MACHINE LEARNING LAB

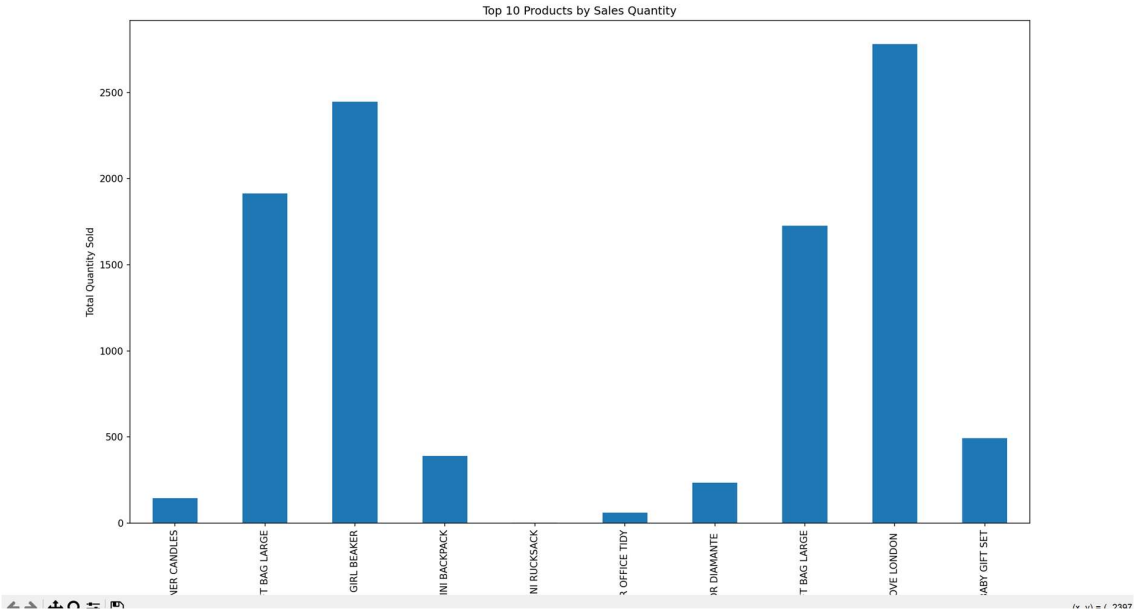
S.KAVIYA – 24BAD 059

SCENARIO – 1 : E – COMMERCE SALES DATA

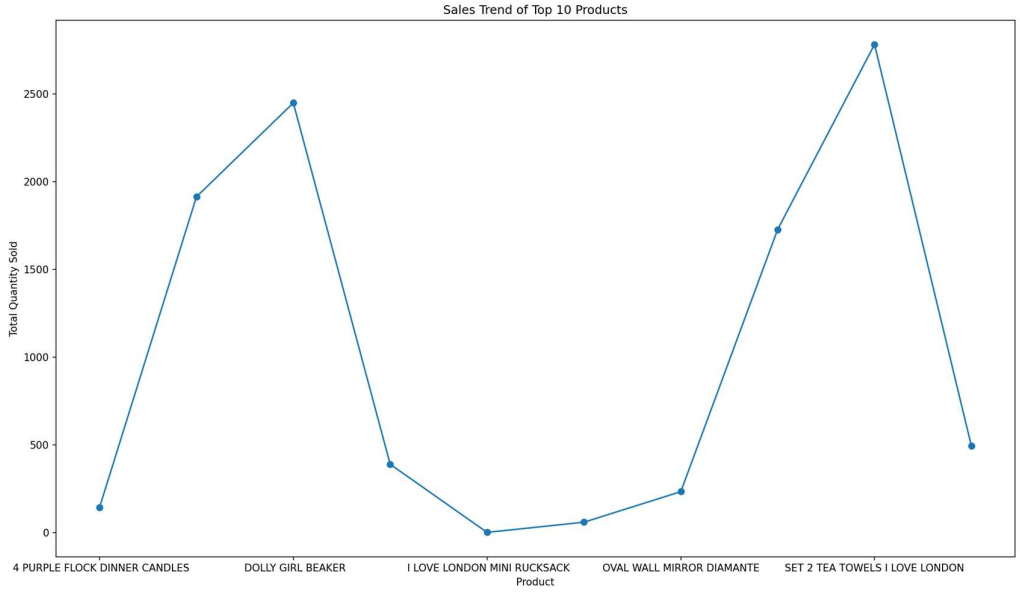
```
S.KAVIYA - 24BAD059
HEAD:
   InvoiceNo StockCode                          Description  ...  UnitPrice CustomerID         Country
0     536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER  ...       2.55    17850.0  United Kingdom
1     536365     71053                  WHITE METAL LANTERN  ...       3.39    17850.0  United Kingdom
2     536365    84406B       CREAM CUPID HEARTS COAT HANGER  ...       2.75    17850.0  United Kingdom
3     536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE  ...       3.39    17850.0  United Kingdom
4     536365    84029E       RED WOOLLY HOTTIE WHITE HEART.  ...       3.39    17850.0  United Kingdom

[5 rows x 8 columns]

TAIL:
        InvoiceNo StockCode                       Description  ...  UnitPrice CustomerID  Country
541904    581587     22613      PACK OF 20 SPACEBOY NAPKINS  ...       0.85    12680.0   France
541905    581587     22899     CHILDREN'S APRON DOLLY GIRL   ...       2.10    12680.0   France
541906    581587     23254    CHILDRENS CUTLERY DOLLY GIRL   ...       4.15    12680.0   France
541907    581587     23255  CHILDRENS CUTLERY CIRCUS PARADE  ...       4.15    12680.0   France
541908    581587     22138    BAKING SET 9 PIECE RETROSPOT   ...       4.95    12680.0   France

[5 rows x 8 columns]

INFO:
<class 'pandas.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    541909 non-null  str
 1   StockCode    541909 non-null  str
 2   Description  540455 non-null  str
 3   Quantity     541909 non-null  int64
 4   InvoiceDate  541909 non-null  str
 5   UnitPrice    541909 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      541909 non-null  str
dtypes: float64(2), int64(1), str(5)
```

```
DESCRIBE:
        InvoiceNo StockCode                      Description  ...      UnitPrice     CustomerID         Country
count      541909    541909                           540455  ...  541909.000000  406829.000000          541909
unique      25900      4070                             4223  ...            NaN            NaN              38
top        573585    85123A  WHITE HANGING HEART T-LIGHT HOLDER  ...            NaN            NaN  United Kingdom
freq         1114      2313                             2369  ...            NaN            NaN          495478
mean          NaN       NaN                              NaN  ...       4.611114   15287.690570             NaN
std           NaN       NaN                              NaN  ...      96.759853    1713.600303             NaN
min           NaN       NaN                              NaN  ...  -11062.060000   12346.000000             NaN
0              NaN
max           NaN       NaN                              NaN  ...   38970.000000   18287.000000             NaN

[11 rows x 8 columns]

Missing Values:
 InvoiceNo          0
StockCode           0
Description      1454
Quantity            0
InvoiceDate         0
UnitPrice           0
CustomerID     135080
Country             0
dtype: int64
```

BAR CHART :



LINE CHART :

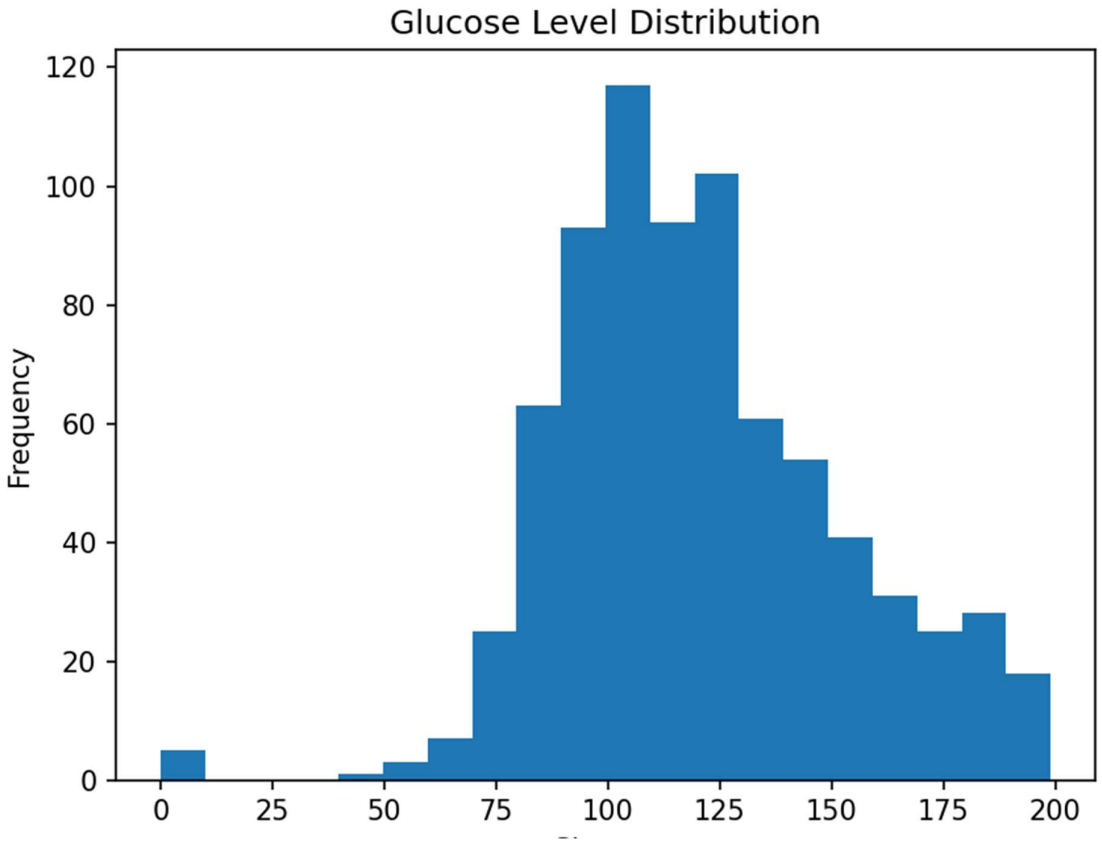


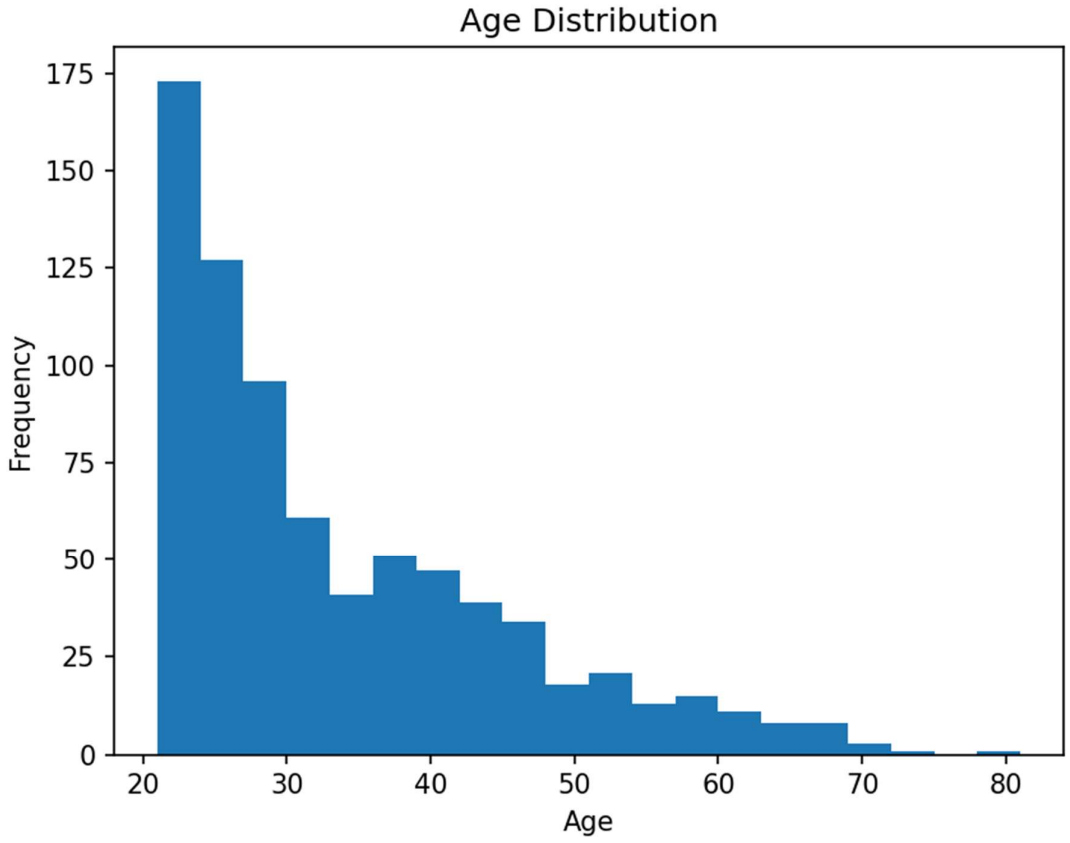
SCENARIO 2 : DIABETES

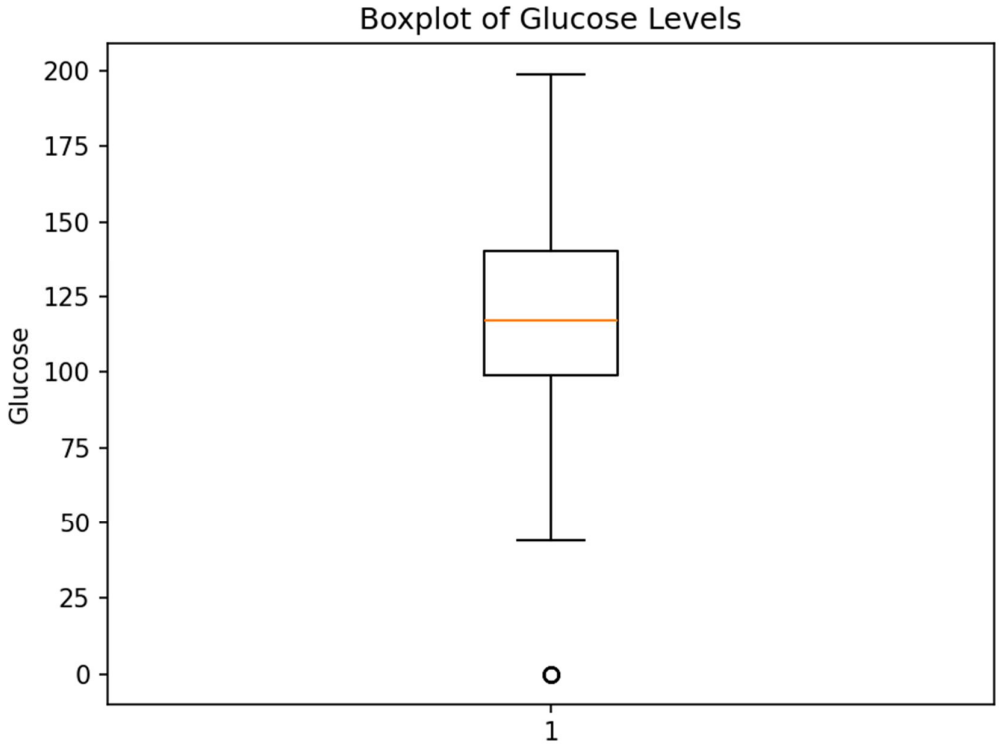
```
S.KAVIYA - 24BAD059
HEAD:
    Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  DiabetesPedigreeFunction  Age  Outcome
0             6      148             72             35        0  33.6                     0.627   50        1
1             1       85             66             29        0  26.6                     0.351   31        0
2             8      183             64              0        0  23.3                     0.672   32        1
3             1       89             66             23       94  28.1                     0.167   21        0
4             0      137             40             35      168  43.1                     2.288   33        1

INFO:
<class 'pandas.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64
 8   Outcome                   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB

DESCRIBE:
        Pregnancies     Glucose  BloodPressure  ...  DiabetesPedigreeFunction         Age     Outcome
count   768.000000  768.000000     768.000000  ...                768.000000  768.000000  768.000000
mean      3.845052  120.894531      69.105469  ...                  0.471876   33.240885    0.348958
std       3.369578   31.972618      19.355807  ...                  0.331329   11.760232    0.476951
min       0.000000    0.000000       0.000000  ...                  0.078000   21.000000    0.000000
25%       1.000000   99.000000      62.000000  ...                  0.243750   24.000000    0.000000
50%       3.000000  117.000000      72.000000  ...                  0.372500   29.000000    0.000000
75%       6.000000  140.250000      80.000000  ...                  0.626250   41.000000    1.000000
max      17.000000  199.000000     122.000000  ...                  2.420000   81.000000    1.000000

[8 rows x 9 columns]
```

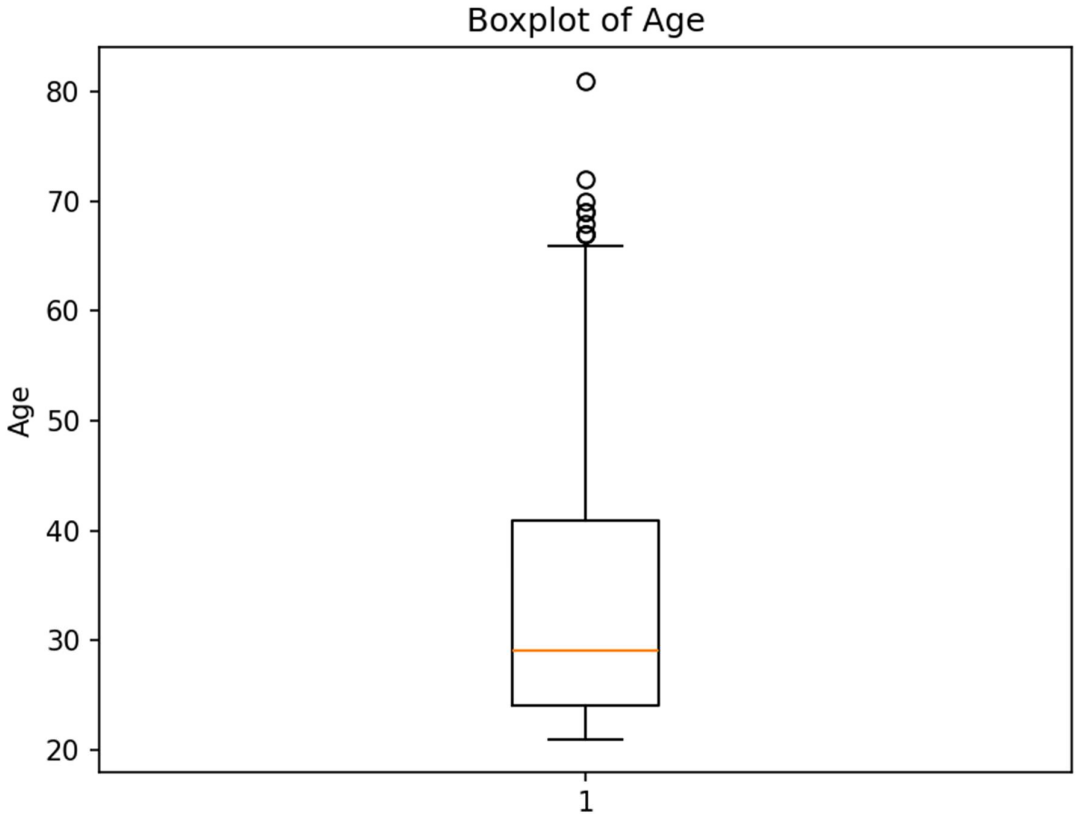
```
Missing Values:
 Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64

Zero Values (treated as missing):

Pregnancies                 111
Glucose                       5
BloodPressure                35
SkinThickness               227
Insulin                     374
BMI                          11
DiabetesPedigreeFunction      0
Age                           0
Outcome                     500
dtype: int64
```

HISTOGRAM


Glucose Level Distribution
Frequency
0
20
40
60
80
100
120
0
25
50
75
100
125
150
175
200


Age Distribution
Frequency
0
25
50
75
100
125
150
175
20
30
40
50
60
70
80
Age


Boxplot of Glucose Levels
Glucose
200
175
150
125
100
75
50
25
0
1


Boxplot of Age
Age
80
70
60
50
40
30
20
1

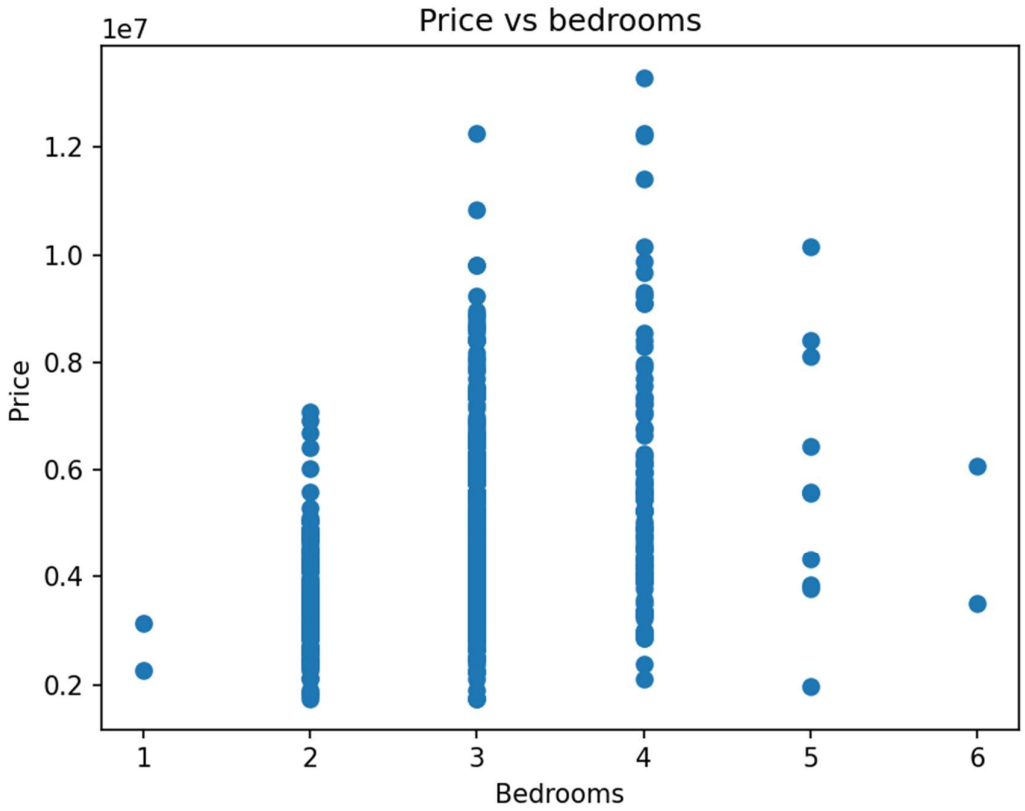
SCENARIO – 3 – HOUSING

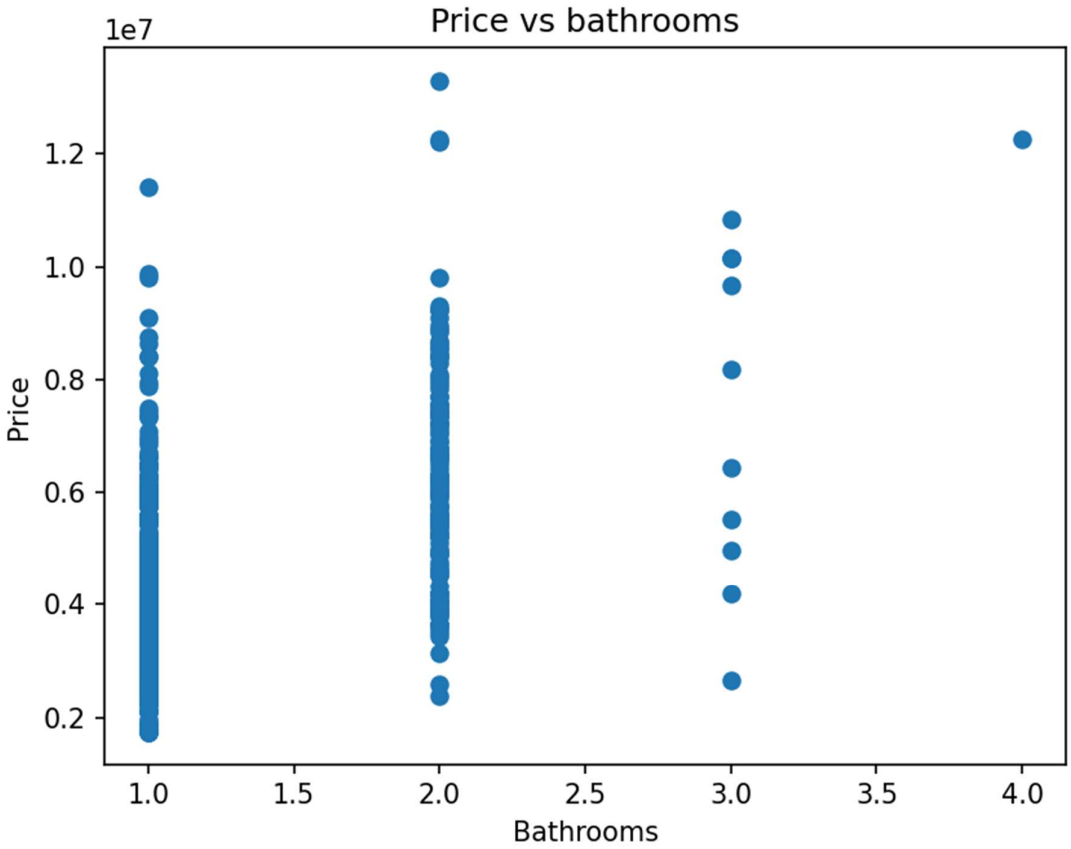
```
S.KAVIYA - 24BAD059
COLUMNS:
 Index(['price', 'area', 'bedrooms', 'bathrooms', 'stories', 'mainroad',
       'guestroom', 'basement', 'hotwaterheating', 'airconditioning',
       'parking', 'prefarea', 'furnishingstatus'],
      dtype='str')

HEAD:
      price  area  bedrooms  bathrooms  stories  ... hotwaterheating airconditioning parking prefarea furnishings
atus
0  13300000  7420         4          2        3  ...              no             yes       2      yes        furni
hed
1  12250000  8960         4          4        4  ...              no             yes       3       no        furni
hed
2  12250000  9960         3          2        2  ...              no              no       2      yes   semi-furni
hed
3  12215000  7500         4          2        2  ...              no             yes       3      yes        furni
hed
4  11410000  7420         4          1        2  ...              no             yes       2       no        furni
hed

[5 rows x 13 columns]

INFO:
<class 'pandas.DataFrame'>
RangeIndex: 545 entries, 0 to 544
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   price             545 non-null    int64
 1   area              545 non-null    int64
 2   bedrooms          545 non-null    int64
 3   bathrooms         545 non-null    int64
 4   stories           545 non-null    int64
 5   mainroad          545 non-null    str
 6   guestroom         545 non-null    str
 7   basement          545 non-null    str
 8   hotwaterheating   545 non-null    str
 9   airconditioning   545 non-null    str
 10  parking           545 non-null    int64
```

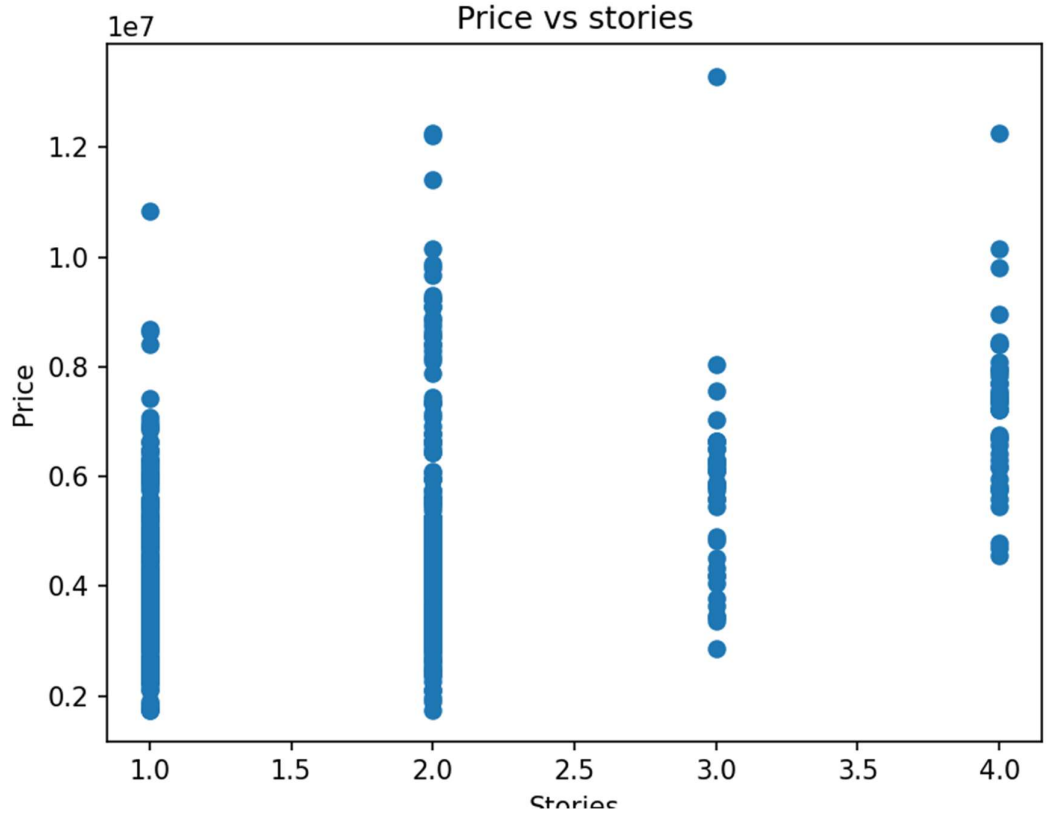
A Java Runtime Environment was not found. The

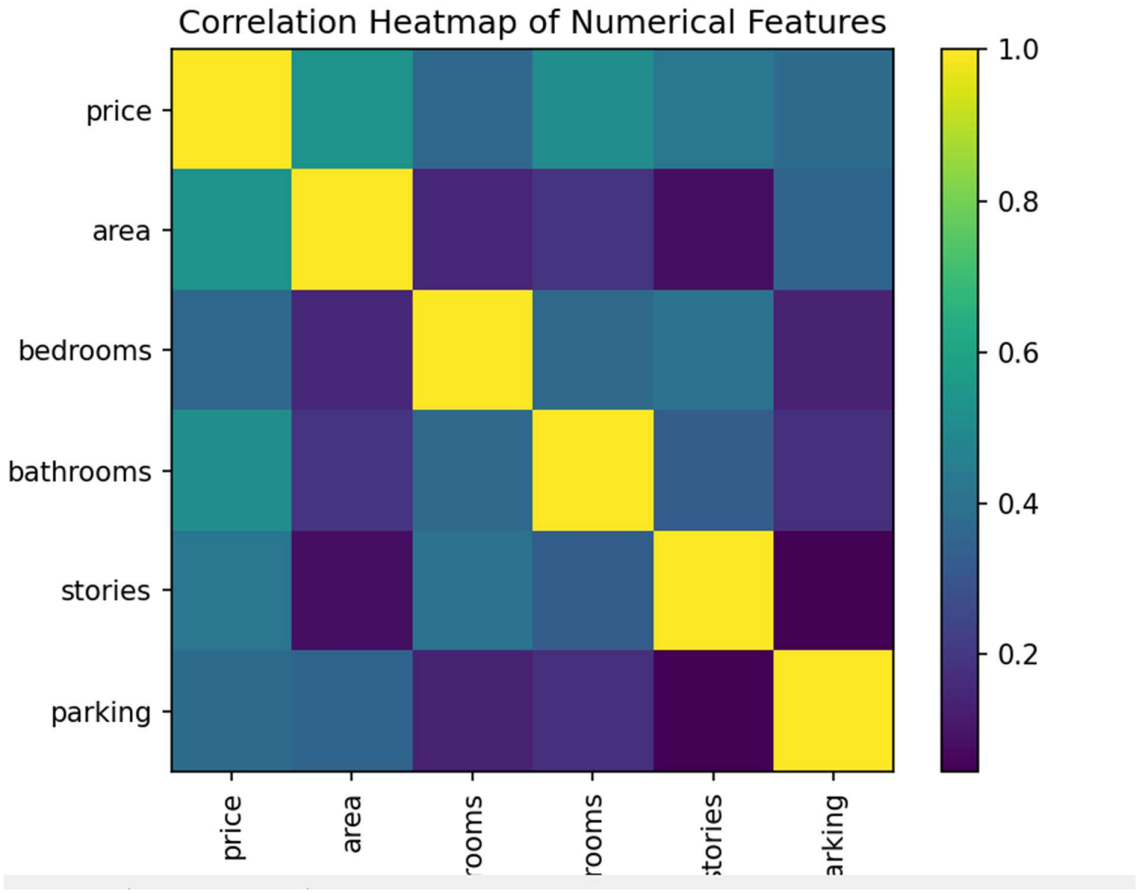
```
INFO:
<class 'pandas.DataFrame'>
RangeIndex: 545 entries, 0 to 544
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   price             545 non-null    int64
 1   area              545 non-null    int64
 2   bedrooms          545 non-null    int64
 3   bathrooms         545 non-null    int64
 4   stories           545 non-null    int64
 5   mainroad          545 non-null    str
 6   guestroom         545 non-null    str
 7   basement          545 non-null    str
 8   hotwaterheating   545 non-null    str
 9   airconditioning   545 non-null    str
 10  parking           545 non-null    int64
 11  prefarea          545 non-null    str
 12  furnishingstatus  545 non-null    str
dtypes: int64(6), str(7)
memory usage: 55.5 KB

DESCRIBE:
              price          area    bedrooms   bathrooms     stories     parking
count  5.450000e+02    545.000000  545.000000  545.000000  545.000000  545.000000
mean   4.766729e+06   5150.541284    2.965138    1.286239    1.805505    0.693578
std    1.870440e+06   2170.141023    0.738064    0.502470    0.867492    0.861586
min    1.750000e+06   1650.000000    1.000000    1.000000    1.000000    0.000000
25%    3.430000e+06   3600.000000    2.000000    1.000000    1.000000    0.000000
50%    4.340000e+06   4600.000000    3.000000    1.000000    2.000000    0.000000
75%    5.740000e+06   6360.000000    3.000000    2.000000    2.000000    1.000000
max    1.330000e+07  16200.000000    6.000000    4.000000    4.000000    3.000000

Missing Values per Column:
 Series([], dtype: int64)
```

SCATTER PLOT :


Price vs bedrooms
1e7
1.2
1.0
0.8
0.6
0.4
0.2
Price
1
2
3
4
5
6
Bedrooms


Price vs bathrooms
1e7
1.2
1.0
0.8
0.6
0.4
0.2
Price
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Bathrooms


Price vs stories
1e7
1.2
1.0
0.8
0.6
0.4
0.2
Price
1.0
1.5
2.0
2.5
3.0
3.5
4.0



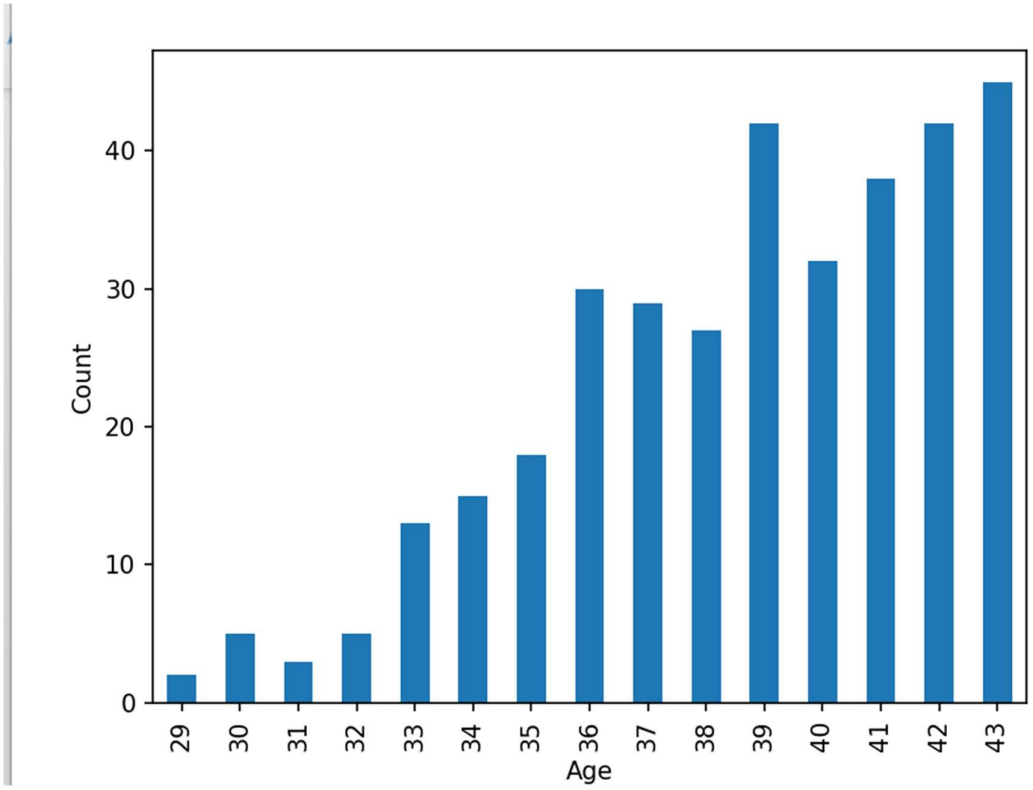
SCENARIO 4 – MARKETING CAMPAIGN

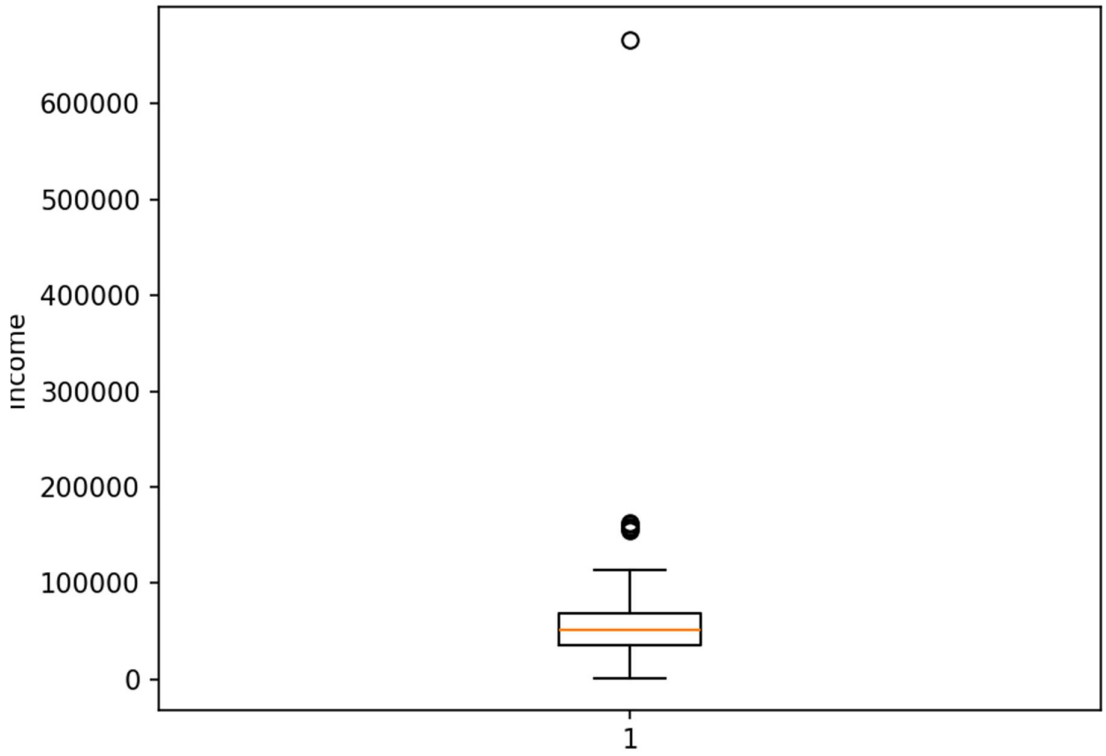
```
PS C:\Users\Kaviya\Desktop\Machine_learning_lab> & C:/Users/Kaviya/AppDa
ython.exe c:/Users/Kaviya/Desktop/Machine_learning_lab/analysis.4.py
     ID  Year_Birth   Education  ... Z_CostContact  Z_Revenue  Response
0  5524        1957  Graduation  ...             3         11         1
1  2174        1954  Graduation  ...             3         11         0
2  4141        1965  Graduation  ...             3         11         0
3  6182        1984  Graduation  ...             3         11         0
4  5324        1981         PhD  ...             3         11         0

[5 rows x 29 columns]
<class 'pandas.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 29 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   ID                   2240 non-null   int64
 1   Year_Birth           2240 non-null   int64
 2   Education            2240 non-null   str
 3   Marital_Status       2240 non-null   str
 4   Income               2216 non-null   float64
 5   Kidhome              2240 non-null   int64
 6   Teenhome             2240 non-null   int64
 7   Dt_Customer          2240 non-null   str
 8   Recency              2240 non-null   int64
 9   MntWines             2240 non-null   int64
 10  MntFruits            2240 non-null   int64
 11  MntMeatProducts      2240 non-null   int64
 12  MntFishProducts      2240 non-null   int64
 13  MntSweetProducts     2240 non-null   int64
 14  MntGoldProds         2240 non-null   int64
 15  NumDealsPurchases    2240 non-null   int64
 16  NumWebPurchases      2240 non-null   int64
 17  NumCatalogPurchases  2240 non-null   int64
 18  NumStorePurchases    2240 non-null   int64
 19  NumWebVisitsMonth    2240 non-null   int64
 20  AcceptedCmp3         2240 non-null   int64
 21  AcceptedCmp4         2240 non-null   int64
 22  AcceptedCmp5         2240 non-null   int64
```

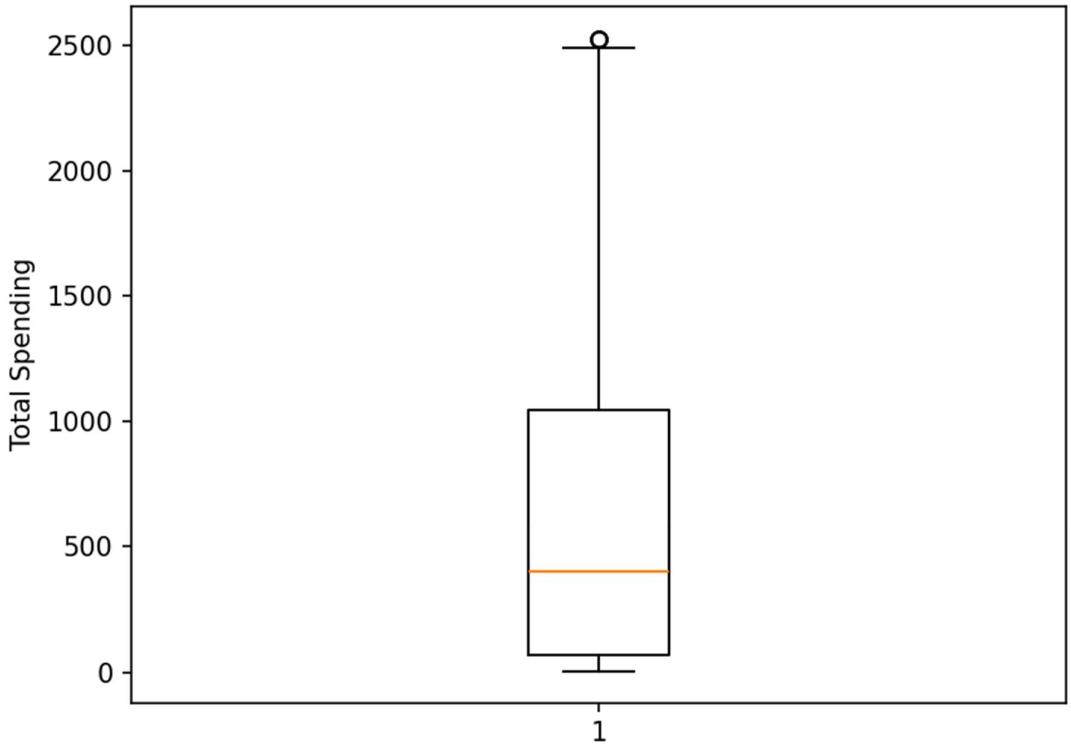
```
 26  Z_CostContact        2240 non-null   int64
 27  Z_Revenue            2240 non-null   int64
 28  Response             2240 non-null   int64
dtypes: float64(1), int64(25), str(3)
memory usage: 507.6 KB
                ID   Year_Birth         Income  ...  Z_CostContact  Z_Revenue     Response
count   2240.000000  2240.000000    2216.000000  ...         2240.0     2240.0  2240.000000
mean    5592.159821  1968.805804   52247.251354  ...            3.0       11.0     0.149107
std     3246.662198    11.984069   25173.076661  ...            0.0        0.0     0.356274
min        0.000000  1893.000000    1730.000000  ...            3.0       11.0     0.000000
25%     2828.250000  1959.000000   35303.000000  ...            3.0       11.0     0.000000
50%     5458.500000  1970.000000   51381.500000  ...            3.0       11.0     0.000000
75%     8427.750000  1977.000000   68522.000000  ...            3.0       11.0     0.000000
max    11191.000000  1996.000000  666666.000000  ...            3.0       11.0     1.000000

[8 rows x 26 columns]
ID                     0
Year_Birth             0
Education              0
Marital_Status         0
Income                24
Kidhome                0
Teenhome               0
Dt_Customer            0
Recency                0
MntWines               0
MntFruits              0
MntMeatProducts        0
MntFishProducts        0
MntSweetProducts       0
MntGoldProds           0
NumDealsPurchases      0
NumWebPurchases        0
NumCatalogPurchases    0
NumStorePurchases      0
NumWebVisitsMonth      0
```

```
NumWebVisitsMonth      0
AcceptedCmp3           0
AcceptedCmp4           0
AcceptedCmp5           0
AcceptedCmp1           0
AcceptedCmp2           0
Complain               0
Z_CostContact          0
Z_Revenue              0
Response               0
dtype: int64
```

BAR CHART :


Count
0
10
20
30
40
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
Age


Income
0
100000
200000
300000
400000
500000
600000
1


2500
2000
1500
1000
500
0
Total Spending
1