

Nama : Muhammad Azfar Wisnu
Kelas : 2IA02
NPM : 50420803

Tugas

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. Univariate analysis untuk kategori hours-per-week dan income

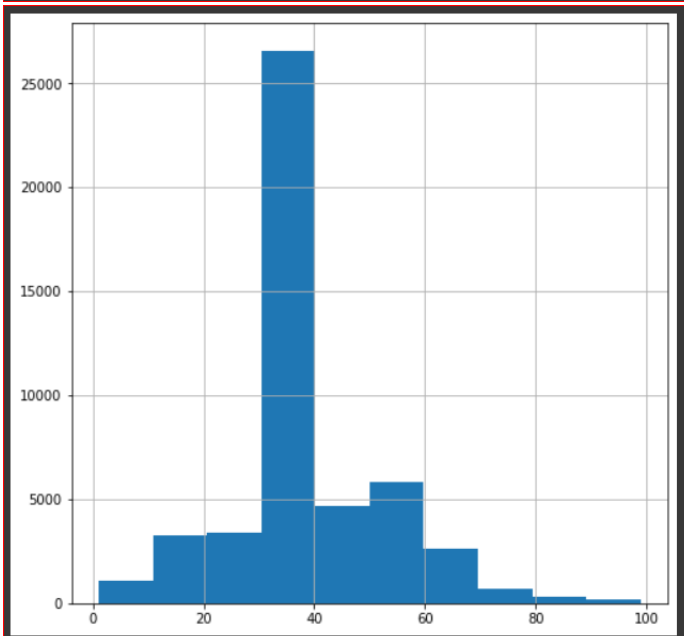
▼ Tugas

1. Univariate analysis

▼ 1.1 hours-per-week

▼ i. Distribusi

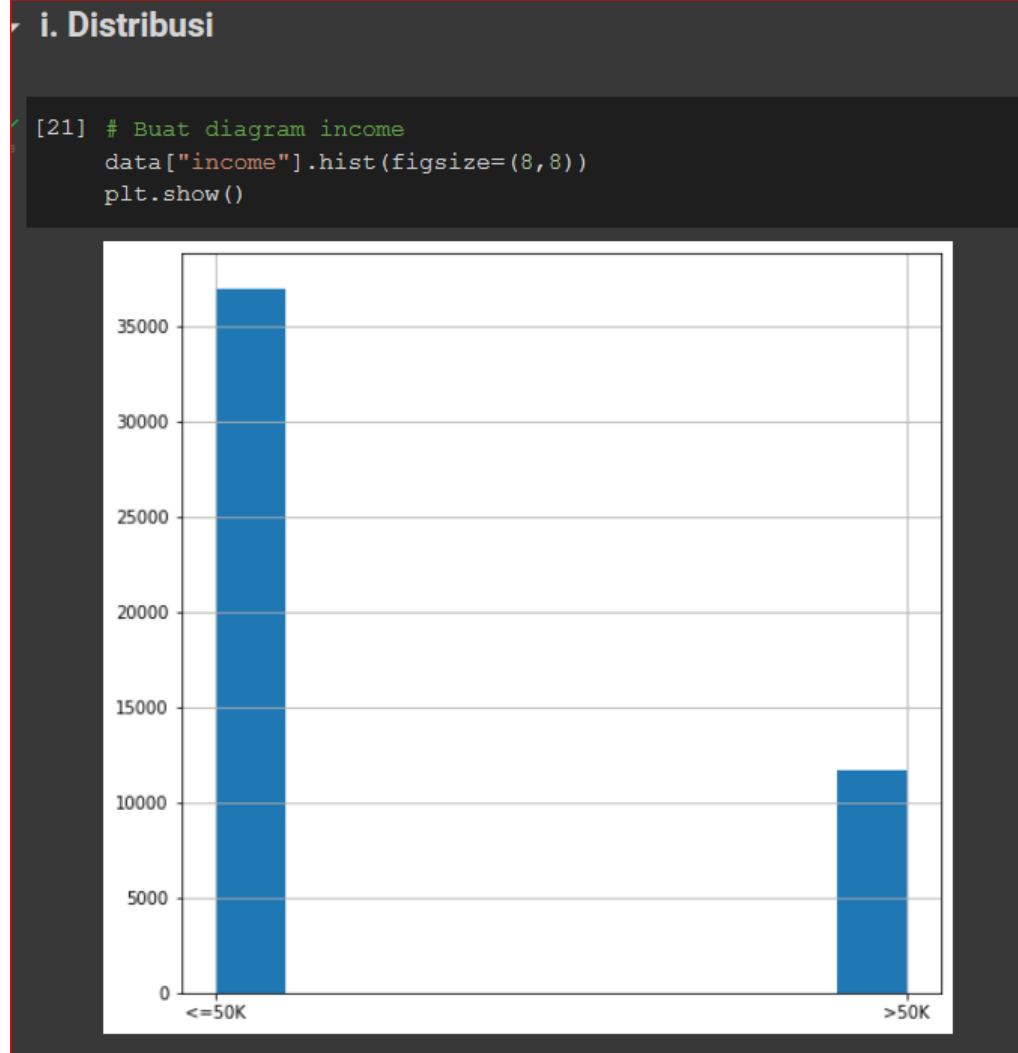
```
✓ [11] # Buat diagram hours-per-week  
data['hours-per-week'].hist(figsize=(8,8))  
plt.show()
```



Jelaskan hasil diagram tersebut

1. Banyak jam kerja hours-per-week berada pada nilai 40
2. rata rata pekerja berkerja dengan jam kerja 40 hour-per-week

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

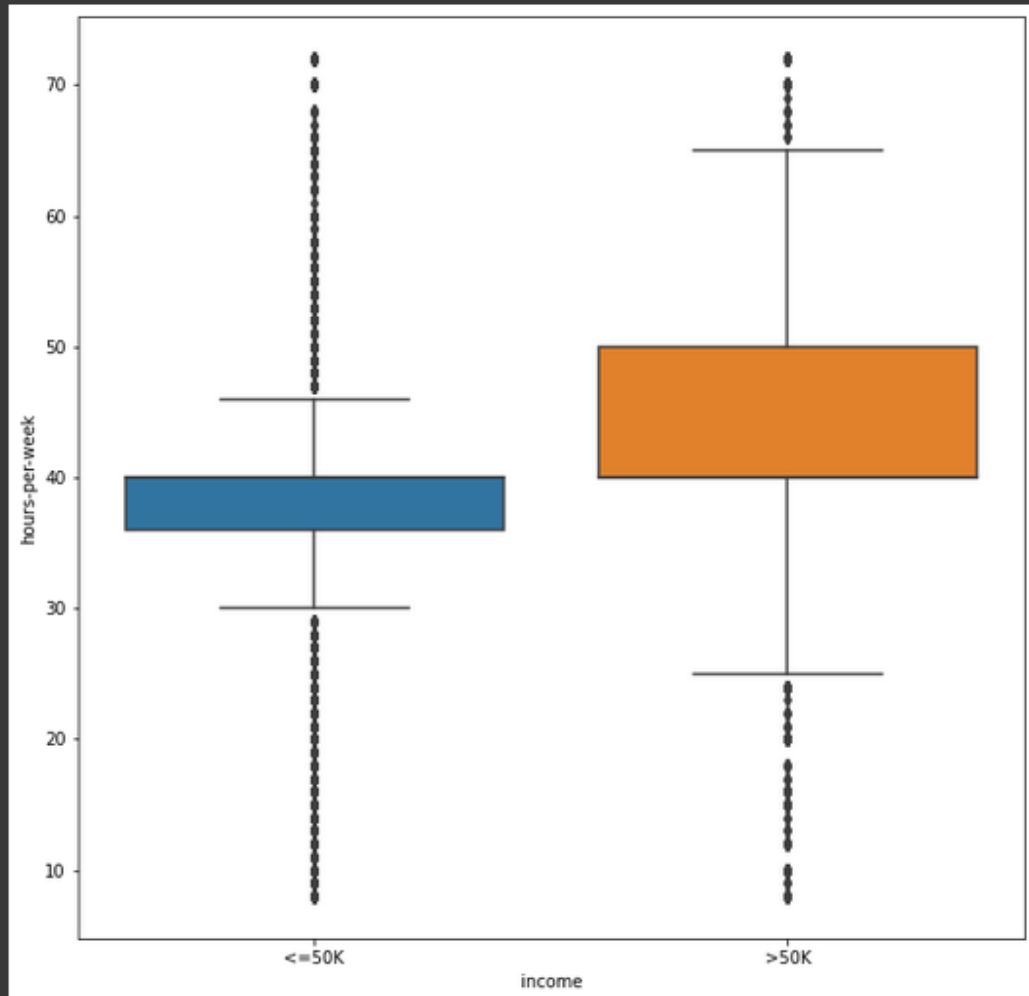


****ii. Deskripsi****

1. income yang didapat signifikan tidak melebihi 50k atau sama dengan 50 kan
2. incoem ditas 50 k mencapai 12000 value

2.1 hours-per-week - income (Hubungan)

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



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

	income	hours-per-week
1	>50K	44.859766
0	<=50K	38.698463

rata rata income >50k adalah 44.859 hours-per-week dan untuk income <= 50k adalah 38.698 hours-per-week

kelompok pendapatan > 50k memiliki median 30 hors-per-week dan untuk pendapatan <= 50k adalah 26 hours-per-week

iii. Uji Hipotesis (untuk menguji hubungan antara hours-per-week & Income)

Dalam contoh ini kita dapat mengatakan bahwa:

Null Hypothesis :- tidak ada perbedaan Rerata kelompok pendapatan >50k dan kelompok pendapatan <=50k.

Alternate Hypothesis :- ada perbedaan Rerata kelompok pendapatan >50k dan kelompok pendapatan <=50k.

```
[31] # Buat uji hipotesis untuk menentukan null atau alternate hypothesis
import random
```

```
data = data[(np.abs(stats.zscore(data["hours-per-week"])) < 3)]
```

```
income_1 = data[data['income']=='<=50K']['hours-per-week']
```

```
income_0 = data[data['income']=='>50K']['hours-per-week']
```

```
income_0 = income_0.values.tolist()
```

```
income_0 = random.sample(income_0, 100)
```

```
income_1 = income_1.values.tolist()
```

```
income_1 = random.sample(income_1, 100)
```

```
[33] from scipy.stats import ttest_ind
ttest,pval = ttest_ind(income_1,income_0,equal_var = False)
print("ttest",ttest)
print('p value',pval)
```

```
if pval <0.05:
    print("null hypothesis ditolak")
else:
    print("null hypothesis diterima")
```

```
ttest -4.350559331799701
p value 2.1802774517325487e-05
null hypothesis ditolak
```

Berikan kesimpulan

dapat disimpulkan bahwa terdapat perbedaan signifikan pada hours-per-week berpenghasilan > 50k dan kelompok berpenghasilan <= 50k
hal ini membuktikan bahwa hours-per-week memberikan peranana dalam membedakan kelompok pendapatan