```
!pip install pandas ydata-sdk
!pip install ydata-sdk
!pip install matplotlib seaborn
!pip install numpy
!pip install sweetviz ipywidgets
```

```
!kaggle datasets download -d "thedevilsdeal/telecom-customer-churn"
```

```
!mkdir -p data
!unzip -o telecom_customer_churn.zip -d data
```

```
df = pd.read_csv("data/Customer_Churn_Predictions.csv")
```

```
df.head()
```

	customer_id	college	job	gender	dependents	occupation	city	customer_rm_category	branch_code	data_since_y
0	1	0.135	0[illegible]	0[illegible]	0.0	self_employed	187.0	2	756	
1	2	0.10	0[illegible]	0[illegible]	0.0	self_employed	1[illegible]	3	1214	
2	4	0[illegible]	0[illegible]	0[illegible]	1.0	salaried	148.0	2	42	
3	5	0[illegible]	0[illegible]	0[illegible]	0.0	self_employed	1[illegible]	1	1[illegible]	
4	6	0.[illegible]	0[illegible]	0[illegible]	0.0	self_employed	1[illegible]	2	348	

5 rows × 21 columns

```
df.describe(include='all')
```

	customer_id	college	job	gender	dependents	occupation	city	customer_rm_category
count	20000.000000	20000.000000	20000.000000	19947	20000.000000	20000	19979.000000	20000.000
unique	NaN	NaN	NaN	2	NaN	3	NaN	5
top	NaN	NaN	NaN	male	NaN	self_employed	NaN	5
freq	NaN	NaN	NaN	10[illegible]	NaN	10[illegible]	NaN	7[illegible]
mean	10140.499857	2304.334448	40.280439	NaN	0.367200	NaN	164.159270	2.5200
std	5[illegible].[illegible]	1[illegible].[illegible]	1[illegible].[illegible]	NaN	0.[illegible]	NaN	1[illegible].[illegible]	1.1[illegible]
min	1.000000	0.000000	1.000000	NaN	0.000000	NaN	1.000000	1.000
25%	5[illegible].[illegible]	1[illegible].[illegible]	30.000000	NaN	0.000000	NaN	4[illegible].000000	2.000
50%	10[illegible].000000	2[illegible].000000	40.000000	NaN	0.000000	NaN	8[illegible].000000	2.000
75%	15[illegible].750000	3[illegible].750000	50.000000	NaN	1.000000	NaN	1[illegible].000000	3.000
max	20[illegible].000000	5[illegible].000000	60.000000	NaN	1.000000	NaN	1[illegible].000000	5.000

11 rows × 21 columns

```
df.info()
```

```
<class'.pandas.core.frame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 21 columns):
 #   Column                 Non-Nulls Count  Dtype
---  ------                 --------------  -----
 0   customer_id            20000 non-null  int64
 1   college                20000 non-null  float64
 2   job                    20000 non-null  int64
 3   gender                 19947 non-null  object
 4   dependents             20000 non-null  int64
 5   occupation             20000 non-null  object
 6   city                   19979 non-null  float64
 7   customer_rm_category   20000 non-null  int64
 8   branch_code            20000 non-null  int64
 9   data_since_yrs         20000 non-null  int64
 10  marital_status         19[illegible] non-null  object
 11  home_ownership         19[illegible] non-null  object
 12  number_of_dependants   19[illegible] non-null  object
 13  property_owned         19[illegible] non-null  object
 14  work_phone             20000 non-null  int64
 15  phone                  20000 non-null  int64
 16  chq_status             20000 non-null  int64
 17  foreign_worker         20000 non-null  int64
 18  job_status             20000 non-null  int64
 19  no_of_mobile           20000 non-null  int64
 20  no_of_dependants       20000 non-null  int64
dtypes: int64(1[illegible]), float64(2), object(5)
memory usage: 3.[illegible] MB
```

```
df['occupation'].value_counts()
```

```
occupation
self_employed    10[illegible]
salaried         7[illegible]
business         1[illegible]
Name: count, dtype: int64
```

```
df['marital_status'].value_counts()
```

```
marital_status
married    1[illegible]
single     [illegible]
divorced   [illegible]
Name: count, dtype: int64
```

```
df['home_ownership'].value_counts()
```

```
home_ownership
owned    1[illegible]
rented   [illegible]
Name: count, dtype: int64
```

```
df['number_of_dependants'].value_counts()
```

```
number_of_dependants
0    [illegible]
1    [illegible]
2    [illegible]
3    [illegible]
4    [illegible]
Name: count, dtype: int64
```

```
df['property_owned'].value_counts()
```

```
property_owned
yes    1[illegible]
no     [illegible]
Name: count, dtype: int64
```

```
df['work_phone'].value_counts()
```

```
work_phone
0    [illegible]
1    [illegible]
Name: count, dtype: int64
```

```
df['phone'].value_counts()
```

```
phone
1    [illegible]
0    [illegible]
Name: count, dtype: int64
```

```
df['chq_status'].value_counts()
```

```
chq_status
0    [illegible]
1    [illegible]
Name: count, dtype: int64
```

```
df['foreign_worker'].value_counts()
```

```
foreign_worker
1    [illegible]
0    [illegible]
Name: count, dtype: int64
```

```
df['job_status'].value_counts()
```

```
job_status
1    [illegible]
0    [illegible]
Name: count, dtype: int64
```

```
df['no_of_mobile'].value_counts()
```

```
no_of_mobile
1    [illegible]
0    [illegible]
Name: count, dtype: int64
```

```
df['no_of_dependants'].value_counts()
```

```
no_of_dependants
0    [illegible]
1    [illegible]
2    [illegible]
3    [illegible]
4    [illegible]
Name: count, dtype: int64
```

```
df['occupation'].value_counts().plot(kind = 'bar')
```

```
matplotlib.axes._subplots.AxesSubplot at 0x7f[illegible]
```

```
df['gender'].value_counts().plot(kind = 'bar')
```

```
matplotlib.axes._subplots.AxesSubplot at 0x7f[illegible]
```

```
df['dependents'].value_counts().plot(kind = 'bar')
```

```
matplotlib.axes._subplots.AxesSubplot at 0x7f[illegible]
```

```
df['home_owned'].value_counts().plot(kind = 'bar')
```

```
matplotlib.axes._subplots.AxesSubplot at 0x7f[illegible]
```

```
df['marital_status'].value_counts().plot(kind = 'bar')
```

```
matplotlib.axes._subplots.AxesSubplot at 0x7f[illegible]
```

```
df.info()
```

```
<class'.pandas.core.frame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 21 columns):
 #   Column                 Non-Nulls Count  Dtype
---  ------                 --------------  -----
 0   customer_id            20000 non-null  int64
 1   college                20000 non-null  float64
 2   job                    20000 non-null  int64
 3   gender                 19947 non-null  object
 4   dependents             20000 non-null  int64
 5   occupation             20000 non-null  object
 6   city                   19979 non-null  float64
 7   customer_rm_category   20000 non-null  int64
 8   branch_code            20000 non-null  int64
 9   data_since_yrs         20000 non-null  int64
 10  marital_status         19[illegible] non-null  object
 11  home_ownership         19[illegible] non-null  object
 12  number_of_dependants   19[illegible] non-null  object
 13  property_owned         19[illegible] non-null  object
 14  work_phone             20000 non-null  int64
 15  phone                  20000 non-null  int64
 16  chq_status             20000 non-null  int64
 17  foreign_worker         20000 non-null  int64
 18  job_status             20000 non-null  int64
 19  no_of_mobile           20000 non-null  int64
 20  no_of_dependants       20000 non-null  int64
dtypes: int64(1[illegible]), float64(2), object(5)
memory usage: 3.[illegible] MB
```

```
df['home_owned'].value_counts()
```

```
home_owned
owned    1[illegible]
rented   [illegible]
Name: count, dtype: int64
```

```
df['job_status'].value_counts()
```

```
job_status
1    [illegible]
0    [illegible]
Name: count, dtype: int64
```

```
df['job'].value_counts()
```

```
job
0    [illegible]
1    [illegible]
Name: count, dtype: int64
```

```
df['college'].value_counts()
```

```
college
0.000000    [illegible]
0.[illegible]    [illegible]
0.[illegible]    [illegible]
0.[illegible]
```