

Praktikum Data Mining

Minggu Ke-8

Faris Saifullah
3124640034

```
[6] import pandas as pd
    from sklearn.cluster import AgglomerativeClustering
    import numpy as np
```

```
#no.1
dataset = pd.read_csv('transaction.csv')
dataset.head()
```

	InvoiceNo	StockCode	Qty	InvoiceDate	CustomerID	Country
0	537626	22725	830	12/7/2010 14:57	12347	Iceland
1	537626	22729	948	12/7/2010 14:57	12347	Iceland
2	537626	22195	695	12/7/2010 14:57	12347	Iceland
3	542237	22725	636	1/26/2011 14:30	12347	Iceland
4	542237	22729	536	1/26/2011 14:30	12347	Iceland

Analisa :
menampilkan data yang ada pada file transaction.csv

```
#no.2  
country_counts = dataset['Country'].value_counts()  
country_counts.head()
```

	count
Country	
Germany	2269
France	2109
EIRE	1620
Netherlands	634
Spain	539

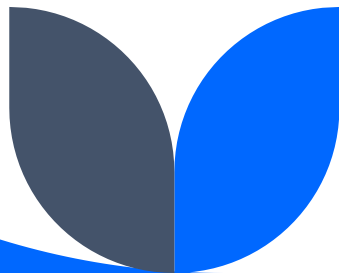
Analisa :
menampilkan data kemunculan tiap negara pada dataset
file transaction.csv



```
#no.3
transaction_counts = dataset.groupby(['Country', 'InvoiceNo'])['InvoiceNo'].nunique()
transaction_counts_by_country = transaction_counts.groupby('Country').sum()
transaction_counts_by_country.head()
```

InvoiceNo	
Country	
Australia	44
Austria	12
Bahrain	1
Belgium	84
Brazil	1

Analisa :
menampilkan data banyaknya transaksi pada tiap negara



```
#no.4
transaction_counts_df = transaction_counts_by_country.reset_index()
X = transaction_counts_df[['InvoiceNo']]
model = AgglomerativeClustering(n_clusters=3, linkage='average')
model.fit(X)
transaction_counts_df['cluster'] = model.labels_
transaction_counts_df.head(36)
```

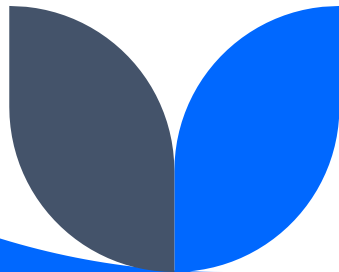
4	Brazil	1	0
5	Canada	3	0
6	Channel Islands	21	0
7	Cyprus	16	0
8	Czech Republic	2	0
9	Denmark	18	0
10	EIRE	224	2
11	European Community	3	0
12	Finland	26	0
13	France	344	1
14	Germany	377	1
15	Greece	5	0
16	Iceland	6	0
17	Israel	4	0

Analisa :
menampilkan clustering pada transaksi dengan
Average Linkage, dengan k=3

```
#no.5  
centroids = transaction_counts_df.groupby('cluster')['InvoiceNo'].mean()  
centroids.head()
```

InvoiceNo	
cluster	
0	18.787879
1	360.500000
2	224.000000

Analisa :
menampilkan posisi centroid dari setiap cluster



```
#no.6
sorted_centroids = centroids.sort_values()
sorted_centroids.head()
```

	InvoiceNo
cluster	
0	18.787879
2	224.000000
1	360.500000

Analisa :
menampilkan pengurutan posisi centroid secara ascending

```
#no.7
low_transaction_cluster = sorted_centroids.index[0]
medium_transaction_cluster = sorted_centroids.index[1]
high_transaction_cluster = sorted_centroids.index[2]

low_transaction_countries = transaction_counts_df[
    transaction_counts_df['cluster'] == low_transaction_cluster]['Country'].tolist()
medium_transaction_countries = transaction_counts_df[
    transaction_counts_df['cluster'] == medium_transaction_cluster]['Country'].tolist()
high_transaction_countries = transaction_counts_df[
    transaction_counts_df['cluster'] == high_transaction_cluster]['Country'].tolist()
print("Negara dengan transaksi rendah:")
print(low_transaction_countries)
print("\nNegara dengan transaksi sedang:")
print(medium_transaction_countries)
print("\nNegara dengan transaksi tinggi:")
print(high_transaction_countries)
```

```
Negara dengan transaksi rendah:
['Australia', 'Austria', 'Bahrain', 'Belgium', 'Brazil', 'Canada', 'Channel Islands', 'Cyprus', 'Czech Republic', 'Denmark', 'European Community', 'Finland', 'Greece', 'Iceland',

Negara dengan transaksi sedang:
['EIRE']

Negara dengan transaksi tinggi:
['France', 'Germany']
```

Analisa :
menampilkan negara mana saja yang transaksinya rendah,
sedang dan tinggi


```
#no.8
import matplotlib.pyplot as plt

transaction_counts_df_sorted = transaction_counts_df.sort_values('InvoiceNo')
plt.figure(figsize=(15, 6))
colors = {
    low_transaction_cluster: 'blue',
    medium_transaction_cluster: 'green',
    high_transaction_cluster: 'red'
}
for index, row in transaction_counts_df_sorted.iterrows():
    plt.scatter(index, row['InvoiceNo'], color=colors[row['cluster']])
plt.xlabel('Urutan Negara')
plt.ylabel('Jumlah Transaksi')
plt.title('Visualisasi Cluster Negara Berdasarkan Jumlah Transaksi')
plt.show()
```

Analisa :
menampilkan Visualisasi dengan warna yang berbeda untuk hasil cluster (no. 7), dimana sumbu x=urutan country dan sumbu y=transaksi

