In [7]:
```
import pandas as pd
```

In [9]:
```
df=pd.read_csv(r"C:\Users\Pooja Shinde\Downloads\Loan_dataset.csv")
```

In [13]:
```
df
```

Out[13]:

	Loan ID	Customer ID	Loan Status	Current Loan Amount	Term	Credit Score	Annual Income	Years in current job	Home Ownership	Purpose	Monthly Debt	H
0	14dd8831-6af5-400b-83ec-68e61888a048	981165ec-3274-42f5-a3b4-d104041a9ca9	Fully Paid	445412.0	Short Term	709.0	1167493.0	8 years	Home Mortgage	Home Improvements	5214.74	
1	4771cc26-131a-45db-b5aa-537ea4ba5342	2de017a3-2e01-49cb-a581-08169e83be29	Fully Paid	262328.0	Short Term	NaN	NaN	10+ years	Home Mortgage	Debt Consolidation	33295.98	
2	4eed4e6a-aa2f-4c91-8651-ce984ee8fb26	5efb2b2b-bf11-4dfd-a572-3761a2694725	Fully Paid	99999999.0	Short Term	741.0	2231892.0	8 years	Own Home	Debt Consolidation	29200.53	
3	77598f7b-32e7-4e3b-a6e5-06ba0d98fe8a	e777faab-98ae-45af-9a86-7ce5b33b1011	Fully Paid	347666.0	Long Term	721.0	806949.0	3 years	Own Home	Debt Consolidation	8741.90	
4	d4062e70-befa-4995-8643-a0de73938182	81536ad9-5ccf-4eb8-befb-47a4d608658e	Fully Paid	176220.0	Short Term	NaN	NaN	5 years	Rent	Debt Consolidation	20639.70	
...	...	...	...	...	...	...	...	...	...	...	...	...
100509	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
100510	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
100511	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
100512	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
100513	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

100514 rows × 19 columns

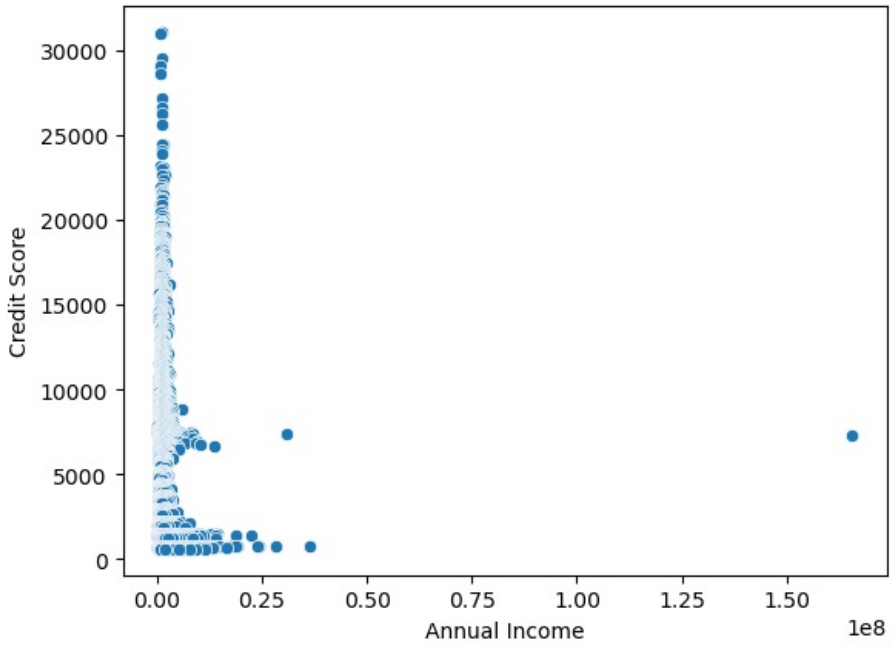
In [ ]:
```
### Python Project :
###Dataset : Loan_Data
###Domain : Finance
###Analyse the data and give the answers of below questions :

# 1.What is the average current loan amount for each loan status?
# 2.How does the credit score vary with the annual income?
#3.Is there a correlation between the number of open accounts and the current credit balance?
#4.What is the distribution of credit scores across different home ownership types?
#5.How does the annual income differ for different purposes of loans?
#6.What is the average monthly debt for each term (short-term vs. long-term)?
#7.Is there a correlation between years of credit history and the current credit balance?
#8.How does the credit score vary with the years in the current job?
#9.What is the relationship between the number of credit problems and the number of open accounts?
#10.What is the distribution of annual income across different loan statuses?
#11.Is there a correlation between the current loan amount and the number of open accounts?
#12.How does the monthly debt vary with the years of credit history?
#13.What is the average annual income for each purpose of loan?
#14.How does the credit score vary with the number of credit problems?
#15.Is there a correlation between the number of credit problems and the current credit balance?
#16.What is the distribution of current loan amounts across different home ownership types?
#17.How does the annual income vary with the years in the current job?
#18.Is there a correlation between the current loan amount and the monthly debt?
#19.What is the average monthly debt for each home ownership type?
#20.How does the credit score vary with the number of open accounts?
#21.What is the distribution of credit scores across different loan statuses?
#22.Is there a correlation between the current loan amount and the years of credit history?
#23.How does the monthly debt vary with the number of credit problems?
#24.What is the average current loan amount for each purpose of loan?
#25.How does the credit score vary with the current credit balance?
#26.Is there a correlation between the annual income and the current credit balance?
#27.What is the distribution of annual income across different terms (short-term vs. long-term)?
#28.How does the credit score vary with the number of credit problems?
#29.Is there a correlation between the current loan amount and the number of credit problems?
#30.What is the relationship between the number of open accounts and the years of credit history?
```

In [16]:
```
# 1.What is the average current loan amount for each loan status?
df.groupby(by='Loan Status')['Current Loan Amount'].mean()
```

Out[16]:
```
Loan Status
Charged Off    3.249752e+05
Fully Paid     1.510694e+07
Name: Current Loan Amount, dtype: float64
```

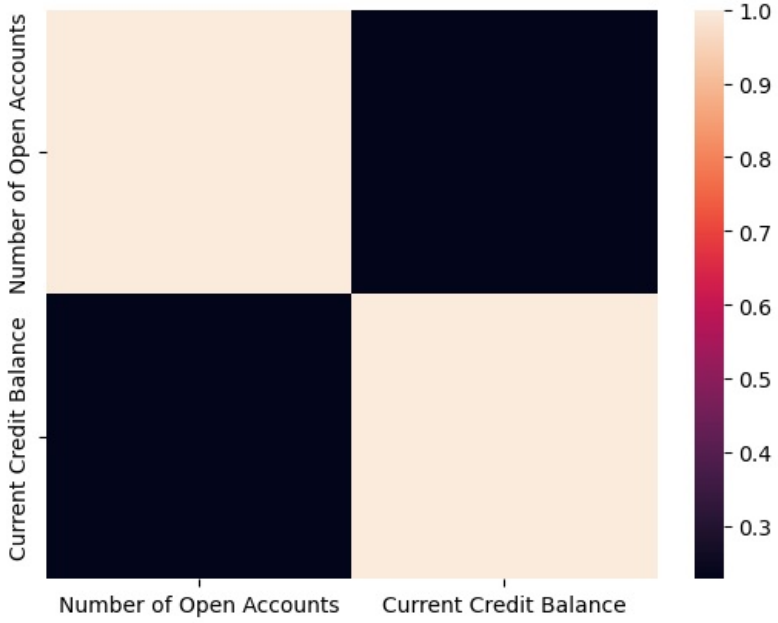
In [19]:
```
import seaborn as sns
# 2.How does the credit score vary with the annual income?
A=df.groupby('Annual Income')['Credit Score'].sum().reset_index()
A=A.sort_values(by='Credit Score',ascending=False)
A
sns.scatterplot(data=A,x='Annual Income',y='Credit Score')
```

Out[19]:
```
<Axes: xlabel='Annual Income', ylabel='Credit Score'>
```



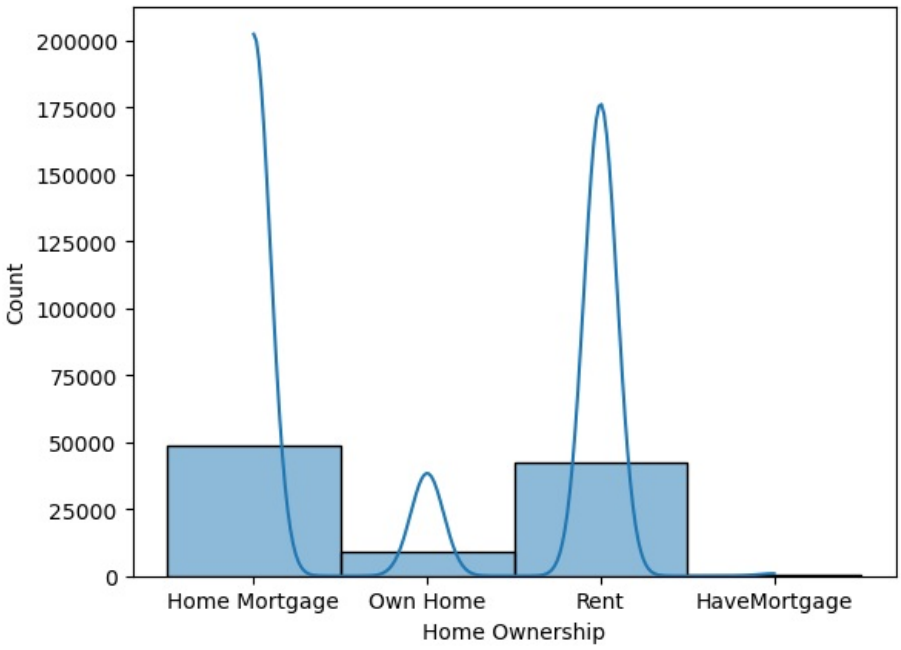
In [36]:
```
#3.Is there a correlation between the number of open accounts and the current credit balance?
sns.heatmap(df[['Number of Open Accounts','Current Credit Balance']].corr())
```

Out[36]:
```
<Axes: >
```



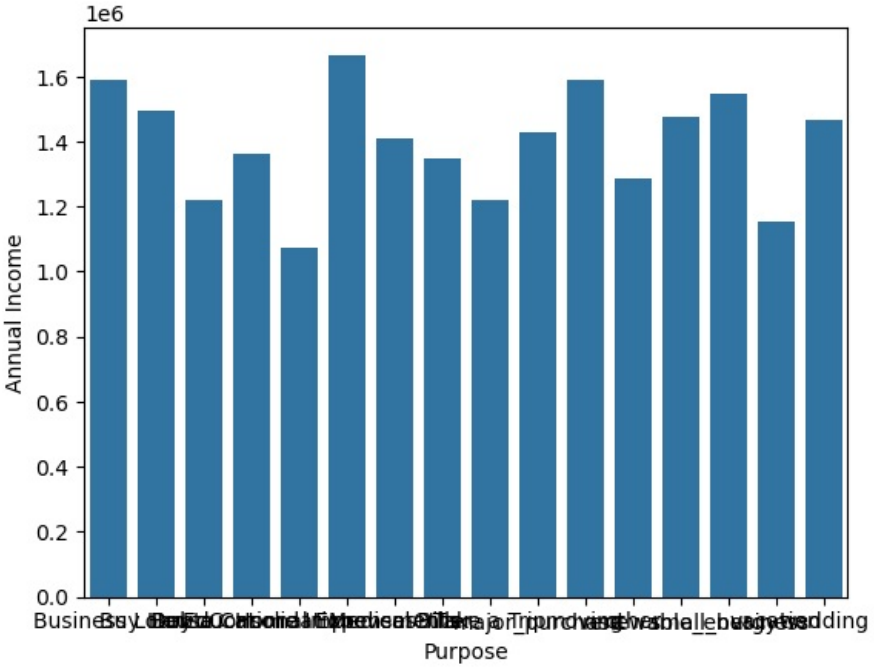
In [33]:
```
#4.What is the distribution of credit scores across different home ownership types?
A=df.groupby(by='Home Ownership')['Credit Score'].sum().reset_index()
sns.histplot(data=df,x='Home Ownership',bins=5,kde=True)
```

Out[33]:
```
<Axes: xlabel='Home Ownership', ylabel='Count'>
```



In [68]:
```
#5.How does the annual income differ for different purposes of loans?
X=df.groupby('Purpose')['Annual Income'].mean().reset_index()
sns.barplot(data=X,x='Purpose',y='Annual Income')
```

Out[68]:
```
<Axes: xlabel='Purpose', ylabel='Annual Income'>
```

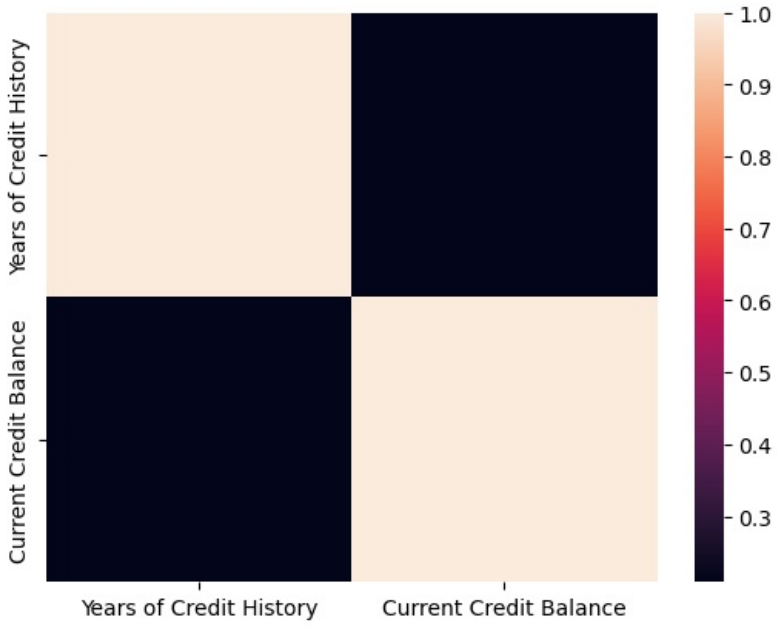


In [70]:
```
#6.What is the average monthly debt for each term (short-term vs. long-term)?
df.groupby(by='Term')['Monthly Debt'].mean()
```

Out[70]:
```
Term
Long Term     21585.664968
Short Term    17274.158442
Name: Monthly Debt, dtype: float64
```

In [37]:
```
#7.Is there a correlation between years of credit history and the current credit balance?
sns.heatmap(df[['Years of Credit History','Current Credit Balance']].corr())
```

Out[37]:
```
<Axes: >
```



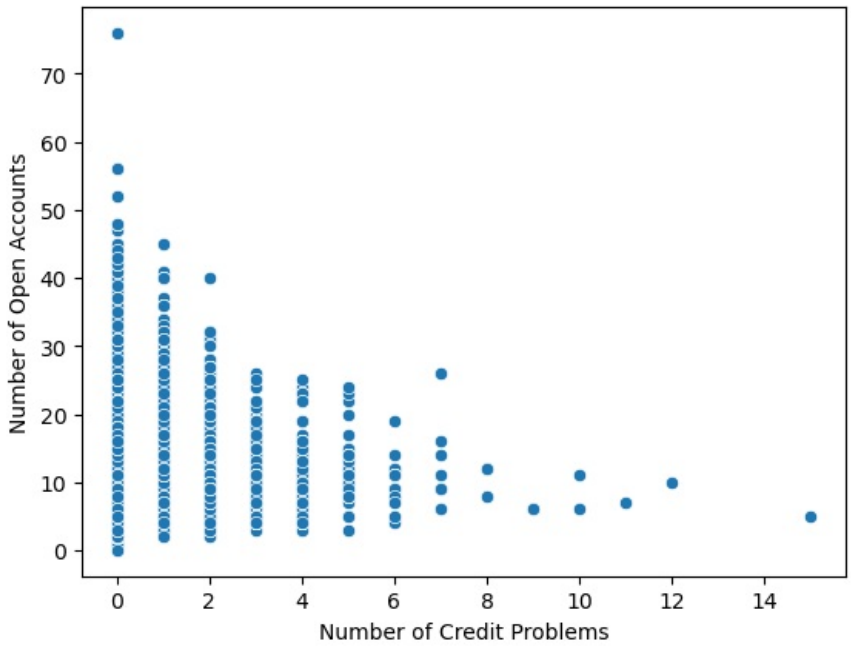
In [41]:
```
#8.How does the credit score vary with the years in the current job?
A=df.groupby('Years in current job')['Credit Score'].sum().reset_index()
A=A.sort_values(by='Credit Score',ascending=False)
A
```

Out[41]:

	Years in current job	Credit Score
1	10+ years	26787599.0
2	2 years	7867323.0
3	3 years	7133806.0
10	< 1 year	7055383.0
0	1 year	5811980.0
5	5 years	5777325.0
4	4 years	5102010.0
6	6 years	5044588.0
7	7 years	5044234.0
8	8 years	3956660.0
9	9 years	3425077.0

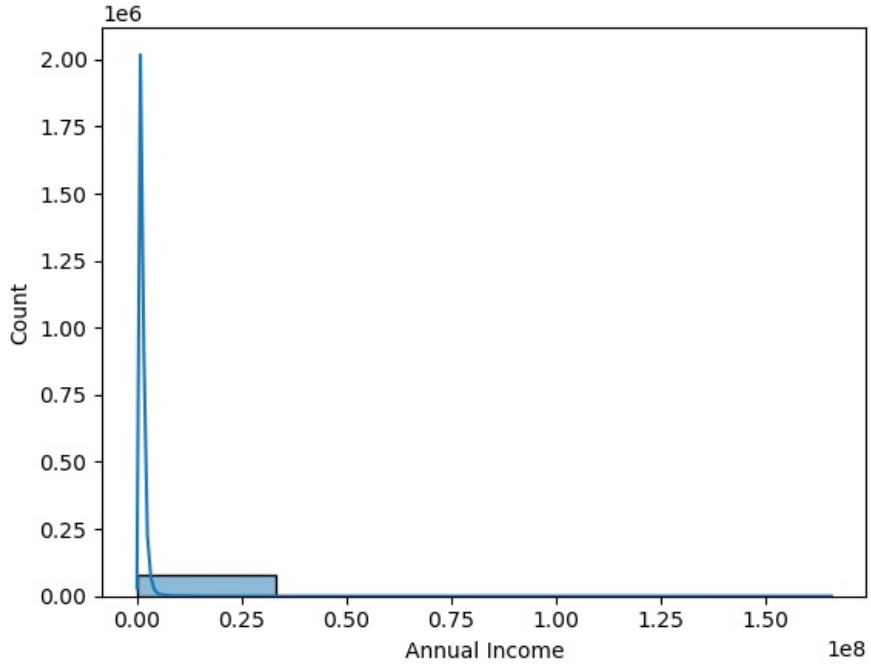
In [90]:
```
#9.What is the relationship between the number of credit problems and the number of open accounts?
sns.scatterplot(data=df,x='Number of Credit Problems',y='Number of Open Accounts')
```

Out[90]:
```
<Axes: xlabel='Number of Credit Problems', ylabel='Number of Open Accounts'>
```



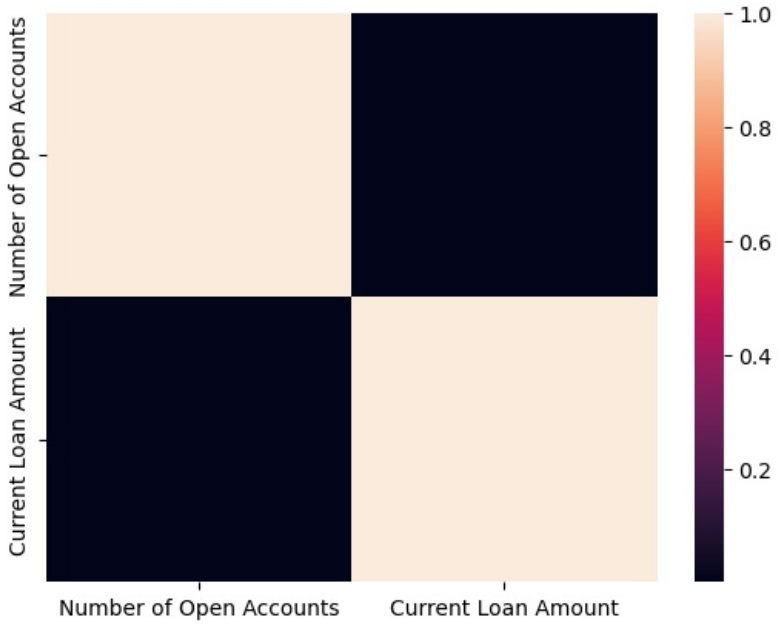
In [45]:
```
#10.What is the distribution of annual income across different loan statuses?
A=df.groupby(by='Loan Status')['Annual Income'].sum().reset_index()
A.sort_values(by='Annual Income', ascending=False)
sns.histplot(data=df,x='Annual Income',bins=5,kde=True)
```

Out[45]:
```
<Axes: xlabel='Annual Income', ylabel='Count'>
```



In [47]:
```
#11.Is there a correlation between the current loan amount and the number of open accounts?
sns.heatmap(df[['Number of Open Accounts','Current Loan Amount']].corr())
```

Out[47]: <Axes: >



In [49]:
```
#12.How does the monthly debt vary with the years of credit history?
A=df.groupby('Years of Credit History')['Monthly Debt'].sum().reset_index()
A=A.sort_values(by='Monthly Debt',ascending=False)
A
```

Out[49]:

	Years of Credit History	Monthly Debt
124	16.0	24586013.87
114	15.0	23805986.73
134	17.0	22881515.13
129	16.5	21236080.36
139	17.5	19879837.56
...	...	...
459	49.8	4551.45
496	57.0	4508.70
488	53.8	3323.29
1	3.7	3148.30
505	70.5	0.00

506 rows × 2 columns

In [104…
```
#13.What is the average annual income for each purpose of loan?
df.groupby(by='Purpose')['Annual Income'].mean()
```

Out[104…
```
Purpose
Business Loan           1.591963e+06
Buy House               1.495251e+06
Buy a Car               1.220353e+06
Debt Consolidation      1.362464e+06
Educational Expenses    1.071051e+06
Home Improvements       1.667437e+06
Medical Bills           1.409286e+06
Other                   1.348516e+06
Take a Trip             1.219899e+06
major_purchase          1.430753e+06
moving                  1.591375e+06
other                   1.286008e+06
renewable_energy        1.473740e+06
small_business          1.547456e+06
vacation                1.155290e+06
wedding                 1.464233e+06
Name: Annual Income, dtype: float64
```

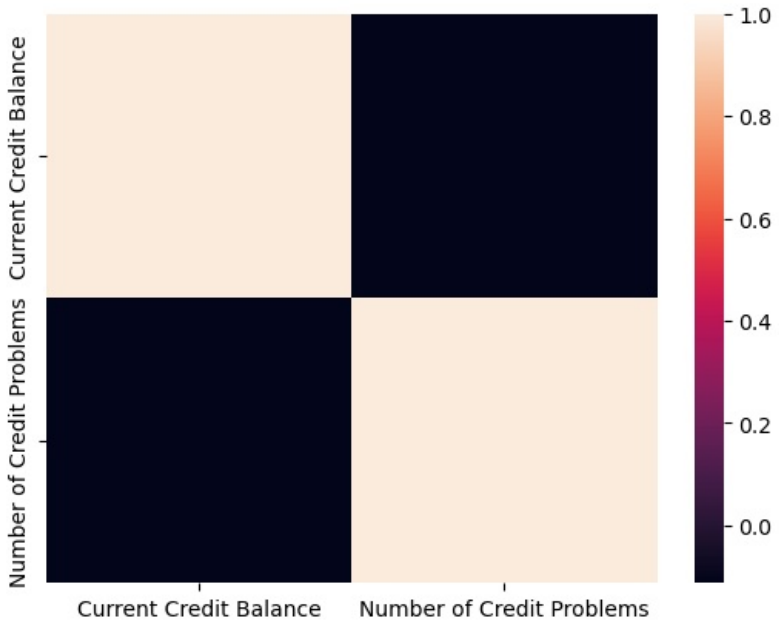
In [51]:
```
#14.How does the credit score vary with the number of credit problems?
A=df.groupby('Number of Credit Problems')['Credit Score'].sum().reset_index()
A=A.sort_values(by='Credit Score',ascending=False)
A
```

Out[51]:

	Number of Credit Problems	Credit Score
0	0.0	74966472.0
1	1.0	10405814.0
2	2.0	1158462.0
3	3.0	288137.0
4	4.0	116036.0
5	5.0	55208.0
6	6.0	17696.0
12	12.0	7200.0
7	7.0	5002.0
8	8.0	2858.0
9	9.0	1402.0
11	11.0	1392.0
13	15.0	746.0
10	10.0	744.0

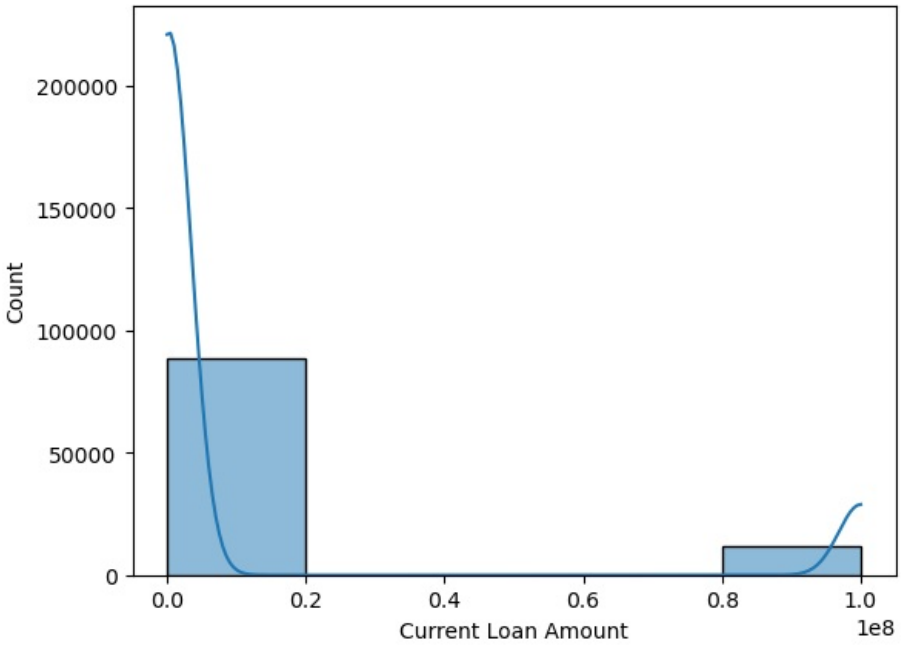
In [53]:
```
#15.Is there a correlation between the number of credit problems and the current credit balance?
sns.heatmap(df[['Current Credit Balance','Number of Credit Problems']].corr())
```

Out[53]: <Axes: >



In [55]:
```
#16.What is the distribution of current loan amounts across different home ownership types?
A=df.groupby(by='Home Ownership')['Current Loan Amount'].sum().reset_index()
A.sort_values(by='Current Loan Amount', ascending=False)
sns.histplot(data=df,x='Current Loan Amount',bins=5,kde=True)
```

Out[55]: <Axes: xlabel='Current Loan Amount', ylabel='Count'>



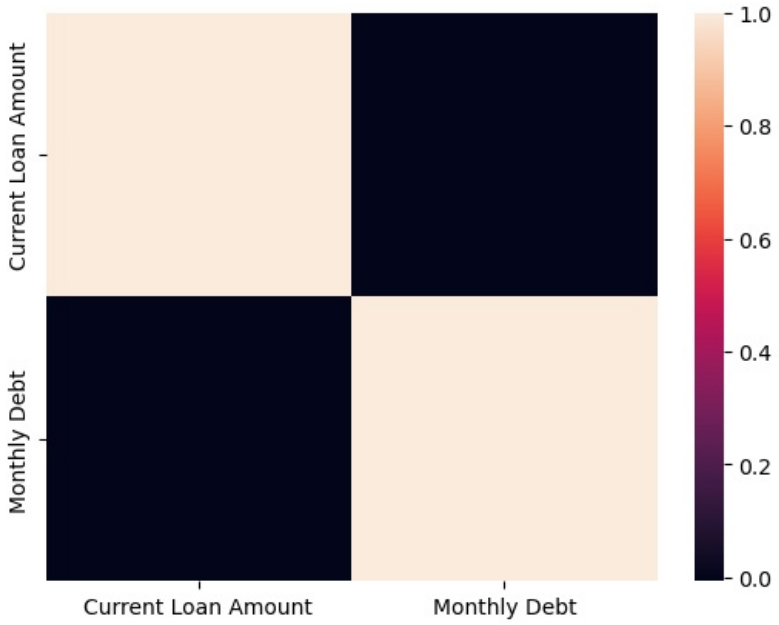
In [59]:
```
#17.How does the annual income vary with the years in the current job?
A=df.groupby('Years in current job')['Annual Income'].sum().reset_index()
A=A.sort_values(by='Annual Income',ascending=False)
A
```

Out[59]:

	Years in current job	Annual Income
1	10+ years	3.844998e+10
2	2 years	9.863079e+09
3	3 years	8.860041e+09
10	< 1 year	8.347987e+09
5	5 years	7.377869e+09
0	1 year	6.713135e+09
4	4 years	6.510239e+09
7	7 years	6.287609e+09
6	6 years	6.161153e+09
8	8 years	5.235659e+09
9	9 years	4.529392e+09

In [61]:
```
#18.Is there a correlation between the current loan amount and the monthly debt?
sns.heatmap(df[['Current Loan Amount','Monthly Debt']].corr())
```

Out[61]: <Axes: >



In [124…
```
#19.What is the average monthly debt for each home ownership type?
df.groupby(by='Home Ownership')['Monthly Debt'].mean()
```

Out[124…
```
Home Ownership
HaveMortgage     18541.776822
Home Mortgage    21072.059467
Own Home         17060.082101
Rent             15796.777284
Name: Monthly Debt, dtype: float64
```

In [63]:
```
#20.How does the credit score vary with the number of open accounts?
A=df.groupby('Number of Open Accounts')['Credit Score'].sum().reset_index()
A=A.sort_values(by='Credit Score',ascending=False)
A
```

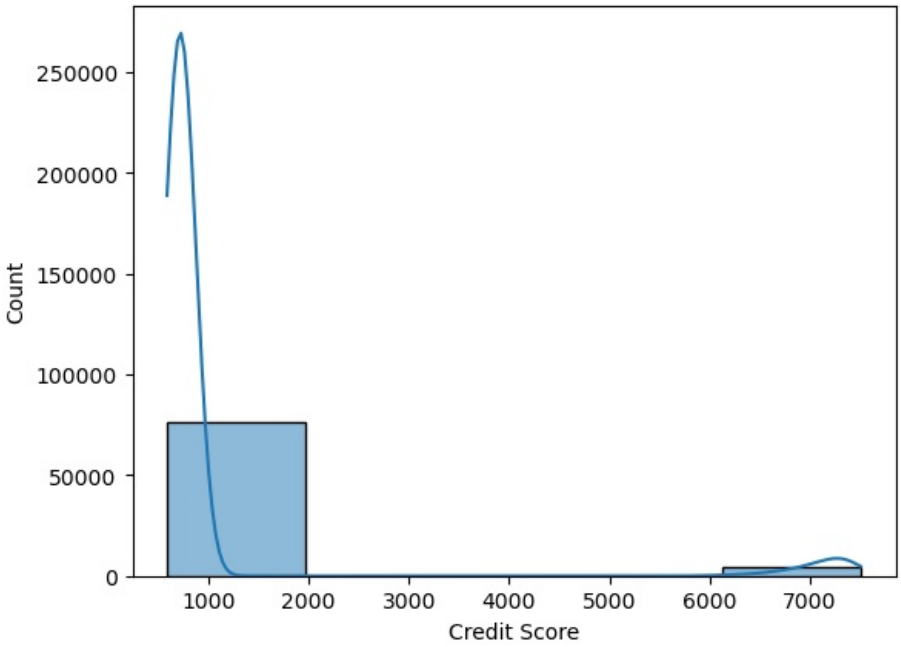
Out[63]:

	Number of Open Accounts	Credit Score
9	9.0	8061922.0
10	10.0	7822070.0
11	11.0	7726628.0
8	8.0	7574712.0
7	7.0	6920616.0
12	12.0	6528665.0
6	6.0	5864478.0
13	13.0	5340277.0
14	14.0	4474586.0
5	5.0	4002617.0
15	15.0	3900959.0
16	16.0	2958957.0
17	17.0	2495796.0
4	4.0	2459523.0
18	18.0	1888948.0
19	19.0	1703142.0
20	20.0	1222865.0
3	3.0	1153628.0
21	21.0	1057866.0
22	22.0	767222.0
23	23.0	576964.0
24	24.0	540062.0

25	25.0	423073.0
2	2.0	396356.0
27	27.0	230990.0
26	26.0	215888.0
28	28.0	163665.0
30	30.0	95123.0
29	29.0	90832.0
31	31.0	71210.0
33	33.0	57324.0
32	32.0	56363.0
34	34.0	48671.0
1	1.0	25590.0
37	37.0	17336.0
36	36.0	13429.0
38	38.0	13244.0
40	40.0	10786.0
35	35.0	10547.0
42	42.0	8954.0
50	76.0	7408.0
43	43.0	6043.0
41	41.0	4224.0
45	45.0	4144.0
39	39.0	3583.0
47	48.0	2816.0
44	44.0	2750.0
48	52.0	1453.0
49	56.0	1442.0
46	47.0	1422.0
0	0.0	0.0

In [65]:
```
#21.What is the distribution of credit scores across different loan statuses?
A=df.groupby(by='Loan Status')['Credit Score'].sum().reset_index()
A.sort_values(by='Credit Score', ascending=False)
sns.histplot(data=df,x='Credit Score',bins=5,kde=True)
```

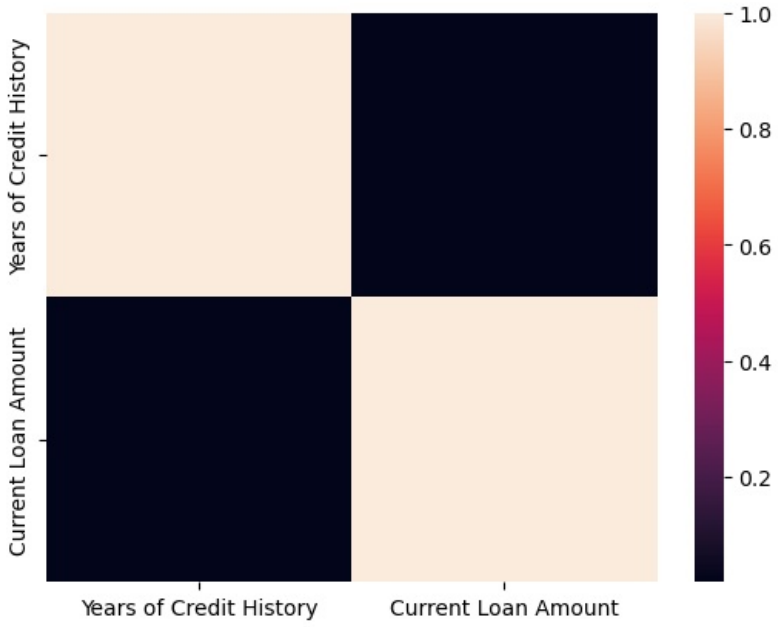
Out[65]: <Axes: xlabel='Credit Score', ylabel='Count'>



In [67]:
```
#22.Is there a correlation between the current loan amount and the years of credit history?
```

```
sns.heatmap(df[['Years of Credit History','Current Loan Amount']].corr())
```

Out[67]: <Axes: >



In [69]:
```
#23.How does the monthly debt vary with the number of credit problems?
A=df.groupby('Number of Credit Problems')['Monthly Debt'].sum().reset_index()
A=A.sort_values(by='Monthly Debt',ascending=False)
A
```

Out[69]:

	Number of Credit Problems	Monthly Debt
0	0.0	1.619387e+09
1	1.0	1.951776e+08
2	2.0	2.209683e+07
3	3.0	6.686827e+06
4	4.0	2.208436e+06
5	5.0	9.635580e+05
6	6.0	3.092047e+05
7	7.0	1.893491e+05
8	8.0	1.217585e+05
9	9.0	2.848936e+04
12	12.0	2.382961e+04
10	10.0	1.804069e+04
13	15.0	1.551331e+04
11	11.0	1.525624e+04

In [138…
```
#24.What is the average current loan amount for each purpose of loan?
df.groupby(by='Purpose')['Current Loan Amount'].mean()
```

Out[138…
```
Purpose
Business Loan           7.837789e+06
Buy House               1.106036e+07
Buy a Car               1.414682e+07
Debt Consolidation      1.180683e+07
Educational Expenses    1.428536e+07
Home Improvements       1.252117e+07
Medical Bills           1.179902e+07
Other                   1.253627e+07
Take a Trip             1.111667e+07
major_purchase          1.298956e+07
moving                  1.350072e+07
other                   1.059243e+07
renewable_energy        1.975534e+05
small_business          9.484232e+06
vacation                7.058680e+06
wedding                 1.760299e+07
Name: Current Loan Amount, dtype: float64
```

In [71]:
```
#25.How does the credit score vary with the current credit balance?
```

```
A=df.groupby('Current Credit Balance')['Credit Score'].sum().reset_index()
A=A.sort_values(by='Credit Score',ascending=False)
A
```

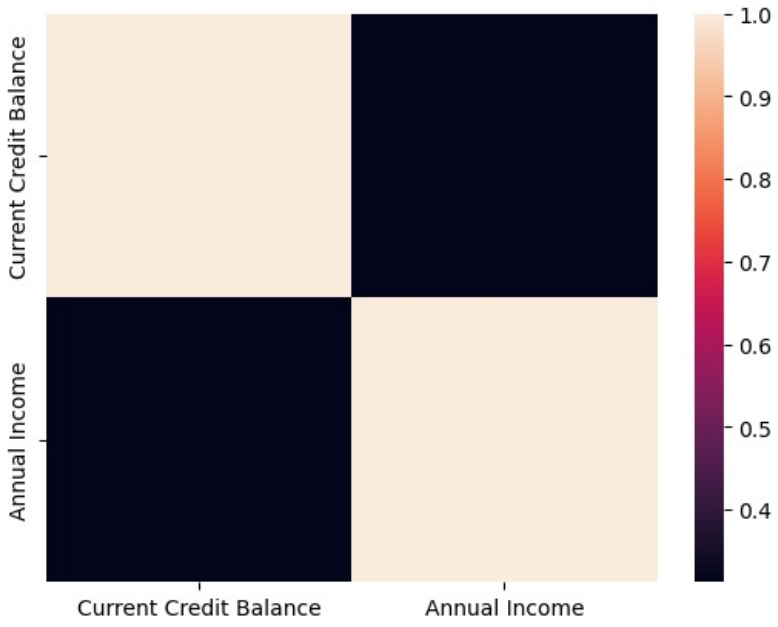
Out[71]:

	Current Credit Balance	Credit Score
0	0.0	458478.0
5919	118009.0	30550.0
8786	173261.0	28178.0
8516	167998.0	25459.0
12995	256006.0	24250.0
...	...	...
9895	194712.0	0.0
17784	358150.0	0.0
8414	166041.0	0.0
30867	1031111.0	0.0
21214	443859.0	0.0

32730 rows × 2 columns

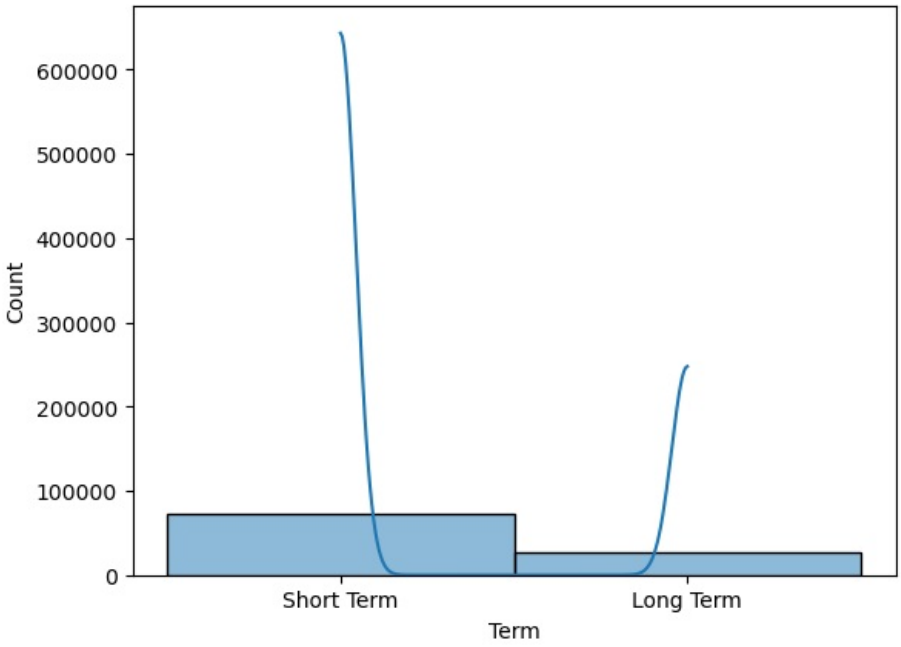
In [73]:
```
#26.Is there a correlation between the annual income and the current credit balance?
sns.heatmap(df[['Current Credit Balance','Annual Income']].corr())
```

Out[73]: <Axes: >



In [75]:
```
#27.What is the distribution of annual income across different terms (short-term vs. long-term)?
A=df.groupby(by='Annual Income')['Term'].sum().reset_index()
A.sort_values(by='Term', ascending=False)
sns.histplot(data=df,x='Term',bins=5,kde=True)
```

Out[75]: <Axes: xlabel='Term', ylabel='Count'>



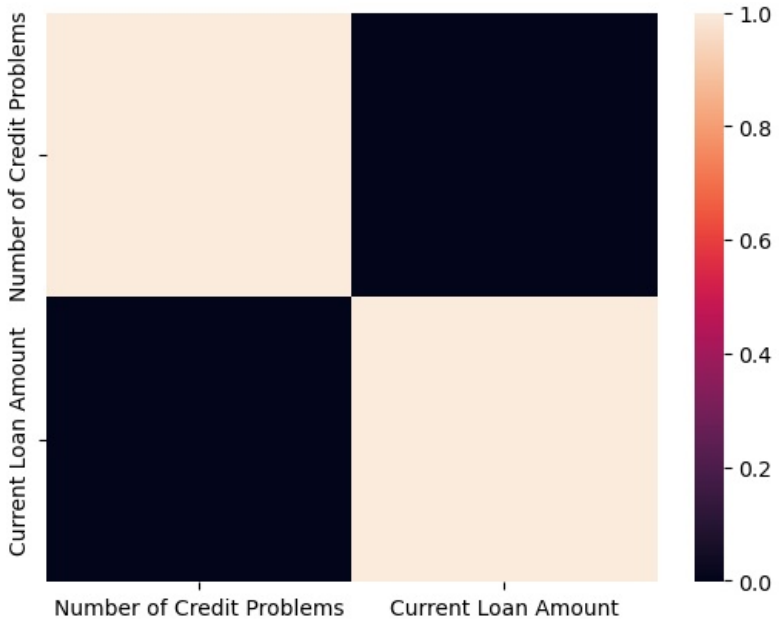
In [77]:
```
#28.How does the credit score vary with the number of credit problems?
A=df.groupby('Number of Credit Problems')['Credit Score'].sum().reset_index()
A=A.sort_values(by='Credit Score',ascending=False)
A
```

Out[77]:

	Number of Credit Problems	Credit Score
0	0.0	74966472.0
1	1.0	10405814.0
2	2.0	1158462.0
3	3.0	288137.0
4	4.0	116036.0
5	5.0	55208.0
6	6.0	17696.0
12	12.0	7200.0
7	7.0	5002.0
8	8.0	2858.0
9	9.0	1402.0
11	11.0	1392.0
13	15.0	746.0
10	10.0	744.0

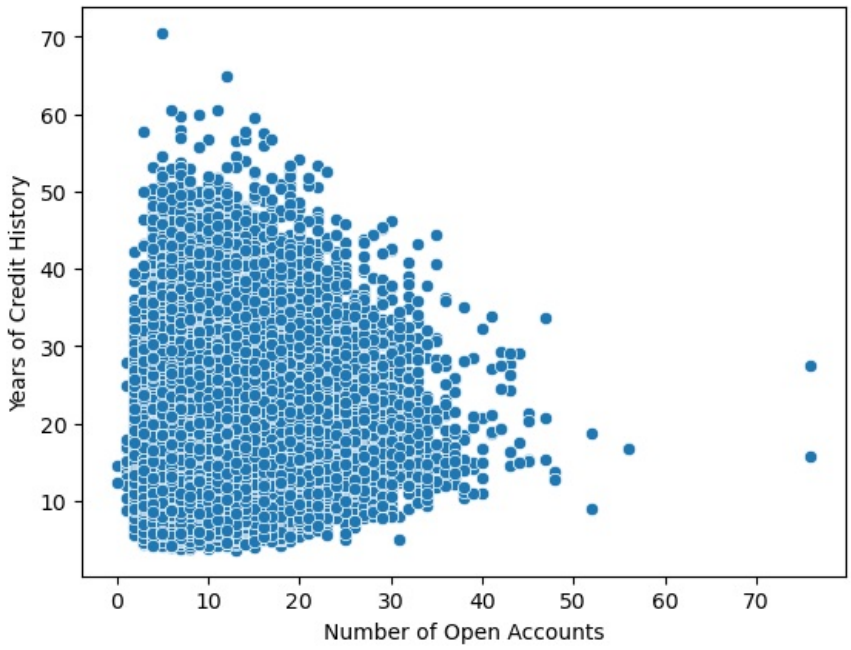
In [79]:
```
#29.Is there a correlation between the current loan amount and the number of credit problems?
sns.heatmap(df[['Number of Credit Problems','Current Loan Amount']].corr())
```

Out[79]: <Axes: >



In [150…
```
#30.What is the relationship between the number of open accounts and the years of credit history?
sns.scatterplot(data=df,x='Number of Open Accounts',y='Years of Credit History')
```

Out[150… <Axes: xlabel='Number of Open Accounts', ylabel='Years of Credit History'>



In [ ]: