**Documentation after applying EDA :**

After extracting the columns ‘A’,‘B’,’C’,’G’,’R’,’X6’ and ’X8’ for different classes EDA is done.

No missed values in any class.

Class A:

* Using Pair plot 🡪 **It is clearly observed that the column 'R' has misclassification data.** ( i.e ,wide spread of data) .
* Columns ‘B’,’C’,’X6’ and ’X8’ have no outliers.
* Columns ‘A’,’G’ and ’R’ have so many outliers.
* Using Heatmap 🡪 It is clearly observed that the column 'B' and 'C' having 0.82 correlation. We should drop any one of those variables.(highly correlated).

Class B:

* + Using Pair plot 🡪 **It is clearly observed that the column 'R' has misclassification data.**
  + Columns ‘B’,’C’,’X6’ and ’X8’ have no outliers.
  + Columns ‘A’,’G’ and ’R’ have so many outliers.
  + Using Heatmap 🡪 It is clearly observed that the column 'B' and 'C' having 0.82 correlation. We should drop any one of those variables. ( highly correlated).

Class C:

* + Using Pair plot 🡪 **It is clearly observed that the column 'R' has misclassification data.**
  + Column ‘B’ have no outliers.
  + Columns ‘A’,’C’,’G’,’R’,’X6’ and ‘X8’ have so many outliers.
  + Using Heatmap 🡪 It is clearly observed that the column 'B' and 'C' having 0.89 correlation. We should drop any one of those variables. ( highly correlated).

Class D:

* + Using Pair plot 🡪 **It is clearly observed that the column 'R' has misclassification data.**
  + Columns ‘A’,‘B’,’G’,’X6’ and ‘X8’ have no outliers.
  + Columns ’C’ and ’R’ have so many outliers.
  + Using Heatmap 🡪 It is clearly observed that the column 'B' and 'C' having 0.86 correlation. We should drop any one of those variables.(highly correlated).

Class E:

* + Using Pair plot 🡪 **It is clearly observed that the column 'R' has misclassification data.**
  + Columns ‘A’,’B’,’C’,’G’,’X6’ and ‘X8’ have no outliers.
  + Column ‘R’ have so many outliers.
  + Using Heatmap 🡪 It is clearly observed that the column 'B' and 'C' having 0.89 correlation. We should drop any one of those variables.(highly correlated).

Class F:

* + Using Pair plot 🡪 **It is clearly observed that the column 'R' has misclassification data.**
  + Column ’B’,’C’,’X6’ and ‘X8’ have no outliers.
  + Columns ‘A’,’G’ and ‘R’ have so many outliers.
  + Using Heatmap 🡪 It is clearly observed that the column 'B' and 'C' having 0.89 correlation. We should drop any one of those variables.(highly correlated).

From all the classes it is clearly observed that column ‘R’ has so many outliers of data and the columns ‘B’ and ‘C’ are highly correlated.

Compared with all the classes Class C is misbehaving because in Class C except column ‘B’ all the columns having outliers.