

Nama : Ricky Agung Sumiranto

Npm : 2013020101

Kelas : 3B

Matkul : Data Maining

```
import pandas as pd
import numpy as np
```

```
#Membaca File CSV
file = pd.read_csv("data-hasil-penertiban-minuman-keras-satpol-pp-jakarta-utara-
periode-juni-2017.csv")

#Menampilkan data
file.head
```

```
<bound method NDFrame.head of
0      Kecamatan Pademangan      Ciu      NaN      Botol
1      Kecamatan Pademangan  Anggur Orang Tua  134.0      Botol
2      Kecamatan Pademangan      Guinness  12.0      Botol
3      Kecamatan Pademangan      Anggur Merah  84.0      Botol
4      Kecamatan Pademangan      Rajawali  84.0      Botol
5      Kecamatan Pademangan      Vodka  4.0      Botol
6      Kecamatan Pademangan      Intisari  9.0      NaN
7      Kecamatan Koja      Anggur Orang Tua  13.0      Botol
8      Kecamatan Koja      Anggur Merah  2.0      Botol
9      Kecamatan Koja      Brandy  60.0      Botol
10     Kecamatan Koja      Mansion  23.0      Botol
11     Kecamatan Koja      Rajawali  55.0      Botol
12     Kecamatan Koja      Anggur Ginseng  132.0      Botol
13     Kecamatan Koja      Vodka  20.0      Botol
14     Kecamatan Cilincing      Anggur Merah  9.0      Botol
15     Kecamatan Cilincing      Brandy  181.0      Botol
16     Kecamatan Cilincing      Rajawali  26.0      Botol
17     Kecamatan Cilincing      Kolesom Kecil  9.0      Botol
18     Kecamatan Cilincing      Kolesom Besar  10.0      Botol
19     Kecamatan Cilincing      Aseng  41.0      Botol
20     KecamatanTanjung Priok      Brandy  60.0      Botol
21     KecamatanTanjung Priok      Rajawali  45.0      Botol
22     KecamatanTanjung Priok      Intisari  3.0      Botol
```

```
#melihat detail data
file.info
```

```
<bound method DataFrame.info of
0      Kecamatan Pademangan      Ciu      NaN      Botol
1      Kecamatan Pademangan      Anggur Orang Tua      134.0      Botol
2      Kecamatan Pademangan      Guinness      12.0      Botol
3      Kecamatan Pademangan      Anggur Merah      84.0      Botol
4      Kecamatan Pademangan      Rajawali      84.0      Botol
5      Kecamatan Pademangan      Vodka      4.0      Botol
6      Kecamatan Pademangan      Intisari      9.0      NaN
7      Kecamatan Koja      Anggur Orang Tua      13.0      Botol
8      Kecamatan Koja      Anggur Merah      2.0      Botol
9      Kecamatan Koja      Brandy      60.0      Botol
10     Kecamatan Koja      Mansion      23.0      Botol
11     Kecamatan Koja      Rajawali      55.0      Botol
12     Kecamatan Koja      Anggur Ginseng      132.0      Botol
13     Kecamatan Koja      Vodka      20.0      Botol
14     Kecamatan Cilincing      Anggur Merah      9.0      Botol
15     Kecamatan Cilincing      Brandy      181.0      Botol
16     Kecamatan Cilincing      Rajawali      26.0      Botol
17     Kecamatan Cilincing      Kolesom Kecil      9.0      Botol
18     Kecamatan Cilincing      Kolesom Besar      10.0      Botol
19     Kecamatan Cilincing      Aseng      41.0      Botol
20     KecamatanTanjung Priok      Brandy      60.0      Botol
21     KecamatanTanjung Priok      Rajawali      45.0      Botol
22     KecamatanTanjung Priok      Intisari      3.0      Botol
```

```
#melihat jumlah missing data / data kosong
file.isnull().sum()
```

```
wilayah      0
jenis_minuman  0
jumlah      1
volume      1
dtype: int64
```

```
#melihat presentase
#jika di baca = file.datakosong.jumlahkan * banyak data 100% di bagi banyak data
31
file.isnull().sum() * 100 / len(file)
```

```
wilayah      0.000000
jenis_minuman  0.000000
jumlah      3.333333
volume      3.333333
dtype: float64
```

```
#mengisi bagian kosong pada kolom jumlah dengan Median
file['jumlah'] = file['jumlah'].fillna(file['jumlah'].median())
```

```
#mengisi bagian kosong pada kolom volume dengan Modus
file['volume'] = file['volume'].fillna(file['volume'].mode())
```

```
#melihat data mising
file.isnull().sum()
```

```
wilayah      0
jenis_minuman 0
jumlah       0
volume       1
dtype: int64
```

```
#menghapus baris/kolom yang terdapat missing
file = file.dropna()
```

```
#melakukan pengecekan apakah masih ada mising
file.isnull().sum()
```

```
wilayah      0
jenis_minuman 0
jumlah       0
volume       0
dtype: int64
```

```
#hapus kolom volume dan tampilkan data
#jika 0 Baris, 1 Kolom
file.drop('volume', axis=1)
```

	wilayah	jenis_minuman	jumlah
0	Kecamatan Pademangan	Ciu	41.0
1	Kecamatan Pademangan	Anggur Orang Tua	134.0
2	Kecamatan Pademangan	Guinnes	12.0
3	Kecamatan Pademangan	Anggur Merah	84.0
4	Kecamatan Pademangan	Rajawali	84.0
5	Kecamatan Pademangan	Vodka	4.0
7	Kecamatan Koja	Anggur Orang Tua	13.0
8	Kecamatan Koja	Anggur Merah	2.0
9	Kecamatan Koja	Brandy	60.0
10	Kecamatan Koja	Mansion	23.0
11	Kecamatan Koja	Rajawali	55.0
12	Kecamatan Koja	Anggur Ginseng	132.0
13	Kecamatan Koja	Vodka	20.0
14	Kecamatan Cilincing	Anggur Merah	9.0
15	Kecamatan Cilincing	Brandy	181.0
16	Kecamatan Cilincing	Rajawali	26.0
17	Kecamatan Cilincing	Kolesom Kecil	9.0

file.head

✓ 0.4s

```
<bound method NDFrame.head of
0      Kecamatan Pademangan      Ciu      41.0  Botol
1      Kecamatan Pademangan  Anggur Orang Tua  134.0  Botol
2      Kecamatan Pademangan      Guinness  12.0  Botol
3      Kecamatan Pademangan      Anggur Merah  84.0  Botol
4      Kecamatan Pademangan      Rajawali  84.0  Botol
5      Kecamatan Pademangan      Vodka  4.0  Botol
7      Kecamatan Koja  Anggur Orang Tua  13.0  Botol
8      Kecamatan Koja  Anggur Merah  2.0  Botol
9      Kecamatan Koja  Brandy  60.0  Botol
10     Kecamatan Koja  Mansion  23.0  Botol
11     Kecamatan Koja  Rajawali  55.0  Botol
12     Kecamatan Koja  Anggur Ginseng  132.0  Botol
13     Kecamatan Koja  Vodka  20.0  Botol
14     Kecamatan Cilincing  Anggur Merah  9.0  Botol
15     Kecamatan Cilincing  Brandy  181.0  Botol
16     Kecamatan Cilincing  Rajawali  26.0  Botol
17     Kecamatan Cilincing  Kolesom Kecil  9.0  Botol
18     Kecamatan Cilincing  Kolesom Besar  10.0  Botol
19     Kecamatan Cilincing  Aseng  41.0  Botol
```