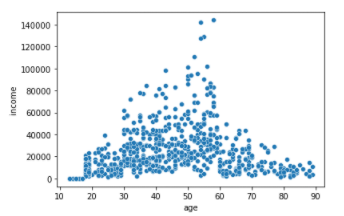


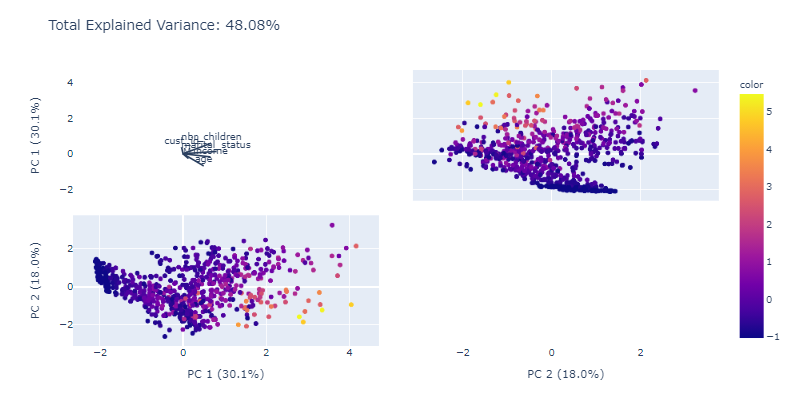
Not a lot of correlation btw variables and income; little bit with age, little bit with # children

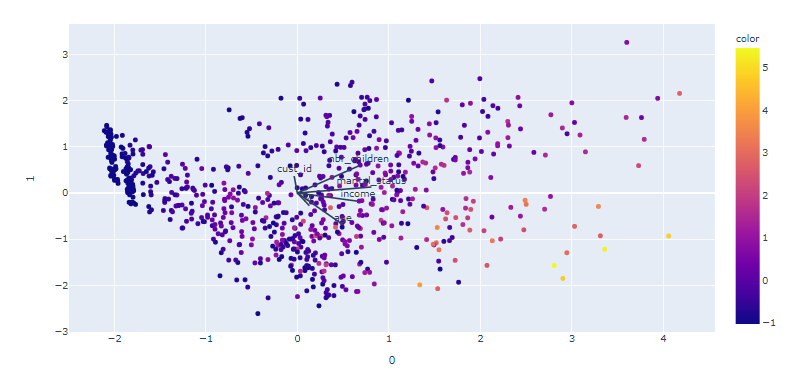


Kind of parabolic upsidedown

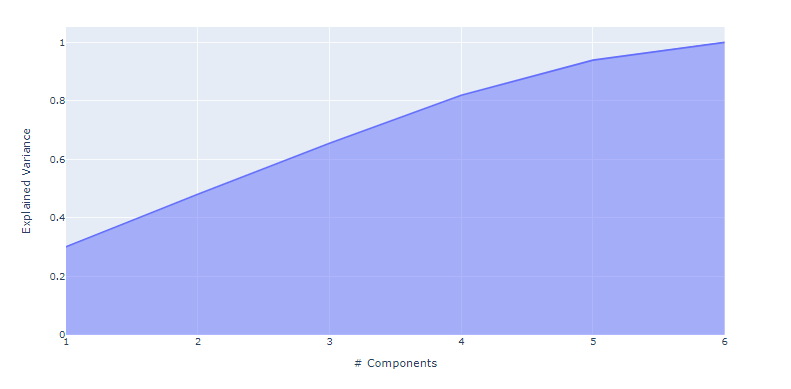
With scaled data

PCA

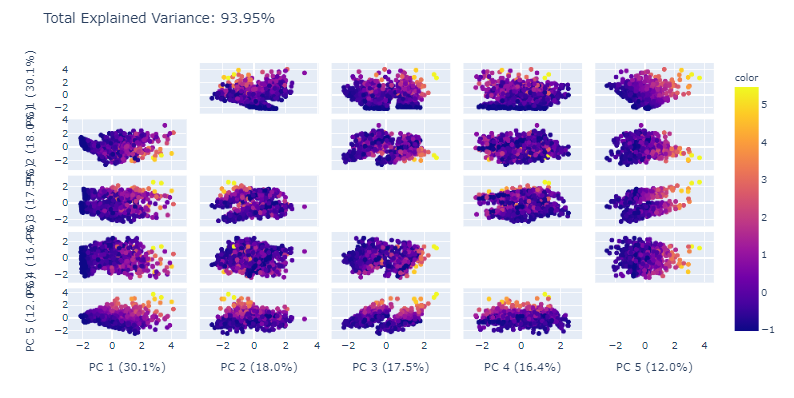




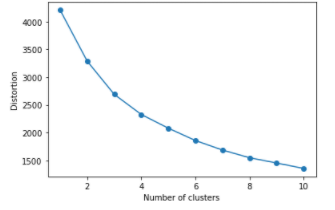
Even though PCA probably not required to explain the data

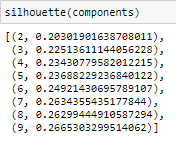


Fit PCA @ 5



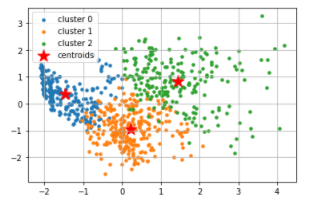
* Kmeans

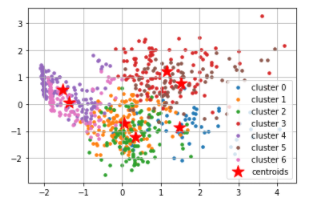




**silhouette** value is a measure of how similar an object is to its own cluster, measures are quite low

Could’ve been 3 or 7

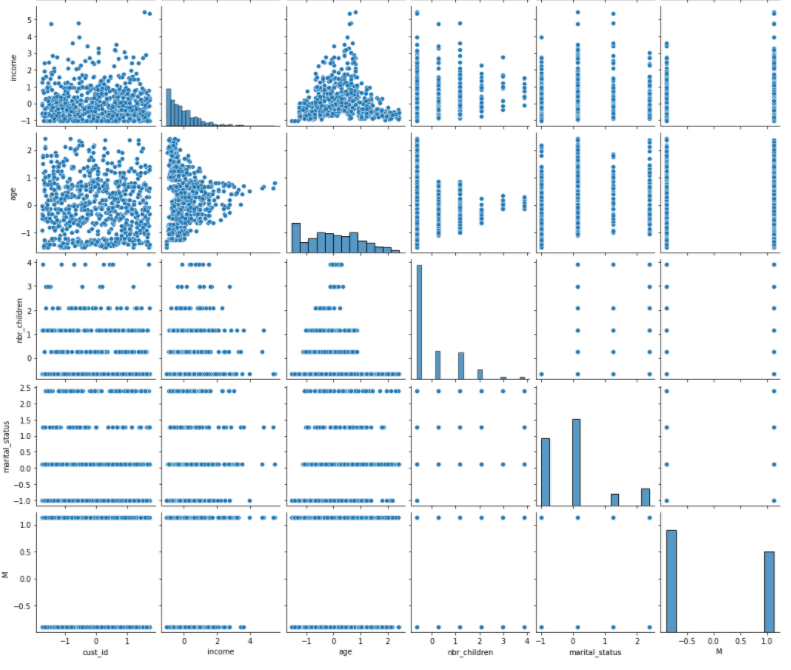




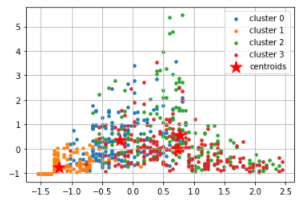
3 looks cleaner/more organized; 7 not so much

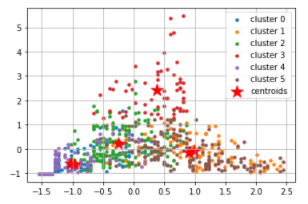
Not a lot of interpretability from clustering PCA’d values

Non-PCA’d; just scaled data



(y = income, x = age); k could have been 4 or 6

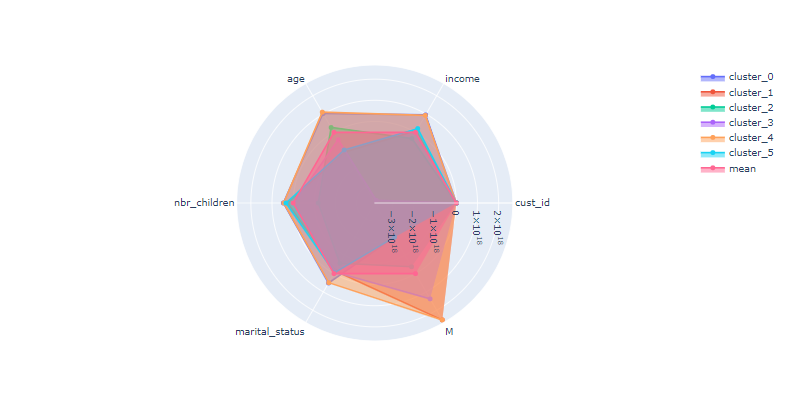




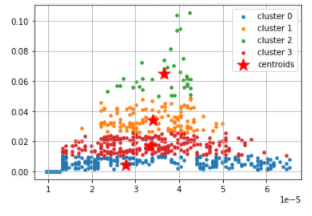
Hard to say how it’s clustered. Seems like it’s clustered by age group, but has a lot of variance/same groupings

* Green is older, has higher income
* Blue seems middle aged, medium
* Yellow is young, lowest grouped income

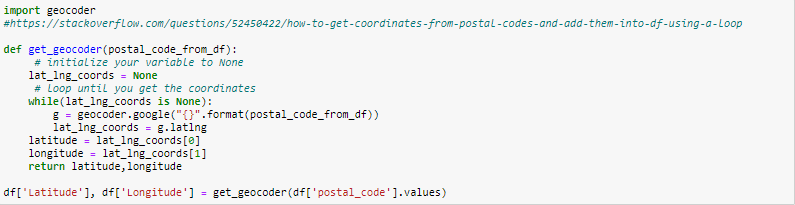
(youngest = 13, oldest = 89; std = 19)



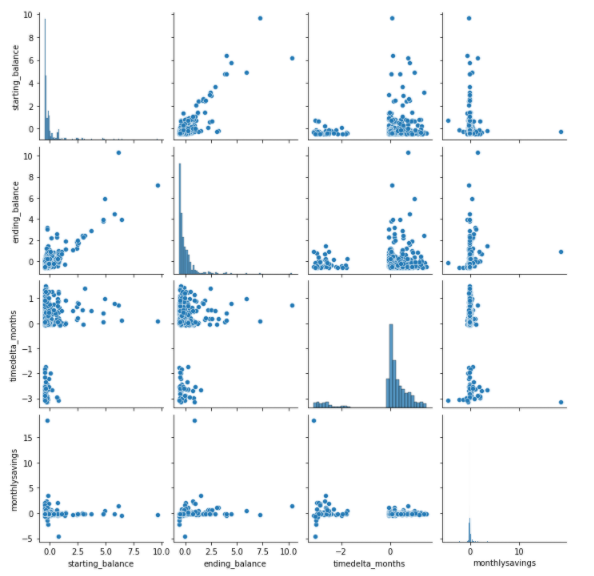
Normalized data

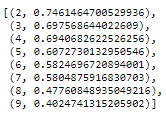
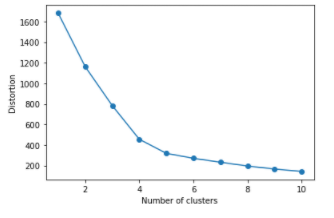


Not enough time:

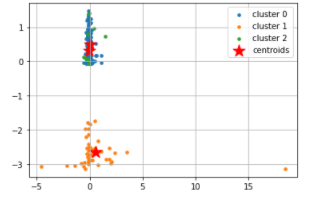


Worked with savings data, engineered age of account in months (timedelta\_months)and avg monthly savings (monthlysavings); data is scaled





Y = Age of acc; x = avg monthly savings



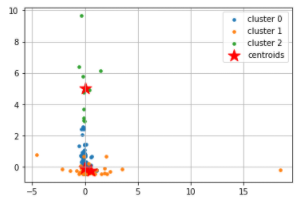
Thought the longer someone had an account, the more savings someone would have; not the case. Clusters are more based on how long someone had an account for; most savings around 0.

Shorter lived accounts had more variation in data but that’s because avg is affect by time

Yellow = short aged accs

Green + blue = long aged accounts (goes up to 30 years)

Y = starting balance, x = avg monthly savings



Thought people with lower starting balance would be incentivized to save more or something along those lines; not true. Again, just clustered around who had more starting balance.

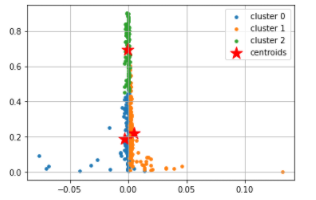
Yellow + blue = low starting balance

Green = high starting balance

Seems like no one is using their savings accounts

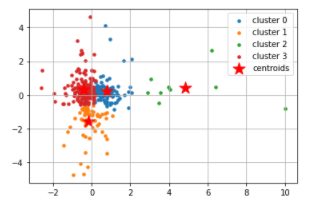
Weird finding: With norm’d data, clusters differently

Y = age of acc, x = avg monthly savings



Engineered total savings vs total debt; basically end savings – start savings/ end debt – start debt

Y = debt gained, x = savings gained



Most people hovered at 0, a lot more variance in debt. People who actually saved seems to be outliers

Clusters @

* positive debt, negative savings (red)
* positive savings positive/neutral debt (kinda) (blue)
* neutral savings/negative debt (yellow)
* outliers SAVERS (green)