```
In [41]: import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         import vaex
         %matplotlib inline
```

```
In [2]: pip install vaex
```

```
Requirement already satisfied: uvicorn[standard] in ./.local/lib/python3.8/
site-packages (from vaex-server<0.9,>=0.8.1->vaex) (0.20.0)
Requirement already satisfied: fastapi in ./.local/lib/python3.8/site-packa
ges (from vaex-server<0.9,>=0.8.1->vaex) (0.88.0)
Requirement already satisfied: tornado>4.1 in ./.local/lib/python3.8/site-p
ackages (from vaex-server<0.9,>=0.8.1->vaex) (6.2)
Requirement already satisfied: cachetools in ./.local/lib/python3.8/site-pa
ckages (from vaex-server<0.9,>=0.8.1->vaex) (5.2.0)
Requirement already satisfied: jinja2 in ./.local/lib/python3.8/site-packag
es (from vaex-ml<0.19,>=0.18.1->vaex) (3.1.2)
Requirement already satisfied: numba in ./.local/lib/python3.8/site-package
s (from vaex-ml<0.19,>=0.18.1->vaex) (0.56.4)
Requirement already satisfied: traitlets in ./.local/lib/python3.8/site-pac
kages (from vaex-ml<0.19,>=0.18.1->vaex) (5.7.1)
Requirement already satisfied: astropy in ./.local/lib/python3.8/site-packa
ges (from vaex-astro<0.10,>=0.9.3->vaex) (5.2)
Requirement already satisfied: ipyvolume>=0.4 in ./.local/lib/python3.8/sit
e-packages (from vaex-jupyter<0.9,>=0.8.1->vaex) (0.5.2)
Requirement already satisfied: ipyvuetify<2,>=1.2.2 in ./.local/lib/python
3.8/site-packages (from vaex-jupyter<0.9,>=0.8.1->vaex) (1.8.4)
```

```
In [14]: dataset=pd.read_csv(r"/home/nirmala/Documents/precog_requirement_task/judges_c
```

```
In [15]: dataset.head(10)
```

Out[15]:

	ddl_judge_id	state_code	dist_code	court_no	judge_position	female_judge	start_date	end_date
0	1	1	1	1	chief judicial magistrate	0 nonfemale	20-09-2013	20-02-2014
1	2	1	1	1	chief judicial magistrate	0 nonfemale	31-10-2013	20-02-2014
2	3	1	1	1	chief judicial magistrate	0 nonfemale	21-02-2014	31-05-2016
3	4	1	1	1	chief judicial magistrate	0 nonfemale	01-06-2016	06-06-2016
4	5	1	1	1	chief judicial magistrate	0 nonfemale	06-06-2016	07-07-2018
5	6	1	1	1	chief judicial magistrate	1 female	09-07-2018	NaN
6	7	1	1	1	civil judge junior division	0 nonfemale	01-06-2011	09-06-2013
7	8	1	1	1	civil judge junior division	0 nonfemale	06-06-2011	08-06-2014
8	9	1	1	1	civil judge junior division	0 nonfemale	11-06-2012	31-05-2015
9	10	1	1	1	civil judge junior division	0 nonfemale	29-08-2013	08-06-2014

```
In [5]:  np.random.seed(0)
```

```
In [16]:  dataset.head()
```

Out[16]:

	ddl_judge_id	state_code	dist_code	court_no	judge_position	female_judge	start_date	end_date
0	1	1	1	1	chief judicial magistrate	0 nonfemale	20-09-2013	20-02-2014
1	2	1	1	1	chief judicial magistrate	0 nonfemale	31-10-2013	20-02-2014
2	3	1	1	1	chief judicial magistrate	0 nonfemale	21-02-2014	31-05-2016
3	4	1	1	1	chief judicial magistrate	0 nonfemale	01-06-2016	06-06-2016
4	5	1	1	1	chief judicial magistrate	0 nonfemale	06-06-2016	07-07-2018

```
In [7]:  dataset.tail(100)
```

Out[7]:

	ddl_judge_id	state_code	dist_code	court_no	judge_position	female_judge	start_date	end_date
98378	98379	30	2	2	criminal cases	1 female	02-06-2014	20-05-2017
98379	98380	30	2	2	criminal cases	1 female	16-10-2014	23-05-2016
98380	98381	30	2	2	criminal cases	1 female	01-12-2014	02-01-2015
98381	98382	30	2	2	criminal cases	1 female	01-12-2014	30-03-2015
98382	98383	30	2	2	criminal cases	1 female	01-12-2014	30-05-2016
...	...	...	...	...	...	...	...	...
98473	98474	30	2	9	criminal cases	1 female	21-04-2004	14-11-2013
98474	98475	30	2	9	criminal cases	1 female	16-01-2015	16-01-2016
98475	98476	30	2	9	criminal cases	1 female	09-12-2016	31-07-2017
98476	98477	30	2	10	criminal cases	1 female	15-05-2017	28-01-2019
98477	98478	30	2	10	criminal cases	0 nonfemale	28-01-2019	NaN

100 rows × 8 columns

```
In [8]:  mssising_values_count=dataset.isnull().sum()
```

```
In [9]:  mssising_values_count[0:10]

Out[9]:  ddl_judge_id            0
         state_code              0
         dist_code               0
         court_no                0
         judge_position          0
         female_judge            1
         start_date              0
         end_date            18158
         dtype: int64

In [10]:  dataset.info()

         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 98478 entries, 0 to 98477
         Data columns (total 8 columns):
          #   Column          Non-Null Count  Dtype
         ---  ------          --------------  -----
          0   ddl_judge_id    98478 non-null  int64
          1   state_code      98478 non-null  int64
          2   dist_code       98478 non-null  int64
          3   court_no        98478 non-null  int64
          4   judge_position  98478 non-null  object
          5   female_judge    98477 non-null  object
          6   start_date      98478 non-null  object
          7   end_date        80320 non-null  object
         dtypes: int64(4), object(4)
         memory usage: 6.0+ MB

In [ ]:  dataset['court_no'].isin([1])

In [63]:  dataset.loc[dataset.court_no ==1,'court_no']=1
          dataset.loc[dataset.state_code==1,'state_code']=1
          dataset.groupby(['court_no','state_code']).size()

Out[63]:  court_no  state_code
          1         1             2377
                    2             1449
                    3              194
                    4             1841
                    5              187

                                   ...
          995       3                7
          996       3                3
          997       3               26
          998       3               22
          999       3               13
          Length: 2350, dtype: int64
```

```
In [32]: print(dataset.describe())
         print(dataset['judge_position'].unique())
```

```
         ddl_judge_id     state_code      dist_code       court_no      difference
count   80320.000000   80320.000000   80320.000000   80320.000000   80320.000000
mean    47574.487936       9.553735      18.627764      48.077490     409.376855
std     28655.437864       7.898909      14.142144     138.433296     457.846601
min         1.000000       1.000000       1.000000       1.000000       1.000000
25%     22211.750000       1.000000       8.000000       3.000000      76.000000
50%     46749.000000       9.000000      16.000000       9.000000     259.000000
75%     72406.250000      14.000000      26.000000      23.000000     616.000000
max     98477.000000      30.000000      75.000000     999.000000    8150.000000
['chief judicial magistrate' 'civil judge junior division'
 'civil judge senior division' 'district and sessions court'
 'criminal cases' 'civil judge' 'judicial magistrate court' 'civil court'
 'motor accidents claims tribunal' 'railway ner' 'small cause court'
 'gram nyayalaya' 'additional district judge' 'juvenile court'
 'additional district and sessions court' 'jmfc' 'commercial court'
 'district and additional sessions court'
 'additional metropolitan magistrate' 'chief metropolitan magistrate'
 'junior division' 'cantonment court' 'municipal court'
 'city district and sessions court' 'fast track court' 'industrial court'
```

```
In [65]: dataset1=dataset[dataset['state_code']==6]
         # dataset1.groupby('judge_position').size()
         print(dataset1.describe())
         print(dataset1.isnull().sum())
```

```
          ddl_judge_id   state_code      dist_code      court_no     difference
count      1183.000000       1183.0    1183.000000   1183.000000    1183.000000
mean      37550.865596          6.0      10.290786      5.896872     544.895182
std         449.652853          0.0       6.958784      5.008066     474.593112
min       36786.000000          6.0       1.000000      1.000000       1.000000
25%       37155.500000          6.0       4.000000      2.000000     201.000000
50%       37552.000000          6.0      10.000000      5.000000     404.000000
75%       37943.500000          6.0      17.000000      8.000000     757.000000
max       38361.000000          6.0      28.000000     37.000000    3600.000000
ddl_judge_id       0
state_code         0
dist_code          0
court_no           0
judge_position     0
Gender             0
start_date         0
end_date           0
difference         0
dtype: int64
```

```
In [66]: def replace_values(value):
             if value == '1 female':
                 return 'Female'
             elif value == '0 nonfemale':
                 return 'Male'
             else:
                 return value

         # # Apply the function to the column
         dataset['female_judge'] = dataset['female_judge'].apply(replace_values)
         dataset.rename(columns={'female_judge': 'Gender'}, inplace=True)
         print(dataset.head(30))
```

```
In [28]: dataset['start_date'] = pd.to_datetime(dataset['start_date'], format='%d-%m-%Y
         dataset['end_date'] = pd.to_datetime(dataset['end_date'], format='%d-%m-%Y')
         dataset['difference']=(dataset['end_date']-dataset['start_date']).dt.days
         dataset=dataset[dataset['difference']>=0]
         print(dataset.head(50))
```

```
        ddl_judge_id  state_code  dist_code  court_no  \
0                  1           1          1         1
1                  2           1          1         1
2                  3           1          1         1
3                  4           1          1         1
4                  5           1          1         1
6                  7           1          1         1
7                  8           1          1         1
8                  9           1          1         1
9                 10           1          1         1
10                11           1          1         1
11                12           1          1         1
12                13           1          1         1
13                14           1          1         1
14                15           1          1         1
15                16           1          1         1
16                17           1          1         1
17                18           1          1         1
19                20           1          1         1
20                21           1          1         1
```
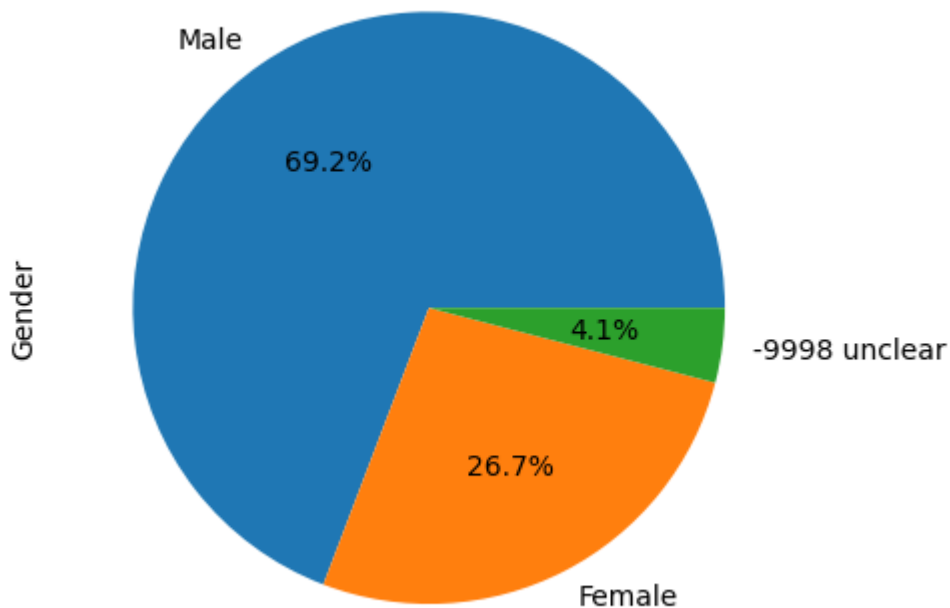
```
In [27]: print(dataset.describe());
```

```
        ddl_judge_id     state_code     dist_code       court_no     difference
count  80320.000000   80320.000000   80320.000000   80320.000000   80320.000000
mean   47574.487936       9.553735      18.627764      48.077490     409.376855
std    28655.437864       7.898909      14.142144     138.433296     457.846601
min        1.000000       1.000000       1.000000       1.000000       1.000000
25%    22211.750000       1.000000       8.000000       3.000000      76.000000
50%    46749.000000       9.000000      16.000000       9.000000     259.000000
75%    72406.250000      14.000000      26.000000      23.000000     616.000000
max    98477.000000      30.000000      75.000000     999.000000    8150.000000
```

```
In [49]:  # dat=vaex.open("/home/nirmala/Documents/precog_requirement_task/judges_clean.
          # value1=dat.groupby('judge_position')
          # counts=value1.count()
          # labels=counts.index.tolist()
          # values=counts['ddl_judge_id'].tolist()
          # plt.pie(values,labels=labels)
          # plt.show()
          dats=dataset['Gender'].value_counts()
          classes=dataset['Gender'].unique()
          dats.plot(kind='pie',autopct="%0.1f%%",labels=classes)
```

Out[49]: <AxesSubplot:ylabel='Gender'>


```
In [67]:  cc=dataset['judge_position'].value_counts()
          cc1=cc.head()
          print(cc1);
```

```
district and sessions court    16475
chief judicial magistrate      11688
civil judge senior division    11369
civil judge junior division     8190
civil court                     4358
Name: judge_position, dtype: int64
```