Hi Team,

We are working on a project where we need to aggregate all the columns based on user defined subsets.

Below process needs to be followed for completing this task,

1. Combine all the data files into a single file
2. Take user defined input for number of subsets
3. Sort the CAGR column in descending order
4. Divide the entire data set into equal sized buckets/subsets based on input
5. Aggregate(mean) all the columns based on this subset
6. Create a formatted output file through code having first tab as complete data and second tab as bucketed output with all the columns aggregated
   * Bucketed data center aligned
   * All the values should be in percentages (%)
   * Header Specific formatting
     + Font size = 10
     + Font color = White
     + Font name = Calibri
     + Font style = Bold
     + Background color = “#000099”

Performing this task manually is not a scalable approach and would take sufficient amount of time so we would need Data Analytics teams support for performing this task.

Let us know if it will be possible to build a code that can perform this task and estimated the time for development.