

PYTHON PROJECT TASK 2



George Billy Nanlohy

03411942000005

08 March 2019

GEOPHYSICAL ENGINEERING DEPARTMENT
FACULTY OF CIVIL, PLANNING, AND EARTH ENGINEERING
SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY

In this second assignment we have been taught how to analyze data in a form CSV or XLS. The following image is a data processing:

```
In [3]: 1 import pandas as pd
```

```
In [4]: 1 import matplotlib.pyplot as plt
```

```
In [5]: 1 iris = pd.read_csv('http://data.jakarta.go.id/dataset/8970f5d3-519d-4def-8cf4-47a71dad3baa/resource/75332e35-199d-45cf-a19f-
< >')
```

```
In [6]: 1 iris.head()
```

Out[6]:

	tahun	kota	jenis_transmigrasi	target	realisasi_kk	realisasi_orang
0	2011	Jakarta Selatan	Transmigrasi Umum	23	21	77
1	2011	Jakarta Selatan	Transmigrasi Swakarsa Mandiri	4	4	13
2	2011	Jakarta Timur	Transmigrasi Umum	23	29	114
3	2011	Jakarta Timur	Transmigrasi Swakarsa Mandiri	4	1	3
4	2011	Jakarta Pusat	Transmigrasi Umum	23	9	36

```
In [7]: 1 iris.tail()
```

Out[7]:

	tahun	kota	jenis_transmigrasi	target	realisasi_kk	realisasi_orang
5	2011	Jakarta Pusat	Transmigrasi Swakarsa Mandiri	4	0	0
6	2011	Jakarta Barat	Transmigrasi Umum	23	32	106
7	2011	Jakarta Barat	Transmigrasi Swakarsa Mandiri	4	6	15
8	2011	Jakarta Utara	Transmigrasi Umum	23	23	106
9	2011	Jakarta Utara	Transmigrasi Swakarsa Mandiri	4	2	8

(Figure.1) Processing data with head and tail syntax

```
In [8]: 1 iris.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 6 columns):
tahun          10 non-null int64
kota           10 non-null object
jenis_transmigrasi  10 non-null object
target         10 non-null int64
realisasi_kk    10 non-null int64
realisasi_orang  10 non-null int64
dtypes: int64(4), object(2)
memory usage: 608.0+ bytes
```

```
In [9]: 1 iris.describe()
```

Out[9]:

	tahun	target	realisasi_kk	realisasi_orang
count	10.0	10.000000	10.000000	10.000000
mean	2011.0	13.500000	12.700000	47.800000
std	0.0	10.013879	12.293178	47.506257
min	2011.0	4.000000	0.000000	0.000000
25%	2011.0	4.000000	2.500000	9.250000
50%	2011.0	13.500000	7.500000	25.500000
75%	2011.0	23.000000	22.500000	98.750000
max	2011.0	23.000000	32.000000	114.000000

(Figure.2) Processing data with info and describe syntax

```
In [10]: 1 iris.groupby('realisasi_kk').mean()
```

```
Out[10]:
```

	tahun	target	realisasi_orang
realisasi_kk			
0	2011	4	0
1	2011	4	3
2	2011	4	8
4	2011	4	13
6	2011	4	15
9	2011	23	36
21	2011	23	77
23	2011	23	106
29	2011	23	114
32	2011	23	106

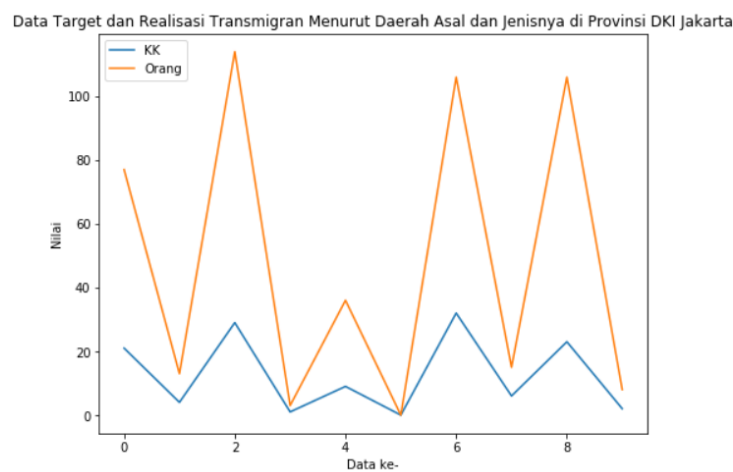
```
In [11]: 1 iris.groupby('realisasi_orang').mean()
```

```
Out[11]:
```

	tahun	target	realisasi_kk
realisasi_orang			
0	2011.0	4.0	0.0
3	2011.0	4.0	1.0
8	2011.0	4.0	2.0
13	2011.0	4.0	4.0
15	2011.0	4.0	6.0
36	2011.0	23.0	9.0
77	2011.0	23.0	21.0
106	2011.0	23.0	27.5
114	2011.0	23.0	29.0

(Figure.3) grouping data processing with syntax groupby.mean

```
In [18]: 1 plt.figure(figsize=(8,6))
2 plt.plot(iris['realisasi_kk'], label='KK')
3 plt.plot(iris['realisasi_orang'], label='Orang')
4
5 plt.xlabel('Data ke-')
6 plt.ylabel('Nilai')
7 plt.title('Data Target dan Realisasi Transmigran Menurut Daerah Asal dan Jenisnya di Provinsi DKI Jakarta')
8 plt.legend()
9 plt.show()
10
```



(Figure.4) Graph processing

Based on the pictures above it can be seen that the graph is the result of processing the target data and the realization of the origin and type of transmigrants in the DKI Jakarta province. The reason I took the data was because I wanted to know the comparative data of the realization of the head of the family with the realization of the transmigrants.

In figure 1 there is a head and tail function. The head function is inputted on that data to show the top 5 data. Meanwhile, the tail function is inputted on that data to show the bottom 5 data. In figure 2 there is an info function. This function is used in the data to show index numbers and data types. In picture 3 there is a describe function. This function is inputted on the data to show statistical summaries such as averages, medians, and quartiles and much more in that column. In Figure 4 there is a mean groupby function. This function is inputted for the average grouping of data inputted. In Figure 5 there is a plot for the graph and there is also a graphical interpretation of the data that I am processing. In making graphs it uses a function in the form of `matplotlib.pyplot` which is imported as `plt`. `Plt` here serves to plot a graph. After completing the plot, a graph interpretation is obtained, like the fifth image. The graph contains 2 graph lines which are graphs of the realization of the number of families and the realization of the number of people. Where the data according to its type in the province of DKI Jakarta.

On the graph, it can be seen that the number of people is more than the number of families. So it can be concluded that the realization of transmigrants in the type of number of people is more engaged in transmigration activities to the DKI Jakarta province.