## Assignment 2

**Part 1: Real-World Applications of K-Means**

Task 1: Choose a Real-World Scenario

**Scenario:**

A marketing perspective of Customer Segmentation Marketing firms employ K-Means clustering to group customers based on similarities in purchasing behaviour, demographics and preferences for targeted marketing strategies.

K-Means groups customers into clusters by minimizing the distance between customer data points and their corresponding cluster centroids.

High value customers could be placed in one cluster while budget conscious buyers belong to another.

This technique is useful because it assists businesses in creating different customer segments for personalized marketing strategies and better resource management and increased customer **happiness.**

**Task 2: The Advantages of Applying K-Means**

**Simplifies Decision-Making:**

K-Means transforms complex, multi-dimensional data into interpretable clusters, making it easier to identify actionable insights; for example, to design marketing campaigns for particular customer segments.

**Efficiency and Scalability:**

K-Means is a fairly fast and effective algorithm that works well with large amounts of data which is important for large companies with large customer bases. Its simplicity and speed enable it to support rapid cycles and timely assessments during rapidly changing conditions.

**Part 2: Challenges and Alternatives**

**Task 1: The Problems of K-Means Clustering**

**Sensitivity to Initial Centroids:**

The algorithm's effectiveness is directly influenced by the random selection of the centroids. The choice of poor initializations can result in suboptimal cluster findings that require several initializations to produce improved results.

**Difficulty Handling Non-Spherical Clusters:**

K-Means assumes that data points are spherical in shape and are distributed homogeneously. The approach is not suitable for those datasets which contain overlapping or non-spherical clusters; for instance, in image segmentation, objects are not round in shape.

**Task 2: When Not to Use K-Means**

**Situation:**

K-Means is not suitable for datasets with overlapping or non-spherical clusters, such as in image segmentation where objects have irregular shapes.

**Alternative Algorithm:**

However, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is more effective for such data. DBSCAN also groups data based on the density of points, and it is therefore effective for non-spherical and arbitrary shaped clusters. Furthermore, it manages noise very well and does not need to know the number of clusters in the data set beforehand, K-Means cannot.