**Task 1: Data Preparation:**

First, we import the necessary libraries and load the dataset. We will drop the "Attack" column as we are not using it in unsupervised learning. We need to do the normalization of the data before the k means and the isolation forest algorithm, which ranges from 0 to 1.

**TASK-2: K-Means Clustering**

The datasets will be clustered using K-means with the value of k set to 2, which represents the two outcomes (Attack=0 and Attack=1). The fitted K-means clusters will be visually displayed using scatter plots, and the centroids of each cluster will be supplied for additional study.

We can observe that the data points were divided into two clusters by K-means clustering based on their similarity. Each cluster's centroids are represented by a black dot. The scatter plots, however, do not clearly distinguish between the Attack and No Attack situations.

Message data=0 when, speed value=0, speed parameter= 1, When the speed value=0 then Message data range= 0.35 to 1. When speed=0, Message=0,  
Message=1, Every dataset value is present in dataset 3. W.r.t speed=1 and message=0.When RPM is 0, then Message data=0,When RPM is 1, then Message value= 0.25 to 1 When RPM value is 1 the values are in the dataset 5 which ranges from 0.4 to 1,  
when speed=0 the values range from 0 to 1When the RPM is 0 for the dataset 6, then the data is present at 0 When RPM is 1 , the dataset will be present from 0 to 1.

**TASK-3 – Isolation Forest Algorithm**

Next, we will utilize the Isolation Forest Algorithm on the datasets to identify data points flagged as anomalies. These identified anomalies will be recorded, and scatter plots will be used to visually display the fitted Isolation Forest clusters. The implementation of the Isolation Forest Algorithm has revealed the presence of anomalies in the datasets, as evident from the yellow dots on the scatter plots. However, the scatter plots do not distinctly differentiate between the Attack and No Attack situations. Despite detecting anomalies, the scatter plots do notprovide clear separation between the different outcomes.

**TASK-4: Discussion**

Unlike supervised learning, K-means clustering is a useful technique for dealing with unlabeled data since I learnt how to utilize it to divide data into discrete groups. This enables effective data analysis and categorization based on similarities among dataset items. The limitations of utilizing unsupervised machine learning in this context include the challenges in determining the actual clusters or anomalies without labeled data and the uncertainty about their relationship with attacks. To enhance the performance of the models, a combination of unsupervised and supervised learning could be considered, where unsupervised techniques identify potential clusters or anomalies, and supervised learning classifies them as Attack or No Attack. Feature engineering can also be employed to extract more meaningful features from the data. Collecting labeled attack data and employing advanced anomaly detection techniques such as neural networks or deep learning may further improve the accuracy of the models. In conclusion, incorporating supervised learning, feature engineering, and collecting labeled data can potentially overcome the limitations of unsupervised learning in accurately predicting attacks on the CAN bus of a vehicle.