**Quiz 8 - K Means**

Consider using the dataset “hdbresale\_cluster.csv” to create distinct groups from the various flats within the data.

We will take the following naming convention for the rest of this quiz:

hdb=read.csv("hdbresale\_cluster.csv")

|  |  |
| --- | --- |
| **Question 1**  Which of the following is/are FALSE. Select all options that apply. | |
| A | K-Means works well if we want to cluster the flats from this dataset based on the flat\_type. |
| B | The number of clusters (or groups) for all the flats in the data set given could be chosen differently from user to user. |
| C | K-Means is a supervised learning algorithm. |
| D | We do NOT need a response variable in the dataset to perform the K-Means algorithm. |

|  |  |
| --- | --- |
| **Question 2**  Which line of code below is the correct way to find the Number of Points (or flats) in each cluster?  Given:  kout <- kmeans(hdb[ , c("floor\_area\_sqm" , "amenities")] , centers=2) | |
| A | kout$withinss |
| B | kout$cluster |
| C | kout$centers |
| D | kout$size |

|  |  |
| --- | --- |
| **Question 3**  We can use the K-means algorithm for categorical variables as long as we declare them as factors. True or False? | |
| A | True |
| B | False |

|  |  |
| --- | --- |
| **Question 4**  Which of the following is a suitable way to check for the Goodness-of-Fit of the K-Means algorithm? | |
| A | Accuracy |
| B | Precision |
| C | Type 2 Error Rate |
| D | None of the above |

|  |  |
| --- | --- |
| **Question 5**  We can also use the KNN algorithm to cluster or group the flats based on their floor\_area\_sqm and amenities instead of using the K-means algorithm. True or False? | |
| A | True |
| B | False |