

KBD_ACT_PERT 3

```
In [4]: # TUGAS  
# Nama : M. Imaduddin AL farobi  
# NPM  : 10123641
```

Overview Dataset

"Data Sensus Penghasilan". Data ini diambil dari database biro Sensus 1994 oleh Ronny Kohavi dan Barry Becker (Data Mining and Visualization, Silicon Graphics).

Data set ini berisikan beberapa kategori sebagai berikut

- age
- sex
- race
- education
- marital-status
- workclass
- occupation
- relationship
- native-country
- hours-per-week
- capital-loss
- capital-gain
- fnlwgt

Instruksi Tugas Praktikum FIKTI dan FTI

1. Buat dan jelaskan univariate analysis untuk kategori hours-per-week dan income
2. Buat dan jelaskan Bivariate analysis untuk hubungan antara hours-per-week dan income

1. Pengambilan Data: DAN IMPORT PACKAGE

```
In [ ]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
from scipy.stats import ttest_ind, ttest_rel
from scipy import stats
```

```
In [ ]: data = pd.read_csv("https://gitlab.com/andreass.bayu/file-directory/-/raw/main/adult.csv")
data.head(10)
```

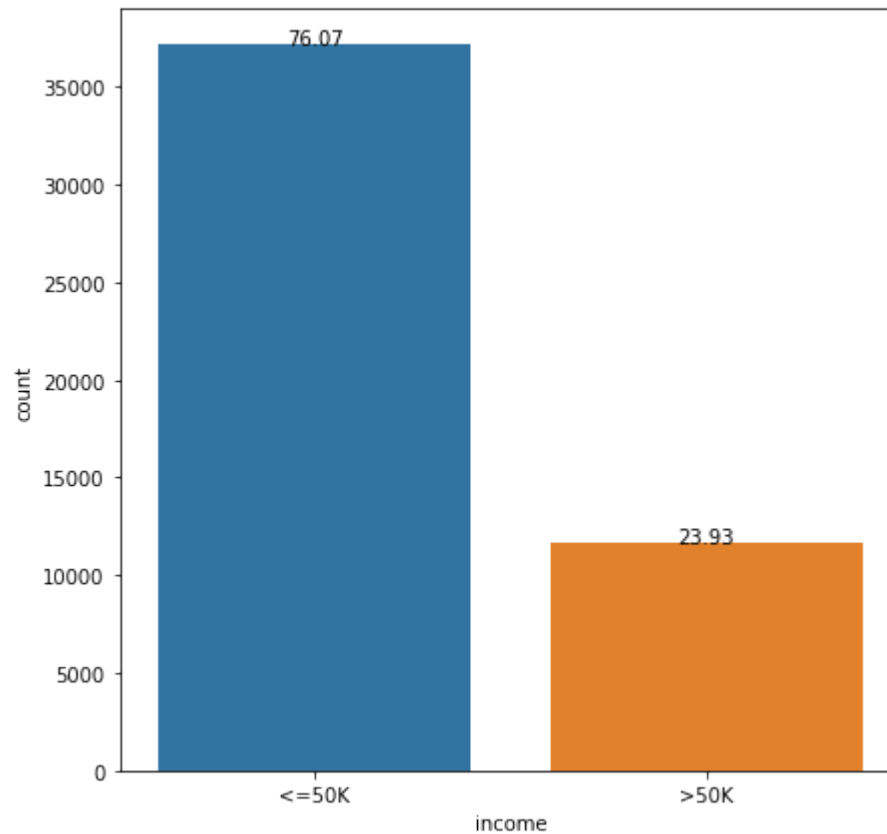
Out[21]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hours-per-week	native-country	income
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male	0	0	40	United-States	<=50K
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	50	United-States	<=50K
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White	Male	0	0	40	United-States	>50K
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male	7688	0	40	United-States	>50K
4	18	?	103497	Some-college	10	Never-married	?	Own-child	White	Female	0	0	30	United-States	<=50K
5	34	Private	198693	10th	6	Never-married	Other-service	Not-in-family	White	Male	0	0	30	United-States	<=50K
6	29	?	227026	HS-grad	9	Never-married	?	Unmarried	Black	Male	0	0	40	United-States	<=50K
7	63	Self-emp-not-inc	104626	Prof-school	15	Married-civ-spouse	Prof-specialty	Husband	White	Male	3103	0	32	United-States	>50K
8	24	Private	369667	Some-college	10	Never-married	Other-service	Unmarried	White	Female	0	0	40	United-States	<=50K
9	55	Private	104996	7th-8th	4	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	10	United-States	<=50K

#1. Buat dan jelaskan univariate analysis untuk kategori hours-per-week dan income

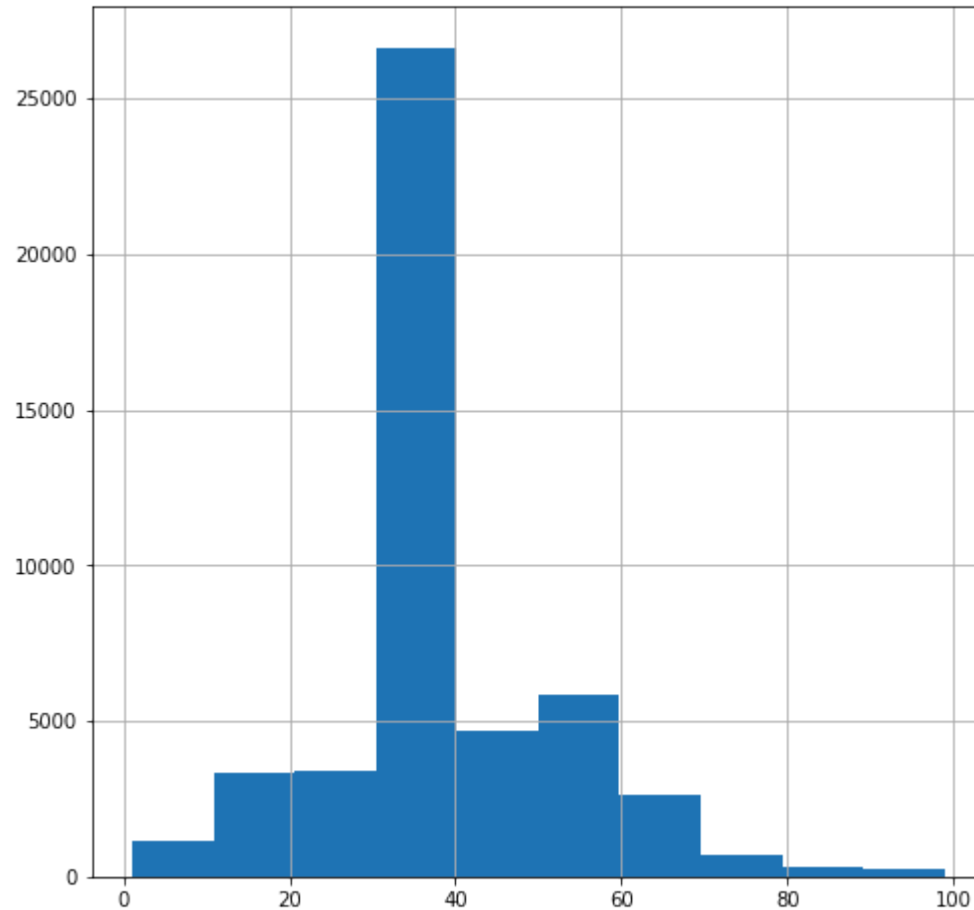
```
In [ ]: plt.figure(figsize=(7,7))
total = float(len(data) )

ax = sns.countplot(x="income", data=data)
for p in ax.patches:
    height = p.get_height()
    ax.text(p.get_x()+p.get_width()/2.,
            height + 3,
            '{:1.2f}'.format((height/total)*100),
            ha="center")
plt.show()
```



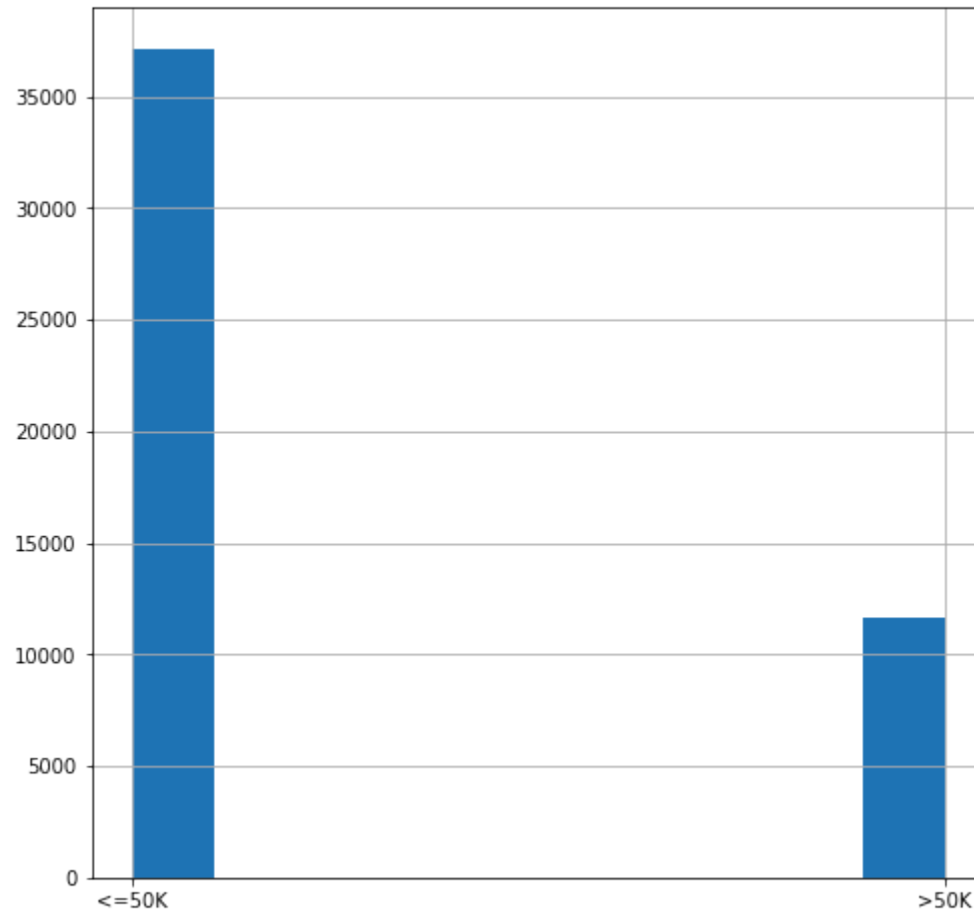
```
In [ ]: # menghitung diagram hours-per-week  
data['hours-per-week'].hist(figsize=(8,8))  
plt.show
```

Out[23]: <function matplotlib.pyplot.show>



histogram diatas dapat menjelaskan bahwa: atribut "hours per week" termasuk simetris stabil (jam kerja paling banyak dihasilkan diperoleh sebanyak 40 jam) Jam kerja minimum dan maksimum adalah 1 dan 99 jam.

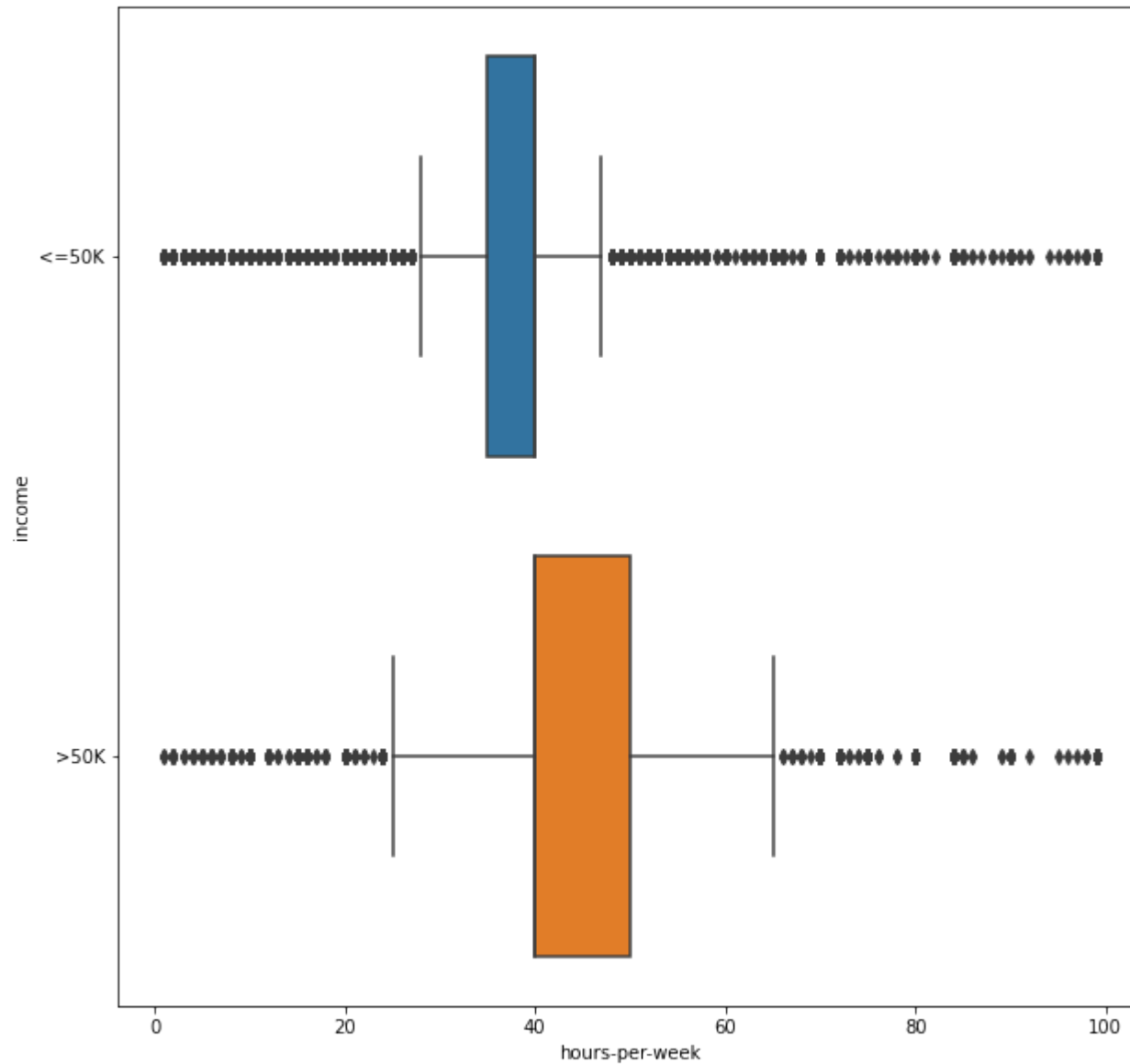
```
In [ ]: # buat diagram income  
data['income'].hist(figsize=(8,8))  
plt.show()
```



histogram diatas dapat menjelaskan bahwa: atribut "income" termasuk simetris tidak stabil (upah yang didapatkan lebih sedikit dari 50 ribu) Upah minimum dan maksimum adalah lebih kecil dari sama dengan 50 dan lebih besar dari 50 ribu

#2. Buat dan jelaskan Bivariate analysis untuk hubungan antara hours-per-week dan income

```
In [ ]: #hubungan hours per week dengan income
fig = plt.figure(figsize=(10,10))
sns.boxplot(x="hours-per-week", y="income", data=data)
plt.show()
```



```
In [ ]: data[['hours-per-week', 'income']].groupby(['income'],as_index=False).mean().sort_values(by='hours-per-week', ascending=True)
```

```
Out[26]:
```

	income	hours-per-week
1	>50K	45.452896
0	<=50K	38.840048

histogram diatas dapat menjelaskan bahwa:
atribut "hours per week - income" termasuk simetris
Upah minimum adalah kurang dari sama dengan 50 ribu dalam kurun waktu 38 jam per minggu.
Upah maksimum adalah lebih dari 50 ribu dalam kurun waktu 44 jam per minggu.

#KESIMPULAN

Disimpulkan bahwa terdapat perbedaan yang signifikan pada rerata jam kerja perminggu pada berpenghasilan >50K dan kelompok berpenghasilan <=50K.
Hal Ini membuktikan bahwa jam kerja mingguan memiliki peranan dalam membedakan kelompok pendapatan gaji paling besar.