In [ ]:
```
# Introduction

"""
In this project, I'm diving into unsupervised learning with a focus on cluster analysis.

I'll be using the famous Iris dataset, which records the sepal and petal measurements
of three different Iris flower species.

My goal is to apply K-Means clustering to see how well it can group the flowers
based on their physical traits without any prior labeling.

Through exploratory data analysis, I'll get familiar with the dataset before I put the K-Means algorithm to work.

I'm looking to see how effective K-Means clustering is and to understand the natural
groupings within the Iris dataset.

"""
```

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import datasets
```

In [2]:
```
iris = datasets.load_iris()
df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
df['target'] = iris.target
```

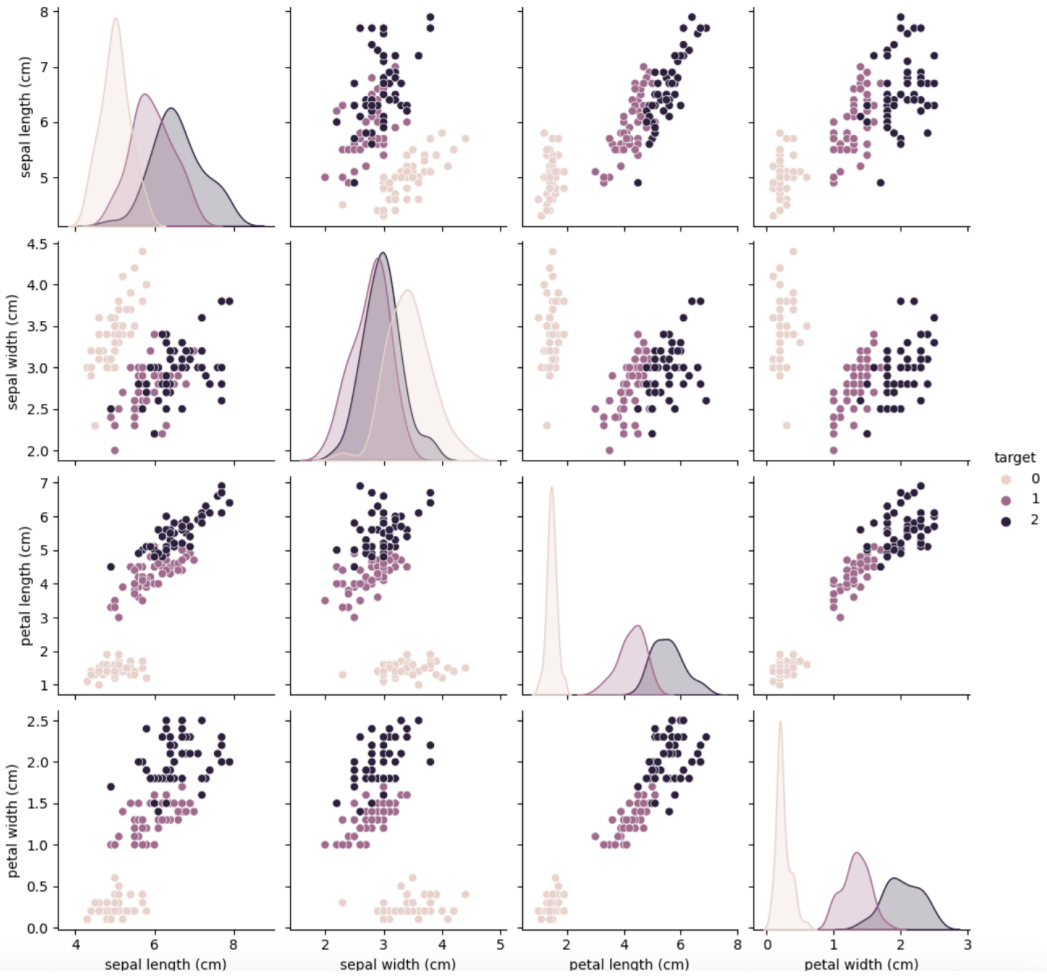
In [3]:
```
df.head()
```

Out[3]:

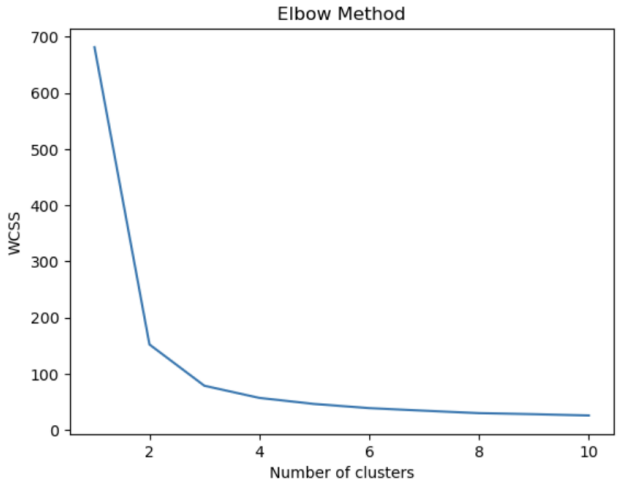
	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

In [5]:
```
df.describe()
df.isnull().sum()

sns.pairplot(df, hue='target')
plt.show()
```

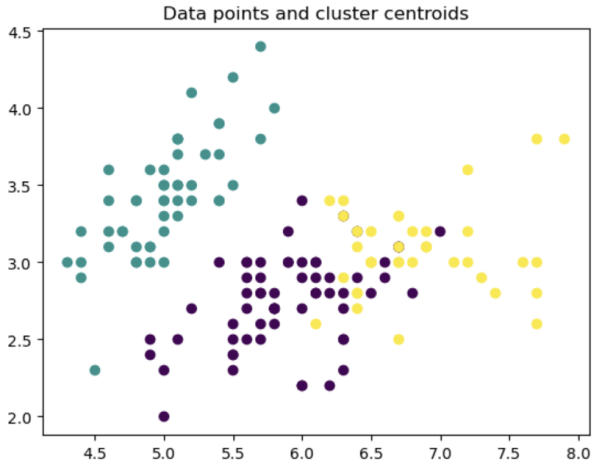
```
/Users/rahilchandra/anaconda3/lib/python3.11/site-packages/seaborn/axisgrid.py:118: UserWarning: The figure layout
has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



In [6]:
```
from sklearn.cluster import KMeans

wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=0)
    kmeans.fit(df.drop('target', axis=1))
    wcss.append(kmeans.inertia_)

plt.plot(range(1, 11), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()

kmeans = KMeans(n_clusters=3, init='k-means++', max_iter=300, n_init=10, random_state=0)
pred_y = kmeans.fit_predict(df.drop('target', axis=1))
df['cluster'] = pred_y
```



In [7]:
```
# Visualize the clusters
plt.scatter(df.iloc[:, 0], df.iloc[:, 1], c=df['cluster'], cmap='viridis')
plt.title('Data points and cluster centroids')
plt.show()
```



In [ ]:
```
# Conclusion

"""
To conclude, using K-Means to analyze the Iris dataset has been quite enlightening.
It's clear that unsupervised learning is a strong tool for spotting patterns and neatly grouping data.

I've found that K-Means is really good at telling the Iris species apart just by looking at their physical features.

Seeing the clusters laid out in the charts really drove home how precise the algorithm is.

This whole experience has shown me how valuable unsupervised learning can be,
especially when you don't have labels on your data. It's got a lot of potential in various fields,
not just in identifying flowers but also in areas like marketing. Like product grouping and targeted marketing.

Looking ahead, I'm thinking of trying out different clustering techniques
and seeing how they stack up against K-Means.
I'm also curious to see if playing around with the features could shine a new light on the data.
"""
```