* Chart, scatter chart, bubble chart

  Description automatically generatedCluster Analysis Using K-Means technique.
* PCA and t-sne used to reduce the number of dimensions, while maintaining 99% of the explained variance.
* Chose 2 clusters for purposes of visualization. Graduation, outcome variable of interest, is binary.

Shape

Description automatically generated

* KNN clusters added to original participant dataframe to anlayze how clusters relate to specific variables of interest.
* KNN clusters vs graduation status of participant.
* Participants from cluster one (yellow) were more likely to be program graduates than participants from cluster 0 (purple).

A picture containing shape

Description automatically generated

* KNN clusters vs. Employment at Entry.
* Cluster one (yellow) most likely to be employed full time at entry.
* Cluster two (purple) most likely to be unemployed at entry.

Table

Description automatically generated

* Clusters differ on basis of incoming age, income, employment, and risk profile.
* Clusters also differ on treatment, drug test, and number of hearings.
* From this data, we can surmise that treatment provided in program in combination with the profile of incoming participant are both important drivers for participant outcomes.