.500000      32730.000000
75%      75.250000   46.250000    80566.250000      47375.750000
max     100.000000   59.000000    89923.000000      59619.000000

       Shopping_Frequency
count          100.000000
mean            10.560000
std              5.693022
min              1.000000
25%              6.000000
50%             11.000000
75%             16.000000
max             19.000000

SHAPE:
(100, 7)

COLUMNS:
Index(['Customer_ID', 'Customer_Name', 'Age', 'Gender', 'Monthly_Income',
       'Monthly_Spending', 'Shopping_Frequency'],
      dtype='object')

DTYPES:
Customer_ID            int64
Customer_Name         object
Age                    int64
Gender                object
Monthly_Income         int64
Monthly_Spending       int64
Shopping_Frequency     int64
dtype: object

AFTER DROPPING Customer_ID:
   Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0         Cust_1   31  Female           86306             31313
1         Cust_2   37    Male           28353              5955
2         Cust_3   38    Male           51011             53346
3         Cust_4   45  Female           80228             25510
4         Cust_5   24    Male           46576             32505
..           ...  ...     ...             ...               ...
95       Cust_96   48  Female           37034             31002
96       Cust_97   28    Male           27696             14714
97       Cust_98   37  Female           84130             39055
98       Cust_99   24    Male           87063             29934
99      Cust_100   53    Male           73577             35450

    Shopping_Frequency
0                    4
1                   18
2                   10
3                   14
4                   19
..                 ...
95                   4
```

```
95                   4
96                  11
97                   7
98                   8
99                   3

[100 rows x 6 columns]

AFTER RENAMING Customer_Name TO Full_Name:
    Customer_ID Full_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1    Cust_1   31  Female           86306             31313
1             2    Cust_2   37    Male           28353              5955
2             3    Cust_3   38    Male           51011             53346
3             4    Cust_4   45  Female           80228             25510
4             5    Cust_5   24    Male           46576             32505
..          ...       ...  ...     ...             ...               ...
95           96   Cust_96   48  Female           37034             31002
96           97   Cust_97   28    Male           27696             14714
97           98   Cust_98   37  Female           84130             39055
98           99   Cust_99   24    Male           87063             29934
99          100  Cust_100   53    Male           73577             35450

    Shopping_Frequency
0                    4
1                   18
2                   10
3                   14
4                   19
..                 ...
95                   4
96                  11
97                   7
98                   8
99                   3

[100 rows x 7 columns]

SORTED BY Monthly_Spending:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
72           73       Cust_73   46    Male           55715              5101
71           72       Cust_72   52    Male           82251              5782
1             2        Cust_2   37    Male           28353              5955
33           34       Cust_34   46  Female           24646              6301
8             9        Cust_9   57  Female           82959              6674
..          ...           ...  ...     ...             ...               ...
87           88       Cust_88   32    Male           27806             58120
85           86       Cust_86   49  Female           82927             58829
94           95       Cust_95   28    Male           80543             59115
19           20       Cust_20   57    Male           44346             59598
29           30       Cust_30   47  Female           16535             59619

    Shopping_Frequency
72                  16
71                   2
1                   18
33                   6
8                    5
..                 ...
87                   7
85                  19
94                  16
19                  11
29                   9

[100 rows x 7 columns]

FILL NA WITH 0:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1        Cust_1   31  Female           86306             31313
1             2        Cust_2   37    Male           28353              5955
2             3        Cust_3   38    Male           51011             53346
3             4        Cust_4   45  Female           80228             25510
4             5        Cust_5   24    Male           46576             32505
..          ...           ...  ...     ...             ...               ...
95           96       Cust_96   48  Female           37034             31002
96           97       Cust_97   28    Male           27696             14714
97           98       Cust_98   37  Female           84130             39055
98           99       Cust_99   24    Male           87063             29934
99          100      Cust_100   53    Male           73577             35450

    Shopping_Frequency
0                    4
1                   18
2                   10
3                   14
4                   19
..                 ...
95                   4
96                  11
97                   7
98                   8
99                   3

[100 rows x 7 columns]

UNIQUE ROWS (DROP DUPLICATES):
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1        Cust_1   31  Female           86306             31313
1             2        Cust_2   37    Male           28353              5955
2             3        Cust_3   38    Male           51011             53346
3             4        Cust_4   45  Female           80228             25510
```

```
    0             1        Cust_1   31  Female           86306             31
    1             2        Cust_2   37    Male           28353              5
    2             3        Cust_3   38    Male           51011          53346
    3             4        Cust_4   45  Female           80228          25510
    4             5        Cust_5   24    Male           46576          32505
    ..          ...           ...  ...     ...             ...            ...
    95           96       Cust_96   48  Female           37034          31002
    96           97       Cust_97   28    Male           27696          14714
    97           98       Cust_98   37  Female           84130          39055
    98           99       Cust_99   24    Male           87063          29934
    99          100      Cust_100   53    Male           73577          35450

        Shopping_Frequency
    0                    4
    1                   18
    2                   10
    3                   14
    4                   19
    ..                 ...
    95                   4
    96                  11
    97                   7
    98                   8
    99                   3

    [100 rows x 7 columns]

    REPLACED Male->M, Female->F:
        Customer_ID Customer_Name  Age Gender  Monthly_Income  Monthly_Spending  \
    0             1        Cust_1   31      F           86306             31313
    1             2        Cust_2   37      M           28353              5955
    2             3        Cust_3   38      M           51011             53346
    3             4        Cust_4   45      F           80228             25510
    4             5        Cust_5   24      M           46576             32505
    ..          ...           ...  ...    ...             ...               ...
    95           96       Cust_96   48      F           37034             31002
    96           97       Cust_97   28      M           27696             14714
    97           98       Cust_98   37      F           84130             39055
    98           99       Cust_99   24      M           87063             29934
    99          100      Cust_100   53      M           73577             35450

        Shopping_Frequency
    0                    4
    1                   18
    2                   10
    3                   14
    4                   19
    ..                 ...
    95                   4
    96                  11
    97                   7
    98                   8
    99                   3

    [100 rows x 7 columns]

    ADDING COLUMN (Name Contains 'u'):
        Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
    0             1        Cust_1   31  Female           86306             31313
    1             2        Cust_2   37    Male           28353              5955
    2             3        Cust_3   38    Male           51011             53346
    3             4        Cust_4   45  Female           80228             25510
    4             5        Cust_5   24    Male           46576             32505
    ..          ...           ...  ...     ...             ...               ...
    95           96       Cust_96   48  Female           37034             31002
    96           97       Cust_97   28    Male           27696             14714
    97           98       Cust_98   37  Female           84130             39055
    98           99       Cust_99   24    Male           87063             29934
    99          100      Cust_100   53    Male           73577             35450

        Shopping_Frequency  Contains_u
    0                    4        True
    1                   18        True
    2                   10        True
    3                   14        True
    4                   19        True
    ..                 ...         ...
    95                   4        True
    96                  11        True
    97                   7        True
    98                   8        True
    99                   3        True

    [100 rows x 8 columns]

    MEAN INCOME BY GENDER:
    Gender
    Female    54900.125000
    Male      55464.433333
    Name: Monthly_Income, dtype: float64

    MAX VALUES GROUPED BY AGE:
         Customer_ID Customer_Name  Gender  Monthly_Income  Monthly_Spending  \
    Age
    18            92       Cust_92    Male           73426             48866
    19            77       Cust_77    Male           84531             48885
    20            36       Cust_36    Male           63296             30243
    21            42       Cust_42    Male           68366             52756
    22            70       Cust_70    Male           62339             56743
    23            74       Cust_74  Female           82174             24127
```

23	74	Cust_74	Female	82174	24127	
24	99	Cust_99	Male	87063	35898	
25	31	Cust_31	Male	89061	47927	
26	32	Cust_32	Female	46109	42258	
27	87	Cust_87	Male	73655	49248	
28	97	Cust_97	Male	80543	59115	
30	94	Cust_94	Male	84071	31731	
31	78	Cust_78	Male	86306	31313	
32	93	Cust_93	Male	49691	58120	
34	69	Cust_69	Male	84392	54623	
35	91	Cust_91	Male	79271	43774	
36	82	Cust_82	Male	85724	38006	
37	98	Cust_98	Male	84130	39055	
38	3	Cust_3	Male	51011	53346	
39	66	Cust_66	Male	44463	50958	
42	51	Cust_51	Male	86616	49536	
43	76	Cust_76	Male	85741	54823	
44	84	Cust_84	Male	89923	54032	
45	90	Cust_90	Male	89866	44113	
46	73	Cust_73	Male	64598	37502	
47	30	Cust_30	Female	16535	59619	
48	96	Cust_96	Male	37034	50025	
49	86	Cust_86	Female	82927	58829	
50	71	Cust_71	Male	72994	39196	
51	8	Cust_8	Male	54829	28291	
52	72	Cust_72	Male	82251	32955	
53	100	Cust_6	Male	86135	53894	
54	81	Cust_81	Male	88298	35796	
56	60	Cust_60	Male	88402	34887	
57	50	Cust_9	Male	82959	59598	
58	75	Cust_75	Male	57933	15757	
59	67	Cust_67	Male	65768	50526	

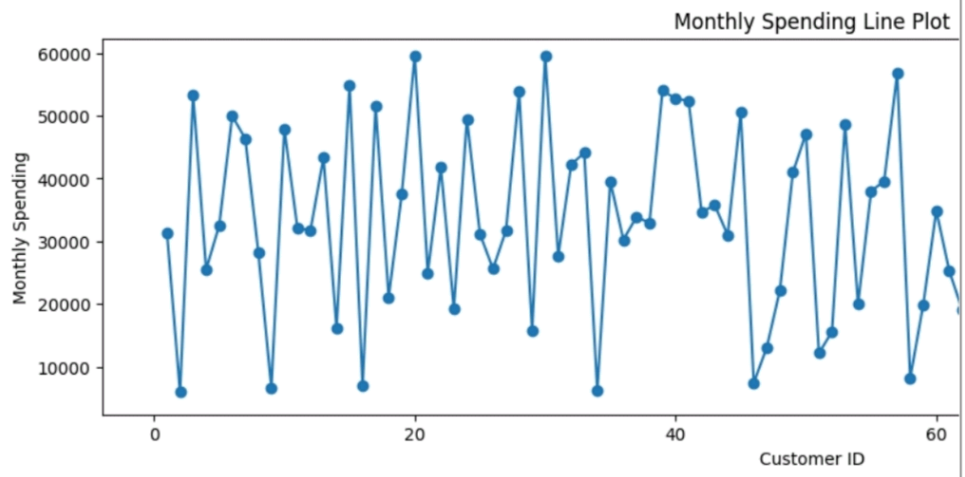
Age	Shopping_Frequency	Contains_u
18	14	True
19	19	True
20	17	True
21	19	True
22	14	True
23	15	True
24	19	True
25	12	True
26	13	True
27	17	True
28	19	True
30	19	True
31	10	True
32	19	True
34	19	True
35	12	True
36	16	True
37	18	True
38	10	True
39	16	True
42	17	True
43	18	True
44	19	True
45	14	True
46	16	True
47	9	True
48	11	True
49	19	True
50	8	True
51	8	True
52	17	True
53	17	True
54	16	True
56	18	True
57	11	True
58	17	True
59	10	True

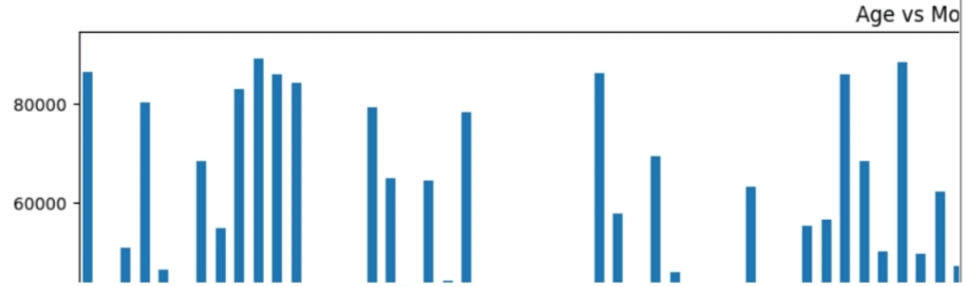
MIN VALUES GROUPED BY AGE:

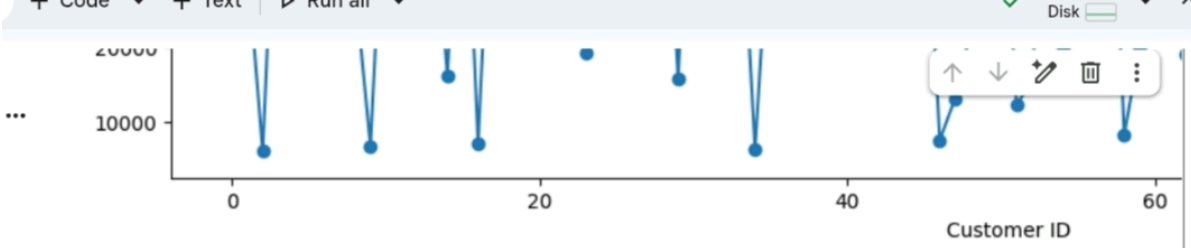
Age	Customer_ID	Customer_Name	Gender	Monthly_Income	Monthly_Spending	\
18	7	Cust_13	Female	25219	19917	
19	25	Cust_25	Female	42956	31169	
20	36	Cust_36	Male	63296	30243	
21	40	Cust_40	Female	56604	34612	
22	46	Cust_46	Female	20061	7420	
23	74	Cust_74	Female	82174	24127	
24	5	Cust_23	Female	30490	19290	
25	10	Cust_10	Female	69477	27674	
26	27	Cust_27	Female	29528	31755	
27	14	Cust_14	Female	23878	14417	
28	18	Cust_18	Female	17293	14714	
30	12	Cust_12	Female	38829	12472	
31	1	Cust_1	Female	76134	22670	
32	54	Cust_54	Male	17371	19990	
34	17	Cust_17	Female	47411	13077	
35	16	Cust_16	Female	24514	7054	
36	11	Cust_11	Male	57010	12922	
37	2	Cust_2	Female	28353	5955	
38	3	Cust_3	Male	51011	53346	
39	37	Cust_37	Female	19968	9007	
42	21	Cust_21	Female	34240	12157	
43	15	Cust_15	Female	30372	47135	
44	39	Cust_39	Female	55424	41361	
45	4	Cust_33	Female	20270	15597	

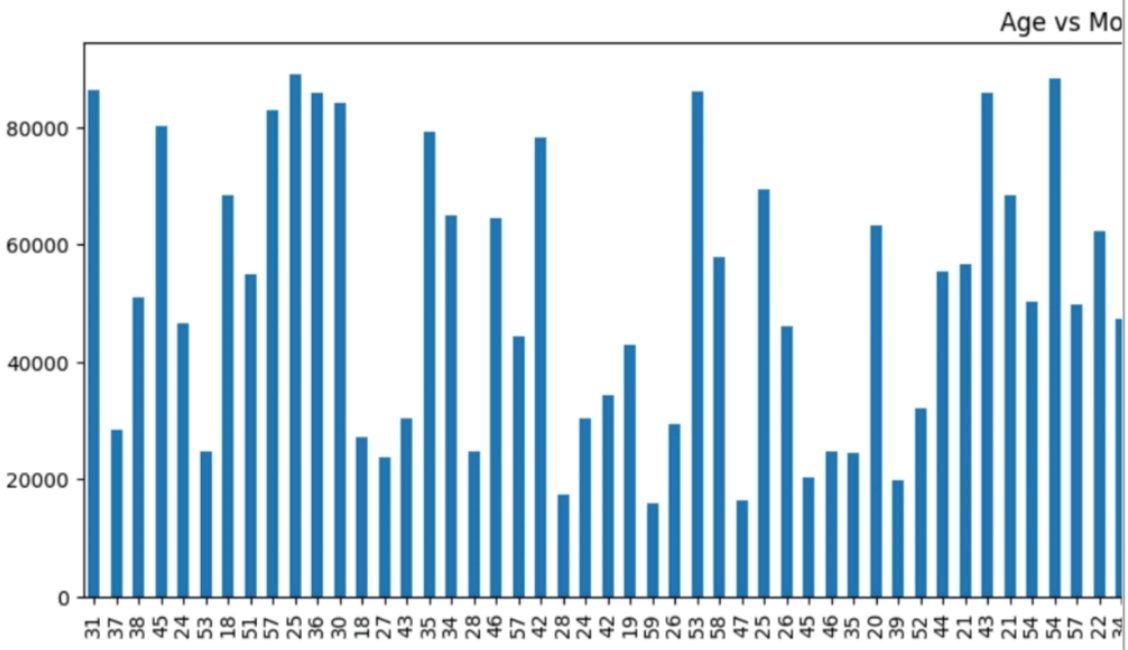
```
45             4       Cust_33  Female           20270             15597
46            19       Cust_19  Female           24646              5101
47            30       Cust_30  Female           16535             59619
48            89       Cust_89  Female           28778             31002
49            86       Cust_86  Female           82927             58829
50            71       Cust_71    Male           72994             39196
51             8        Cust_8    Male           54829             28291
52            38       Cust_38    Male           29516              5782
53             6      Cust_100    Male           24694             35450
54            43       Cust_43  Female           15523             23920
56            60       Cust_60    Male           88402             34887
57             9       Cust_20  Female           44346              6674
58            29       Cust_29  Female           19555             14204
59            26       Cust_26  Female           16017             25596

     Shopping_Frequency  Contains_u
Age
18                    2        True
19                   18        True
20                   17        True
21                    5        True
22                    6        True
23                   15        True
24                    6        True
25                    3        True
26                    6        True
27                    1        True
28                    1        True
30                    1        True
31                    4        True
32                    7        True
34                   11        True
35                    1        True
36                    8        True
37                    7        True
38                   10        True
39                    7        True
42                   11        True
43                   15        True
44                    7        True
45                    1        True
46                    4        True
47                    9        True
48                    4        True
49                   19        True
50                    8        True
51                    8        True
52                    2        True
53                    2        True
54                    7        True
56                   18        True
57                    5        True
58                    4        True
59                    2        True

COUNT OF GENDERS:
Gender
Male      60
Female    40
Name: count, dtype: int64

DROPPING NA VALUES:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1        Cust_1   31  Female           86306             31313
1             2        Cust_2   37    Male           28353              5955
2             3        Cust_3   38    Male           51011             53346
3             4        Cust_4   45  Female           80228             25510
4             5        Cust_5   24    Male           46576             32505
..          ...           ...  ...     ...             ...               ...
95           96       Cust_96   48  Female           37034             31002
96           97       Cust_97   28    Male           27696             14714
97           98       Cust_98   37  Female           84130             39055
98           99       Cust_99   24    Male           87063             29934
99          100      Cust_100   53    Male           73577             35450

    Shopping_Frequency  Contains_u
0                    4        True
1                   18        True
2                   10        True
3                   14        True
4                   19        True
..                 ...         ...
95                   4        True
96                  11        True
97                   7        True
98                   8        True
99                   3        True

[100 rows x 8 columns]

FILL NA WITH 0 AGAIN:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1        Cust_1   31  Female           86306             31313
1             2        Cust_2   37    Male           28353              5955
2             3        Cust_3   38    Male           51011             53346
3             4        Cust_4   45  Female           80228             25510
4             5        Cust_5   24    Male           46576             32505
..          ...           ...  ...     ...             ...               ...
95           96       Cust_96   48  Female           37034             31002
96           97       Cust_97   28    Male           27696             14714
```

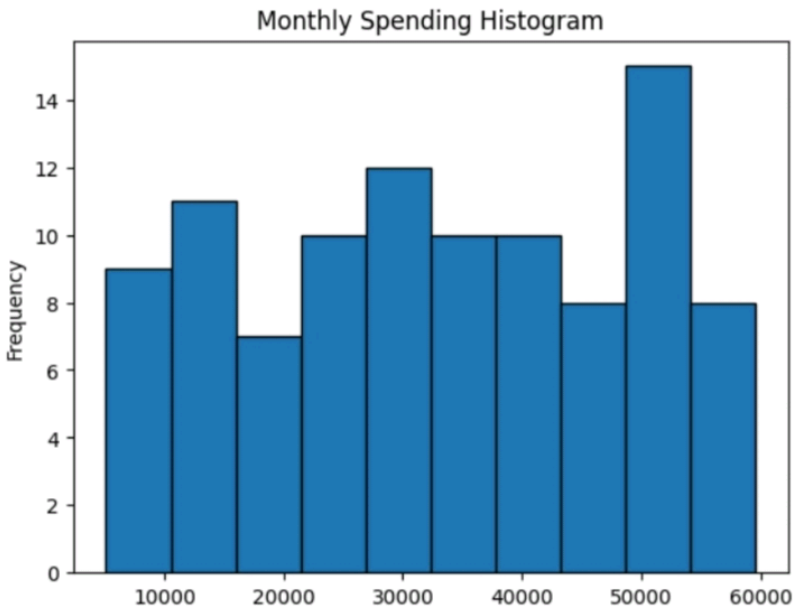
```
96                  11        True
97                   7        True
98                   8        True
99                   3        True

[100 rows x 8 columns]

FILL NA WITH 0 AGAIN:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1        Cust_1   31  Female           86306             31313
1             2        Cust_2   37    Male           28353              5955
2             3        Cust_3   38    Male           51011             53346
3             4        Cust_4   45  Female           80228             25510
4             5        Cust_5   24    Male           46576             32505
..          ...           ...  ...     ...             ...               ...
95           96       Cust_96   48  Female           37034             31002
96           97       Cust_97   28    Male           27696             14714
97           98       Cust_98   37  Female           84130             39055
98           99       Cust_99   24    Male           87063             29934
99          100      Cust_100   53    Male           73577             35450

    Shopping_Frequency  Contains_u
0                    4        True
1                   18        True
2                   10        True
3                   14        True
4                   19        True
..                 ...         ...
95                   4        True
96                  11        True
97                   7        True
98                   8        True
99                   3        True

[100 rows x 8 columns]

REPLACE NaN STRING WITH 0:
    Customer_ID Customer_Name  Age  Gender  Monthly_Income  Monthly_Spending  \
0             1        Cust_1   31  Female           86306             31313
1             2        Cust_2   37    Male           28353              5955
2             3        Cust_3   38    Male           51011             53346
3             4        Cust_4   45  Female           80228             25510
4             5        Cust_5   24    Male           46576             32505
..          ...           ...  ...     ...             ...               ...
95           96       Cust_96   48  Female           37034             31002
96           97       Cust_97   28    Male           27696             14714
97           98       Cust_98   37  Female           84130             39055
98           99       Cust_99   24    Male           87063             29934
99          100      Cust_100   53    Male           73577             35450

    Shopping_Frequency  Contains_u
0                    4        True
1                   18        True
2                   10        True
3                   14        True
4                   19        True
..                 ...         ...
95                   4        True
96                  11        True
97                   7        True
98                   8        True
99                   3        True

[100 rows x 8 columns]
```

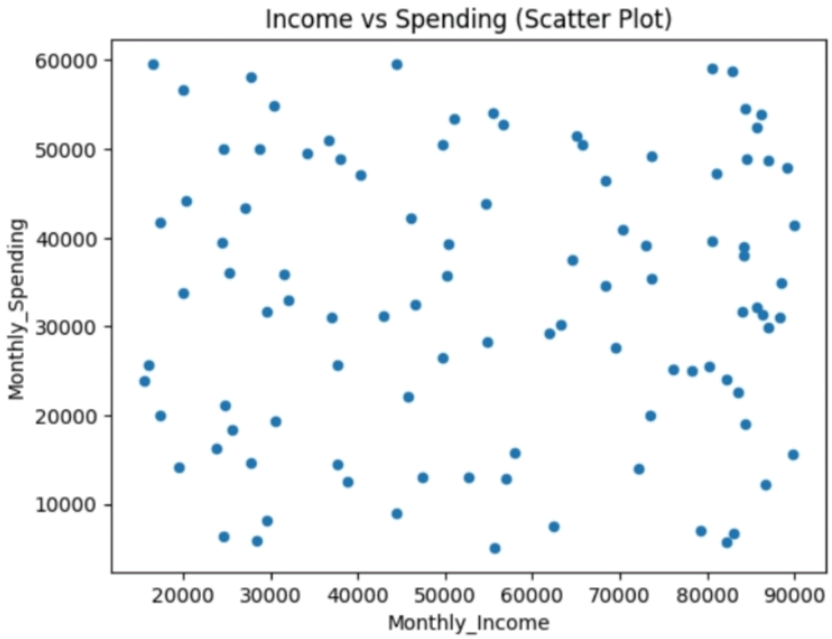





10000
0
20
40
60
Customer ID


Age vs Mo
80000
60000
40000
20000
0


Monthly Spending Histogram
Frequency
14
12
10
8
6
4
2
0
10000
20000
30000
40000
50000
60000


Income vs Spending (Scatter Plot)
Monthly_Spending
60000
50000
40000
30000
20000
10000
20000
30000
40000
50000
60000
70000
80000
90000
Monthly_Income