Measuring Clustering Performance with PyCaret

```
!pip install pycaret
```

- Silhouette: Silhouette Coefficient or silhouette score is a metric used to calculate the goodness of a clustering technique. Its value ranges from -1 to 1. If s=1: Means clusters are well apart from each other and clearly distinguished.
- The Calinski-Harabasz index also known as the Variance Ratio Criterion, is the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters. The higher the score, the better the performance. The score is higher when clusters are dense and well separated.
- Homogeneity: A clustering result satisfies homogeneity if all of its clusters contain only data points which are members of a single class. This metric is independent of the absolute values of the labels: a permutation of the class or cluster label values won't change the score value in any way.
- Rand Index: The Rand Index computes a similarity measure between two clusterings by considering all pairs of samples and counting pairs that are assigned in the same or different clusters in the predicted and true clusterings.
- Completeness: A clustering result satisfies completeness if all the data points that are members of a given class are elements of the same cluster. This metric is independent of the absolute values of the labels: a permutation of the class or cluster label values won't change the score value in any way.

```
# Importing dataset
from pycaret.datasets import get_data
jewellery = get_data('jewellery')
# Importing module and initializing setup
from pycaret.clustering import *
clu1 = setup(data = jewellery)
# check the model library to see all models
models()
# training kmeans model
kmeans = create_model('kmeans')
# training kmodes model
kmodes = create_model('kmodes')
```

```
   Silhouette  Calinski-Harabasz  Davies-Bouldin  Homogeneity  Rand Index  Completeness
0     -0.3819            20.2973          4.7314            0           0             0
```

✓ 13s completed at 12:00 PM