

Nama : Moh Wildan Haikhal

Kelas : 5MIM

Saya menggunakan dataset Employee Sample Data untuk mencari jumlah laki –laki dan perempuan pada dataset tersebut dan juga jumlah orang berdasarkan jenis pekerjaannya

- Import library python yang diperlukan untuk melakukan pre processing dan visualisasi

```
# default
import numpy as np
import pandas as pd

# visualization
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import missingno as msno
import networkx as nx
import plotly.graph_objects as go
import country_converter as coco
from wordcloud import WordCloud
import pycountry
```

- Melakukan proses pre processing pada dataset untuk menghilangkan missing value, duplikat dan lainnya

```
data_na = (df.isnull().sum() / len(df)) * 100
data_na = data_na.drop(data_na[data_na == 0].index).sort_values(ascending=False)[:30]
missing_data = pd.DataFrame({'Missing Ratio' :data_na})
missing_data.head(20)
```

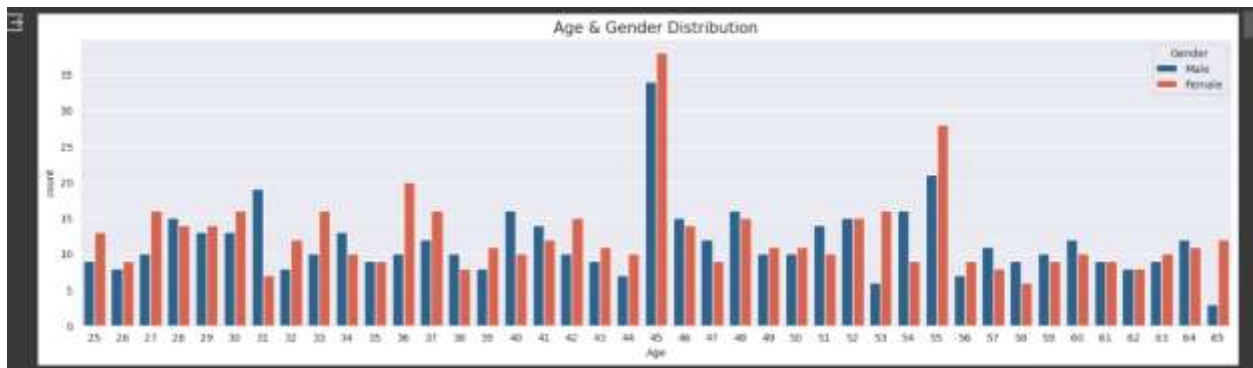
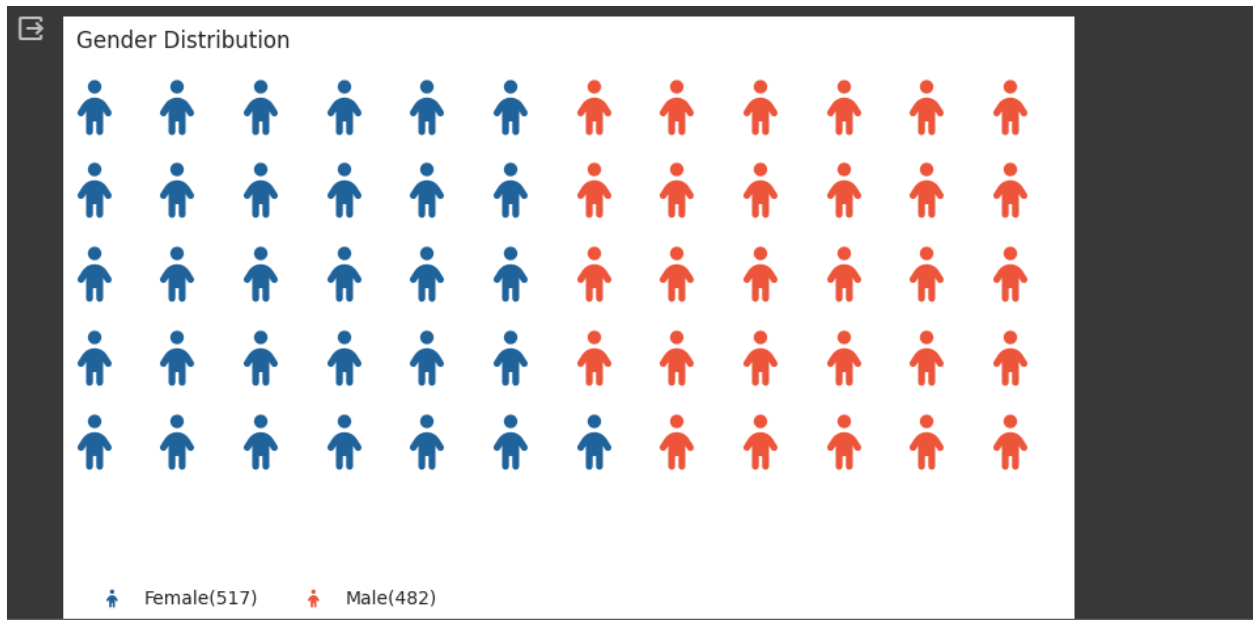
	Missing Ratio
Exit Date	91.5

```
[129] duplicate_rows_data = df[df.duplicated()]
print("number of duplicate rows: ", duplicate_rows_data.shape)

number of duplicate rows: (0, 14)

[130] df.drop(0, axis=0, inplace=True)
```

- Melakukan visuilisasi pada dataset employee sample data



Visualisasi gender distribution untuk menampilkan jumlah laki –laki dan perempuan pada dataset tersebut