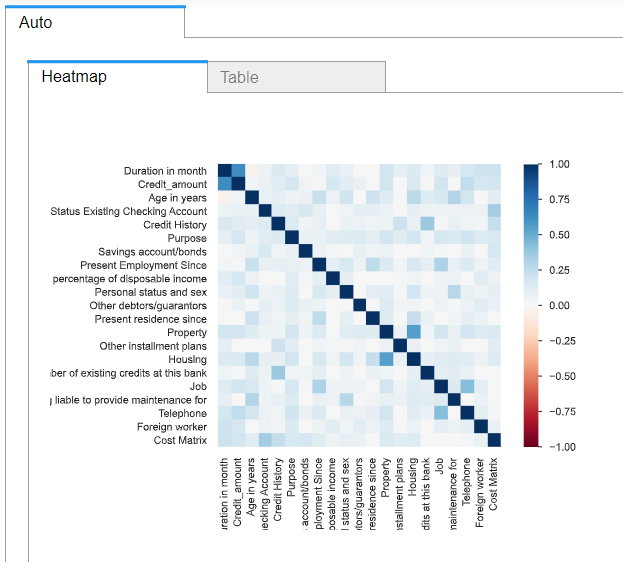
**Highly Correlated Data**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



Duration in month ~ Credit Amount (very high)

Status Existing Checking Account ~ Cost Matrix (med / important?)

Credit History ~ Number of Existing Credits at this bank (high)

Present Employment Since ~ Job (med)

Personal Status and sex ~ Liable to provide maintenance for (med)

Property ~ Housing (high-very high)

Job ~ Telephone (med-high)

**Features to use, amount of values and total columns post one-hot**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1 Duration in month

2 Credit Amount

3 Property (4)

4 Housing (3)

5 Credit History (5)

6 Number of Existing Credits at this bank

7 Job (4)

8 Telephone (2) \*binary

9 Status Existing Checking Account (4)

10 Cost Matrix \*binary

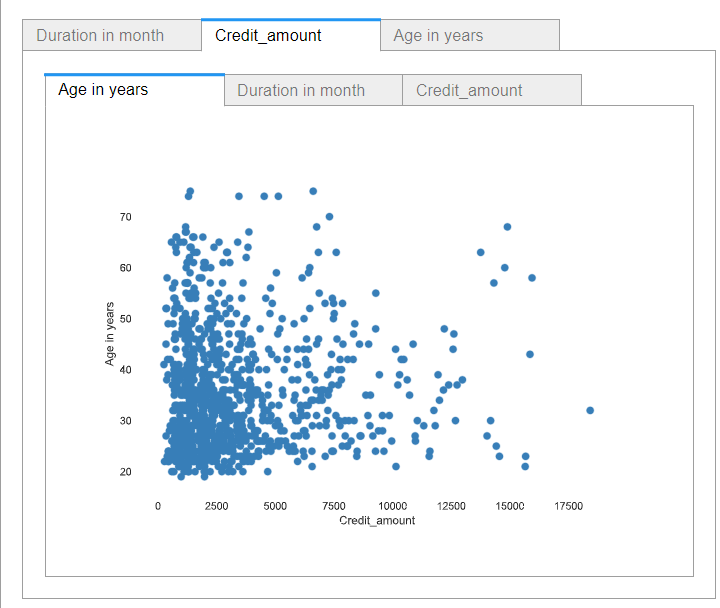
11 Present Employment Since (5)

12 Personal Status and sex (5)

13 Liable to provide maintenance for

14 Age in years

Dimension / Total columns after one hot-encoding = 44



Numerical

1 - Age in Years\*

2 - Credit Amount\*

3 - Duration in month\*

4 - Number of Existing Credits at this bank\*

5 - Number of people being liable to provide maintenance for\*

6 - President Residence Since

\* - using