

```
Ввод [1]: import pandas as pd
import codecs
import numpy as np
with codecs.open('Copy of Выданная сумма по кредиту.xlsx', 'r', encoding='utf-8',
                errors='ignore') as fdata:
    data = pd.read_excel('Copy of Выданная сумма по кредиту.xlsx')
data = data.drop(['Unnamed: 0', 'Unnamed: 4', 'Unnamed: 5'],axis=1)
data = data.replace('н/д', np.NaN)
```

```
Ввод [2]: data = data[data['DogovorID'].isna() == False]
data['DogovorID'] = data['DogovorID'].astype(int)
```

```
Ввод [3]: data
```

```
Out[3]:
```

	DogovorID	ПИНФЛ	ДАТА	Выданная сумма по кредиту
2	397626	-	NaN	NaN
3	397626	-	20.09.2017	11,982,000.00
4	397672	-	NaN	NaN
5	397672	-	22.12.2016	13,000,000.00
6	397806	-	19.07.2016	3,420,000.00
...
11056	441459	40801734310030	08.09.2018	18,000,000.00
11057	441486	40101794160076	25.07.2019	1,000,000.00
11058	441930	31912907010034	03.05.2019	5,000,000.00
11059	422398	42410754310016	18.03.2020	4,000,000.00
11060	422398	42410754310016	20.11.2019	13,000,000.00

8105 rows × 4 columns

```
Ввод [4]: data["Выданная сумма по кредиту"] = [float(str(i).replace(",","")) for i in data["Выданная сумма по кредиту"]]
```

```
Ввод [5]: data=data[data['Выданная сумма по кредиту'].isna() == False]
data['Выданная сумма по кредиту']=data['Выданная сумма по кредиту'].astype(float).ast
```

C:\Users\TR11LeR\AppData\Local\Temp\ipykernel_11160\2786920534.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
data['Выданная сумма по кредиту']=data['Выданная сумма по кредиту'].astype(float).astype(int)
```

Ввод [6]: data

Out[6]:

	DogovorID	ПИНФЛ	ДАТА	Выданная сумма по кредиту
3	397626	-	20.09.2017	11982000
5	397672	-	22.12.2016	13000000
6	397806	-	19.07.2016	3420000
9	397853	-	02.10.2014	6000000
14	398051	-	25.02.2018	3125000
...
11056	441459	40801734310030	08.09.2018	18000000
11057	441486	40101794160076	25.07.2019	1000000
11058	441930	31912907010034	03.05.2019	5000000
11059	422398	42410754310016	18.03.2020	4000000
11060	422398	42410754310016	20.11.2019	13000000

6810 rows × 4 columns

Ввод [7]: data.groupby('DogovorID')['ДАТА'].value_counts()

Out[7]:

DogovorID	ДАТА	
396015	08.01.2019	2
	08.08.2018	2
397201	05.09.2018	1
397204	22.09.2016	1
	26.08.2015	1
	..	
453831	12.05.2021	1
454735	10.08.2021	1
455538	06.05.2019	1
456264	23.01.2020	1
	23.11.2020	1

Name: ДАТА, Length: 6620, dtype: int64

Ввод [8]: data = data[data['DogovorID'] != 396015]

Ввод [9]: data.groupby('DogovorID')['Выданная сумма по кредиту'].sum()

Out[9]:

DogovorID	
397201	18000000
397204	8721000
397207	10000000
397213	6326500
397218	16381000
...	
452962	12000000
453831	5000000
454735	20000000
455538	327272
456264	18200000

Name: Выданная сумма по кредиту, Length: 4305, dtype: int32

Ввод [10]: data.groupby('DogovorID')['Выданная сумма по кредиту'].sum().to_excel('NEW_Выданная с

