Improving Decision Making in Credit and Lending

•••

Project 2
Rüdiger Hass
Johannes Pastorek
2 July 2020

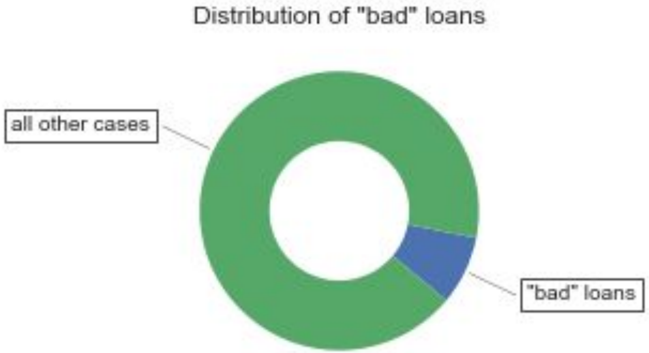
Our Goal

“lower loan risk by identifying patterns from within historical data using machine learning models.”

Historical Data

no. of loans:	365,255
client properties:	123
datapoints:	43,819,365
missing values:	10,605,628
missings in %:	24

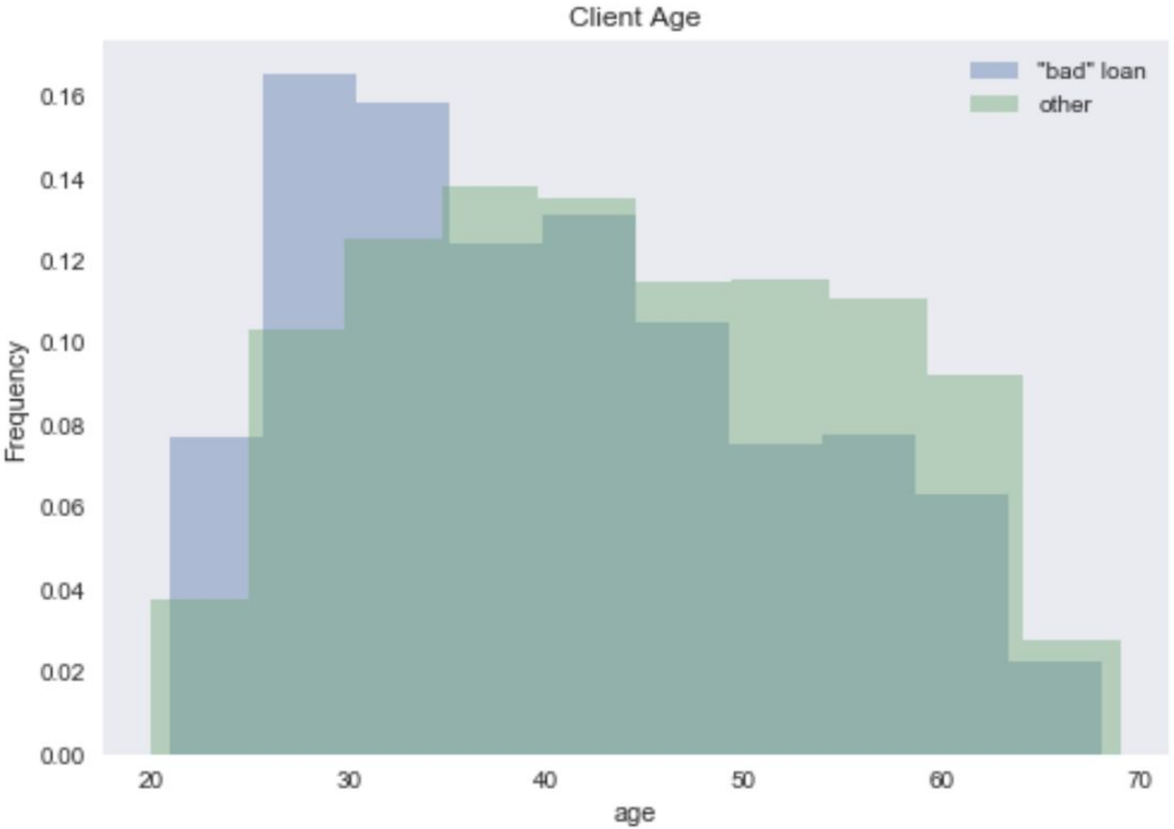
Historical Data



“client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample.”

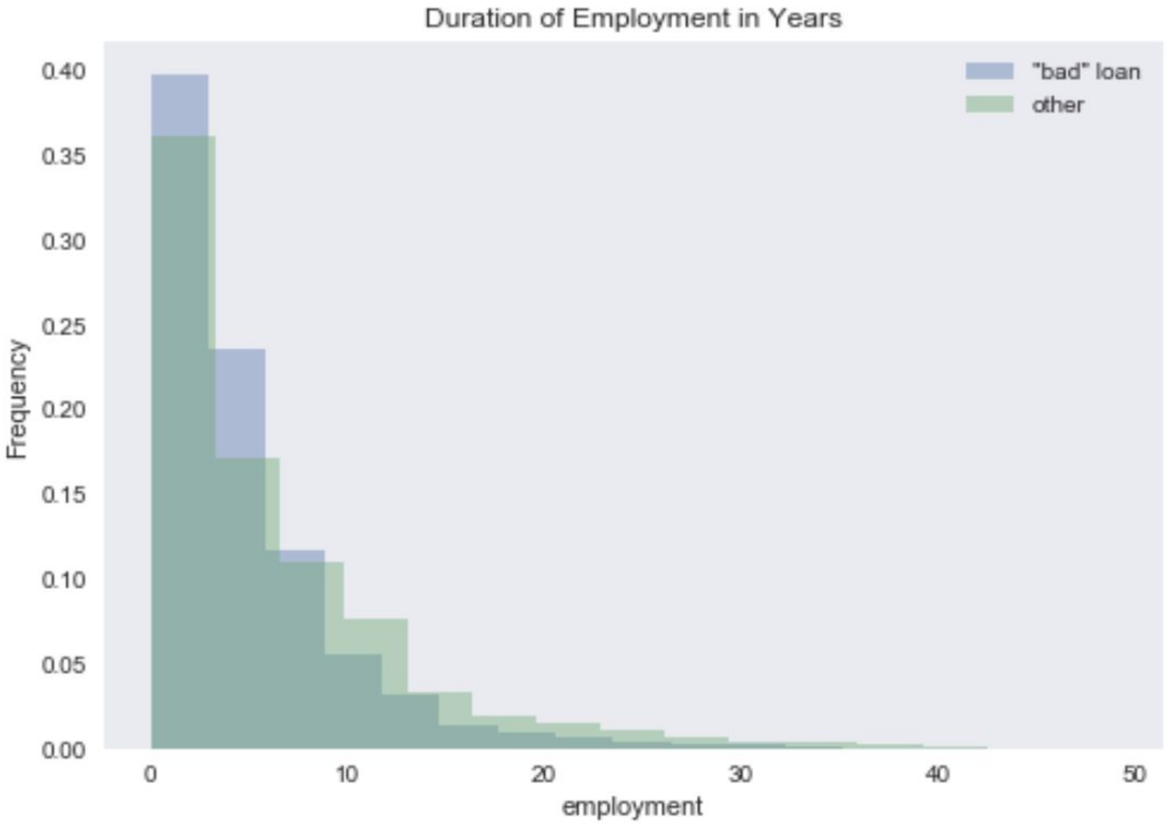
Interesting Features:

- ❑ Age
- ❑ Years employed
- ❑ Gender
- ❑ Education
- ❑ Age of car
- ❑ External Source 1
- ❑ External Source 2
- ❑ External Source 3



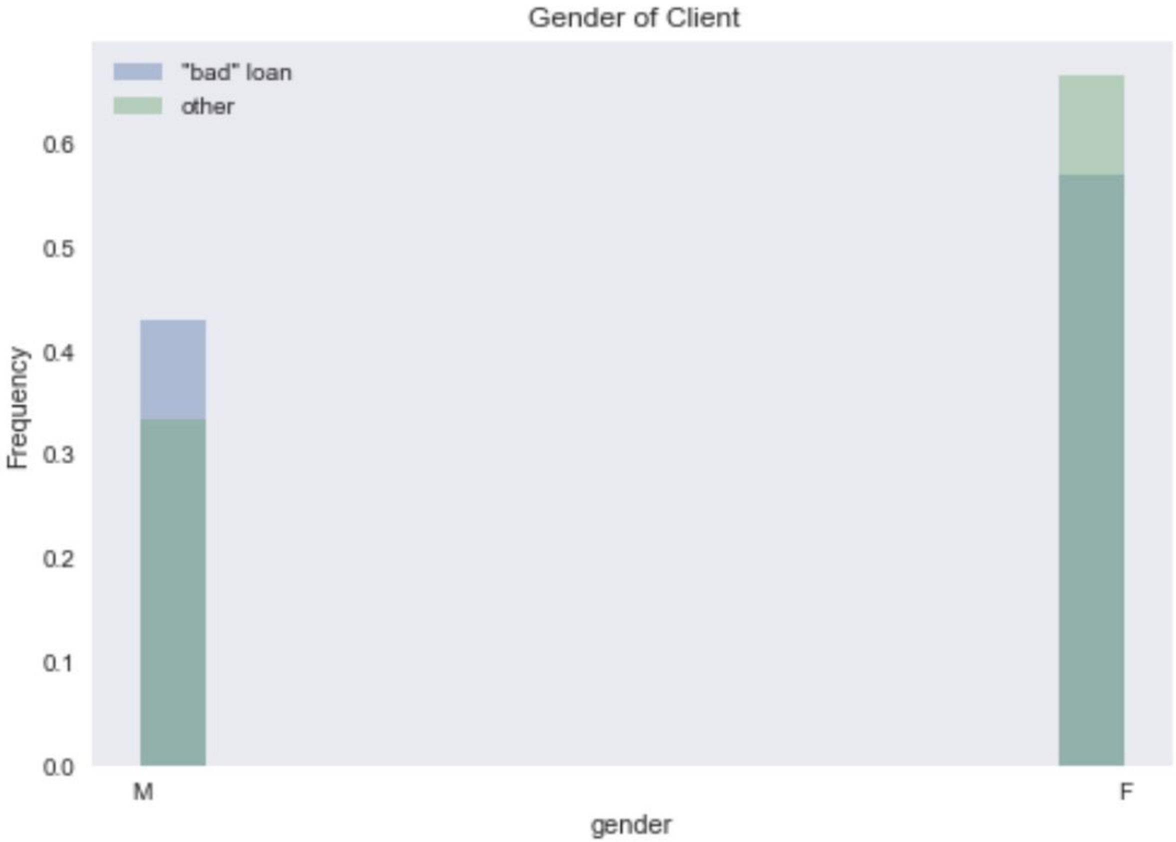
Interesting Features:

- ❑ Age
- ❑ Years employed
- ❑ Gender
- ❑ Education
- ❑ Age of car
- ❑ External Source 1
- ❑ External Source 2
- ❑ External Source 3



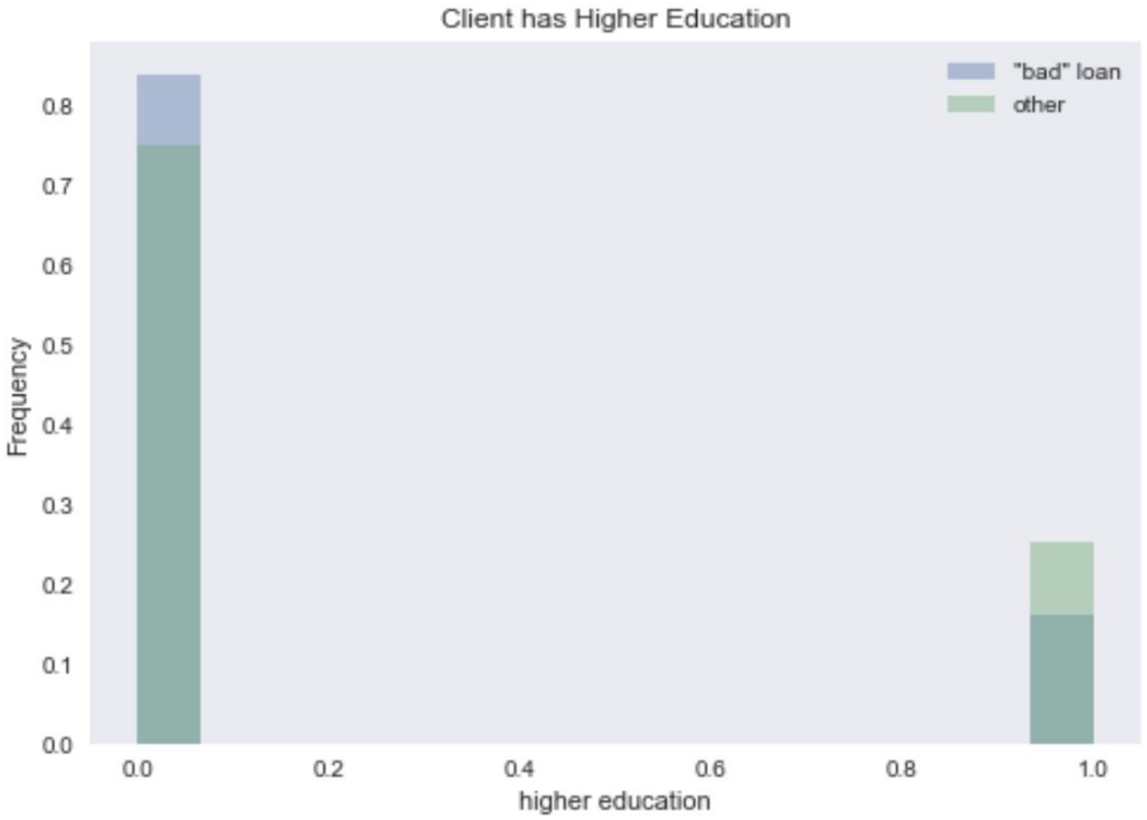
Interesting Features:

- ❑ Age
- ❑ Years employed
- ❑ Gender
- ❑ Education
- ❑ Age of car
- ❑ External Source 1
- ❑ External Source 2
- ❑ External Source 3



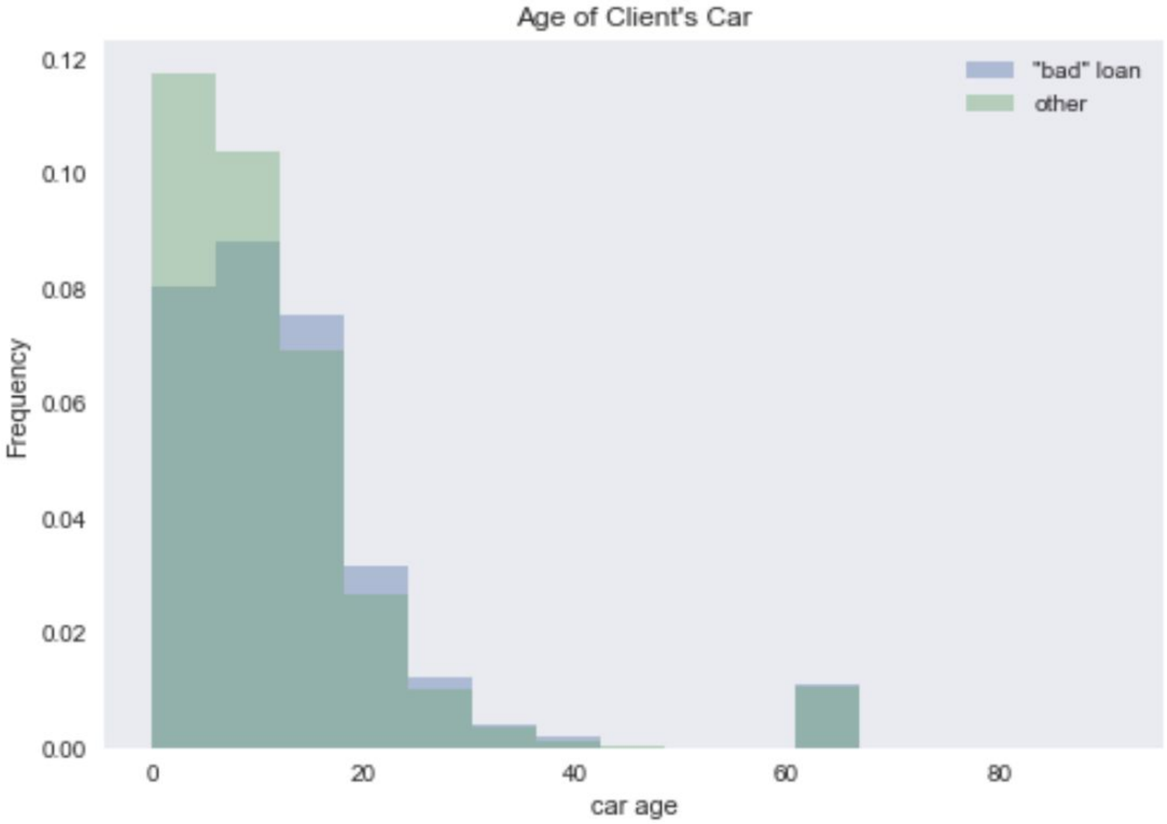
Interesting Features:

- ❏ Age
- ❏ Years employed
- ❏ Gender
- ❏ Education
- ❏ Age of car
- ❏ External Source 1
- ❏ External Source 2
- ❏ External Source 3



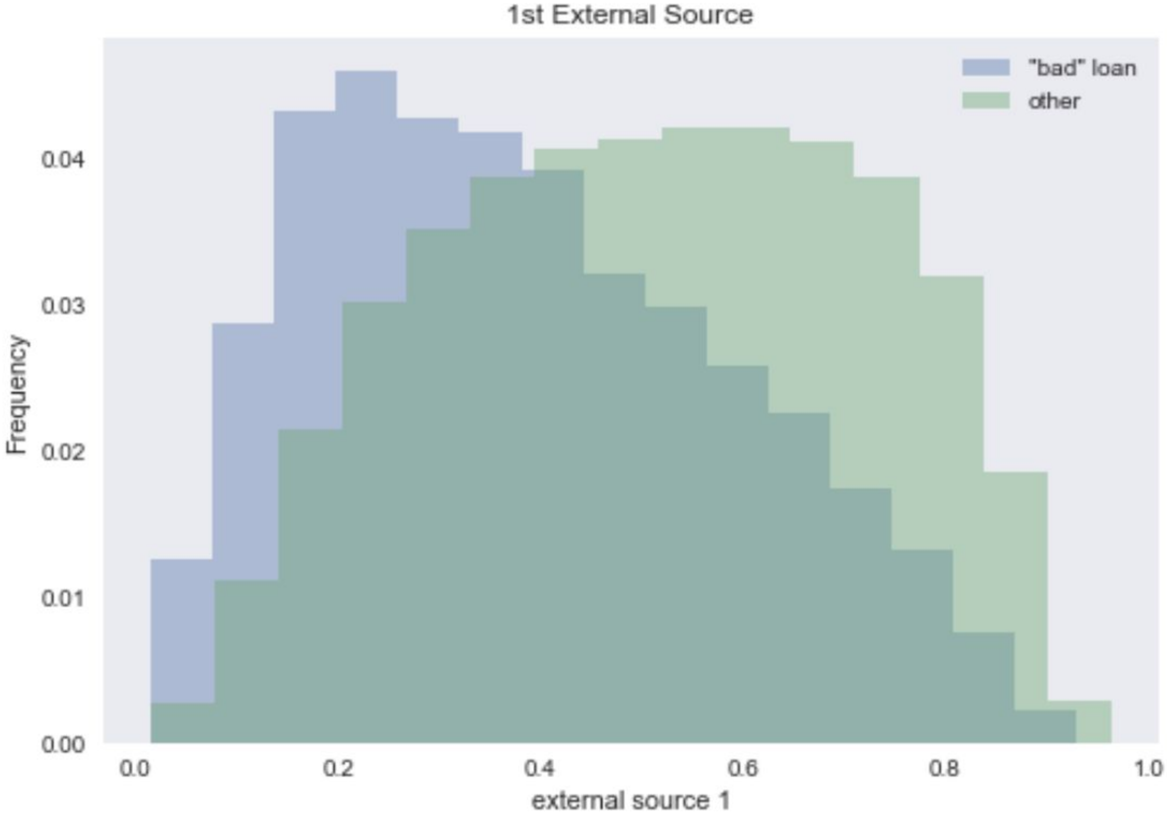
Interesting Features:

- ❏ Age
- ❏ Years employed
- ❏ Gender
- ❏ Education
- ❏ Age of car
- ❏ External Source 1
- ❏ External Source 2
- ❏ External Source 3



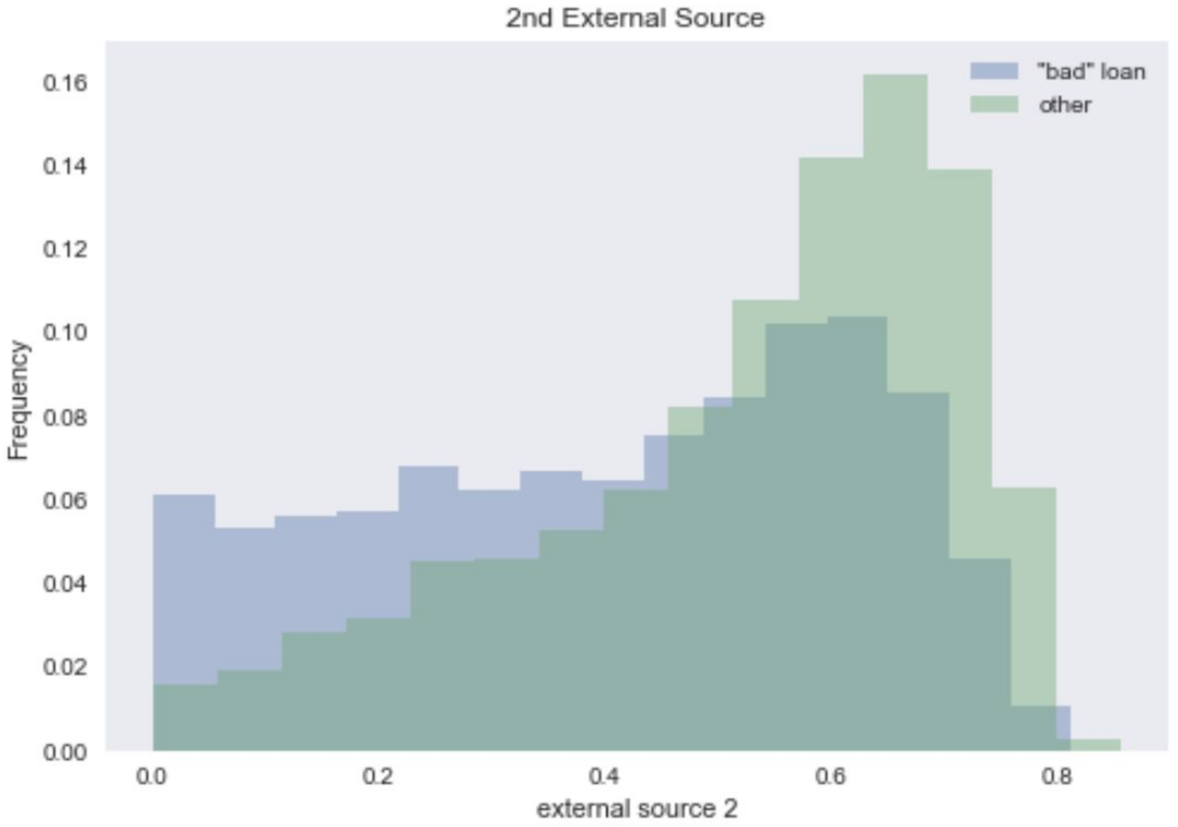
Interesting Features:

- ❑ Age
- ❑ Years employed
- ❑ Gender
- ❑ Education
- ❑ Age of car
- ❑ External Source 1
- ❑ External Source 2
- ❑ External Source 3



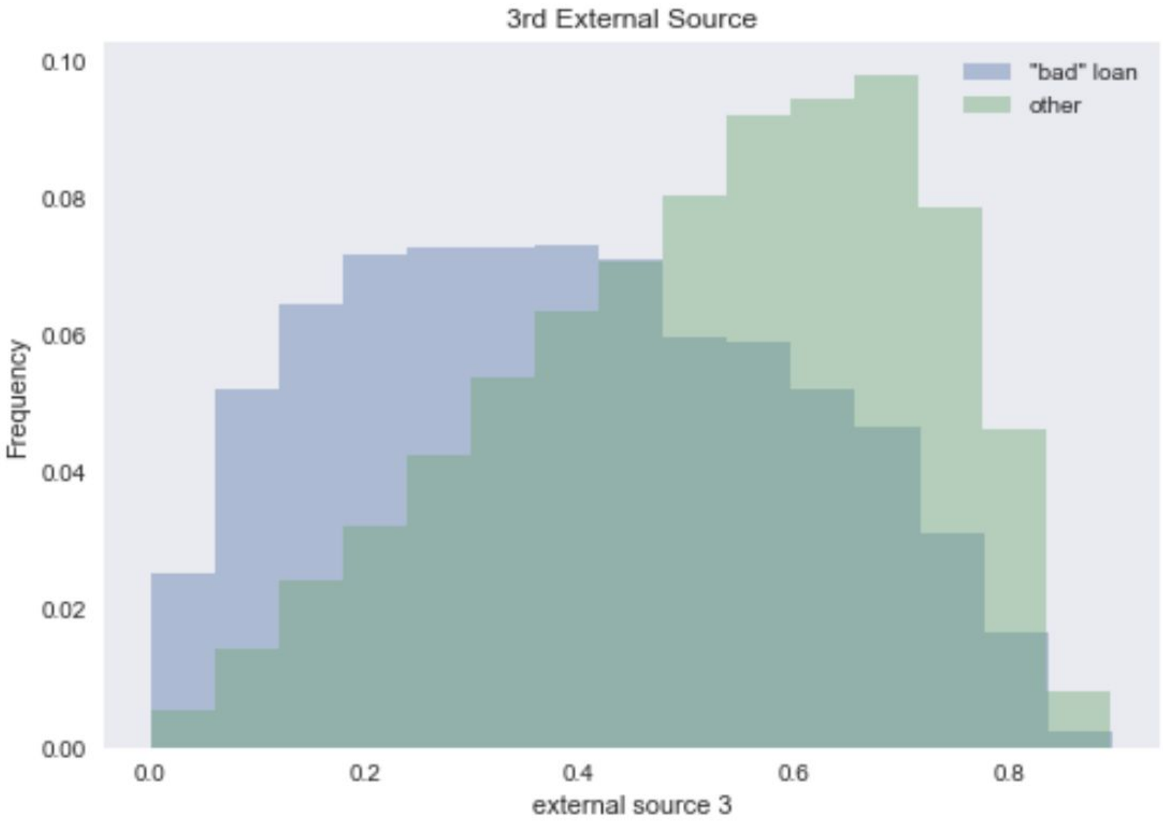
Interesting Features:

- ❏ Age
- ❏ Years employed
- ❏ Gender
- ❏ Education
- ❏ Age of car
- ❏ External Source 1
- ❏ External Source 2
- ❏ External Source 3

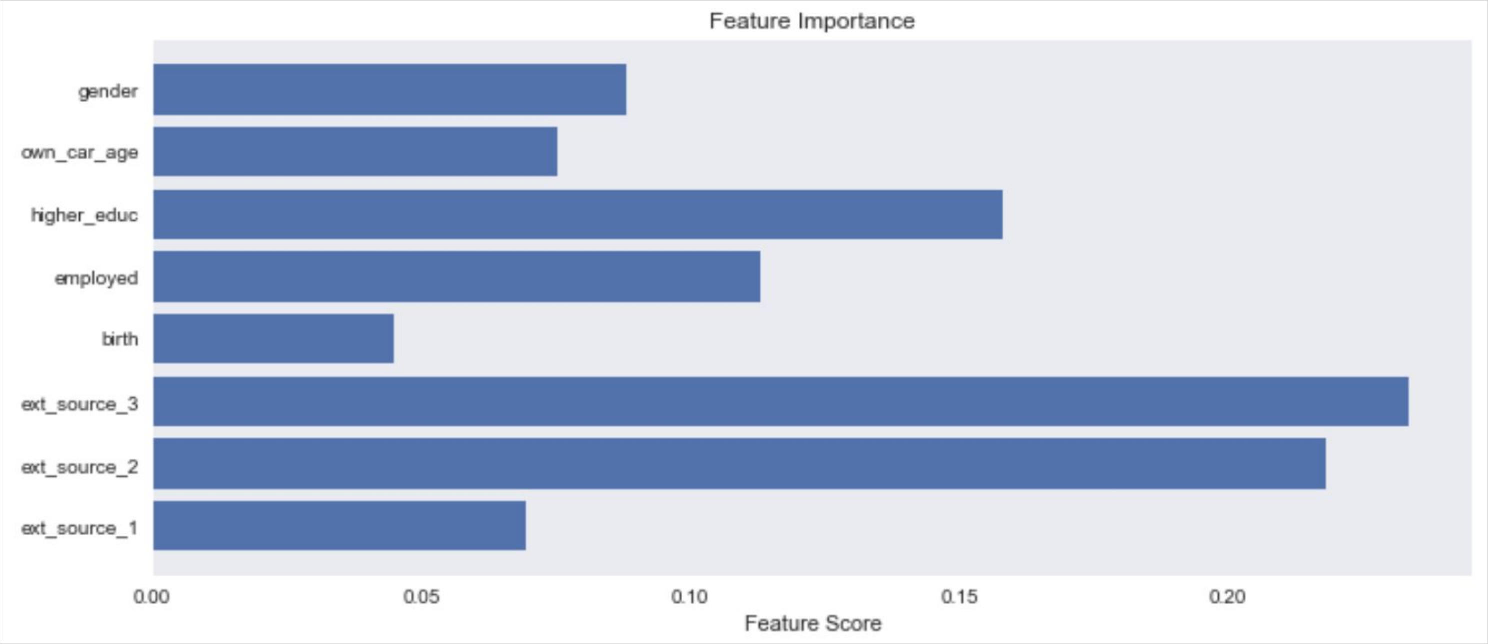


Interesting Features:

- ❑ Age
- ❑ Years employed
- ❑ Gender
- ❑ Education
- ❑ Age of car
- ❑ External Source 1
- ❑ External Source 2
- ❑ External Source 3



What are the driving factors?



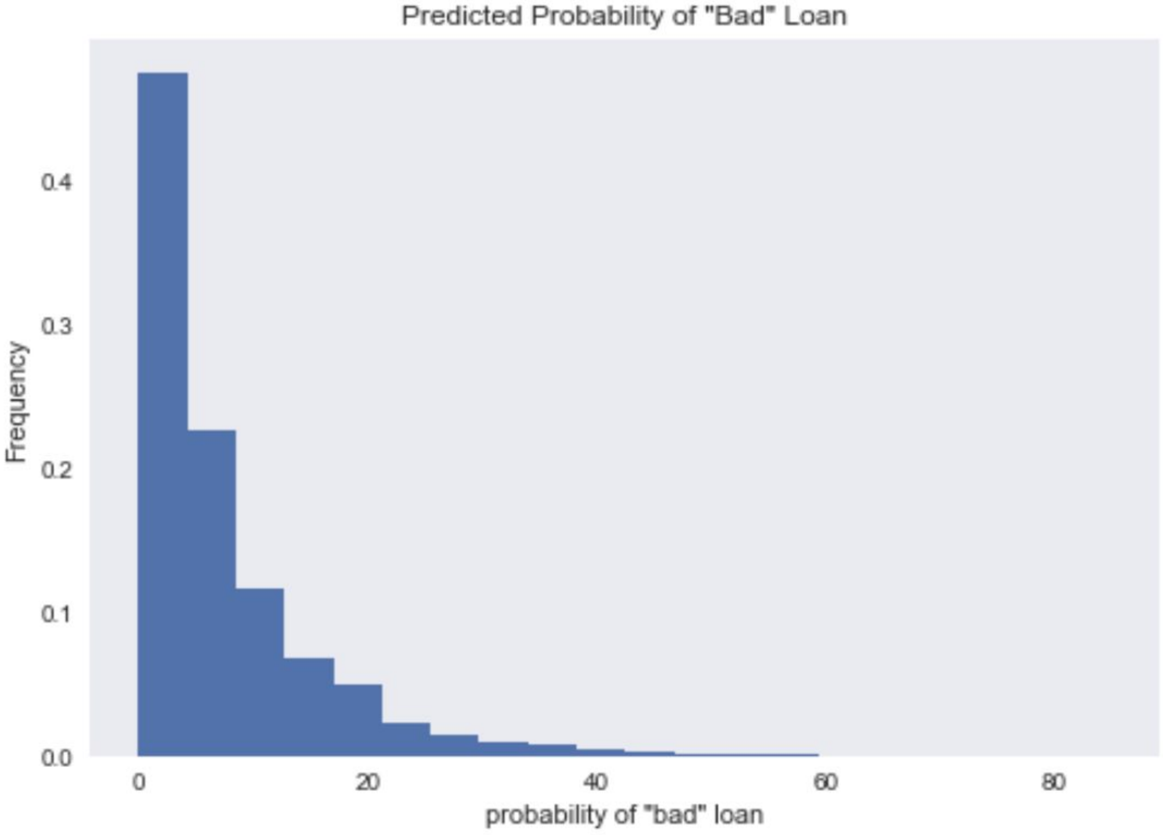
How does the model perform?

	“good” loan	“bad” loan
“good” loan	92 %	35 %
“bad” loan	8 %	65 %

PREDICTED

ACTUAL

Final Prediction



Discussion

blind spots in data

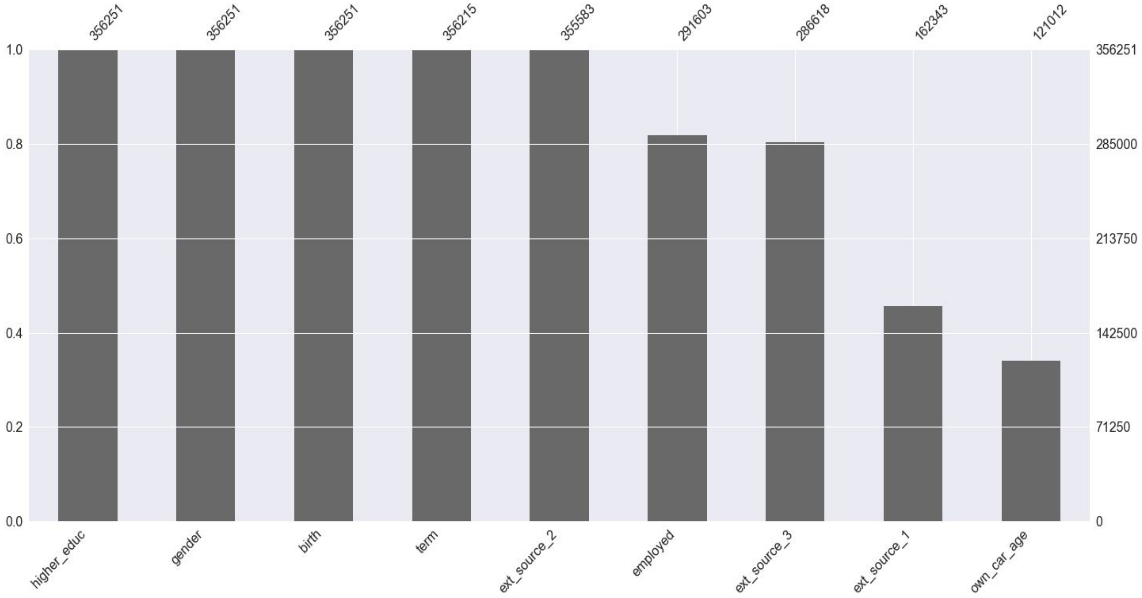
❏ Imbalanced dataset

❏ Missing Values:
frequency and distribution

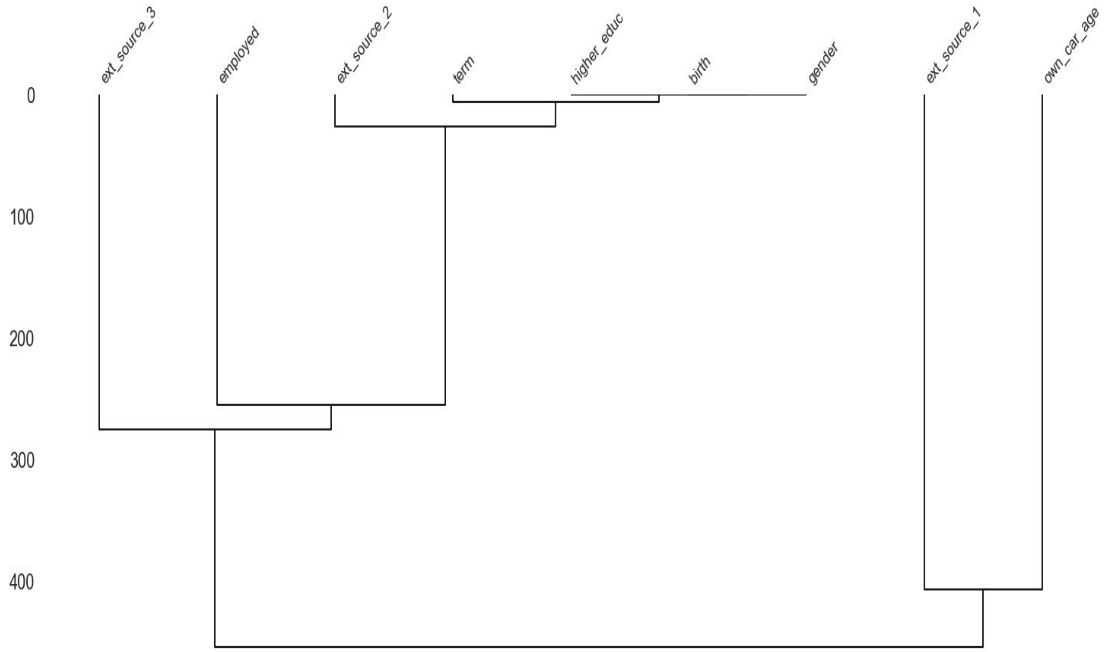
❏ Traces of multicollinearity

❏ Little knowledge about features

Missing Values

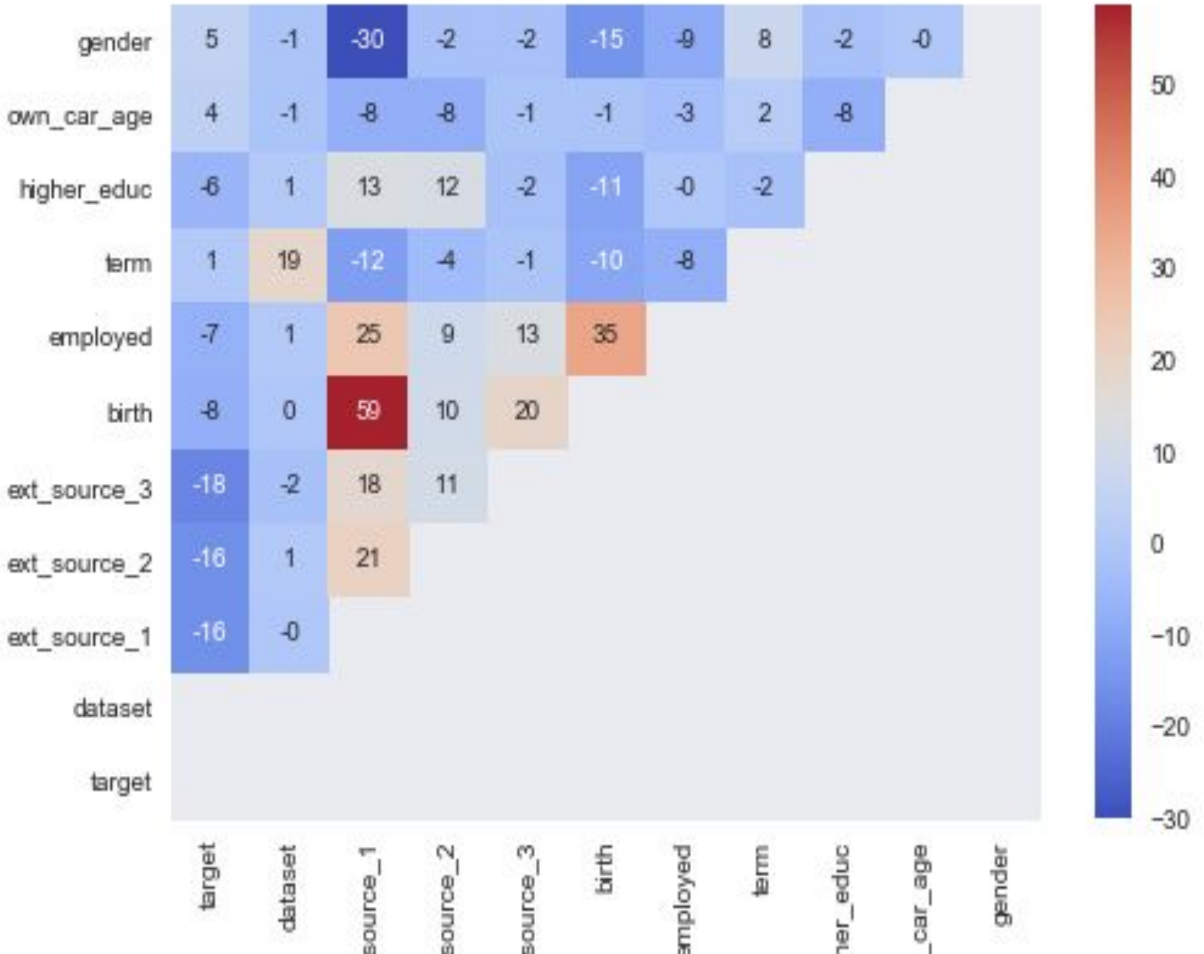


Missing Values



Heat Map

Correlation between variables



Future Work

❑ improve missing value handling
❑ add more data (external)
❑ create more new features
❑ improve trade-off scores

The Team

Johannes Pastorek

~~doesn’t like hyper parameter tuning that turns out to be worthless.~~ “Hyperparameter tuning with a fast computer is great!”

Rüdiger Hass

~~doesn’t like imbalanced data sets.~~

~~“this data set sucks and you know it, Dirk!”~~

“Great project, loved the data set!”