In [13]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
pd.set_option('display.max_columns', None)
%matplotlib inline
sns.set_context('notebook')
sns.set_style('whitegrid')
sns.set_palette('Blues_r')

# turn off warnings for final notebook
import warnings
warnings.filterwarnings('ignore')
```

In [14]:

```
df=pd.read_csv("ifood_df.csv")
```

In [15]:

```
df
```

Out[15]:

	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldPro
0	58138.0	0	0	58	635	88	546	172	88	
1	46344.0	1	1	38	11	1	6	2	1	
2	71613.0	0	0	26	426	49	127	111	21	
3	26646.0	1	0	26	11	4	20	10	3	
4	58293.0	1	0	94	173	43	118	46	27	
...	...	...	...	...	...	...	...	...	...	
2200	61223.0	0	1	46	709	43	182	42	118	2
2201	64014.0	2	1	56	406	0	30	0	0	
2202	56981.0	0	0	91	908	48	217	32	12	
2203	69245.0	0	1	8	428	30	214	80	30	
2204	52869.0	1	1	40	84	3	61	2	1	

2205 rows × 39 columns

In [16]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2205 entries, 0 to 2204
Data columns (total 39 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Income                2205 non-null   float64
 1   Kidhome               2205 non-null   int64  
 2   Teenhome              2205 non-null   int64  
 3   Recency               2205 non-null   int64  
 4   MntWines              2205 non-null   int64  
 5   MntFruits             2205 non-null   int64  
 6   MntMeatProducts       2205 non-null   int64  
 7   MntFishProducts       2205 non-null   int64  
 8   MntSweetProducts      2205 non-null   int64  
 9   MntGoldProds          2205 non-null   int64  
 10  NumDealsPurchases     2205 non-null   int64  
 11  NumWebPurchases       2205 non-null   int64  
 12  NumCatalogPurchases   2205 non-null   int64  
 13  NumStorePurchases     2205 non-null   int64  
 14  NumWebVisitsMonth     2205 non-null   int64  
 15  AcceptedCmp3          2205 non-null   int64  
 16  AcceptedCmp4          2205 non-null   int64  
 17  AcceptedCmp5          2205 non-null   int64  
 18  AcceptedCmp1          2205 non-null   int64  
 19  AcceptedCmp2          2205 non-null   int64  
 20  Complain              2205 non-null   int64  
 21  Z_CostContact         2205 non-null   int64  
 22  Z_Revenue             2205 non-null   int64  
 23  Response              2205 non-null   int64  
 24  Age                   2205 non-null   int64  
 25  Customer_Days         2205 non-null   int64  
 26  marital_Divorced      2205 non-null   int64  
 27  marital_Married       2205 non-null   int64  
 28  marital_Single        2205 non-null   int64  
 29  marital_Together      2205 non-null   int64  
 30  marital_Widow         2205 non-null   int64  
 31  education_2n Cycle    2205 non-null   int64  
 32  education_Basic       2205 non-null   int64  
 33  education_Graduation  2205 non-null   int64  
 34  education_Master      2205 non-null   int64  
 35  education_PhD         2205 non-null   int64  
 36  MntTotal              2205 non-null   int64
 37  MntRegularProds       2205 non-null   int64
 38  AcceptedCmpOverall    2205 non-null   int64
dtypes: float64(1), int64(38)
memory usage: 672.0 KB
```

In [17]:
```
df.describe()
```

Out[17]:

	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetPro
count	2205.000000	2205.000000	2205.000000	2205.000000	2205.000000	2205.000000	2205.000000	2205.000000	2205.0
mean	51622.094785	0.442177	0.506576	49.009070	306.164626	26.403175	165.312018	37.756463	27.1
std	20713.063826	0.537132	0.544380	28.932111	337.493839	39.784484	217.784507	54.824635	41.1
min	1730.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
25%	35196.000000	0.000000	0.000000	24.000000	24.000000	2.000000	16.000000	3.000000	1.0
50%	51287.000000	0.000000	0.000000	49.000000	178.000000	8.000000	68.000000	12.000000	8.0
75%	68281.000000	1.000000	1.000000	74.000000	507.000000	33.000000	232.000000	50.000000	34.0
max	113734.000000	2.000000	2.000000	99.000000	1493.000000	199.000000	1725.000000	259.000000	262.0

In [18]:
```
df.shape
```

Out[18]:
```
(2205, 39)
```

In [19]:
```
df.transpose
```

Out[19]:

```
<bound method DataFrame.transpose of       Income  Kidhome  Teenhome  Recency  MntWines  MntFruits  \
0     58138.0        0         0       58       635         88
1     46344.0        1         1       38        11          1
2     71613.0        0         0       26       426         49
3     26646.0        1         0       26        11          4
4     58293.0        1         0       94       173         43
...       ...      ...       ...      ...       ...        ...
2200  61223.0        0         1       46       709         43
2201  64014.0        2         1       56       406          0
2202  56981.0        0         0       91       908         48
2203  69245.0        0         1        8       428         30
2204  52869.0        1         1       40        84          3

      MntMeatProducts  MntFishProducts  MntSweetProducts  MntGoldProds  \
0                 546              172                88            88
1                   6                2                 1             6
2                 127              111                21            42
3                  20               10                 3             5
4                 118               46                27            15
...               ...              ...               ...           ...
2200              182               42               118           247
2201               30                0                 0             8
2202              217               32                12            24
2203              214               80                30            61
2204               61                2                 1            21

      NumDealsPurchases  NumWebPurchases  NumCatalogPurchases  \
0                     3                8                   10
1                     2                1                    1
2                     1                8                    2
3                     2                2                    0
4                     5                5                    3
...                 ...              ...                  ...
2200                  2                9                    3
2201                  7                8                    2
2202                  1                2                    3
2203                  2                6                    5
2204                  3                3                    1

      NumStorePurchases  NumWebVisitsMonth  AcceptedCmp3  AcceptedCmp4  \
0                     4                  7             0             0
1                     2                  5             0             0
2                    10                  4             0             0
3                     4                  6             0             0
4                     6                  5             0             0
...                 ...                ...           ...           ...
2200                  4                  5             0             0
2201                  5                  7             0             0
2202                 13                  6             0             1
2203                 10                  3             0             0
2204                  4                  7             0             0

      AcceptedCmp5  AcceptedCmp1  AcceptedCmp2  Complain  Z_CostContact  \
0                0             0             0         0              3   
1                0             0             0         0              3   
2                0             0             0         0              3   
3                0             0             0         0              3   
4                0             0             0         0              3   
...            ...           ...           ...       ...            ...   
2200             0             0             0         0              3   
2201             0             1             0         0              3   
2202             0             0             0         0              3   
2203             0             0             0         0              3   
2204             0             0             0         0              3   

      Z_Revenue  Response  Age  Customer_Days  marital_Divorced  \
0            11         1   63           2822                 0   
1            11         0   66           2272                 0   
2            11         0   55           2471                 0   
3            11         0   36           2298                 0   
4            11         0   39           2320                 0   
...         ...       ...  ...            ...               ...   
2200         11         0   53           2540                 0   
2201         11         0   74           2178                 0   
2202         11         0   39           2314                 1   
2203         11         0   64           2315                 0   
2204         11         1   66           2781                 0   

      marital_Married  marital_Single  marital_Together  marital_Widow  \
0                   0               1                 0              0   
1                   0               1                 0              0   
```

```
2                   0               0                 1              0
3                   0               0                 1              0
4                   1               0                 0              0
...               ...             ...               ...            ...
2200                1               0                 0              0
2201                0               0                 1              0
2202                0               0                 0              0
2203                0               0                 1              0
2204                1               0                 0              0

      education_2n Cycle  education_Basic  education_Graduation  \
0                      0                0                     1   
1                      0                0                     1   
2                      0                0                     1   
3                      0                0                     1   
4                      0                0                     0   
...                  ...              ...                   ...   
2200                   0                0                     1   
2201                   0                0                     0   
2202                   0                0                     1   
2203                   0                0                     0   
2204                   0                0                     0   

      education_Master  education_PhD  MntTotal  MntRegularProds  \
0                    0              0      1529             1441   
1                    0              0        21               15   
2                    0              0       734              692   
3                    0              0        48               43   
4                    0              1       407              392   
...                ...            ...       ...              ...   
2200                 0              0      1094              847   
2201                 0              1       436              428   
2202                 0              0      1217             1193   
2203                 1              0       782              721   
2204                 0              1       151              130   

      AcceptedCmpOverall  
0                      0  
1                      0  
2                      0  
3                      0  
```

```
4                      0
...                  ...
2200                   0
2201                   1
2202                   1
2203                   0
2204                   0

[2205 rows x 39 columns]>
```

In [10]:
```
df.head()
```

Out[10]:

	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds
0	58138.0	0	0	58	635	88	546	172	88	88
1	46344.0	1	1	38	11	1	6	2	1	6
2	71613.0	0	0	26	426	49	127	111	21	42
3	26646.0	1	0	26	11	4	20	10	3	5
4	58293.0	1	0	94	173	43	118	46	27	15

5 rows × 39 columns

In [12]:
```
df.shape[0]
```

Out[12]:
```
2205
```

In [20]:
```
df.isnull().sum().sort_values(ascending=True)
```

```
Out[20]: Income                  0
         Z_CostContact           0
         Z_Revenue               0
         Response                0
         Age                     0
         Customer_Days           0
         marital_Divorced        0
         marital_Married         0
         Complain                0
         marital_Single          0
         marital_Widow           0
         education_2n Cycle      0
         education_Basic         0
         education_Graduation    0
         education_Master        0
         education_PhD           0
         MntTotal                0
         marital_Together        0
         MntRegularProds         0
         AcceptedCmp2            0
         AcceptedCmp5            0
         Kidhome                 0
         Teenhome                0
         Recency                 0
         MntWines                0
         MntFruits               0
         MntMeatProducts         0
         MntFishProducts         0
         AcceptedCmp1            0
         MntSweetProducts        0
         NumDealsPurchases       0
         NumWebPurchases         0
         NumCatalogPurchases     0
         NumStorePurchases       0
         NumWebVisitsMonth       0
         AcceptedCmp3            0
         AcceptedCmp4            0
         MntGoldProds            0
         AcceptedCmpOverall      0
         dtype: int64
```

```
In [21]: df.isnull().sum().sort_values(ascending=False)
```

```
Out[21]: Income                   0
         marital_Together         0
         Z_Revenue                0
         Response                 0
         Age                      0
         Customer_Days            0
         marital_Divorced         0
         marital_Married          0
         marital_Single           0
         marital_Widow            0
         Complain                 0
         education_2n Cycle       0
         education_Basic          0
         education_Graduation     0
         education_Master         0
         education_PhD            0
         MntTotal                 0
         MntRegularProds          0
         Z_CostContact            0
         AcceptedCmp2             0
         Kidhome                  0
         MntGoldProds             0
         Teenhome                 0
         Recency                  0
         MntWines                 0
         MntFruits                0
         MntMeatProducts          0
         MntFishProducts          0
         MntSweetProducts         0
         NumDealsPurchases        0
         AcceptedCmp1             0
         NumWebPurchases          0
         NumCatalogPurchases      0
         NumStorePurchases        0
         NumWebVisitsMonth        0
         AcceptedCmp3             0
         AcceptedCmp4             0
         AcceptedCmp5             0
         AcceptedCmpOverall       0
         dtype: int64
```

In [24]:

```
df = df.dropna()
```

In [72]:

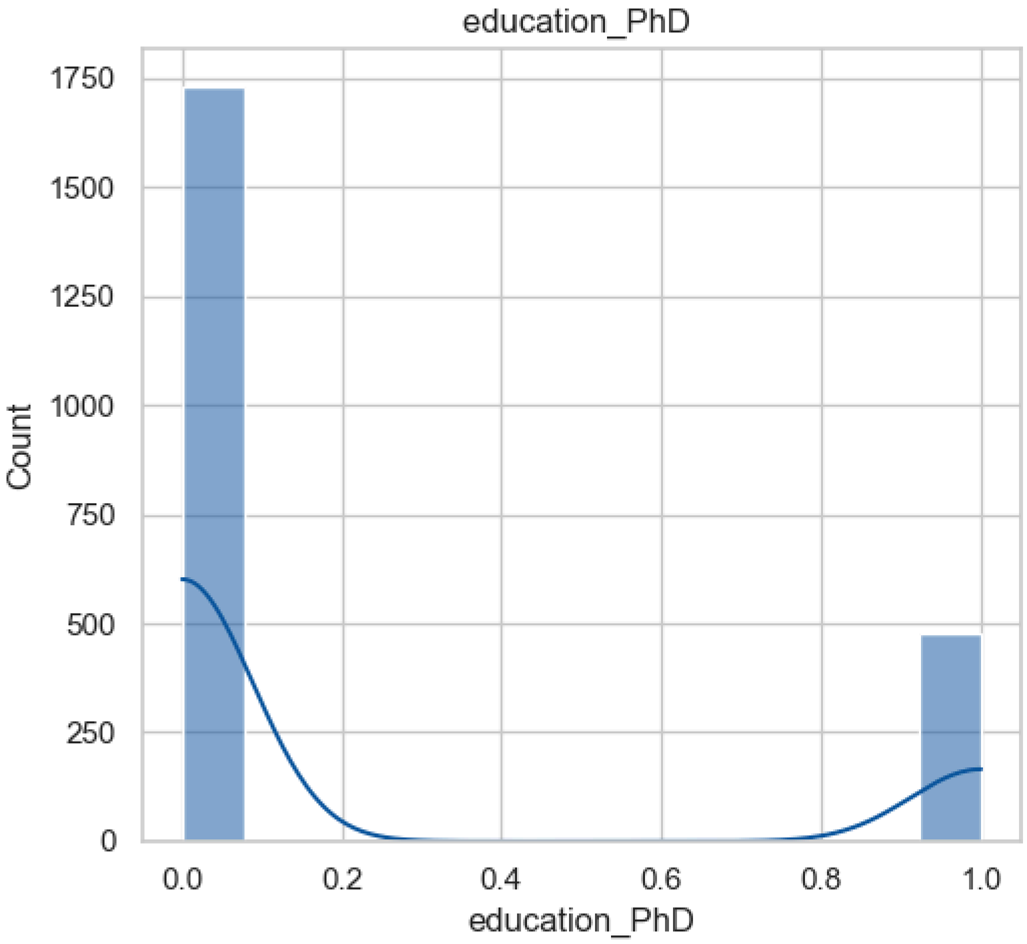
```
print(df_cleaned.describe())
```

```
           Kidhome     Teenhome      Recency     MntWines    MntFruits  \
count  2205.000000  2205.000000  2205.000000  2205.000000  2205.000000
mean      0.442177     0.506576    49.009070   306.164626    26.403175
std       0.537132     0.544380    28.932111   337.493839    39.784484
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.000000     0.000000    24.000000    24.000000     2.000000
50%       0.000000     0.000000    49.000000   178.000000     8.000000
75%       1.000000     1.000000    74.000000   507.000000    33.000000
max       2.000000     2.000000    99.000000  1493.000000   199.000000

       MntMeatProducts  MntFishProducts  MntSweetProducts  MntGoldProds  \
count      2205.000000      2205.000000       2205.000000   2205.000000
mean        165.312018        37.756463         27.128345     44.057143
std         217.784507        54.824635         41.130468     51.736211
min           0.000000         0.000000          0.000000      0.000000
25%          16.000000         3.000000          1.000000      9.000000
50%          68.000000        12.000000          8.000000     25.000000
75%         232.000000        50.000000         34.000000     56.000000
max        1725.000000       259.000000        262.000000    321.000000

       NumDealsPurchases  NumWebPurchases  NumCatalogPurchases  \
count        2205.000000      2205.000000          2205.000000
mean            2.318367         4.100680             2.645351
std             1.886107         2.737424             2.798647
min             0.000000         0.000000             0.000000
25%             1.000000         2.000000             0.000000
50%             2.000000         4.000000             2.000000
75%             3.000000         6.000000             4.000000
max            15.000000        27.000000            28.000000

       NumStorePurchases  NumWebVisitsMonth  AcceptedCmp3  AcceptedCmp4  \
count        2205.000000        2205.000000   2205.000000   2205.000000
mean            5.823583           5.336961      0.073923      0.074376
std             3.241796           2.413535      0.261705      0.262442
min             0.000000           0.000000      0.000000      0.000000
25%             3.000000           3.000000      0.000000      0.000000
50%             5.000000           6.000000      0.000000      0.000000
75%             8.000000           7.000000      0.000000      0.000000
max            13.000000          20.000000      1.000000      1.000000

       AcceptedCmp5  AcceptedCmp1  AcceptedCmp2     Complain  Z_CostContact  \
count   2205.000000   2205.000000   2205.000000  2205.000000          2205.0
mean       0.073016      0.064399      0.013605     0.009070             3.0
std        0.260222      0.245518      0.115872     0.094827             0.0
min        0.000000      0.000000      0.000000     0.000000             3.0
25%        0.000000      0.000000      0.000000     0.000000             3.0
50%        0.000000      0.000000      0.000000     0.000000             3.0
75%        0.000000      0.000000      0.000000     0.000000             3.0
max        1.000000      1.000000      1.000000     1.000000             3.0

       Z_Revenue    Response          Age  Customer_Days  marital_Divorced  \
count     2205.0  2205.00000  2205.000000    2205.000000       2205.000000   
mean        11.0     0.15102    51.095692    2512.718367          0.104308   
std          0.0     0.35815    11.705801     202.563647          0.305730   
min         11.0     0.00000    24.000000    2159.000000          0.000000   
25%         11.0     0.00000    43.000000    2339.000000          0.000000   
50%         11.0     0.00000    50.000000    2515.000000          0.000000   
75%         11.0     0.00000    61.000000    2688.000000          0.000000   
max         11.0     1.00000    80.000000    2858.000000          1.000000   

       marital_Married  marital_Single  marital_Together  marital_Widow  \
count      2205.000000     2205.000000       2205.000000    2205.000000   
mean          0.387302        0.216327          0.257596       0.034467   
std           0.487244        0.411833          0.437410       0.182467   
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%           1.000000        0.000000          1.000000       0.000000   
max           1.000000        1.000000          1.000000       1.000000   

       education_2n Cycle  education_Basic  education_Graduation  \
count         2205.000000      2205.000000           2205.000000   
mean             0.089796         0.024490              0.504762   
std              0.285954         0.154599              0.500091   
min              0.000000         0.000000              0.000000   
25%              0.000000         0.000000              0.000000   
50%              0.000000         0.000000              1.000000   
75%              0.000000         0.000000              1.000000   
max              1.000000         1.000000              1.000000   

       education_Master  education_PhD     MntTotal  MntRegularProds  \
count       2205.000000    2205.000000  2205.000000      2205.000000   
```

```
mean           0.165079       0.215873   562.764626       518.707483
std            0.371336       0.411520   575.936911       553.847248
min            0.000000       0.000000     4.000000      -283.000000
25%            0.000000       0.000000    56.000000        42.000000
50%            0.000000       0.000000   343.000000       288.000000
75%            0.000000       0.000000   964.000000       884.000000
max            1.000000       1.000000  2491.000000      2458.000000

       AcceptedCmpOverall      Cluster
count          2205.00000  2205.000000
mean              0.29932     1.494331
std               0.68044     1.073349
min               0.00000     0.000000
25%               0.00000     1.000000
50%               0.00000     1.000000
75%               0.00000     2.000000
max               4.00000     3.000000
```

In [77]:

```
plt.figure(figsize=(10, 5))

plt.subplot(1, 2, 1)
sns.histplot(df_cleaned['education_PhD'], kde=True)
plt.title('education_PhD ')

plt.tight_layout()
plt.show()
```



```
In [27]: df_cleaned = df.drop(columns=['Income'], axis=1)

         # View cleaned dataset
         print(df_cleaned.head())
```

```
   Kidhome  Teenhome  Recency  MntWines  MntFruits  MntMeatProducts  \
0        0         0       58       635         88              546
1        1         1       38        11          1                6
2        0         0       26       426         49              127
3        1         0       26        11          4               20
4        1         0       94       173         43              118

   MntFishProducts  MntSweetProducts  MntGoldProds  NumDealsPurchases  \
0              172                88            88                  3
1                2                 1             6                  2
2              111                21            42                  1
3               10                 3             5                  2
4               46                27            15                  5

   NumWebPurchases  NumCatalogPurchases  NumStorePurchases  NumWebVisitsMonth  \
0                8                   10                  4                  7
1                1                    1                  2                  5
2                8                    2                 10                  4
3                2                    0                  4                  6
4                5                    3                  6                  5

   AcceptedCmp3  AcceptedCmp4  AcceptedCmp5  AcceptedCmp1  AcceptedCmp2  \
0             0             0             0             0             0
1             0             0             0             0             0
2             0             0             0             0             0
3             0             0             0             0             0
4             0             0             0             0             0

   Complain  Z_CostContact  Z_Revenue  Response  Age  Customer_Days  \
0         0              3         11         1   63           2822
1         0              3         11         0   66           2272
2         0              3         11         0   55           2471
3         0              3         11         0   36           2298
4         0              3         11         0   39           2320

   marital_Divorced  marital_Married  marital_Single  marital_Together  \
0                 0                0               1                 0
1                 0                0               1                 0
2                 0                0               0                 1
3                 0                0               0                 1
4                 0                1               0                 0
```

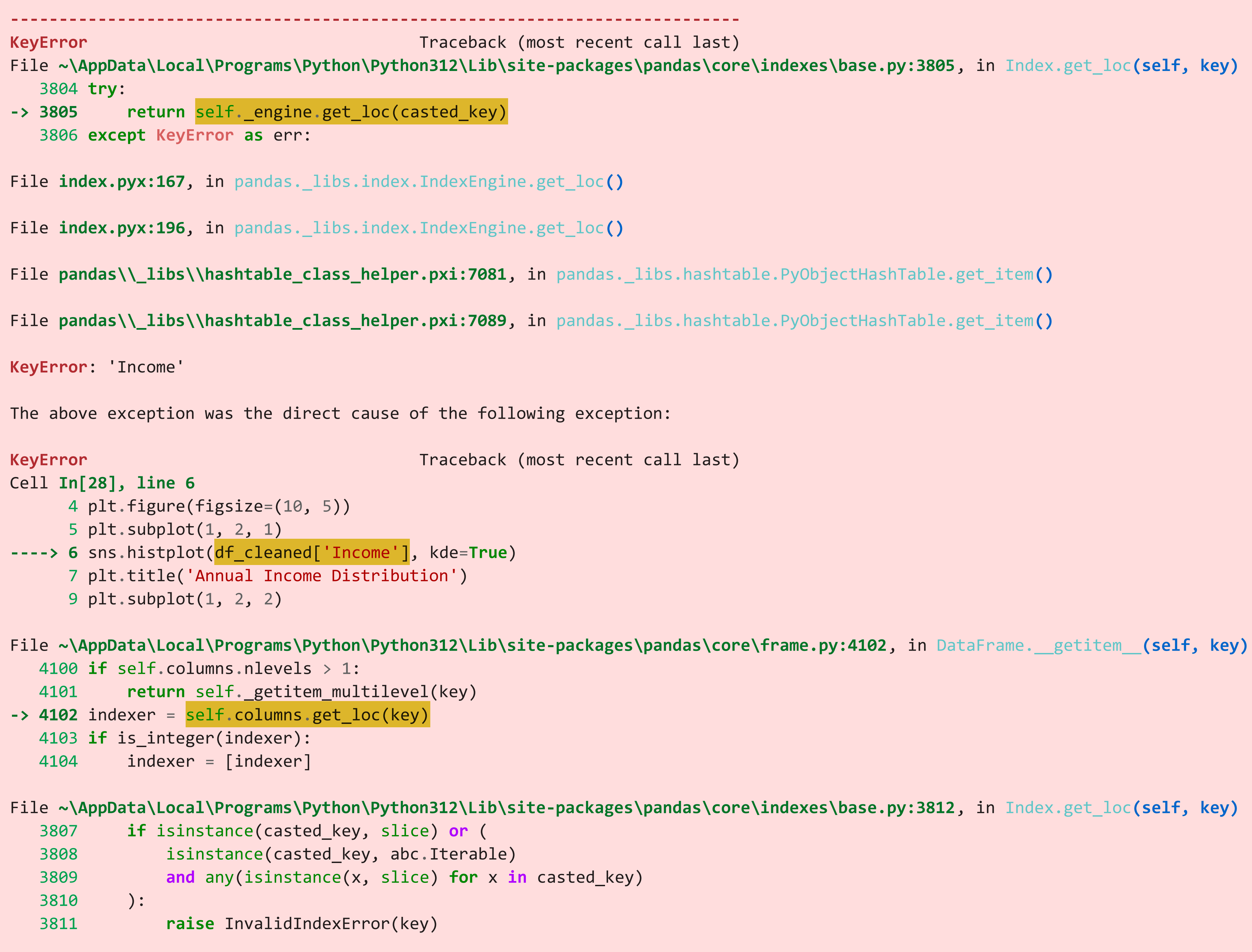
```
   marital_Widow  education_2n Cycle  education_Basic  education_Graduation  \
0              0                   0                0                     1   
1              0                   0                0                     1   
2              0                   0                0                     1   
3              0                   0                0                     1   
4              0                   0                0                     0   

   education_Master  education_PhD  MntTotal  MntRegularProds  \
0                 0              0      1529             1441   
1                 0              0        21               15   
2                 0              0       734              692   
3                 0              0        48               43   
4                 0              1       407              392   

   AcceptedCmpOverall  
0                   0  
1                   0  
2                   0  
3                   0  
4                   0  
```

In [28]:

```
print(df_cleaned.describe())

# Visualizing distribution of annual income and spending score
plt.figure(figsize=(10, 5))
plt.subplot(1, 2, 1)
sns.histplot(df_cleaned['Income'], kde=True)
plt.title('Annual Income Distribution')

plt.subplot(1, 2, 2)
sns.histplot(df_cleaned['NumStorePurchases'], kde=True)
plt.title('Spending Score Distribution')

plt.tight_layout()
plt.show()
```

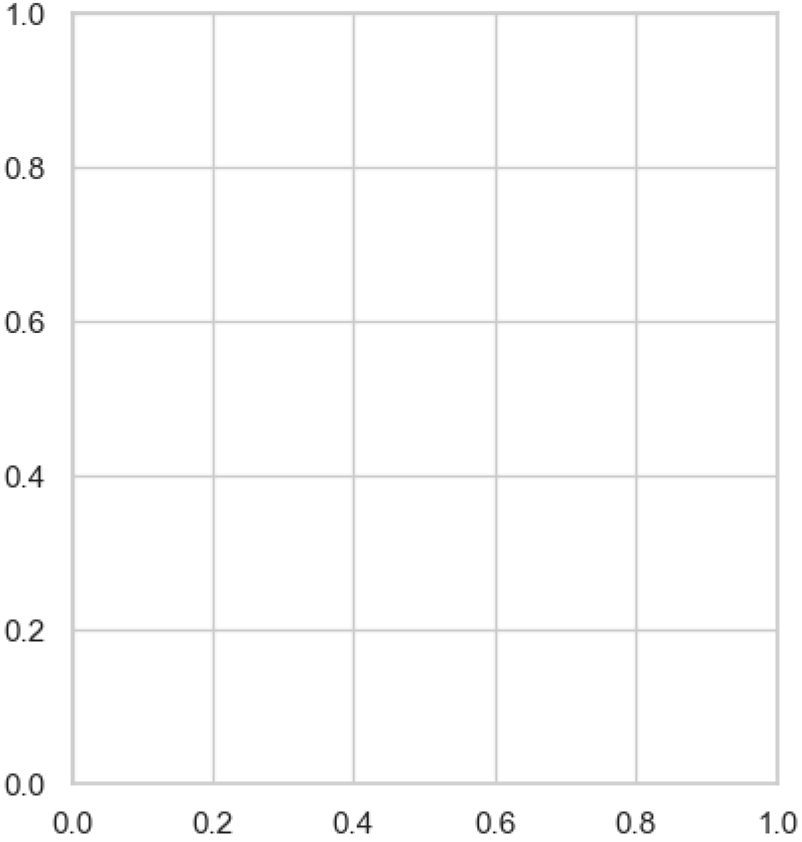
```
           Kidhome     Teenhome      Recency     MntWines    MntFruits  \
count  2205.000000  2205.000000  2205.000000  2205.000000  2205.000000
mean      0.442177     0.506576    49.009070   306.164626    26.403175
std       0.537132     0.544380    28.932111   337.493839    39.784484
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.000000     0.000000    24.000000    24.000000     2.000000
50%       0.000000     0.000000    49.000000   178.000000     8.000000
75%       1.000000     1.000000    74.000000   507.000000    33.000000
max       2.000000     2.000000    99.000000  1493.000000   199.000000

       MntMeatProducts  MntFishProducts  MntSweetProducts  MntGoldProds  \
count      2205.000000      2205.000000       2205.000000   2205.000000
mean        165.312018        37.756463         27.128345     44.057143
std         217.784507        54.824635         41.130468     51.736211
min           0.000000         0.000000          0.000000      0.000000
25%          16.000000         3.000000          1.000000      9.000000
50%          68.000000        12.000000          8.000000     25.000000
75%         232.000000        50.000000         34.000000     56.000000
max        1725.000000       259.000000        262.000000    321.000000

       NumDealsPurchases  NumWebPurchases  NumCatalogPurchases  \
count        2205.000000      2205.000000          2205.000000
mean            2.318367         4.100680             2.645351
std             1.886107         2.737424             2.798647
min             0.000000         0.000000             0.000000
25%             1.000000         2.000000             0.000000
50%             2.000000         4.000000             2.000000
75%             3.000000         6.000000             4.000000
max            15.000000        27.000000            28.000000

       NumStorePurchases  NumWebVisitsMonth  AcceptedCmp3  AcceptedCmp4  \
count        2205.000000        2205.000000   2205.000000   2205.000000
mean            5.823583           5.336961      0.073923      0.074376
std             3.241796           2.413535      0.261705      0.262442
min             0.000000           0.000000      0.000000      0.000000
25%             3.000000           3.000000      0.000000      0.000000
50%             5.000000           6.000000      0.000000      0.000000
75%             8.000000           7.000000      0.000000      0.000000
max            13.000000          20.000000      1.000000      1.000000

       AcceptedCmp5  AcceptedCmp1  AcceptedCmp2     Complain  Z_CostContact  \
count   2205.000000   2205.000000   2205.000000  2205.000000         2205.0
mean       0.073016      0.064399      0.013605     0.009070            3.0
std        0.260222      0.245518      0.115872     0.094827            0.0
min        0.000000      0.000000      0.000000     0.000000            3.0
25%        0.000000      0.000000      0.000000     0.000000            3.0
50%        0.000000      0.000000      0.000000     0.000000            3.0
75%        0.000000      0.000000      0.000000     0.000000            3.0
max        1.000000      1.000000      1.000000     1.000000            3.0

       Z_Revenue    Response          Age  Customer_Days  marital_Divorced  \
count     2205.0  2205.00000  2205.000000    2205.000000       2205.000000   
mean        11.0     0.15102    51.095692    2512.718367          0.104308   
std          0.0     0.35815    11.705801     202.563647          0.305730   
min         11.0     0.00000    24.000000    2159.000000          0.000000   
25%         11.0     0.00000    43.000000    2339.000000          0.000000   
50%         11.0     0.00000    50.000000    2515.000000          0.000000   
75%         11.0     0.00000    61.000000    2688.000000          0.000000   
max         11.0     1.00000    80.000000    2858.000000          1.000000   

       marital_Married  marital_Single  marital_Together  marital_Widow  \
count      2205.000000     2205.000000       2205.000000    2205.000000   
mean          0.387302        0.216327          0.257596       0.034467   
std           0.487244        0.411833          0.437410       0.182467   
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%           1.000000        0.000000          1.000000       0.000000   
max           1.000000        1.000000          1.000000       1.000000   

       education_2n Cycle  education_Basic  education_Graduation  \
count         2205.000000      2205.000000           2205.000000   
mean             0.089796         0.024490              0.504762   
std              0.285954         0.154599              0.500091   
min              0.000000         0.000000              0.000000   
25%              0.000000         0.000000              0.000000   
50%              0.000000         0.000000              1.000000   
75%              0.000000         0.000000              1.000000   
max              1.000000         1.000000              1.000000   

       education_Master  education_PhD     MntTotal  MntRegularProds  \
count       2205.000000    2205.000000  2205.000000      2205.000000   
```

```
mean           0.165079       0.215873   562.764626       518.707483
std            0.371336       0.411520   575.936911       553.847248
min            0.000000       0.000000     4.000000      -283.000000
25%            0.000000       0.000000    56.000000        42.000000
50%            0.000000       0.000000   343.000000       288.000000
75%            0.000000       0.000000   964.000000       884.000000
max            1.000000       1.000000  2491.000000      2458.000000

       AcceptedCmpOverall
count          2205.00000
mean              0.29932
std               0.68044
min               0.00000
25%               0.00000
50%               0.00000
75%               0.00000
max               4.00000
```



```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
File ~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\indexes\base.py:3805, in Index.get_loc(self, key)
   3804 try:
-> 3805     return self._engine.get_loc(casted_key)
   3806 except KeyError as err:

File index.pyx:167, in pandas._libs.index.IndexEngine.get_loc()

File index.pyx:196, in pandas._libs.index.IndexEngine.get_loc()

File pandas\\_libs\\hashtable_class_helper.pxi:7081, in pandas._libs.hashtable.PyObjectHashTable.get_item()

File pandas\\_libs\\hashtable_class_helper.pxi:7089, in pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'Income'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
Cell In[28], line 6
      4 plt.figure(figsize=(10, 5))
      5 plt.subplot(1, 2, 1)
----> 6 sns.histplot(df_cleaned['Income'], kde=True)
      7 plt.title('Annual Income Distribution')
      9 plt.subplot(1, 2, 2)

File ~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\frame.py:4102, in DataFrame.__getitem__(self, key)
   4100 if self.columns.nlevels > 1:
   4101     return self._getitem_multilevel(key)
-> 4102 indexer = self.columns.get_loc(key)
   4103 if is_integer(indexer):
   4104     indexer = [indexer]

File ~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\indexes\base.py:3812, in Index.get_loc(self, key)
   3807     if isinstance(casted_key, slice) or (
   3808         isinstance(casted_key, abc.Iterable)
   3809         and any(isinstance(x, slice) for x in casted_key)
   3810     ):
   3811         raise InvalidIndexError(key)
```

```
-> 3812     raise KeyError(key) from err
   3813 except TypeError:
   3814     # If we have a listlike key, _check_indexing_error will raise
   3815     #  InvalidIndexError. Otherwise we fall through and re-raise
   3816     #  the TypeError.
   3817     self._check_indexing_error(key)

KeyError: 'Income'
```



In [30]:

```
from sklearn.preprocessing import StandardScaler
```

In [32]:

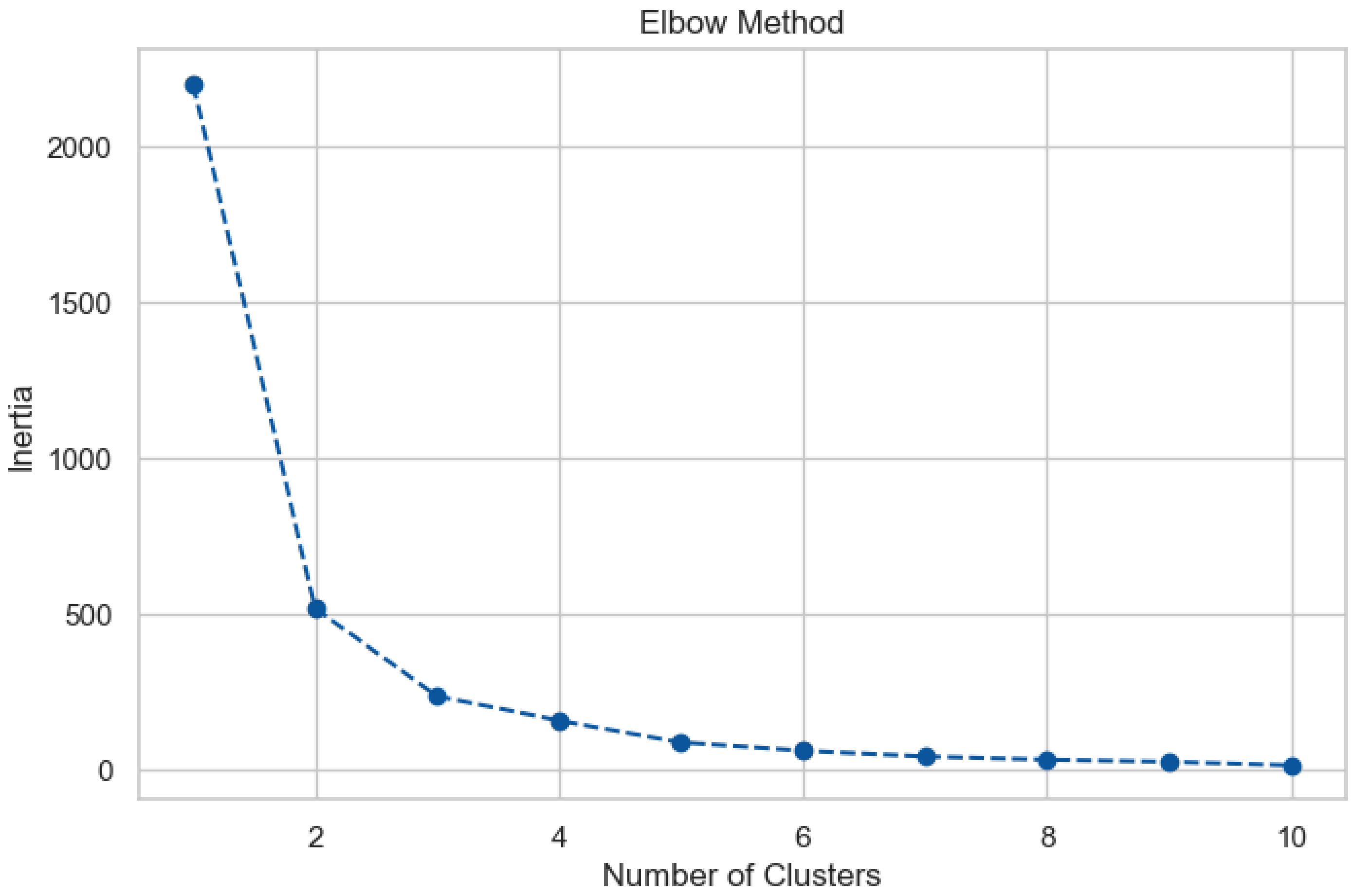
```
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df_cleaned[['NumStorePurchases']])
```

In [35]:

```
from sklearn.cluster import KMeans
```

In [36]:

```
inertia = []
for n in range(1, 11):
    kmeans = KMeans(n_clusters=n, init='k-means++', random_state=42)
    kmeans.fit(df_scaled)
    inertia.append(kmeans.inertia_)

plt.figure(figsize=(8, 5))
plt.plot(range(1, 11), inertia, marker='o', linestyle='--')
plt.title('Elbow Method')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia')
plt.show()

kmeans = KMeans(n_clusters=4, init='k-means++', random_state=42)
df_cleaned['Cluster'] = kmeans.fit_predict(df_scaled)
```

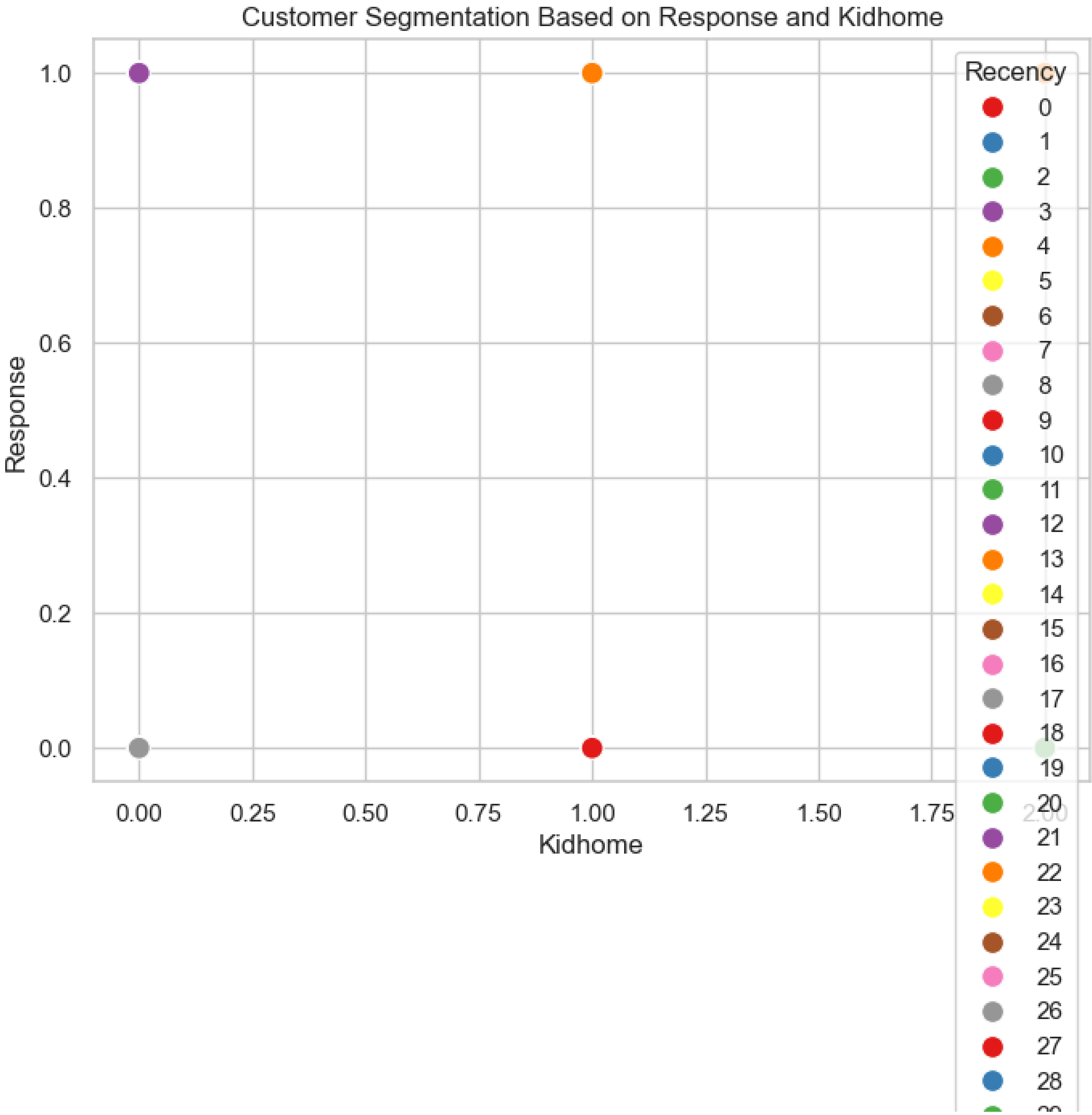


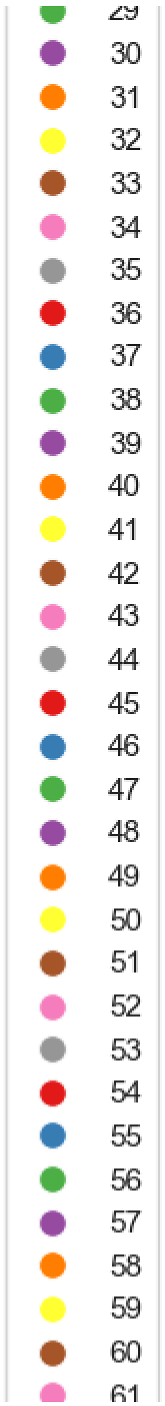
```
In [40]: print(df_cleaned.columns)
```

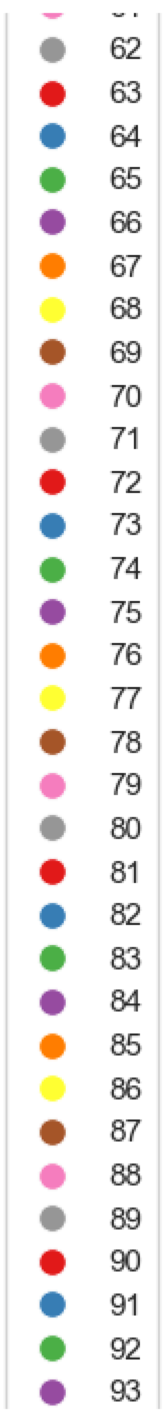
```
Index(['Kidhome', 'Teenhome', 'Recency', 'MntWines', 'MntFruits',
       'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts',
       'MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases',
       'NumCatalogPurchases', 'NumStorePurchases', 'NumWebVisitsMonth',
       'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5', 'AcceptedCmp1',
       'AcceptedCmp2', 'Complain', 'Z_CostContact', 'Z_Revenue', 'Response',
       'Age', 'Customer_Days', 'marital_Divorced', 'marital_Married',
       'marital_Single', 'marital_Together', 'marital_Widow',
       'education_2n Cycle', 'education_Basic', 'education_Graduation',
       'education_Master', 'education_PhD', 'MntTotal', 'MntRegularProds',
       'AcceptedCmpOverall', 'Cluster'],
      dtype='object')
```

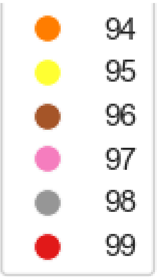
In [43]:

```
plt.figure(figsize=(8, 6))

# Use correct column names based on your dataset
sns.scatterplot(x='Kidhome', y='Response', hue='Recency', palette='Set1', data=df_cleaned, s=100)

plt.title('Customer Segmentation Based on Response and Kidhome')
plt.show()
```


Customer Segmentation Based on Response and Kidhome
Response
1.0
0.8
0.6
0.4
0.2
0.0
Kidhome
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Recency
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

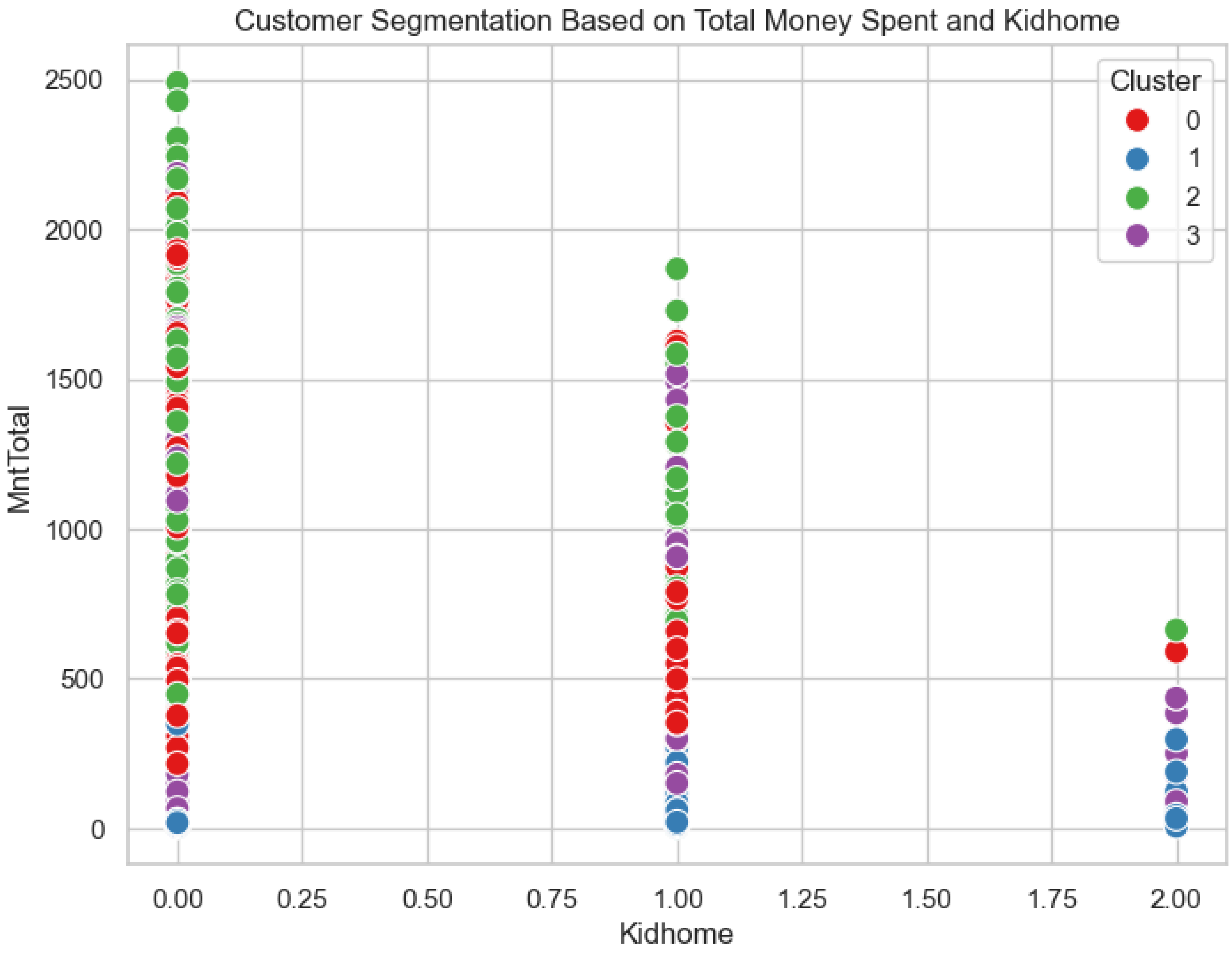

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61


61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93



In [44]:

```
plt.figure(figsize=(8, 6))

# Plot using 'Kidhome' and 'MntTotal' (total money spent)
sns.scatterplot(x='Kidhome', y='MntTotal', hue='Cluster', palette='Set1', data=df_cleaned, s=100)

plt.title('Customer Segmentation Based on Total Money Spent and Kidhome')
plt.show()
```



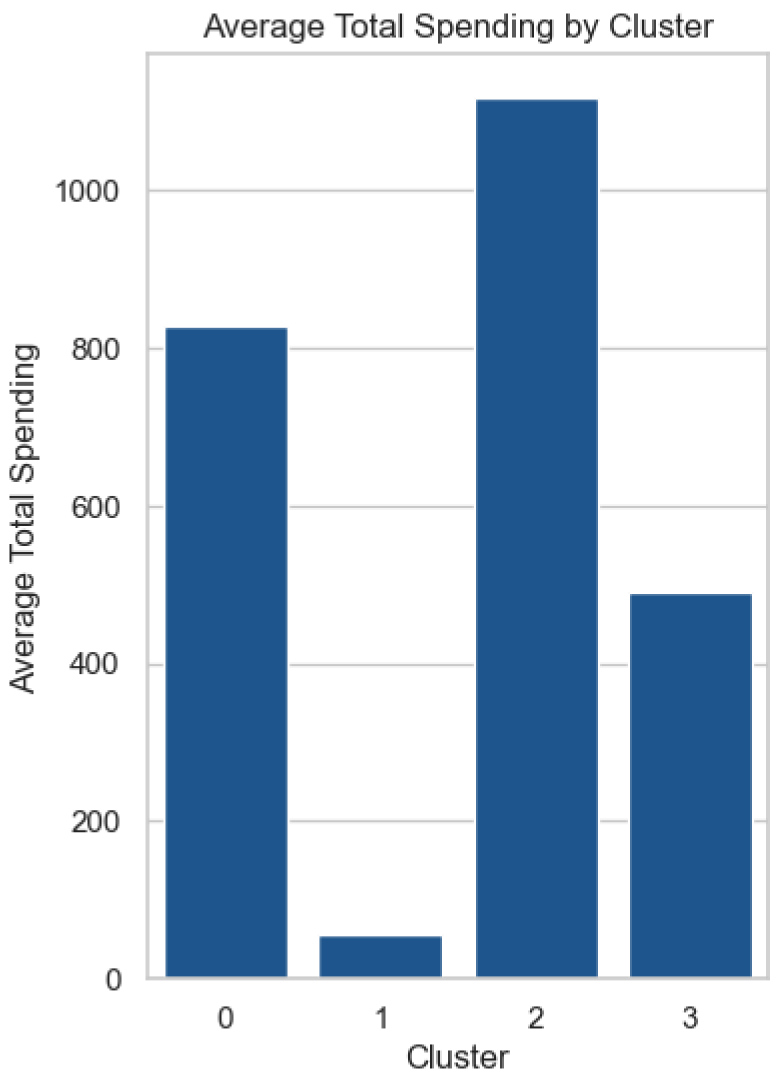
In [45]:
```
cluster_summary = df_cleaned.groupby('Cluster').mean()
print(cluster_summary)
```

```
          Kidhome  Teenhome    Recency    MntWines  MntFruits  \
Cluster
0        0.191398  0.552688  46.627957  449.677419  40.236559
1        0.829132  0.420168  48.791317   25.651261   3.596639
2        0.094567  0.505030  50.138833  612.746479  51.358149
3        0.466919  0.584121  50.334594  270.591682  21.580340

         MntMeatProducts  MntFishProducts  MntSweetProducts  MntGoldProds  \
Cluster
0             245.529032        54.283871         39.544086     61.690323
1              17.338936         5.448179          4.128852     16.232493
2             326.040241        73.175050         53.798793     69.788732
3             143.516068        33.559546         22.200378     41.937618

         NumDealsPurchases  NumWebPurchases  NumCatalogPurchases  \
Cluster
0                 2.688172         5.608602             3.651613
1                 1.802521         1.858543             0.530812
2                 2.315895         5.734406             4.859155
3                 2.691871         4.266541             2.534972

         NumStorePurchases  NumWebVisitsMonth  AcceptedCmp3  AcceptedCmp4  \
Cluster
0                 6.935484           4.498925      0.070968      0.135484
1                 2.644258           6.767507      0.103641      0.002801
2                10.867203           3.943662      0.052314      0.130785
3                 4.398866           5.451796      0.056711      0.064272

         AcceptedCmp5  AcceptedCmp1  AcceptedCmp2  Complain  Z_CostContact  \
Cluster
0            0.133333      0.105376      0.019355  0.006452            3.0
1            0.000000      0.000000      0.002801  0.014006            3.0
2            0.140845      0.124748      0.028169  0.008048            3.0
3            0.054820      0.058601      0.009452  0.005671            3.0

         Z_Revenue  Response        Age  Customer_Days  marital_Divorced  \
Cluster
0             11.0  0.204301  52.449462    2519.612903          0.094624
1             11.0  0.109244  47.844538    2479.571429          0.102241
2             11.0  0.154930  52.875252    2545.945674          0.110664
3             11.0  0.156900  52.621928    2520.179584          0.109641
```

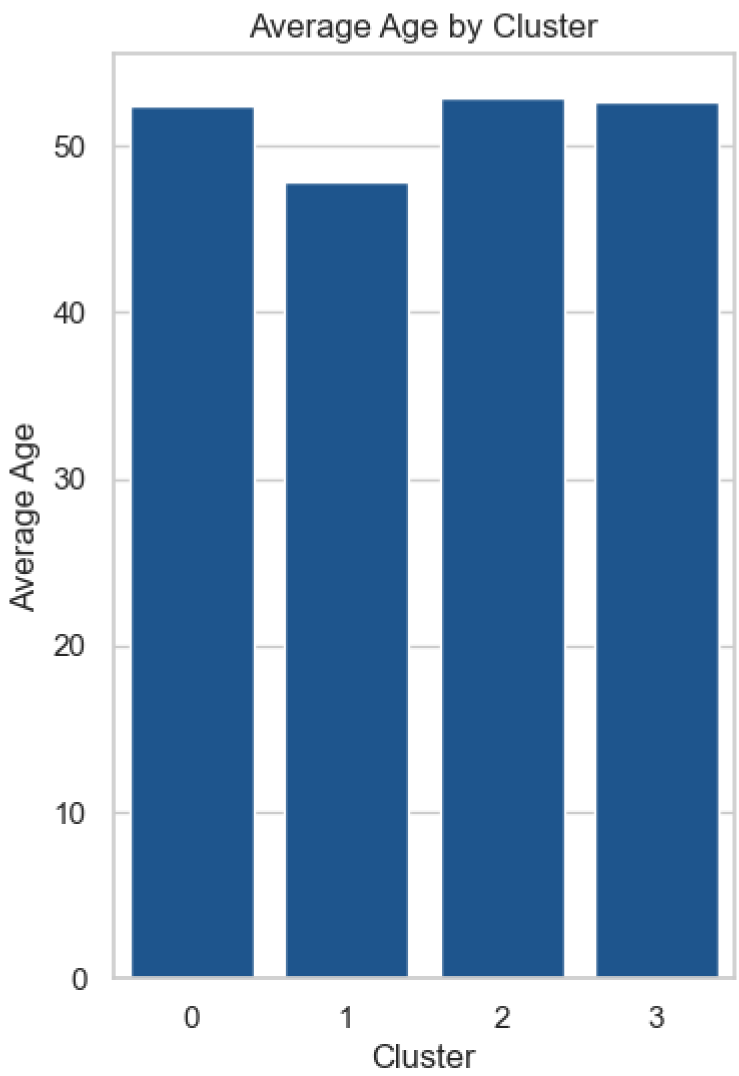
```
         marital_Married  marital_Single  marital_Together  marital_Widow  \
Cluster
0               0.384946        0.208602          0.264516       0.047312
1               0.378151        0.236695          0.256303       0.026611
2               0.396378        0.211268          0.243461       0.038229
3               0.393195        0.200378          0.266541       0.030246

         education_2n Cycle  education_Basic  education_Graduation  \
Cluster
0                  0.083871         0.002151              0.537634
1                  0.109244         0.068627              0.488796
2                  0.082495         0.000000              0.490946
3                  0.075614         0.007561              0.510397

         education_Master  education_PhD     MntTotal  MntRegularProds  \
Cluster
0                0.159140       0.217204   829.270968       767.580645
1                0.159664       0.173669    56.163866        39.931373
2                0.181087       0.245473  1117.118712      1047.329980
3                0.162571       0.243856   491.448015       449.510397

         AcceptedCmpOverall
Cluster
0                  0.464516
1                  0.109244
2                  0.476861
3                  0.243856
```

In [49]:

```
plt.figure(figsize=(4, 6))
sns.barplot(x=cluster_summary.index, y='MntTotal', data=cluster_summary)
plt.title('Average Total Spending by Cluster')
plt.ylabel('Average Total Spending')
plt.xlabel('Cluster')
plt.show()
```

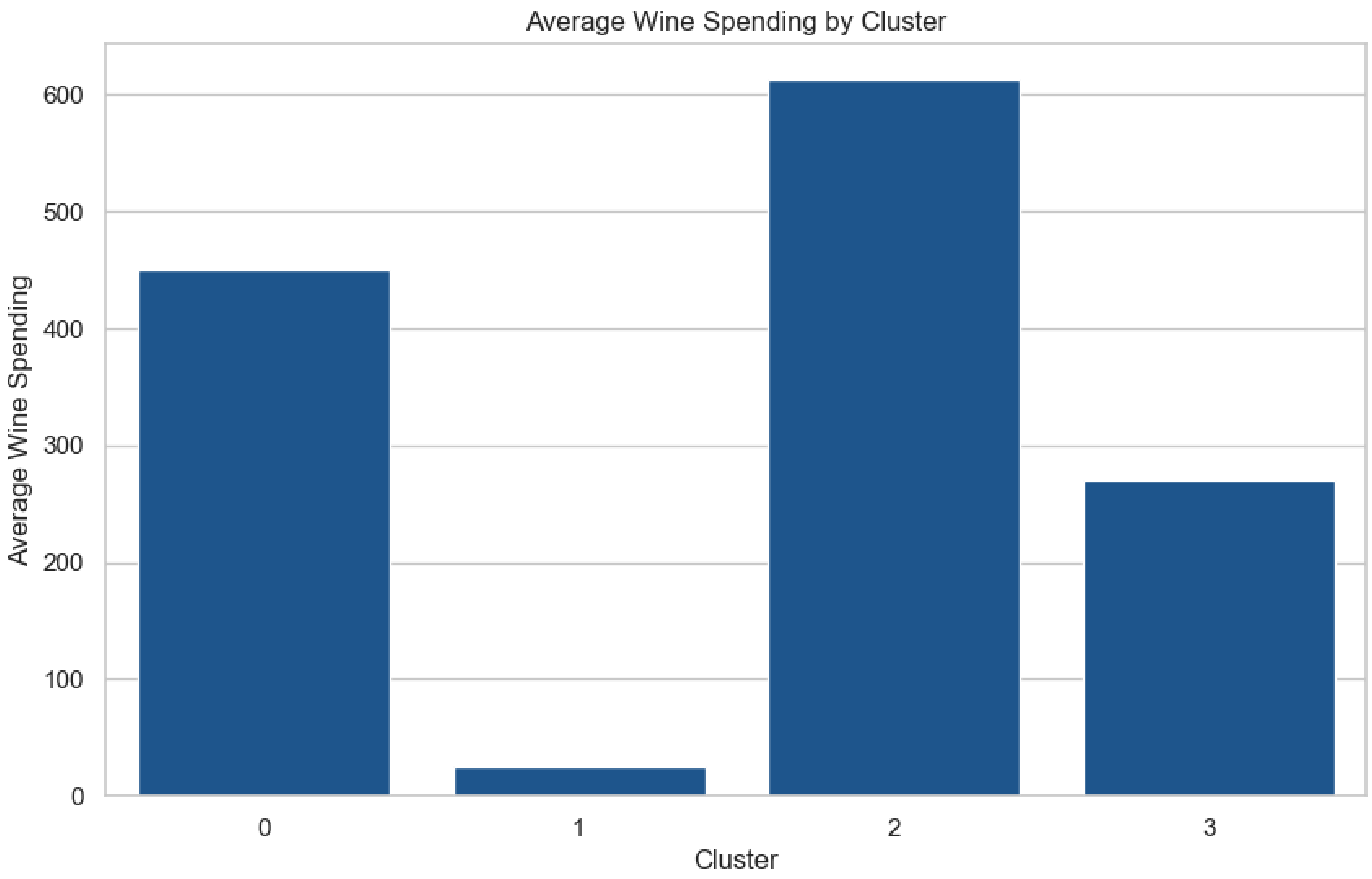


In [48]:
```
plt.figure(figsize=(4, 6))
sns.barplot(x=cluster_summary.index, y='Age', data=cluster_summary)
plt.title('Average Age by Cluster')
plt.ylabel('Average Age')
plt.xlabel('Cluster')
plt.show()
```



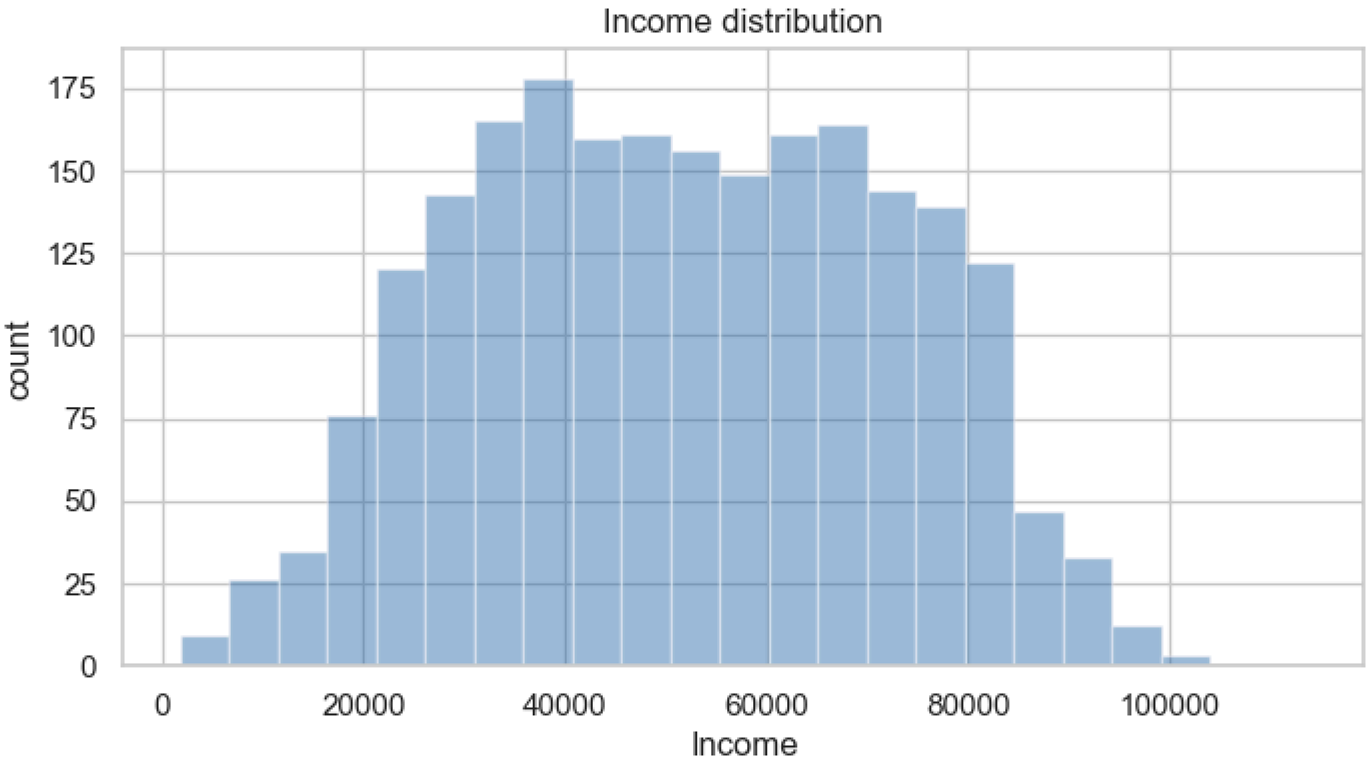
In [50]:

```
plt.figure(figsize=(10, 6))
sns.barplot(x=cluster_summary.index, y='MntWines', data=cluster_summary)
plt.title('Average Wine Spending by Cluster')
plt.ylabel('Average Wine Spending')
plt.xlabel('Cluster')
plt.show()
```



In [81]:

```
plt.figure(figsize=(8,4))
sns.distplot(df['Income'], kde=False, hist=True)
plt.title('Income distribution', size=12)
plt.ylabel('count');
```



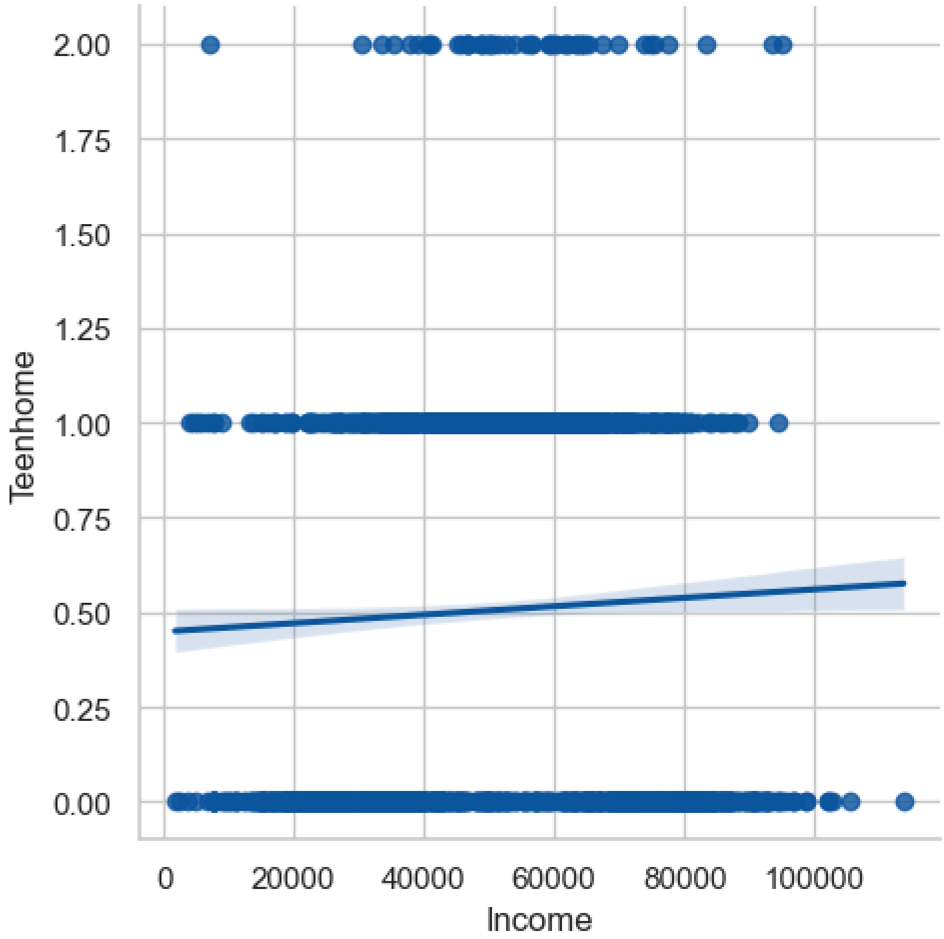
```
In [51]: df['Income'] = df['Income'].fillna(df['Income'].median())
```

```
In [62]: list(df.columns)
```

```
Out[62]: ['Income',
          'Kidhome',
          'Teenhome',
          'Recency',
          'MntWines',
          'MntFruits',
          'MntMeatProducts',
          'MntFishProducts',
          'MntSweetProducts',
          'MntGoldProds',
          'NumDealsPurchases',
          'NumWebPurchases',
          'NumCatalogPurchases',
          'NumStorePurchases',
          'NumWebVisitsMonth',
          'AcceptedCmp3',
          'AcceptedCmp4',
          'AcceptedCmp5',
          'AcceptedCmp1',
          'AcceptedCmp2',
          'Complain',
          'Z_CostContact',
          'Z_Revenue',
          'Response',
          'Age',
          'Customer_Days',
          'marital_Divorced',
          'marital_Married',
          'marital_Single',
          'marital_Together',
          'marital_Widow',
          'education_2n Cycle',
          'education_Basic',
          'education_Graduation',
          'education_Master',
          'education_PhD',
          'MntTotal',
          'MntRegularProds',
          'AcceptedCmpOverall',
          'Dependents']
```

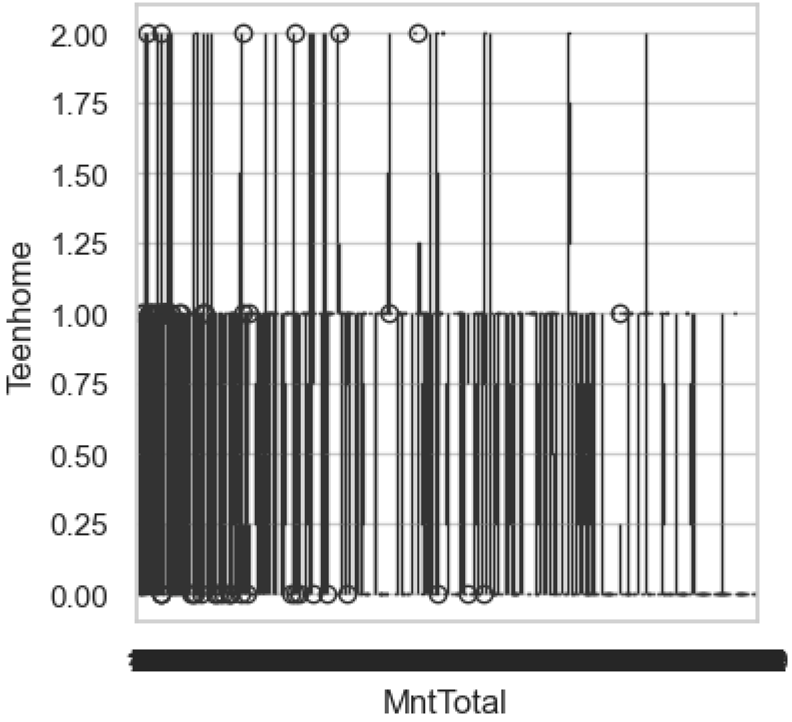
In [58]:

```
sns.lmplot(x='Income', y='Teenhome', data=df[df['Income'] < 200000]);
```



In [60]:

```
plt.figure(figsize=(4,4))
sns.boxplot(x='MntTotal', y='Teenhome', data=df);
```


Teenhome
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
MntTotal