credit-eda-assignment

August 22, 2023

0.1 import the libraries

```
[1]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
```

0.2 read the csv files

```
[2]: inp0 = pd.read_csv("application_data.csv")
     inp0
```

```
[2]:         SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
     0           100002       1         Cash loans           M            N
     1           100003       0         Cash loans           F            N
     2           100004       0    Revolving loans           M            Y
     3           100006       0         Cash loans           F            N
     4           100007       0         Cash loans           M            N
     ...            ...     ...                ...         ...          ...
     307506      456251       0         Cash loans           M            N
     307507      456252       0         Cash loans           F            N
     307508      456253       0         Cash loans           F            N
     307509      456254       1         Cash loans           F            N
     307510      456255       0         Cash loans           F            N

            FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
     0                    Y             0          202500.0    406597.5
     1                    N             0          270000.0   1293502.5
     2                    Y             0           67500.0    135000.0
     3                    Y             0          135000.0    312682.5
     4                    Y             0          121500.0    513000.0
     ...                ...           ...               ...         ...
     307506               N             0          157500.0    254700.0
     307507               Y             0           72000.0    269550.0
     307508               Y             0          153000.0    677664.0
     307509               Y             0          171000.0    370107.0
     307510               N             0          157500.0    675000.0
```

```
        AMT_ANNUITY  …  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20  \
0           24700.5  …                0                0                0
1           35698.5  …                0                0                0
2            6750.0  …                0                0                0
3           29686.5  …                0                0                0
4           21865.5  …                0                0                0
…               …  …              …              …              …
307506      27558.0  …                0                0                0
307507      12001.5  …                0                0                0
307508      29979.0  …                0                0                0
307509      20205.0  …                0                0                0
307510      49117.5  …                0                0                0

        FLAG_DOCUMENT_21 AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
0                      0                        0.0                       0.0
1                      0                        0.0                       0.0
2                      0                        0.0                       0.0
3                      0                        NaN                       NaN
4                      0                        0.0                       0.0
…                    …                        …                       …
307506                 0                        NaN                       NaN
307507                 0                        NaN                       NaN
307508                 0                        1.0                       0.0
307509                 0                        0.0                       0.0
307510                 0                        0.0                       0.0

        AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                             0.0                        0.0
1                             0.0                        0.0
2                             0.0                        0.0
3                             NaN                        NaN
4                             0.0                        0.0
…                           …                        …
307506                        NaN                        NaN
307507                        NaN                        NaN
307508                        0.0                        1.0
307509                        0.0                        0.0
307510                        0.0                        2.0

        AMT_REQ_CREDIT_BUREAU_QRT AMT_REQ_CREDIT_BUREAU_YEAR
0                             0.0                        1.0
1                             0.0                        0.0
2                             0.0                        0.0
3                             NaN                        NaN
4                             0.0                        0.0
…                           …                        …
307506                        NaN                        NaN
```

```
307507                      NaN                          NaN
307508                      0.0                          1.0
307509                      0.0                          0.0
307510                      0.0                          1.0

[307511 rows x 122 columns]
```

[3]: `inp0.shape`

[3]: (307511, 122)

[4]: `inp0.dtypes`

[4]:
```
SK_ID_CURR                    int64
TARGET                        int64
NAME_CONTRACT_TYPE           object
CODE_GENDER                  object
FLAG_OWN_CAR                 object
                              ...
AMT_REQ_CREDIT_BUREAU_DAY    float64
AMT_REQ_CREDIT_BUREAU_WEEK   float64
AMT_REQ_CREDIT_BUREAU_MON    float64
AMT_REQ_CREDIT_BUREAU_QRT    float64
AMT_REQ_CREDIT_BUREAU_YEAR   float64
Length: 122, dtype: object
```

[5]: `inp0.isna().sum()`

[5]:
```
SK_ID_CURR                       0
TARGET                           0
NAME_CONTRACT_TYPE               0
CODE_GENDER                      0
FLAG_OWN_CAR                     0
                              ...
AMT_REQ_CREDIT_BUREAU_DAY    41519
AMT_REQ_CREDIT_BUREAU_WEEK   41519
AMT_REQ_CREDIT_BUREAU_MON    41519
AMT_REQ_CREDIT_BUREAU_QRT    41519
AMT_REQ_CREDIT_BUREAU_YEAR   41519
Length: 122, dtype: int64
```

[6]:
```
inp1 = inp0.dropna()
inp1
```

[6]:
```
     SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
71       100083       0         Cash loans           M            Y
124      100145       0         Cash loans           F            Y
```

			Cash loans		
152	100179	0	Cash loans	F	Y
161	100190	0	Cash loans	M	Y
255	100295	1	Cash loans	M	Y
...	...	...		...	...
307358	456083	0	Cash loans	F	Y
307359	456084	0	Cash loans	F	Y
307407	456140	1	Cash loans	F	Y
307456	456195	0	Cash loans	F	Y
307482	456226	0	Cash loans	F	Y

	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	\
71	Y	0	103500.0	573628.5	
124	Y	1	202500.0	260725.5	
152	N	0	202500.0	675000.0	
161	N	0	162000.0	263686.5	
255	N	1	225000.0	1019205.0	
...	...	...	...	...	
307358	Y	2	112500.0	361462.5	
307359	Y	1	99000.0	675000.0	
307407	Y	1	261000.0	711454.5	
307456	Y	0	94500.0	270000.0	
307482	Y	0	225000.0	500566.5	

	AMT_ANNUITY	...	FLAG_DOCUMENT_18	FLAG_DOCUMENT_19	FLAG_DOCUMENT_20	\
71	24435.0	...	0	0	0	
124	16789.5	...	0	0	0	
152	53329.5	...	0	0	0	
161	24781.5	...	0	0	0	
255	31032.0	...	1	0	0	
...	...	...	...	...	...	
307358	16051.5	...	0	0	0	
307359	21906.0	...	0	0	0	
307407	47673.0	...	0	0	0	
307456	15075.0	...	0	0	0	
307482	34969.5	...	0	0	0	

	FLAG_DOCUMENT_21	AMT_REQ_CREDIT_BUREAU_HOUR	AMT_REQ_CREDIT_BUREAU_DAY	\
71	0	0.0	0.0	
124	0	0.0	0.0	
152	0	0.0	0.0	
161	0	0.0	0.0	
255	0	0.0	0.0	
...	...	...	...	
307358	0	0.0	0.0	
307359	0	0.0	0.0	
307407	0	0.0	0.0	
307456	0	0.0	0.0	

```
307482                            0                      0.0                          0.0

            AMT_REQ_CREDIT_BUREAU_WEEK   AMT_REQ_CREDIT_BUREAU_MON  \
71                                 0.0                         0.0
124                                0.0                         0.0
152                                0.0                         1.0
161                                0.0                         0.0
255                                0.0                         6.0
...                                ...                         ...
307358                             0.0                         2.0
307359                             0.0                         0.0
307407                             0.0                         1.0
307456                             0.0                         0.0
307482                             0.0                         0.0

            AMT_REQ_CREDIT_BUREAU_QRT   AMT_REQ_CREDIT_BUREAU_YEAR
71                                0.0                          3.0
124                               0.0                          3.0
152                               0.0                          4.0
161                               0.0                          3.0
255                               0.0                          1.0
...                               ...                          ...
307358                            0.0                          0.0
307359                            0.0                          0.0
307407                            0.0                          1.0
307456                            0.0                          3.0
307482                            0.0                          1.0

[8602 rows x 122 columns]
```

[7]: `inp1.shape`

[7]: (8602, 122)

[8]: `inp1.isnull().sum()`

[8]:
```
SK_ID_CURR                    0
TARGET                        0
NAME_CONTRACT_TYPE            0
CODE_GENDER                   0
FLAG_OWN_CAR                  0
                             ..
AMT_REQ_CREDIT_BUREAU_DAY     0
AMT_REQ_CREDIT_BUREAU_WEEK    0
AMT_REQ_CREDIT_BUREAU_MON     0
AMT_REQ_CREDIT_BUREAU_QRT     0
AMT_REQ_CREDIT_BUREAU_YEAR    0
```

```
Length: 122, dtype: int64
```

[9]: `inp1.dtypes`

[9]:
```
SK_ID_CURR                  int64
TARGET                      int64
NAME_CONTRACT_TYPE         object
CODE_GENDER                object
FLAG_OWN_CAR               object
                           ...
AMT_REQ_CREDIT_BUREAU_DAY    float64
AMT_REQ_CREDIT_BUREAU_WEEK   float64
AMT_REQ_CREDIT_BUREAU_MON    float64
AMT_REQ_CREDIT_BUREAU_QRT    float64
AMT_REQ_CREDIT_BUREAU_YEAR   float64
Length: 122, dtype: object
```

[10]: `inp1`

[10]:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR \
71	100083	0	Cash loans	M	Y
124	100145	0	Cash loans	F	Y
152	100179	0	Cash loans	F	Y
161	100190	0	Cash loans	M	Y
255	100295	1	Cash loans	M	Y
...	...	...	...	...	...
307358	456083	0	Cash loans	F	Y
307359	456084	0	Cash loans	F	Y
307407	456140	1	Cash loans	F	Y
307456	456195	0	Cash loans	F	Y
307482	456226	0	Cash loans	F	Y

	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT \
71	Y	0	103500.0	573628.5
124	Y	1	202500.0	260725.5
152	N	0	202500.0	675000.0
161	N	0	162000.0	263686.5
255	N	1	225000.0	1019205.0
...	...	...	...	...
307358	Y	2	112500.0	361462.5
307359	Y	1	99000.0	675000.0
307407	Y	1	261000.0	711454.5
307456	Y	0	94500.0	270000.0
307482	Y	0	225000.0	500566.5

	AMT_ANNUITY	...	FLAG_DOCUMENT_18	FLAG_DOCUMENT_19	FLAG_DOCUMENT_20 \
71	24435.0	...	0	0	0

124	16789.5	…	0	0	0
152	53329.5	…	0	0	0
161	24781.5	…	0	0	0
255	31032.0	…	1	0	0
…	…	…	…	…	…
307358	16051.5	…	0	0	0
307359	21906.0	…	0	0	0
307407	47673.0	…	0	0	0
307456	15075.0	…	0	0	0
307482	34969.5	…	0	0	0

	FLAG_DOCUMENT_21	AMT_REQ_CREDIT_BUREAU_HOUR	AMT_REQ_CREDIT_BUREAU_DAY \
71	0	0.0	0.0
124	0	0.0	0.0
152	0	0.0	0.0
161	0	0.0	0.0
255	0	0.0	0.0
…	…	…	…
307358	0	0.0	0.0
307359	0	0.0	0.0
307407	0	0.0	0.0
307456	0	0.0	0.0
307482	0	0.0	0.0

	AMT_REQ_CREDIT_BUREAU_WEEK	AMT_REQ_CREDIT_BUREAU_MON \
71	0.0	0.0
124	0.0	0.0
152	0.0	1.0
161	0.0	0.0
255	0.0	6.0
…	…	…
307358	0.0	2.0
307359	0.0	0.0
307407	0.0	1.0
307456	0.0	0.0
307482	0.0	0.0

	AMT_REQ_CREDIT_BUREAU_QRT	AMT_REQ_CREDIT_BUREAU_YEAR
71	0.0	3.0
124	0.0	3.0
152	0.0	4.0
161	0.0	3.0
255	0.0	1.0
…	…	…
307358	0.0	0.0
307359	0.0	0.0
307407	0.0	1.0

```
307456                      0.0                        3.0
307482                      0.0                        1.0

[8602 rows x 122 columns]
```

[11]: `inp1.describe`

[11]: <bound method NDFrame.describe of   SK_ID_CURR   TARGET NAME_CONTRACT_TYPE
CODE_GENDER FLAG_OWN_CAR \
71   100083   0   Cash loans   M   Y
124   100145   0   Cash loans   F   Y
152   100179   0   Cash loans   F   Y
161   100190   0   Cash loans   M   Y
255   100295   1   Cash loans   M   Y

...   .
307358   456083   0   Cash loans   F   Y
307359   456084   0   Cash loans   F   Y
307407   456140   1   Cash loans   F   Y
307456   456195   0   Cash loans   F   Y
307482   456226   0   Cash loans   F   Y

          FLAG_OWN_REALTY   CNT_CHILDREN   AMT_INCOME_TOTAL   AMT_CREDIT \
71   Y   0   103500.0   573628.5
124   Y   1   202500.0   260725.5
152   N   0   202500.0   675000.0
161   N   0   162000.0   263686.5
255   N   1   225000.0   1019205.0

...   .
307358   Y   2   112500.0   361462.5
307359   Y   1   99000.0   675000.0
307407   Y   1   261000.0   711454.5
307456   Y   0   94500.0   270000.0
307482   Y   0   225000.0   500566.5

          AMT_ANNUITY   ... FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20 \
71   24435.0   ...   0   0   0
124   16789.5   ...   0   0   0
152   53329.5   ...   0   0   0
161   24781.5   ...   0   0   0
255   31032.0   ...   1   0   0

...   .
307358   16051.5   ...   0   0   0
307359   21906.0   ...   0   0   0
307407   47673.0   ...   0   0   0
307456   15075.0   ...   0   0   0
307482   34969.5   ...   0   0   0

```
       FLAG_DOCUMENT_21  AMT_REQ_CREDIT_BUREAU_HOUR  AMT_REQ_CREDIT_BUREAU_DAY  \
71                    0                        0.0                        0.0
124                   0                        0.0                        0.0
152                   0                        0.0                        0.0
161                   0                        0.0                        0.0
255                   0                        0.0                        0.0
...                 ...                        ...                        ...
307358                0                        0.0                        0.0
307359                0                        0.0                        0.0
307407                0                        0.0                        0.0
307456                0                        0.0                        0.0
307482                0                        0.0                        0.0

       AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
71                            0.0                        0.0
124                           0.0                        0.0
152                           0.0                        1.0
161                           0.0                        0.0
255                           0.0                        6.0
...                           ...                        ...
307358                        0.0                        2.0
307359                        0.0                        0.0
307407                        0.0                        1.0
307456                        0.0                        0.0
307482                        0.0                        0.0

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
71                           0.0                        3.0
124                          0.0                        3.0
152                          0.0                        4.0
161                          0.0                        3.0
255                          0.0                        1.0
...                          ...                        ...
307358                       0.0                        0.0
307359                       0.0                        0.0
307407                       0.0                        1.0
307456                       0.0                        3.0
307482                       0.0                        1.0

[8602 rows x 122 columns]>
```

[29]: `inp1.columns`

[29]: Index(['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER',
        'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN', 'AMT_INCOME_TOTAL',
        'AMT_CREDIT', 'AMT_ANNUITY',
        …

```
          'FLAG_DOCUMENT_18', 'FLAG_DOCUMENT_19', 'FLAG_DOCUMENT_20',
          'FLAG_DOCUMENT_21', 'AMT_REQ_CREDIT_BUREAU_HOUR',
          'AMT_REQ_CREDIT_BUREAU_DAY', 'AMT_REQ_CREDIT_BUREAU_WEEK',
          'AMT_REQ_CREDIT_BUREAU_MON', 'AMT_REQ_CREDIT_BUREAU_QRT',
          'AMT_REQ_CREDIT_BUREAU_YEAR'],
        dtype='object', length=122)
```

[41]:
```python
plt.figure(figsize = (5,4))
sns.barplot(data = inp1, x = 'CODE_GENDER', y = 'AMT_INCOME_TOTAL')
plt.xlabel("GENDER")
plt.ylabel("Amount Income Total")
plt.show()
```
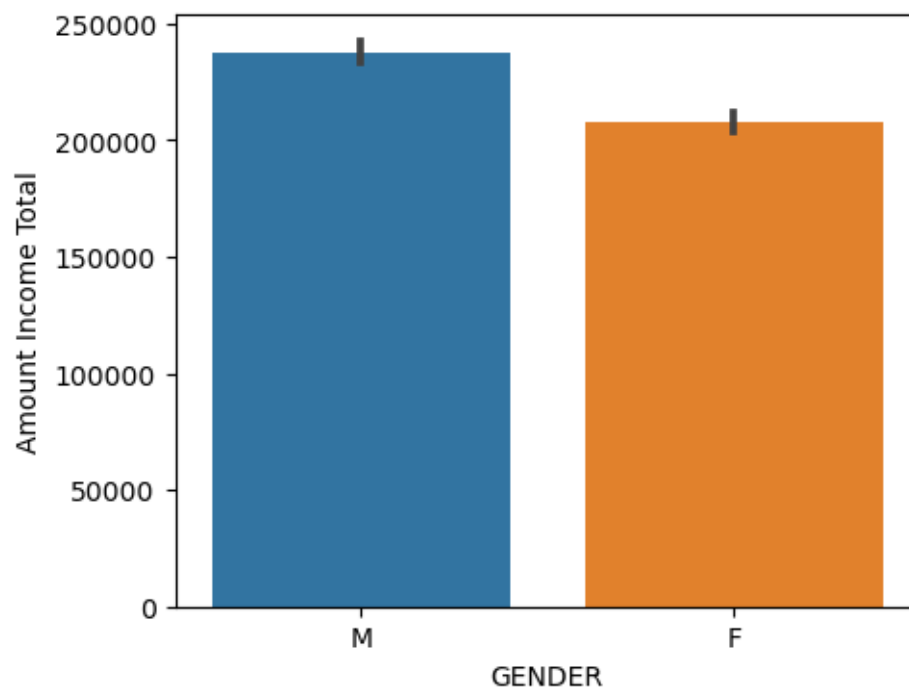

[12]:
```python
plt.figure(figsize = (5,4))
plt.scatter(inp1.AMT_INCOME_TOTAL, inp1.AMT_CREDIT)
plt.xlabel("Amount Income")
plt.ylabel("Amount Credit")
plt.grid()
plt.show()
```
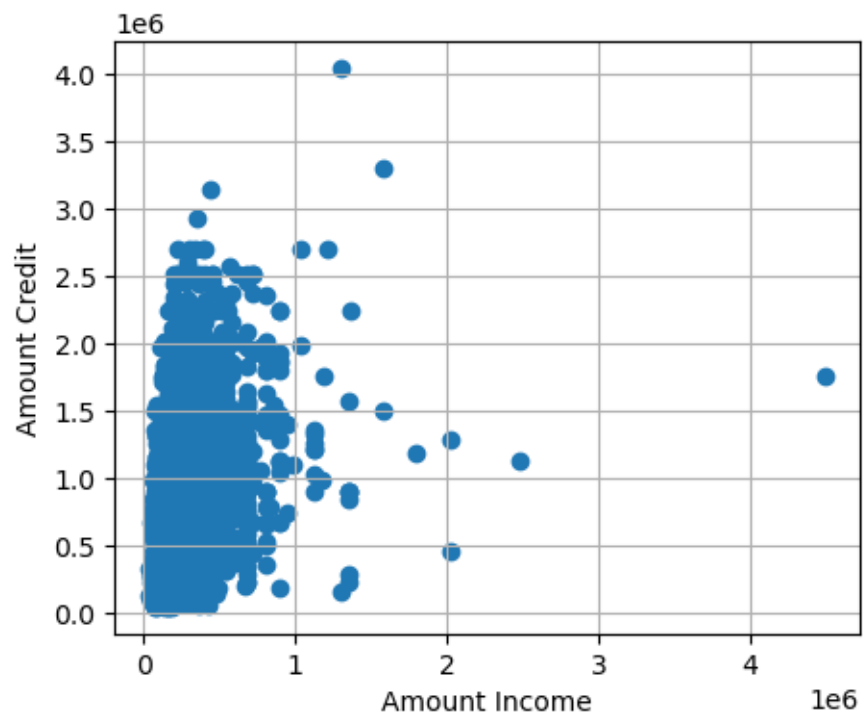
```
[13]: plt.figure(figsize = (5,4))
      sns.barplot(data = inp1, x = 'CODE_GENDER', y = 'AMT_ANNUITY',
                  hue = 'TARGET')
      plt.show()
```
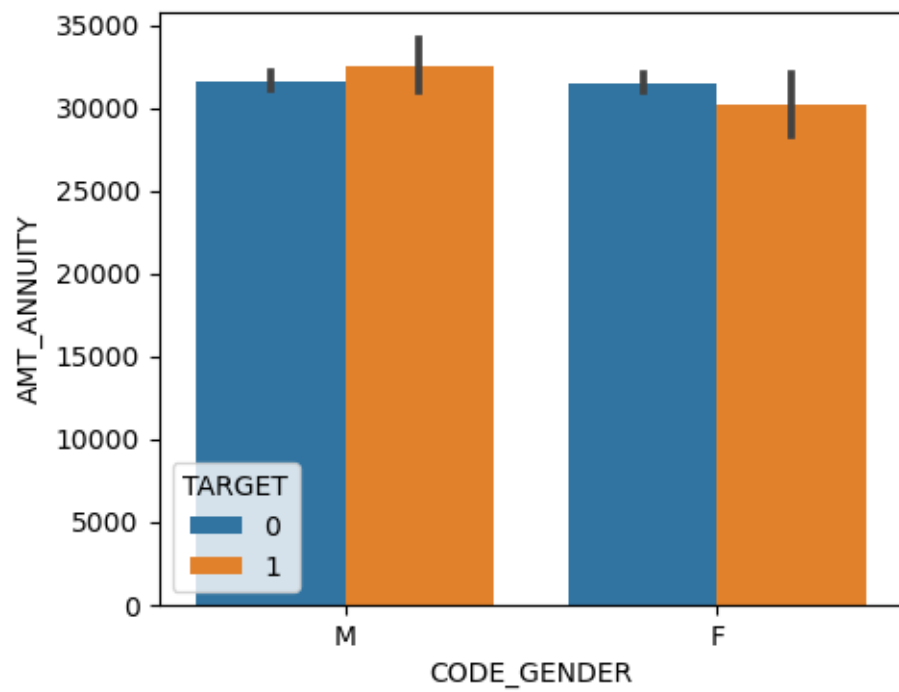
```
[14]: plt.figure(figsize = (5,5))
      sns.barplot(data = inp1, x = 'CODE_GENDER', y = 'AMT_ANNUITY',
                  hue = 'FLAG_OWN_REALTY')
      plt.show()
```
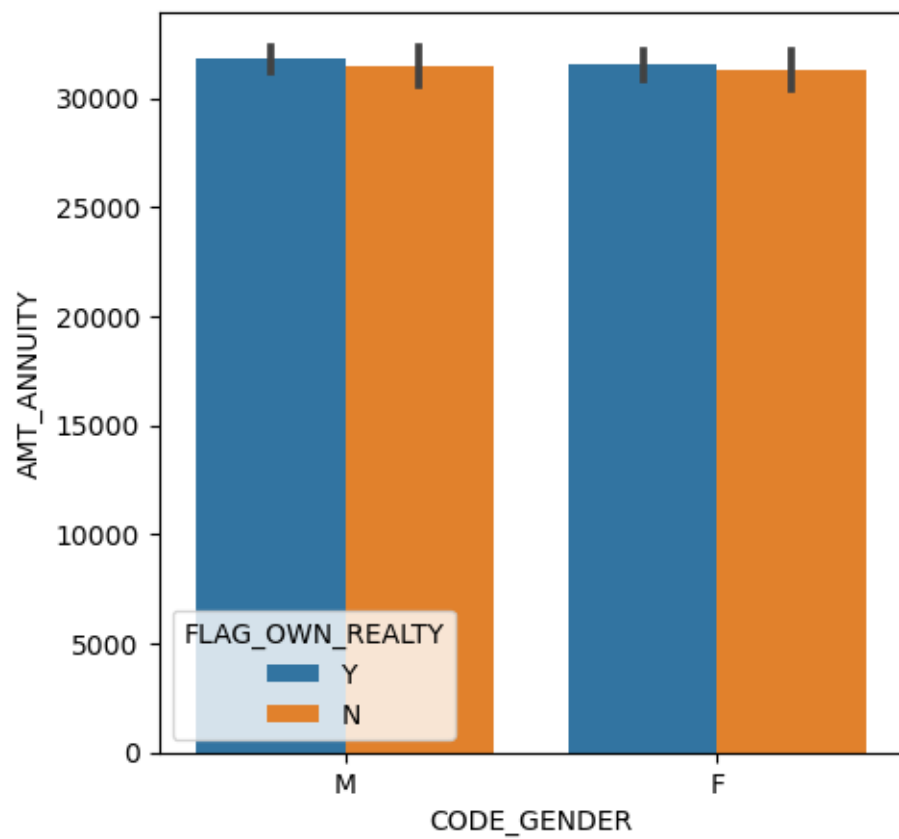
```
[81]: inp1.NAME_CONTRACT_TYPE.value_counts().plot.barh()
```

```
[81]: <Axes: >
```
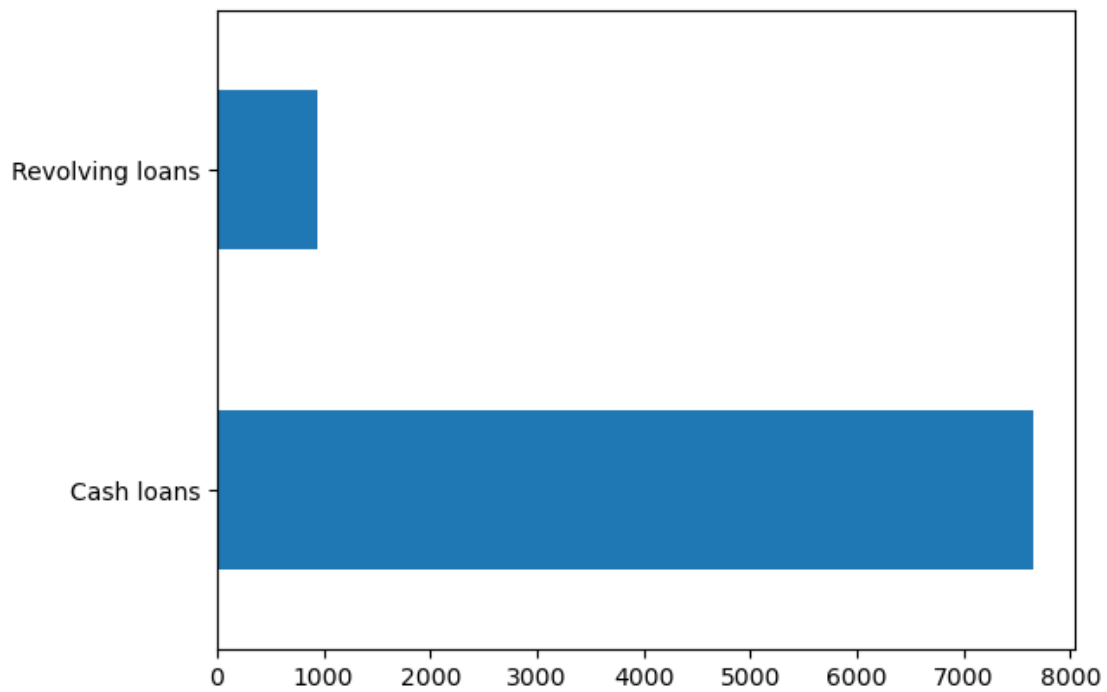
```
[56]: sns.boxplot(data = inp1, x = 'AMT_ANNUITY')
      plt.show()
```
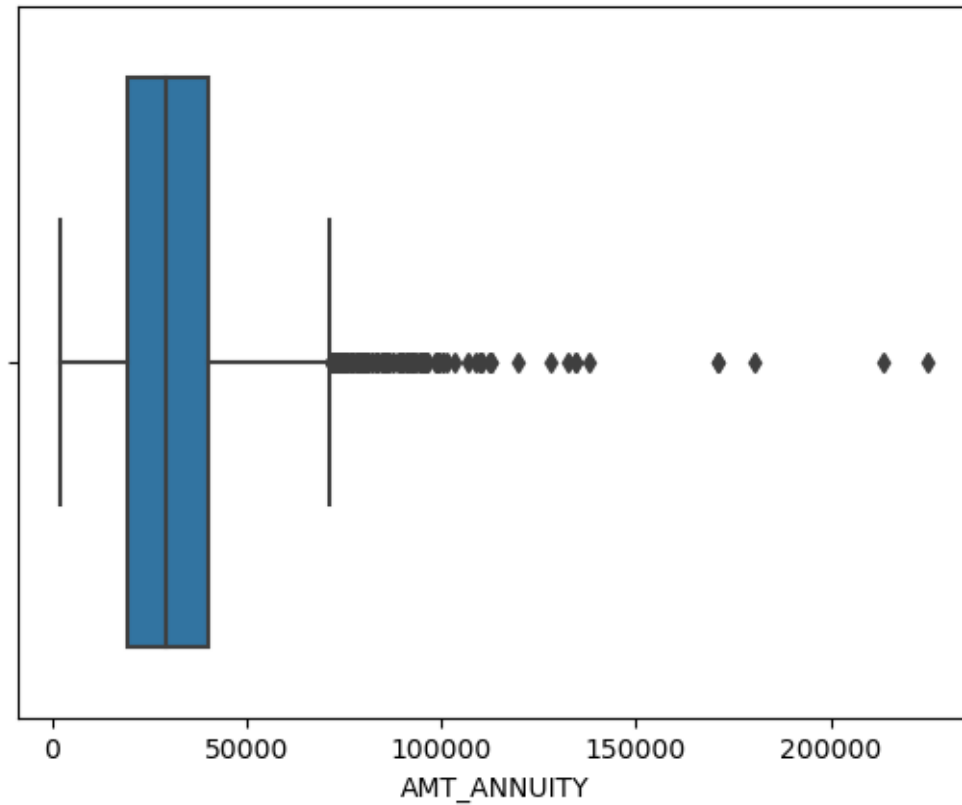
```
[15]:  res = pd.pivot_table(data = inp1, columns= "NAME_CONTRACT_TYPE",
                    values = 'AMT_INCOME_TOTAL',
                    index = 'CODE_GENDER')
       res
```

```
[15]:  NAME_CONTRACT_TYPE      Cash loans   Revolving loans
       CODE_GENDER
       F                    210113.183777     188669.616052
       M                    238121.004231     231755.360707
```

```
[16]:  plt.figure(figsize=(5,3))
       sns.heatmap(res, cmap = "Reds", annot = True)
       plt.show()
```
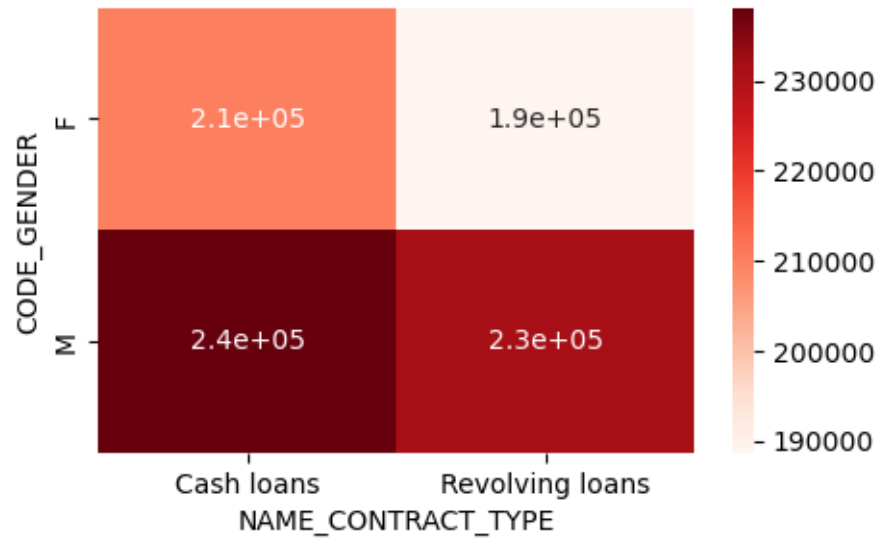
```
[17]: res = pd.pivot_table(data = inp1, columns= "NAME_CONTRACT_TYPE",
                  values = 'AMT_CREDIT',
                  index = 'CODE_GENDER')
      res
```

```
[17]: NAME_CONTRACT_TYPE      Cash loans   Revolving loans
      CODE_GENDER
      F                     758523.749601     369517.353579
      M                     723141.546923     371600.831601
```

```
[18]: plt.figure(figsize=(5,3))
      sns.heatmap(res, cmap = "Greens", annot = True)
      plt.show()
```
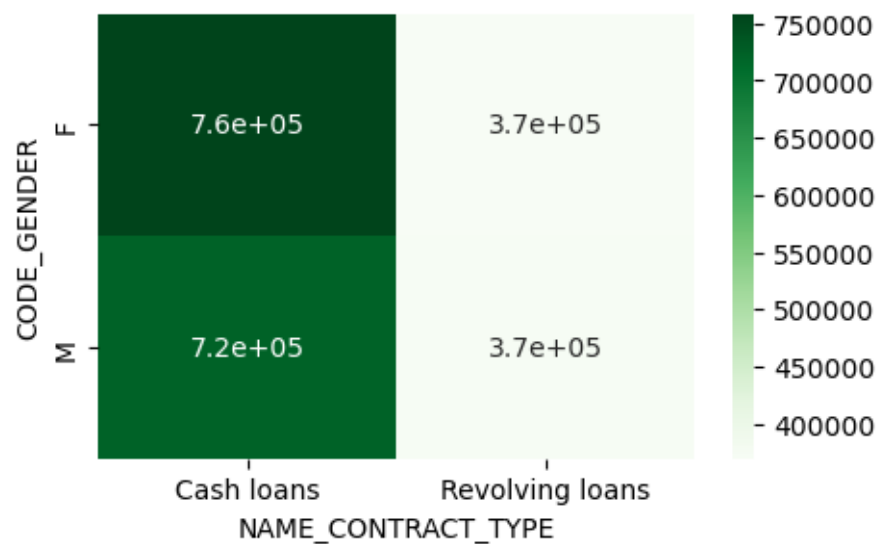
```
[19]: inp2 = pd.read_csv("previous_application.csv")
      inp2
```

[19]:

	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY
0	2030495	271877	Consumer loans	1730.430
1	2802425	108129	Cash loans	25188.615
2	2523466	122040	Cash loans	15060.735
3	2819243	176158	Cash loans	47041.335
4	1784265	202054	Cash loans	31924.395
...	...	...	...	...
1670209	2300464	352015	Consumer loans	14704.290
1670210	2357031	334635	Consumer loans	6622.020
1670211	2659632	249544	Consumer loans	11520.855
1670212	2785582	400317	Cash loans	18821.520
1670213	2418762	261212	Cash loans	16431.300

	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE
0	17145.0	17145.0	0.0	17145.0
1	607500.0	679671.0	NaN	607500.0
2	112500.0	136444.5	NaN	112500.0
3	450000.0	470790.0	NaN	450000.0
4	337500.0	404055.0	NaN	337500.0
...	...	...	...	...
1670209	267295.5	311400.0	0.0	267295.5
1670210	87750.0	64291.5	29250.0	87750.0
1670211	105237.0	102523.5	10525.5	105237.0
1670212	180000.0	191880.0	NaN	180000.0
1670213	360000.0	360000.0	NaN	360000.0

17

```
         WEEKDAY_APPR_PROCESS_START  HOUR_APPR_PROCESS_START  …  \
0                        SATURDAY                        15  …
1                        THURSDAY                        11  …
2                         TUESDAY                        11  …
3                          MONDAY                         7  …
4                        THURSDAY                         9  …
…                             …                          …  …  …
1670209                 WEDNESDAY                        12  …
1670210                   TUESDAY                        15  …
1670211                    MONDAY                        12  …
1670212                 WEDNESDAY                         9  …
1670213                    SUNDAY                        10  …


            NAME_SELLER_INDUSTRY  CNT_PAYMENT  NAME_YIELD_GROUP  \
0                   Connectivity         12.0            middle
1                            XNA         36.0        low_action
2                            XNA         12.0              high
3                            XNA         12.0            middle
4                            XNA         24.0              high
…                              …            …                 …
1670209                Furniture         30.0        low_normal
1670210                Furniture         12.0            middle
1670211      Consumer electronics       10.0        low_normal
1670212                      XNA         12.0        low_normal
1670213                      XNA         48.0            middle


                  PRODUCT_COMBINATION  DAYS_FIRST_DRAWING  DAYS_FIRST_DUE  \
0                POS mobile with interest            365243.0           -42.0
1                       Cash X-Sell: low            365243.0          -134.0
2                      Cash X-Sell: high            365243.0          -271.0
3                    Cash X-Sell: middle            365243.0          -482.0
4                       Cash Street: high                NaN             NaN
…                                     …                  …               …
1670209        POS industry with interest          365243.0          -508.0
1670210        POS industry with interest          365243.0         -1604.0
1670211        POS household with interest         365243.0         -1457.0
1670212                   Cash X-Sell: low         365243.0         -1155.0
1670213                Cash X-Sell: middle         365243.0         -1163.0


         DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE  DAYS_TERMINATION  \
0                            300.0          -42.0             -37.0
1                            916.0       365243.0          365243.0
2                             59.0       365243.0          365243.0
3                           -152.0         -182.0            -177.0
4                              NaN            NaN               NaN
…                                …              …                 …
```

```
1670209                          362.0        -358.0           -351.0
1670210                        -1274.0       -1304.0          -1297.0
1670211                        -1187.0       -1187.0          -1181.0
1670212                         -825.0        -825.0           -817.0
1670213                          247.0        -443.0           -423.0

          NFLAG_INSURED_ON_APPROVAL
0                            0.0
1                            1.0
2                            1.0
3                            1.0
4                            NaN
…                            …
1670209                      0.0
1670210                      0.0
1670211                      0.0
1670212                      1.0
1670213                      0.0

[1670214 rows x 37 columns]
```

[ ]:

[20]: `inp2.isnull().sum()`

```
[20]: SK_ID_PREV                         0
      SK_ID_CURR                         0
      NAME_CONTRACT_TYPE                 0
      AMT_ANNUITY                   372235
      AMT_APPLICATION                    0
      AMT_CREDIT                         1
      AMT_DOWN_PAYMENT              895844
      AMT_GOODS_PRICE               385515
      WEEKDAY_APPR_PROCESS_START         0
      HOUR_APPR_PROCESS_START            0
      FLAG_LAST_APPL_PER_CONTRACT        0
      NFLAG_LAST_APPL_IN_DAY             0
      RATE_DOWN_PAYMENT             895844
      RATE_INTEREST_PRIMARY        1664263
      RATE_INTEREST_PRIVILEGED     1664263
      NAME_CASH_LOAN_PURPOSE             0
      NAME_CONTRACT_STATUS               0
      DAYS_DECISION                      0
      NAME_PAYMENT_TYPE                  0
      CODE_REJECT_REASON                 0
      NAME_TYPE_SUITE               820405
      NAME_CLIENT_TYPE                   0
```

```
NAME_GOODS_CATEGORY              0
NAME_PORTFOLIO                   0
NAME_PRODUCT_TYPE                0
CHANNEL_TYPE                     0
SELLERPLACE_AREA                 0
NAME_SELLER_INDUSTRY             0
CNT_PAYMENT                 372230
NAME_YIELD_GROUP                 0
PRODUCT_COMBINATION            346
DAYS_FIRST_DRAWING          673065
DAYS_FIRST_DUE              673065
DAYS_LAST_DUE_1ST_VERSION   673065
DAYS_LAST_DUE               673065
DAYS_TERMINATION            673065
NFLAG_INSURED_ON_APPROVAL   673065
dtype: int64
```

[21]:
```
inp3 = inp2.dropna()
inp3
```

[21]:

	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY \
598	2388655	414811	Consumer loans	14152.545
21366	1184010	252161	Consumer loans	3136.275
24027	2144692	423348	Consumer loans	2640.195
43927	2697394	178347	Consumer loans	10324.665
115115	2403906	268507	Consumer loans	13452.660
...	...	...	...	...
1603346	1928485	386819	Consumer loans	45418.500
1619458	1347931	336203	Consumer loans	9207.180
1644524	2002593	168701	Consumer loans	3518.460
1645311	2396619	341729	Consumer loans	17179.380
1663414	1328802	105065	Consumer loans	6357.375

	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE \
598	153387.0	138046.5	15340.5	153387.0
21366	29781.0	29781.0	0.0	29781.0
24027	26145.0	26014.5	2614.5	26145.0
43927	101002.5	101002.5	0.0	101002.5
115115	145800.0	131220.0	14580.0	145800.0
...	...	...	...	...
1603346	562500.0	450000.0	112500.0	562500.0
1619458	113400.0	90720.0	22680.0	113400.0
1644524	38524.5	34668.0	3856.5	38524.5
1645311	171477.0	167571.0	17149.5	171477.0
1663414	68553.0	61695.0	6858.0	68553.0

	WEEKDAY_APPR_PROCESS_START	HOUR_APPR_PROCESS_START	... \

```
598               THURSDAY                    9   …
21366               MONDAY                   12   …
24027             THURSDAY                   14   …
43927               MONDAY                   14   …
115115              FRIDAY                    9   …
…                      …                     …   …
1603346             SUNDAY                   14   …
1619458          WEDNESDAY                   19   …
1644524             FRIDAY                   20   …
1645311             MONDAY                   20   …
1663414           SATURDAY                   21   …


          NAME_SELLER_INDUSTRY  CNT_PAYMENT NAME_YIELD_GROUP  \
598              Construction          12.0          middle
21366            Construction          12.0          middle
24027            Connectivity          12.0          middle
43927                Clothing          12.0          middle
115115           Construction          12.0          middle
…                         …             …               …
1603346               Tourism          12.0          middle
1619458          Connectivity          12.0          middle
1644524          Connectivity          12.0          middle
1645311              Clothing          12.0          middle
1663414          Connectivity          12.0          middle


              PRODUCT_COMBINATION  DAYS_FIRST_DRAWING DAYS_FIRST_DUE  \
598       POS industry with interest          365243.0         -659.0
21366     POS industry with interest          365243.0        365243.0
24027       POS mobile with interest          365243.0         -656.0
43927     POS industry with interest          365243.0         -151.0
115115    POS industry with interest          365243.0         -844.0
…                              …                  …               …
1603346      POS other with interest          365243.0         -399.0
1619458     POS mobile with interest          365243.0         -373.0
1644524     POS mobile with interest          365243.0         -599.0
1645311   POS industry with interest          365243.0         -462.0
1663414     POS mobile with interest          365243.0          -27.0


          DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE DAYS_TERMINATION  \
598                          -329.0         -329.0           -323.0
21366                         349.0       365243.0         365243.0
24027                        -326.0         -566.0           -557.0
43927                         179.0          -61.0            -56.0
115115                       -514.0         -724.0           -717.0
…                                …              …                …
1603346                       -69.0         -339.0           -332.0
1619458                       -43.0         -283.0           -274.0
```

```
1644524                              -269.0        -599.0         -592.0
1645311                              -132.0        -342.0         -339.0
1663414                               303.0      365243.0       365243.0


          NFLAG_INSURED_ON_APPROVAL
598                             0.0
21366                           0.0
24027                           0.0
43927                           0.0
115115                          0.0
...                             ...
1603346                         0.0
1619458                         0.0
1644524                         0.0
1645311                         1.0
1663414                         0.0


[71 rows x 37 columns]
```

[22]: `inp3.describe`

[22]:
```
<bound method NDFrame.describe of            SK_ID_PREV  SK_ID_CURR
NAME_CONTRACT_TYPE   AMT_ANNUITY  \
598           2388655      414811   Consumer loans   14152.545
21366         1184010      252161   Consumer loans    3136.275
24027         2144692      423348   Consumer loans    2640.195
43927         2697394      178347   Consumer loans   10324.665
115115        2403906      268507   Consumer loans   13452.660
...               ...         ...             ...         ...
1603346       1928485      386819   Consumer loans   45418.500
1619458       1347931      336203   Consumer loans    9207.180
1644524       2002593      168701   Consumer loans    3518.460
1645311       2396619      341729   Consumer loans   17179.380
1663414       1328802      105065   Consumer loans    6357.375


          AMT_APPLICATION   AMT_CREDIT   AMT_DOWN_PAYMENT   AMT_GOODS_PRICE  \
598               153387.0     138046.5            15340.5          153387.0
21366              29781.0      29781.0                0.0           29781.0
24027              26145.0      26014.5             2614.5           26145.0
43927             101002.5     101002.5                0.0          101002.5
115115            145800.0     131220.0            14580.0          145800.0
...                    ...          ...                ...               ...
1603346           562500.0     450000.0           112500.0          562500.0
1619458           113400.0      90720.0            22680.0          113400.0
1644524            38524.5      34668.0             3856.5           38524.5
1645311           171477.0     167571.0            17149.5          171477.0
1663414            68553.0      61695.0             6858.0           68553.0
```

	WEEKDAY_APPR_PROCESS_START	HOUR_APPR_PROCESS_START	…
598	THURSDAY	9	…
21366	MONDAY	12	…
24027	THURSDAY	14	…
43927	MONDAY	14	…
115115	FRIDAY	9	…
…	…	…	…
1603346	SUNDAY	14	…
1619458	WEDNESDAY	19	…
1644524	FRIDAY	20	…
1645311	MONDAY	20	…
1663414	SATURDAY	21	…

	NAME_SELLER_INDUSTRY	CNT_PAYMENT	NAME_YIELD_GROUP
598	Construction	12.0	middle
21366	Construction	12.0	middle
24027	Connectivity	12.0	middle
43927	Clothing	12.0	middle
115115	Construction	12.0	middle
…	…	…	…
1603346	Tourism	12.0	middle
1619458	Connectivity	12.0	middle
1644524	Connectivity	12.0	middle
1645311	Clothing	12.0	middle
1663414	Connectivity	12.0	middle

	PRODUCT_COMBINATION	DAYS_FIRST_DRAWING	DAYS_FIRST_DUE
598	POS industry with interest	365243.0	-659.0
21366	POS industry with interest	365243.0	365243.0
24027	POS mobile with interest	365243.0	-656.0
43927	POS industry with interest	365243.0	-151.0
115115	POS industry with interest	365243.0	-844.0
…	…	…	…
1603346	POS other with interest	365243.0	-399.0
1619458	POS mobile with interest	365243.0	-373.0
1644524	POS mobile with interest	365243.0	-599.0
1645311	POS industry with interest	365243.0	-462.0
1663414	POS mobile with interest	365243.0	-27.0

	DAYS_LAST_DUE_1ST_VERSION	DAYS_LAST_DUE	DAYS_TERMINATION
598	-329.0	-329.0	-323.0
21366	349.0	365243.0	365243.0
24027	-326.0	-566.0	-557.0
43927	179.0	-61.0	-56.0
115115	-514.0	-724.0	-717.0
…	…	…	…

```
1603346                   -69.0        -339.0        -332.0
1619458                   -43.0        -283.0        -274.0
1644524                  -269.0        -599.0        -592.0
1645311                  -132.0        -342.0        -339.0
1663414                   303.0      365243.0      365243.0

        NFLAG_INSURED_ON_APPROVAL
598                           0.0
21366                         0.0
24027                         0.0
43927                         0.0
115115                        0.0
…                             …
1603346                       0.0
1619458                       0.0
1644524                       0.0
1645311                       1.0
1663414                       0.0

[71 rows x 37 columns]>
```

[23]: `inp3.isna().sum()`

[23]:
```
SK_ID_PREV                     0
SK_ID_CURR                     0
NAME_CONTRACT_TYPE             0
AMT_ANNUITY                    0
AMT_APPLICATION                0
AMT_CREDIT                     0
AMT_DOWN_PAYMENT               0
AMT_GOODS_PRICE                0
WEEKDAY_APPR_PROCESS_START     0
HOUR_APPR_PROCESS_START        0
FLAG_LAST_APPL_PER_CONTRACT    0
NFLAG_LAST_APPL_IN_DAY         0
RATE_DOWN_PAYMENT              0
RATE_INTEREST_PRIMARY          0
RATE_INTEREST_PRIVILEGED       0
NAME_CASH_LOAN_PURPOSE         0
NAME_CONTRACT_STATUS           0
DAYS_DECISION                  0
NAME_PAYMENT_TYPE              0
CODE_REJECT_REASON             0
NAME_TYPE_SUITE                0
NAME_CLIENT_TYPE               0
NAME_GOODS_CATEGORY            0
NAME_PORTFOLIO                 0
```

```
NAME_PRODUCT_TYPE               0
CHANNEL_TYPE                    0
SELLERPLACE_AREA                0
NAME_SELLER_INDUSTRY            0
CNT_PAYMENT                     0
NAME_YIELD_GROUP                0
PRODUCT_COMBINATION             0
DAYS_FIRST_DRAWING              0
DAYS_FIRST_DUE                  0
DAYS_LAST_DUE_1ST_VERSION       0
DAYS_LAST_DUE                   0
DAYS_TERMINATION                0
NFLAG_INSURED_ON_APPROVAL       0
dtype: int64
```

[24]: `inp3.dtypes`

[24]:
```
SK_ID_PREV                     int64
SK_ID_CURR                     int64
NAME_CONTRACT_TYPE            object
AMT_ANNUITY                  float64
AMT_APPLICATION              float64
AMT_CREDIT                   float64
AMT_DOWN_PAYMENT             float64
AMT_GOODS_PRICE              float64
WEEKDAY_APPR_PROCESS_START    object
HOUR_APPR_PROCESS_START        int64
FLAG_LAST_APPL_PER_CONTRACT   object
NFLAG_LAST_APPL_IN_DAY         int64
RATE_DOWN_PAYMENT            float64
RATE_INTEREST_PRIMARY        float64
RATE_INTEREST_PRIVILEGED     float64
NAME_CASH_LOAN_PURPOSE        object
NAME_CONTRACT_STATUS          object
DAYS_DECISION                  int64
NAME_PAYMENT_TYPE             object
CODE_REJECT_REASON            object
NAME_TYPE_SUITE               object
NAME_CLIENT_TYPE              object
NAME_GOODS_CATEGORY           object
NAME_PORTFOLIO                object
NAME_PRODUCT_TYPE             object
CHANNEL_TYPE                  object
SELLERPLACE_AREA               int64
NAME_SELLER_INDUSTRY          object
CNT_PAYMENT                  float64
NAME_YIELD_GROUP              object
```

```
PRODUCT_COMBINATION              object
DAYS_FIRST_DRAWING              float64
DAYS_FIRST_DUE                 float64
DAYS_LAST_DUE_1ST_VERSION      float64
DAYS_LAST_DUE                  float64
DAYS_TERMINATION               float64
NFLAG_INSURED_ON_APPROVAL      float64
dtype: object
```

[90]:
```python
inp3.columns.tolist
```

[90]:
```
<bound method IndexOpsMixin.tolist of Index(['SK_ID_PREV', 'SK_ID_CURR',
       'NAME_CONTRACT_TYPE', 'AMT_ANNUITY',
       'AMT_APPLICATION', 'AMT_CREDIT', 'AMT_DOWN_PAYMENT', 'AMT_GOODS_PRICE',
       'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START',
       'FLAG_LAST_APPL_PER_CONTRACT', 'NFLAG_LAST_APPL_IN_DAY',
       'RATE_DOWN_PAYMENT', 'RATE_INTEREST_PRIMARY',
       'RATE_INTEREST_PRIVILEGED', 'NAME_CASH_LOAN_PURPOSE',
       'NAME_CONTRACT_STATUS', 'DAYS_DECISION', 'NAME_PAYMENT_TYPE',
       'CODE_REJECT_REASON', 'NAME_TYPE_SUITE', 'NAME_CLIENT_TYPE',
       'NAME_GOODS_CATEGORY', 'NAME_PORTFOLIO', 'NAME_PRODUCT_TYPE',
       'CHANNEL_TYPE', 'SELLERPLACE_AREA', 'NAME_SELLER_INDUSTRY',
       'CNT_PAYMENT', 'NAME_YIELD_GROUP', 'PRODUCT_COMBINATION',
       'DAYS_FIRST_DRAWING', 'DAYS_FIRST_DUE', 'DAYS_LAST_DUE_1ST_VERSION',
       'DAYS_LAST_DUE', 'DAYS_TERMINATION', 'NFLAG_INSURED_ON_APPROVAL'],
      dtype='object')>
```

[88]:
```python
plt.figure(figsize=(8,4))
sns.barplot(data = inp3, x = 'WEEKDAY_APPR_PROCESS_START', y = 'AMT_CREDIT')
plt.xticks(rotation = 45)
plt.show()
```
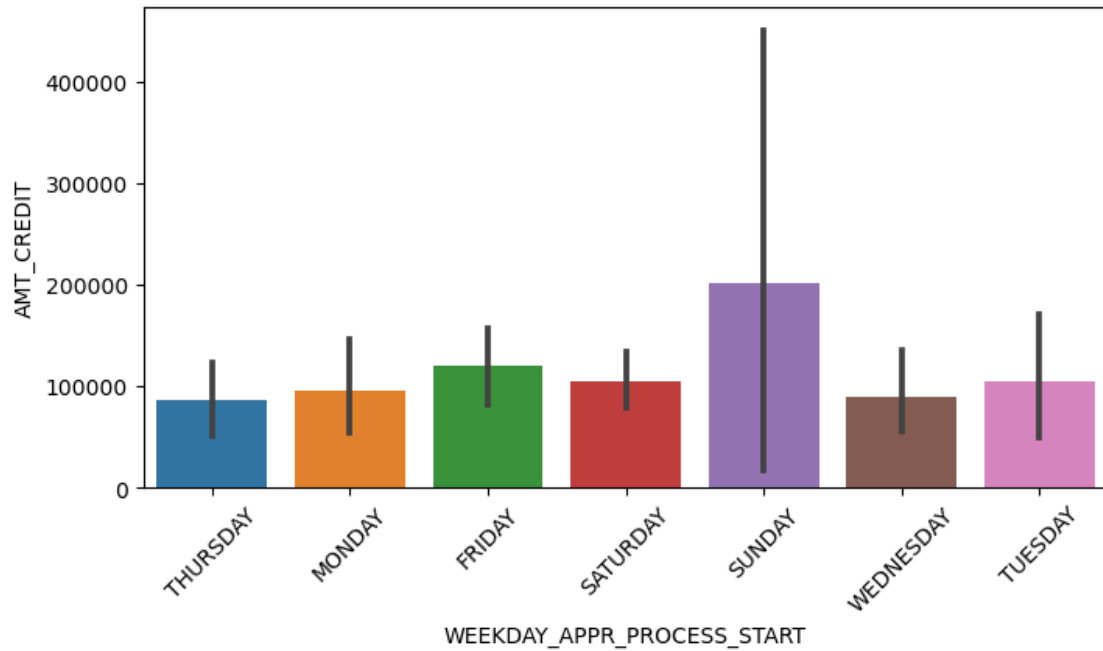
```python
[116]: plt.figure(figsize=(8,4))
       inp3.groupby('NAME_SELLER_INDUSTRY')['RATE_INTEREST_PRIVILEGED'].mean().
         ↪sort_values(ascending = False).plot.barh()
       plt.show()
```
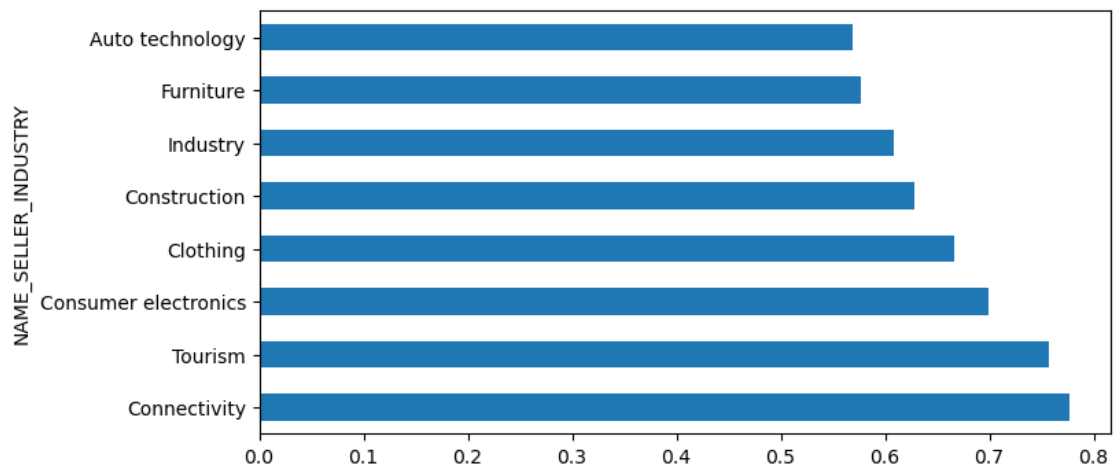


```python
[25]: sns.pairplot(data = inp3, vars = ['AMT_DOWN_PAYMENT', 'AMT_GOODS_PRICE',
                                         'AMT_GOODS_PRICE'])
       plt.show()
```
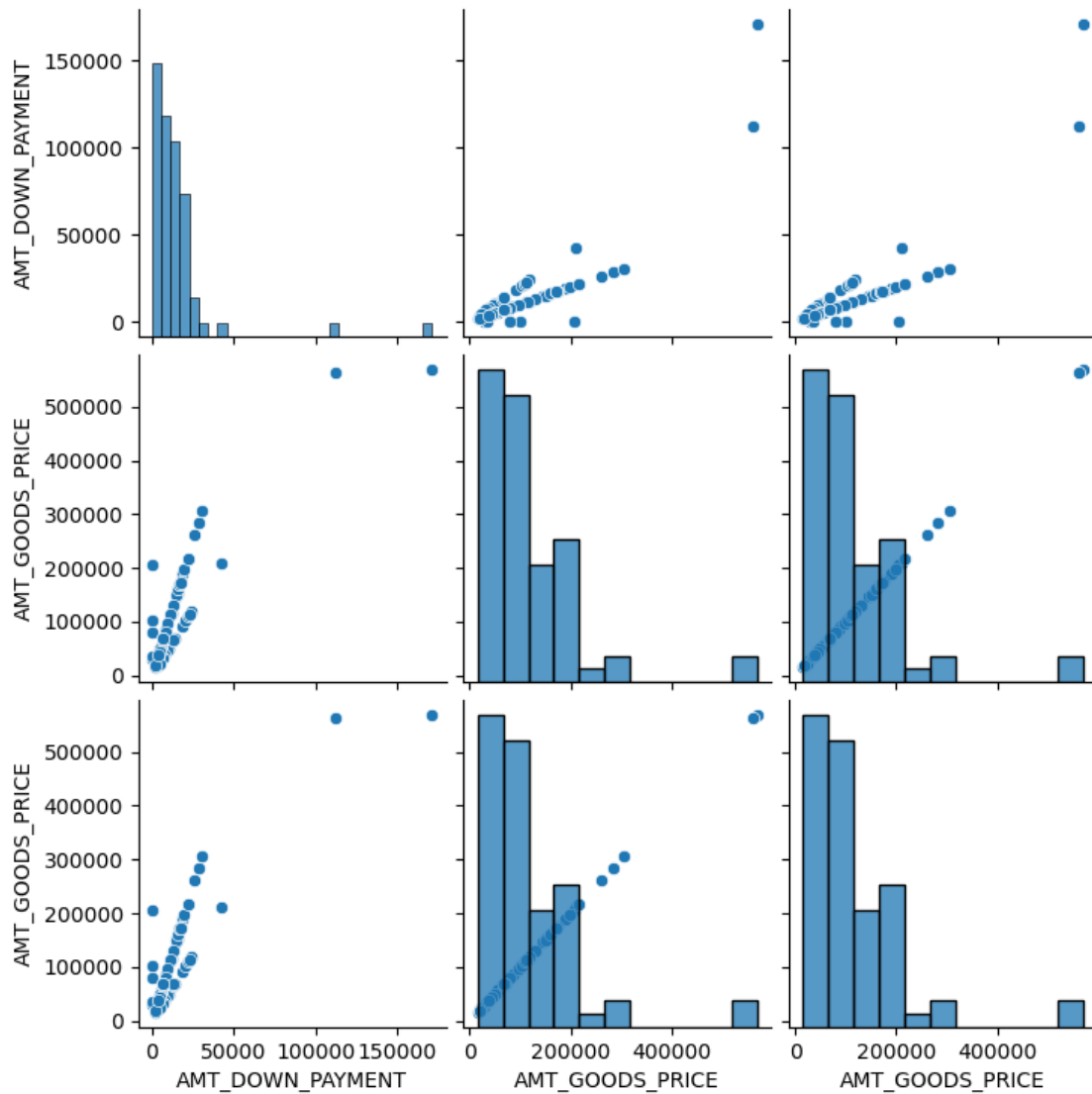
```
[26]: sns.boxplot(data = inp3, x = 'AMT_GOODS_PRICE')
      plt.show()
```
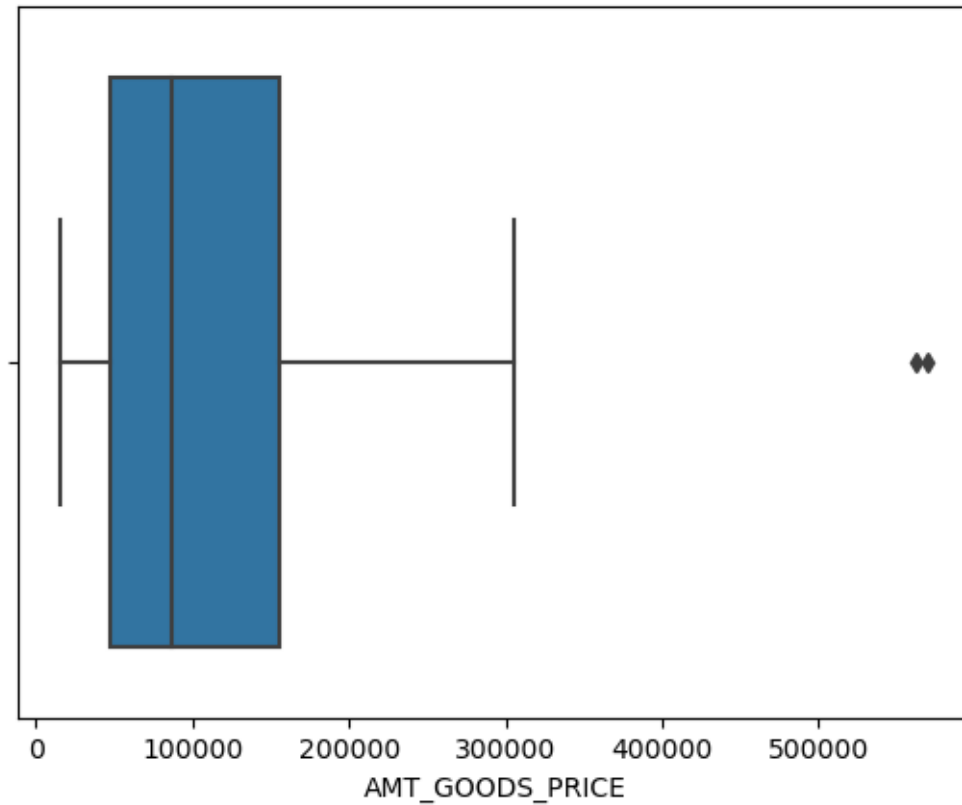
```
[27]: res = pd.pivot_table(data = inp3, columns= "WEEKDAY_APPR_PROCESS_START",
                   values = 'AMT_DOWN_PAYMENT',
                   index = 'NAME_SELLER_INDUSTRY')
      res
```

```
[27]: WEEKDAY_APPR_PROCESS_START    FRIDAY      MONDAY   SATURDAY     SUNDAY   THURSDAY  \
      NAME_SELLER_INDUSTRY
      Auto technology                 NaN         NaN  11272.500        NaN        NaN
      Clothing                        NaN    8574.750  13072.500        NaN    9735.75
      Connectivity                11818.5    9119.250  14944.500     3924.0    7947.00
      Construction                16230.0       0.000  15061.500        NaN   13702.50
      Consumer electronics            NaN    7752.375  16501.500    15115.5    3453.30
      Furniture                    8640.0         NaN  16142.625        NaN        NaN
      Industry                        NaN         NaN        NaN        NaN        NaN
      Tourism                         0.0   30600.000        NaN   112500.0        NaN

      WEEKDAY_APPR_PROCESS_START   TUESDAY      WEDNESDAY
      NAME_SELLER_INDUSTRY
      Auto technology                 NaN            NaN
      Clothing                        NaN            NaN
      Connectivity                 6673.5   12219.300000
```

29

```
Construction              18900.0   13423.500000
Consumer electronics      15210.0   10540.285714
Furniture                  4590.0    8776.500000
Industry                   9535.5   86877.000000
Tourism                       NaN            NaN
```

[28]:
```python
plt.figure(figsize=(5,3))
sns.heatmap(res, cmap = "Greens", annot = True)
plt.show()
```
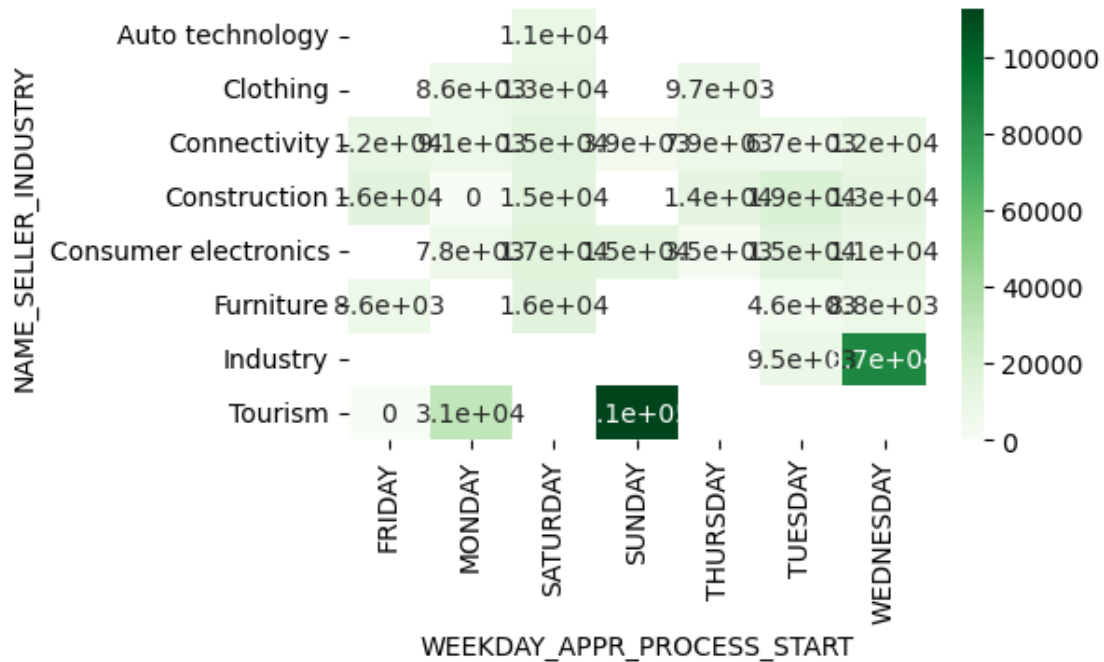

[127]:
```python
plt.figure(figsize=(10,3))
sns.histplot(inp3['AMT_DOWN_PAYMENT'], bins = 30, color = 'purple')
plt.show()
```
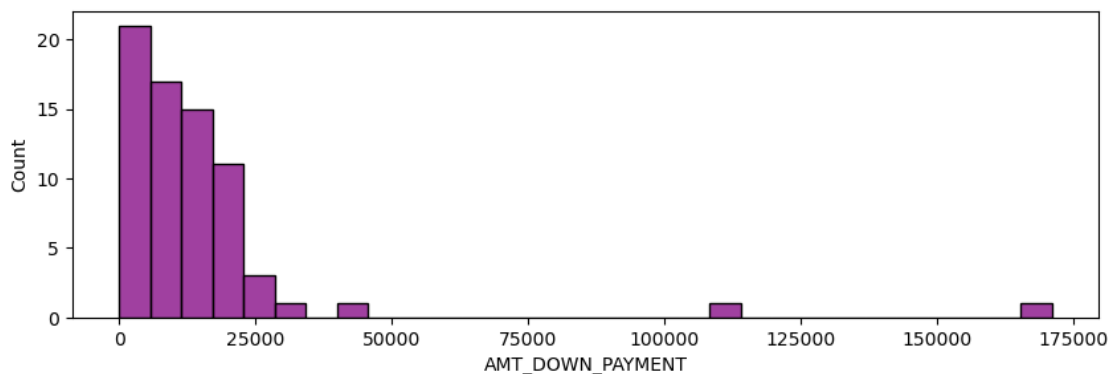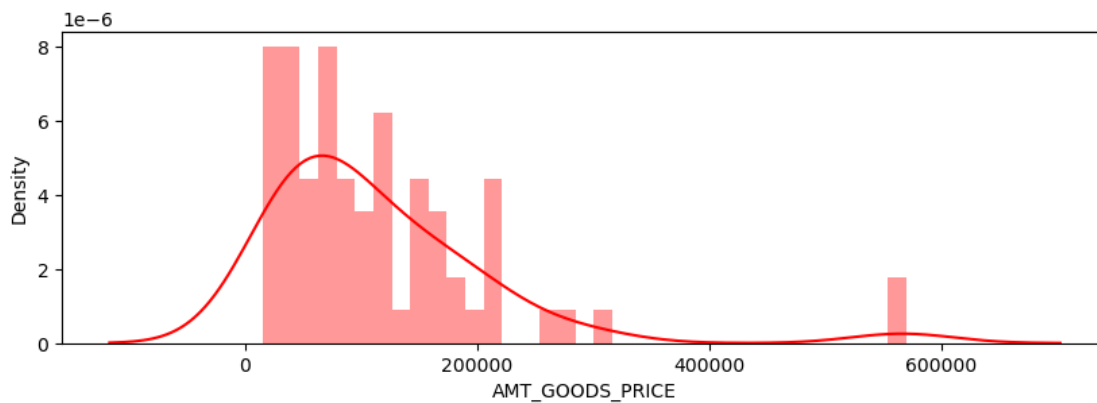
```
[128]: plt.figure(figsize=(10,3))
       sns.histplot(inp3[inp3['AMT_DOWN_PAYMENT']<25000]['AMT_DOWN_PAYMENT'], bins =␣
        ↪30, color = 'purple')
       plt.show()
```
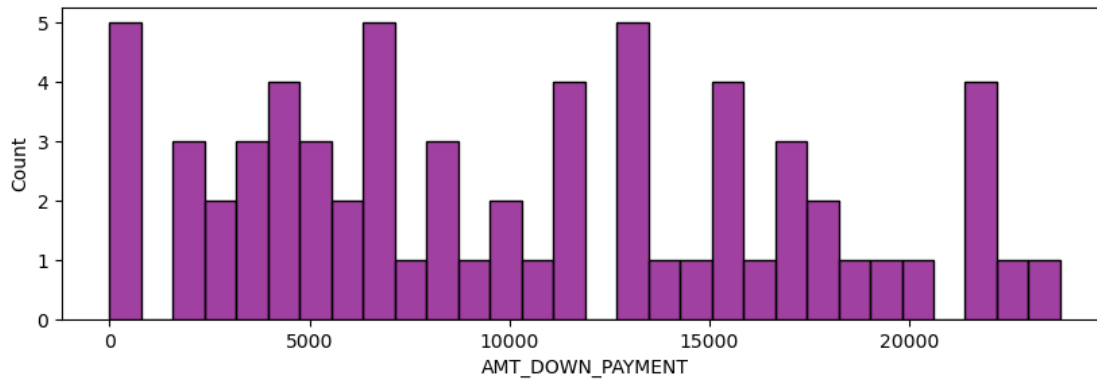


```
[131]: import warnings
       warnings.filterwarnings('ignore')
```

```
[133]: plt.figure(figsize=(10,3))
       sns.distplot(inp3.AMT_GOODS_PRICE, bins = 35, kde = True, color = 'Red')
       plt.show()
```
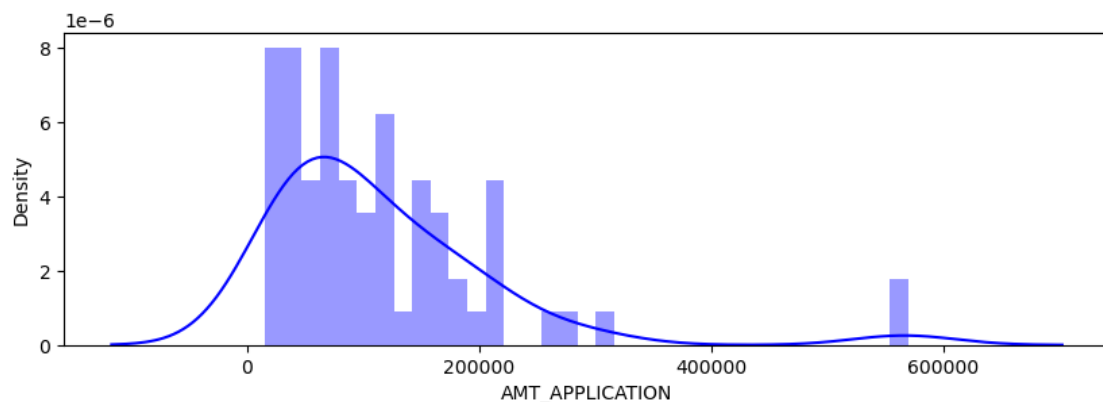


```
[135]: plt.figure(figsize=(10,3))
       sns.distplot(inp3.AMT_CREDIT, bins = 35, kde = True, color = 'purple')
       plt.show()
```
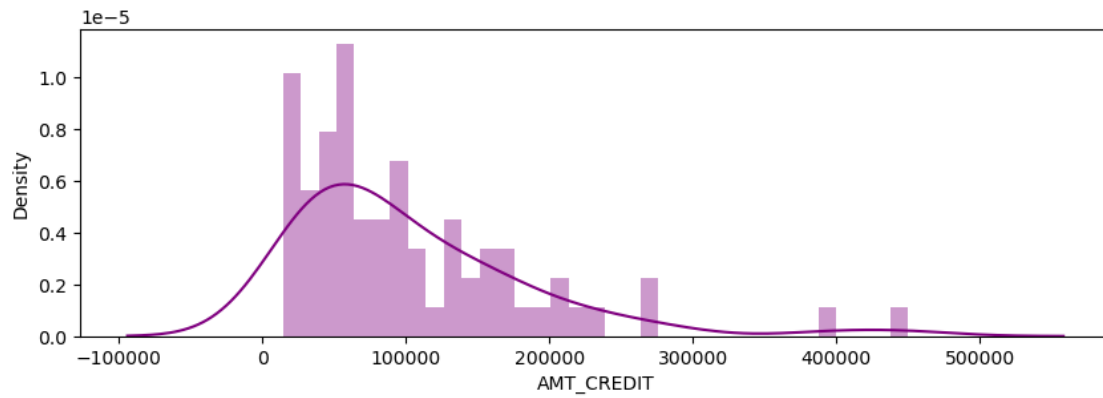
```
[137]: plt.figure(figsize=(10,3))
       sns.distplot(inp3.AMT_ANNUITY, bins = 35, kde = True, color = 'Green')
       plt.show()
```
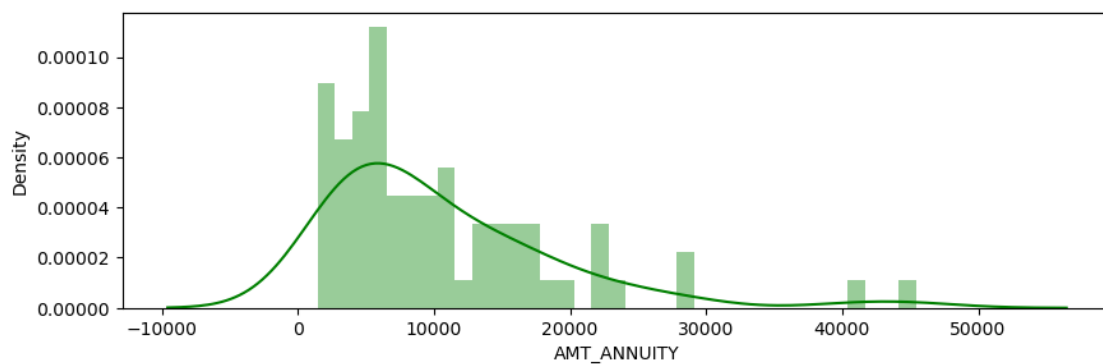


```
[146]: plt.figure(figsize=(10,3))
       sns.distplot(inp3.AMT_APPLICATION, bins = 35, kde = True, color = 'blue')
       plt.show()
```

[ ]: