```python
import  numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv("investment_survey.csv")

df.head()
```

```
   Gender  Age  Working_professional  Annual_income  \
0    Male   19                     0              0
1  Female   24                     0              0
2    Male   24                     1         400000
3    Male   22                     1         150000
4    Male   25                     1         250000


            Mode_of_investment Investment_per_month  \
0               Banking - RD, FD                  200
1               Banking - RD, FD                 1000
2  Stocks - Intraday, long term                 5000
3                   Mutual Funds                 2000
4                   Mutual Funds                 2000


            Motivation_cause                  Resources_used  \
0             Family member         Family members/Friends
1      Social media/ Articles            Mobile applications
2      Social media/ Articles  News articles / Current affairs
3  Agents/Investment brokers  News articles / Current affairs
4             Self Interest                          Books


                           Goal_for_investment
Duration_to_save(in_Years)    \
0                                Wealth generation
10
1                                Wealth generation
3
2  Planning for finacial freedom after Retirement
10
3                                Wealth generation
10
4                                Wealth generation
15


   Unnamed: 10
0          NaN
1          NaN
2          NaN
3          NaN
4          NaN
```

```python
df.tail()
```

```
     Gender  Age  Working_professional  Annual_income  \
95    Male   22                     0              0
96    Male   29                     1         250000
97  Female   23                     1         150000
98  Female   22                     0              0
99    Male   26                     1         250000

               Mode_of_investment  Investment_per_month  \
95  Stocks - Intraday, long term                    500
96                  Mutual Funds                   2000
97              Banking - RD, FD                   1000
98                  Mutual Funds                   1000
99  Stocks - Intraday, long term                   1500

            Motivation_cause                     Resources_used  \
95       Social media/ Articles              Mobile applications
96                 Through Bank  News articles / Current affairs
97                Family member         Family members/Friends
98                 Through Bank              Mobile applications
99      Agents/Investment brokers         Family members/Friends

                             Goal_for_investment  \
95                             Wealth generation
96                               Personal Savings
97     Construction of own House / Buying a new Car
98                             Wealth generation
99   Planning for finacial freedom after Retirement

   Duration_to_save(in_Years)  Unnamed: 10
95                          10          NaN
96                           2          NaN
97                          10          NaN
98                           6          NaN
99                          10          NaN
```

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 11 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   Gender                 100 non-null     object
 1   Age                    100 non-null     int64
 2   Working_professional   100 non-null     int64
 3   Annual_income          100 non-null     int64
 4   Mode_of_investment     100 non-null     object
 5   Investment_per_month   100 non-null     object
 6   Motivation_cause       100 non-null     object
 7   Resources_used         100 non-null     object
```

```
 8    Goal_for_investment          100 non-null     object
 9    Duration_to_save(in_Years)   100 non-null     object
 10   Unnamed: 10                  0 non-null       float64
dtypes: float64(1), int64(3), object(7)
memory usage: 8.7+ KB
```

df.shape

(100, 11)

df.describe(include = "all")

```
        Gender         Age  Working_professional    Annual_income  \
count      100  100.000000            100.000000       100.000000
unique       2         NaN                   NaN              NaN
top       Male         NaN                   NaN              NaN
freq        64         NaN                   NaN              NaN
mean       NaN   25.370000              0.590000    166440.800000
std        NaN    5.781641              0.494311    161676.882399
min        NaN   18.000000              0.000000         0.000000
25%        NaN   22.000000              0.000000         0.000000
50%        NaN   24.000000              1.000000    150000.000000
75%        NaN   27.000000              1.000000    300000.000000
max        NaN   56.000000              1.000000    600000.000000

                      Mode_of_investment Investment_per_month
Motivation_cause  \
count                                100                  100
100
unique                                 9                   20
11
top        Stocks - Intraday, long term                 1000     Family
member
freq                                  29                   19
24
mean                                 NaN                  NaN
NaN
std                                  NaN                  NaN
NaN
min                                  NaN                  NaN
NaN
25%                                  NaN                  NaN
NaN
50%                                  NaN                  NaN
NaN
75%                                  NaN                  NaN
NaN
max                                  NaN                  NaN
NaN
```

	Resources_used	Goal_for_investment
count	100	100
unique	7	8
top	Family members/Friends	Wealth generation
freq	48	28
mean	NaN	NaN
std	NaN	NaN
min	NaN	NaN
25%	NaN	NaN
50%	NaN	NaN
75%	NaN	NaN
max	NaN	NaN

	Duration_to_save(in_Years)	Unnamed: 10
count	100	0.0
unique	15	NaN
top	10	NaN
freq	30	NaN
mean	NaN	NaN
std	NaN	NaN
min	NaN	NaN
25%	NaN	NaN
50%	NaN	NaN
75%	NaN	NaN
max	NaN	NaN

```
df.isnull().sum()
```

```
Gender                        0
Age                           0
Working_professional          0
Annual_income                 0
Mode_of_investment            0
Investment_per_month          0
Motivation_cause              0
Resources_used                0
Goal_for_investment           0
Duration_to_save(in_Years)    0
Unnamed: 10                 100
dtype: int64
```

```
df.duplicated()
```

```
0     False
1     False
2     False
3     False
4     False
      ...
95    False
```

```
96    False
97    False
98    False
99    False
Length: 100, dtype: bool
```

```
df .size
```

```
1100
```

```
df.columns
```

```
Index(['Gender', 'Age', 'Working_professional', 'Annual_income',
       'Mode_of_investment', 'Investment_per_month',
'Motivation_cause',
       'Resources_used', 'Goal_for_investment',
'Duration_to_save(in_Years) ',
       'Unnamed: 10'],
      dtype='object')
```

```
df.sample(4)
```

```
    Gender  Age  Working_professional  Annual_income  \
99    Male   26                     1         250000
78  Female   26                     1         350000
2     Male   24                     1         400000
38    Male   22                     0              0

                           Mode_of_investment Investment_per_month  \
99            Stocks - Intraday, long term                    1500
78  Gold / Any other Materialistic investment                2500
2             Stocks - Intraday, long term                    5000
38  Gold / Any other Materialistic investment                2000

            Motivation_cause                       Resources_used  \
99  Agents/Investment brokers           Family members/Friends
78                    Schemes           Family members/Friends
2      Social media/ Articles  News articles / Current affairs
38                    Friends           Family members/Friends

                          Goal_for_investment  \
99  Planning for finacial freedom after Retirement
78                            Personal Savings
2   Planning for finacial freedom after Retirement
38                                All the above

   Duration_to_save(in_Years)   Unnamed: 10
99                          10          NaN
78                           5          NaN
2                           10          NaN
38                           3          NaN
```

```
df.head()

    Gender  Age  Working_professional  Annual_income  \
0     Male   19                     0              0
1   Female   24                     0              0
2     Male   24                     1         400000
3     Male   22                     1         150000
4     Male   25                     1         250000

              Mode_of_investment Investment_per_month  \
0               Banking - RD, FD                  200
1               Banking - RD, FD                 1000
2  Stocks - Intraday, long term                 5000
3                   Mutual Funds                 2000
4                   Mutual Funds                 2000

              Motivation_cause                   Resources_used  \
0               Family member         Family members/Friends
1        Social media/ Articles            Mobile applications
2        Social media/ Articles  News articles / Current affairs
3   Agents/Investment brokers  News articles / Current affairs
4               Self Interest                          Books

                               Goal_for_investment
Duration_to_save(in_Years)    \
0                                 Wealth generation
10
1                                 Wealth generation
3
2  Planning for finacial freedom after Retirement
10
3                                 Wealth generation
10
4                                 Wealth generation
15

    Unnamed: 10
0           NaN
1           NaN
2           NaN
3           NaN
4           NaN

k = df.groupby(["Age"])["Annual_income"].sum()
k

Age
18      360000
19           0
20          80
```

```
21        240000
22        270000
23       1200000
24       1400000
25       2160000
26       2950000
27       2184000
28       2570000
29        930000
30        180000
31        200000
32        300000
35        300000
36        600000
37        400000
45        200000
48             0
56        200000
Name: Annual_income, dtype: int64

sns.countplot(x = "Age",data  = df )
plt.show()
```
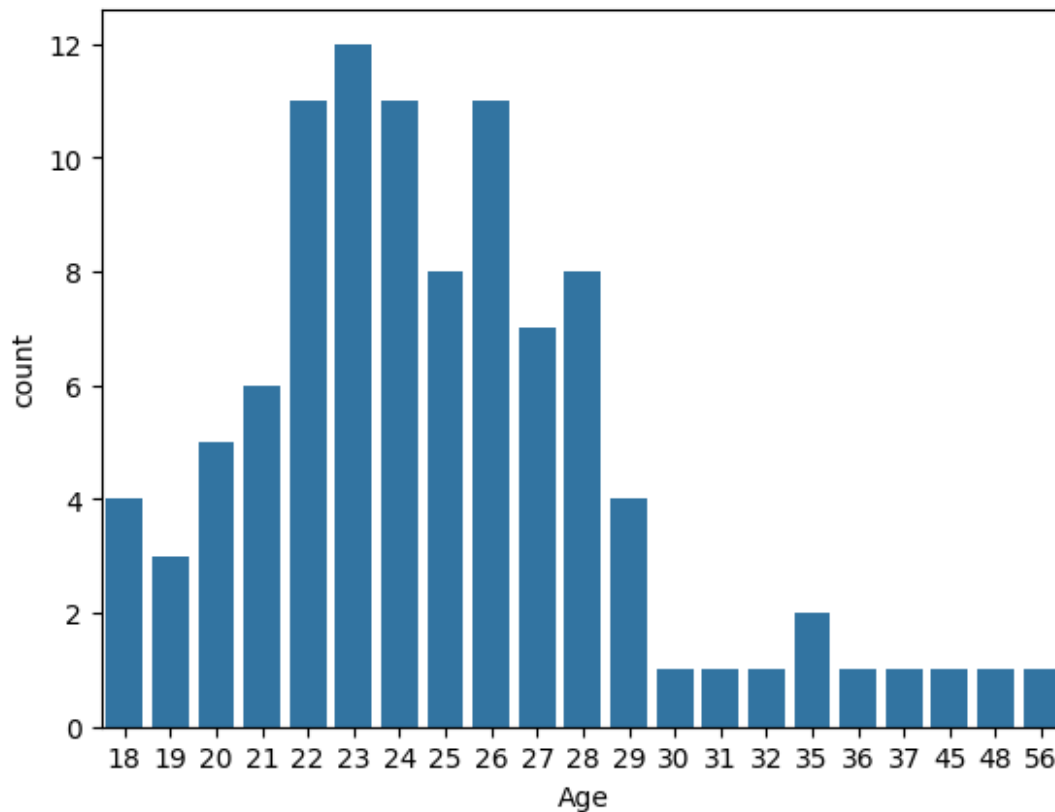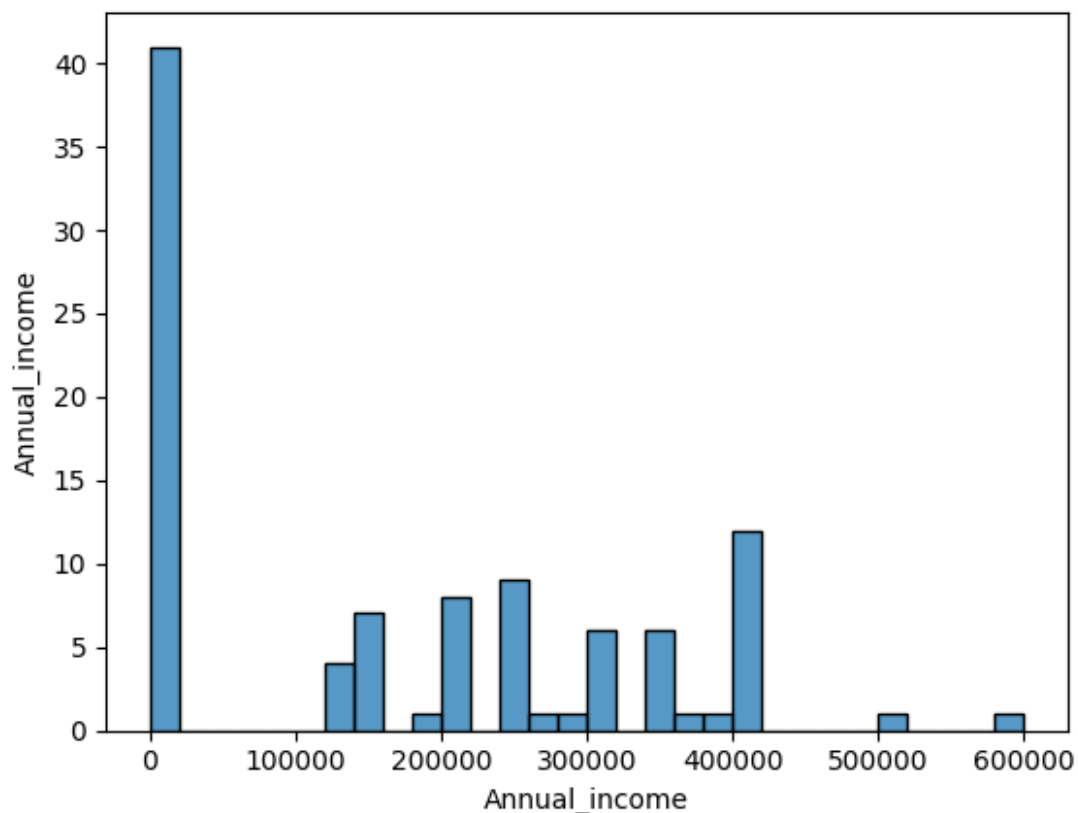
```
sns.histplot(df["Age"],bins = 30,)
plt.title("Age distribution")
plt.ylabel('Frequency')
plt.show()
```


Age distribution
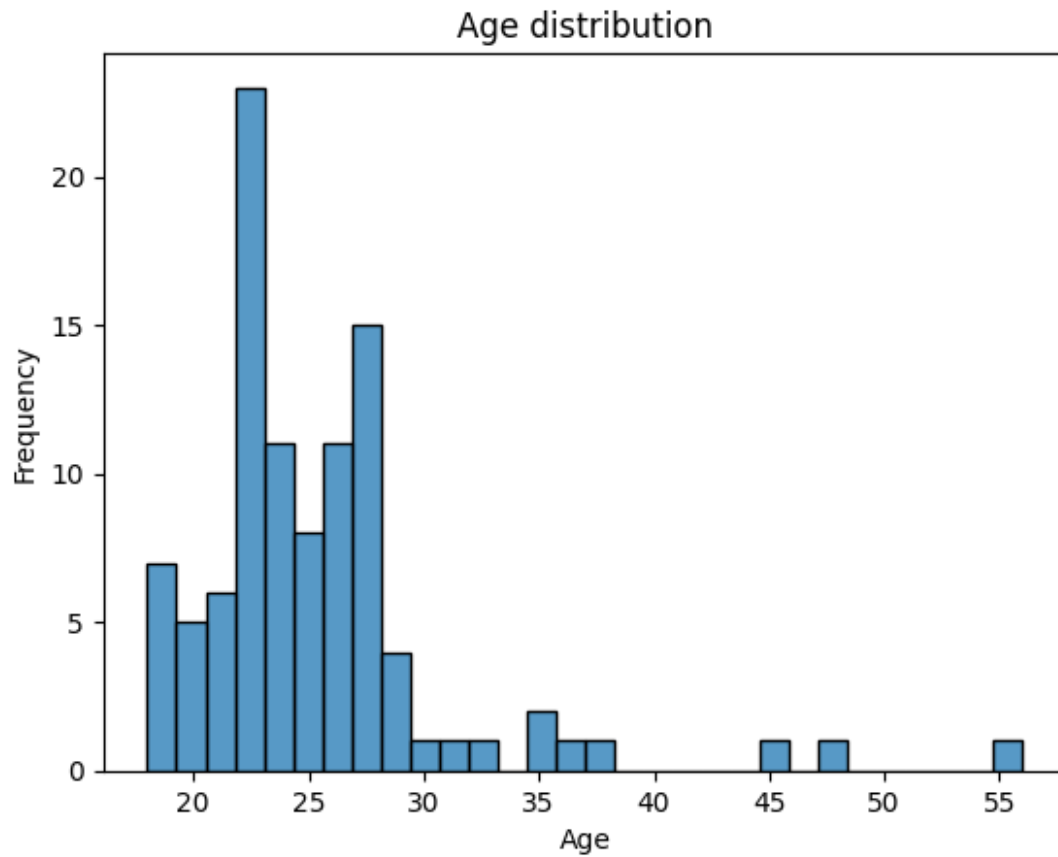
```
sns.histplot(df["Annual_income"],bins = 30)
plt.ylabel("Annual_income")
plt.show()
```

```
df.head()

    Gender  Age  Working_professional  Annual_income  \
0    Male   19                       0              0
1  Female   24                       0              0
2    Male   24                       1         400000
3    Male   22                       1         150000
4    Male   25                       1         250000

                Mode_of_investment  Investment_per_month  \
0               Banking - RD, FD                     200
1               Banking - RD, FD                    1000
2     Stocks - Intraday, long term                  5000
3                   Mutual Funds                    2000
4                   Mutual Funds                    2000

              Motivation_cause                  Resources_used  \
0               Family member          Family members/Friends
1        Social media/ Articles            Mobile applications
2        Social media/ Articles  News articles / Current affairs
3  Agents/Investment brokers     News articles / Current affairs
4               Self Interest                           Books

                      Goal_for_investment
```

```
Duration_to_save(in_Years)     \
0                              Wealth generation
10
1                              Wealth generation
3
2  Planning for finacial freedom after Retirement
10
3                              Wealth generation
10
4                              Wealth generation
15

    Unnamed: 10
0          NaN
1          NaN
2          NaN
3          NaN
4          NaN

sns.countplot(x = "Gender", data = df)
plt.show()
```
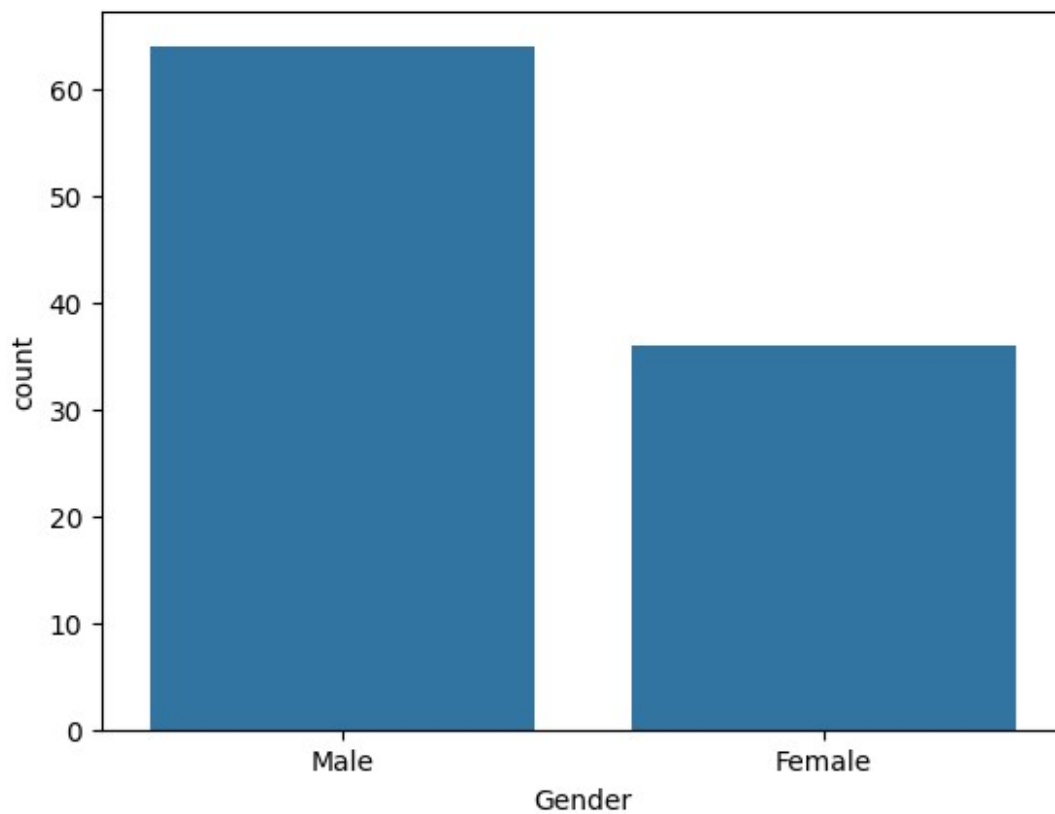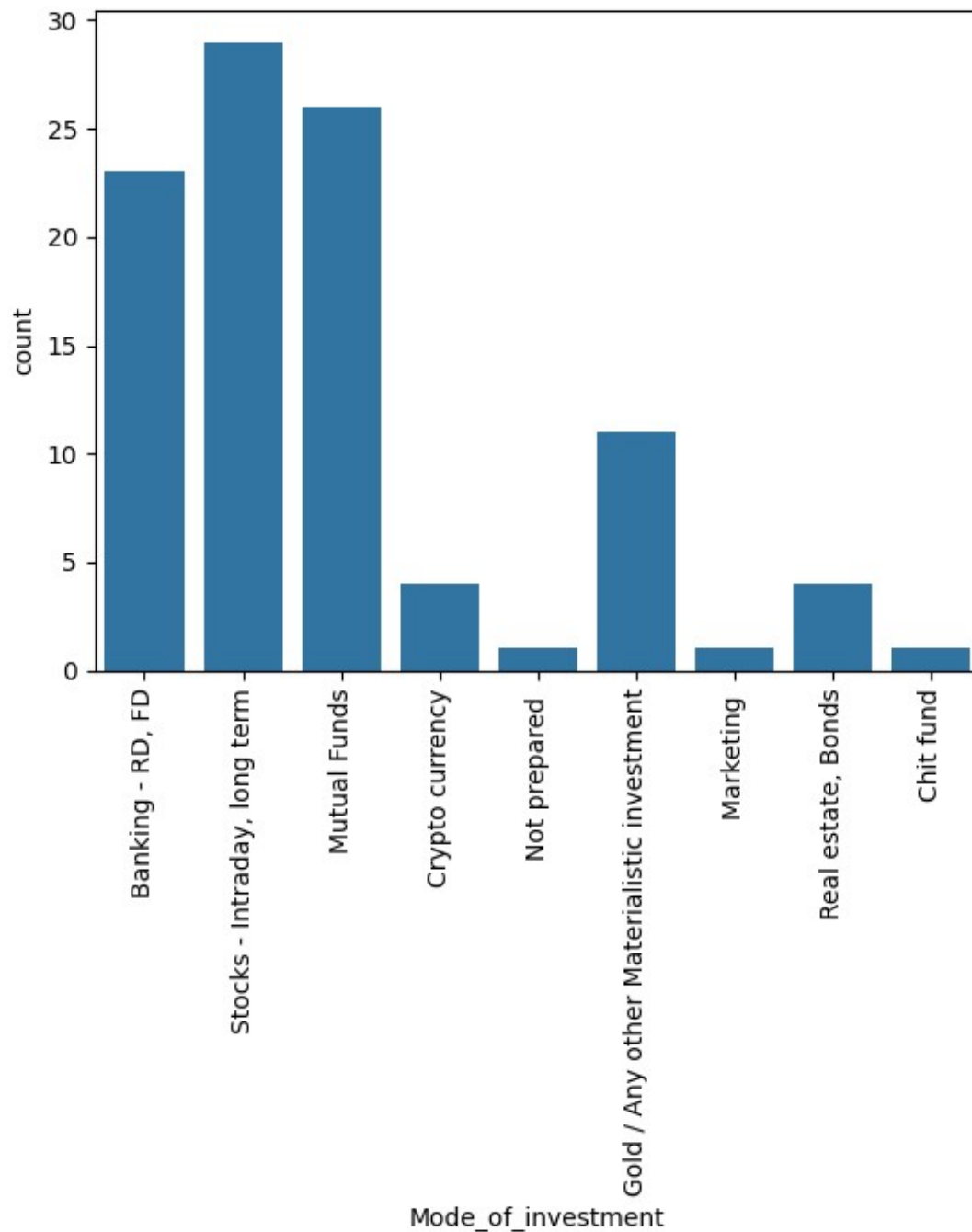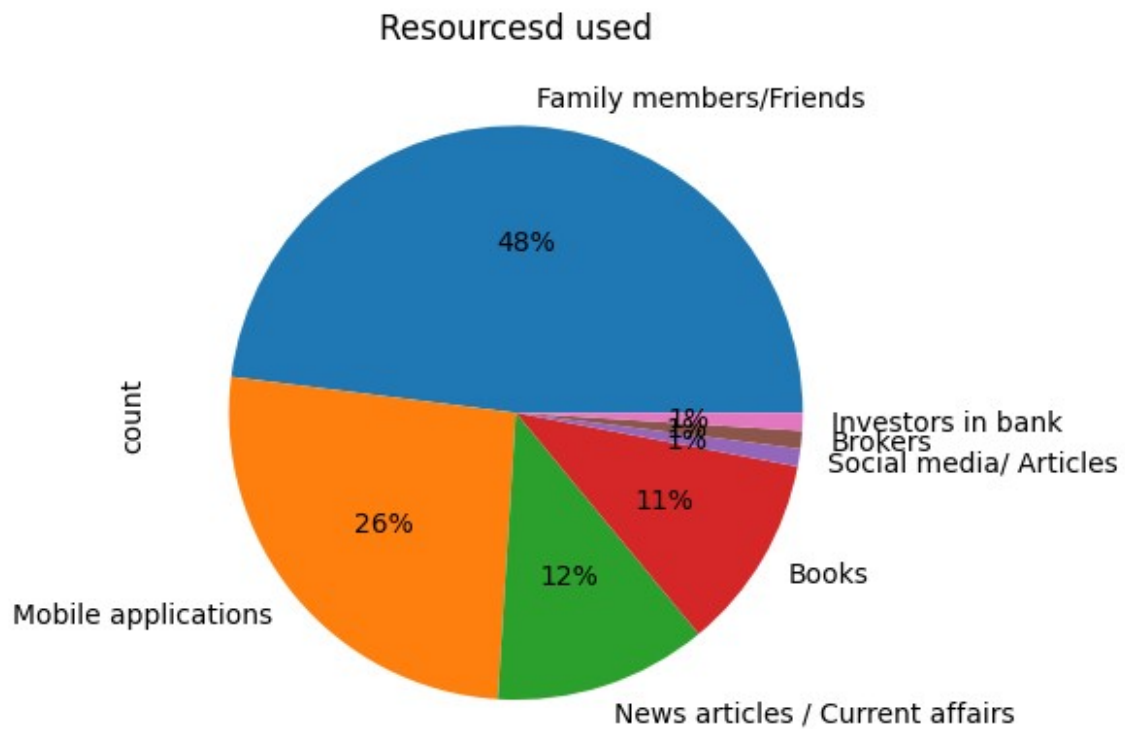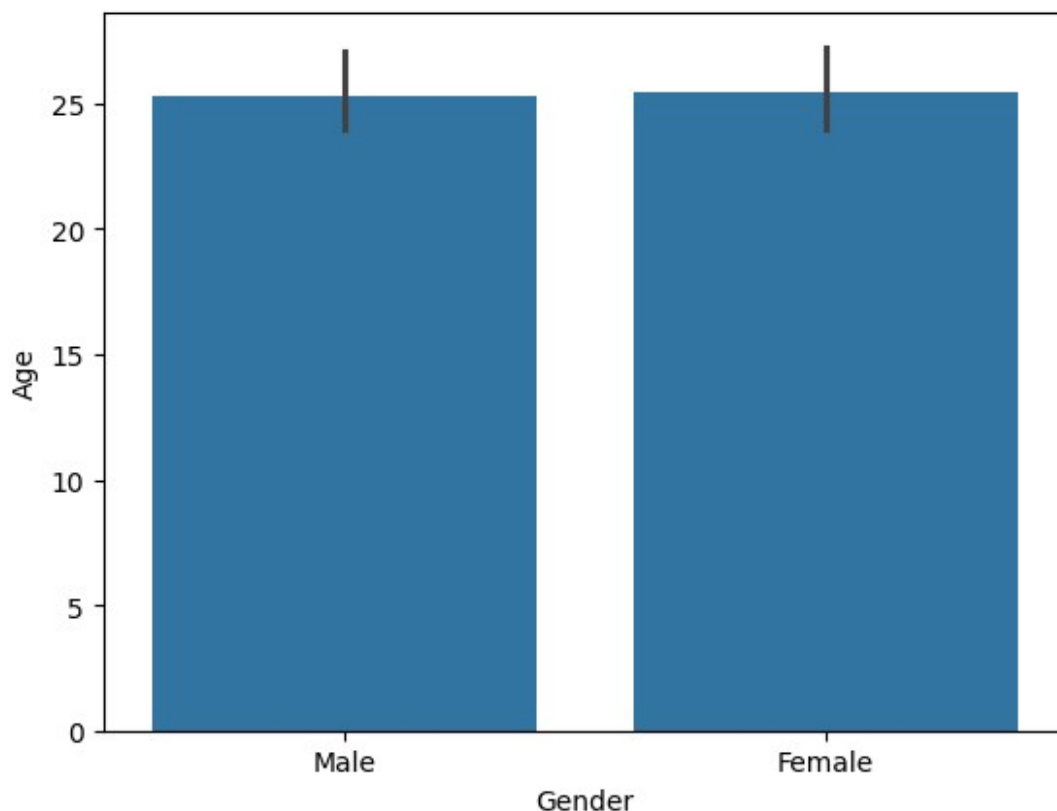
```
sns.countplot(x = "Mode_of_investment", data = df)
plt.xticks(rotation =90)
plt.show()
```



```
df["Resources_used"].value_counts().plot.pie(autopct = "%1.0f%%")
plt.title("Resourcesd used")
plt.show()
```

Resourcesd used


```
sns.barplot(x = "Gender",y = "Age",data = df)
plt.show()
```

```
df.head()

   Gender  Age  Working_professional  Annual_income  \
0    Male   19                     0              0
1  Female   24                     0              0
2    Male   24                     1         400000
3    Male   22                     1         150000
4    Male   25                     1         250000

             Mode_of_investment  Investment_per_month  \
0              Banking - RD, FD                   200
1              Banking - RD, FD                  1000
2     Stocks - Intraday, long term              5000
3                  Mutual Funds                  2000
4                  Mutual Funds                  2000

              Motivation_cause                  Resources_used  \
0                 Family member       Family members/Friends
1           Social media/ Articles        Mobile applications
2           Social media/ Articles  News articles / Current affairs
3     Agents/Investment brokers  News articles / Current affairs
4                 Self Interest                          Books

                            Goal_for_investment
Duration_to_save(in_Years)    \
```

```
0                            Wealth generation
10
1                            Wealth generation
3
2  Planning for finacial freedom after Retirement
10
3                            Wealth generation
10
4                            Wealth generation
15

    Unnamed: 10
0         NaN
1         NaN
2         NaN
3         NaN
4         NaN
```
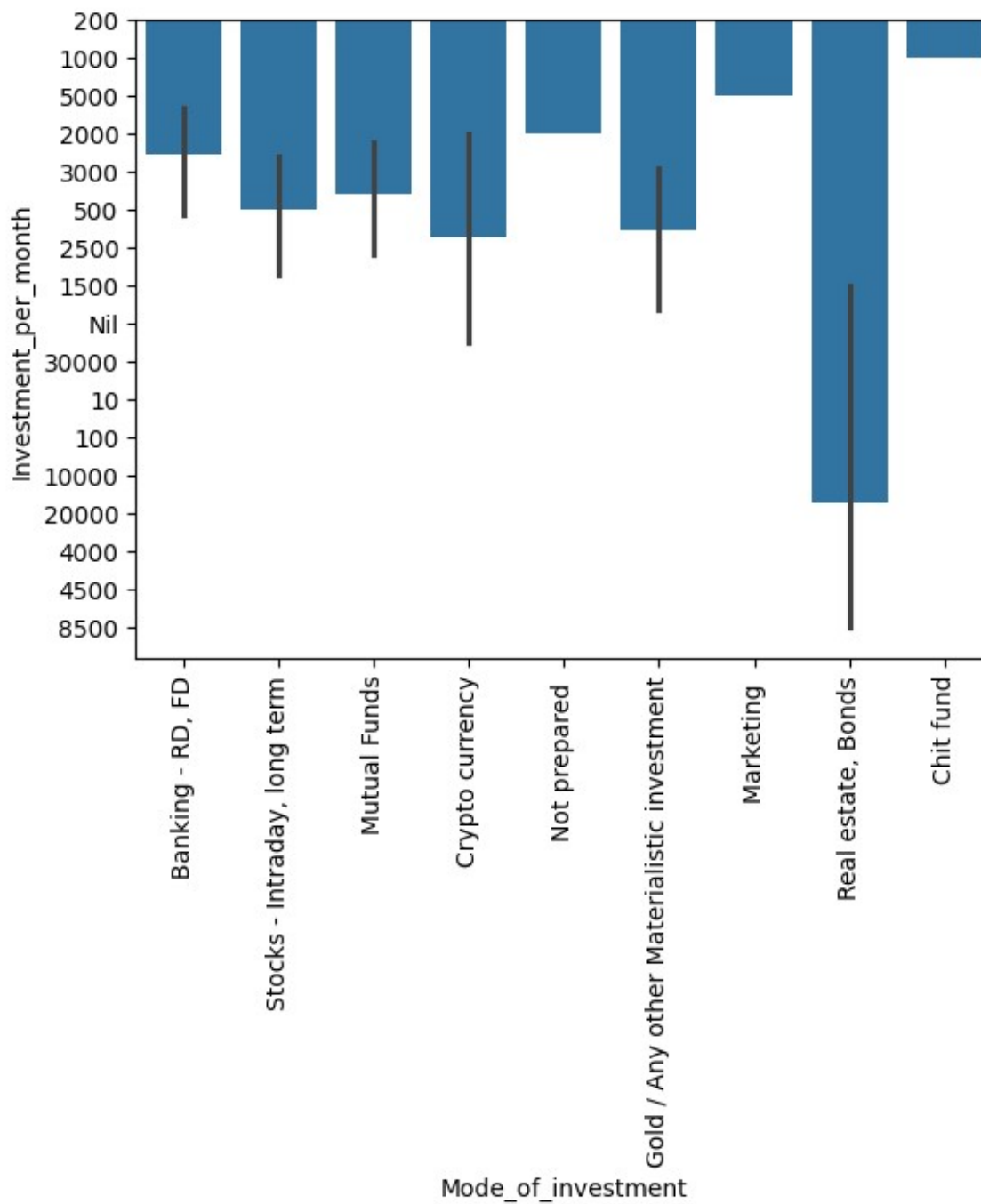
```python
sns.barplot( x = "Mode_of_investment", y = "Investment_per_month",data
= df)
plt.xticks(rotation = 90)
plt.show()
```

```
df.head()

    Gender  Age  Working_professional   Annual_income  \
0     Male   19                      0               0
1   Female   24                      0               0
2     Male   24                      1          400000
3     Male   22                      1          150000
4     Male   25                      1          250000
```

```
              Mode_of_investment Investment_per_month  \
0                 Banking - RD, FD                   200
1                 Banking - RD, FD                  1000
2        Stocks - Intraday, long term               5000
3                     Mutual Funds                  2000
4                     Mutual Funds                  2000

             Motivation_cause                   Resources_used  \
0               Family member          Family members/Friends
1        Social media/ Articles             Mobile applications
2        Social media/ Articles  News articles / Current affairs
3      Agents/Investment brokers  News articles / Current affairs
4                 Self Interest                           Books

                                 Goal_for_investment
Duration_to_save(in_Years)    \
0                                 Wealth generation
10
1                                 Wealth generation
3
2    Planning for finacial freedom after Retirement
10
3                                 Wealth generation
10
4                                 Wealth generation
15

    Unnamed: 10
0           NaN
1           NaN
2           NaN
3           NaN
4           NaN

sns.barplot(x = "Gender",y = "Investment_per_month",data = df)
plt.show

<function matplotlib.pyplot.show(close=None, block=None)>
```
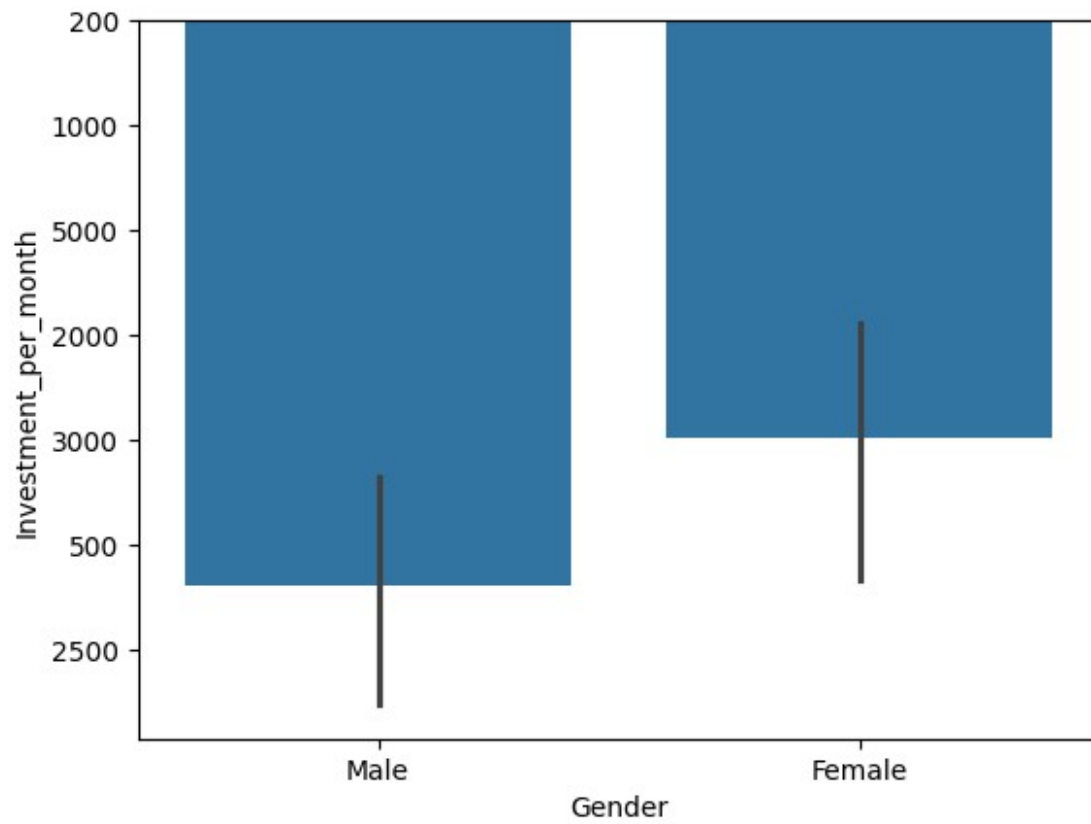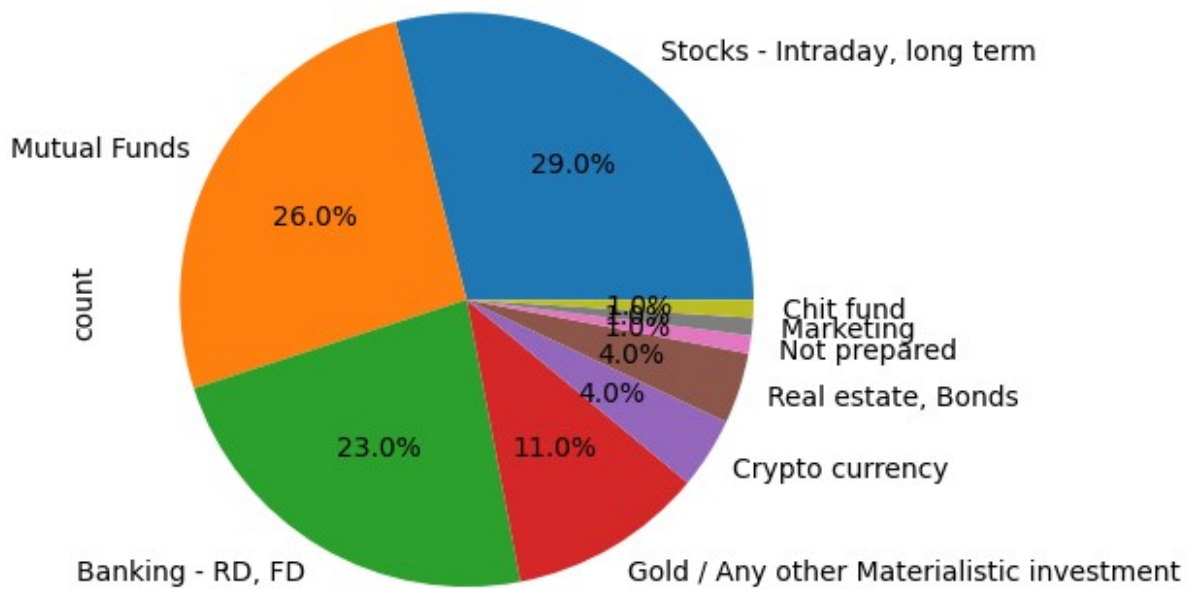
```
df["Mode_of_investment"].value_counts().plot(kind = "pie",autopct =
"%1.1f%%")
plt.title("Mode_of_investment")
plt.show()
```

Mode_of_investment


```
df.head()

   Gender  Age  Working_professional  Annual_income  \
0    Male   19                     0              0
1  Female   24                     0              0
2    Male   24                     1         400000
3    Male   22                     1         150000
4    Male   25                     1         250000


               Mode_of_investment  Investment_per_month  \
0               Banking - RD, FD                     200
1               Banking - RD, FD                    1000
2  Stocks - Intraday, long term                    5000
3                  Mutual Funds                    2000
4                  Mutual Funds                    2000


             Motivation_cause                  Resources_used  \
0               Family member       Family members/Friends
1       Social media/ Articles          Mobile applications
2       Social media/ Articles  News articles / Current affairs
3  Agents/Investment brokers  News articles / Current affairs
4               Self Interest                          Books


                           Goal_for_investment
Duration_to_save(in_Years)     \
0                              Wealth generation
```

```
10
1                                    Wealth generation
3
2   Planning for finacial freedom after Retirement
10
3                                    Wealth generation
10
4                                    Wealth generation
15

    Unnamed: 10
0           NaN
1           NaN
2           NaN
3           NaN
4           NaN
```
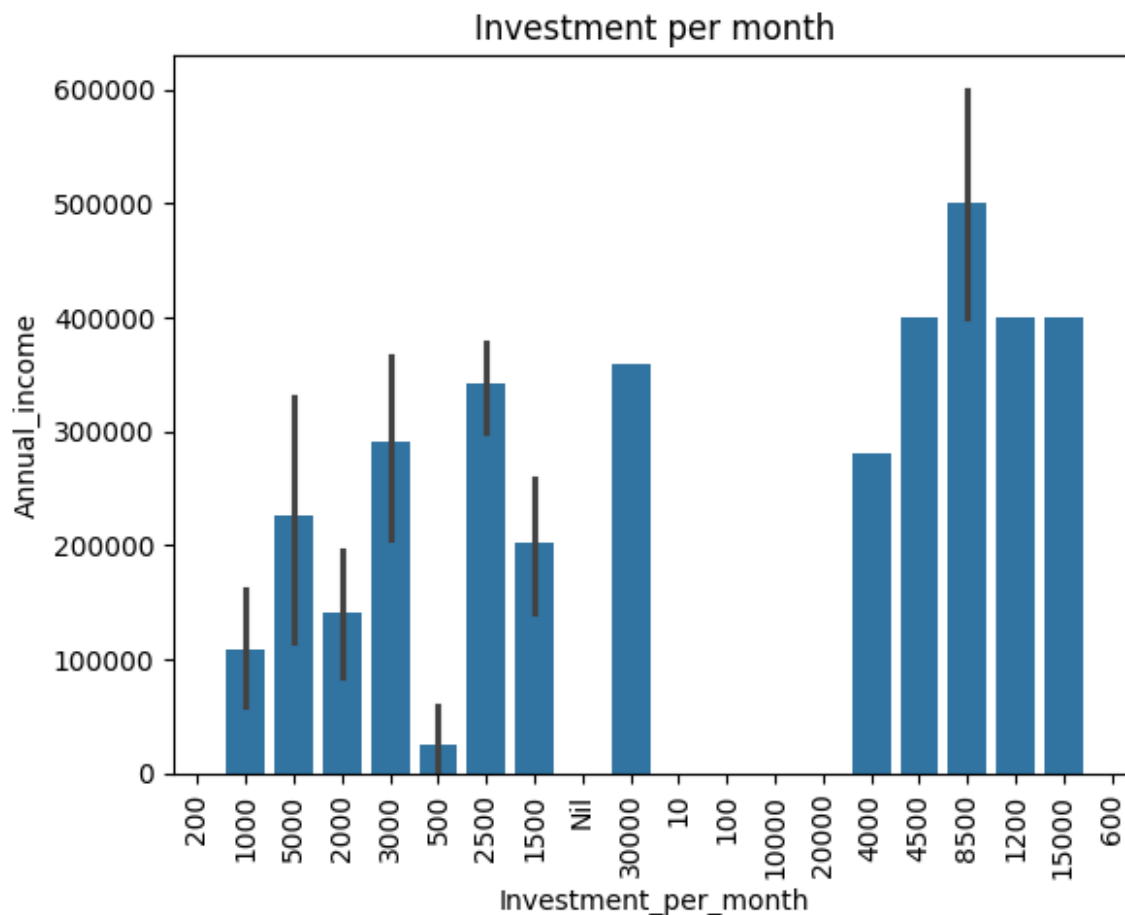
```python
sns.barplot (y = "Annual_income",x  = "Investment_per_month",data =
df)
plt.xticks(rotation = 90)
plt.title("Investment per month")
plt.show()
```

Investment per month


```
df["Goal_for_investment"].value_counts().plot.pie(autopct = "%1.2f%%")
plt.title ("Goal of investment")
plt.show()
```

Goal of investment