



In [164]:

```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 pd.set_option('display.max_columns', 80)
6 pd.set_option('display.max_rows', 20000)
7 pd.options.display.float_format = "{:.2f}".format
8 zbior = pd.read_csv('C:/Users/Laptop/Desktop/Stat/DATASET/train_LTFS.csv')
9
10 #usuniecie kropek z nazw zmiennych
11 zbior.rename(columns={'Date.of.Birth':'Date_of_Birth','Employment.Type':'Employment_Ty
12
13 #usuwa wartości "n/a"
14 #zbior.dropna(inplace=True)
15
16 #edycja datu urodzenia do wieku (z pominięciem metody "pd.DatetimeIndex(zbior['Date_of
17 zbior['rok_ur2'] = zbior['Date_of_Birth'].str[-2:]
18 zbior['rok_ur2'] = zbior['rok_ur2'].astype(str).astype(int)
19 zbior['wiek'] = 120 - zbior['rok_ur2']
20 #korekta przeliczenia wieku
21 zbior.wiek.replace([120], [20], inplace=True)
22
23 #edycja zmiennej "AVERAGE_ACCT_AGE" z typu "object" do wartości wyrazonej w miesiącach
24 zbior['acc_age_rok'] = zbior['AVERAGE_ACCT_AGE'].str[:2]
25 #usuwa wszystkie znaki, które nie są cyframi
26 zbior['acc_age_rok'].replace(regex=True, inplace= True, to_replace=r'\D', value=r'')
27 zbior['acc_age_mies'] = zbior['AVERAGE_ACCT_AGE'].str[-5:-3]
28 zbior['acc_age_rok'] = zbior['acc_age_rok'].astype(str).astype(int)
29 zbior['acc_age_mies'] = zbior['acc_age_mies'].astype(str).astype(int)
30 zbior['AVERAGE_ACCT_AGE_mth'] = zbior['acc_age_rok'] *12 + zbior['acc_age_mies']
31
32 #edycja zmiennej "CREDIT_HISTORY_LENGTH" z typu "object" do wartości wyrazonej w miesiącach
33 zbior['credit_hist_rok'] = zbior['CREDIT_HISTORY_LENGTH'].str[:2]
34 #usuwa wszystkie znaki, które nie są cyframi
35 zbior['credit_hist_rok'].replace(regex=True, inplace= True, to_replace=r'\D', value=r'')
36 zbior['credit_hist_mies'] = zbior['CREDIT_HISTORY_LENGTH'].str[-5:-3]
37 zbior['credit_hist_mies'].replace(regex=True, inplace= True, to_replace=r'\D', value=r'')
38 zbior['credit_hist_rok'] = zbior['credit_hist_rok'].astype(str).astype(int)
39 zbior['credit_hist_mies'] = zbior['credit_hist_mies'].astype(str).astype(int)
40 zbior['CREDIT_HISTORY_LENGTH_mth'] = zbior['credit_hist_rok'] *12 + zbior['credit_hist
41
42 #usuwa kolumny robocze oraz zmienne "MobileNo_Avl_Flag" (zmienna przyjmuje tylko jedną
43 #zbior.drop(['MobileNo_Avl_Flag', 'rok_ur2', 'acc_age_rok', 'acc_age_mies', 'credit_hist
44
45 #przekształcenie zmiennej "Employment_Type" z typu object do int64
46 zbior.Employment_Type.replace(['Salaried', 'Self employed'], [1, 2], inplace=True)
47
48 #przekształca zmienne "PERFORM_CNS_SCORE_DESCRIPTION" w szereg zmiennych binarnych
49 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'A-Very Low Risk', 'SCORE_DESC_A_V
50 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'B-Very Low Risk', 'SCORE_DESC_B_V
51 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'C-Very Low Risk', 'SCORE_DESC_C_V
52 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'D-Very Low Risk', 'SCORE_DESC_D_V
53 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'E-Low Risk', 'SCORE_DESC_E_LR'] =
54 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'F-Low Risk', 'SCORE_DESC_F_LR'] =
55 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'G-Low Risk', 'SCORE_DESC_G_LR'] =
56 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'H-Medium Risk', 'SCORE_DESC_H_MR'
57 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'I-Medium Risk', 'SCORE_DESC_I_MR'
58 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'J-High Risk', 'SCORE_DESC_J_HR'] =
59 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'K-High Risk', 'SCORE_DESC_K_HR'] =

```

```
60 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'L-Very High Risk', 'SCORE_DESC_L_'
61 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'M-Very High Risk', 'SCORE_DESC_M_'
62 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'No Bureau History Available', 'SC
63 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'Not Scored: Sufficient History No
64 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'Not Scored: Not Enough Info avail
65 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'Not Scored: No Activity seen on t
66 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'Not Scored: No Updates available
67 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'Not Scored: Only a Guarantor', 'S
68 zbior.loc[zbior['PERFORM_CNS_SCORE_DESCRIPTION'] == 'Not Scored: More than 50 active A
69 zbior.fillna(0, inplace=True)
70
71 #agregowanie kategorii zmiennej "PERFORM_CNS_SCORE_DESCRIPTION"
72 zbior['PERFORM_CNS_SCORE_DESCRIPTION2'] = zbior.PERFORM_CNS_SCORE_DESCRIPTION
73 zbior.PERFORM_CNS_SCORE_DESCRIPTION2.replace(['No Bureau History Available',
74 'C-Very Low Risk',
75 'A-Very Low Risk',
76 'D-Very Low Risk',
77 'B-Very Low Risk',
78 'M-Very High Risk',
79 'F-Low Risk',
80 'K-High Risk',
81 'H-Medium Risk',
82 'E-Low Risk',
83 'I-Medium Risk',
84 'G-Low Risk',
85 'Not Scored: Sufficient History Not Avai
86 'J-High Risk',
87 'Not Scored: Not Enough Info available o
88 'Not Scored: No Activity seen on the cus
89 'Not Scored: No Updates available in las
90 'L-Very High Risk',
91 'Not Scored: Only a Guarantor',
92 'Not Scored: More than 50 active Account
93
94 [
95     'Not Scored',
96     'Very Low Risk',
97     'Very Low Risk',
98     'Very Low Risk',
99     'Very Low Risk',
100    'Very High Risk',
101    'Low Risk',
102    'High Risk',
103    'Medium Risk',
104    'Low Risk',
105    'Medium Risk',
106    'Low Risk',
107    'Not Scored',
108    'High Risk',
109    'Not Scored',
110    'Not Scored',
111    'Not Scored',
112    'Very High Risk',
113    'Not Scored',
114    'Not Scored' ],
115
116 #kod pomocniczy:
117 #zbior['PERFORM_CNS_SCORE_DESCRIPTION2'].value_counts()
118 #zbior.info()
119 #for col in zbior.columns:
120 #    print(col)
```

121

In [ ]:

```
1 #wydzielenie identyfikatorów klientów do ograniczenia zmiennej "Date_of_Birth"
2 zbior2 = zbior[['UniqueID', 'Date_of_Birth']]
3 zbior2.info()
4 zbior2.sort_values('Date_of_Birth')
5 count = zbior2.groupby(['Date_of_Birth']).count()
6 print(count)
7 count_df = pd.DataFrame(count)
8 count_df = count_df.reset_index()
9 sort_by = count_df.sort_values('count', ascending=False)
10 top138 = sort_by.head(138)
11 print(top138)
12 top138.info()
13 top138.to_excel('top138.xlsx')
14
15 #wyliczenie wartości unikalnych
16 zbior2 = zbior[['UniqueID', 'disbursed_amount', 'asset_cost', 'ltv', 'branch_id', 'sup
17 ]]
18 zbior2.nunique()
```

In [165]:

```
1 #przekształcenie logarytmiczne zmiennych ciągłych kredytu
2 zbior['disbursed_amount_log'] = np.log1p(zbior['disbursed_amount'])
3 zbior['asset_cost_log'] = np.log1p(zbior['asset_cost'])
4 zbior['ltv_log'] = np.log1p(zbior['ltv'])
5
6 #wykresy zmiennych kredytu, przed i po przekształceniu log.
7 plt.rcParams['figure.figsize'] = (16, 4)
8
9 plt.subplot(1, 3, 1)
10 sns.distplot(zbior['disbursed_amount_log'], color = 'grey')
11 plt.title('Kwota kredytu')
12
13 plt.subplot(1, 3, 2)
14 sns.distplot(zbior['asset_cost_log'], color = 'brown')
15 plt.title('Koszt całkowity kredytu')
16
17 plt.subplot(1, 3, 3)
18 sns.distplot(zbior['ltv_log'], color = 'red')
19 plt.title('Wskaźnik LTV')
20
21 plt.rcParams['figure.figsize'] = (16, 4)
22
23 plt.subplot(1, 3, 1)
24 sns.distplot(zbior['disbursed_amount'], color = 'grey')
25 plt.title('Kwota kredytu')
26
27 plt.subplot(1, 3, 2)
28 sns.distplot(zbior['asset_cost'], color = 'brown')
29
30 plt.title('Koszt całkowity kredytu')
31
32 plt.subplot(1, 3, 3)
33 sns.distplot(zbior['ltv'], color = 'red')
34 plt.title('Wskaźnik LTV')
35
```

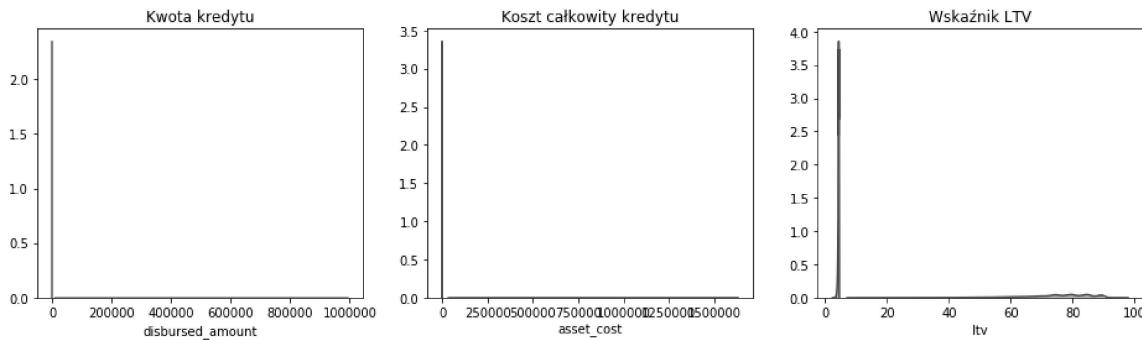
C:\Users\ Laptop\Anaconda3\lib\site-packages\ipykernel\_launcher.py:23: MatplotlibDeprecationWarning: Adding an axes using the same arguments as a previous axes currently reuses the earlier instance. In a future version, a new instance will always be created and returned. Meanwhile, this warning can be suppressed, and the future behavior ensured, by passing a unique label to each axes instance.

C:\Users\ Laptop\Anaconda3\lib\site-packages\ipykernel\_launcher.py:27: MatplotlibDeprecationWarning: Adding an axes using the same arguments as a previous axes currently reuses the earlier instance. In a future version, a new instance will always be created and returned. Meanwhile, this warning can be suppressed, and the future behavior ensured, by passing a unique label to each axes instance.

C:\Users\ Laptop\Anaconda3\lib\site-packages\ipykernel\_launcher.py:32: MatplotlibDeprecationWarning: Adding an axes using the same arguments as a previous axes currently reuses the earlier instance. In a future version, a new instance will always be created and returned. Meanwhile, this warning can be suppressed, and the future behavior ensured, by passing a unique label to each axes instance.

Out[165]:

Text(0.5, 1.0, 'Wskaźnik LTV')



In [171]:

```
1 #przekształcenie logarytmiczne zmiennych ciągłych kredytu
2 zbior['disbursed_amount_log'] = np.log1p(zbior['disbursed_amount'])
3 zbior['asset_cost_log'] = np.log1p(zbior['asset_cost'])
4 zbior['ltv_log'] = np.log1p(zbior['ltv'])
```

In [173]:

1 zbiór.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 233154 entries, 0 to 233153
Data columns (total 73 columns):
UniqueID                      233154 non-null int64
disbursed_amount               233154 non-null int64
asset_cost                     233154 non-null int64
ltv                            233154 non-null float64
branch_id                      233154 non-null int64
supplier_id                    233154 non-null int64
manufacturer_id                233154 non-null int64
Current_pincode_ID             233154 non-null int64
Date_of_Birth                  233154 non-null object
Employment_Type                233154 non-null float64
DisbursalDate                 233154 non-null object
State_ID                       233154 non-null int64
Employee_code_ID               233154 non-null int64
MobileNo_Avl_Flag              233154 non-null int64
Aadhar_flag                    233154 non-null int64
PAN_flag                       233154 non-null int64
VoterID_flag                  233154 non-null int64
Driving_flag                   233154 non-null int64
Passport_flag                  233154 non-null int64
PERFORM_CNS_SCORE              233154 non-null int64
PERFORM_CNS_SCORE_DESCRIPTION  233154 non-null object
PRI_NO_OF_ACCTS                233154 non-null int64
PRI_ACTIVE_ACCTS               233154 non-null int64
PRI_OVERDUE_ACCTS              233154 non-null int64
PRI_CURRENT_BALANCE            233154 non-null int64
PRI_SANCTIONED_AMOUNT          233154 non-null int64
PRI_DISBURSED_AMOUNT           233154 non-null int64
SEC_NO_OF_ACCTS                233154 non-null int64
SEC_ACTIVE_ACCTS               233154 non-null int64
SEC_OVERDUE_ACCTS              233154 non-null int64
SEC_CURRENT_BALANCE            233154 non-null int64
SEC_SANCTIONED_AMOUNT          233154 non-null int64
SEC_DISBURSED_AMOUNT           233154 non-null int64
PRIMARY_INSTAL_AMT             233154 non-null int64
SEC_INSTAL_AMT                 233154 non-null int64
NEW_ACCTS_IN_LAST_SIX_MONTHS   233154 non-null int64
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS 233154 non-null int64
AVERAGE_ACCT_AGE               233154 non-null object
CREDIT_HISTORY_LENGTH           233154 non-null object
NO_OF_INQUIRIES                 233154 non-null int64
loan_default                   233154 non-null int64
rok_ur2                        233154 non-null int32
wiek                           233154 non-null int64
acc_age_rok                    233154 non-null int32
acc_age_mies                   233154 non-null int32
AVERAGE_ACCT_AGE_mth           233154 non-null int32
credit_hist_rok                233154 non-null int32
credit_hist_mies                233154 non-null int32
CREDIT_HISTORY_LENGTH_mth       233154 non-null int32
SCORE_DESC_A_VLR                233154 non-null float64
SCORE_DESC_B_VLR                233154 non-null float64
SCORE_DESC_C_VLR                233154 non-null float64
SCORE_DESC_D_VLR                233154 non-null float64
SCORE_DESC_E_LR                 233154 non-null float64
```

```
SCORE_DESC_F_LR          233154 non-null float64
SCORE_DESC_G_LR          233154 non-null float64
SCORE_DESC_H_MR          233154 non-null float64
SCORE_DESC_I_MR          233154 non-null float64
SCORE_DESC_J_HR          233154 non-null float64
SCORE_DESC_K_HR          233154 non-null float64
SCORE_DESC_L_VHR         233154 non-null float64
SCORE_DESC_M_VHR         233154 non-null float64
SCORE_DESC_N1             233154 non-null float64
SCORE_DESC_N2             233154 non-null float64
SCORE_DESC_N3             233154 non-null float64
SCORE_DESC_N4             233154 non-null float64
SCORE_DESC_N5             233154 non-null float64
SCORE_DESC_N6             233154 non-null float64
SCORE_DESC_N7             233154 non-null float64
PERFORM_CNS_SCORE_DESCRIPTION2 233154 non-null object
disbursed_amount_log     233154 non-null float64
asset_cost_log           233154 non-null float64
ltv_log                  233154 non-null float64
dtypes: float64(25), int32(7), int64(35), object(6)
memory usage: 123.6+ MB
```

In [7]:

```
1 zbor.wiek.replace([120], [20], inplace=True)
2 zbor['wiek'].value_counts()
```

Out[7]:

```
25    10584
26    10530
28    10390
24    10211
30    10116
27    10007
29     9385
32     9232
31     8903
33     8633
34     8441
35     7922
23     7489
36     7481
37     7140
38     7018
40     6788
39     6178
42     5883
45     5627
41     5614
44     5581
43     5310
46     4638
47     4466
48     4366
50     4063
49     3721
51     3147
52     3081
53     2522
55     2397
54     2268
22     2188
56     1773
21     1604
57     1556
58     1478
60     1241
59     1173
61      816
20      632
62      621
63      412
64      337
65      169
66       21
71        1
Name: wiek, dtype: int64
```

In [284]:

```
1 zbor.PERFORM_CNS_SCORE_DESCRIPTION2.replace(['No Bureau History Available'], ['Not Sc'])
```

Out[284]:

```
0          Not Scored
1          Medium Risk
2          Not Scored
3      Very High Risk
4          Not Scored
...
233149    Very Low Risk
233150    Very Low Risk
233151    Not Scored
233152    Not Scored
233153    Not Scored
Name: PERFORM_CNS_SCORE_DESCRIPTION2, Length: 225493, dtype: object
```

In [3]:

```
1 zbor['PERFORM_CNS_SCORE_DESCRIPTION2'].value_counts()
```

Out[3]:

Not Scored	124253
Very Low Risk	49671
Low Risk	17906
Medium Risk	12135
High Risk	11774
Very High Risk	9754

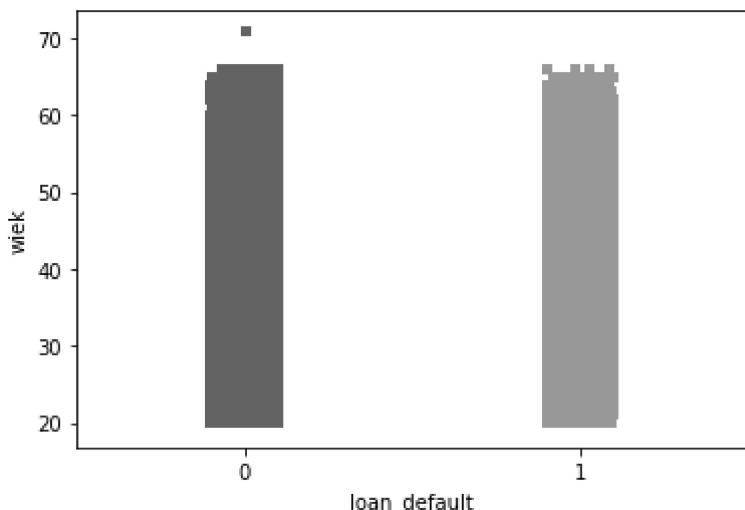
Name: PERFORM\_CNS\_SCORE\_DESCRIPTION2, dtype: int64

In [8]:

```
1 sns.stripplot(y='wiek', x='loan_default', data=zbor, jitter=True, dodge=True, marker=
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x22e710f7f08>
```



In [41]:

```
1 zbor2= zbor[['Date_of_Birth', 'wiek']]
2 zbor2.to_excel('zbor2.xlsx')
```

In [92]:

```
1 for col in zbior.columns:  
2     print(col)
```

UniqueID  
disbursed\_amount  
asset\_cost  
ltv  
branch\_id  
supplier\_id  
manufacturer\_id  
Current\_pincode\_ID  
Date\_of\_Birth  
Employment\_Type  
DisbursalDate  
State\_ID  
Employee\_code\_ID  
MobileNo\_Avl\_Flag  
Aadhar\_flag  
PAN\_flag  
VoterID\_flag  
Driving\_flag  
Passport\_flag  
PERFORM\_CNS\_SCORE  
PERFORM\_CNS\_SCORE\_DESCRIPTION  
PRI\_NO\_OF\_ACCTS  
PRI\_ACTIVE\_ACCTS  
PRI\_OVERDUE\_ACCTS  
PRI\_CURRENT\_BALANCE  
PRI\_SANCTIONED\_AMOUNT  
PRI\_DISBURSED\_AMOUNT  
SEC\_NO\_OF\_ACCTS  
SEC\_ACTIVE\_ACCTS  
SEC\_OVERDUE\_ACCTS  
SEC\_CURRENT\_BALANCE  
SEC\_SANCTIONED\_AMOUNT  
SEC\_DISBURSED\_AMOUNT  
PRIMARY\_INSTAL\_AMT  
SEC\_INSTAL\_AMT  
NEW\_ACCTS\_IN\_LAST\_SIX\_MONTHS  
DELINQUENT\_ACCTS\_IN\_LAST\_SIX\_MONTHS  
AVERAGE\_ACCT\_AGE  
CREDIT\_HISTORY\_LENGTH  
NO\_OF\_INQUIRIES  
loan\_default  
rok\_ur2  
wiek  
acc\_age\_rok  
acc\_age\_mies  
AVERAGE\_ACCT\_AGE\_mth  
credit\_hist\_rok  
credit\_hist\_mies  
CREDIT\_HISTORY\_LENGTH\_mth  
SCORE\_DESC\_A\_VLR  
SCORE\_DESC\_B\_VLR  
SCORE\_DESC\_C\_VLR  
SCORE\_DESC\_D\_VLR  
SCORE\_DESC\_E\_LR  
SCORE\_DESC\_F\_LR  
SCORE\_DESC\_G\_LR

```
SCORE_DESC_H_MR
SCORE_DESC_I_MR
SCORE_DESC_J_MR
SCORE_DESC_K_MR
SCORE_DESC_L_VHR
SCORE_DESC_M_VHR
SCORE_DESC_N1
SCORE_DESC_N2
SCORE_DESC_N3
SCORE_DESC_N4
SCORE_DESC_N5
SCORE_DESC_N6
SCORE_DESC_N7
PERFORM_CNS_SCORE_DESCRIPTION2
disbursed_amount_log
asset_cost_log
ltv_log
```

In [32]:

```
1 zbor.AVERAGE_ACCT_AGE.head()
```

Out[32]:

```
0    0yrs 0mon
1    1yrs 11mon
2    0yrs 0mon
3    0yrs 8mon
4    0yrs 0mon
Name: AVERAGE_ACCT_AGE, dtype: object
```

In [69]:

```
1 urodz = pd.DataFrame(zbor['Date_of_Birth'].value_counts())
2 print(urodz)
```

Date_of_Birth	
01-01-88	2173
01-01-90	2170
01-01-87	2127
01-01-86	2063
01-01-85	2005
01-01-91	1985
01-01-89	1962
01-01-93	1930
01-01-95	1924
01-01-92	1924
01-01-94	1867
01-01-83	1841
01-01-80	1821
01-01-84	1768
01-01-81	1764
01-01-96	1759
01-01-82	1691
01-01-78	1646
01-01-75	1644

In [ ]:

```
1
```

In [57]:

```
1 urodz.head()
```

Out[57]:

Date_of_Birth	
01-01-88	2173
01-01-90	2170
01-01-87	2127
01-01-86	2063
01-01-85	2005

In [74]:

```
1 plt.hist(zbior['Date_of_Birth'])
```

Out[74]:

```
(array([88812., 30110., 24685., 21202., 19519., 16118., 13321., 9869.,
       6328., 3190.]),
 array([
     0. , 1543.2, 3086.4, 4629.6, 6172.8, 7716. , 9259.2,
    10802.4, 12345.6, 13888.8, 15432. ]),
 <a list of 10 Patch objects>)
```

Error in callback <function flush\_figures at 0x0000022E73D9C798> (for post\_execute):

In [58]:

```
1 urodz.to_excel('urodz.xlsx')
```

In [65]:

```
1 urodz.groupby(['Date_of_Birth']).agg(sum)
```

Out[65]:

Date\_of\_Birth

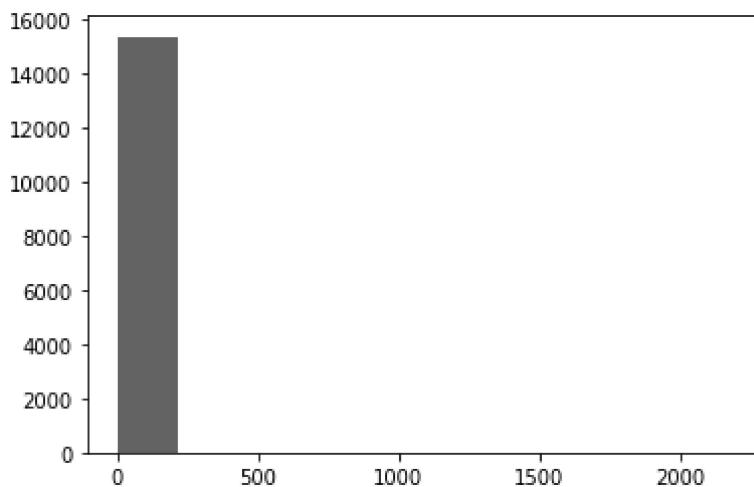
1  
2  
3  
4  
5  
6  
7  
8  
9  
10

In [59]:

```
1 plt.hist(urodz.Date_of_Birth)  
2 #plt.xticks([20, 25, 30, 35, 40, 45, 50, 55, 60, 65])  
3 #plt.title('Wiek klientów')
```

Out[59]:

```
(array([1.5386e+04, 1.2000e+01, 3.0000e+00, 2.0000e+00, 2.0000e+00,  
       4.0000e+00, 3.0000e+00, 5.0000e+00, 9.0000e+00, 7.0000e+00]),  
 array([1.0000e+00, 2.1820e+02, 4.3540e+02, 6.5260e+02, 8.6980e+02,  
       1.0870e+03, 1.3042e+03, 1.5214e+03, 1.7386e+03, 1.9558e+03,  
       2.1730e+03]),  
<a list of 10 Patch objects>)
```



In [66]:

```
1 zbiór.acc_age_mth.to_excel('ACC AGE.xlsx')
```

In [65]:

```
1 zbior['acc_age_mth'].value_counts()
```

Out[65]:

```
0      114135
6      5907
7      5254
11     5110
10     5005
...
182     1
227     1
270     1
169     1
167     1
Name: acc_age_mth, Length: 192, dtype: int64
```

In [44]:

```
1 zbior['CREDIT_HISTORY_LENGTH_mth'].value_counts()
```

Out[44]:

```
0      119127
6      4761
25     4745
7      4017
24     3833
...
296     1
339     1
249     1
274     1
288     1
Name: CREDIT_HISTORY_LENGTH_mth, Length: 294, dtype: int64
```

In [12]:

```
1 #zbior['rok_ur'] = pd.DatetimeIndex(zbior['Date_of_Birth']).year" - metoda nie dziala
```

In [1]:

```
1 zbior.wiek2.describe()
```

```
-----
NameError                                 Traceback (most recent call last)
<ipython-input-1-d9f103ad4658> in <module>
----> 1 zbior.wiek2.describe()
```

```
NameError: name 'zbior' is not defined
```

In [2]:

1

Out[2]:

```
25    10584
26    10530
28    10390
24    10211
30    10116
27    10007
29    9385
32    9232
31    8903
33    8633
34    8441
35    7922
23    7489
36    7481
37    7140
38    7018
40    6788
39    6178
```

In [98]:

```
1 zbor2 = zbor[['UniqueID', 'disbursed_amount', 'asset_cost', 'ltv', 'branch_id', 'sup
2   ]]
```

In [99]:

1 zbior2.nunique()

Out[99]:

UniqueID	233154
disbursed_amount	24565
asset_cost	46252
ltv	6579
branch_id	82
supplier_id	2953
manufacturer_id	11
Current_pincode_ID	6698
Date_of_Birth	15433
Employment_Type	3
DisbursalDate	84
State_ID	22
Employee_code_ID	3270
PERFORM_CNS_SCORE	573
PERFORM_CNS_SCORE_DESCRIPTION	20
PRI_NO_OF_ACCTS	108
PRI_ACTIVE_ACCTS	40
PRI_OVERDUE_ACCTS	22
PRI_CURRENT_BALANCE	71341
PRI_SANCTIONED_AMOUNT	44390
PRI_DISBURSED_AMOUNT	47909
SEC_NO_OF_ACCTS	37
SEC_ACTIVE_ACCTS	23
SEC_OVERDUE_ACCTS	9
SEC_CURRENT_BALANCE	3246
SEC_SANCTIONED_AMOUNT	2223
SEC_DISBURSED_AMOUNT	2553
PRIMARY_INSTAL_AMT	28067
SEC_INSTAL_AMT	1918
NEW_ACCTS_IN_LAST_SIX_MONTHS	26
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS	14
AVERAGE_ACCT_AGE	192
CREDIT_HISTORY_LENGTH	294
NO_OF_INQUIRIES	25
wiek	48
acc_age_rok	22
acc_age_mies	12
AVERAGE_ACCT_AGE_mth	192
CREDIT_HISTORY_LENGTH_mth	294
PERFORM_CNS_SCORE_DESCRIPTION2	6
dtype:	int64

In [95]:

```

1 zbiór.nunique()
2 #nazwa kolumny nie moze zawierac kropki, inaczej nunique nie dziala (trakruje wartosc )

```

Out[95]:

UniqueID	233154
disbursed_amount	24565
asset_cost	46252
ltv	6579
branch_id	82
supplier_id	2953
manufacturer_id	11
Current_pincode_ID	6698
Date_of_Birth	15433
Employment_Type	3
DisbursalDate	84
State_ID	22
Employee_code_ID	3270
MobileNo_Avl_Flag	1
Aadhar_flag	2
PAN_flag	2
VoterID_flag	2
Driving_flag	2
Passport_flag	2
PERFORM_CNS_SCORE	573
PERFORM_CNS_SCORE_DESCRIPTION	20
PRI_NO_OF_ACCTS	108
PRI_ACTIVE_ACCTS	40
PRI_OVERDUE_ACCTS	22
PRI_CURRENT_BALANCE	71341
PRI_SANCTIONED_AMOUNT	44390
PRI_DISBURSED_AMOUNT	47909
SEC_NO_OF_ACCTS	37
SEC_ACTIVE_ACCTS	23
SEC_OVERDUE_ACCTS	9
SEC_CURRENT_BALANCE	3246
SEC_SANCTIONED_AMOUNT	2223
SEC_DISBURSED_AMOUNT	2553
PRIMARY_INSTAL_AMT	28067
SEC_INSTAL_AMT	1918
NEW_ACCTS_IN_LAST_SIX_MONTHS	26
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS	14
AVERAGE_ACCT_AGE	192
CREDIT_HISTORY_LENGTH	294
NO_OF_INQUIRIES	25
loan_default	2
rok_ur2	48
wiek	48
acc_age_rok	22
acc_age_mies	12
AVERAGE_ACCT_AGE_mth	192
credit_hist_rok	35
credit_hist_mies	12
CREDIT_HISTORY_LENGTH_mth	294
SCORE_DESC_A_VLR	2
SCORE_DESC_B_VLR	2
SCORE_DESC_C_VLR	2
SCORE_DESC_D_VLR	2
SCORE_DESC_E_LR	2

```
SCORE_DESC_F_LR          2
SCORE_DESC_G_LR          2
SCORE_DESC_H_MR          2
SCORE_DESC_I_MR          2
SCORE_DESC_J_HR          2
SCORE_DESC_K_HR          2
SCORE_DESC_L_VHR         2
SCORE_DESC_M_VHR         2
SCORE_DESC_N1             2
SCORE_DESC_N2             2
SCORE_DESC_N3             2
SCORE_DESC_N4             2
SCORE_DESC_N5             2
SCORE_DESC_N6             2
SCORE_DESC_N7             2
PERFORM_CNS_SCORE_DESCRIPTION2 6
disbursed_amount_log     24565
asset_cost_log           46252
ltv_log                  6579
dtype: int64
```

In [115]:

```
1 Q1 = zbior.disbursed_amount_log.quantile(0,25)
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-115-f7fc07b8b23> in <module>
----> 1 Q1 = zbior.disbursed_amount_log.quantile(0,25)

~\Anaconda3\lib\site-packages\pandas\core\series.py in quantile(self, q, interpolation)
2397         df = self.to_frame()
2398
-> 2399         result = df.quantile(q=q, interpolation=interpolation, numeric_only=False)
2400         if result.ndim == 2:
2401             result = result.iloc[:, 0]

~\Anaconda3\lib\site-packages\pandas\core\frame.py in quantile(self, q, axis, numeric_only, interpolation)
8279
8280         result = data._data.quantile(
-> 8281             qs=q, axis=1, interpolation=interpolation, transposed=is_transposed
8282         )
8283

~\Anaconda3\lib\site-packages\pandas\core\internals\managers.py in quantile(self, axis, consolidate, transposed, interpolation, qs, numeric_only)
497         axes, blocks = [], []
498         for b in self.blocks:
--> 499             block = b.quantile(axis=axis, qs=qs, interpolation=interpolation)
500
501         axe = get_axe(b, qs, axes=self.axes)

~\Anaconda3\lib\site-packages\pandas\core\internals\blocks.py in quantile(self, qs, interpolation, axis)
1563             mask=mask,
1564             ndim=self.ndim,
-> 1565             interpolation=interpolation,
1566         )
1567

~\Anaconda3\lib\site-packages\pandas\core\nanops.py in nanpercentile(values, q, axis, na_value, mask, ndim, interpolation)
1413         return result
1414     else:
-> 1415         return np.percentile(values, q, axis=axis, interpolation=interpolation)

~\Anaconda3\lib\site-packages\numpy\lib\function_base.py in percentile(a, q, axis, out, overwrite_input, interpolation, keepdims)
3705         raise ValueError("Percentiles must be in the range [0, 100]")
)
3706     return _quantile_unchecked(
-> 3707         a, q, axis, out, overwrite_input, interpolation, keepdims)
3708
3709

~\Anaconda3\lib\site-packages\numpy\lib\function_base.py in _quantile_unchecked
```

```
ked(a, q, axis, out, overwrite_input, interpolation, keepdims)
3824     r, k = _ureduce(a, func=_quantile_ureduce_func, q=q, axis=axis,
out=out,
3825                     overwrite_input=overwrite_input,
-> 3826                     interpolation=interpolation)
3827     if keepdims:
3828         return r.reshape(q.shape + k)

~\Anaconda3\lib\site-packages\numpy\lib\function_base.py in _ureduce(a, fun
c, **kwargs)
3403     keepdim = (1,) * a.ndim
3404
-> 3405     r = func(a, **kwargs)
3406     return r, keepdim
3407

~\Anaconda3\lib\site-packages\numpy\lib\function_base.py in _quantile_ureduc
e_func(a, q, axis, out, overwrite_input, interpolation, keepdims)
3885     else:
3886         raise ValueError(
-> 3887             "interpolation can only be 'linear', 'lower' 'higher', "
3888             "'midpoint', or 'nearest'")
3889

ValueError: interpolation can only be 'linear', 'lower' 'higher', 'midpoin
t', or 'nearest'
```

In [249]:

```
1 korel = zbor3[['loan_default',
2                 'wiek',
3                 'Employment_Type',
4                 'Aadhar_flag',
5                 'PAN_flag',
6                 'VoterID_flag',
7                 'Driving_flag',
8                 'Passport_flag',
9                 'disbursed_amount',
10                'disbursed_amount_log',
11                'asset_cost',
12                'asset_cost_log',
13                'ltv',
14                'ltv_log',
15                'branch_id',
16                'supplier_id',
17                'manufacturer_id',
18                'State_ID',
19                'Employee_code_ID',
20                'PERFORM_CNS_SCORE',
21                'PRI_NO_OF_ACCTS',
22                'PRI_ACTIVE_ACCTS',
23                'PRI_OVERDUE_ACCTS',
24                'PRI_CURRENT_BALANCE',
25                'PRI_SANCTIONED_AMOUNT',
26                'PRI_DISBURSED_AMOUNT',
27                'SEC_NO_OF_ACCTS',
28                'SEC_ACTIVE_ACCTS',
29                'SEC_OVERDUE_ACCTS',
30                'SEC_CURRENT_BALANCE',
31                'SEC_SANCTIONED_AMOUNT',
32                'SEC_DISBURSED_AMOUNT',
33                'PRIMARY_INSTAL_AMT',
34                'SEC_INSTAL_AMT',
35                'NEW_ACCTS_IN_LAST_SIX_MONTHS',
36                'DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS',
37                'NO_OF_INQUIRIES',
38                'AVERAGE_ACCT_AGE_mth',
39                'CREDIT_HISTORY_LENGTH_mth',
40                'SCORE_DESC_A_VLR',
41                'SCORE_DESC_B_VLR',
42                'SCORE_DESC_C_VLR',
43                'SCORE_DESC_D_VLR',
44                'SCORE_DESC_E_LR',
45                'SCORE_DESC_F_LR',
46                'SCORE_DESC_G_LR',
47                'SCORE_DESC_H_MR',
48                'SCORE_DESC_I_MR',
49                'SCORE_DESC_J_HR',
50                'SCORE_DESC_K_HR',
51                'SCORE_DESC_L_VHR',
52                'SCORE_DESC_M_VHR',
53                'SCORE_DESC_N1',
54                'SCORE_DESC_N2',
55                'SCORE_DESC_N3',
56                'SCORE_DESC_N4',
57                'SCORE_DESC_N5',
58                'SCORE_DESC_N6',
59                'SCORE_DESC_N7', ]]
```



In [254]:

```

1 corr = korel.corr()
2 corr.style.background_gradient(cmap='coolwarm').set_precision(2)
3
4 #corr.info()

```

Out[254]:

	loan_default	wiek	Employment_Type	Aadhar_flag
loan_default	1	-0.037	0.025	
wiek	-0.037	1	0.17	
Employment_Type	0.025	0.17	1	
Aadhar_flag	-0.042	0.018		-0.094
PAN_flag	0.0023	-0.013		0.022
VoterID_flag	0.045	-0.016		0.093
Driving_flag	-0.0063	-0.0043		0.01
Passport_flag	-0.0074	-0.0071		-0.0046
disbursed_amount	0.089	-0.051		-0.048
disbursed_amount_log	0.091	-0.047		-0.049
asset_cost	0.0095	-0.13		0.018
asset_cost_log	0.0098	-0.13		0.016
ltv	0.1	0.088		-0.076
ltv_log	0.098	0.083		-0.072
branch_id	0.03	-0.0035		-0.0097
supplier_id	0.026	-0.033		0.079
manufacturer_id	-0.025	0.056		-0.027
State_ID	0.05	0.018		-0.056
Employee_code_ID	0.021	0.012		0.038
PERFORM_CNS_SCORE	-0.058	0.17		-0.009
PRI_NO_OF_ACCTS	-0.035	0.17		0.018
PRI_ACTIVE_ACCTS	-0.041	0.15		0.012
PRI_OVERDUE_ACCTS	0.04	0.14		0.032
PRI_CURRENT_BALANCE	-0.027	0.086		0.025
PRI_SANCTIONED_AMOUNT	-0.011	0.047		0.015
PRI_DISBURSED_AMOUNT	-0.011	0.047		0.015
SEC_NO_OF_ACCTS	-0.0081	0.023		0.001
SEC_ACTIVE_ACCTS	-0.0057	0.015		4.4e-05
SEC_OVERDUE_ACCTS	-0.0016	0.015		0.0026
SEC_CURRENT_BALANCE	-0.0053	0.01		0.00099
SEC_SANCTIONED_AMOUNT	-0.0061	0.012		0.0012
SEC_DISBURSED_AMOUNT	-0.006	0.012		0.0011

	loan_default	wiek	Employment_Type	Aadhar
<b>PRIMARY_INSTAL_AMT</b>	-0.01	0.052	0.0095	
<b>SEC_INSTAL_AMT</b>	-0.0014	0.0091	-0.0039	
<b>NEW_ACCTS_IN_LAST_SIX_MONTHS</b>	-0.029	0.034	0.0052	
<b>DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS</b>	0.035	0.08	0.015	
<b>NO_OF_INQUIRIES</b>	0.044	-0.00021	0.01	
<b>AVERAGE_ACCT_AGE_mth</b>	-0.025	0.25	0.021	
<b>CREDIT_HISTORY_LENGTH_mth</b>	-0.042	0.28	0.012	
<b>SCORE_DESC_A_VLR</b>	-0.031	0.035	-0.01	
<b>SCORE_DESC_B_VLR</b>	-0.042	0.054	-0.018	
<b>SCORE_DESC_C_VLR</b>	-0.029	0.025	-0.0059	
<b>SCORE_DESC_D_VLR</b>	-0.037	0.041	-0.02	
<b>SCORE_DESC_E_LR</b>	-0.017	0.035	-0.0038	
<b>SCORE_DESC_F_LR</b>	-0.015	0.028	0.0027	
<b>SCORE_DESC_G_LR</b>	-0.0058	0.037	0.0049	
<b>SCORE_DESC_H_MR</b>	0.011	0.031	0.0078	
<b>SCORE_DESC_I_MR</b>	0.021	0.031	0.0098	
<b>SCORE_DESC_J_HR</b>	0.011	0.035	0.01	
<b>SCORE_DESC_K_HR</b>	0.029	0.074	0.016	
<b>SCORE_DESC_L_VHR</b>	0.01	0.025	0.011	
<b>SCORE_DESC_M_VHR</b>	0.041	0.095	0.027	
<b>SCORE_DESC_N1</b>	0.035	-0.2	-0.0053	
<b>SCORE_DESC_N2</b>	0.013	-0.03	-0.0058	
<b>SCORE_DESC_N3</b>	-0.0023	-0.021	-0.0017	
<b>SCORE_DESC_N4</b>	-0.0094	0.041	0.0064	
<b>SCORE_DESC_N5</b>	-0.0051	0.041	0.0028	
<b>SCORE_DESC_N6</b>	-0.0018	0.015	0.013	
<b>SCORE_DESC_N7</b>	-0.0016	0.0029	0.0026	

In [257]:

```
1 corr.to_excel('kor.xlsx')
```

In [157]:

1 corr.info()

```
<class 'pandas.core.frame.DataFrame'>
Index: 59 entries, loan_default to SCORE_DESC_N7
Data columns (total 59 columns):
loan_default                59 non-null float64
wiek                         59 non-null float64
Employment_Type              59 non-null float64
Aadhar_flag                  59 non-null float64
PAN_flag                      59 non-null float64
VoterID_flag                 59 non-null float64
Driving_flag                  59 non-null float64
Passport_flag                 59 non-null float64
disbursed_amount              59 non-null float64
disbursed_amount_log          59 non-null float64
asset_cost                    59 non-null float64
asset_cost_log                59 non-null float64
ltv                           59 non-null float64
ltv_log                        59 non-null float64
branch_id                     59 non-null float64
supplier_id                   59 non-null float64
manufacturer_id               59 non-null float64
State_ID                       59 non-null float64
Employee_code_ID               59 non-null float64
PERFORM_CNS_SCORE              59 non-null float64
PRI_NO_OF_ACCTS                59 non-null float64
PRI_ACTIVE_ACCTS               59 non-null float64
PRI_OVERDUE_ACCTS              59 non-null float64
PRI_CURRENT_BALANCE             59 non-null float64
PRI_SANCTIONED_AMOUNT           59 non-null float64
PRI_DISBURSED_AMOUNT             59 non-null float64
SEC_NO_OF_ACCTS                59 non-null float64
SEC_ACTIVE_ACCTS               59 non-null float64
SEC_OVERDUE_ACCTS               59 non-null float64
SEC_CURRENT_BALANCE              59 non-null float64
SEC_SANCTIONED_AMOUNT             59 non-null float64
SEC_DISBURSED_AMOUNT              59 non-null float64
PRIMARY_INSTAL_AMT               59 non-null float64
SEC_INSTAL_AMT                  59 non-null float64
NEW_ACCTS_IN_LAST_SIX_MONTHS      59 non-null float64
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS 59 non-null float64
NO_OF_INQUIRIES                  59 non-null float64
AVERAGE_ACCT_AGE_mth              59 non-null float64
CREDIT_HISTORY_LENGTH_mth          59 non-null float64
SCORE_DESC_A_VLR                  59 non-null float64
SCORE_DESC_B_VLR                  59 non-null float64
SCORE_DESC_C_VLR                  59 non-null float64
SCORE_DESC_D_VLR                  59 non-null float64
SCORE_DESC_E_LR                  59 non-null float64
SCORE_DESC_F_LR                  59 non-null float64
SCORE_DESC_G_LR                  59 non-null float64
SCORE_DESC_H_MR                  59 non-null float64
SCORE_DESC_I_MR                  59 non-null float64
SCORE_DESC_J_HR                  59 non-null float64
SCORE_DESC_K_HR                  59 non-null float64
SCORE_DESC_L_VHR                  59 non-null float64
SCORE_DESC_M_VHR                  59 non-null float64
SCORE_DESC_N1                     59 non-null float64
SCORE_DESC_N2                     59 non-null float64
```

```
SCORE_DESC_N3          59 non-null float64
SCORE_DESC_N4          59 non-null float64
SCORE_DESC_N5          59 non-null float64
SCORE_DESC_N6          59 non-null float64
SCORE_DESC_N7          59 non-null float64
dtypes: float64(59)
memory usage: 30.2+ KB
```

In [255]:

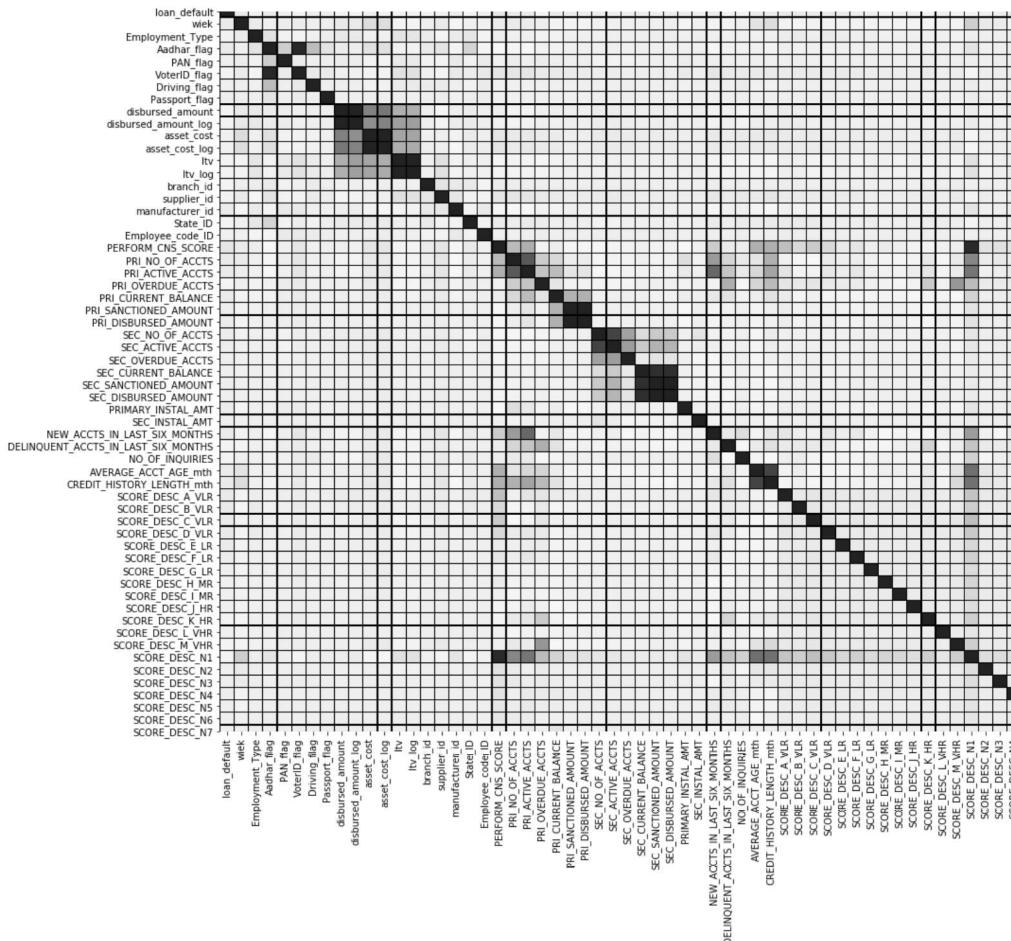
```
1 corr2 = pd.DataFrame(corr)
```

In [29]:

```
1 corr2.to_html('table.html')
```

In [256]:

```
1 #import seaborn as sns
2 plt.figure(figsize = (20, 14))
3 corr2 = sns.heatmap(corr, annot=False, linewidths=0.2, linecolor = 'black', cmap='RdBu')
4 plt.savefig('1.png')
```

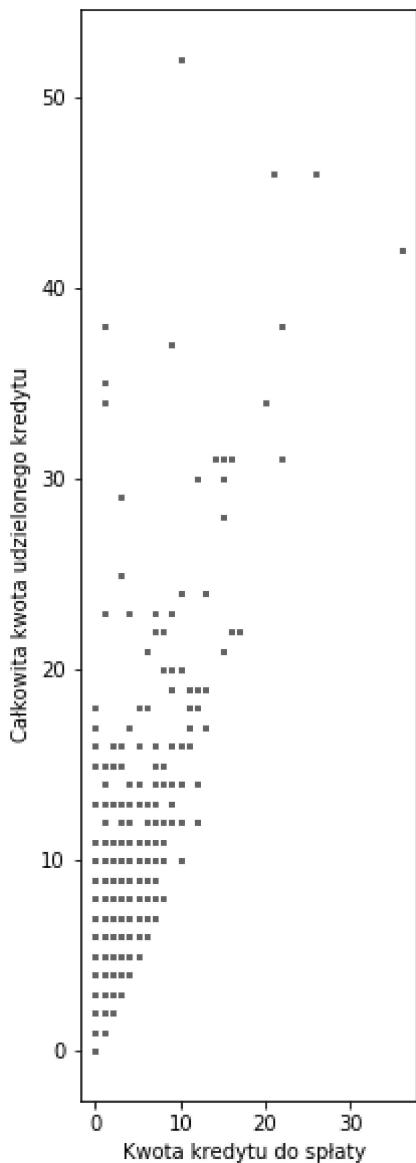


In [272]:

```
1 zbor.to_excel('zbor.xlsx')
```

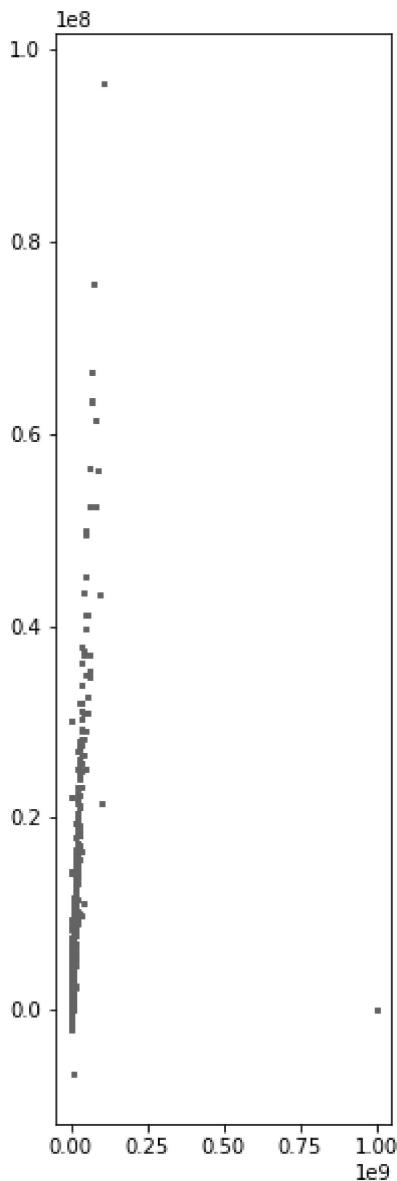
In [302]:

```
1 fig, ax = plt.subplots()
2 plt.scatter(zbior.SEC_ACTIVE_ACCTS , zbior.SEC_NO_OF_ACCTS , s=5, norm=False, alpha=1
3 #ax.legend()
4 #ax.add_artist()
5 plt.ylabel('Całkowita kwota udzielonego kredytu')
6 plt.xlabel('Kwota kredytu do spłaty ')
7 plt.gcf().set_size_inches((3,10))
```



In [86]:

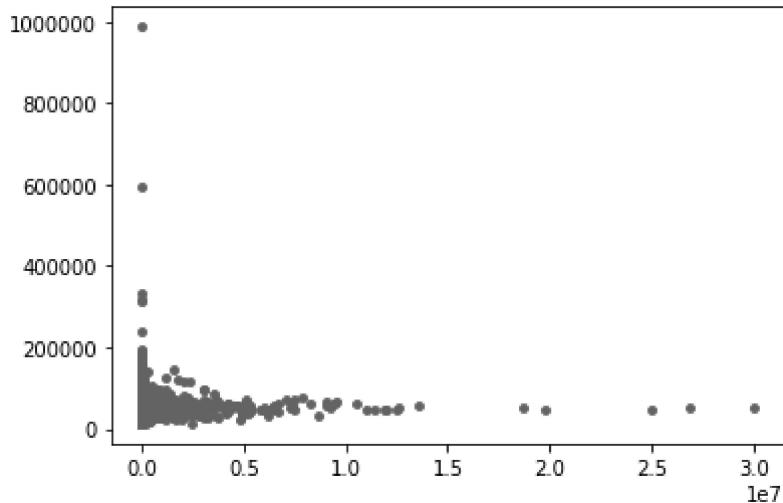
```
1 plt.scatter(zbior.PRI_DISBURSED_AMOUNT, zbior.PRI_CURRENT_BALANCE, s=5, norm=False, alpha=0.5)
2 plt.gcf().set_size_inches((3,10))
```



In [107]:

```
1 #plt.figure(figsize=(30,10))
2 fig, ax = plt.scatter(zbior.SEC_DISBURSED_AMOUNT, zbior.disbursed_amount, s=15, norm=False)
3 #subplot.axes.ticklabel_format(style='plain', useOffset=False, axis='both')
4 #ax.ticklabel_format(style='plain')
5 #plt.title("xxx")
6 #plt.gcf().set_size_inches((10,10))
```

```
-----  
TypeError                                     Traceback (most recent call last)  
<ipython-input-107-7e44c045b93c> in <module>  
      1 #plt.figure(figsize=(30,10))  
----> 2 fig, ax = plt.scatter(zbior.SEC_DISBURSED_AMOUNT, zbior.disbursed_amount, s=15, norm=False, alpha=1, marker="o")  
      3 #subplot.axes.ticklabel_format(style='plain', useOffset=False, axis='both')  
      4 #ax.ticklabel_format(style='plain')  
      5 #plt.title("xxx")  
  
TypeError: cannot unpack non-iterable PathCollection object
```



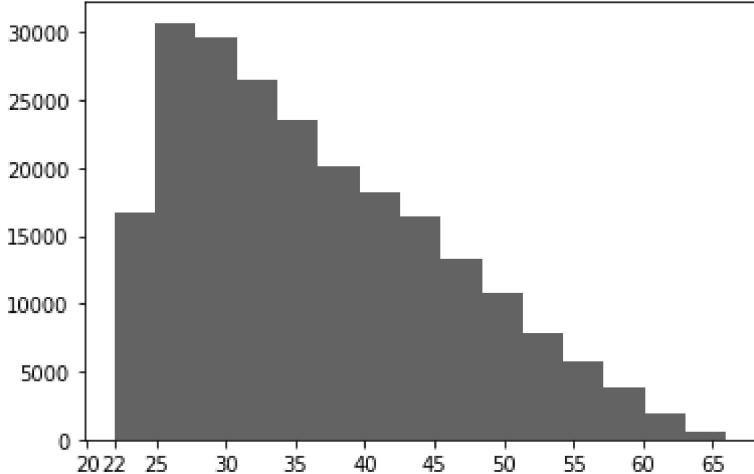
In [51]:

```
1 plt.hist(zbior.wiek2, bins=15)
2 plt.xticks([20, 25, 30, 35, 40, 45, 50, 55, 60, 65])
3 plt.title('Wiek klientów')
```

Out[51]:

Text(0.5, 1.0, 'Wiek klientów')

Wiek klientów

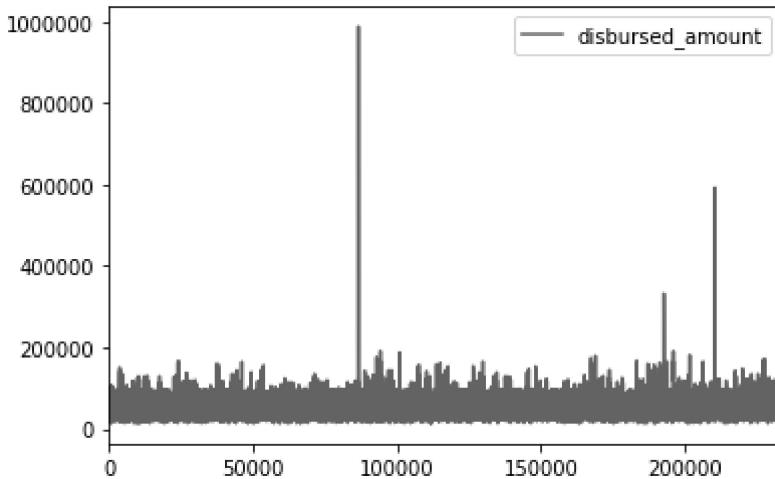


In [86]:

```
1 zbior[['disbursed_amount']].plot()
```

Out[86]:

&lt;matplotlib.axes.\_subplots.AxesSubplot at 0x1f10a90f9c8&gt;



In [123]:

```
1 corr.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 13 entries, disbursed_amount to SEC_SANCTIONED_AMOUNT
Data columns (total 13 columns):
disbursed_amount           13 non-null float64
wiek2                      13 non-null float64
asset_cost                  13 non-null float64
loan_default                13 non-null float64
CREDIT_HISTORY_LENGTH_mth   13 non-null float64
AVERAGE_ACCT_AGE_mth       13 non-null float64
NO_OF_INQUIRIES             13 non-null float64
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS 13 non-null float64
NEW_ACCTS_IN_LAST_SIX_MONTHS    13 non-null float64
SEC_INSTAL_AMT              13 non-null float64
PRIMARY_INSTAL_AMT          13 non-null float64
SEC_DISBURSED_AMOUNT         13 non-null float64
SEC_SANCTIONED_AMOUNT        13 non-null float64
dtypes: float64(13)
memory usage: 2.0+ KB
```

In [8]:

```
1 zbiор2 = zbiор[['UniqueID', 'Date_of_Birth']]
2 zbiор2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 233154 entries, 0 to 233153
Data columns (total 2 columns):
UniqueID      233154 non-null int64
Date_of_Birth 233154 non-null object
dtypes: int64(1), object(1)
memory usage: 3.6+ MB
```

In [37]:

```
1 zbor2.sort_values('Date_of_Birth')
```

Out[37]:

	UniqueID	Date_of_Birth	count
90910	448427	01-01-00	1
55521	592974	01-01-00	1
170105	432993	01-01-00	1
61660	632098	01-01-00	1
147799	562688	01-01-00	1
...	...	...	...
224088	551003	31-12-99	0
133150	422086	31-12-99	0
97330	579575	31-12-99	0
110565	470782	31-12-99	0
144455	653306	31-12-99	0

233154 rows × 3 columns

In [38]:

```
1 zbor2.to_excel('zbor2.xlsx')
```

In [14]:

```
1 count = zbor2.groupby(['Date_of_Birth']).count()
2 print(count)
```

Date_of_Birth	UniqueID	count
01-01-00	135	135
01-01-55	22	22
01-01-56	75	75
01-01-57	95	95
01-01-58	173	173
01-01-59	213	213
01-01-60	379	379
01-01-61	358	358
01-01-62	364	364
01-01-63	439	439
01-01-64	441	441
01-01-65	702	702
01-01-66	642	642
01-01-67	683	683
01-01-68	911	911
01-01-69	875	875
01-01-70	1203	1203
01-01-71	1142	1142

In [54]:

```
1 count = count_df.groupby(['Date_of_Birth']).count()
2 print(count)
```

Date_of_Birth	UniqueID	count
01-01-00	1	1
01-01-55	1	1
01-01-56	1	1
01-01-57	1	1
01-01-58	1	1
01-01-59	1	1
01-01-60	1	1
01-01-61	1	1
01-01-62	1	1
01-01-63	1	1
01-01-64	1	1
01-01-65	1	1
01-01-66	1	1
01-01-67	1	1
01-01-68	1	1
01-01-69	1	1
01-01-70	1	1
.. .. ..	..	..

In [48]:

```
1 count_df = pd.DataFrame(count)
```

In [51]:

```
1 count_df = count_df.reset_index()
```

In [55]:

```
1 sort_by = count_df.sort_values('count', ascending=False)
```

In [64]:

```
1 top138 = sort_by.head(138)
2 print(top138)
```

	Date_of_Birth	UniqueID	count
34	01-01-88	2173	2173
36	01-01-90	2170	2170
33	01-01-87	2127	2127
32	01-01-86	2063	2063
31	01-01-85	2005	2005
37	01-01-91	1985	1985
35	01-01-89	1962	1962
39	01-01-93	1930	1930
38	01-01-92	1924	1924
41	01-01-95	1924	1924
40	01-01-94	1867	1867
29	01-01-83	1841	1841
26	01-01-80	1821	1821
30	01-01-84	1768	1768
27	01-01-81	1764	1764
42	01-01-96	1759	1759
28	01-01-82	1691	1691
24	01-01-78	1646	1646
..	.. .. ..	.. .. ..	.. .. ..

In [68]:

```
1 top138.info()
2 top138.to_excel('top138.xlsx')
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 138 entries, 34 to 76
Data columns (total 3 columns):
Date_of_Birth    138 non-null object
UniqueID         138 non-null int64
count            138 non-null int64
dtypes: int64(2), object(1)
memory usage: 4.3+ KB
```

In [67]:

```
1 zbor3 = zbor([zbor.Date_of_Birth != top138.Date_of_Birth])
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-67-b5f4d7899298> in <module>
----> 1 zbor3 = zbor([zbor.Date_of_Birth != top138.Date_of_Birth])

~\Anaconda3\lib\site-packages\pandas\core\ops\__init__.py in wrapper(self, other, axis)
    1140
    1141      elif isinstance(other, ABCSeries) and not self._indexed_same
(other):
-> 1142          raise ValueError("Can only compare identically-labeled "
"Series objects")
    1143
    1144      elif is_categorical_dtype(self):
```

ValueError: Can only compare identically-labeled Series objects

In [53]:

```
1 count_df.to_excel('count.xlsx')
```

In [61]:

```
1 count_sorted = count.sort_values('count', ascending = False)
2 count_sorted.head()
```

Out[61]:

Date_of_Birth	UniqueID	count
01-01-00	1	1
20-12-63	1	1
20-12-95	1	1
20-12-96	1	1
20-12-97	1	1

In [35]:

```
1 count.info()
```

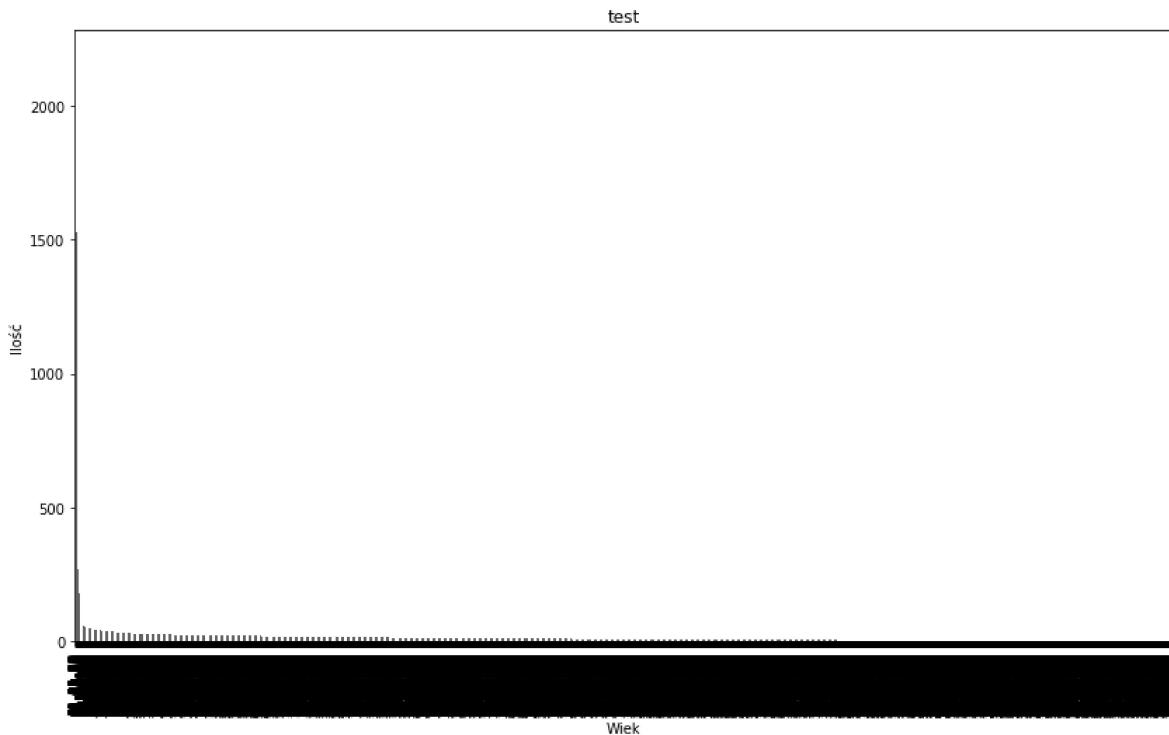
```
<class 'pandas.core.frame.DataFrame'>
Index: 15433 entries, 01-01-00 to 31-12-99
Data columns (total 2 columns):
UniqueID    15433 non-null int64
count       15433 non-null int64
dtypes: int64(2)
memory usage: 361.7+ KB
```

In [28]:

```
1 wykr = zbior['Date_of_Birth'].value_counts().plot(kind='bar', figsize=(10,8), title="t")
2 wykr.set_xlabel("Wiek")
3 wykr.set_ylabel("Ilość")
4 wykr()
```

```
-----
TypeError                                                 Traceback (most recent call last)
<ipython-input-28-8b846d2e3378> in <module>
      2 wykr.set_xlabel("Wiek")
      3 wykr.set_ylabel("Ilość")
----> 4 wykr()

TypeError: 'AxesSubplot' object is not callable
```



In [53]:

```
1 defaults = zbior[zbior['loan_default'] == 1]
2 non-defaults = zbior[zbior['loan_default'] == 0]
3 #print(defaults)
4 #defaults.to_excel("defaults.xlsx")
5 # tworzy oddzielny plik z defaultami
```

In [67]:

```
1 defaults.head()
2 defaults.describe()
```

Out[67]:

	UniqueID	disbursed_amount	asset_cost	ltv	branch_id	supplier_id	manufacture
count	48967.00	48967.00	48967.00	48967.00	48967.00	48967.00	48967.00
mean	540045.62	56163.95	76180.60	76.94	77.13	19828.49	68
std	70841.11	12099.97	18592.90	10.32	72.22	3455.09	22
min	417439.00	13369.00	37000.00	15.30	1.00	10524.00	45
25%	477396.00	49303.00	65800.00	72.13	16.00	16680.00	45
50%	544061.00	55359.00	71205.00	79.12	64.00	20672.00	51
75%	603702.00	61923.00	79950.00	84.71	135.00	23118.00	86
max	658671.00	191392.00	281164.00	95.00	261.00	24803.00	153

In [135]:

```
1 wyliczenia = zbior[[ 'PRI_NO_OF_ACCTS', 'PRI_ACTIVE_ACCTS', 'PRI_OVERDUE_ACCTS', 'PRI_O
```

In [136]:

```
1 wyliczenia.to_excel('wyliczenia.xlsx')
```

In [117]:

```
1 zbior.wiek.describe()
```

Out[117]:

```
count    233154.00
mean      36.10
std       9.81
min      20.00
25%      28.00
50%      34.00
75%      43.00
max      71.00
Name: wiek, dtype: float64
```

In [118]:

```
1 zbior.disbursed_amount_log.describe()
```

Out[118]:

```
count    233154.00
mean      10.88
std       0.23
min       9.50
25%      10.76
50%      10.89
75%      11.01
max      13.81
Name: disbursed_amount_log, dtype: float64
```

In [15]:

```
1 non_defaults = zbior[zbior['loan_default']==0]
2 non_defaults.to_excel("non_defaults.xlsx")
```

In [258]:

```
1 sns.pairplot(zbior, vars=['PRI_DISBURSED_AMOUNT', 'SEC_DISBURSED_AMOUNT', 'SEC_SANCTIONED_AMOU
```

Out[258]:

```
<seaborn.axisgrid.PairGrid at 0x22e075c7c08>
```

```
Error in callback <function flush_figures at 0x0000022E73D9C798> (for post
_execute):
```

In [88]:

```
1 sns.pairplot(korel)
```

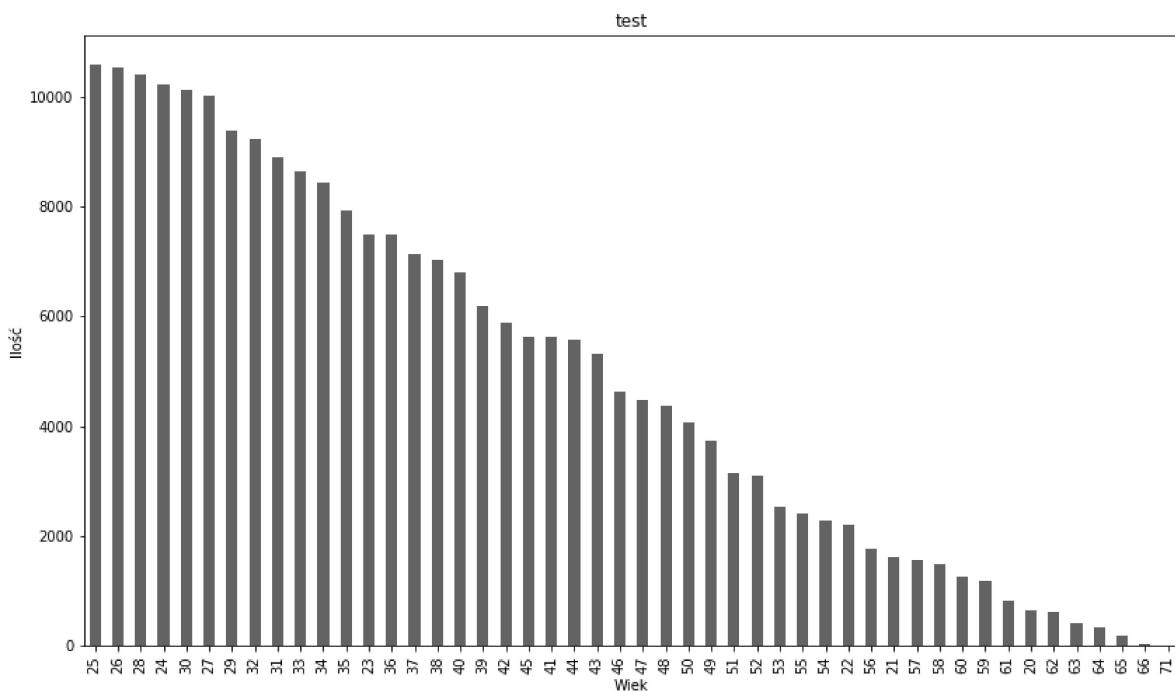
```
-----  
-  
KeyboardInterrupt                                     Traceback (most recent call last)  
<ipython-input-88-4db852752a8f> in <module>  
----> 1 sns.pairplot(korel)  
  
~\Anaconda3\lib\site-packages\seaborn\axisgrid.py in pairplot(data, hue, hue_order, palette, vars, x_vars, y_vars, kind, diag_kind, markers, height, aspect, dropna, plot_kws, diag_kws, grid_kws, size)  
    2119         if kind == "scatter":  
    2120             from .relational import scatterplot # Avoid circular import  
rt  
-> 2121         plotter(scatterplot, **plot_kws)  
    2122     elif kind == "reg":  
    2123         from .regression import regplot # Avoid circular import  
  
~\Anaconda3\lib\site-packages\seaborn\axisgrid.py in map_offdiag(self, fun  
c, **kwargs)  
    4744
```

In [10]:

```
1 Score = zbior['wiek'].value_counts().plot(kind='bar', figsize=(14,8), title="test", x :
2 Score.set_xlabel("Wiek")
3 Score.set_ylabel("Ilość")
4 Score()
```

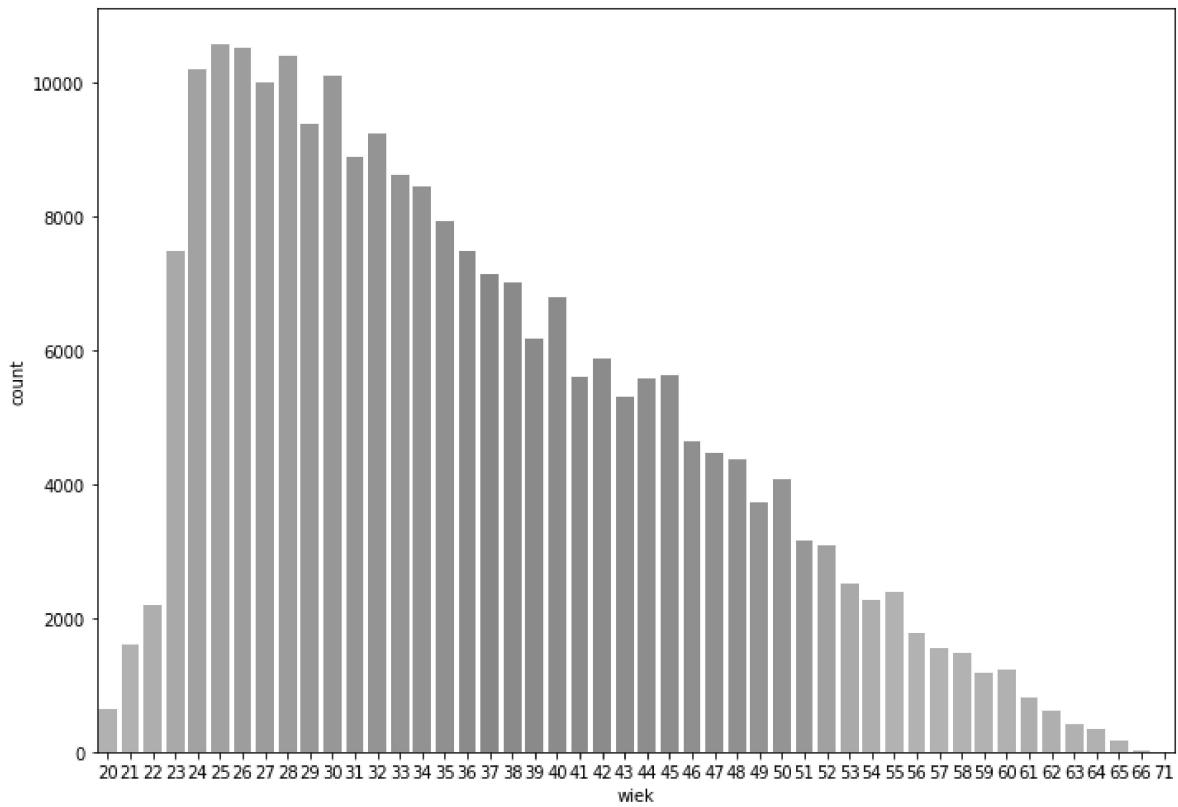
```
-----
TypeError                                                 Traceback (most recent call last)
<ipython-input-10-a61e20eacb36> in <module>
      2 Score.set_xlabel("Wiek")
      3 Score.set_ylabel("Ilość")
----> 4 Score()
```

```
TypeError: 'AxesSubplot' object is not callable
```



In [114]:

```
1 figsize = (40,20)
2 sns.countplot(zbior.wiek)
3 plt.gcf().set_size_inches(11.7, 8.27)
```



In [105]:

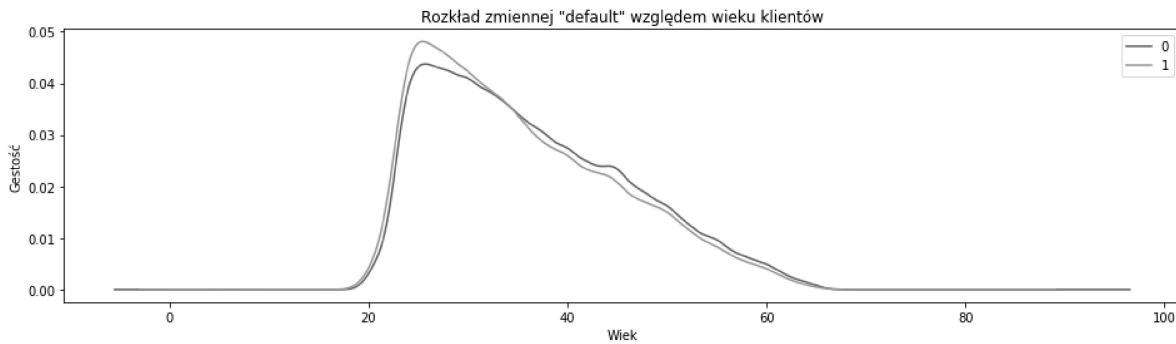
```

1 #zbior.wiek2.plot(kind='kde')
2 zbior.groupby('loan_default').wiek.plot(kind='kde')
3 plt.legend()
4 plt.ylabel("Gęstość")
5 plt.xlabel('Wiek')
6 plt.title('Rozkład zmiennej "default" względem wieku klientów')

```

Out[105]:

Text(0.5, 1.0, 'Rozkład zmiennej "default" względem wieku klientów')



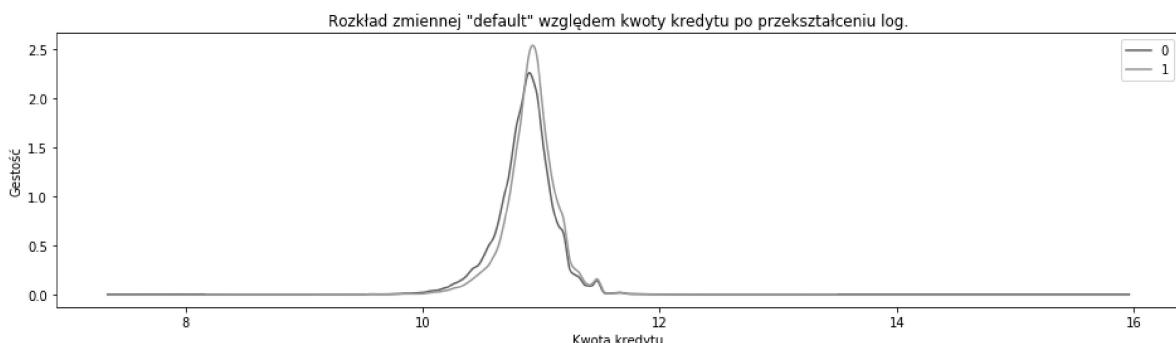
In [111]:

```

1 zbior.groupby('loan_default').disbursed_amount_log.plot(kind='kde')
2 plt.legend()
3 plt.ylabel("Gęstość")
4 plt.xlabel('Kwota kredytu')
5 plt.title('Rozkład zmiennej "default" względem kwoty kredytu po przekształceniu log.')

```

Out[111]:

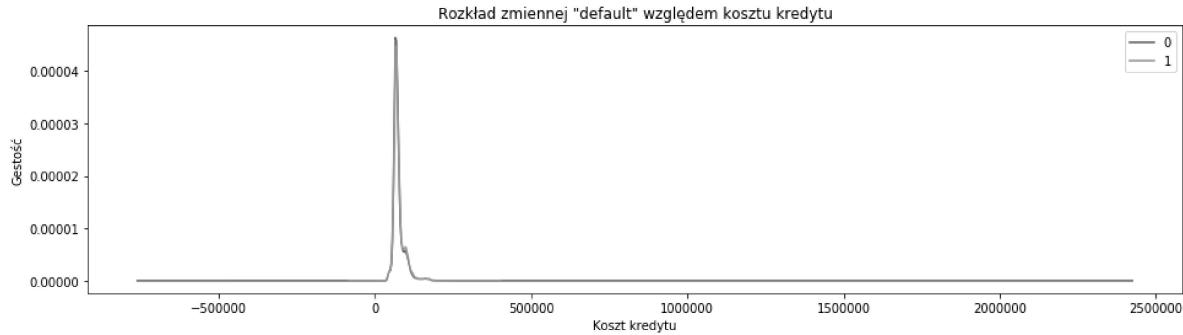
Text(0.5, 1.0, 'Rozkład zmiennej "default" względem kwoty kredytu po przekształceniu log.')  


In [108]:

```
1 zbiór.groupby('loan_default').asset_cost.plot(kind='kde')
2 plt.legend()
3 plt.ylabel("Gęstość")
4 plt.xlabel('Koszt kredytu')
5 plt.title('Rozkład zmiennej "default" względem kosztu kredytu')
```

Out[108]:

Text(0.5, 1.0, 'Rozkład zmiennej "default" względem kosztu kredytu')

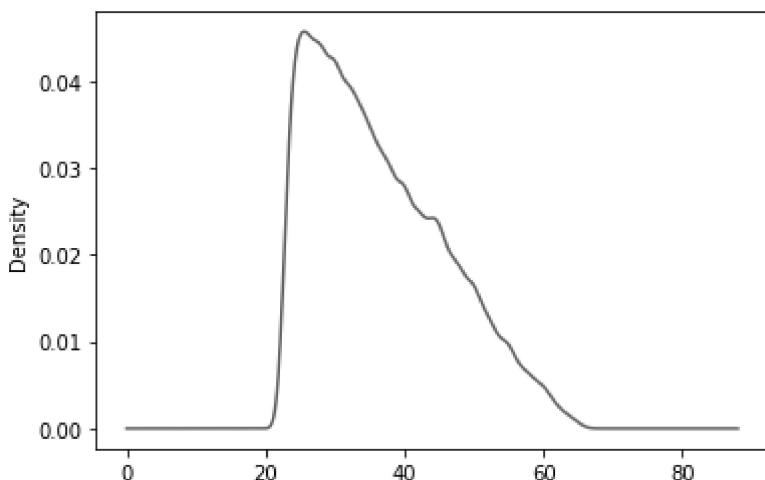


In [17]:

```
1 zbiór.groupby('MobileNo_Avl_Flag').wiek2.plot(kind='kde')
```

Out[17]:

```
MobileNo_Avl_Flag
1 AxesSubplot(0.125,0.125;0.775x0.755)
Name: wiek2, dtype: object
```

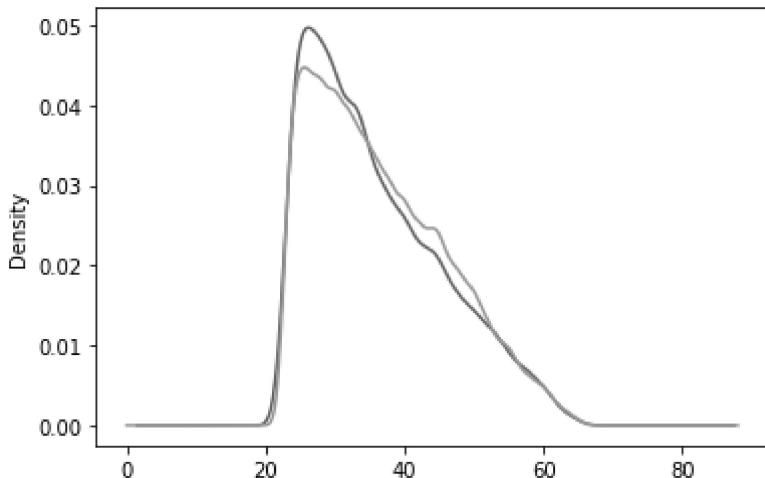


In [18]:

```
1 zbiór.groupby('Aadhar_flag').wiek2.plot(kind='kde')
```

Out[18]:

Aadhar\_flag  
0 AxesSubplot(0.125,0.125;0.775x0.755)  
1 AxesSubplot(0.125,0.125;0.775x0.755)  
Name: wiek2, dtype: object

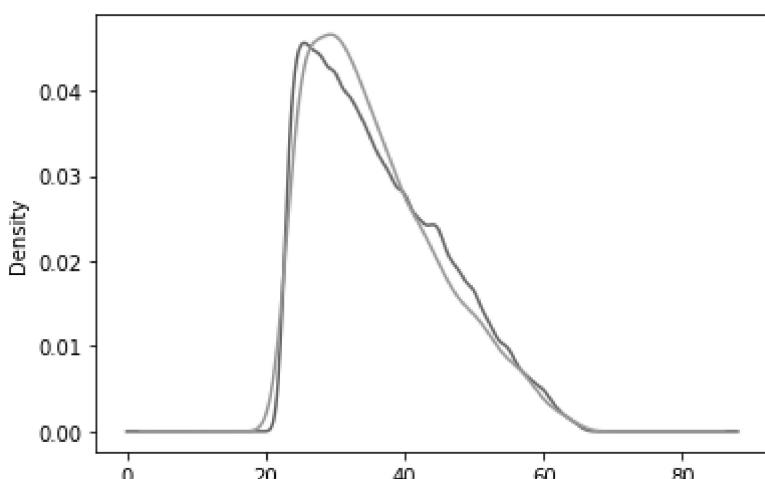


In [19]:

```
1 zbiór.groupby('Driving_flag').wiek2.plot(kind='kde')
```

Out[19]:

Driving\_flag  
0 AxesSubplot(0.125,0.125;0.775x0.755)  
1 AxesSubplot(0.125,0.125;0.775x0.755)  
Name: wiek2, dtype: object

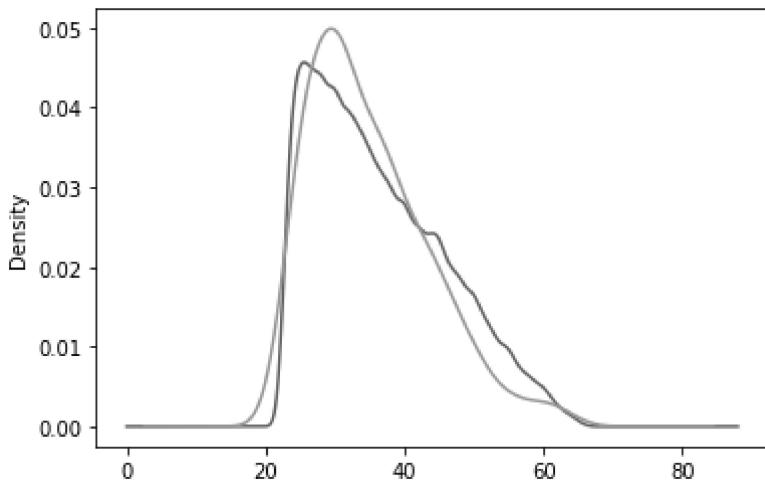


In [20]:

```
1 zbor.groupby('Passport_flag').wiek2.plot(kind='kde')
```

Out[20]:

Passport\_flag  
0 AxesSubplot(0.125,0.125;0.775x0.755)  
1 AxesSubplot(0.125,0.125;0.775x0.755)  
Name: wiek2, dtype: object

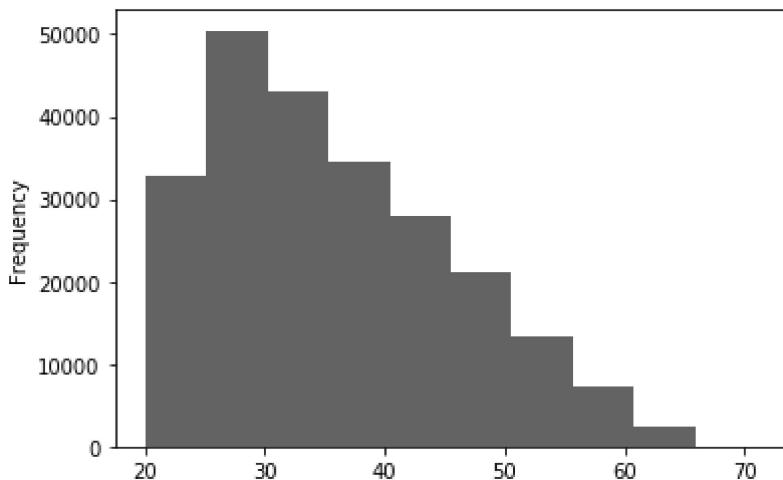


In [12]:

```
1 zbor['wiek'].plot.hist()  
2 #zbor['loan_default'].plot.hist()
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x22e014c6d08>
```

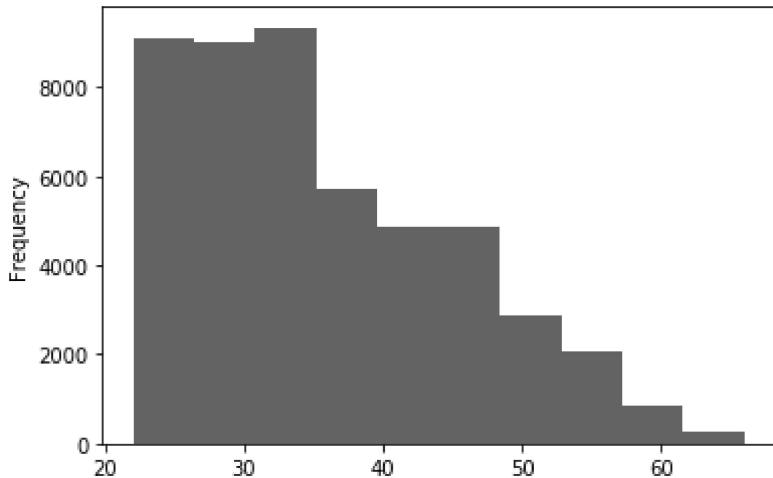


In [22]:

```
1 zbor["wiek2"].plot(kind="hist", weights=zbor["loan_default"])
```

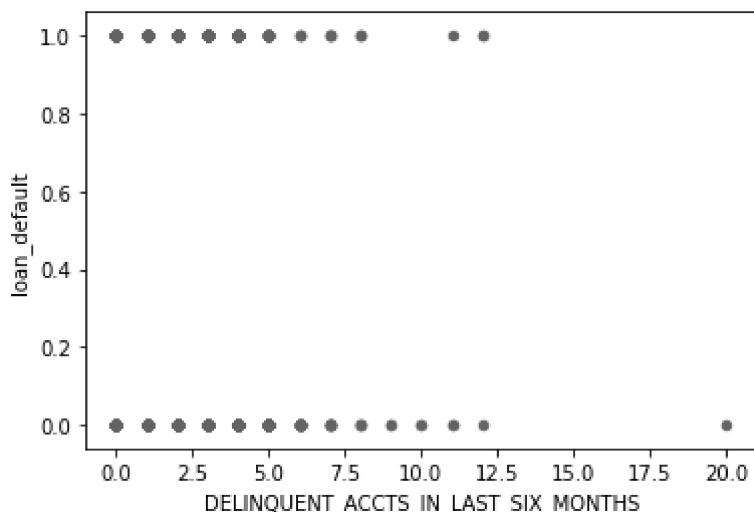
Out[22]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a960b3f0c8>
```



In [31]:

```
1 zbor.plot.scatter('DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS', 'loan_default')
2 #plt.xscale('Log')
3 #plt.xticks([10, 100, 1000])
4 plt.show()
```



In [218]:

```

1 Q1 = zbior.disbursed_amount_log.quantile(0.25) #10.76
2 Q3 = zbior.disbursed_amount_log.quantile(0.75) #11.00
3 IQR = Q3 - Q1 #0.24
4 O1 = Q1 -1.5*0.24
5 O2 = Q3 + 1.5*0.24
6 print(O1, O2)

```

10.401004448831893 11.36897614514367

In [226]:

```
1 zbior2 = zbior[zbior.disbursed_amount_log > 10.40]
```

In [228]:

```

1 zbior3 = zbior2[zbior2.disbursed_amount_log < 11.36]
2 # -

```

In [232]:

```

1 o = 233154 - 220349
2 o #12805

```

Out[232]:

12805

In [243]:

```

1 Q1 = zbior.disbursed_amount.quantile(0.25) #47145.0
2 Q3 = zbior.disbursed_amount.quantile(0.75) #60413.0
3 IQR = Q3 - Q1 #13268
4 O1 = Q1 -1.5*13268 #27243
5 O2 = Q3 + 1.5*13268 #80315
6 print(O1, O2)

```

27243.0 80315.0

In [244]:

```
1 zbior2 = zbior[zbior.disbursed_amount > 27243]
```

In [246]:

```
1 zbior3 = zbior2[zbior2.disbursed_amount < 80315]
```

In [248]:

```

1 o = 233154 - 223153
2 o #10001

```

Out[248]:

10001

In [247]:

1 zbiror3.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 223285 entries, 0 to 233153
Data columns (total 73 columns):
UniqueID                      223285 non-null int64
disbursed_amount               223285 non-null int64
asset_cost                     223285 non-null int64
ltv                            223285 non-null float64
branch_id                      223285 non-null int64
supplier_id                    223285 non-null int64
manufacturer_id                223285 non-null int64
Current_pincode_ID             223285 non-null int64
Date_of_Birth                  223285 non-null object
Employment_Type                223285 non-null float64
DisbursalDate                 223285 non-null object
State_ID                       223285 non-null int64
Employee_code_ID               223285 non-null int64
MobileNo_Avl_Flag              223285 non-null int64
Aadhar_flag                    223285 non-null int64
PAN_flag                       223285 non-null int64
VoterID_flag                  223285 non-null int64
Driving_flag                   223285 non-null int64
Passport_flag                  223285 non-null int64
PERFORM_CNS_SCORE              223285 non-null int64
PERFORM_CNS_SCORE_DESCRIPTION  223285 non-null object
PRI_NO_OF_ACCTS                223285 non-null int64
PRI_ACTIVE_ACCTS               223285 non-null int64
PRI_OVERDUE_ACCTS              223285 non-null int64
PRI_CURRENT_BALANCE            223285 non-null int64
PRI_SANCTIONED_AMOUNT          223285 non-null int64
PRI_DISBURSED_AMOUNT           223285 non-null int64
SEC_NO_OF_ACCTS                223285 non-null int64
SEC_ACTIVE_ACCTS               223285 non-null int64
SEC_OVERDUE_ACCTS              223285 non-null int64
SEC_CURRENT_BALANCE            223285 non-null int64
SEC_SANCTIONED_AMOUNT          223285 non-null int64
SEC_DISBURSED_AMOUNT           223285 non-null int64
PRIMARY_INSTAL_AMT             223285 non-null int64
SEC_INSTAL_AMT                 223285 non-null int64
NEW_ACCTS_IN_LAST_SIX_MONTHS   223285 non-null int64
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS 223285 non-null int64
AVERAGE_ACCT_AGE               223285 non-null object
CREDIT_HISTORY_LENGTH            223285 non-null object
NO_OF_INQUIRIES                 223285 non-null int64
loan_default                   223285 non-null int64
rok_ur2                        223285 non-null int32
wiek                           223285 non-null int64
acc_age_rok                    223285 non-null int32
acc_age_mies                   223285 non-null int32
AVERAGE_ACCT_AGE_mth           223285 non-null int32
credit_hist_rok                 223285 non-null int32
credit_hist_mies                223285 non-null int32
CREDIT_HISTORY_LENGTH_mth       223285 non-null int32
SCORE_DESC_A_VLR                223285 non-null float64
SCORE_DESC_B_VLR                223285 non-null float64
SCORE_DESC_C_VLR                223285 non-null float64
SCORE_DESC_D_VLR                223285 non-null float64
SCORE_DESC_E_LR                 223285 non-null float64
```

```
SCORE_DESC_F_LR          223285 non-null float64
SCORE_DESC_G_LR          223285 non-null float64
SCORE_DESC_H_MR          223285 non-null float64
SCORE_DESC_I_MR          223285 non-null float64
SCORE_DESC_J_HR          223285 non-null float64
SCORE_DESC_K_HR          223285 non-null float64
SCORE_DESC_L_VHR         223285 non-null float64
SCORE_DESC_M_VHR         223285 non-null float64
SCORE_DESC_N1             223285 non-null float64
SCORE_DESC_N2             223285 non-null float64
SCORE_DESC_N3             223285 non-null float64
SCORE_DESC_N4             223285 non-null float64
SCORE_DESC_N5             223285 non-null float64
SCORE_DESC_N6             223285 non-null float64
SCORE_DESC_N7             223285 non-null float64
PERFORM_CNS_SCORE_DESCRIPTION2 223285 non-null object
disbursed_amount_log      223285 non-null float64
asset_cost_log            223285 non-null float64
ltv_log                   223285 non-null float64
dtypes: float64(25), int32(7), int64(35), object(6)
memory usage: 120.1+ MB
```

In [206]:

```
1 zbiор['disbursed_amount_log'].describe()
```

Out[206]:

```
count    233154.00
mean      10.88
std       0.23
min       9.50
25%      10.76
50%      10.89
75%      11.01
max      13.81
Name: disbursed_amount_log, dtype: float64
```

In [196]:

1 disbursed\_amount\_log\_o.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 0 entries
Data columns (total 73 columns):
UniqueID                      0 non-null int64
disbursed_amount               0 non-null int64
asset_cost                      0 non-null int64
ltv                            0 non-null float64
branch_id                       0 non-null int64
supplier_id                     0 non-null int64
manufacturer_id                0 non-null int64
Current_pincode_ID              0 non-null int64
Date_of_Birth                   0 non-null object
Employment_Type                 0 non-null float64
DisbursalDate                  0 non-null object
State_ID                        0 non-null int64
Employee_code_ID                0 non-null int64
MobileNo_Avl_Flag               0 non-null int64
Aadhar_flag                      0 non-null int64
PAN_flag                         0 non-null int64
VoterID_flag                    0 non-null int64
Driving_flag                     0 non-null int64
Passport_flag                   0 non-null int64
PERFORM_CNS_SCORE               0 non-null int64
PERFORM_CNS_SCORE_DESCRIPTION    0 non-null object
PRI_NO_OF_ACCTS                 0 non-null int64
PRI_ACTIVE_ACCTS                0 non-null int64
PRI_OVERDUE_ACCTS               0 non-null int64
PRI_CURRENT_BALANCE              0 non-null int64
PRI_SANCTIONED_AMOUNT            0 non-null int64
PRI_DISBURSED_AMOUNT              0 non-null int64
SEC_NO_OF_ACCTS                 0 non-null int64
SEC_ACTIVE_ACCTS                0 non-null int64
SEC_OVERDUE_ACCTS               0 non-null int64
SEC_CURRENT_BALANCE              0 non-null int64
SEC_SANCTIONED_AMOUNT             0 non-null int64
SEC_DISBURSED_AMOUNT              0 non-null int64
PRIMARY_INSTAL_AMT               0 non-null int64
SEC_INSTAL_AMT                  0 non-null int64
NEW_ACCTS_IN_LAST_SIX_MONTHS     0 non-null int64
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS 0 non-null int64
AVERAGE_ACCT_AGE                 0 non-null object
CREDIT_HISTORY_LENGTH              0 non-null object
NO_OF_INQUIRIES                  0 non-null int64
loan_default                     0 non-null int64
rok_ur2                          0 non-null int32
wiek                            0 non-null int64
acc_age_rok                      0 non-null int32
acc_age_mies                     0 non-null int32
AVERAGE_ACCT_AGE_mth              0 non-null int32
credit_hist_rok                  0 non-null int32
credit_hist_mies                  0 non-null int32
CREDIT_HISTORY_LENGTH_mth          0 non-null int32
SCORE_DESC_A_VLR                  0 non-null float64
SCORE_DESC_B_VLR                  0 non-null float64
SCORE_DESC_C_VLR                  0 non-null float64
SCORE_DESC_D_VLR                  0 non-null float64
SCORE_DESC_E_LR                  0 non-null float64
```

```
SCORE_DESC_F_LR          0 non-null float64
SCORE_DESC_G_LR          0 non-null float64
SCORE_DESC_H_MR          0 non-null float64
SCORE_DESC_I_MR          0 non-null float64
SCORE_DESC_J_HR          0 non-null float64
SCORE_DESC_K_HR          0 non-null float64
SCORE_DESC_L_VHR         0 non-null float64
SCORE_DESC_M_VHR         0 non-null float64
SCORE_DESC_N1             0 non-null float64
SCORE_DESC_N2             0 non-null float64
SCORE_DESC_N3             0 non-null float64
SCORE_DESC_N4             0 non-null float64
SCORE_DESC_N5             0 non-null float64
SCORE_DESC_N6             0 non-null float64
SCORE_DESC_N7             0 non-null float64
PERFORM_CNS_SCORE_DESCRIPTION2 0 non-null object
disbursed_amount_log     0 non-null float64
asset_cost_log           0 non-null float64
ltv_log                  0 non-null float64
dtypes: float64(25), int32(7), int64(35), object(6)
memory usage: 0.0+ bytes
```

In [175]:

1 zbiór.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 233154 entries, 0 to 233153
Data columns (total 73 columns):
UniqueID           233154 non-null int64
disbursed_amount  233154 non-null int64
asset_cost         233154 non-null int64
ltv               233154 non-null float64
branch_id          233154 non-null int64
supplier_id        233154 non-null int64
manufacturer_id   233154 non-null int64
Current_pincode_ID 233154 non-null int64
Date_of_Birth     233154 non-null object
Employment_Type   233154 non-null float64
DisbursalDate    233154 non-null object
State_ID          233154 non-null int64
Employee_code_ID  233154 non-null int64
MobileNo_Avl_Flag 233154 non-null int64
Aadhar_flag       233154 non-null int64
PAN_flag          233154 non-null int64
VoterID_flag     233154 non-null int64
Driving_flag      233154 non-null int64
Passport_flag    233154 non-null int64
PERFORM_CNS_SCORE 233154 non-null int64
PERFORM_CNS_SCORE_DESCRIPTION 233154 non-null object
PRI_NO_OF_ACCTS  233154 non-null int64
PRI_ACTIVE_ACCTS 233154 non-null int64
PRI_OVERDUE_ACCTS 233154 non-null int64
PRI_CURRENT_BALANCE 233154 non-null int64
PRI_SANCTIONED_AMOUNT 233154 non-null int64
PRI_DISBURSED_AMOUNT 233154 non-null int64
SEC_NO_OF_ACCTS  233154 non-null int64
SEC_ACTIVE_ACCTS 233154 non-null int64
SEC_OVERDUE_ACCTS 233154 non-null int64
SEC_CURRENT_BALANCE 233154 non-null int64
SEC_SANCTIONED_AMOUNT 233154 non-null int64
SEC_DISBURSED_AMOUNT 233154 non-null int64
PRIMARY_INSTAL_AMT 233154 non-null int64
SEC_INSTAL_AMT   233154 non-null int64
NEW_ACCTS_IN_LAST_SIX_MONTHS 233154 non-null int64
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS 233154 non-null int64
AVERAGE_ACCT_AGE  233154 non-null object
CREDIT_HISTORY_LENGTH 233154 non-null object
NO_OF_INQUIRIES   233154 non-null int64
loan_default      233154 non-null int64
rok_ur2           233154 non-null int32
wiek              233154 non-null int64
acc_age_rok      233154 non-null int32
acc_age_mies     233154 non-null int32
AVERAGE_ACCT_AGE_mth 233154 non-null int32
credit_hist_rok  233154 non-null int32
credit_hist_mies 233154 non-null int32
CREDIT_HISTORY_LENGTH_mth 233154 non-null int32
SCORE_DESC_A_VLR 233154 non-null float64
SCORE_DESC_B_VLR 233154 non-null float64
SCORE_DESC_C_VLR 233154 non-null float64
SCORE_DESC_D_VLR 233154 non-null float64
SCORE_DESC_E_LR   233154 non-null float64
```

```
SCORE_DESC_F_LR          233154 non-null float64
SCORE_DESC_G_LR          233154 non-null float64
SCORE_DESC_H_MR          233154 non-null float64
SCORE_DESC_I_MR          233154 non-null float64
SCORE_DESC_J_HR          233154 non-null float64
SCORE_DESC_K_HR          233154 non-null float64
SCORE_DESC_L_VHR         233154 non-null float64
SCORE_DESC_M_VHR         233154 non-null float64
SCORE_DESC_N1             233154 non-null float64
SCORE_DESC_N2             233154 non-null float64
SCORE_DESC_N3             233154 non-null float64
SCORE_DESC_N4             233154 non-null float64
SCORE_DESC_N5             233154 non-null float64
SCORE_DESC_N6             233154 non-null float64
SCORE_DESC_N7             233154 non-null float64
PERFORM_CNS_SCORE_DESCRIPTION2 233154 non-null object
disbursed_amount_log     233154 non-null float64
asset_cost_log           233154 non-null float64
ltv_log                  233154 non-null float64
dtypes: float64(25), int32(7), int64(35), object(6)
memory usage: 123.6+ MB
```

In [153]:

1 wiek\_o.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 22 entries, 1184 to 225982
Data columns (total 73 columns):
UniqueID                      22 non-null int64
disbursed_amount               22 non-null int64
asset_cost                     22 non-null int64
ltv                            22 non-null float64
branch_id                      22 non-null int64
supplier_id                    22 non-null int64
manufacturer_id                22 non-null int64
Current_pincode_ID             22 non-null int64
Date_of_Birth                  22 non-null object
Employment_Type                22 non-null float64
DisbursalDate                 22 non-null object
State_ID                       22 non-null int64
Employee_code_ID               22 non-null int64
MobileNo_Avl_Flag              22 non-null int64
Aadhar_flag                    22 non-null int64
PAN_flag                       22 non-null int64
VoterID_flag                  22 non-null int64
Driving_flag                   22 non-null int64
Passport_flag                  22 non-null int64
PERFORM_CNS_SCORE              22 non-null int64
PERFORM_CNS_SCORE_DESCRIPTION  22 non-null object
PRI_NO_OF_ACCTS                22 non-null int64
PRI_ACTIVE_ACCTS               22 non-null int64
PRI_OVERDUE_ACCTS              22 non-null int64
PRI_CURRENT_BALANCE            22 non-null int64
PRI_SANCTIONED_AMOUNT          22 non-null int64
PRI_DISBURSED_AMOUNT           22 non-null int64
SEC_NO_OF_ACCTS                22 non-null int64
SEC_ACTIVE_ACCTS               22 non-null int64
SEC_OVERDUE_ACCTS              22 non-null int64
SEC_CURRENT_BALANCE            22 non-null int64
SEC_SANCTIONED_AMOUNT          22 non-null int64
SEC_DISBURSED_AMOUNT           22 non-null int64
PRIMARY_INSTAL_AMT             22 non-null int64
SEC_INSTAL_AMT                 22 non-null int64
NEW_ACCTS_IN_LAST_SIX_MONTHS   22 non-null int64
DELINQUENT_ACCTS_IN_LAST_SIX_MONTHS 22 non-null int64
AVERAGE_ACCT_AGE               22 non-null object
CREDIT_HISTORY_LENGTH           22 non-null object
NO_OF_INQUIRIES                 22 non-null int64
loan_default                   22 non-null int64
rok_ur2                        22 non-null int32
wiek                           22 non-null int64
acc_age_rok                    22 non-null int32
acc_age_mies                   22 non-null int32
AVERAGE_ACCT_AGE_mth           22 non-null int32
credit_hist_rok                22 non-null int32
credit_hist_mies                22 non-null int32
CREDIT_HISTORY_LENGTH_mth       22 non-null int32
SCORE_DESC_A_VLR                22 non-null float64
SCORE_DESC_B_VLR                22 non-null float64
SCORE_DESC_C_VLR                22 non-null float64
SCORE_DESC_D_VLR                22 non-null float64
SCORE_DESC_E_LR                 22 non-null float64
```

```

SCORE_DESC_F_LR          22 non-null float64
SCORE_DESC_G_LR          22 non-null float64
SCORE_DESC_H_MR          22 non-null float64
SCORE_DESC_I_MR          22 non-null float64
SCORE_DESC_J_HR          22 non-null float64
SCORE_DESC_K_HR          22 non-null float64
SCORE_DESC_L_VHR         22 non-null float64
SCORE_DESC_M_VHR         22 non-null float64
SCORE_DESC_N1             22 non-null float64
SCORE_DESC_N2             22 non-null float64
SCORE_DESC_N3             22 non-null float64
SCORE_DESC_N4             22 non-null float64
SCORE_DESC_N5             22 non-null float64
SCORE_DESC_N6             22 non-null float64
SCORE_DESC_N7             22 non-null float64
PERFORM_CNS_SCORE_DESCRIPTION2 22 non-null object
disbursed_amount_log     22 non-null float64
asset_cost_log            22 non-null float64
ltv_log                  22 non-null float64
dtypes: float64(25), int32(7), int64(35), object(6)
memory usage: 12.1+ KB

```

In [155]:

```
1 wiek_o.describe()
```

Out[155]:

	UniqueID	disbursed_amount	asset_cost	ltv	branch_id	supplier_id	manufacturer_id
count	22.00	22.00	22.00	22.00	22.00	22.00	22.00
mean	537255.00	44907.68	65946.14	70.57	69.64	19247.41	74.77
std	77533.82	10050.42	8114.19	12.31	56.59	3871.42	21.49
min	420977.00	26829.00	50816.00	46.04	2.00	13931.00	45.00
25%	464081.00	39679.75	61514.25	66.19	20.00	15561.25	46.50
50%	546572.00	45699.50	63608.50	71.06	62.00	18694.50	86.00
75%	596066.25	52487.25	70522.75	79.04	112.75	22878.50	86.00
max	652399.00	61736.00	82853.00	89.99	162.00	24496.00	120.00

In [39]:

```
1 wiek_out = pd.DataFrame(zbior.wiek2 > 65.5)
2 wiek_out.head()
```

Out[39]:

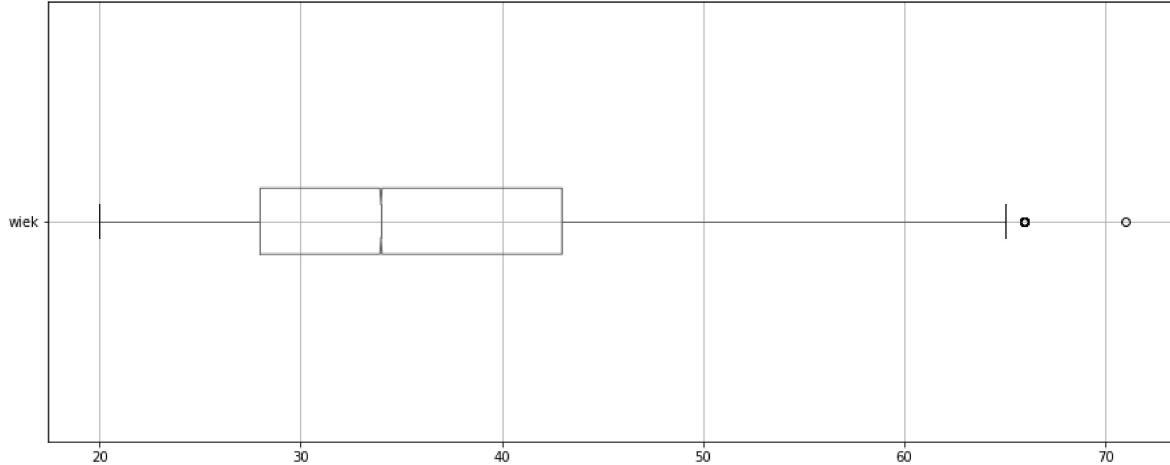
wiek2
0 False
1 False
2 False
3 False
4 False

In [31]:

```

1 zbor[['wiek']].boxplot(vert=False,
2                             figsize=(15,6),
3                             notch=True,
4                             patch_artist=False
5                             #boxdrops=dict(facecolor=c, color=c)
6 )
7 #plt.xlim(50)
8 #plt.xticks([10, 50])
9 #plt.subplots_adjust(left=0,25)
10 plt.show()
11

```

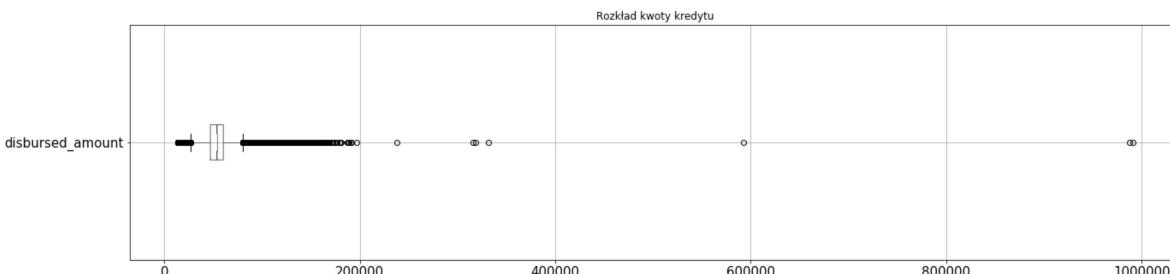


In [84]:

```

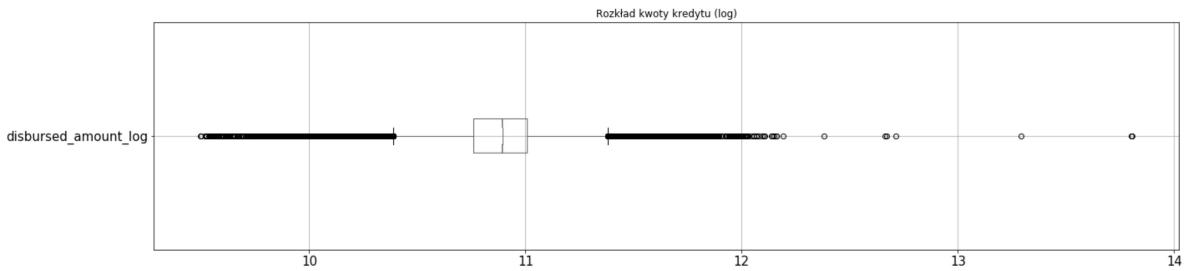
1 zbor[['disbursed_amount']].boxplot(vert=False,
2                                     figsize=(22,5),
3                                     notch=True,
4                                     patch_artist=False,
5                                     fontsize=15)
6 #plt.xlim(50)
7 #plt.xticks([10, 50])
8 #plt.subplots_adjust(left=0,25)
9 plt.title('Rozkład kwoty kredytu')
10 xlabel = ''
11 plt.show()

```



In [112]:

```
1 zbiror[['disbursed_amount_log']].boxplot(vert=False,
2                                     figsize=(22,5),
3                                     notch=True,
4                                     patch_artist=False,
5                                     fontsize=15)
6 #plt.xlim(50)
7 #plt.xticks([10, 50])
8 #plt.subplots_adjust(left=0,25)
9 plt.title('Rozkład kwoty kredytu (log)')
10 plt.show()
```

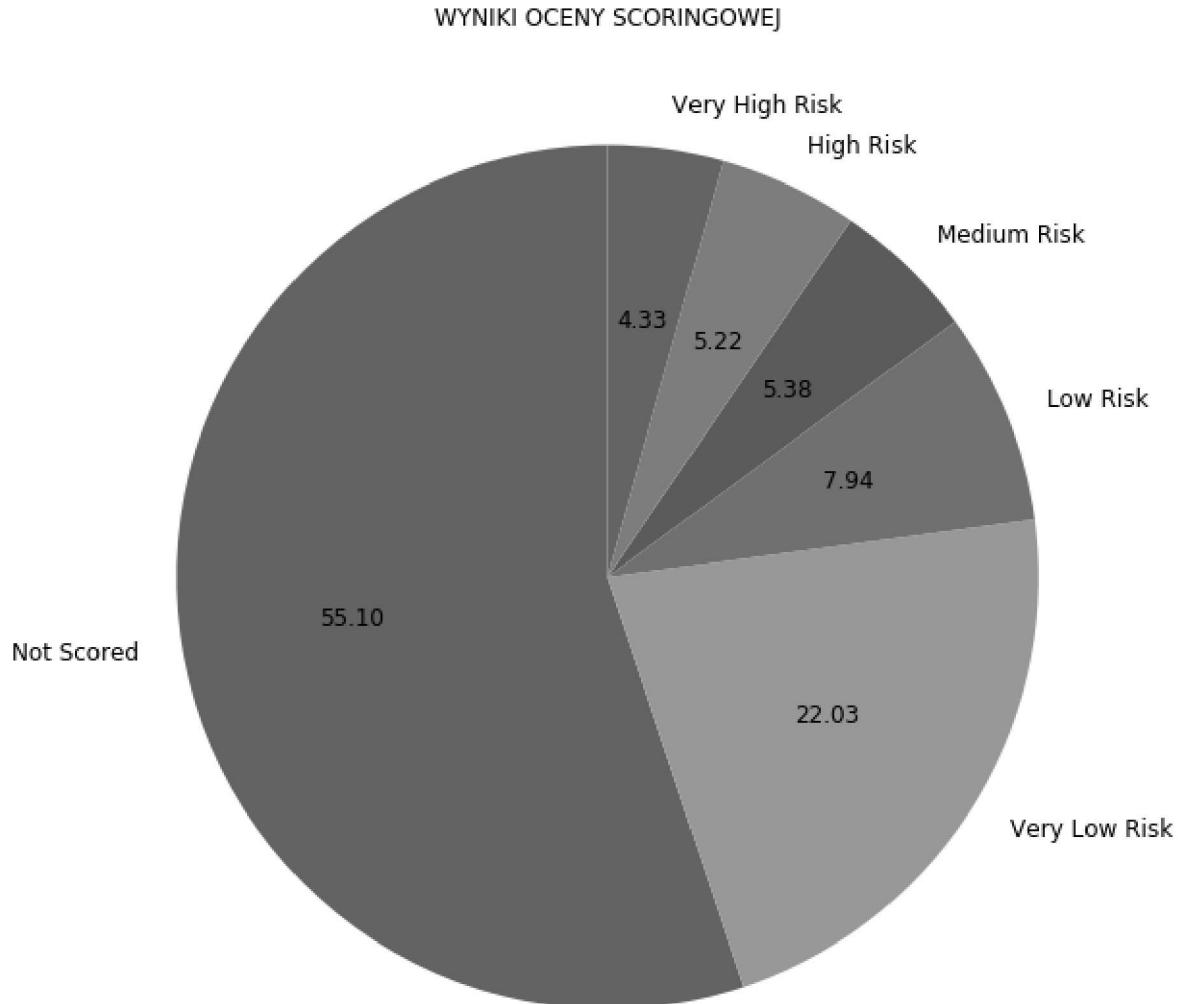


In [296]:

```
1 #wykres "WYNIKI OCENY SCORINGOWEJ"
2 plt.figure(figsize=(10,25))
3 zbiór.PERFORM_CNS_SCORE_DESCRIPTION2.value_counts().plot(kind='pie', startangle=90, au
4
5
```

Out[296]:

&lt;matplotlib.axes.\_subplots.AxesSubplot at 0x13f5fa61588&gt;



In [121]:

```
1 zbiór.wiek2.mean()
```

Out[121]:

36.448009472577866

In [122]:

```
1 zbor.wiek2.std()
```

Out[122]:

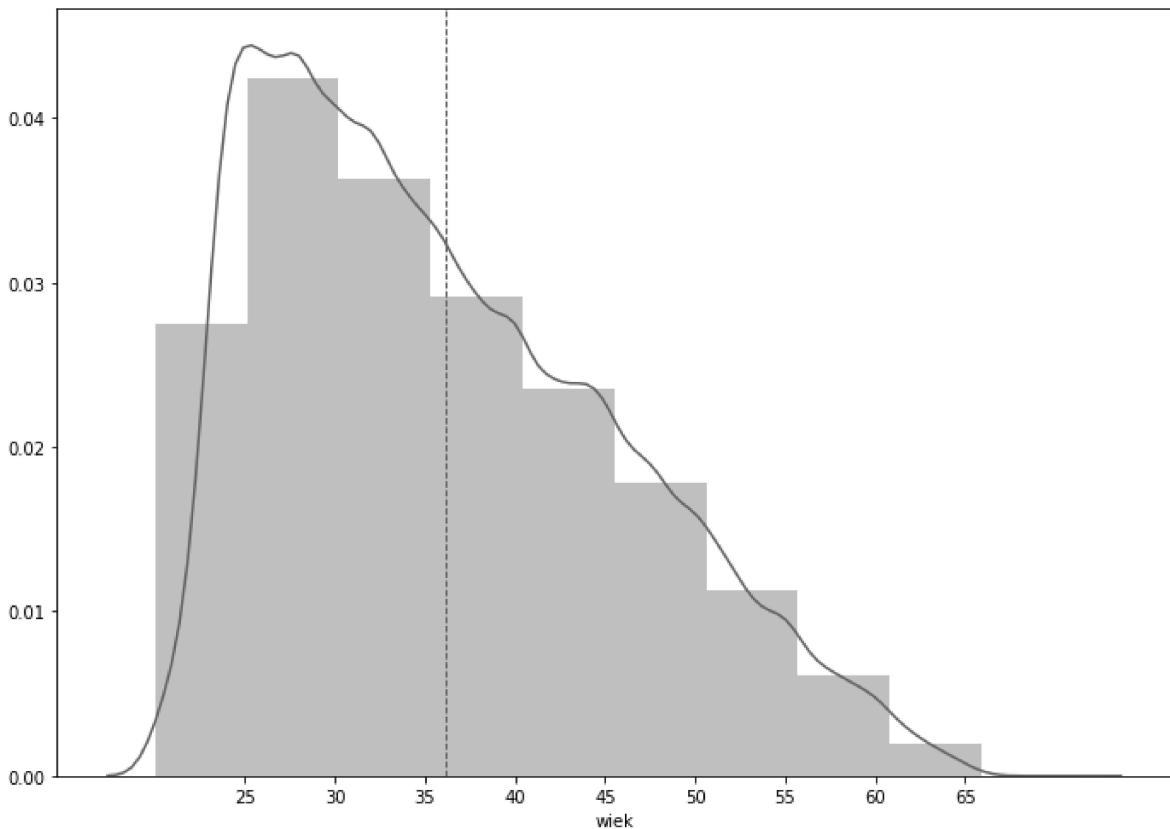
9.666254472381578

In [130]:

```
1 sns.distplot(zbor.wiek, bins=10, kde=True, hist=True, norm_hist=False)
2 plt.gcf().set_size_inches(11.7, 8.27)
3 plt.xticks([25, 30, 35, 40, 45, 50, 55, 60, 65])
4 plt.axvline(zbor.wiek.mean(), linestyle='dashed', linewidth=1)
5 #plt.axvline(zbor.wiek2.std())
```

Out[130]:

&lt;matplotlib.lines.Line2D at 0x22e0f3b5f08&gt;

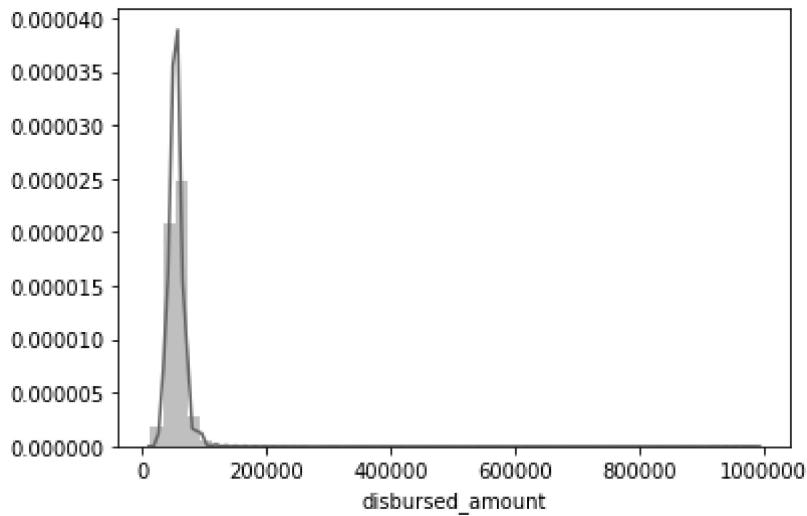


In [29]:

```
1 sns.distplot(zbior.disbursed_amount)
```

Out[29]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x22e000a0508>
```



```
1 sns.distplot(zbior.wiek2);
2 show()
```