

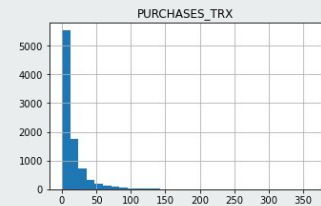
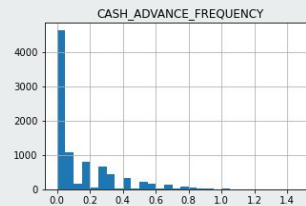
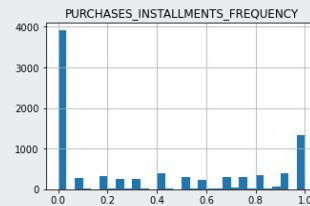
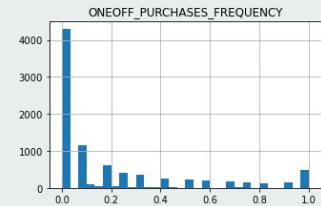
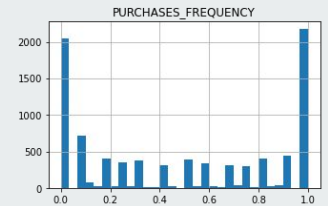
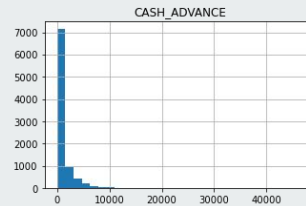
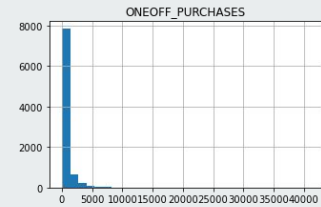
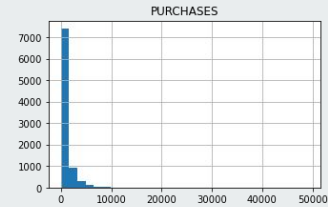
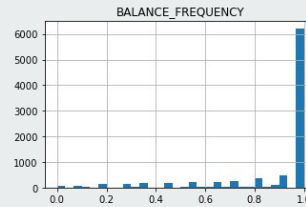
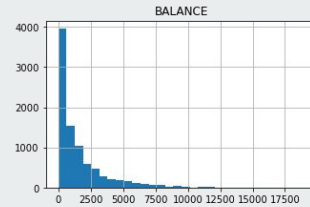


Credit card data analysis

Haiyan Saadi & Malte Åhman

Credit card dataset

- Range: $[0,1]$ or $[0,\text{inf}]$
- 17 variables
- Clustering and classification

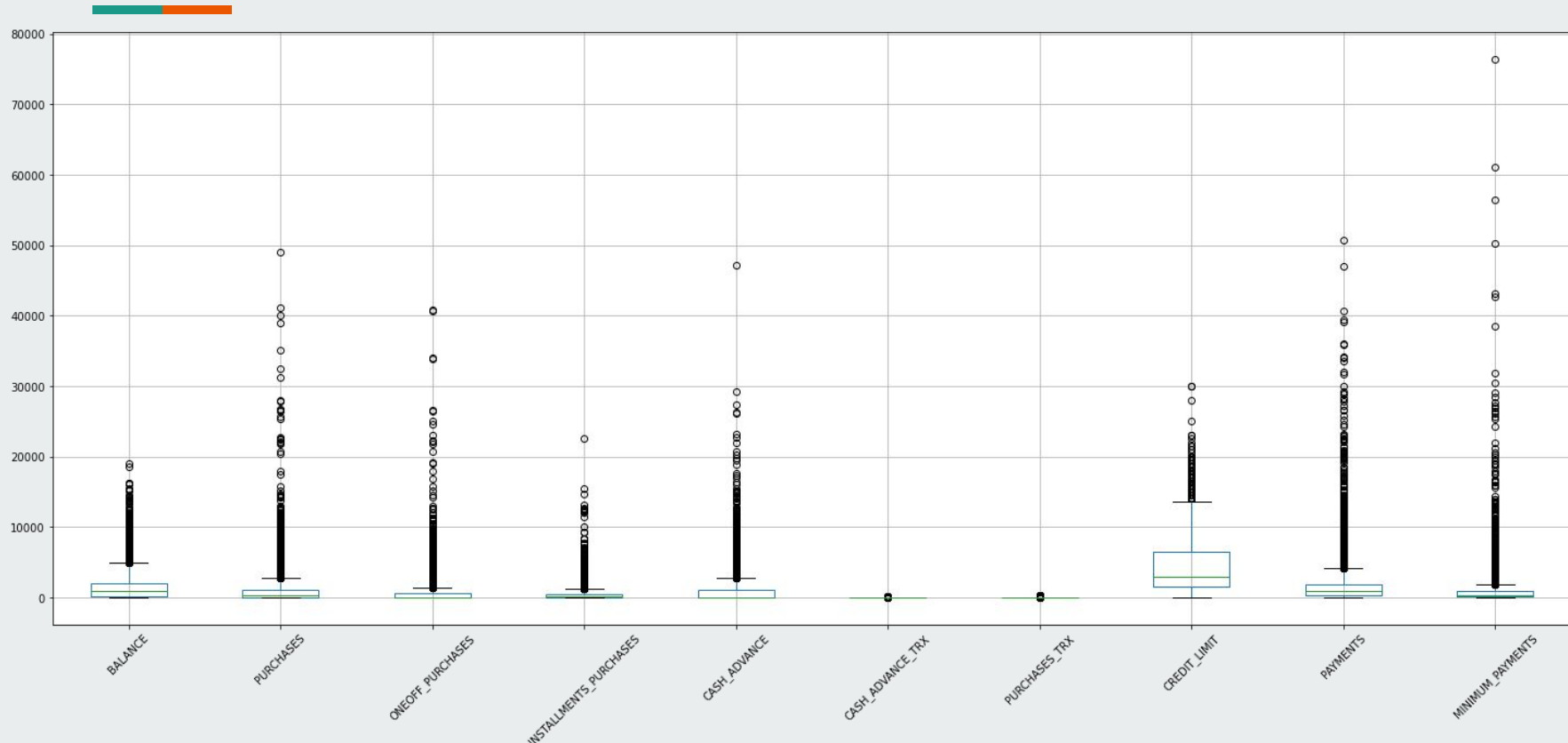


Pre-processing

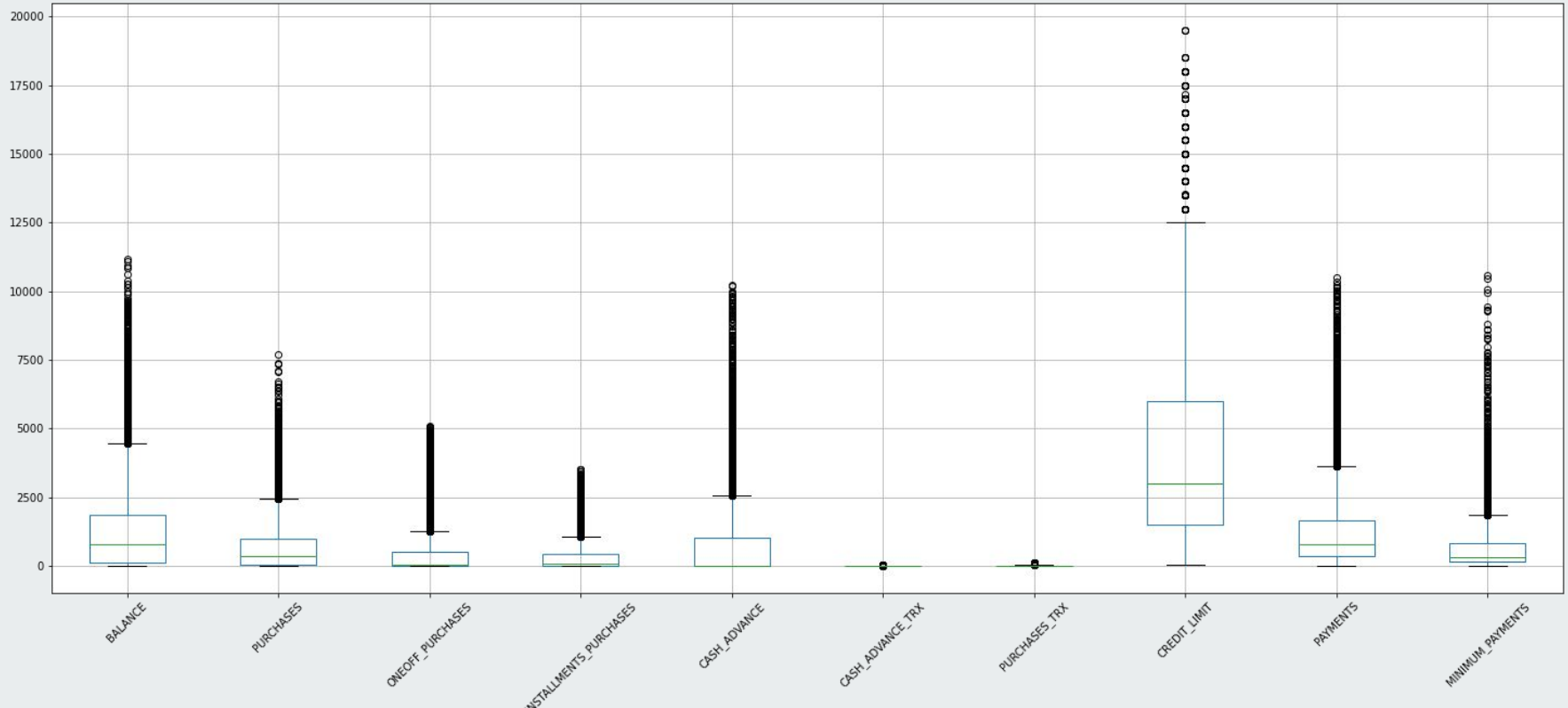


- Outlier detection and removal
- Correlation matrices
- Normalization of data
- Some sampling

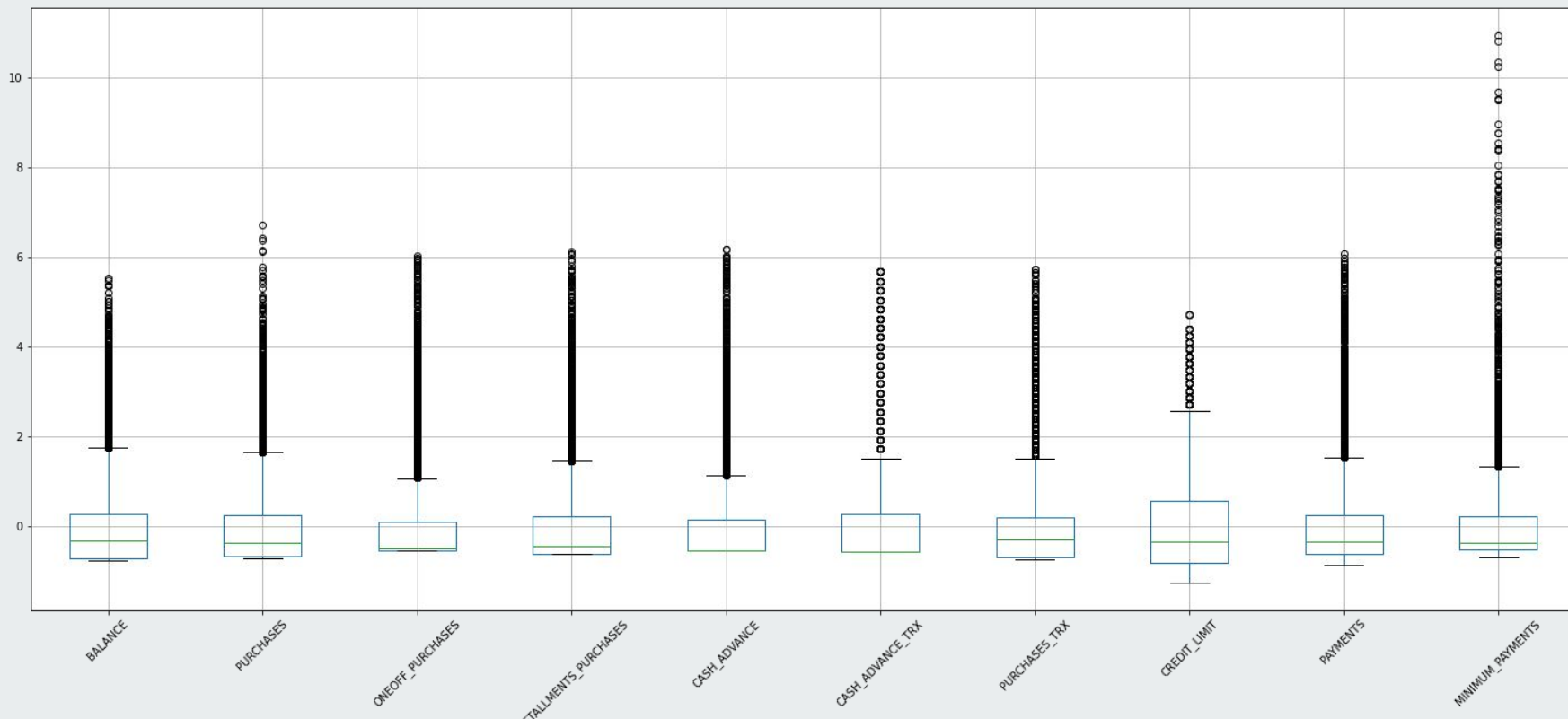
Box plots for raw data



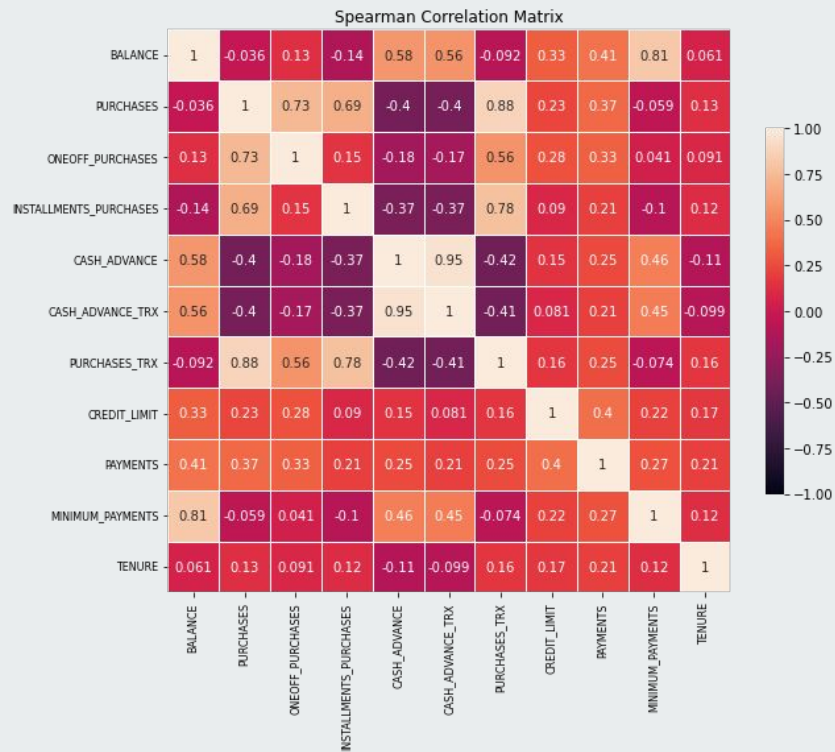
After removing outliers



With z-score normalization

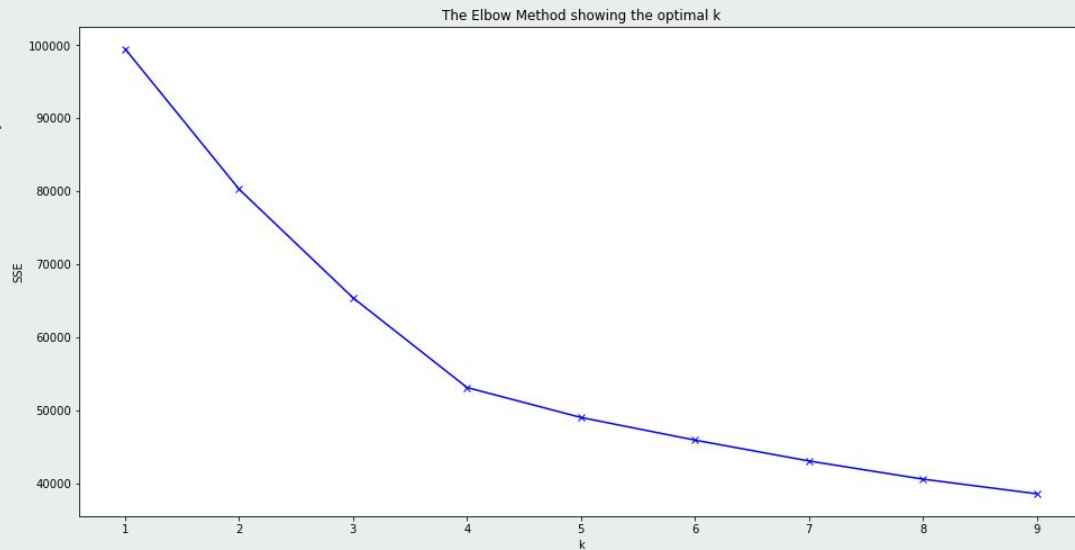


Correlation heatmaps

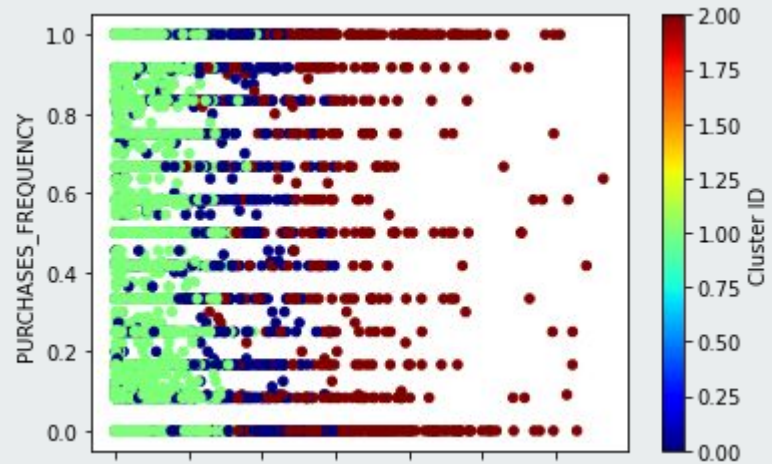
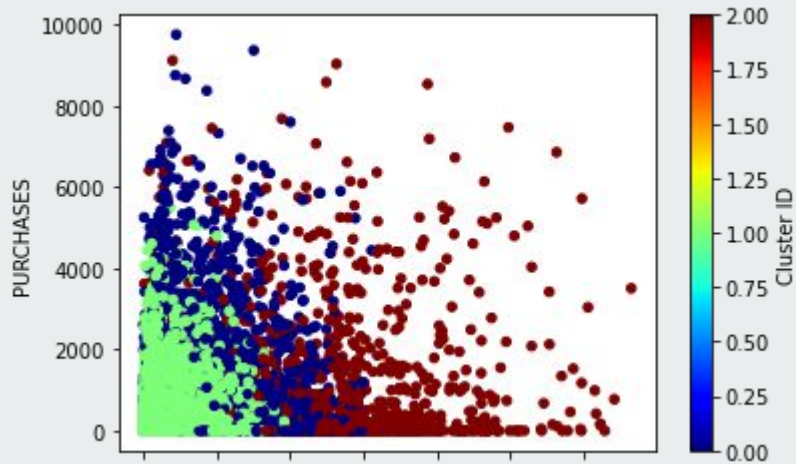


Clustering with K-means

- Elbow method to find optimal k
- Changed when normalizing: $3 \rightarrow 4$

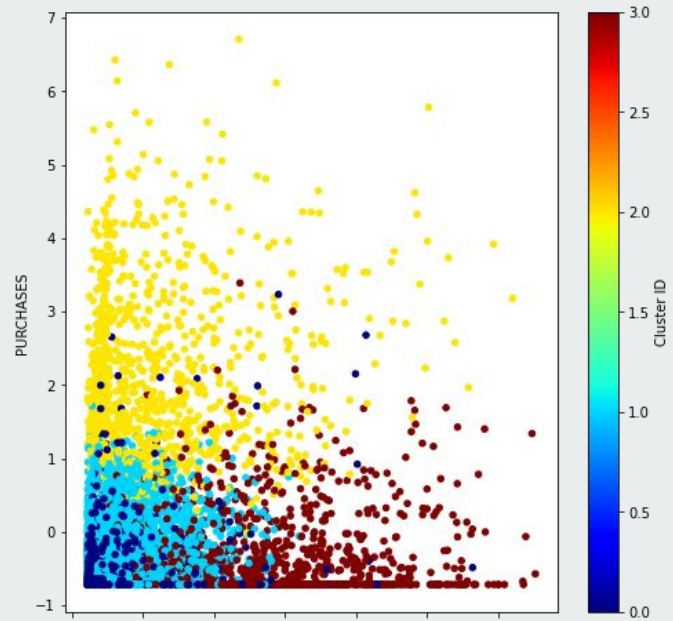


Clustering



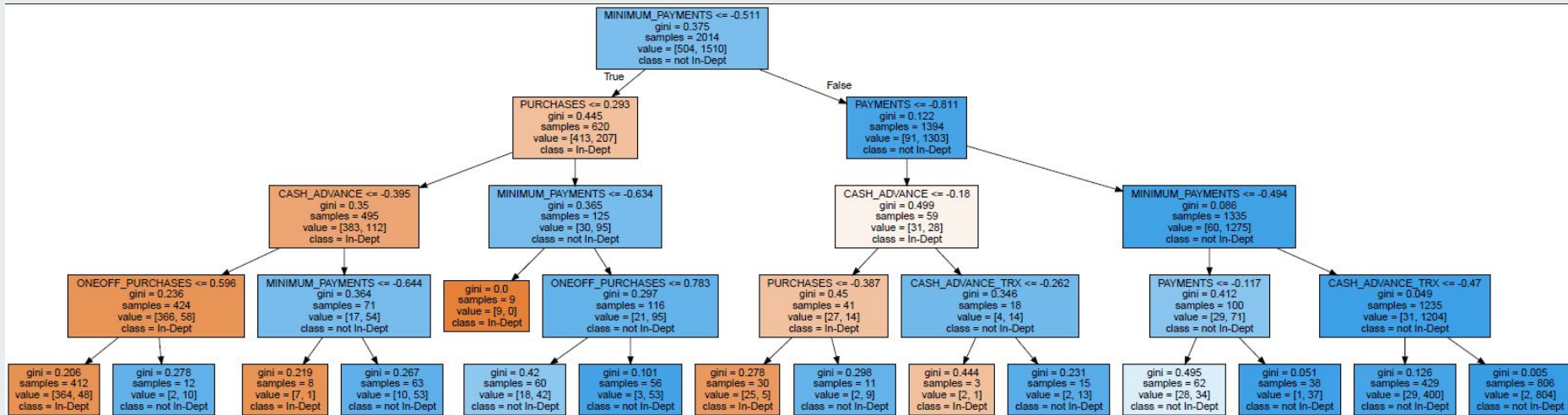
On normalized data

- Optimal number of clusters increased
- Still hard to interpret...



Classification

- decision tree classifier (Accuracy 92.84%)





Thank you

Questions?