inc-exp

August 8, 2023

0.1 LOAD

[1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.style as style
import seaborn as sns
import itertools
%matplotlib inline
```

[2]:
```python
import warnings
warnings.filterwarnings('ignore')
```

[3]:
```python
df=pd.read_csv(r"C:\Users\B YLR SRI VAISHNAVI\Downloads\Inc_Exp_Data.csv")
df
```

[3]:
```
    Mthly_HH_Income  Mthly_HH_Expense  No_of_Fly_Members  Emi_or_Rent_Amt  \
0              5000              8000                  3             2000
1              6000              7000                  2             3000
2             10000              4500                  2                0
3             10000              2000                  1                0
4             12500             12000                  2             3000
5             14000              8000                  2                0
6             15000             16000                  3            35000
7             18000             20000                  5             8000
8             19000              9000                  2                0
9             20000              9000                  4                0
10            20000             18000                  4             8000
11            22000             25000                  6            12000
12            23400              5000                  3                0
13            24000             10500                  6                0
14            24000             10000                  4                0
15            25000             12300                  3                0
16            25000             20000                  3             3500
17            25000             10000                  6                0
18            29000              6600                  2             2000
19            30000             13000                  4                0
20            30500             25000                  5             5000
```

```
21            32000             15000                  4                0
22            34000             19000                  6                0
23            34000             25000                  3             4000
24            35000             12000                  3                0
25            35000             25000                  4                0
26            39000              8000                  4                0
27            40000             10000                  4                0
28            42000             15000                  4                0
29            43000             12000                  4                0
30            45000             25000                  6                0
31            45000             40000                  6             3500
32            45000             10000                  2             1000
33            45000             22000                  4             2500
34            46000             25000                  5             3500
35            47000             15000                  7                0
36            50000             20000                  4                0
37            50500             20000                  3                0
38            55000             45000                  6            12000
39            60000             10000                  3                0
40            60000             50000                  6            10000
41            65000             20000                  4             5000
42            70000              9000                  2                0
43            80000             20000                  4                0
44            85000             25000                  5                0
45            90000             48000                  7                0
46            98000             25000                  5                0
47           100000             30000                  6                0
48           100000             50000                  4            20000
49           100000             40000                  6            10000

    Annual_HH_Income Highest_Qualified_Member  No_of_Earning_Members
0              64200           Under-Graduate                      1
1              79920               Illiterate                      1
2             112800           Under-Graduate                      1
3              97200               Illiterate                      1
4             147000                 Graduate                      1
5             196560                 Graduate                      1
6             167400            Post-Graduate                      1
7             216000                 Graduate                      1
8             218880           Under-Graduate                      1
9             220800           Under-Graduate                      2
10            278400           Under-Graduate                      2
11            279840               Illiterate                      1
12            292032               Illiterate                      1
13            316800                 Graduate                      2
14            244800                 Graduate                      2
15            246000                 Graduate                      1
```

```
16             261000                 Graduate                      1
17             258000           Under-Graduate                      3
18             348000                 Graduate                      1
19             385200                 Graduate                      1
20             351360           Under-Graduate                      1
21             445440             Professional                      1
22             330480             Professional                      1
23             469200             Professional                      1
24             466200                 Graduate                      1
25             449400             Professional                      2
26             556920           Under-Graduate                      1
27             412800           Under-Graduate                      1
28             488880                 Graduate                      1
29             619200                 Graduate                      1
30             523800                 Graduate                      3
31             507600             Professional                      2
32             437400            Post-Graduate                      1
33             610200            Post-Graduate                      1
34             596160                 Graduate                      1
35             456840             Professional                      4
36             570000             Professional                      1
37             581760             Professional                      2
38             600600                 Graduate                      2
39             590400            Post-Graduate                      1
40             590400                 Graduate                      1
41             647400               Illiterate                      2
42             756000                 Graduate                      1
43            1075200                 Graduate                      1
44            1142400           Under-Graduate                      2
45             885600            Post-Graduate                      3
46            1152480             Professional                      2
47            1404000                 Graduate                      3
48            1032000             Professional                      2
49            1320000            Post-Graduate                      1
```

0.2 ANALYZE

[4]:

```python
df.shape
```

[4]:

```
(50, 7)
```

[5]:

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 7 columns):
 #   Column                     Non-Null Count  Dtype
 ---  ------                     --------------  -----
  0   Mthly_HH_Income            50 non-null     int64
  1   Mthly_HH_Expense           50 non-null     int64
  2   No_of_Fly_Members          50 non-null     int64
  3   Emi_or_Rent_Amt            50 non-null     int64
  4   Annual_HH_Income           50 non-null     int64
  5   Highest_Qualified_Member   50 non-null     object
  6   No_of_Earning_Members      50 non-null     int64
 dtypes: int64(6), object(1)
 memory usage: 2.9+ KB
```

[6]:
```
df.describe()
```

[6]:
```
       Mthly_HH_Income  Mthly_HH_Expense  No_of_Fly_Members  Emi_or_Rent_Amt  \
count        50.000000         50.000000          50.000000        50.000000
mean      41558.000000      18818.000000           4.060000      3060.000000
std       26097.908979      12090.216824           1.517382      6241.434948
min        5000.000000       2000.000000           1.000000         0.000000
25%       23550.000000      10000.000000           3.000000         0.000000
50%       35000.000000      15500.000000           4.000000         0.000000
75%       50375.000000      25000.000000           5.000000      3500.000000
max      100000.000000      50000.000000           7.000000     35000.000000

       Annual_HH_Income  No_of_Earning_Members
count      5.000000e+01              50.000000
mean       4.900190e+05               1.460000
std        3.201358e+05               0.734291
min        6.420000e+04               1.000000
25%        2.587500e+05               1.000000
50%        4.474200e+05               1.000000
75%        5.947200e+05               2.000000
max        1.404000e+06               4.000000
```

[7]:
```
df.describe().T  #transpose
```

[7]:
```
                       count       mean            std      min       25%  \
Mthly_HH_Income         50.0   41558.00   26097.908979   5000.0   23550.0
Mthly_HH_Expense        50.0   18818.00   12090.216824   2000.0   10000.0
No_of_Fly_Members       50.0       4.06       1.517382      1.0       3.0
Emi_or_Rent_Amt         50.0    3060.00    6241.434948      0.0       0.0
Annual_HH_Income        50.0  490019.04  320135.792123  64200.0  258750.0
No_of_Earning_Members   50.0       1.46       0.734291      1.0       1.0

                           50%       75%        max
Mthly_HH_Income        35000.0   50375.0   100000.0
Mthly_HH_Expense       15500.0   25000.0    50000.0
No_of_Fly_Members          4.0       5.0        7.0
```

```
Emi_or_Rent_Amt              0.0     3500.0     35000.0
Annual_HH_Income        447420.0   594720.0   1404000.0
No_of_Earning_Members        1.0        2.0         4.0
```

```
[8]: df.isnull().any()
```

```
[8]: Mthly_HH_Income            False
     Mthly_HH_Expense           False
     No_of_Fly_Members          False
     Emi_or_Rent_Amt            False
     Annual_HH_Income           False
     Highest_Qualified_Member   False
     No_of_Earning_Members      False
     dtype: bool
```

```
[9]: df.isnull().sum()
```

```
[9]: Mthly_HH_Income            0
     Mthly_HH_Expense           0
     No_of_Fly_Members          0
     Emi_or_Rent_Amt            0
     Annual_HH_Income           0
     Highest_Qualified_Member   0
     No_of_Earning_Members      0
     dtype: int64
```

0.3 Mean

```
[10]: df["Mthly_HH_Expense"].mean()
```

```
[10]: 18818.0
```

0.4 Median

```
[11]: df['Mthly_HH_Expense'].median()
```

```
[11]: 15500.0
```

0.5 Monthly expense

```
[12]: Mon_Exp=pd.crosstab(index=df["Mthly_HH_Expense"],columns='count')
```

```
[13]: Mon_Exp
```

```
[13]: col_0             count
      Mthly_HH_Expense
```

```
2000                  1
4500                  1
5000                  1
6600                  1
7000                  1
8000                  3
9000                  3
10000                 5
10500                 1
12000                 3
12300                 1
13000                 1
15000                 3
16000                 1
18000                 1
19000                 1
20000                 6
22000                 1
25000                 8
30000                 1
40000                 2
45000                 1
48000                 1
50000                 2
```

[14]:
```
Mon_Exp.reset_index(inplace=True)
Mon_Exp
```

[14]:
```
col_0  Mthly_HH_Expense  count
0                  2000      1
1                  4500      1
2                  5000      1
3                  6600      1
4                  7000      1
5                  8000      3
6                  9000      3
7                 10000      5
8                 10500      1
9                 12000      3
10                12300      1
11                13000      1
12                15000      3
13                16000      1
14                18000      1
15                19000      1
16                20000      6
17                22000      1
```

```
18                 25000      8
19                 30000      1
20                 40000      2
21                 45000      1
22                 48000      1
23                 50000      2
```

[15]:

```
Mon_Exp[Mon_Exp['count']==df.Mthly_HH_Expense.value_counts().max()]
```

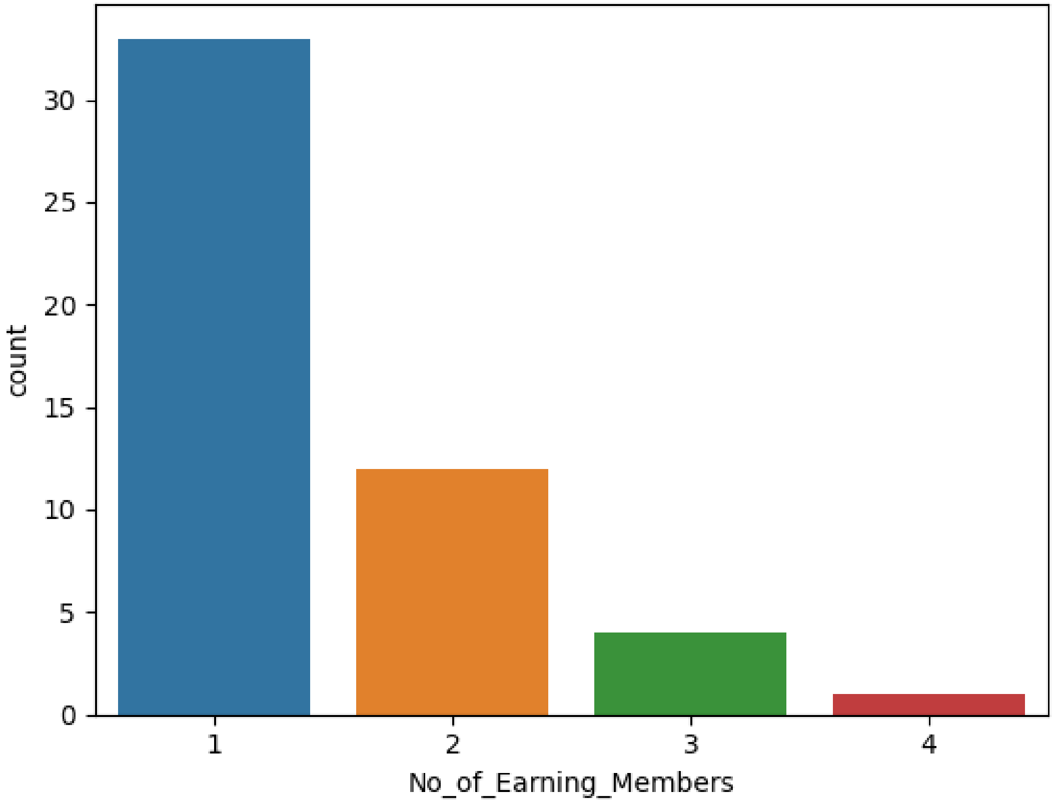
[15]:

```
col_0  Mthly_HH_Expense  count
18                 25000      8
```

0.6 Earning Members

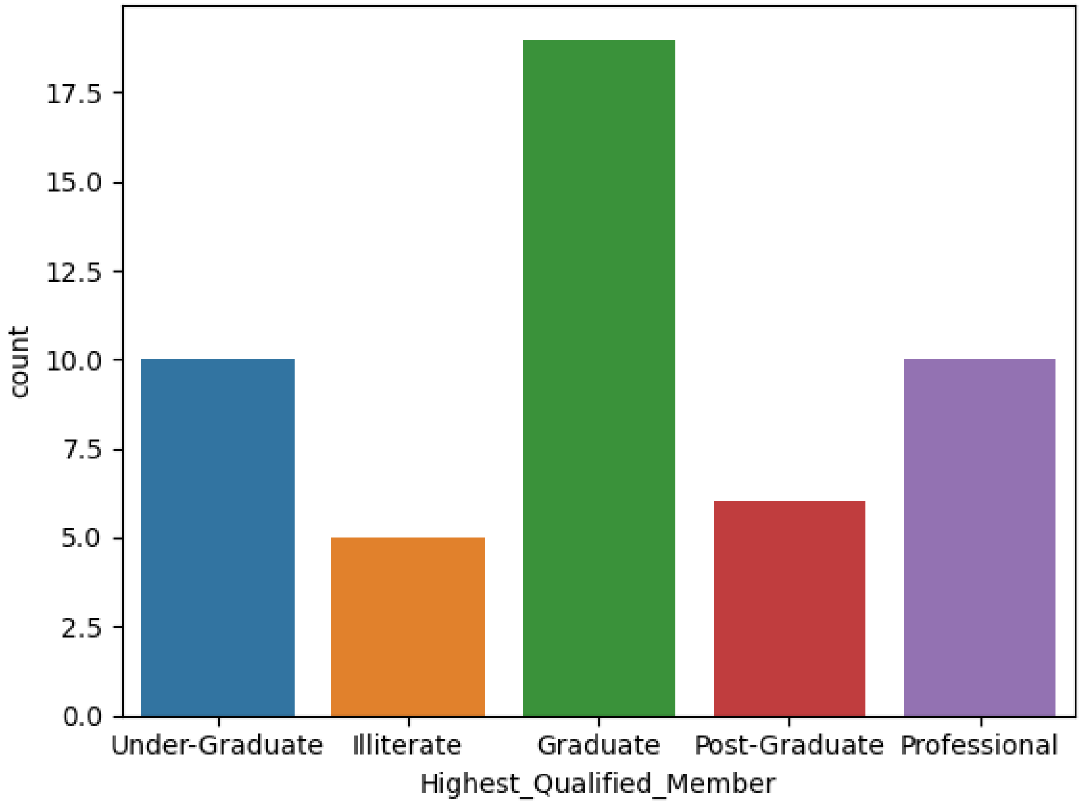
[16]:

```
vis2=sns.countplot(data=df,x='No_of_Earning_Members')
plt.show()
```

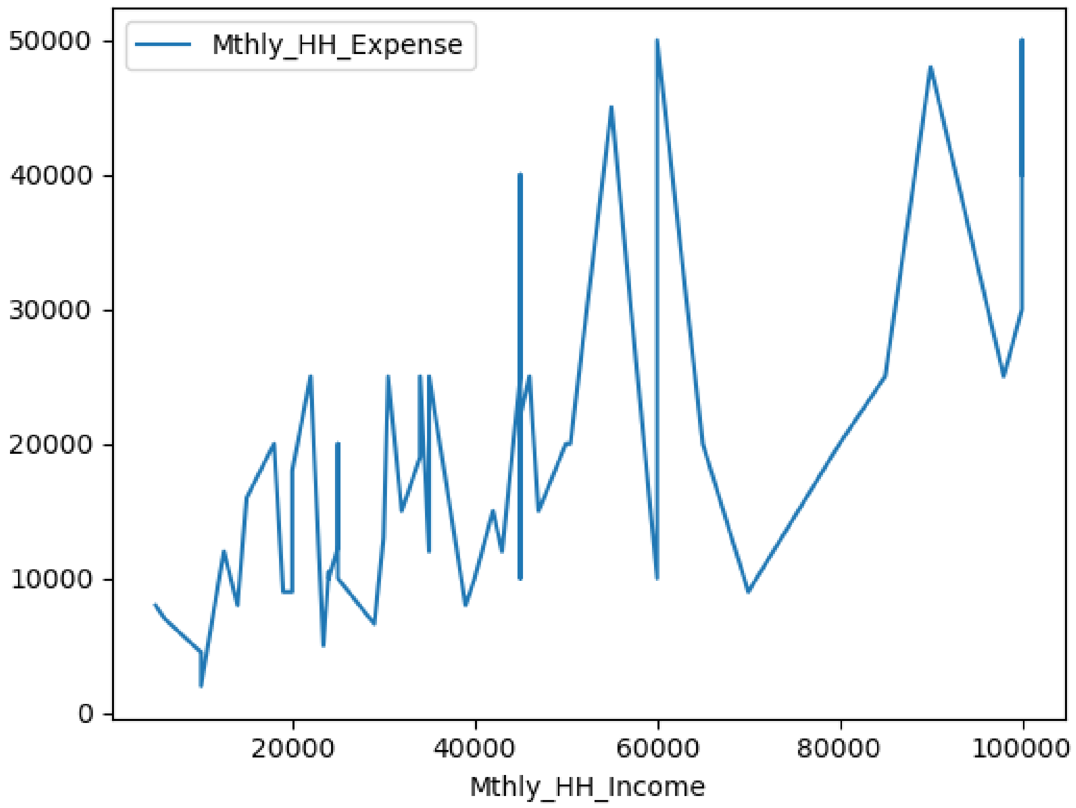


0.7 Highest Qualified Member

[17]:
```
vis=sns.countplot(data=df,x='Highest_Qualified_Member')
```

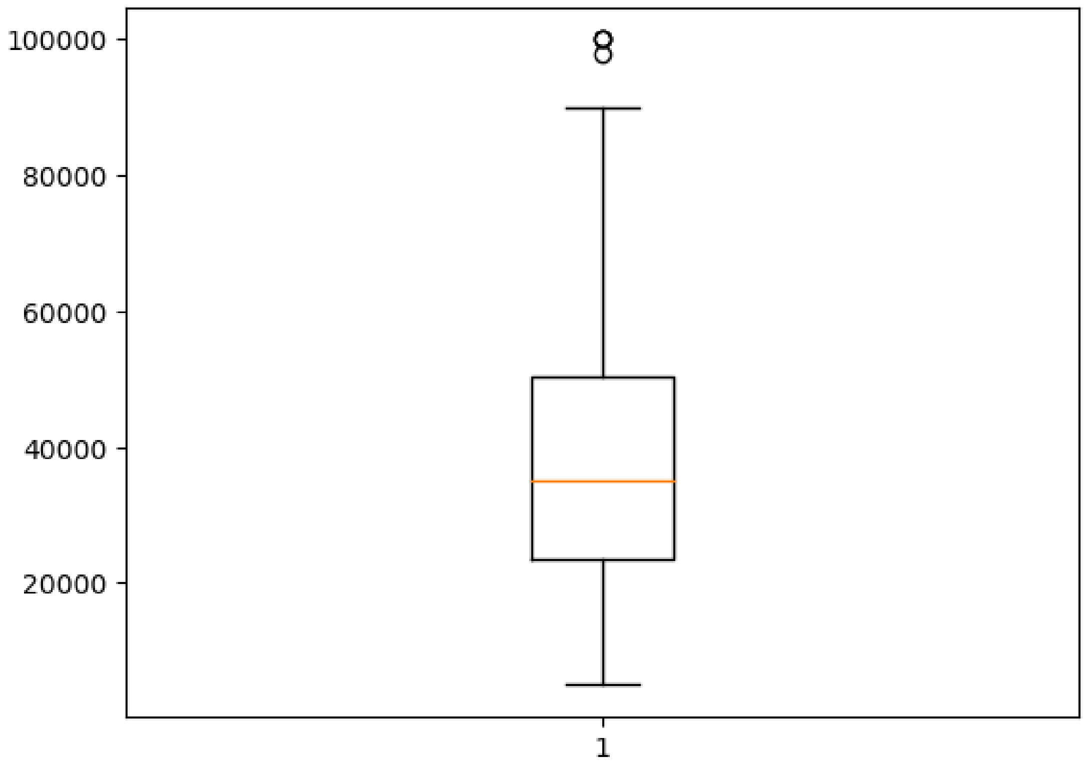


[18]:
```
df.plot(x="Mthly_HH_Income", y="Mthly_HH_Expense");
```



[19]:

```
plt.boxplot(df['Mthly_HH_Income'])
plt.show()
```



0.8 Count Of HQM

```
[20]: df["Highest_Qualified_Member"].value_counts().to_frame().T
```

```
[20]:                           Graduate  Under-Graduate  Professional  \
      Highest_Qualified_Member        19              10            10

                                Post-Graduate  Illiterate
      Highest_Qualified_Member              6           5
```

0.9 IQR

```
[21]: IQR=df["Mthly_HH_Expense"].quantile(0.75)-df["Mthly_HH_Expense"].quantile(0.25)
```

```
[22]: IQR
```

```
[22]: 15000.0
```

0.10 standard Deviation for 1st 10 columns

```
[23]: pd.DataFrame(df.iloc[:,0:10].std().to_frame()).T
```

```
[23]:    Mthly_HH_Income  Mthly_HH_Expense  No_of_Fly_Members  Emi_or_Rent_Amt  \
      0     26097.908979      12090.216824           1.517382      6241.434948

         Annual_HH_Income  No_of_Earning_Members
      0     320135.792123               0.734291
```

0.11 variance for 1st 5

```
[24]: pd.DataFrame(df.iloc[:,0:5].var().to_frame()).T
```

```
[24]:    Mthly_HH_Income  Mthly_HH_Expense  No_of_Fly_Members  Emi_or_Rent_Amt  \
      0     6.811009e+08      1.461733e+08           2.302449     3.895551e+07

         Annual_HH_Income
      0      1.024869e+11
```