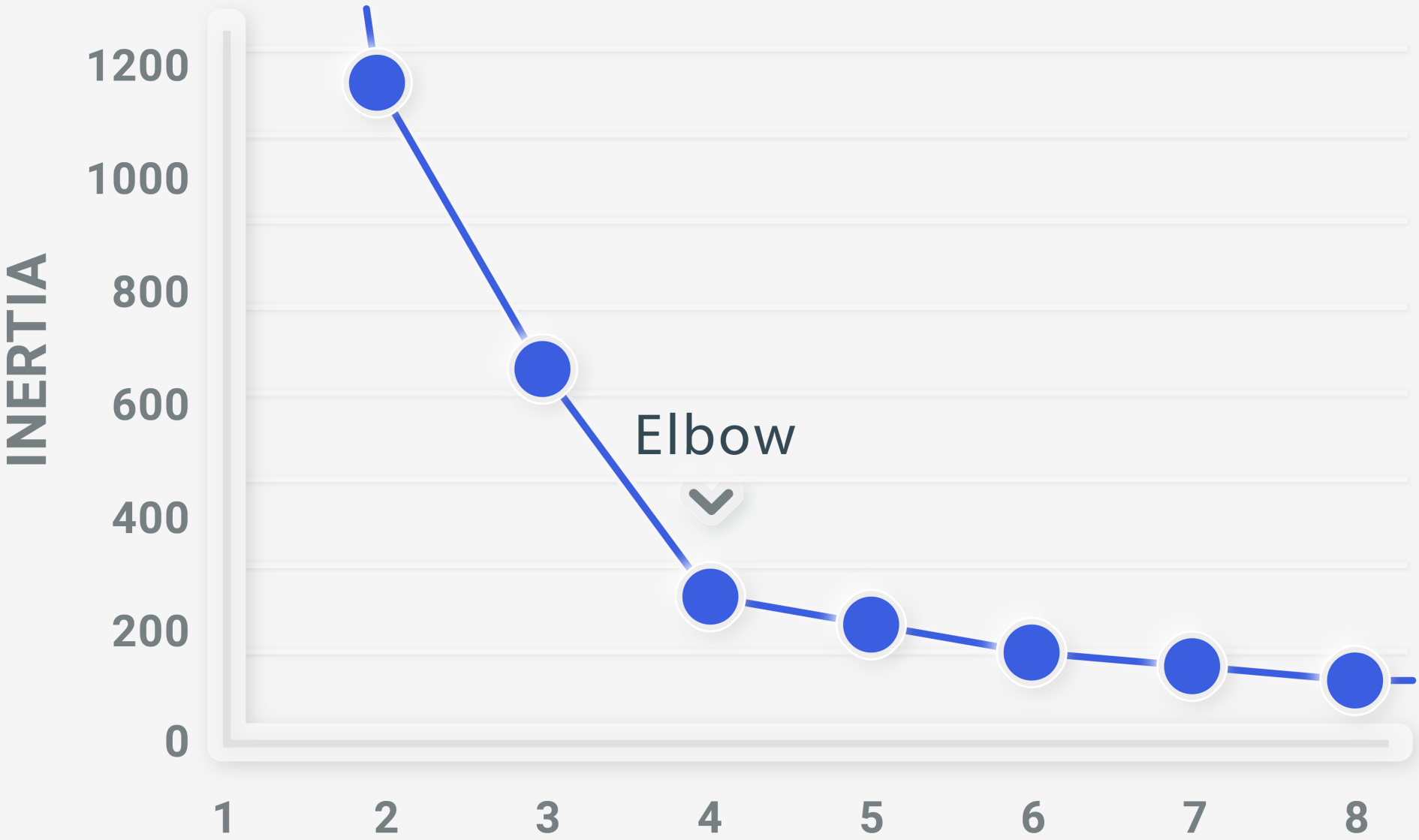
ELBOW CURVE TO FIND "K" CLUSTERS

When plotting the inertia as a function of the number of clusters k, the curve often contains an inflexion point called the “elbow”



DATA-DRIVEN SCIENCE

Elbow Curve to find "K" Clusters

The intuition behind the Elbow curve is that the explained variation changes rapidly until the number of groups you have in the data and then it slows down leading to an **elbow formation** in the graph as shown.

The Elbow point is the **number of clusters** you should use for your K-Means algorithm.

Elbow Curve to find "K" Clusters

You can use the library, **Yellowbrick** which can help you plot the Elbow curve with just 1 line of code. It is a wrapper around Scikit-Learn and hence integrate well with it.

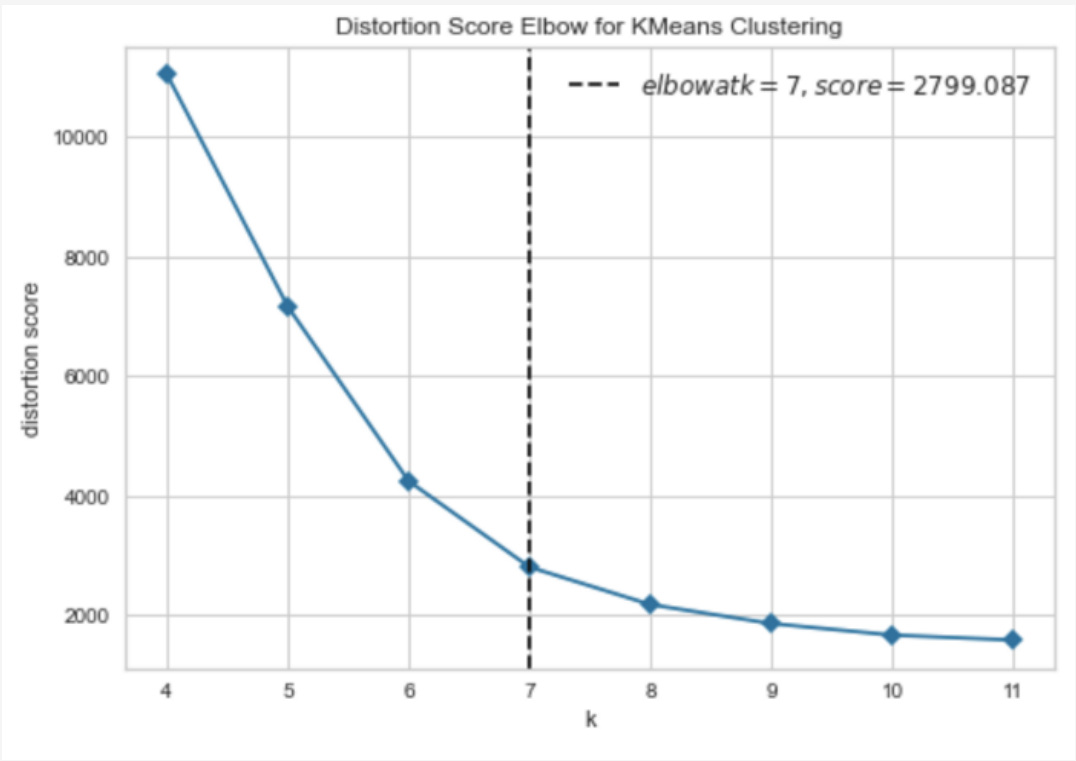
```
# Import ElbowVisualizer
from yellowbrick.cluster import KElbowVisualizer

model = KMeans()
# k is range of number of clusters.
visualizer = KElbowVisualizer(model, k=(4,12),
timings=False)

visualizer.fit(X)        # Fit the data to the visualizer
visualizer.show()        # Finalize and render the figure
```

Source: towardsdatascience.com

DATA-DRIVEN
SCIENCE

Elbow Curve to find "K" Clusters

The above code will generate this graph with all details



Source: towardsdatascience.com

By default, it uses Distortion Score as a metric that computes the sum of squared distances from each point to its assigned center.

DATA-DRIVEN SCIENCE

Get daily updates on
Data Science concepts

FOLLOW NOW

DATA-DRIVEN
SCIENCE