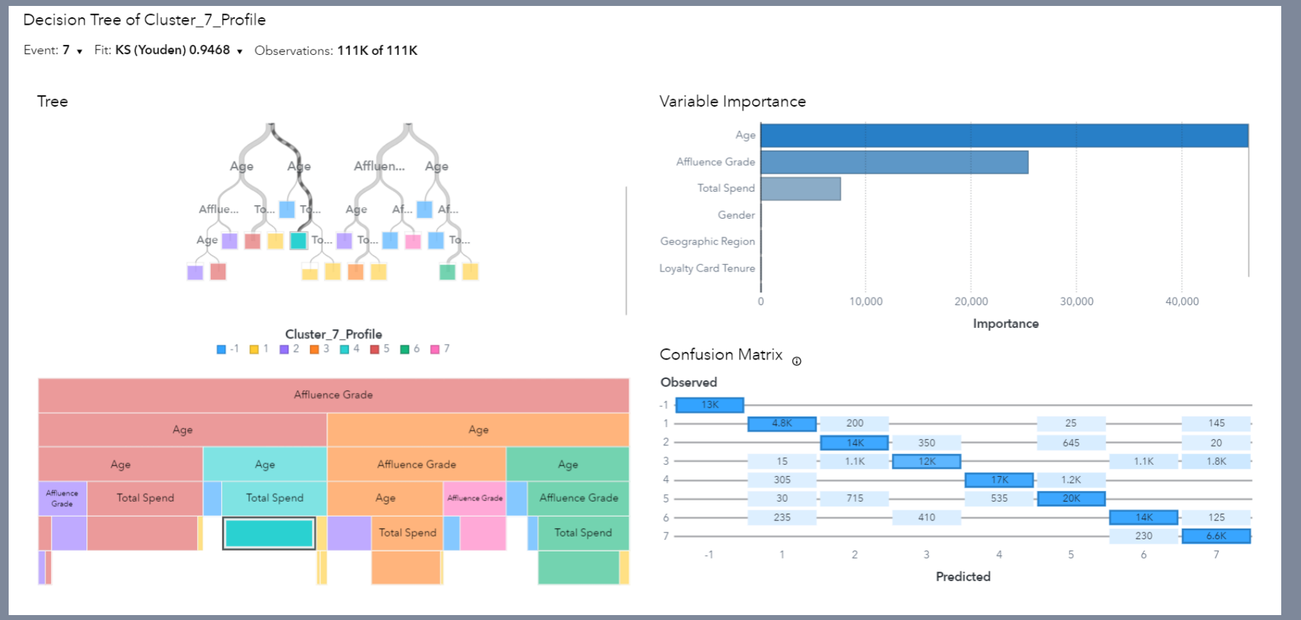
Non-Predictive Use of Decision Tree and Friends

How supervised machine learning models can help you beyond the usual task of prediction and classification

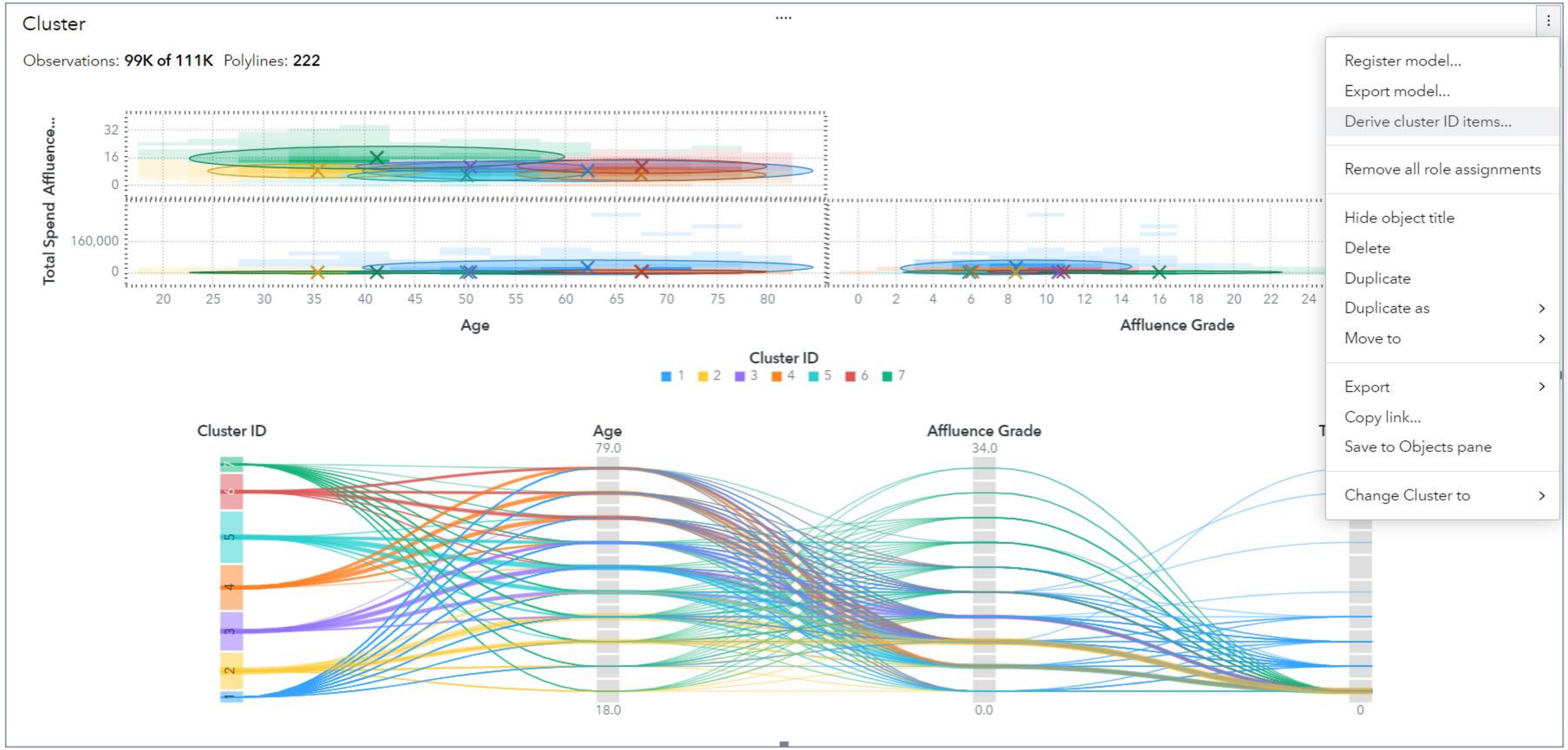
Gerhard Svolba, Data Scientist, SAS Austria

Data Scientist @SAS - Medium | LinkedIn | Github | SAS-Books | SAS Articles
Youtube: DataPreparation4DataScience | Data Science Use Cases

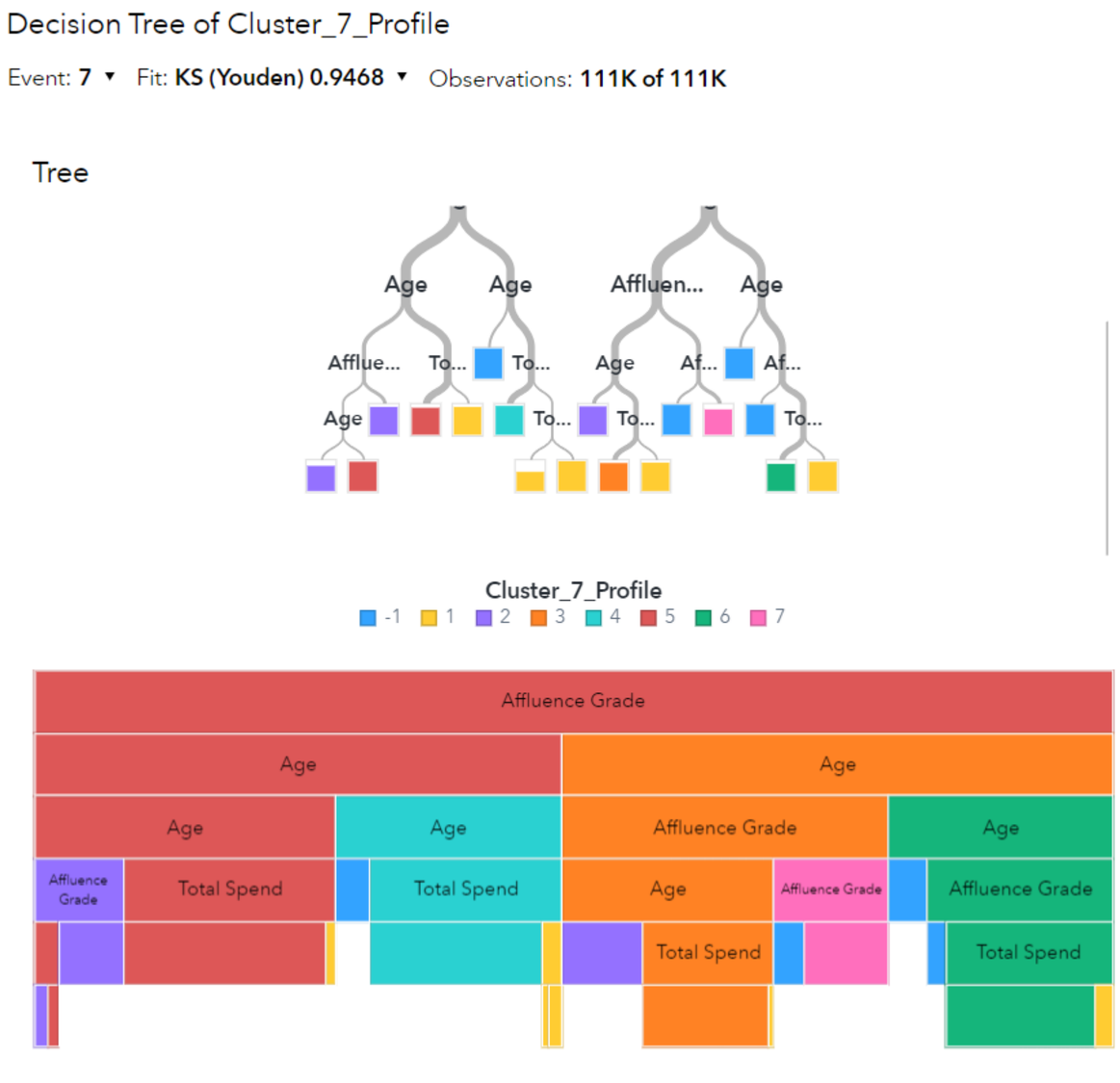
#4
Profiling the nature of clusters and segments

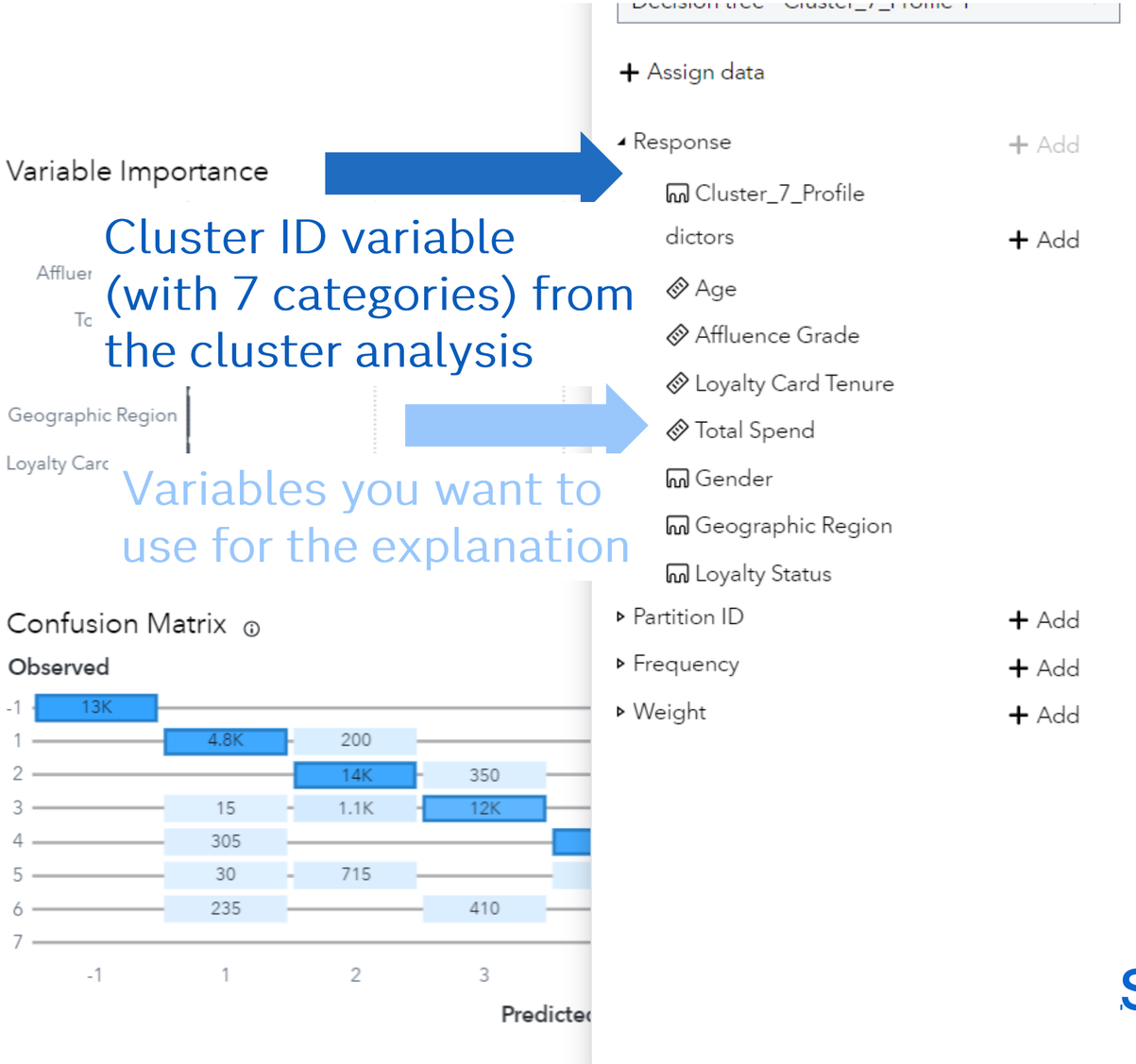


Step 1: Derive the ClusterIDs from your cluster model

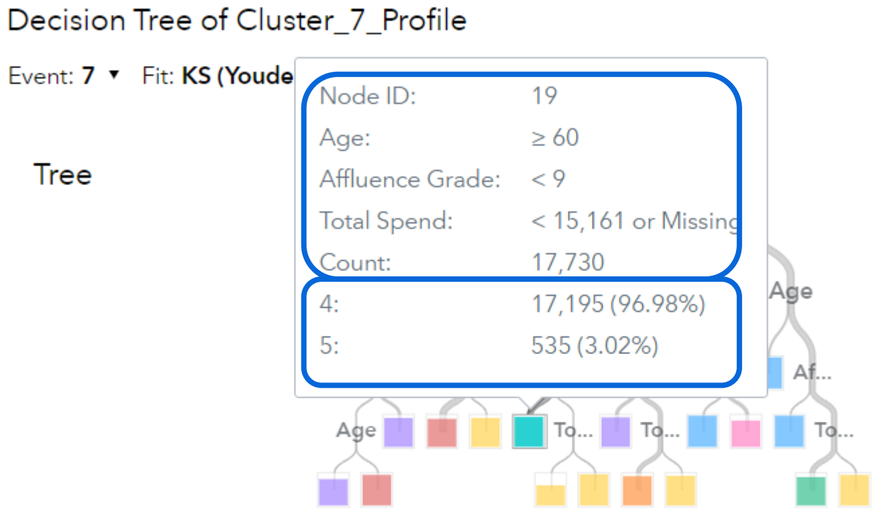


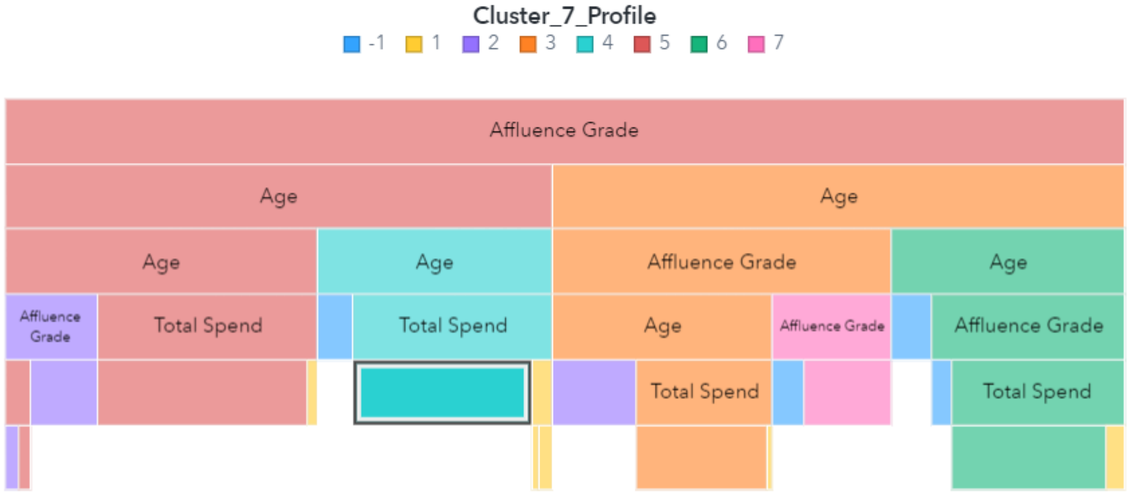
Step 2: Build a decision tree to “explain” why the differences between the clusters





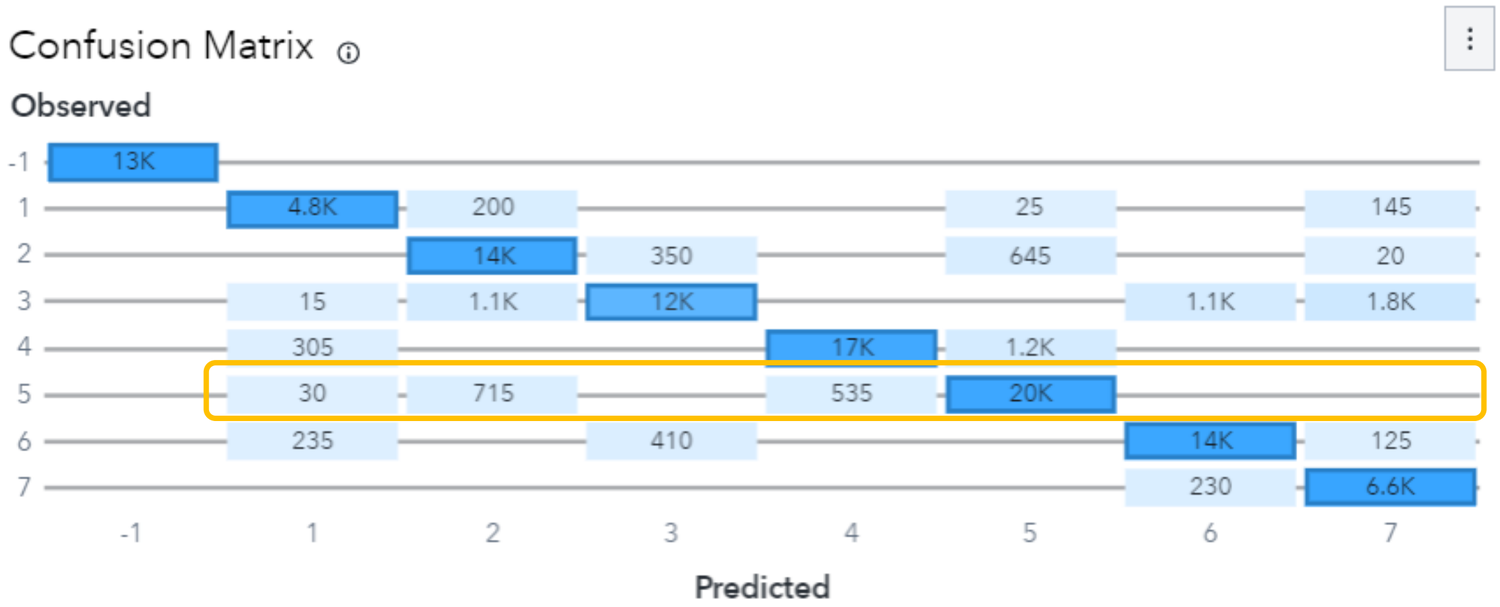
Step 3: Review individual leaves of the tree for the rules and the “predicted” clusters



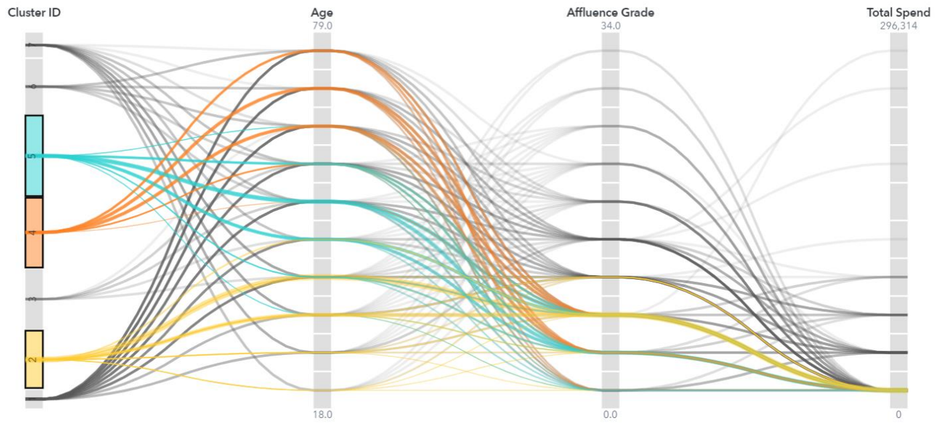


- Node 19 represents
  - 17,730 observations who are
  - older 60 years,
  - have a total spent of < 15,000 and
  - are less affluent
- 97% are found in cluster 4
- 3% are found in cluster 5

Step 4: Review the confusion matrix to understand which clusters are “similar” to each other as some of the analysis subjects are assigned to the “wrong” (= “nearby”) cluster



- Analysis subjects from cluster 5 are
  - Mostly assigned to cluster 5
  - Some them however to cluster 2 and 4



Application Recommendations

- Preferred Method: Decision Tree
- Recommended SAS Tool: SAS Visual Analytics

- This is not limited to cluster models built in SAS Visual Analytics!
- You can use the cluster/segment assignments from any model
  - SAS Model Studio
  - Cluster Models from SAS Procedures
  - Segmentations in general