# Module 4 Assignment: K-Means Python Application

Adetomiwa Adedeji

Nexford University

September, 2024

**Executive Summary**

**Purpose of Application:**

The purpose of this application is to implement the K-Means clustering algorithm to group similar data points together within a selected dataset. K-Means is a popular unsupervised learning algorithm used for clustering that partitions data into K distinct, non-overlapping subsets.

**Dataset Chosen:**

The dataset used for this analysis was obtained from Registry of Open Data on AWS. It contains various features relevant to our clustering goals. The dataset was preprocessed by cleaning, normalizing, and selecting relevant features to ensure effective clustering.

**Results of Clustering:**

The K-Means algorithm was applied to the dataset with K=3 clusters. The clustering results were evaluated using the silhouette score, which measures how similar an object is to its own cluster compared to other clusters. The achieved silhouette score was 0.597, indicating acceptable quality of clustering.

**Insights and Conclusions:**

The clustering results revealed distinct groups within the data, which can provide valuable insights into the patterns and structures present. The clustering analysis helps in understanding the underlying data distribution and can guide further analysis or decision-making processes.

The K-Means clustering approach effectively grouped the data, and the results suggest that the chosen number of clusters is appropriate for the given dataset.