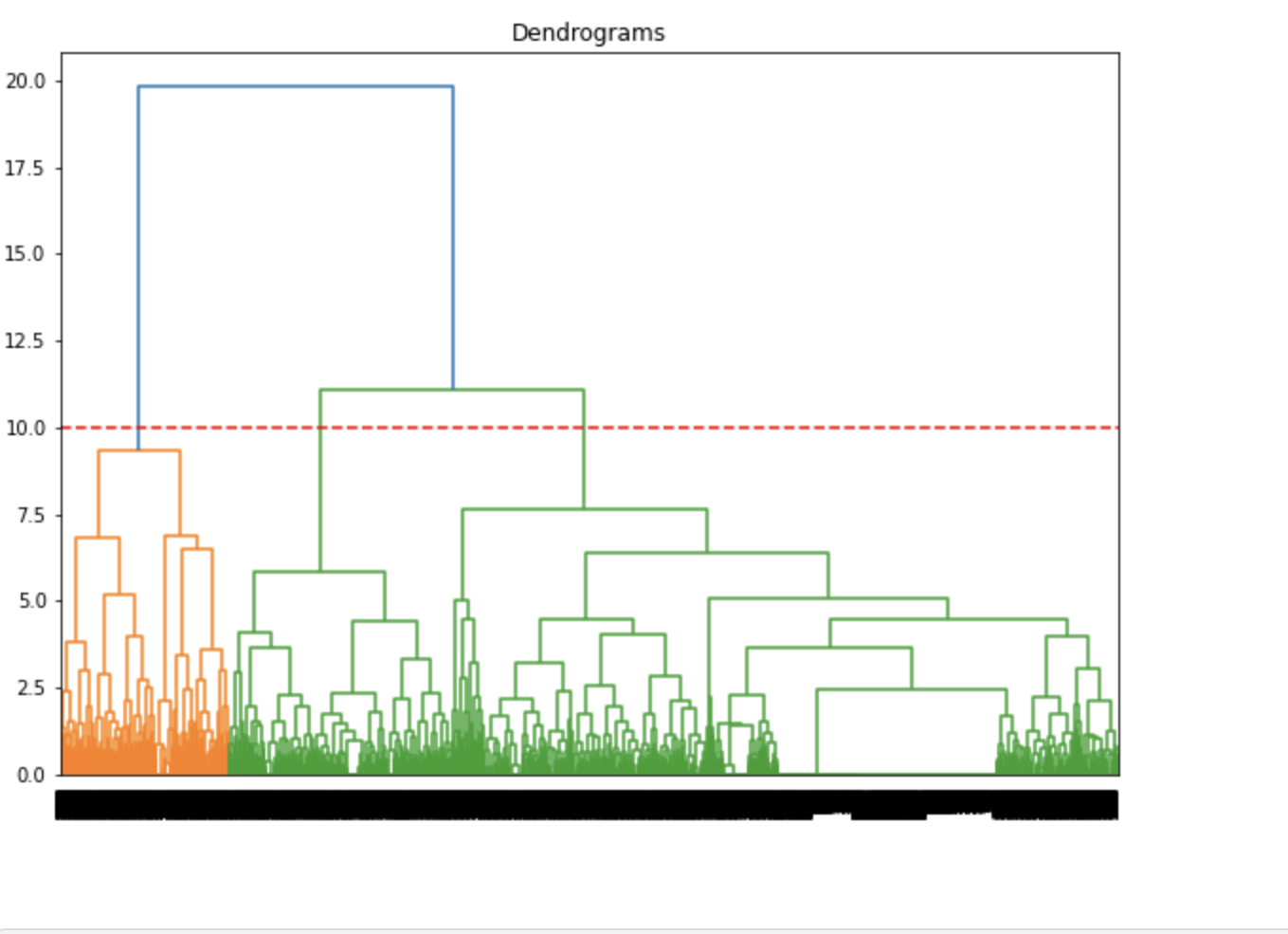
# Our Approach

We have tried clustering on the dataset.

1. Hierarchical clustering to understand the number of clusters.

From the clustering output we could see that k=3 is a good number to begin with



2.) We performed K means clustering with k=3 and below are the results

The segments could be translated as below from the average scores of the 7 variables

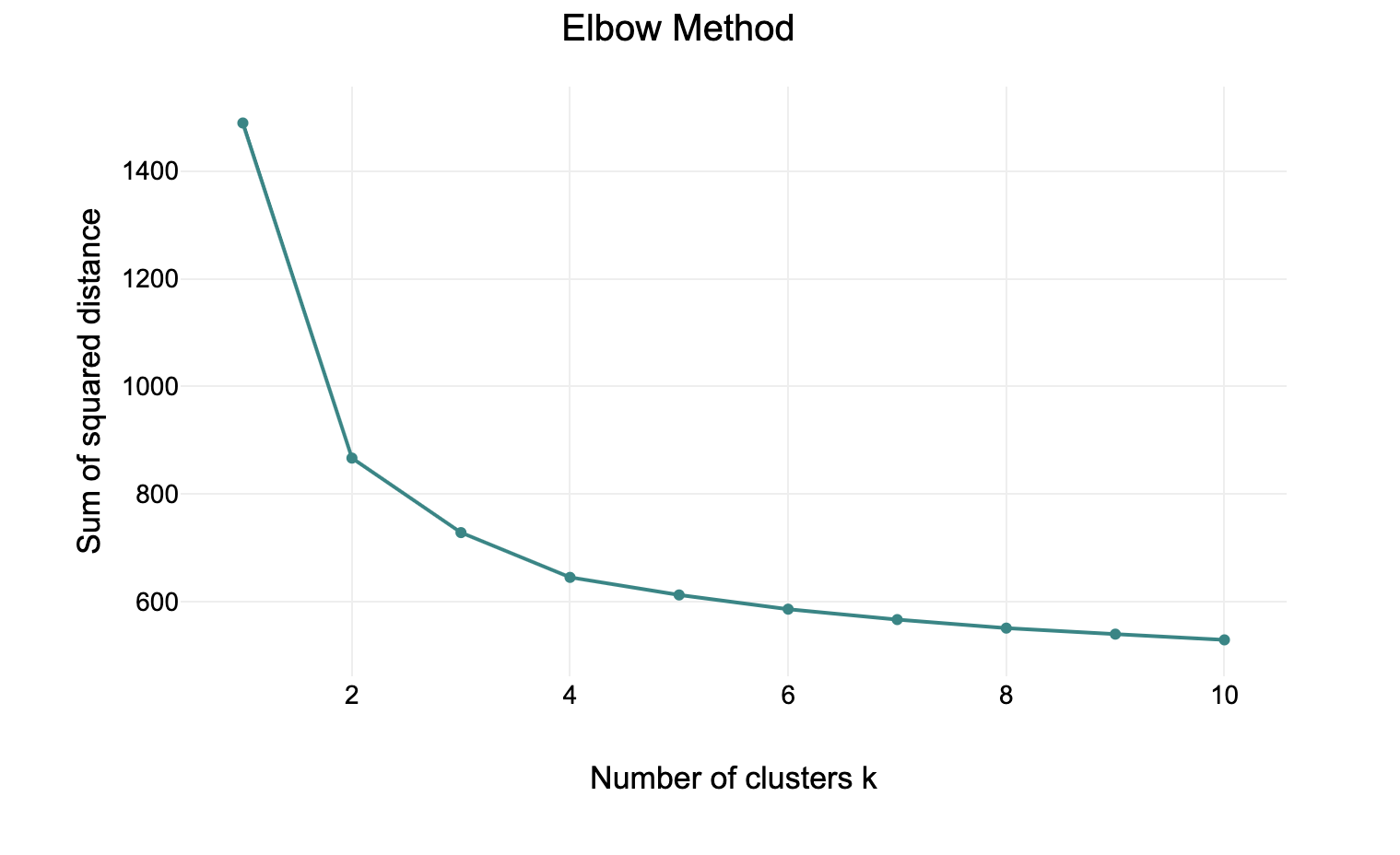
cluster 1- burned out

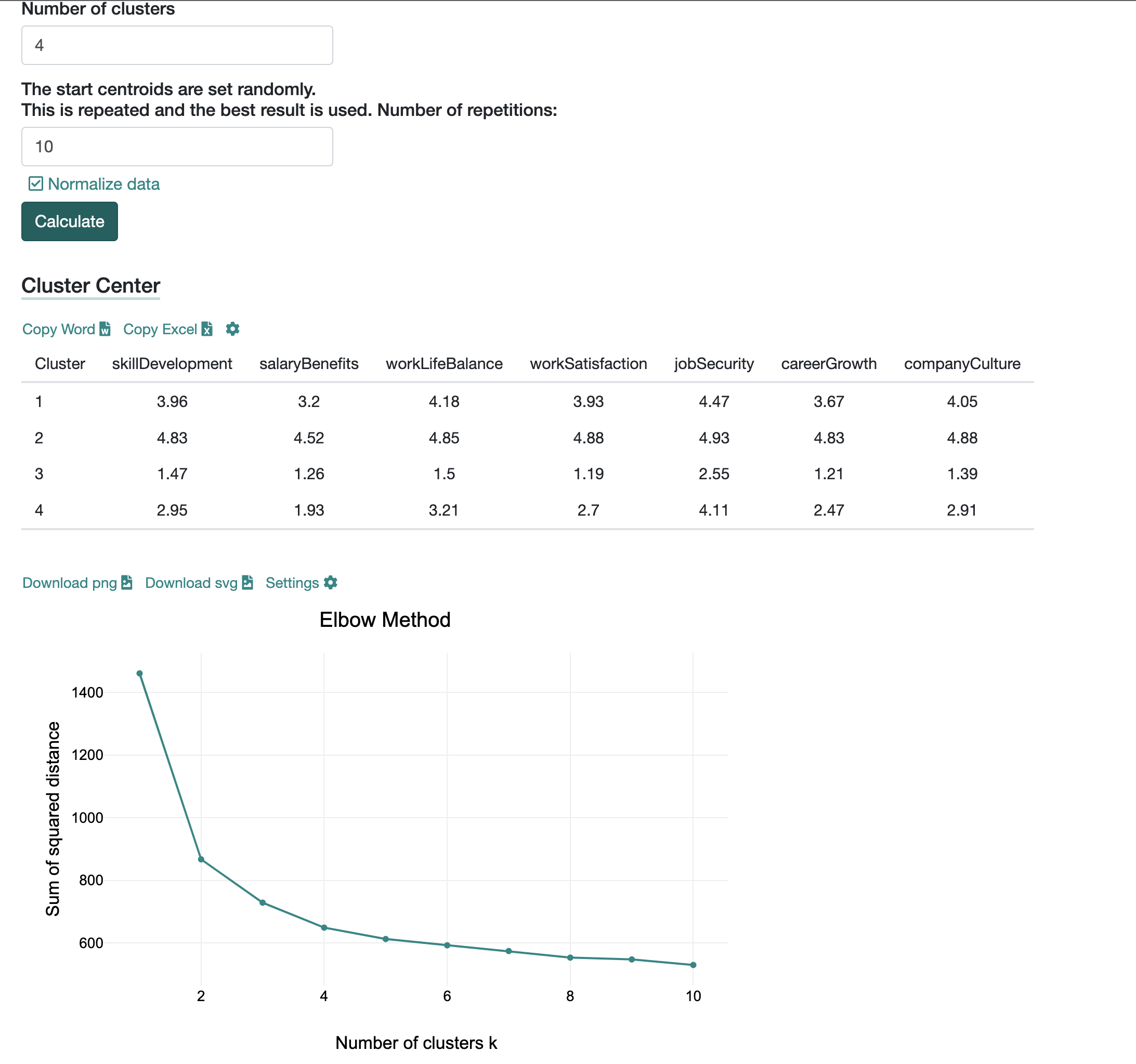
cluster 2-burned out

cluster 3-prone to be burnt out

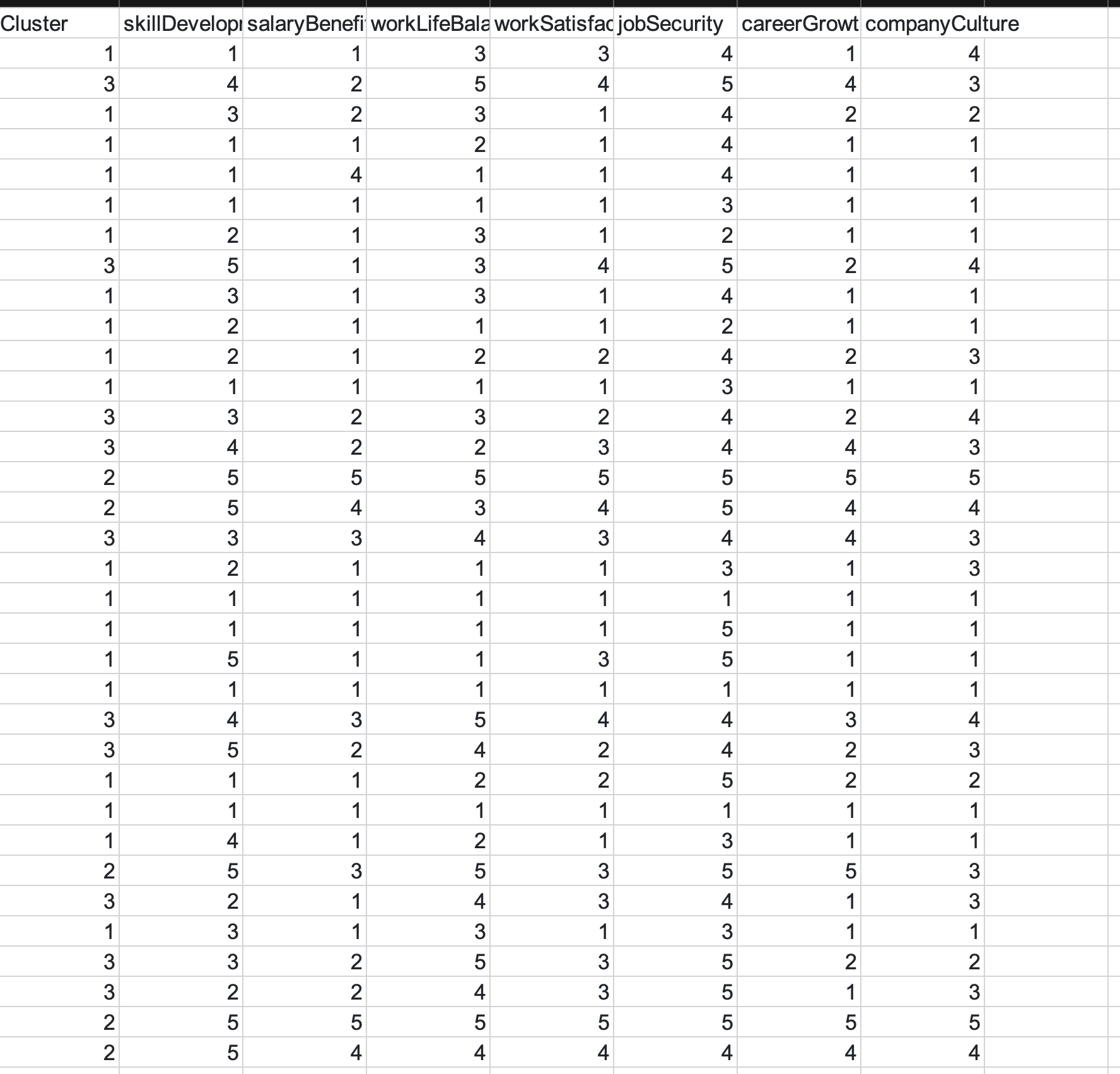


From the elbow plot we are also tempted to use k=4??? (doubt)





The cluster allocation



# Factor Analysis Output

## Exploratory Factor Analysis

* With the Job security variable, the below latent constructs

A picture containing table

Description automatically generatedDiagram

Description automatically generated

* Without Job security variable below are the latent constructs

A picture containing calendar

Description automatically generated

Diagram

Description automatically generated

## Confirmatory Factor Analysis

Diagram

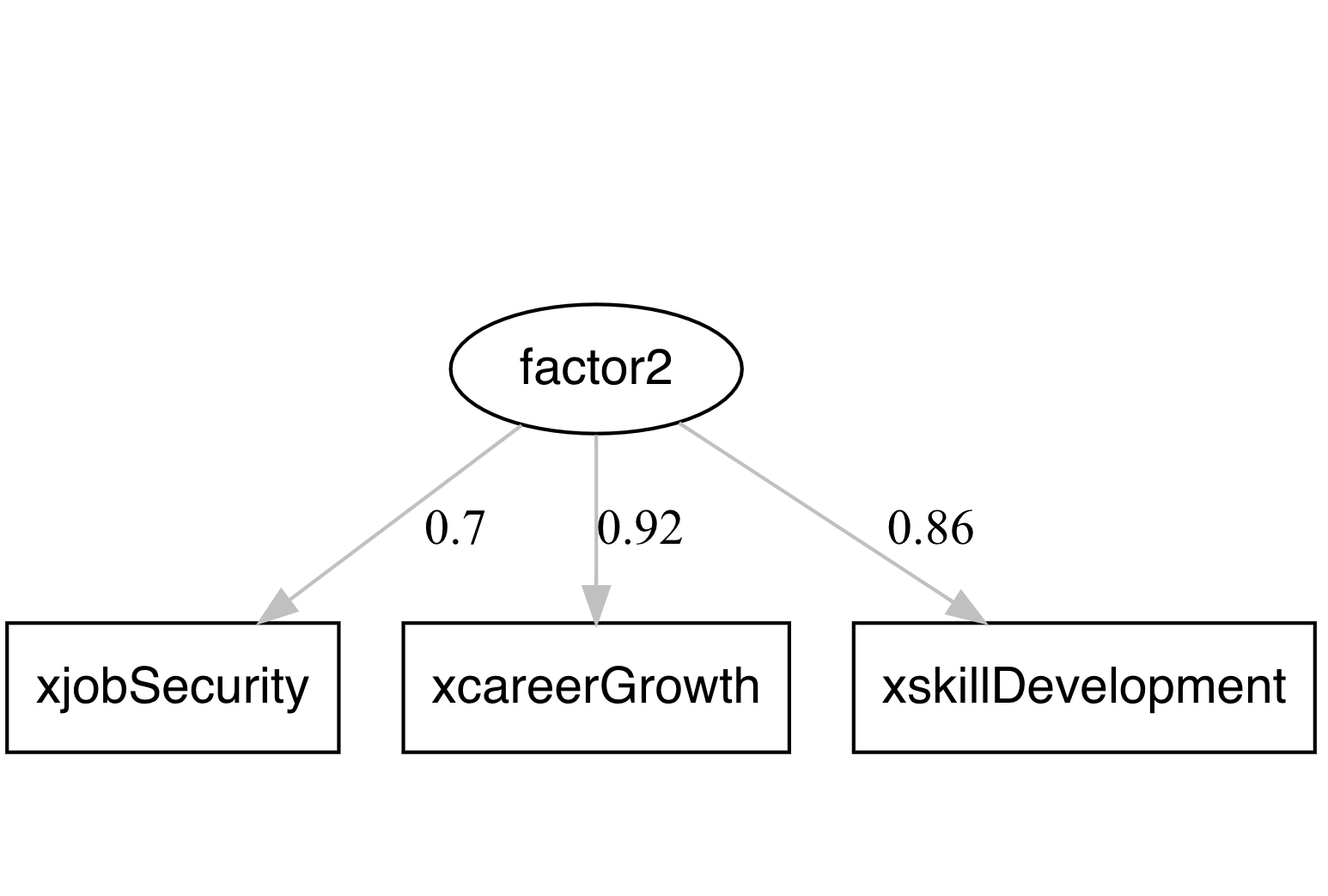
Description automatically generated

Diagram

Description automatically generatedText

Description automatically generated

Diagram

Description automatically generated 

From the above Factor Analysis output we could see that there are three latent constructs (factors) from the 7 variables.

Naming those 3 Factors requires domain expertise.From the research we did from our end we thought these factors could be any of the 3 axis (De personalization,Exhaustion,Lack of achievement).Or should we leave them as is.Is this the right approach to follow???.

Table

Description automatically generated

# FUTURE STEPS

* Given our approach of clustering into 3 segments and the factor analysis .we would like to proceed using the clustered category text for further analysis.

(Ex.Analysing cluster 1 text and topic modelling of cluster one helps us in understanding the sentiment of the people who have given very less ratings to the company )

* We would like to translate this to a recommendation system or a Multiclass classification approach (As we have data that is labelled into clusters now)
* Converting the text to a sentiment score and validating our approach with text and without text scores