In 2018,

around 8.1%

of International Home Credit's client are experiencing default payments.

$13,846,851,949.5

worth of credit that they need to cover.

Default payments put us, the creditor in a huge disadvantages.

Take action!

Economic Update 2022

Alasan OJK Mint ... Keuangan Agresif Naik ...

MARKET - CNBC Indones ...

19 August 2022 16:53

Jakarta, CNBC In ... ngungkapkan tren restrukturisasi per ... lihan ekonomi.Per Juni ... ta restrukturisasi piutang pembiaya ... 78,8 triliun menjadi Rp 23,7 t ...

OJK terus mendor ... KPN) lebih agresif ke perbankan dan pe ... .

Seperti apa arah strategi ... ng tepat guna dan tepat sasaran? Selengkapnya sima ... oner Otoritas Jasa Keuangan (OJK), Friderica Widyasari Dewi ... pdate ,CNBC Indonesia (Jum'at, 19/08/2022)

HOME CREDIT

Perusahaan pembiayaan

HOME CREDIT

R. Dissa Shafira

Hacktiv8 Talent Fair

Credit Default Prediction

Jadi bisa!

Default Payment?

A missed or multiple missed payments by client.

Detected by:

Clients who exhibit a pattern of delayed payments.

Clients who paid only by the minimum/ interest payments (bunga) considered at higher risk of default.

Clients paying pattern past a certain threshold that can be considered default.

How? do we prevent more loss?

We can surely **handle** them, by:

1. Set appropriate credit **limits**.
2. Encourage payment reminders.
3. Enforce late **payment fees.**
4. Enable automated payments.
5. Continuously monitor and manage credit portfolios.

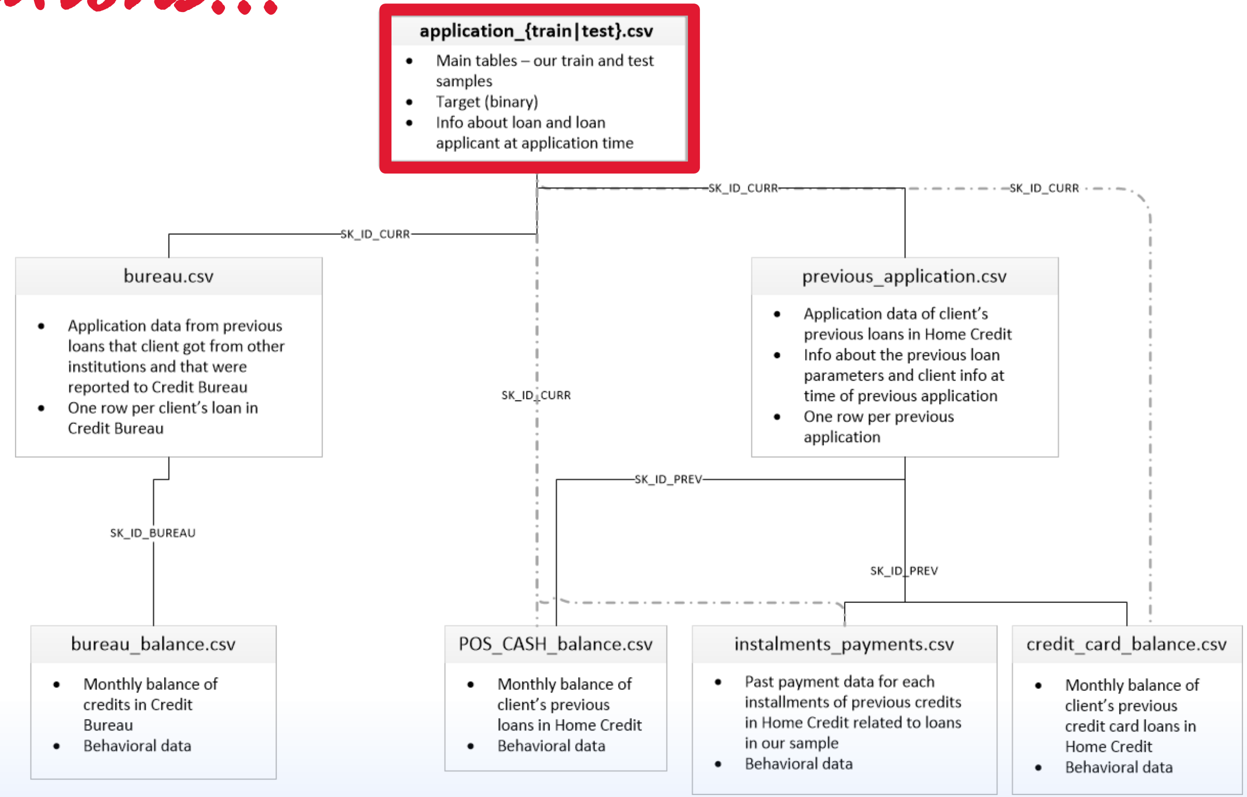
But will those be enough?

No, predict to prevent

- Loan loss provisioning (Cadangan Kerugian Penurunan Nilai) as decentralized approach
- Make informed lending decisions
- Develop effective strategies and meet regulatory obligations.

Look through our dataset

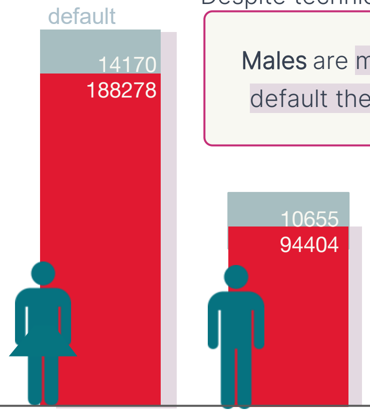
Limitations...

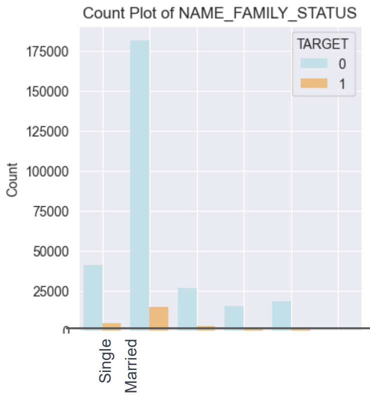


Who..?

Despite technical imbalance,

Males are more likely to default their payment.



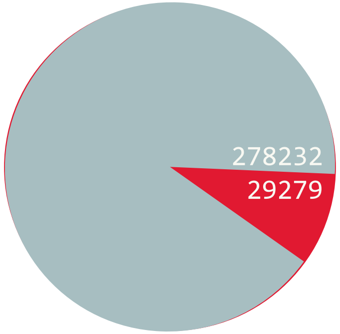


While most clients are married,

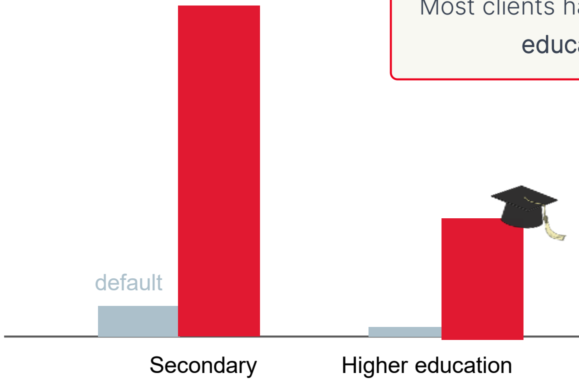
Clients who are **single** have a higher chance to default (0.098 compared 0.007)

Also, most clients don't have children

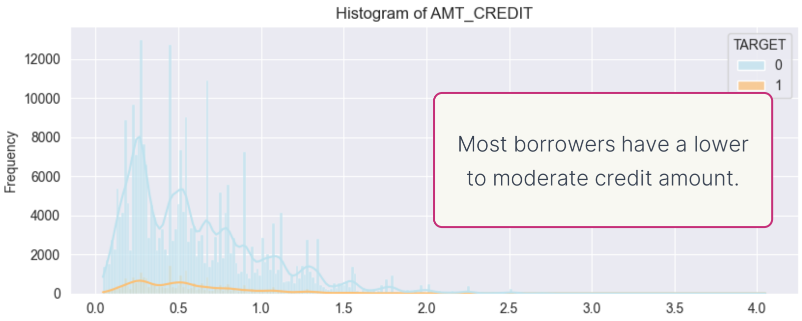
Most clients opt for cash loans rather than revolving loans.



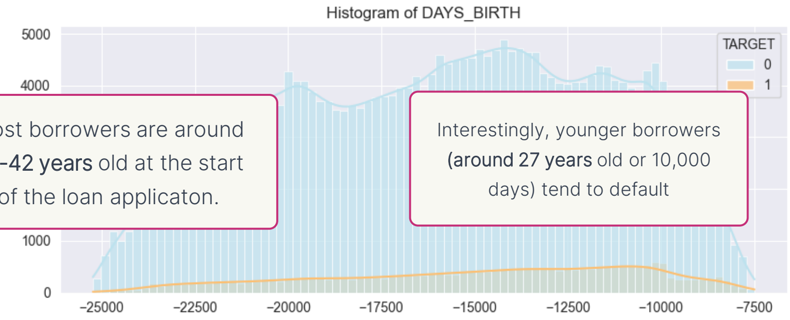
Most clients have **secondary education.**



More about them...

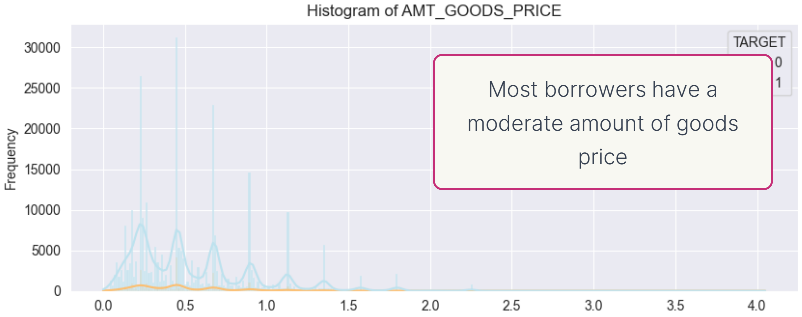


Most borrowers have a lower to moderate credit amount.

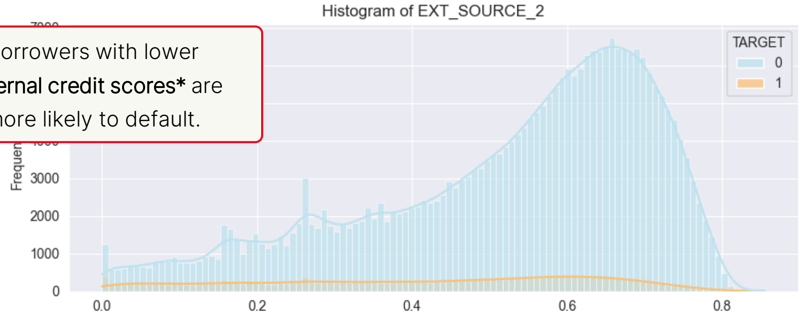


Most borrowers are around **37-42 years** old at the start of the loan applicaton.

Interestingly, younger borrowers **(around 27 years** old or 10,000 days) tend to default



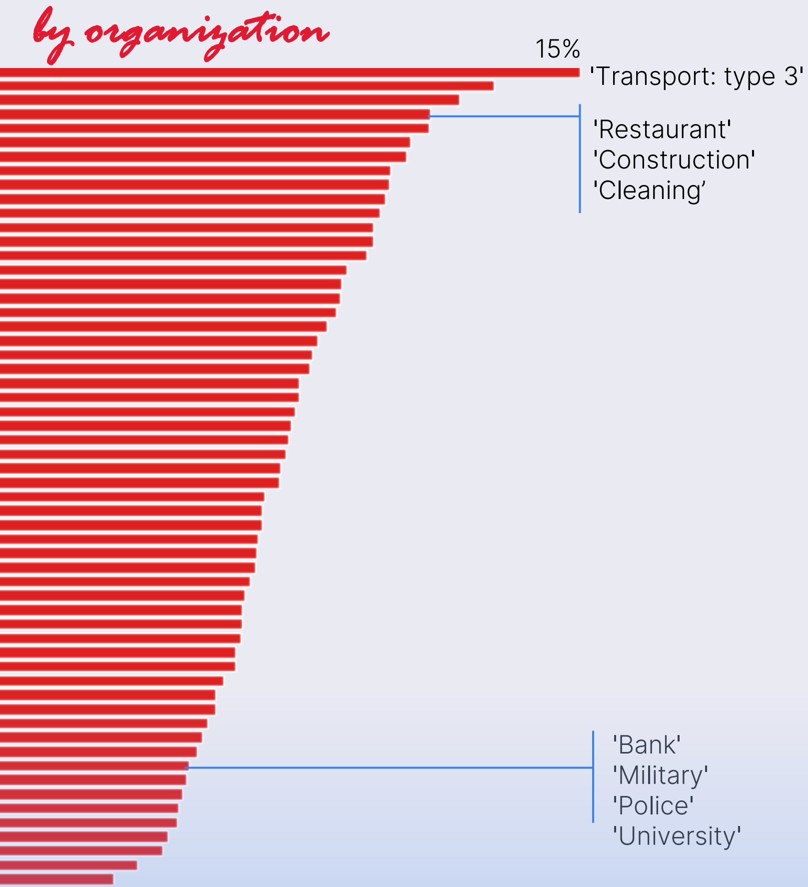
Most borrowers have a moderate amount of goods price

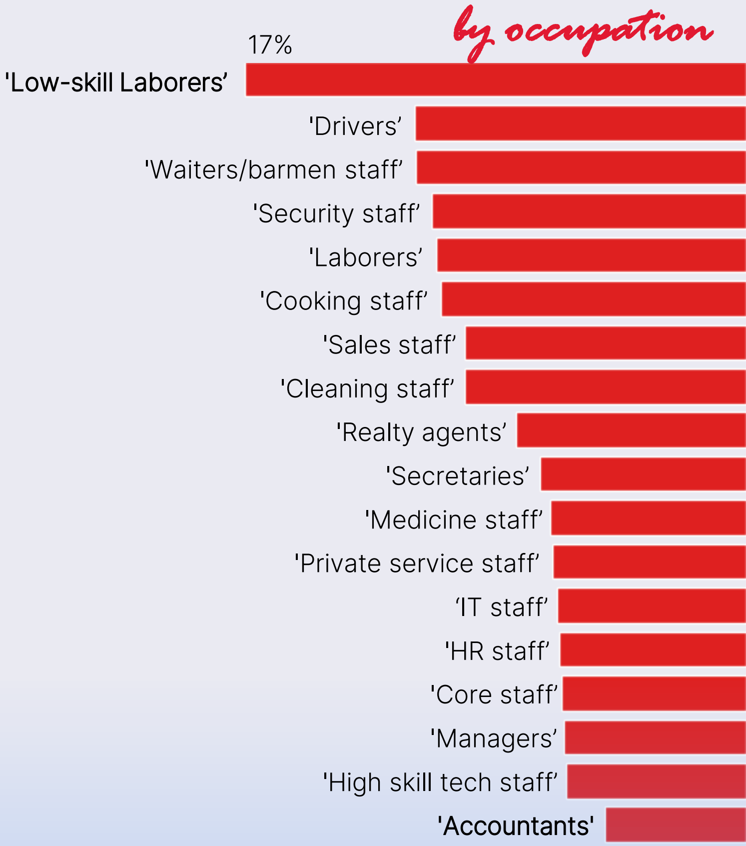


borrowers with lower **external credit scores*** are more likely to default.

*)The US credit score rating indicates higher score for a better credit score. Different from our country (categorical score 1-5)

Default %


by organization
15%
'Transport: type 3'
'Restaurant'
'Construction'
'Cleaning'
'Bank'
'Military'
'Police'
'University'


by occupation
17%
'Low-skill Laborers'
'Drivers'
'Waiters/barmen staff'
'Security staff'
'Laborers'
'Cooking staff'
'Sales staff'
'Cleaning staff'
'Realty agents'
'Secretaries'
'Medicine staff'
'Private service staff'
'IT staff'
'HR staff'
'Core staff'
'Managers'
'High skill tech staff'
'Accountants'

We have seen the big picture of our data...

Now how do we make a prediction out of so many information?

Look through our model

How you can us

list of 'features' need to be filled. ▶ Input clients info ▶ Predict! ▶

Default Risk Prediction

To predict client's default risk, please fill out the form with the client's application information.

The form consists of several sections. The first section requests general information about the client, followed by their income and credit, region information, and finally, any documents they may have.

Client's Demography

Client ID

0

Contract Type

Cash loans

Contract Type

- F
- M

Level of highest education the client achieved

Higher education

Prediction : Li

Thank you for trying out ou
appreciate your participati
insightful and valuable.

Access the model on t
https://huggingface.co
e-credit-default-risk

We have implemente
Ensemble Machine Le
has yielded improved

Cleaning missing values

Drop columns who has more than 40% missing values to avoid false prediction if we impute unrepresentative data.

Feature selection

Split target
Split train, test

Scale

Encode

Ordinal and nominal categories

Missing value handling — Impute mean, median, modus

Outlier handling — IQR

Undersampling — RandomUnder Sampler

Catboost | LogReg | XGBoost | Random forest

Result

Will employing this model enable us to **predict** and **minimize** losses caused by default clients?

unfortunately, *no.*

not yet.

The model is currently in its early stages and still has potential for **improvement**.

Look through the report

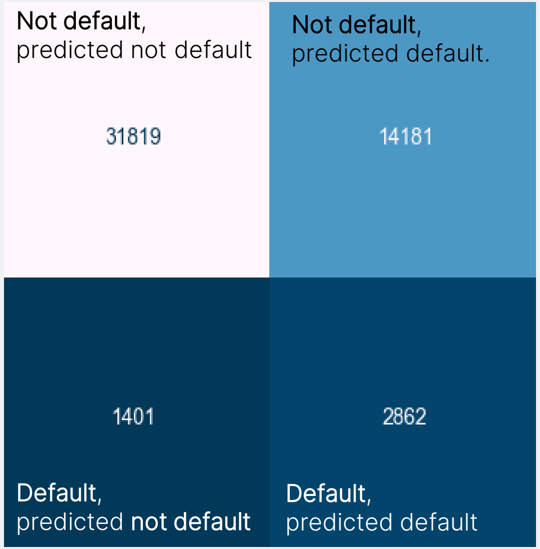
With ROC-AUC score of **74%,**

and accuracy of **68%**

The model tends to predict a non-default clients better than **default**--which are the ones **we're keeping an eye on.**

	precision	recall
Not Default	0.96	0.69
Default	0.17	**0.67**

now



The model can successfully predicted **only 67%** of the defaults.

This also means that the model **missed 33%** **of the actual defaults**, and **incorrectly predicted** them **as non-defaults.**

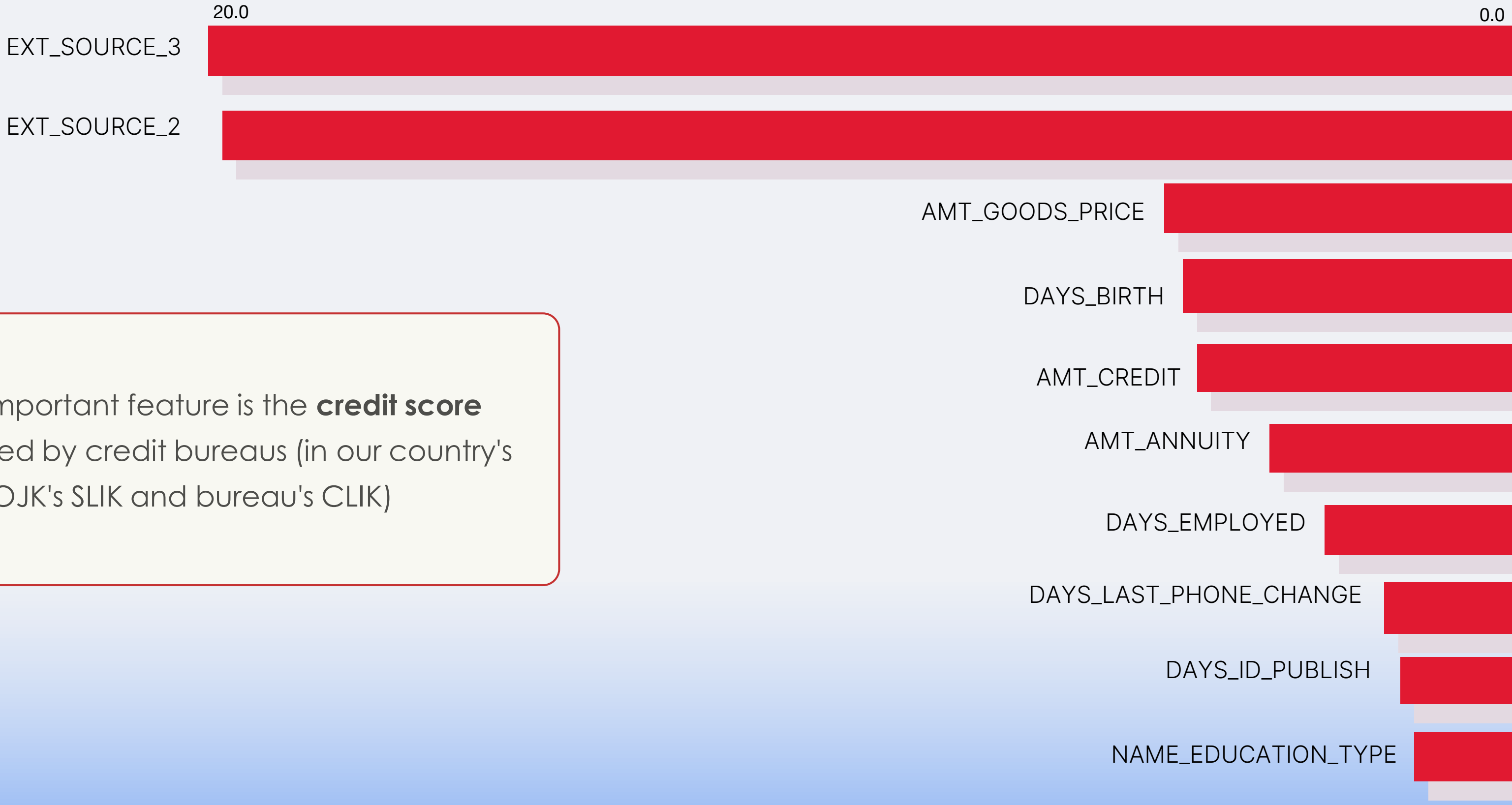
Financial loss

Shortfall in provisioning (kekurangan pencadangan)

We can detect the model's **feature importance**, which may be different from our subjective perception.

evaluate our **dataset weaknesses** and analysis from the model **result.**

Feature Importance



most important feature is the **credit score** provided by credit bureaus (in our country's case, OJK's SLIK and bureau's CLIK)

Result dissection

Most important features:
EXT_SOURCE_3
EXT_SOURCE_2
AMT_GOODS_PRICE
DAYS_BIRTH
AMT_CREDIT

Not default, predicted not default 31819	**Not default**, predicted default. 14181
1401 **Default**, predicted **not default**	2862 **Default**, predicted default

borrowers with significantly higher 'AMT_CREDIT' compared to 'AMT_GOODS_PRICE' tend to have a tendency **to get predicted as default.**

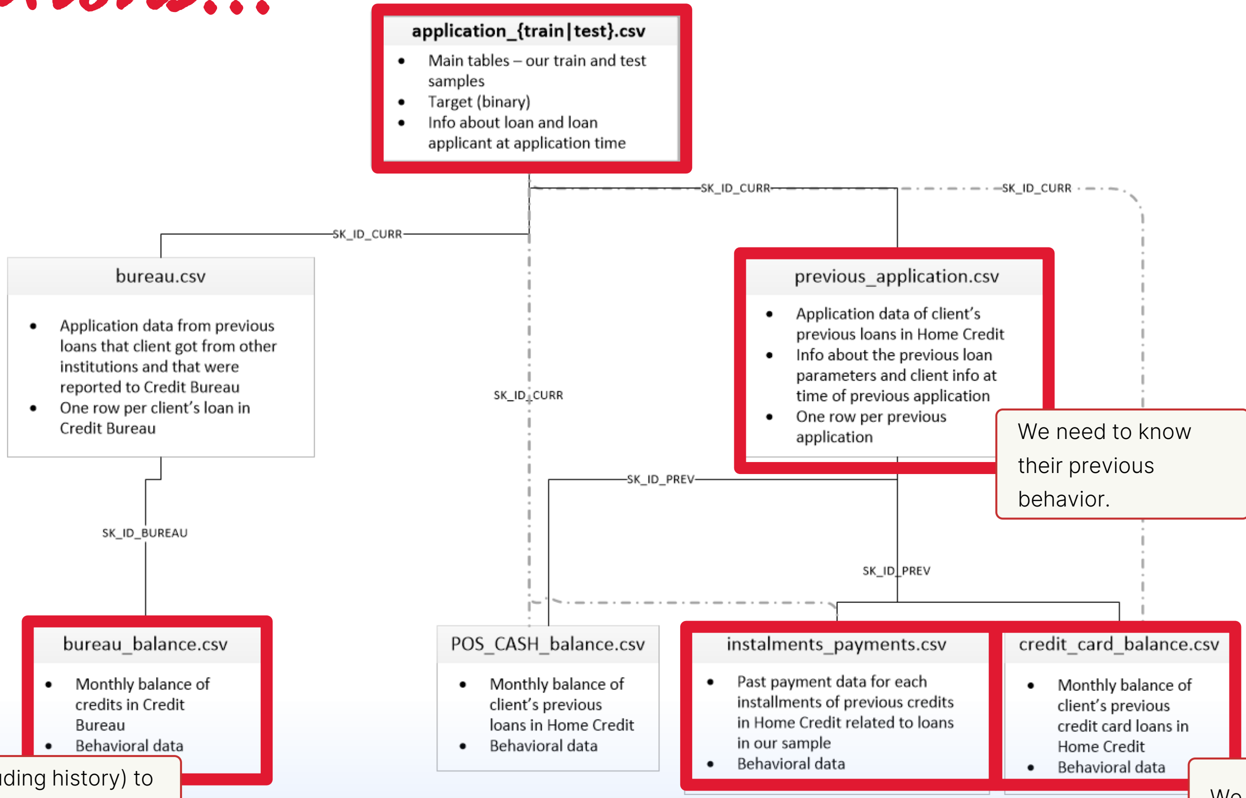
Higher score on external 2, lower on external 3 (imbalance)
Tend to get **predicted as non-default**

Two or more **strong but conflicting** features can lead to calculations that do not align with the actual result.

If we're still going to use the model, the next time we're going to predict a new set of data, consider the correlation of these strong features to get an idea of the error.

Limitations...

Most important features:
EXT_SOURCE_3
EXT_SOURCE_2
AMT_GOODS_PRICE
DAYS_BIRTH
AMT_CREDIT



We need to know their previous behavior.

access data (including history) to determine **credit score** (our most important features yet)

We need to know their credit balance.

Prioritize

For suggestions, we consider both the insights gained from our data analysis and the importance of features identified by our model.

Credit History

Even though debtor **may appear safe in terms of portfolio and assets**, there is still a **potential for them to have a tendency** to be unable to meet debt payment obligations if their score is low (they may be able to, but they just don’t repay)

Goods Price

Credit Amount

Carefully assess the financial capability of debtors before approving a loan.

Age

It is known that **younger** debtors has a higher tendency to default. This may also be influenced by social factors such as economic stability, as well as psychological factors related to an individual's maturity.

Missing 33% of default prediction is *Bad.*

We can miss out on opportunities and erode **clients' confidence** by incorrectly predicting non-defaults.

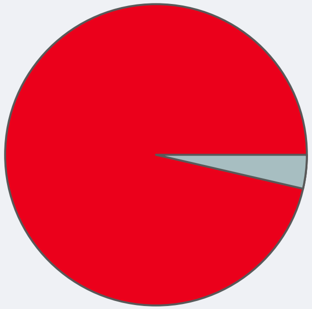
But even worse, this can lead to a **huge financial loss** if we keep on lending those <u>who fail to pay</u>. A **damaged reputation** is also destructive in the long term.

But for now,

We can evaluate the model result to gain insight, as well as attempting to improve the model by:

- ❏ Use more tables and features
- ❏ Reduce dimensionality
- ❏ Do parameter tuning
- ❏ Pre-processing stage; balancing data, and applying other transformations.

In conclusion..



We have noticed a high amount of clients with default payments.

▶ To address this issue, we tried to develop and introduce a <u>model</u> to predict and assess clients' credit behavior.

▶ Despite our efforts to create the model,

Our model is currently in a **development** and still got a lot of room for **improvement**.

If we need the model right now, consider the false tendencies present in **certain types of data**. (Strong but conflicting features)

▶ we can also **enhance our monitoring** and credit **management** by implementing **targeted handling** based on previous **data exploration**.

Thank you!