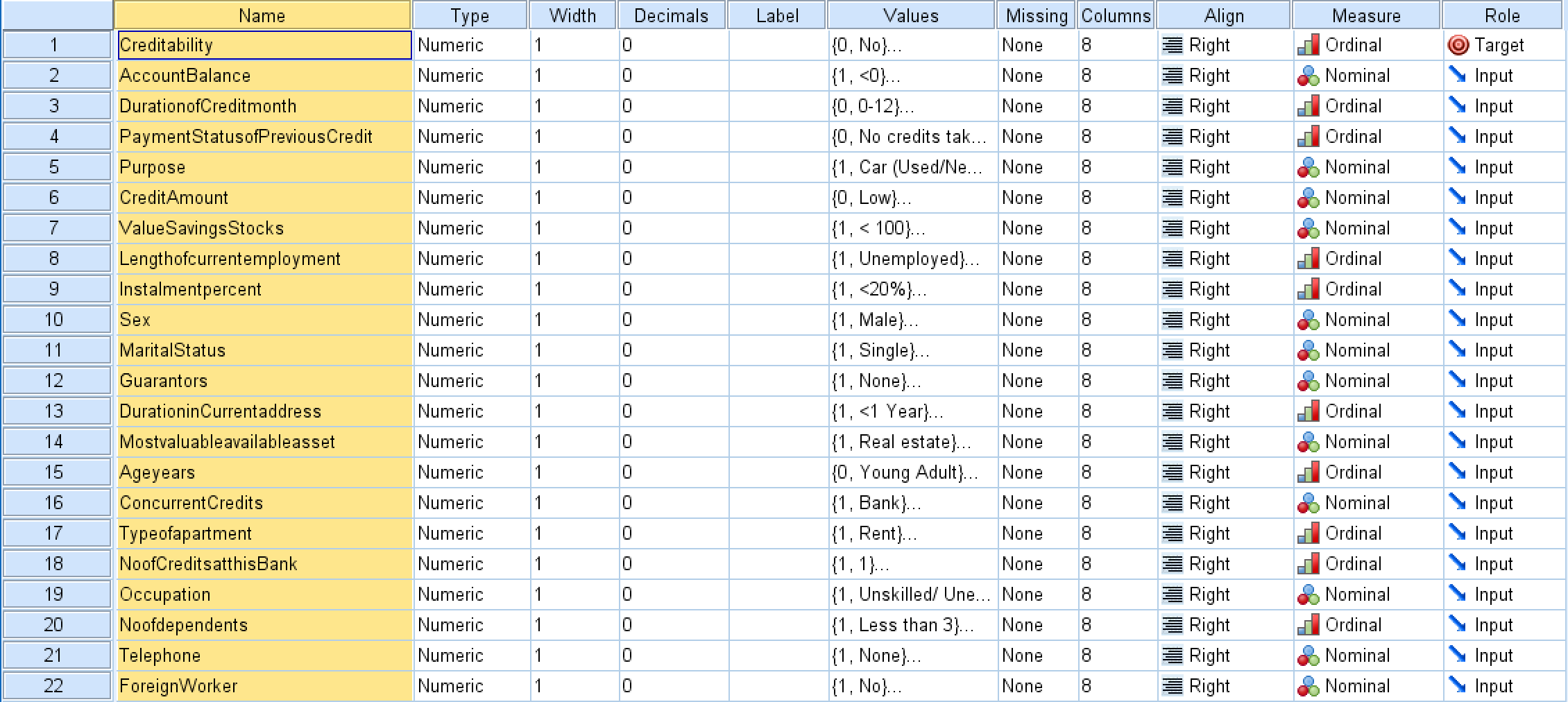
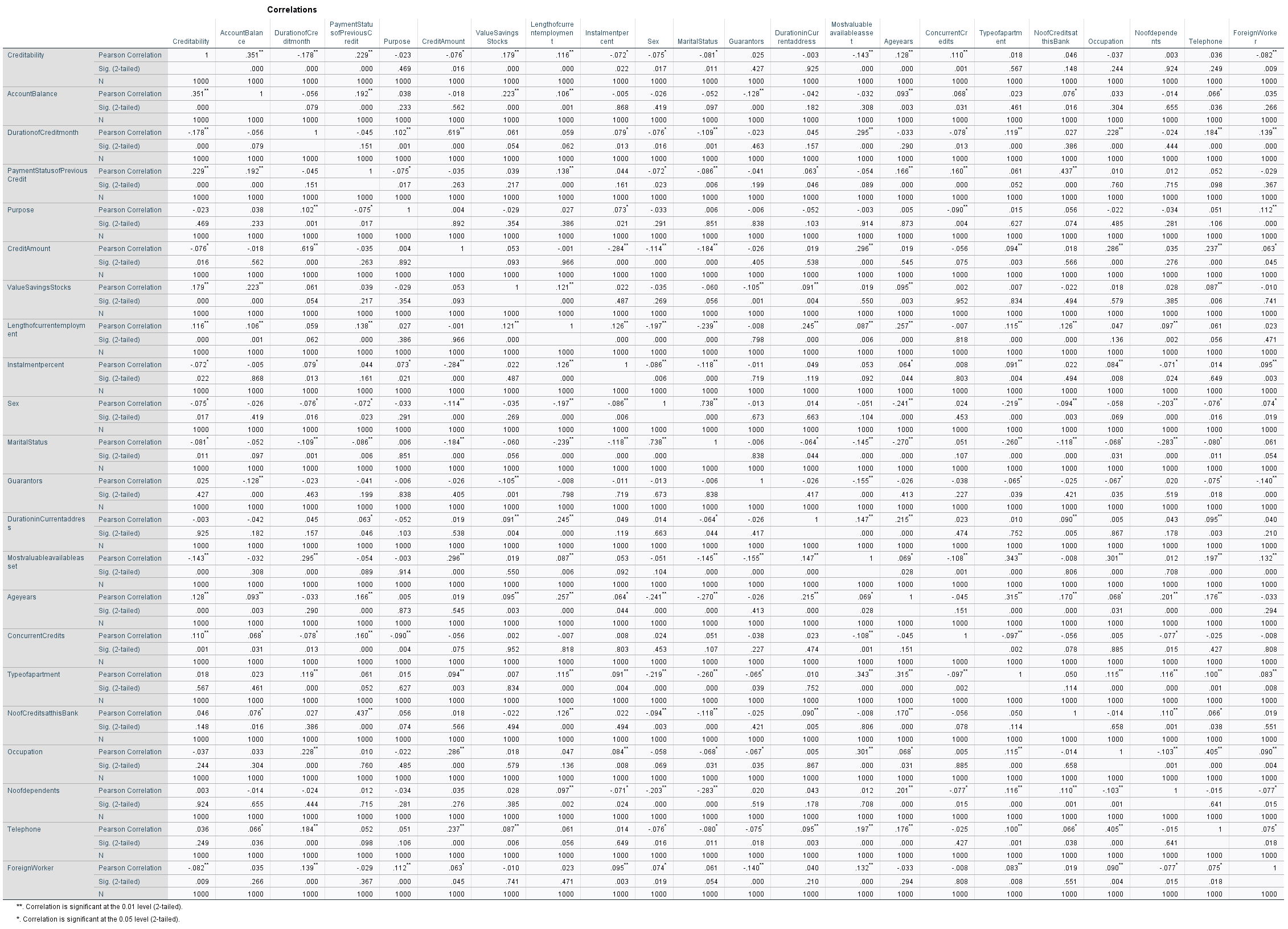
**German Data Bivariate and Data Reduction in SPSS**

Adjusting the measures of each variables & labelling the values



Performing Pearson’s Chi Square Correlations (Spreadsheets will be provided) (haha so smol)





Findings:

Variables that are **not** significant in determining Creditability:

* Duration in current address
* No. of dependents
* Telephone
* Occupation
* No of credits at this bank
* Type of Apartment
* Guarantors
* Purpose



Variables that are highly correlated with Creditability, ranked from highest positive correlation to highest negative correlation:

1. Account Balance .351
2. Payment Status of Previous Credit .229
3. Value Savings Stocks .179
4. AgeYears .128
5. Length of Current Employment .116
6. Concurrent Credits .110
7. Foreign Worker .082
8. Instalment Percent -.072
9. Sex -.075
10. Credit Amount -.076
11. Marital Status -.081
12. Most valuable available asset -.143
13. Duration of Credit Month -.178

Correlation of Independent Variables Findings



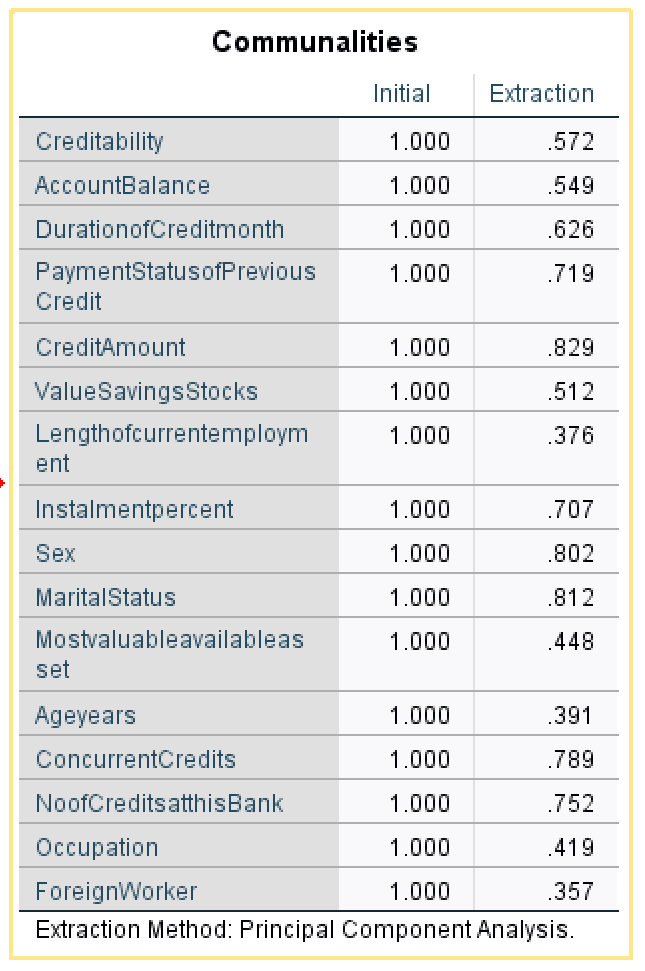
* Sex and Marital Status : .738
* Credit Amount with Duration of Credit Month : 0.619
* No. of Credit at this bank & Payment Status of Previous Credit : 0.437
* Other than that, not much significant high correlation between the independent variables

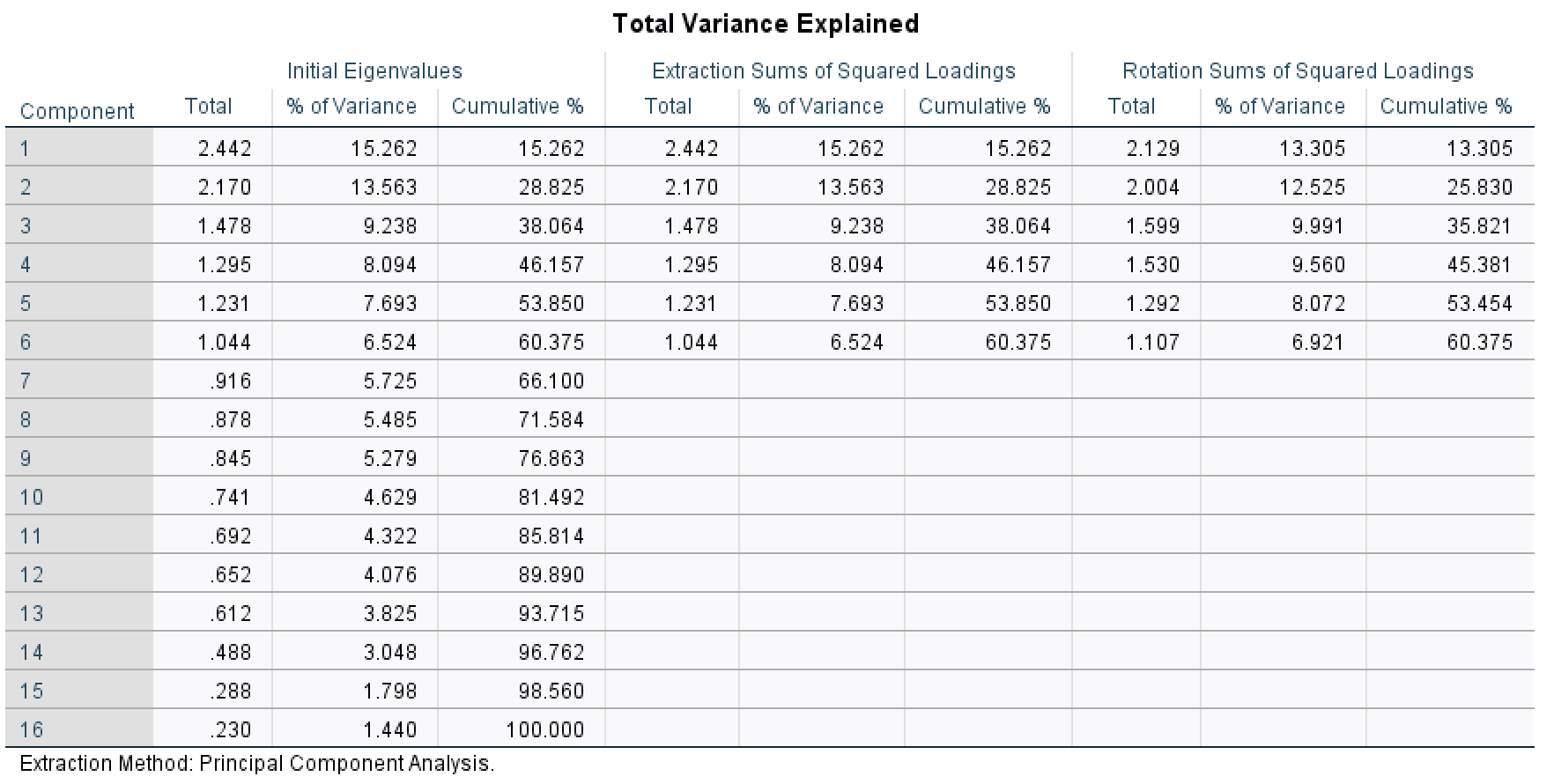
**Experiment on Data Reduction**

Variables that I have chosen to not be included in PCA:

* Duration in current address
* No. of dependents
* Telephone
* Type of Apartment
* Guarantors
* Purpose

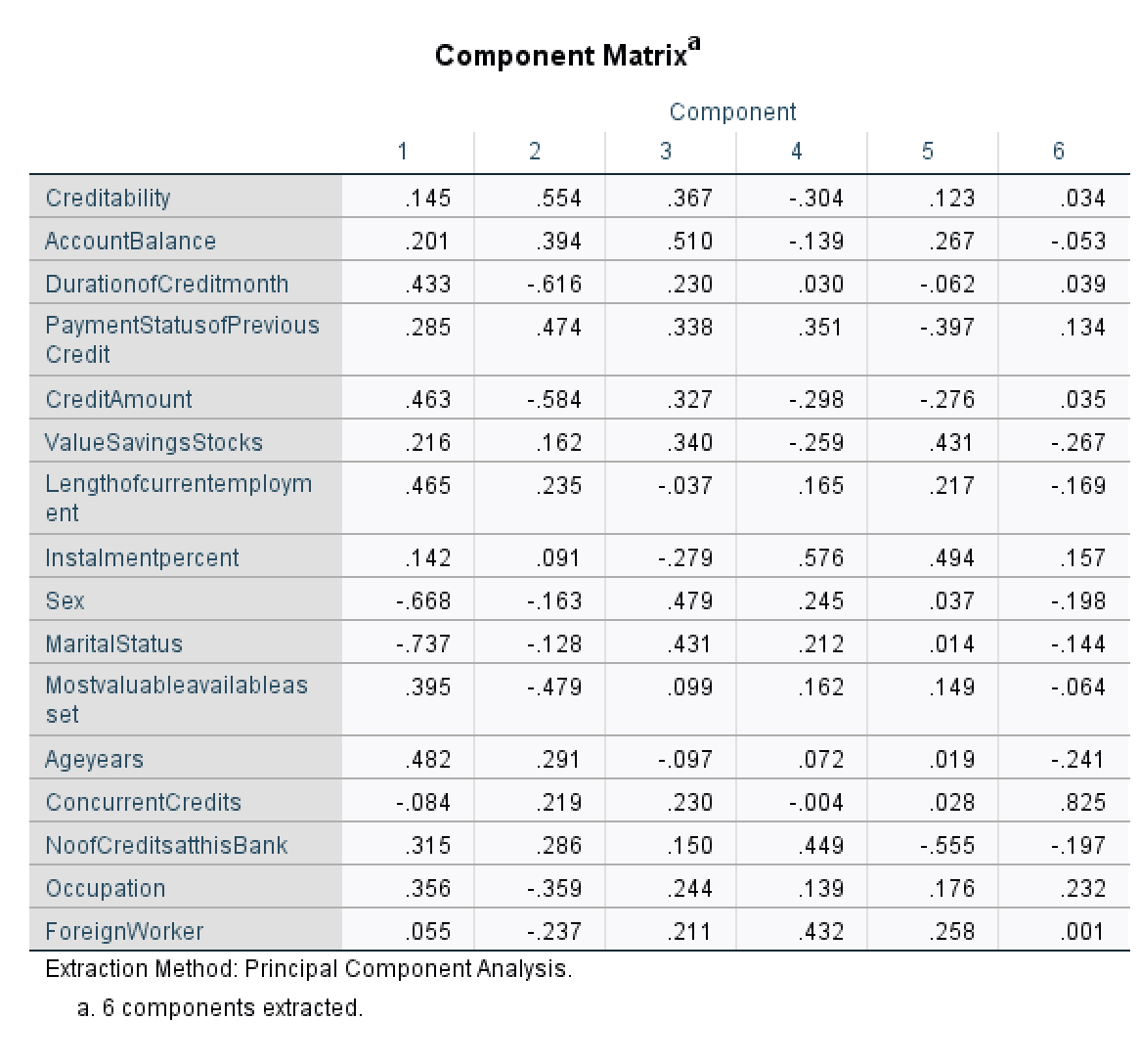
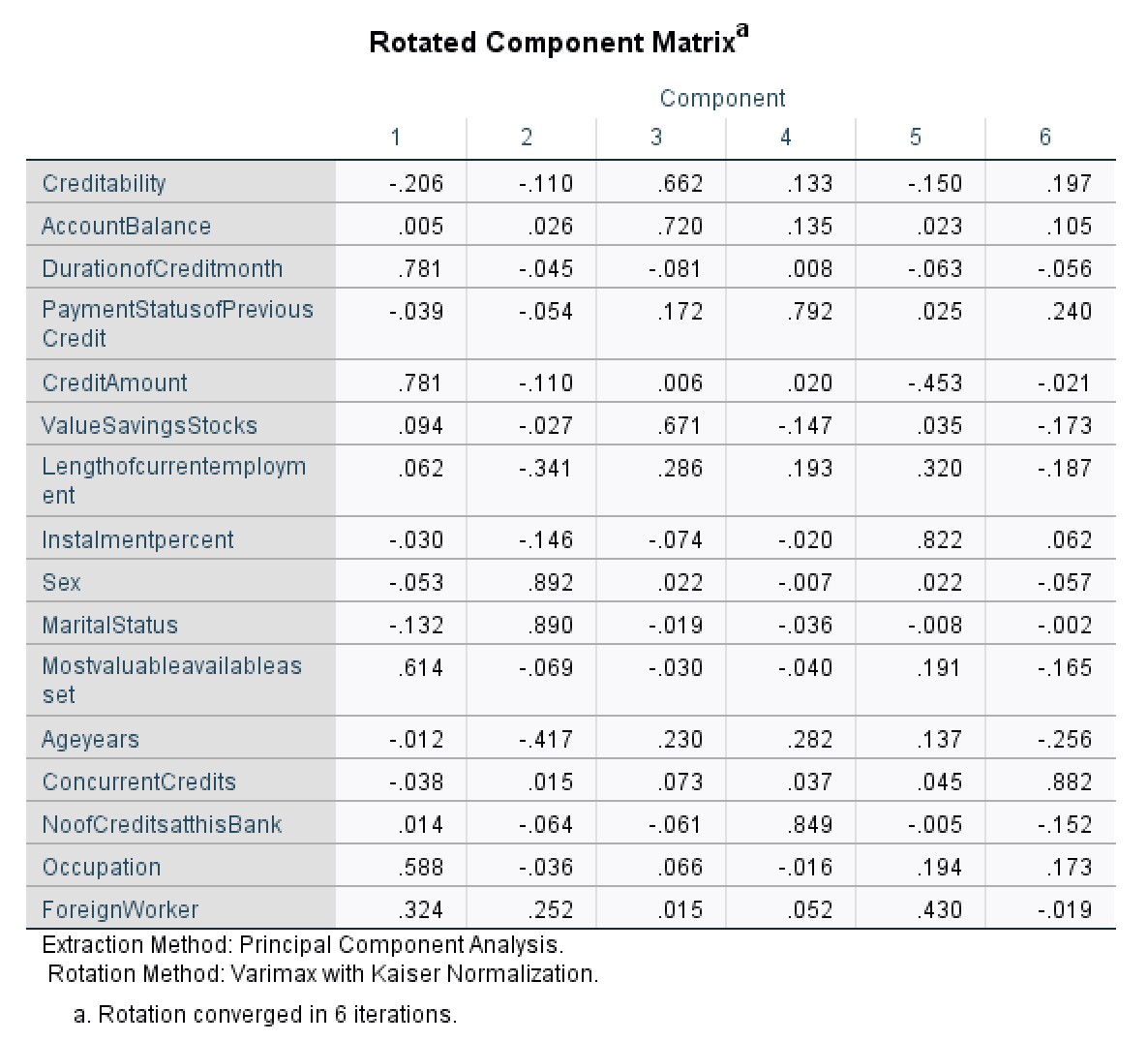
Performing PCA (with 16 variables)







Based on the Scree Plot, the line starts to flatten when it reaches 3PCs, but I have chosen 6 PCs based on the criterion of Eigenvalue > 1.

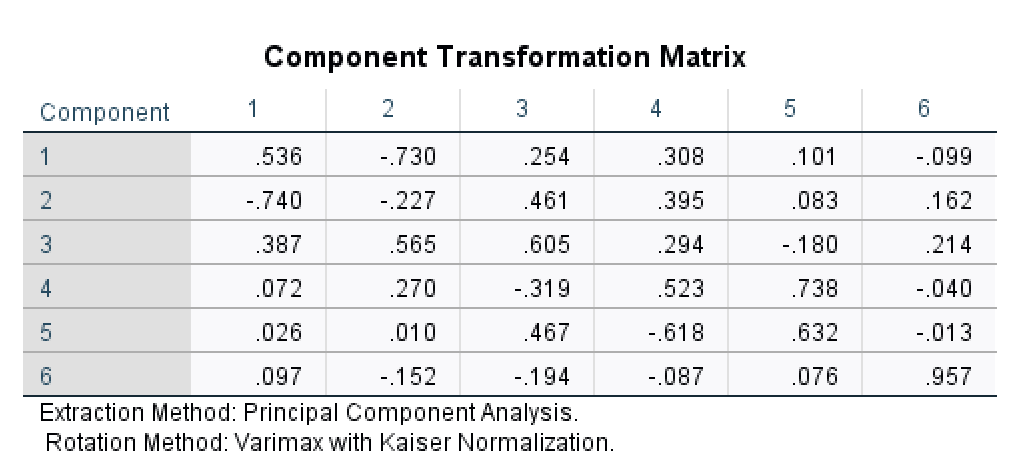
Based on the Rotated scores:

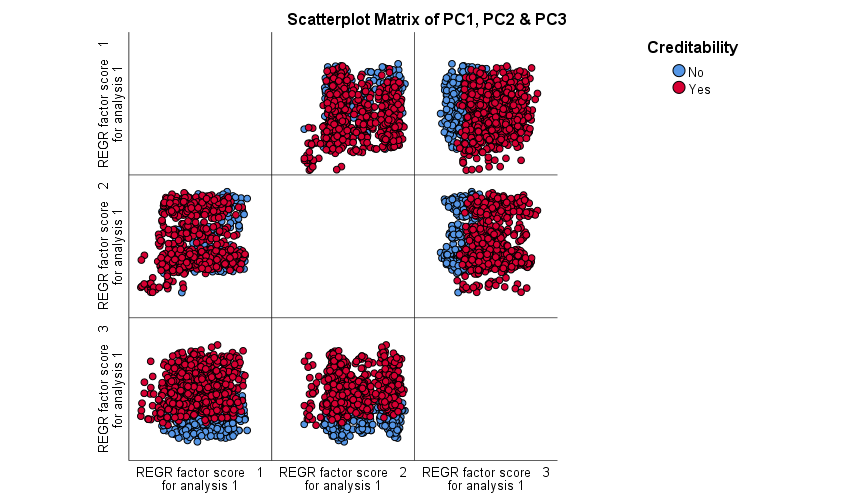
PC 1 mainly accounts for the variance in Credit Amt, Duration in Credit month, Most valuable available assets, Occupation 🡪 Financial criteria; makes sense because these are the more important factors to look at.

PC 2 mainly accounts for the variance in Sex, Marital status & Age Years 🡪 Demographics

PC 3 mainly accounts for the variance in Account Balance & Value savings stocks 🡪 Looking at how much money they actually have in hand

PC 4 mainly accounts for the variance in No of Credit at this bank & Payment status of previous credit 🡪 to see their credit status





There seems to be no visible cluster based on the scatterplot matrix. It seems that those that are not creditable have low PC3 scores.

**Performing Correspondence Analysis**



Based on the total Inertia score of each independent variable we can see that there are certain variables that do not have ability to offer conclusions eg Telephone, No of dependents, Type of apartment, duration in current address…