Zad1

```
import pandas as pd
import numpy as np
```

```
data = pd.read_csv('penguins.csv')
data
```

	rowid	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm
0	1	Adelie	Torgersen	39.1	18.7	181.0
1	2	Adelie	Torgersen	39.5	17.4	186.0
2	3	Adelie	Torgersen	40.3	18.0	195.0
3	4	Adelie	Torgersen	NaN	NaN	NaN
4	5	Adelie	Torgersen	36.7	19.3	193.0
...	...	...	...	...	...	...
339	340	Chinstrap	Dream	55.8	19.8	207.0
340	341	Chinstrap	Dream	43.5	18.1	202.0
341	342	Chinstrap	Dream	49.6	18.2	193.0
342	343	Chinstrap	Dream	50.8	19.0	210.0
343	344	Chinstrap	Dream	50.2	18.7	198.0

344 rows × 9 columns

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 344 entries, 0 to 343
Data columns (total 9 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   rowid              344 non-null    int64  
 1   species            344 non-null    object 
 2   island             344 non-null    object 
 3   bill_length_mm     342 non-null    float64
 4   bill_depth_mm      342 non-null    float64
 5   flipper_length_mm  342 non-null    float64
 6   body_mass_g        342 non-null    float64
 7   sex                333 non-null    object 
 8   year               344 non-null    int64  
dtypes: float64(4), int64(2), object(3)
memory usage: 24.3+ KB
```

```
data.describe()
```

	rowid	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	
count	344.000000	342.000000	342.000000	342.000000	342.000000	
mean	172.500000	43.921930	17.151170	200.915205	4201.754386	2
std	99.448479	5.459584	1.974793	14.061714	801.954536	
min	1.000000	32.100000	13.100000	172.000000	2700.000000	2
25%	86.750000	39.225000	15.600000	190.000000	3550.000000	2
50%	172.500000	44.450000	17.300000	197.000000	4050.000000	2
75%	258.250000	48.500000	18.700000	213.000000	4750.000000	2
max	344.000000	59.600000	21.500000	231.000000	6300.000000	2

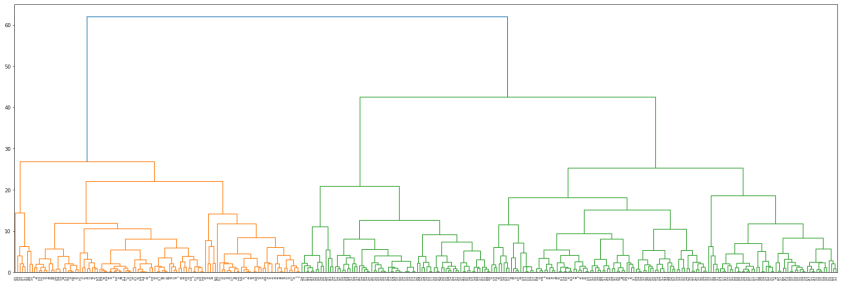
```
dataset = data[["bill_length_mm", "flipper_length_mm"]]
```

```
dataset = dataset.dropna(axis=0)
dataset
```

	bill_length_mm	flipper_length_mm
0	39.1	181.0
1	39.5	186.0
2	40.3	195.0
4	36.7	193.0
5	39.3	190.0
...	...	...
339	55.8	207.0
340	43.5	202.0
341	49.6	193.0

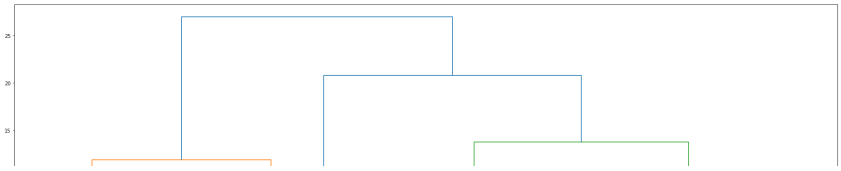
```
import matplotlib.pyplot as plt
from scipy.cluster import hierarchy
```

```
clusters = hierarchy.linkage(dataset, method="complete")
plt.figure(figsize=(30, 10))
dendrogram = hierarchy.dendrogram(clusters)
#plt.axhline(6, color='red');
```



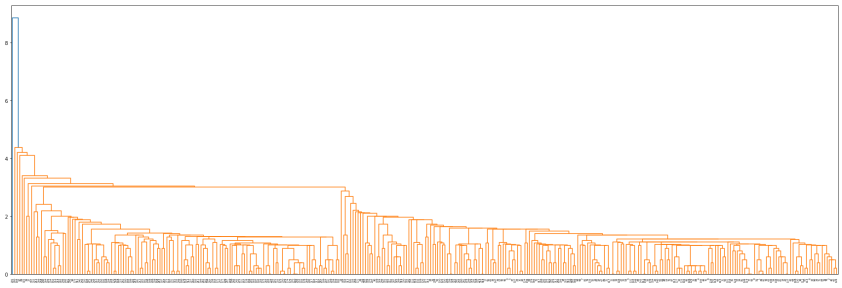
Liczba klastrów: 2

```
clusters = hierarchy.linkage(dataset, method="average")
plt.figure(figsize=(30, 10))
dendrogram = hierarchy.dendrogram(clusters)
#plt.axhline(6, color='red');
```



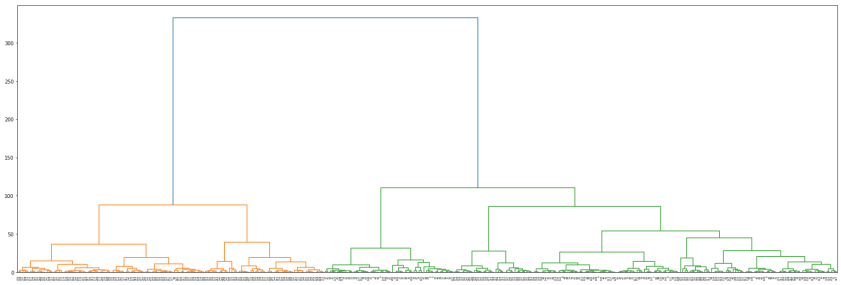
Liczba klastrów: 3

```
clusters = hierarchy.linkage(dataset, method="single")
plt.figure(figsize=(30, 10))
dendrogram = hierarchy.dendrogram(clusters)
#plt.axhline(6, color='red');
```



Liczba klastrów: 1

```
clusters = hierarchy.linkage(dataset, method="ward")
plt.figure(figsize=(30, 10))
dendrogram = hierarchy.dendrogram(clusters)
#plt.axhline(6, color='red');
```



Liczba klastrów: 2

```
from sklearn.cluster import AgglomerativeClustering
```

```
k = 3
model = AgglomerativeClustering(linkage='complete',n_clusters=k,affinity='euclidean', distance_threshold=None)
model.fit(dataset)
```

```
AgglomerativeClustering(linkage='complete', n_clusters=3)
```

```
model.labels_
```

```
array([0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1,
       0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0,
       0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1,
       1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,
       1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 2, 1,
       2, 2, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 1,
       2, 2, 2, 1, 2, 2, 1, 1, 2, 2, 2, 1, 2, 1, 1, 1, 2, 1, 2, 2, 2, 1,
       2, 1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 2,
       2, 1, 2, 1, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 1, 1, 2, 2, 1,
       2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1,
       1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1])
```

```
complete=model.labels_
```

```
k = 3
model = AgglomerativeClustering(linkage='average',n_clusters=k,affinity='euclidean', distance_threshold=None)
model.fit(dataset)
```

```
AgglomerativeClustering(linkage='average', n_clusters=3)
```

```
model.labels_
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0])
```

```
average=model.labels_
```

```
k = 3
model = AgglomerativeClustering(linkage='single',n_clusters=k,affinity='euclidean', distance_threshold=None)
model.fit(dataset)
```

```
AgglomerativeClustering(linkage='single', n_clusters=3)
```

```
model.labels_
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

```
single=model.labels_
```

```
k = 3
model = AgglomerativeClustering(linkage='ward',n_clusters=k,affinity='euclidean', distance_threshold=None)
model.fit(dataset)
```

```
AgglomerativeClustering(n_clusters=3)
```

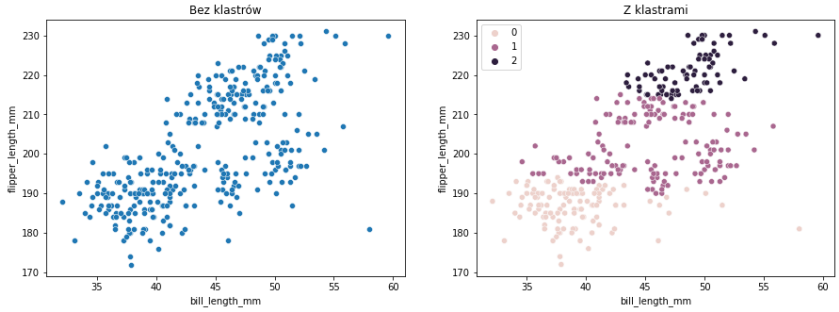
```
model.labels_
```

```
array([2, 2, 1, 1, 1, 2, 1, 1, 1, 2, 2, 2, 1, 1, 2, 1, 1, 2, 1, 2, 2, 1,
       2, 2, 2, 2, 1, 2, 2, 2, 2, 1, 2, 1, 1, 1, 2, 2, 2, 2, 1, 2, 1, 2,
       1, 2, 2, 1, 1, 2, 1, 1, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 2, 2, 2, 0, 1, 1, 2, 1, 1, 1, 2, 1, 1, 2, 1, 1, 2, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 1, 0, 1, 1, 1, 1,
       1, 2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 2, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1])
```

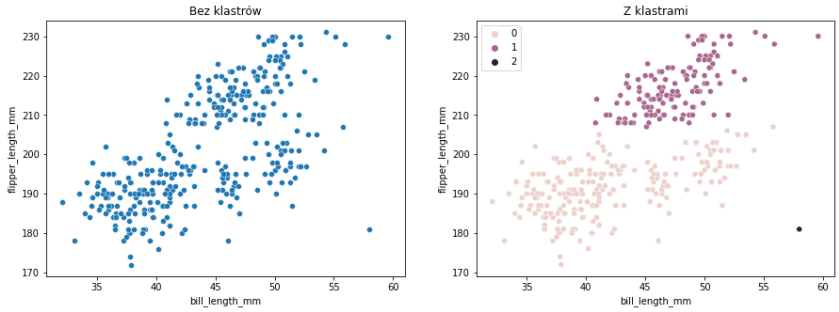
```
ward=model.labels_
```

```
import seaborn as sns
```

```
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15,5))
sns.scatterplot(ax=axes[0], data=dataset, x='bill_length_mm', y='flipper_length_mm').set_title('Bez klastrów')
sns.scatterplot(ax=axes[1], data=dataset, x='bill_length_mm', y='flipper_length_mm', hue=complete).set_title('Z klastrami');
```

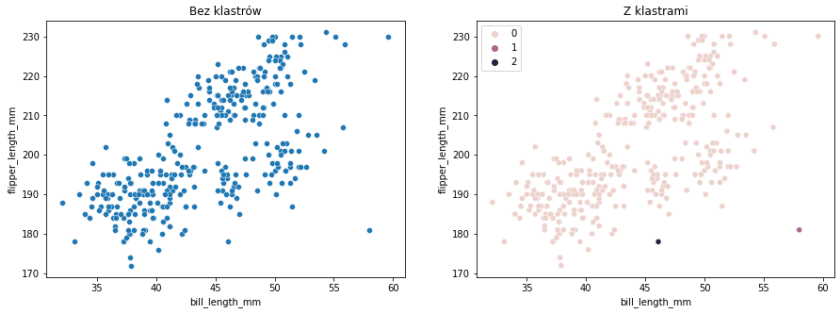



```
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15,5))
sns.scatterplot(ax=axes[0], data=dataset, x='bill_length_mm', y='flipper_length_mm').set_title('Bez klastrów')
sns.scatterplot(ax=axes[1], data=dataset, x='bill_length_mm', y='flipper_length_mm', hue=average).set_title('Z klastrami');
```




```
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15,5))
sns.scatterplot(ax=axes[0], data=dataset, x='bill_length_mm', y='flipper_length_mm').set_title('Bez klastrów')
```

```
sns.scatterplot(ax=axes[1], data=dataset, x='bill_length_mm', y='flipper_length_mm', hue=single).set_title('Z klastrami');
```




```
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15,5))
sns.scatterplot(ax=axes[0], data=dataset, x='bill_length_mm', y='flipper_length_mm').set_title('Bez klastrów')
sns.scatterplot(ax=axes[1], data=dataset, x='bill_length_mm', y='flipper_length_mm', hue=ward).set_title('Z klastrami');
```