Because VALP (property value) and HINCP (Household Income) use the same unit and almost the same scare, I will not use the standard procedure to standardize the variables. Next, I run FASTCLUS with 15 clusters, 20 clusters and 25 clusters.

State top 3 differences between the k=20 and k=15 runs.

1. The cluster numbers are different.
2. The initial seeds are different and the cluster means are different. Then, the distribution of clusters is different.
3. The value ranges are different. For example,
   In 20 Clusters, the range of cluster means for VALP is 55,158.43 ~ 2,267,000.00.
   In 15 Clusters, the range of cluster means for VALP is 85,229.22 ~ 2,267,000.00.

State your next steps in the clustering analysis (not profiling steps). These will be steps that address what you will do next with the 15 clusters:

-Will you proceed with k=15 or k=20? Will you run for a different k value? State reasons why?

I will proceed with k=25. Actually, I have run three different k values, k=15, k=20, k=25. Because this cluster analysis is about tax strategy, I hope there are more clusters which have less than 100 members. Only small group is easier to be added tax. As we see, in the 25 clusters, there are 8 clusters which have less than 100 members. Another reason is I hope the difference between cluster means is small as possible. In 15 Clusters, the cluster means of VALP are from 85,229.22 to 2,267,000 and the cluster means of HINCP are from 16,387.65 to 1,365,333. In 20 Clusters, the cluster means of VALP are from 55,158.43 to 2,267,000 and the cluster means of HINCP are from 16,365.63 to 1,365,333. In 25 Clusters, the cluster means of VALP are from 68,225.25 to 2,267,000 and the cluster means of HINCP are from 21,305.23 to 1,425,000. Comparing them, I will choose k=25.

-Will you perform sub-clustering on the clustered results? State reasons why? If yes, which clusters will you additionally segment? Which clusters will you keep as is?

Yes, I will perform sub-clustering on the clustered results because the frequency of some clusters is bigger than 10,000. In this kind of dense clusters, I will use sub-clustering. These clusters include cluster 14 which has 14,130 members and cluster 22 which has 15,997 members. I will keep the rest clusters as is.

What variables will you consider in profiling the clusters? Just list the variables chosen.

These variables are,
BLD (Type of building),
MRGP (First mortgage payment (monthly amount)),
SMP (Total payment on all second and junior mortgages and home equity loans (monthly amount)),
TEN (Tenure),
VALP (Property value),
YBL (When structure first built),
HINCP (Household income (past 12 months)),

TAXP (Property taxes (yearly amount)),
WIF (Workers in family during the past 12 months),
MV (When moved into this house or apartment),
WORKSTAT (Work status of householder or spouse in family households)