Tugas 3 Data Sains dan Analisis

Menelaah Data dengan Visualisasi

Visualisasi Variabel

Menggunakan data dari tugas 2, pada tugas 3 kali ini data dapat ditelaah dengan beberapa visualisasi, diantaranya:

1. **Bar Chart**

Source Code

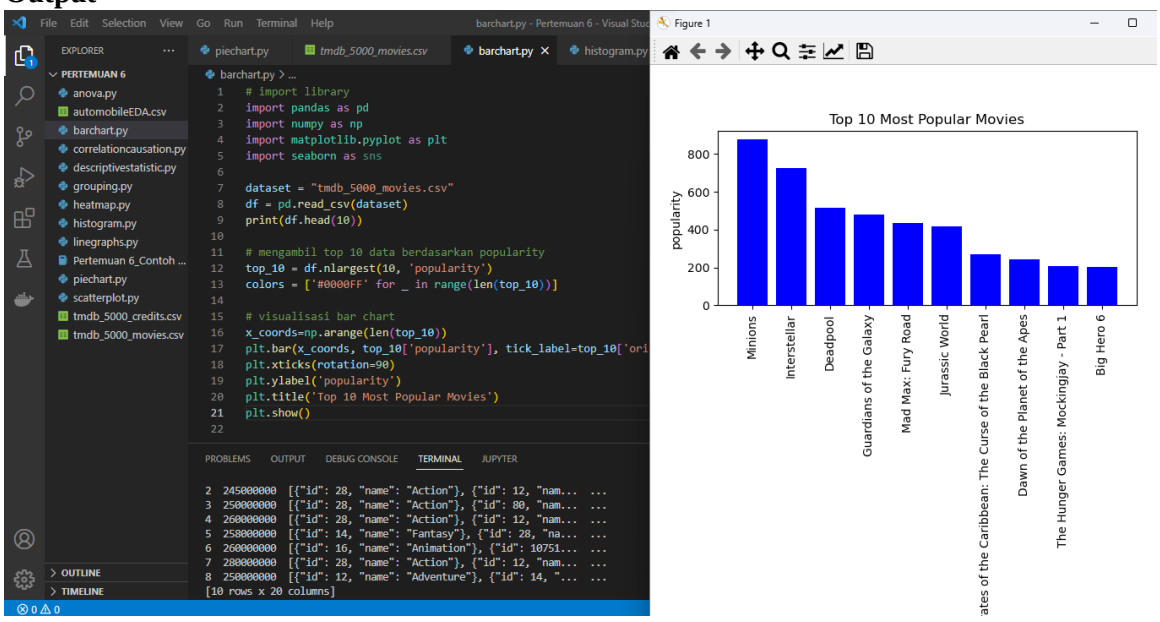
```
# import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

dataset = "tmdb_5000_movies.csv"
df = pd.read_csv(dataset)
print(df.head(10))

# mengambil top 10 data berdasarkan popularity
top_10 = df.nlargest(10, 'popularity')
colors = ['#0000FF' for _ in range(len(top_10))]

# visualisasi bar chart
x_coords=np.arange(len(top_10))
plt.bar(x_coords, top_10['popularity'],
tick_label=top_10['original_title'], color=colors)
plt.xticks(rotation=90)
plt.ylabel('popularity')
plt.title('Top 10 Most Popular Movies')
plt.show()
```

Output



2. Line Chart

Source Code

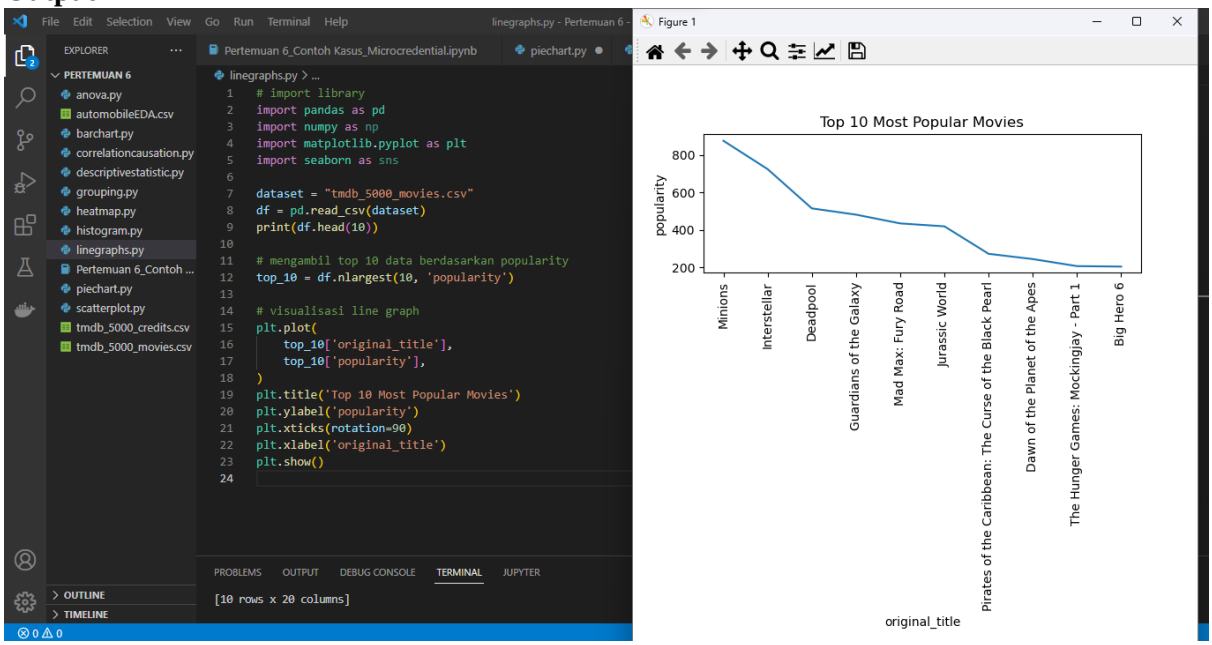
```
# import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

dataset = "tmdb_5000_movies.csv"
df = pd.read_csv(dataset)
print(df.head(10))

# mengambil top 10 data berdasarkan popularity
top_10 = df.nlargest(10, 'popularity')

# visualisasi line graph
plt.plot(
    top_10['original_title'],
    top_10['popularity'],
)
plt.title('Top 10 Most Popular Movies')
plt.ylabel('popularity')
plt.xticks(rotation=90)
plt.xlabel('original_title')
plt.show()
```

Output



Visualisasi Statistik

1. Histogram

Source Code

```
# import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# memuat dataset
dataset = "tmdb_5000_movies.csv"
df = pd.read_csv(dataset)

# visualisasi histogram
plt.hist(df['budget'], bins=50)
plt.title('Movie Budget Distribution')
plt.xlabel('Budget')
plt.ylabel('Count')
plt.show()
```

Output