Elmo Allistair - 12118220 - 4KA17

In [2]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [15]:

```
df = pd.read_csv("https://gitlab.com/andreass.bayu/file-directory/-/raw/main/adult.csv")
df.head()
```

Out[15]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hou p we
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male	0	0	
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White	Male	0	0	
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male	7688	0	
4	18	?	103497	Some-college	10	Never-married	?	Own-child	White	Female	0	0	

In [18]:

```
data = df.replace("?", np.nan)
data.dropna(inplace=True)
data.isnull().sum()
```

Out[18]:

```
age                0
workclass          0
fnlwgt             0
education          0
educational-num    0
marital-status     0
occupation         0
relationship       0
race               0
gender             0
capital-gain       0
capital-loss       0
hours-per-week     0
native-country     0
income             0
dtype: int64
```

In [19]:

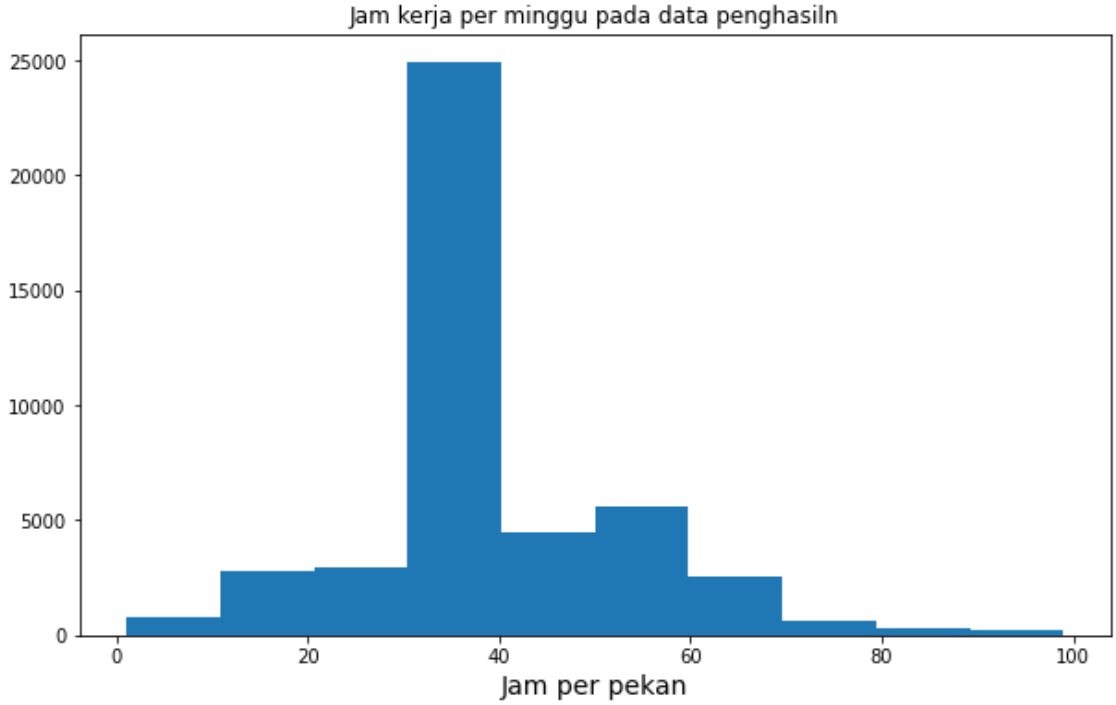
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 45222 entries, 0 to 48841
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   age              45222 non-null  int64
 1   workclass        45222 non-null  object
 2   fnlwgt           45222 non-null  int64
 3   education        45222 non-null  object
 4   educational-num  45222 non-null  int64
 5   marital-status   45222 non-null  object
 6   occupation       45222 non-null  object
 7   relationship     45222 non-null  object
 8   race             45222 non-null  object
 9   gender           45222 non-null  object
 10  capital-gain     45222 non-null  int64
 11  capital-loss     45222 non-null  int64
 12  hours-per-week   45222 non-null  int64
 13  native-country   45222 non-null  object
 14  income           45222 non-null  object
dtypes: int64(6), object(9)
memory usage: 5.5+ MB
```

Univariate Analysis

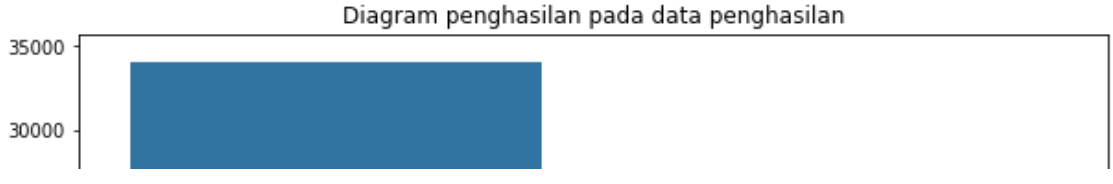
In [24]:

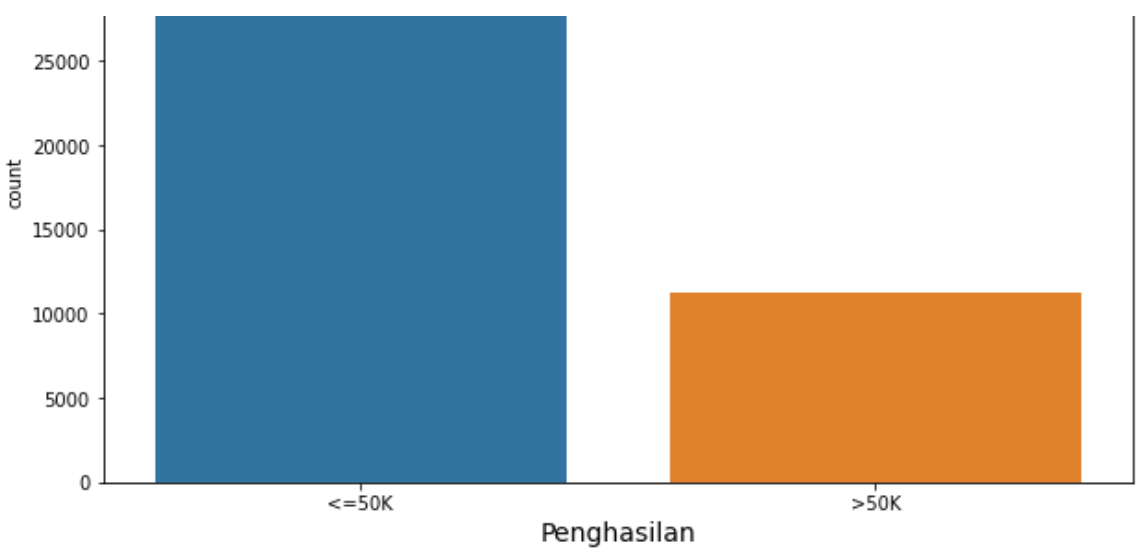
```
plt.figure(figsize=(10,6))
plt.hist(x=data['hours-per-week'])
plt.title("Jam kerja per minggu pada data penghasiln")
plt.xlabel("Jam per pekan", size=14)
plt.show()
```



In [27]:

```
plt.figure(figsize=(10,6))
sns.countplot(x=data.income)
plt.title("Diagram penghasilan pada data penghasilan")
plt.xlabel("Penghasilan", size=14)
plt.show()
```





Bivariate Analysis

In [30]:

```
plt.figure(figsize=(10,6))
sns.barplot(x=data.income, y=data['hours-per-week'])
plt.title("Diagram penghasilan dan waktu jam kerja per minggu")
plt.xlabel("Penghasilan", size=14)
plt.ylabel("Jam per minggu", size=14)
plt.show()
```