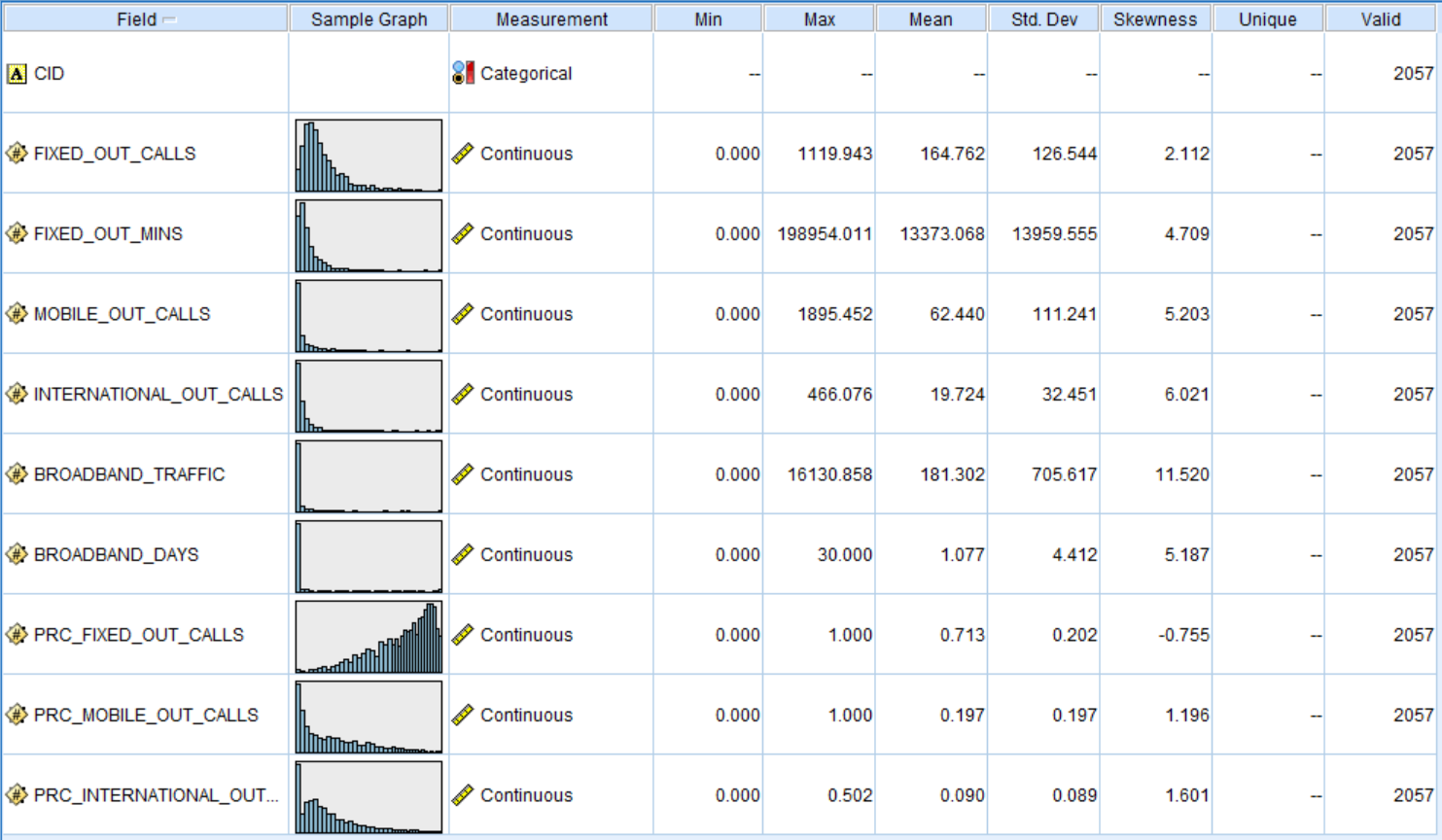
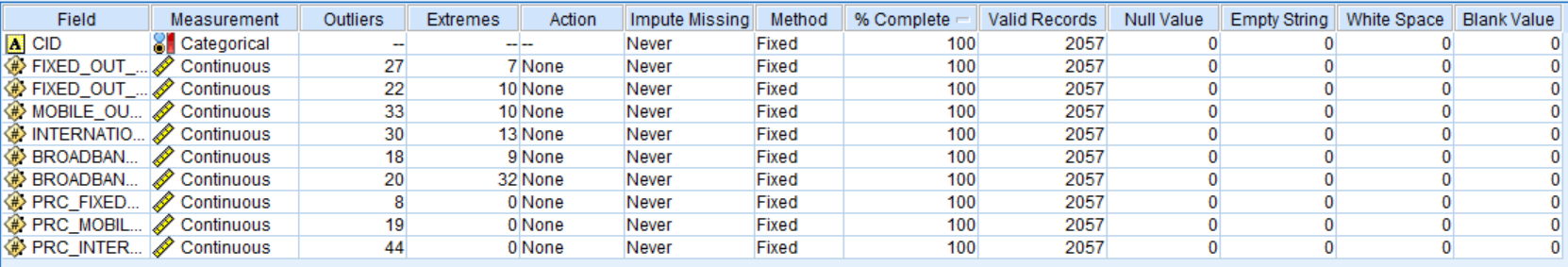
Behavioral Segmentation Model Summary

Behavioral segmentation modeling begins with the CRISP-DM methodology and the establishment of a solid business understanding. It is in this process that the type of segmentation and the related supporting data is identified. For this project, the focus was developing a customer behavioral segmentation based on out-call activity for market development. The initial data source was supplied.

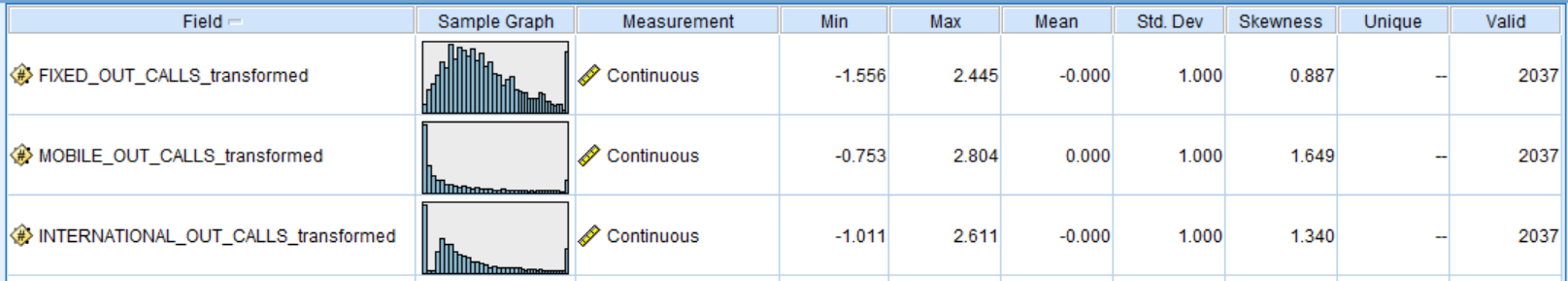
The core data analysis started with a focus on understanding the source data file. Initial inspection of the data, see below, showed that there were data fields that could be removed from the analysis and the out-call fields showed a significant difference in scale. In addition, extreme outliers were detected in all three out-call fields.





These challenges were addressed through several data preparation steps. The first major step was to run the fields through an Anomaly node to detect the most extreme records. Extreme outliers can have a negative impact on clustering models, so they should be reviewed and handled appropriately. (Chorianopoulos, 2016, pp. 118 - 119) In this case, twenty records were detected and removed.

The next major data preparation step was to normalize the data scale. This was done using SPSS Modeler’s Auto Data Prep node. Specifically, the three outcall fields were rescaled using a Z-transformation algorithm. The chart below shows the data field’s new scale through the descriptive statistics.



The final data preparation step concluded with using a PCA node to reduce the number of inputs into the modeling stage. An additional advantage of using solutions like PCA is that it ensures that the inputs will contribute equally to the segmentation models. (Chorianopoulos, 2016, p. 124) This balancing can be see with the resulting PCA factor equation.

0.491 \* FIXED\_OUT\_CALLS\_transformed +

0.4679 \* MOBILE\_OUT\_CALLS\_transformed +

0.4707 \* INTERNATIONAL\_OUT\_CALLS\_transformed +

+ 0.000000000000002523

The modeling phase involved the development of several models. Three finalist models were built that had Silhouette coefficients scores at 0.7. The chosen model was a hand tuned K-Means model with 4 four reasonably sized clusters and good grouping around the centroids. Cluster 2 showed the greatest variation from the centroid with a mean value of 2.5. However, this cluster grouping showed similar challenges with another of the finalist models.

Chosen Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | |  | |
|  | | | Cluster Graph | |
| Cluster 1 | Cluster 2 | Cluster 3 | | Cluster 4 |

In the deployment stage, the model was converted from an unsupervised K-Means model to a decision tree. This conversion to a decision tree provided transparency of the customer scoring by showing the decision rules, shown below. The results showed that the tree has 100% accuracy and perfect gains and lift. This would normally generate a concern of a perfect predictor. In this situation, it is acceptable because the K-Means model is being converted.

|  |  |
| --- | --- |
| C5 Model | Rules |
| Gain | Lift |
|  |  |

This project summary has shown how the project has flowed through the CRISP-DM process, starting with a business goal through to the deployment of a customer scoring decision tree. While the process was presented in a linear fashion, in reality previous steps were revisited to improve the models. Additional ongoing performance monitoring and refresh step will need to be taken as the model executes in a production environment. These final steps are necessary to ensure ongoing accurate results. (Chorianopoulos, 2016, p. 140)

# References

Chorianopoulos, A. (2016). *Effective CRM Using Predictive Analytics.* West Sussex, United Kingdom: John Wiley & Sons, Ltd.