

TASK 2

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: data = pd.read_csv('loan.csv')
```

```
In [3]: data.head()
```

```
Out[3]:
```

	customer_id	disbursed_amount	interest	market	employment	time_employed	householder	ii
0	0	23201.5	15.4840	C	Teacher	<=5 years	RENT	8
1	1	7425.0	11.2032	B	Accountant	<=5 years	OWNER	10
2	2	11150.0	8.5100	A	Statistician	<=5 years	RENT	6
3	3	7600.0	5.8656	A	Other	<=5 years	RENT	10
4	4	31960.0	18.7392	E	Bus driver	>5 years	RENT	9

```
In [4]: # task 1
mean1 = data['disbursed_amount'].mean()
mean2 = data['interest'].mean()
```

```
In [5]: mean1
```

```
Out[5]: 14132.2755
```

```
In [6]: mean2
```

```
Out[6]: 12.678819440000039
```

```
In [7]: # task 2
# number of variables with int64 datatype i.e discrete values
data.select_dtypes(include=['int64']).shape[1]
# variables are :- customer_id , target
# their count is 2
```

```
Out[7]: 2
```

```
In [8]: # task 3
data['customer_id'].unique()
```

```
Out[8]: array([ 0, 1, 2, ..., 9997, 9998, 9999], dtype=int64)
```

```
In [9]: data['target'].unique()
```

```
Out[9]: array([0, 1], dtype=int64)
```

```
In [10]: # task 4  
data['date_issued_dt'] = pd.to_datetime(data['date_issued'])
```

```
In [11]: data['month'] = data['date_issued_dt'].dt.month
```

```
In [12]: data['month'].value_counts()  
# in the month of "October" the number of Loan issued is the maximum and equals t
```

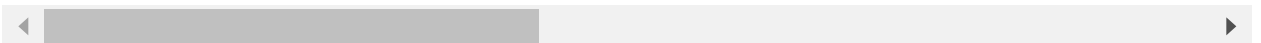
```
Out[12]: 10    1277  
        7     1066  
        11    1017  
        12     882  
        8     852  
        4     816  
        5     749  
        9     734  
        6     700  
        1     700  
        3     623  
        2     584  
Name: month, dtype: int64
```

```
In [13]: # task 5
x = data.query('employment == "Teacher" and householder == "OWNER"')
x
```

Out[13]:

	customer_id	disbursed_amount	interest	market	employment	time_employed	householder
71	71	10230.00	7.8900	A	Teacher	>5 years	OWNER
85	85	3713.00	14.3068	C	Teacher	>5 years	OWNER
171	171	19646.25	14.3904	C	Teacher	<=5 years	OWNER
672	672	15200.00	10.8801	B	Teacher	>5 years	OWNER
1024	1024	12144.00	14.2008	C	Teacher	<=5 years	OWNER
...
9154	9154	9900.00	13.3855	B	Teacher	>5 years	OWNER
9172	9172	22080.00	23.6509	E	Teacher	>5 years	OWNER
9433	9433	4750.00	10.4405	B	Teacher	<=5 years	OWNER
9684	9684	6650.00	7.4955	A	Teacher	>5 years	OWNER
9769	9769	16005.00	12.5631	C	Teacher	>5 years	OWNER

69 rows × 16 columns



```
In [14]: x.shape[0]
# 69 teachers are owners
```

Out[14]: 69

```
In [15]: # task 6
data.query('householder == "RENT"')['employment'].value_counts()
# "Civil Servant" and "Teacher" are the employers who mostly rents
```

```
Out[15]: Civil Servant      371
Teacher      371
Bus driver   360
Nurse        358
Secretary    355
Dentist       355
Other         353
Statistician  342
Accountant    322
Taxi driver   316
Software developer  315
Name: employment, dtype: int64
```