```
import pandas as pd            # to work with tables (dataframes)
import matplotlib.pyplot as plt # to make graphs
import seaborn as sns           # for prettier graphs

df = pd.read_csv(r"E:\DA Project\
cleaned_customer_personality_analysis.xlsx.csv") # load file
df.head() #head() shows the first 5 rows.
```

```
      id  year_birth  age education marital_status  income  kidhome
teenhome  \
0  11004        1893  132  2n cycle         single   60182        0
1
1   1150        1899  126       phd        married   83532        0
0
2   7829        1900  125  2n cycle       divorced   36640        1
0
3   6663        1940   85       phd         single   51141        0
0
4   6932        1941   84       phd        married   93027        0
0

   family_size dt_customer  ...  acceptedcmp3\t  acceptedcmp4\t  \
0            1  17-05-2014  ...               0               0
1            0  26-09-2013  ...               0               0
2            1  26-09-2013  ...               0               0
3            0  08-07-2013  ...               0               0
4            0  13-04-2013  ...               0               0

   acceptedcmp5\t  acceptedcmp1\t  acceptedcmp2\t
total_campaign_accepted  \
0               0               0               0
0
1               1               0               0
1
2               0               0               0
0
3               0               0               0
0
4               1               0               0
1

   complain\t  z_costcontact\t  z_revenue\t  response
0           0                3           11         0
1           0                3           11         0
2           1                3           11         0
3           0                3           11         0
4           0                3           11         0
```

```
[5 rows x 33 columns]
```

```
df.shape         # Shows count of  (rows, columns)
```

```
(2240, 33)
```

```
df.info()  #info() shows column names, types, and missing values.
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 33 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       2240 non-null   int64
 1   year_birth               2240 non-null   int64
 2   age                      2240 non-null   int64
 3   education                2240 non-null   object
 4   marital_status           2240 non-null   object
 5   income                   2240 non-null   int64
 6   kidhome                  2240 non-null   int64
 7   teenhome                 2240 non-null   int64
 8   family_size              2240 non-null   int64
 9   dt_customer              2240 non-null   object
 10  recency                  2240 non-null   int64
 11  mntWines                 2240 non-null   int64
 12  mntFruits                2240 non-null   int64
 13  mntmeatproducts          2240 non-null   int64
 14  mntfishproducts          2240 non-null   int64
 15  mntsweetproducts         2240 non-null   int64
 16  mntgoldprods             2240 non-null   int64
 17  total_spend              2240 non-null   int64
 18  numdealspurchases        2240 non-null   int64
 19  numwebpurchases          2240 non-null   int64
 20  numcatalogpurchases      2240 non-null   int64
 21  numstorepurchases        2240 non-null   int64
 22  numwebvisitsmonth            2240 non-null   int64
 23  acceptedcmp3                 2240 non-null   int64
 24  acceptedcmp4                 2240 non-null   int64
 25  acceptedcmp5                 2240 non-null   int64
 26  acceptedcmp1                 2240 non-null   int64
 27  acceptedcmp2                 2240 non-null   int64
 28  total_campaign_accepted  2240 non-null   int64
 29  complain                   2240 non-null   int64
 30  z_costcontact               2240 non-null   int64
 31  z_revenue                 2240 non-null   int64
 32  response                 2240 non-null   int64
dtypes: int64(30), object(3)
memory usage: 577.6+ KB
```

```
df.describe() #df.describe → shows the function itself
               #df.describe() → runs the function and gives summary
statistics

                  id    year_birth           age        income  
kidhome  \
count    2240.000000  2240.000000  2240.000000    2240.000000   
2240.000000   
mean     5592.159821  1968.805804    56.194196   51754.601786   
0.444196   
std      3246.662198    11.984069    11.984069   25481.939355   
0.538398   
min         0.000000  1893.000000    29.000000       0.000000   
0.000000   
25%      2828.250000  1959.000000    48.000000   34722.000000   
0.000000   
50%      5458.500000  1970.000000    55.000000   51075.000000   
0.000000   
75%      8427.750000  1977.000000    66.000000   68289.750000   
1.000000   
max     11191.000000  1996.000000   132.000000  666666.000000   
2.000000   

          teenhome  family_size      recency     mntWines    mntFruits  
...  \
count  2240.000000  2240.000000  2240.000000  2240.000000  2240.000000  
...   
mean      0.506250     0.950446    49.109375   303.935714    26.302232  
...   
std       0.544538     0.751803    28.962453   336.597393    39.773434  
...   
min       0.000000     0.000000     0.000000     0.000000     0.000000  
...   
25%       0.000000     0.000000    24.000000    23.750000     1.000000  
...   
50%       0.000000     1.000000    49.000000   173.500000     8.000000  
...   
75%       1.000000     1.000000    74.000000   504.250000    33.000000  
...   
max       2.000000     3.000000    99.000000  1493.000000   199.000000  
...   

       acceptedcmp3\t  acceptedcmp4\t  acceptedcmp5\t  acceptedcmp1\t  
\
count     2240.000000     2240.000000     2240.000000     2240.000000   

mean         0.072768        0.074554        0.072768        0.064286   

std          0.259813        0.262728        0.259813        0.245316   
```

```
min          0.000000        0.000000        0.000000        0.000000

25%          0.000000        0.000000        0.000000        0.000000

50%          0.000000        0.000000        0.000000        0.000000

75%          0.000000        0.000000        0.000000        0.000000

max          1.000000        1.000000        1.000000        1.000000


       acceptedcmp2\t  total_campaign_accepted   complain\t  
z_costcontact\t  \
count     2240.000000              2240.000000  2240.000000   
2240.0
mean         0.013393                 0.297768     0.009375   
3.0
std          0.114976                 0.678381     0.096391   
0.0
min          0.000000                 0.000000     0.000000   
3.0
25%          0.000000                 0.000000     0.000000   
3.0
50%          0.000000                 0.000000     0.000000   
3.0
75%          0.000000                 0.000000     0.000000   
3.0
max          1.000000                 4.000000     1.000000   
3.0

       z_revenue\t     response
count       2240.0  2240.000000
mean          11.0     0.149107
std            0.0     0.356274
min           11.0     0.000000
25%           11.0     0.000000
50%           11.0     0.000000
75%           11.0     0.000000
max           11.0     1.000000

[8 rows x 30 columns]
```

```
df.isnull().sum()  #Shows how many missing values per column.
```

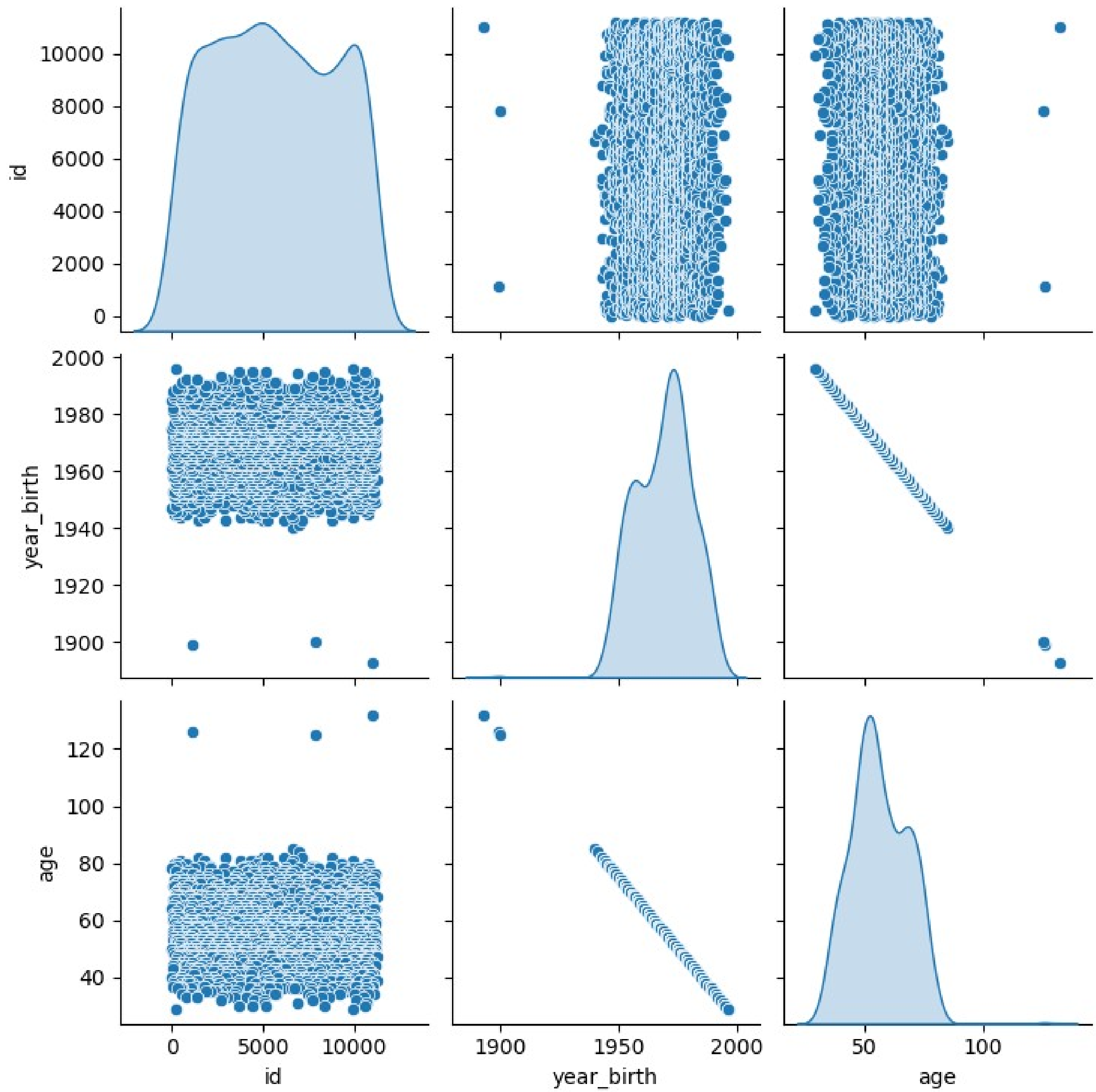
```
id                         0
year_birth                 0
age                        0
education                  0
marital_status             0
income                     0
```

```
kidhome                    0
teenhome                   0
family_size                0
dt_customer                0
recency                    0
mntWines                   0
mntFruits                  0
mntmeatproducts            0
mntfishproducts            0
mntsweetproducts           0
mntgoldprods               0
total_spend                0
numdealspurchases          0
numwebpurchases            0
numcatalogpurchases        0
numstorepurchases          0
numwebvisitsmonth\t        0
acceptedcmp3\t             0
acceptedcmp4\t             0
acceptedcmp5\t             0
acceptedcmp1\t             0
acceptedcmp2\t             0
total_campaign_accepted    0
complain\t                 0
z_costcontact\t            0
z_revenue\t                0
response                   0
dtype: int64
```

```python
num_cols = df.select_dtypes(include=['number']).columns.tolist()
cat_cols = df.select_dtypes(include=['object']).columns.tolist()

print("Numeric columns:", num_cols)
print("Categorical columns:", cat_cols)

#Makes two lists – one for numeric columns, one for text-based columns.
```

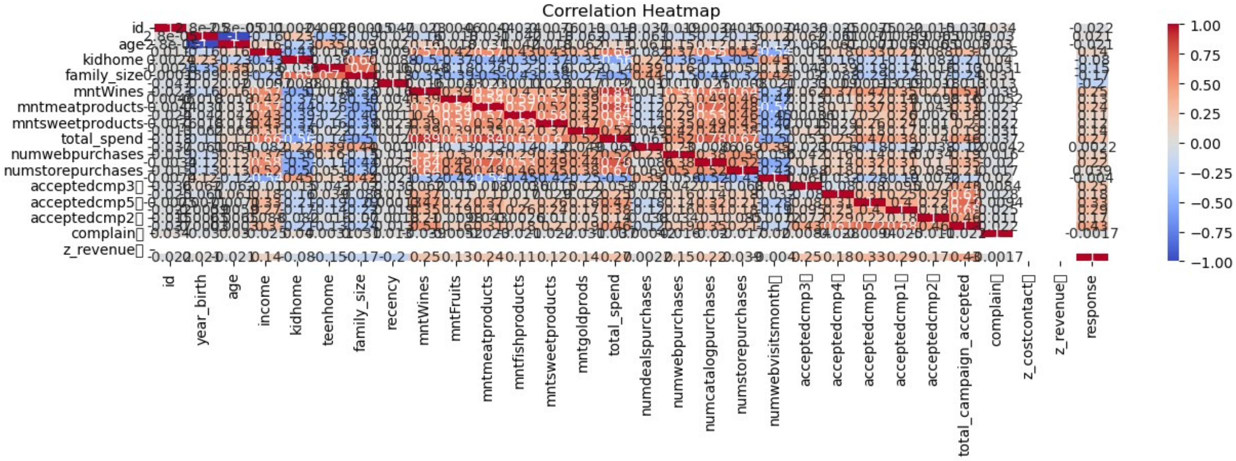
```
Numeric columns: ['id', 'year_birth', 'age', 'income', 'kidhome',
'teenhome', 'family_size', 'recency', 'mntWines', 'mntFruits',
'mntmeatproducts', 'mntfishproducts', 'mntsweetproducts',
'mntgoldprods', 'total_spend', 'numdealspurchases', 'numwebpurchases',
'numcatalogpurchases', 'numstorepurchases', 'numwebvisitsmonth\t',
'acceptedcmp3\t', 'acceptedcmp4\t', 'acceptedcmp5\t', 'acceptedcmp1\
t', 'acceptedcmp2\t', 'total_campaign_accepted', 'complain\t',
'z_costcontact\t', 'z_revenue\t', 'response']
Categorical columns: ['education', 'marital_status', 'dt_customer']
```

```
# Sample if dataset is large
sns.pairplot(df[num_cols[:3]], diag_kind='kde')  # first 5 numeric
cols for clarity
plt.show()
```

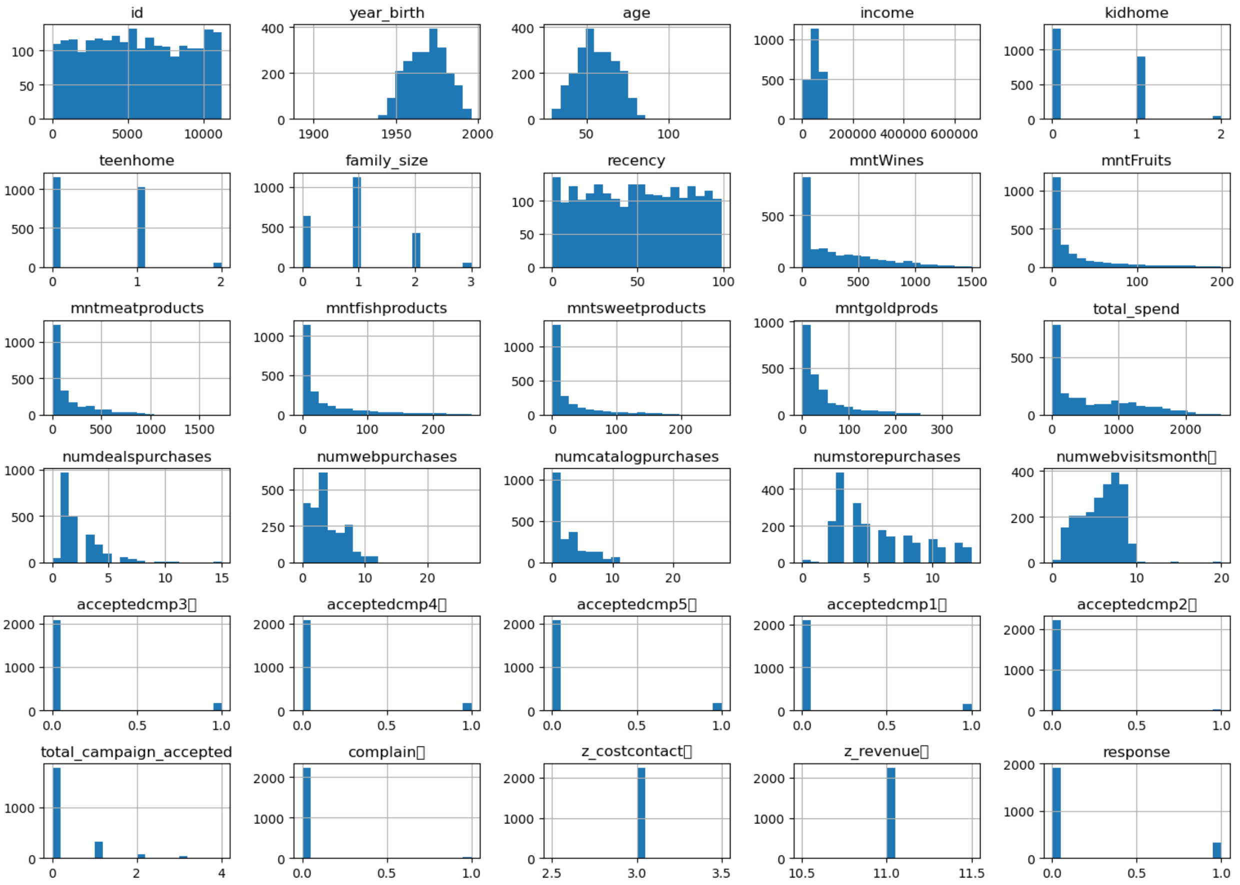


```
plt.figure(figsize=(15,3))
sns.heatmap(df[num_cols].corr(), annot=True, cmap='coolwarm',
center=0)
plt.title("Correlation Heatmap")
plt.show()
```

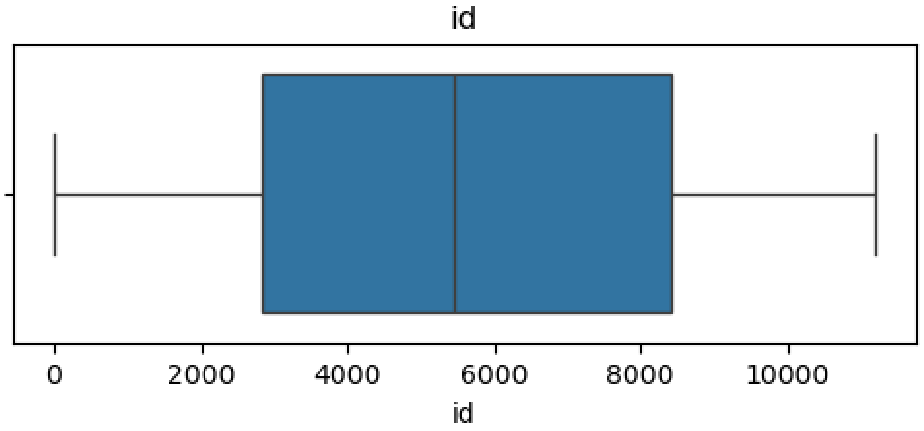
```
C:\Users\ADMIN\anaconda3\Lib\site-packages\seaborn\utils.py:61:
UserWarning: Glyph 9 (      ) missing from font(s) DejaVu Sans.
  fig.canvas.draw()
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```

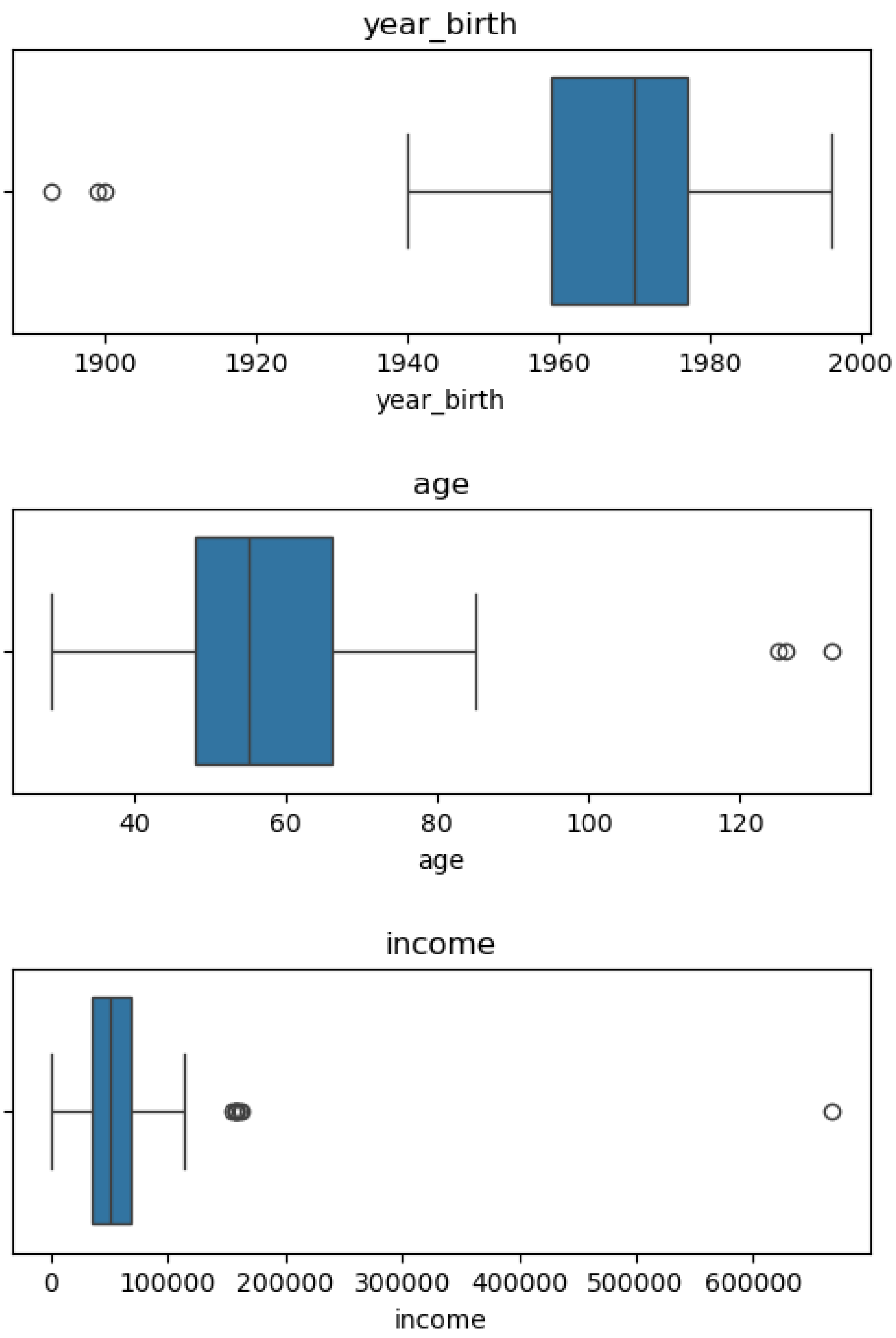


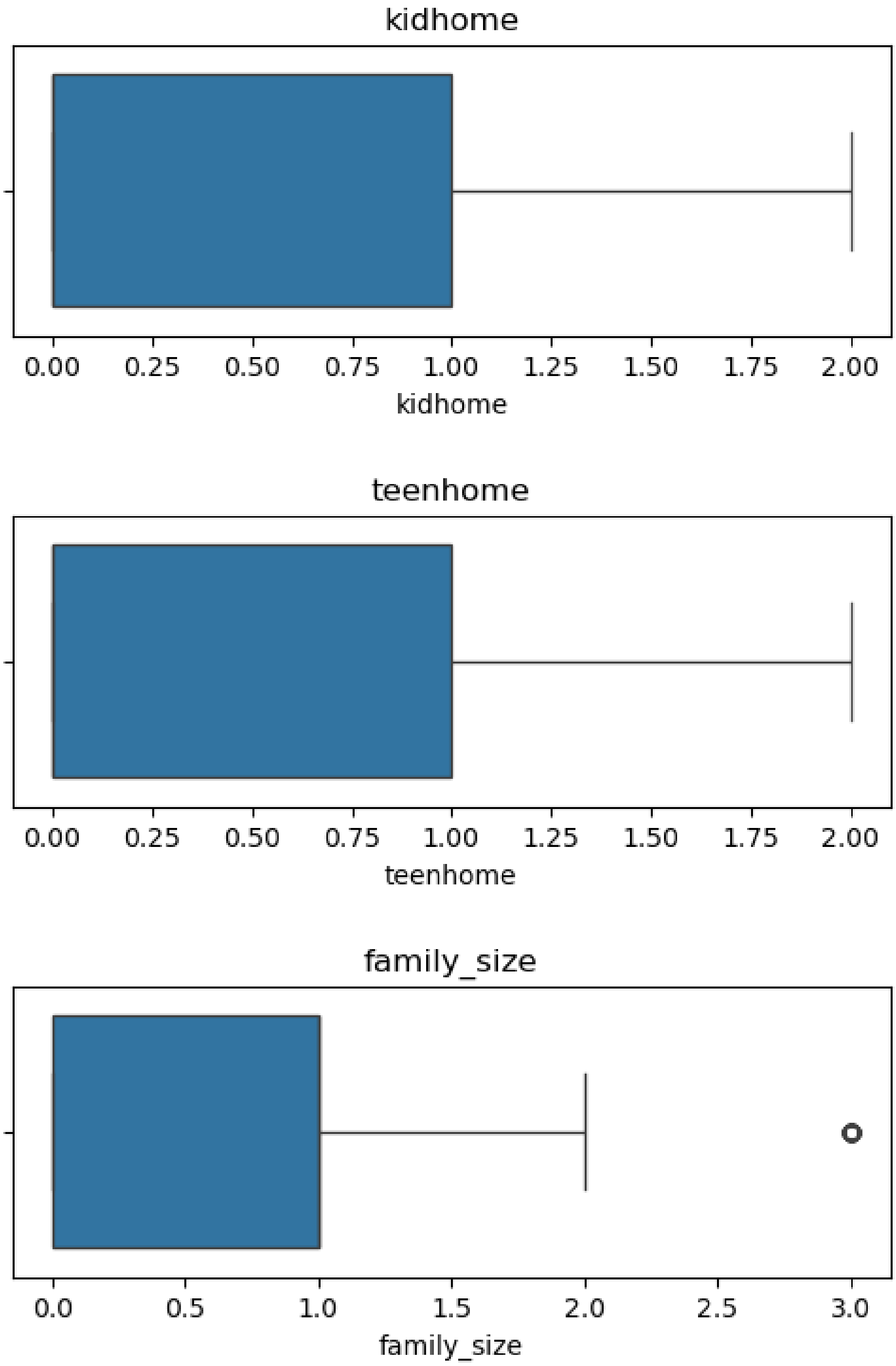
```
df[num_cols].hist(bins=20, figsize=(14,10))
plt.tight_layout()
plt.show()

C:\Users\ADMIN\AppData\Local\Temp\ipykernel_15384\2530790824.py:2:
UserWarning: Glyph 9 (      ) missing from font(s) DejaVu Sans.
  plt.tight_layout()
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



```
for col in num_cols:
    plt.figure(figsize=(6,2))
    sns.boxplot(x=df[col])
    plt.title(col)
    plt.show()
```




year_birth
1900
1920
1940
1960
1980
2000
year_birth
age
40
60
80
100
120
age
income
0
100000
200000
300000
400000
500000
600000
income

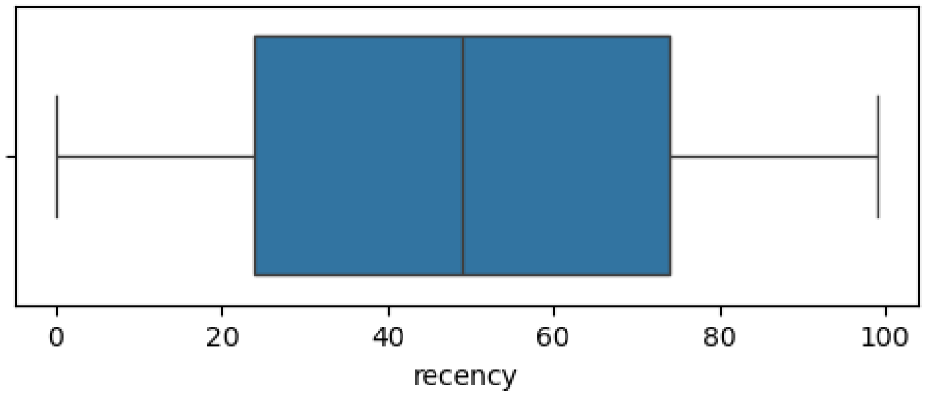


kidhome

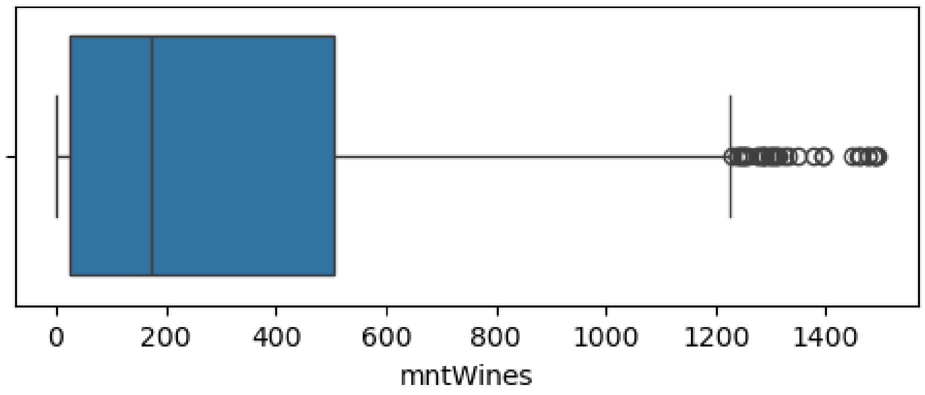
teenhome

family_size

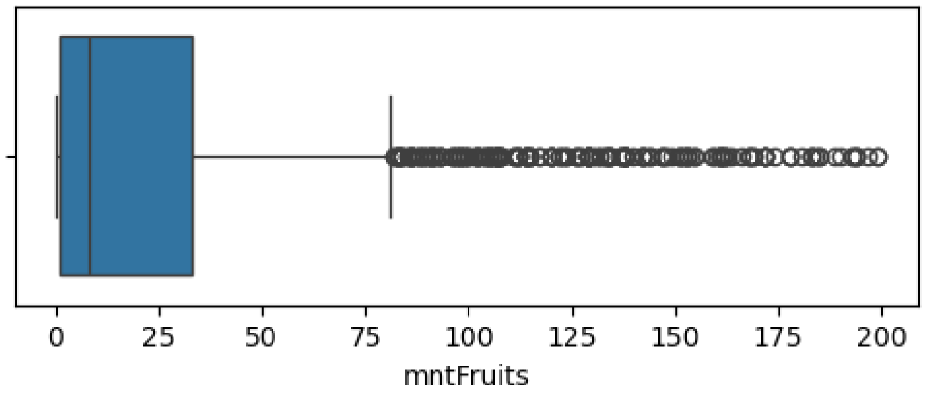
recency

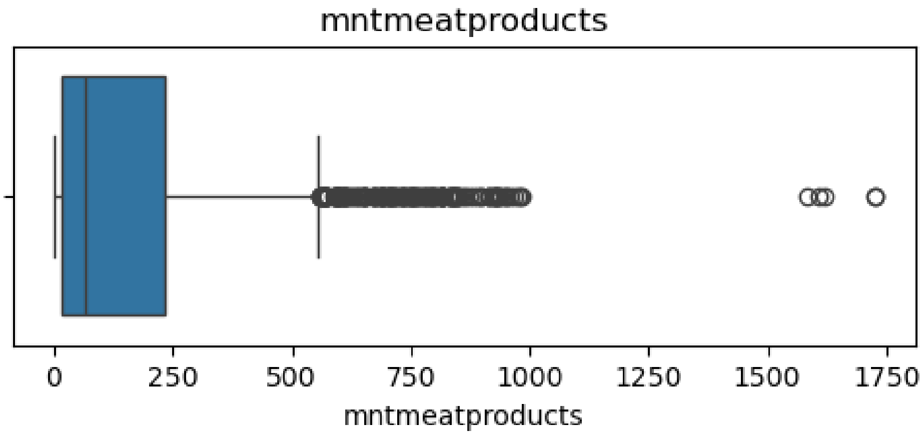


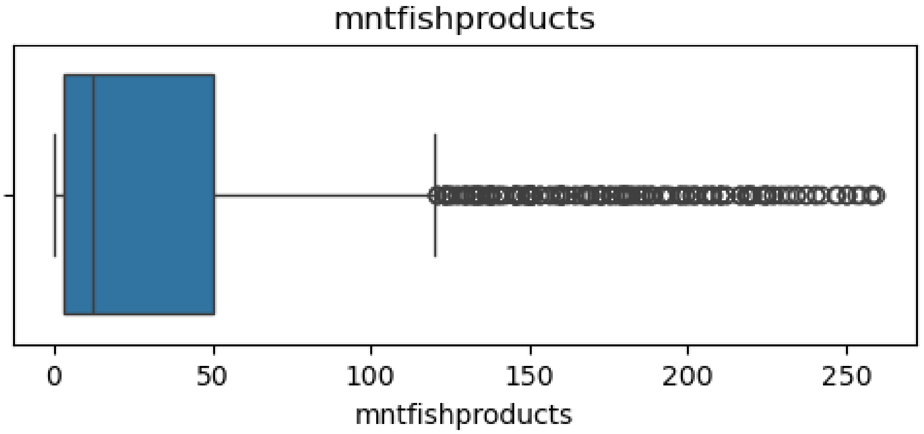
mntWines

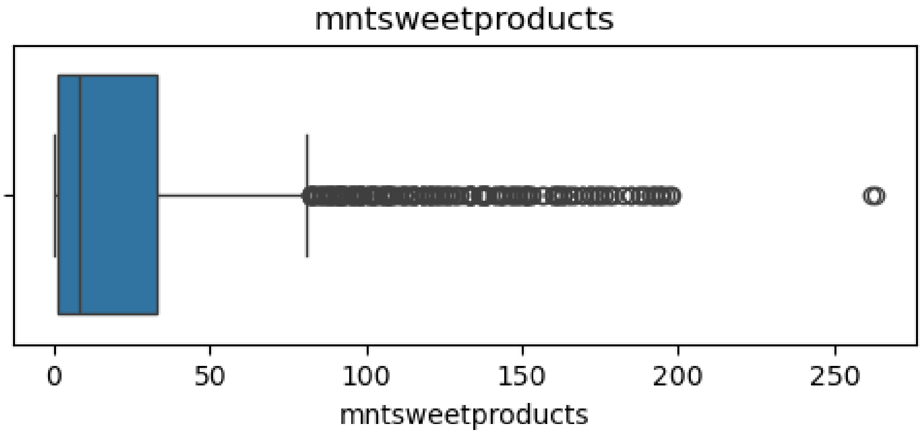


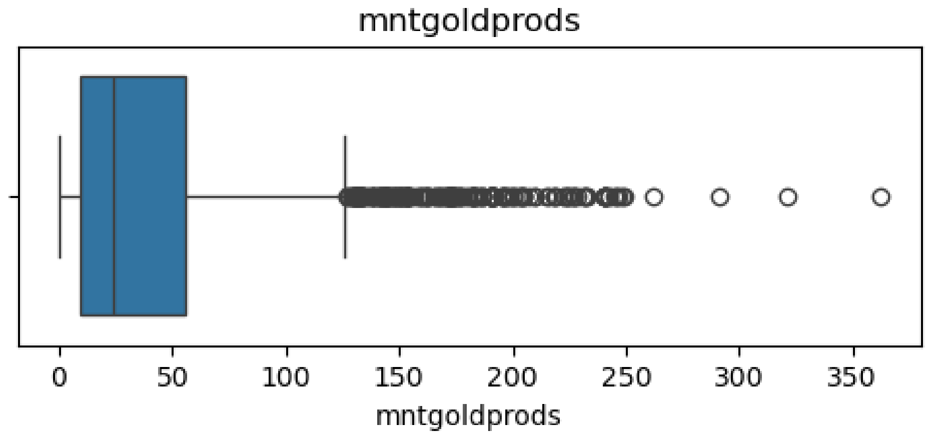
mntFruits

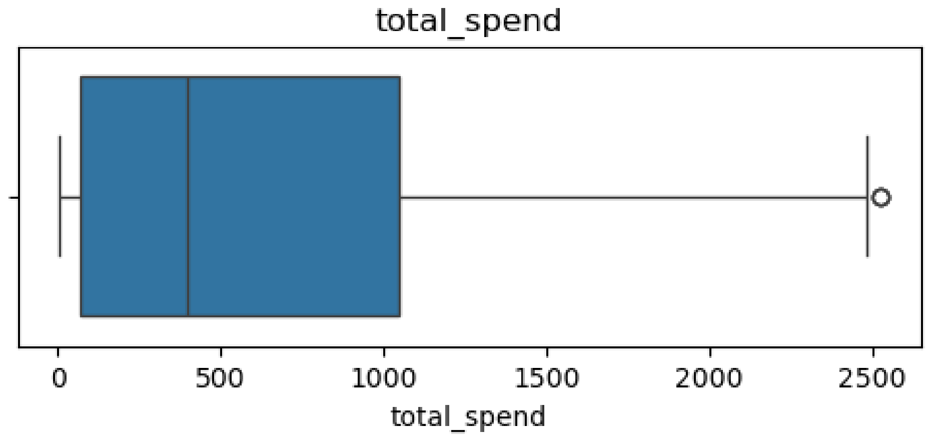


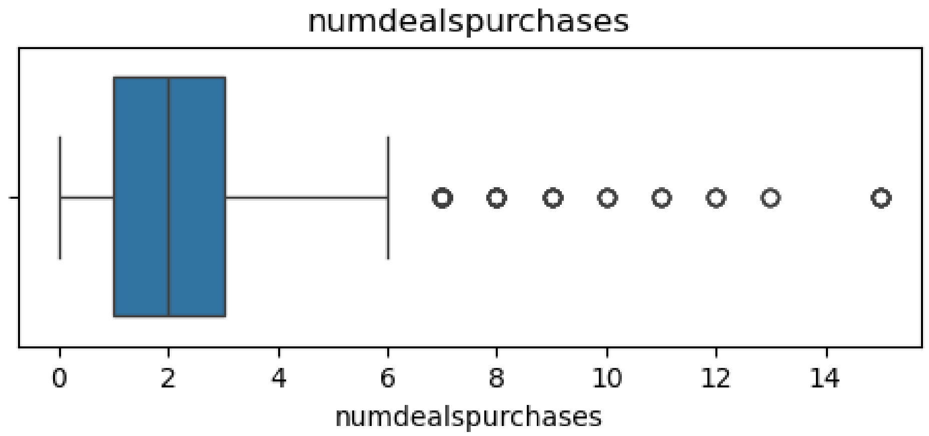

mntmeatproducts
0
250
500
750
1000
1250
1500
1750
mntmeatproducts

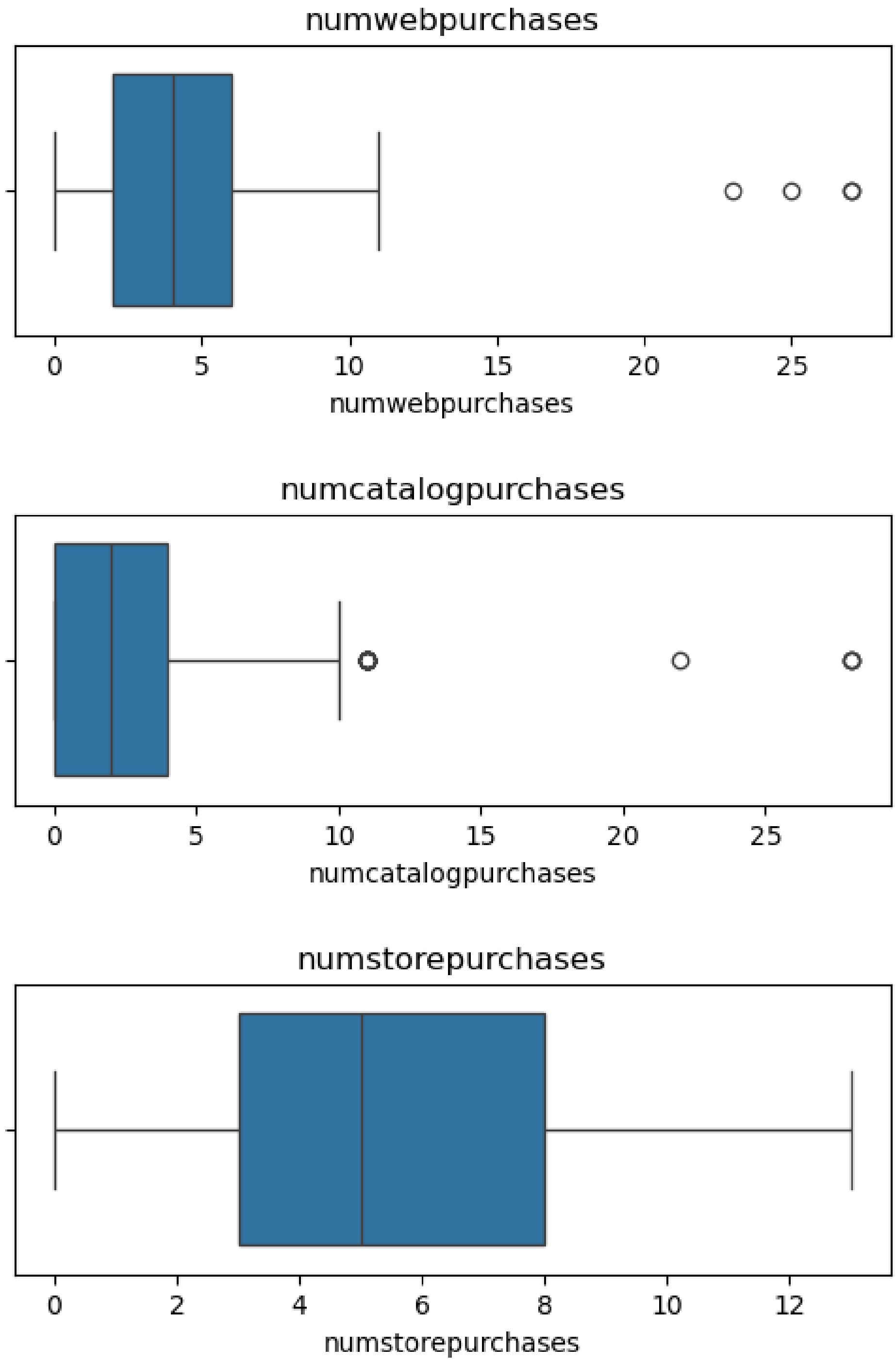

mntfishproducts
0
50
100
150
200
250
mntfishproducts


mntsweetproducts
0
50
100
150
200
250
mntsweetproducts

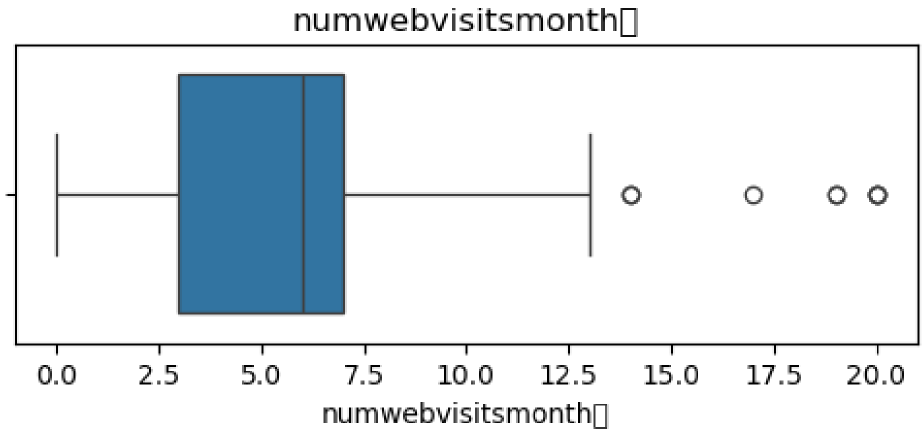

mntgoldprods
0
50
100
150
200
250
300
350
mntgoldprods


total_spend
0
500
1000
1500
2000
2500
total_spend

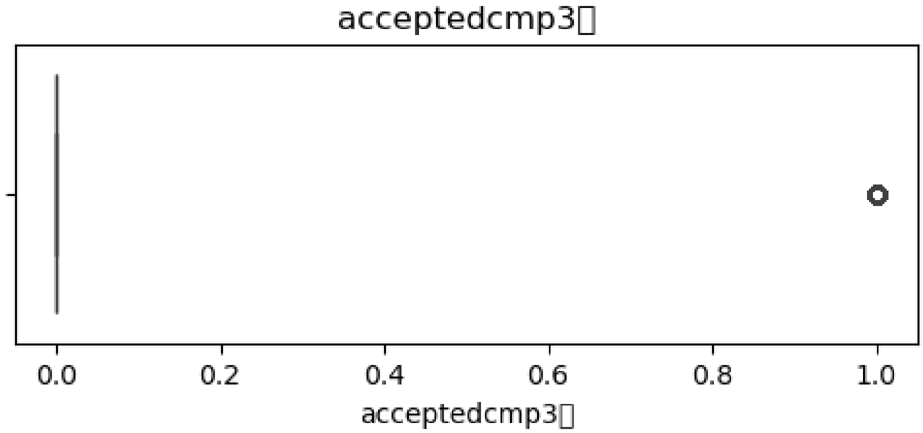

numdealspurchases
0
2
4
6
8
10
12
14
numdealspurchases



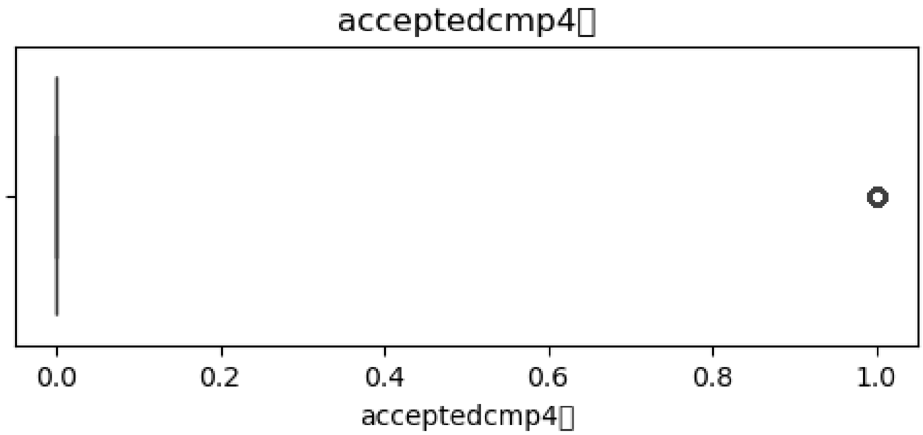
```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



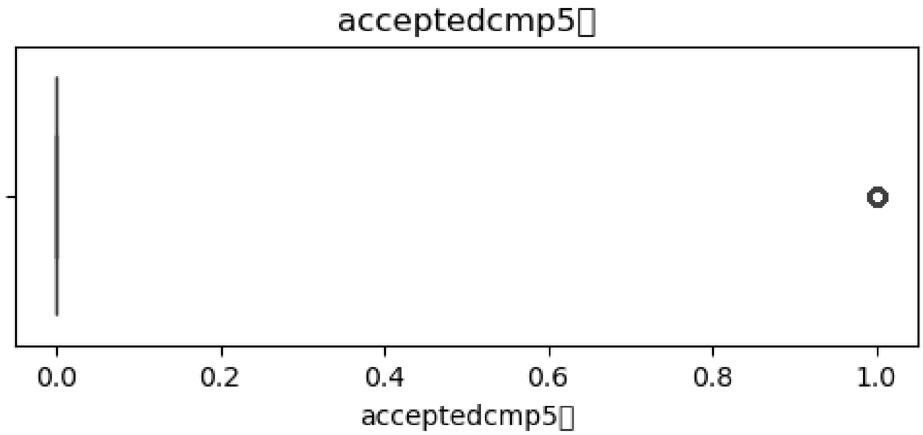
```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



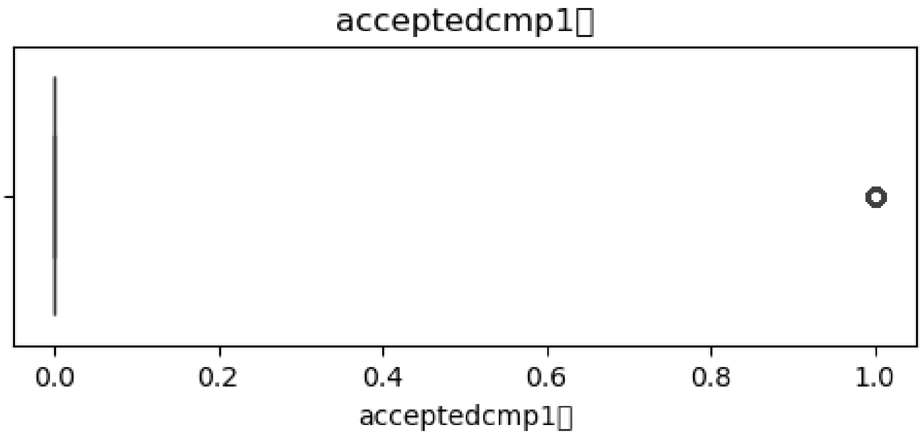
```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



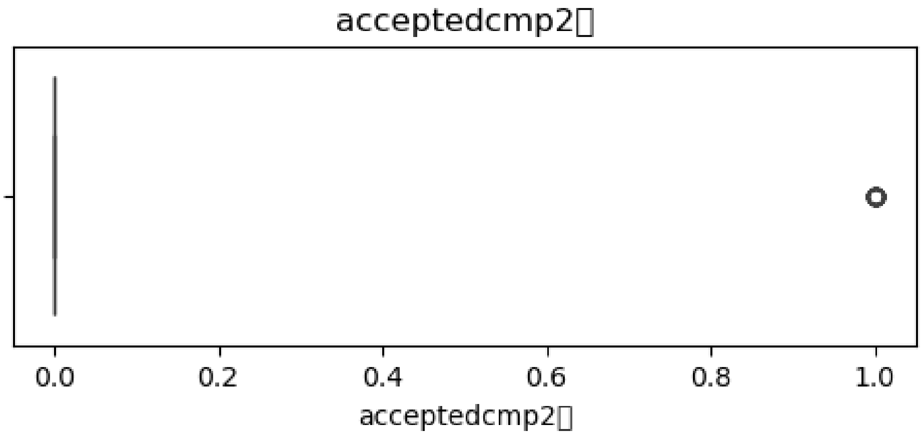
```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```

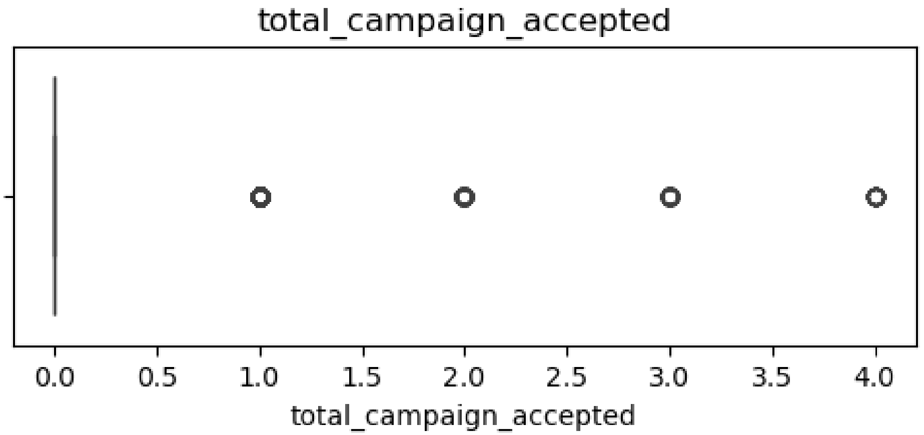


```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```

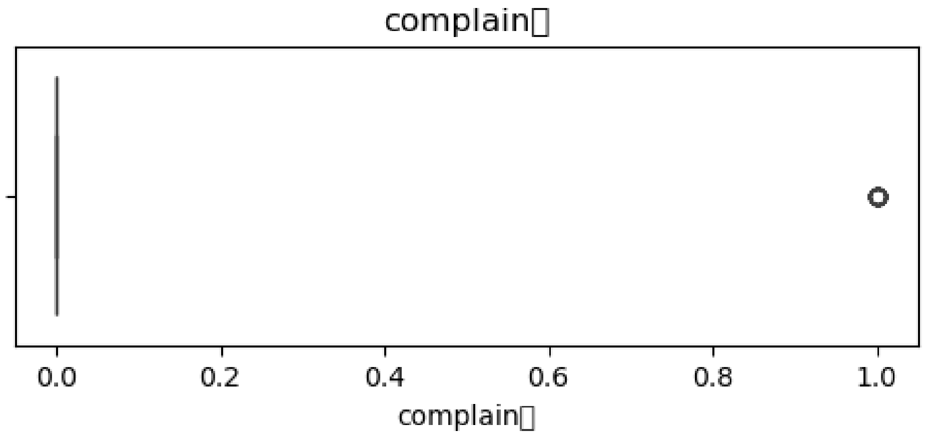


```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (	) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```

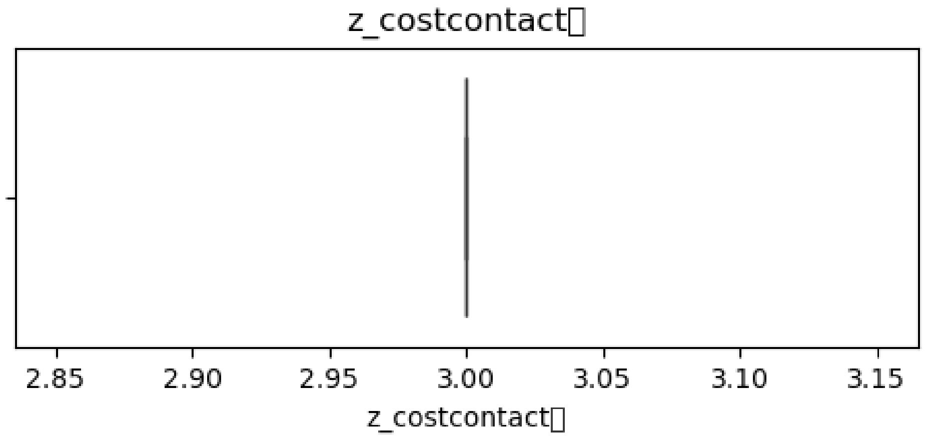




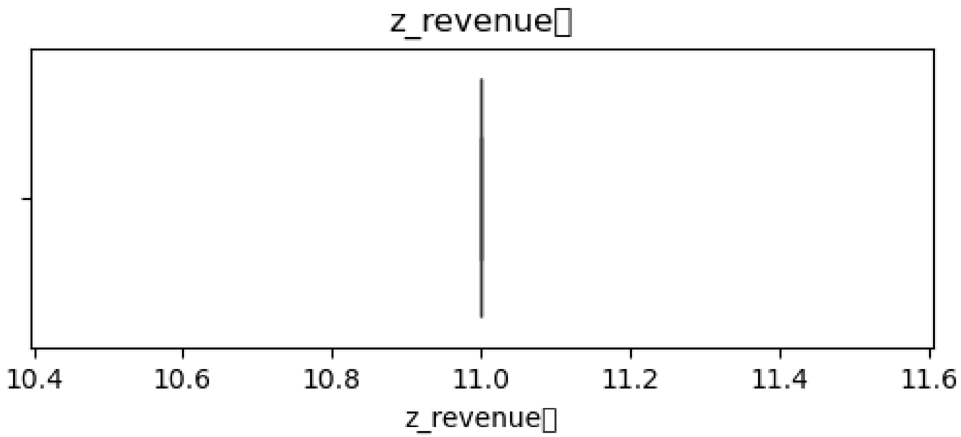
```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```

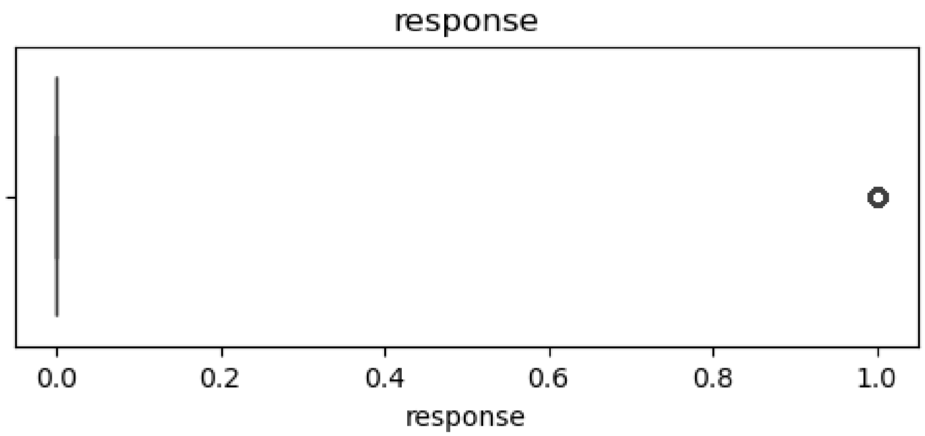


```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```

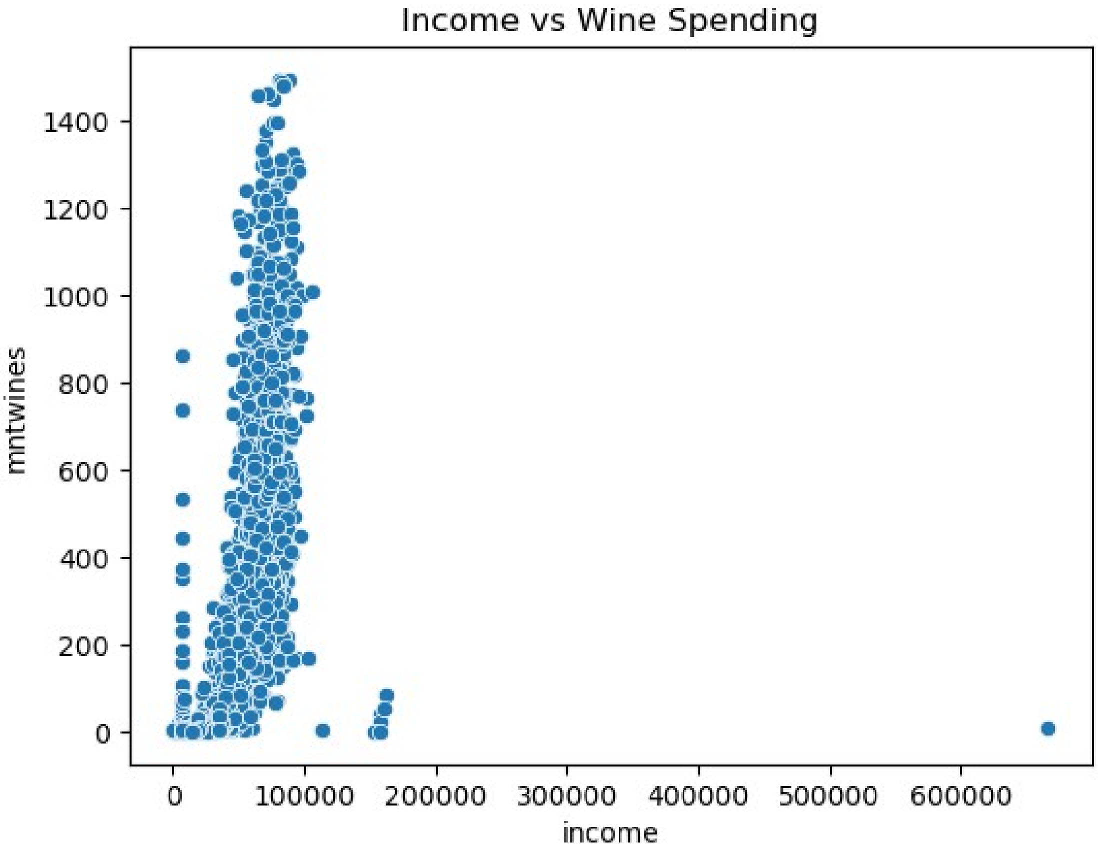


```
C:\Users\ADMIN\anaconda3\Lib\site-packages\IPython\core\
pylabtools.py:170: UserWarning: Glyph 9 (   ) missing from font(s)
DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```





```
print(df.columns.tolist())

['id', 'year_birth', 'age', 'education', 'marital_status', 'income',
'kidhome', 'teenhome', 'family_size', 'dt_customer', 'recency',
'mntWines', 'mntFruits', 'mntmeatproducts', 'mntfishproducts',
'mntsweetproducts', 'mntgoldprods', 'total_spend',
'numdealspurchases', 'numwebpurchases', 'numcatalogpurchases',
'numstorepurchases', 'numwebvisitsmonth\t', 'acceptedcmp3\t',
'acceptedcmp4\t', 'acceptedcmp5\t', 'acceptedcmp1\t', 'acceptedcmp2\
t', 'total_campaign_accepted', 'complain\t', 'z_costcontact\t',
'z_revenue\t', 'response']

df.columns = df.columns.str.strip().str.lower()

sns.scatterplot(data=df, x='income', y='mntwines')
plt.title("Income vs Wine Spending")
plt.show()
```


Income vs Wine Spending
mntwines
1400
1200
1000
800
600
400
200
0
0
100000
200000
300000
400000
500000
600000
income