Client Behavior on Credit Repayments

at a bank in the Eastern Europe

Aliaksandr Nekrashevich
Winter 2022

Smith School of Business | Queen's University

The Story

- Initial version was a project at Scotiabank in Summer 2018.
- Due to the nature of data and NDA limitations, this cannot be disclosed.
- However, after a while, a dataset with similar structure from another bank has appeared openly at a hackathon competition in my home country.
- Therefore, similar methodology can be safely applied and demonstrated here.
- It was further refined when the presenter was studying MGMT-962 at Smith School of Business in Winter 2021.
- Note that this is still a toy example, i.e. it was implemented much faster than the industrial one and with many questionable simplifications. However, it can still be useful as a proof of concept.

Problem Formulation

- We focus on the business line responsible for car credits.
- Although credit scoring is working, clients behave differently.
- The natural question is then how to use this variability in client behavior to make the bank more profitable?
- Financial atmosphere and economic ecology are very risky in Belarus, which adds initial challenges to the problem.
- Ethical concerns: for personal loans, any credit direction except mortgages is questionable.
- The problem and dataset were formulated at Datathon 2019. It is a competition in Data Science, typically held at Imaguru Startup Hub in Minsk, Belarus.

Image Source: https://www.flickr.com/photos/juanelo242a/11280390513
Available on the Internet under the Creative Commons license.

Industrial Information

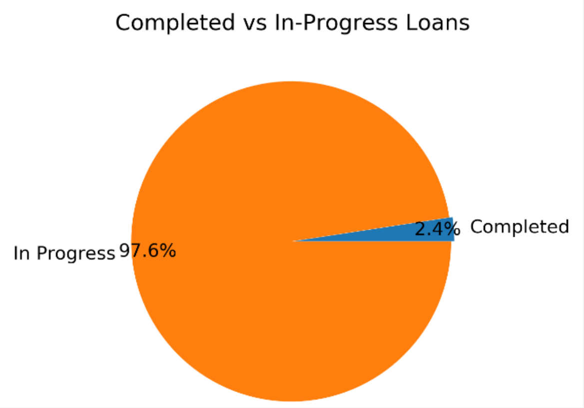
- Belarusky Narodny Bank (BNB-bank) – a small bank located in the Eastern Europe, Republic of Belarus.
- Key business directions: small and medium business, and personal finance.
- Financial investment of the Republic of Georgia in the Republic of Belarus.
- Main credit directions: car loans (9 credit lines) and mortgages (4 credit lines).
- URL: https://www.bnb.by

Image Source: https://www.flickr.com/photos/27665395@N05/45604387065
Available on the Internet under the Creative Commons license.

Dataset Description

DYNAMIC information contains the repayments information of the main loan body (without interest rate).

- **CONTRACT_ID** – contract identifier
- **PERIOD_ID** -- the month after the contract was issued. When PERIOD_ID = 1, it is the first month after the loan was provided
- **REPAYMENT_SCHEDULED** – amount of payment at the current period according to the contract
- **REPAYMENT_ACTUAL** -- factual payment by the client in the current period. When NULL, it means this period has not yet arrived



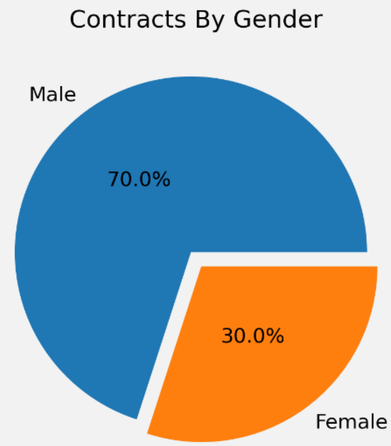
Amount of completed contracts: 91
Amount of in-progress contracts: 3701
Total amount of contracts: 3792
Maximal length of a completed contract: 18

	CONTRACT_ID	PERIOD_ID	REPAYMENT_SCHEDULED	REPAYMENT_ACTUAL
0	17228104	1	19.76	19.76
1	17228104	2	19.76	19.76
2	17228104	3	19.76	19.76
3	17228104	4	19.76	19.76
4	17228104	5	19.76	172.76
5	17228104	6	19.76	268.41

Dataset Description

STATIC information contains initial information about the client and the contract.

- **LOAN_TO_INCOME** – the ratio of the credit amount to the monthly client revenue
- **PAYMENT_TO_INCOME** – the ratio of monthly client payment to the monthly client revenue
- **DOWNPAYMENT** -- ratio of client self-participation in the car purchase. DOWNPAYMENT = 1 - (CONTRACT_SUM / cost of the automobile)
- **GRACE_PERIOD** – the length of grace period in months. At the beginning of the contract during this period, the interest rate is lower than the regular one afterwards. If GRACE_PERIOD = 0, there is no period with discounted interest rate
- **RATE_CHANGE_AFTER_GRACE** – how the interest rate changes after the grace period is over



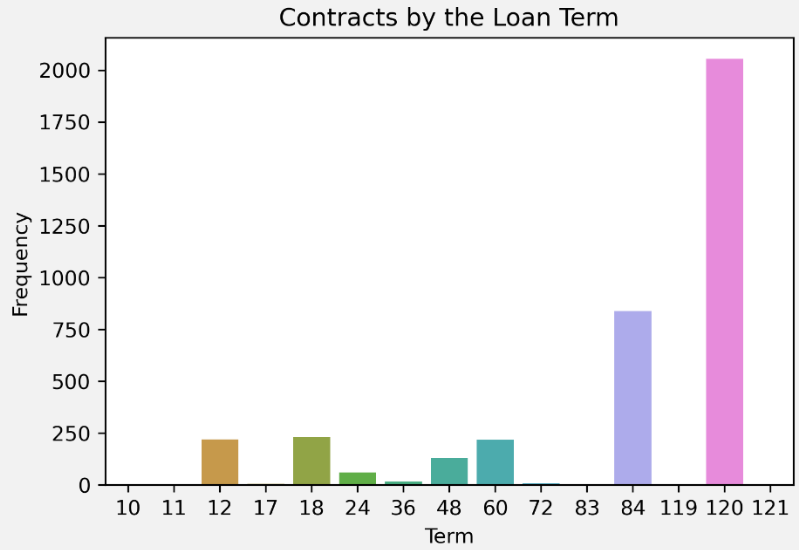
Male Gender: 2655
Female Gender: 1137

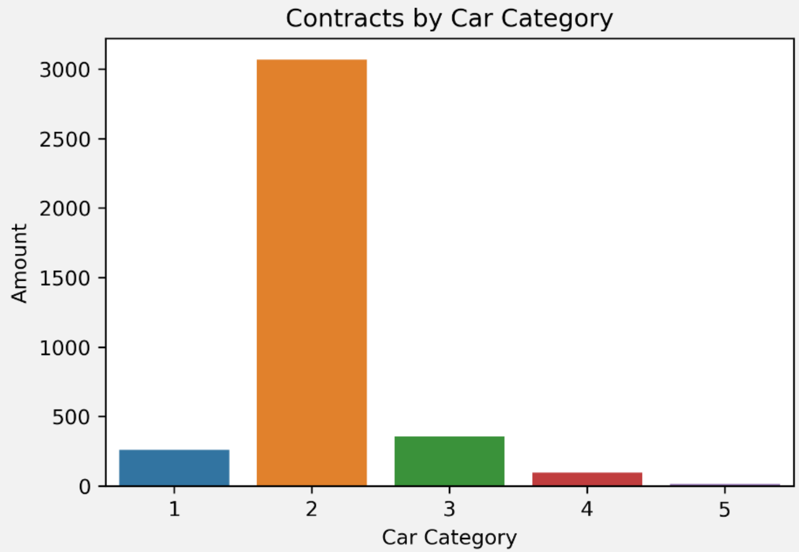
	CONTRACT_ID	CLIENT_ID	TERM	CONTRACT_SUM	GENDER	AGE	LOAN_TO_INCOME	PAYMENT_TO_INCOME	DOWNPAYMENT	CAR_CATEGORY	GRACE_PERIOD	RATE_CHANGE_AFTER_GRACE
0	17228104	251471	60	1185.75	M	32	10	0.22	0.4	2	6	13
1	17237409	251501	18	512.47	M	30	7	0.37	0.7	2	18	15
2	17276280	251669	60	1529.24	M	36	10	0.23	0.1	2	6	13
3	17282809	251684	60	906.53	M	27	6	0.15	0.3	1	6	13
4	17283247	251692	60	1593.5	F	50	18	0.42	0.1	2	6	13
5	17294333	251746	60	1442.03	M	26	16	0.36	0.2	2	6	13
6	17306398	251779	60	971.55	M	51	13	0.29	0.4	2	6	13
7	17320168	251852	60	1748.79	F	54	27	0.62	0.1	2	6	13

Dataset Description

Static information contains initial information about the client and the contract.

- **CONTRACT_ID** – contract identifier. The contract is a loan for a car purchase.
- **CLIENT_ID** -- client identifier
- **CONTRACT_SUM** – credit amount (in a hidden currency)
- **GENDER** – the gender of the client (M - Male, F - Female)
- **CAR_CATEGORY** – category of the purchased automobile. There are five different categories. The minimal budget category is 1, the maximal premium is 5.
- **TERM** -- amount of months for credit repayment
- **AGE** – age of the client at the time the loan was issued





Framework Overview

Client Behavior

Payment Time Series Clustering

1. Cluster time series with at least 12 months after the issue date.

Only small ratio of contracts have been issued at least a year before the dataset was provided. Majority of them are not over.

2. Fill time series based on dataset information.

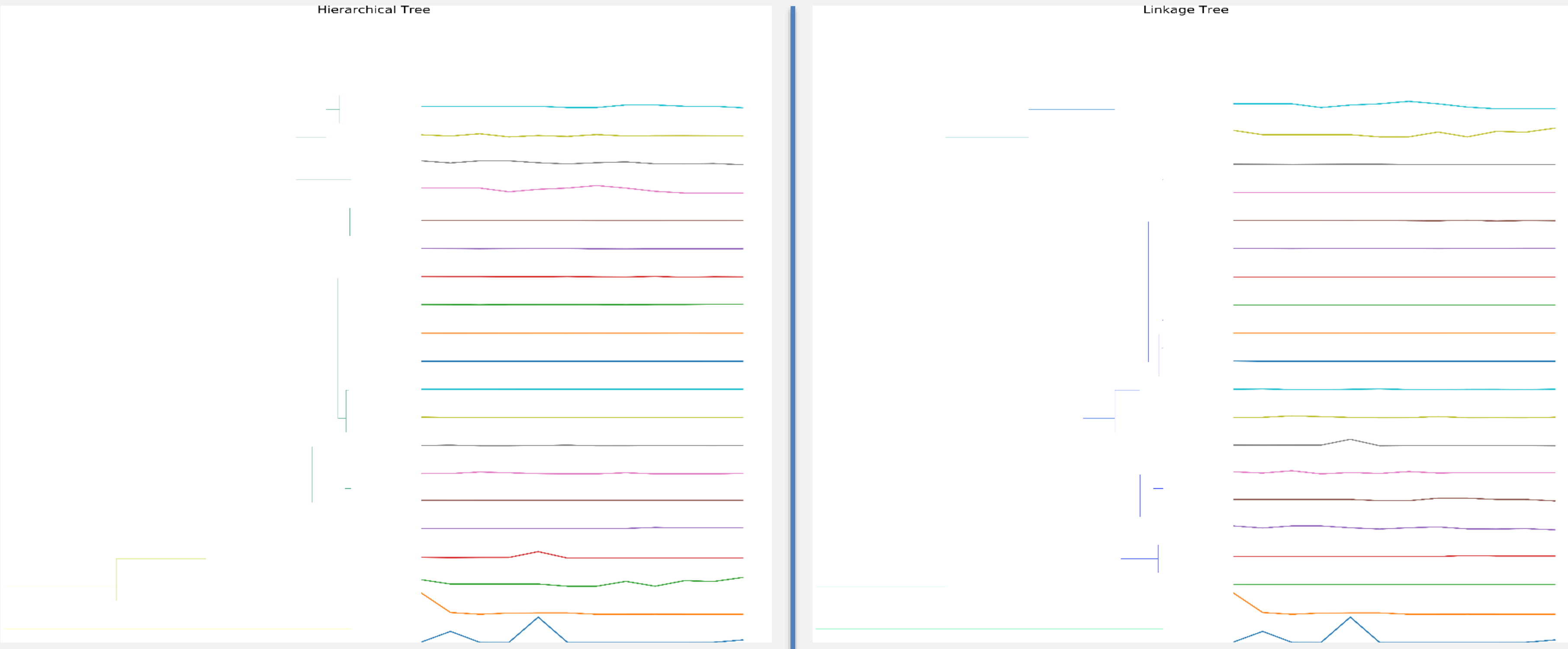
3. Assign remaining time series to a cluster after completion

Client Reliability

A client is **reliable** if the actual cumulative repayment sum is at least as big as the cumulative scheduled repayment sum at any period while the contract is not completed.

Payment Time Series Clustering

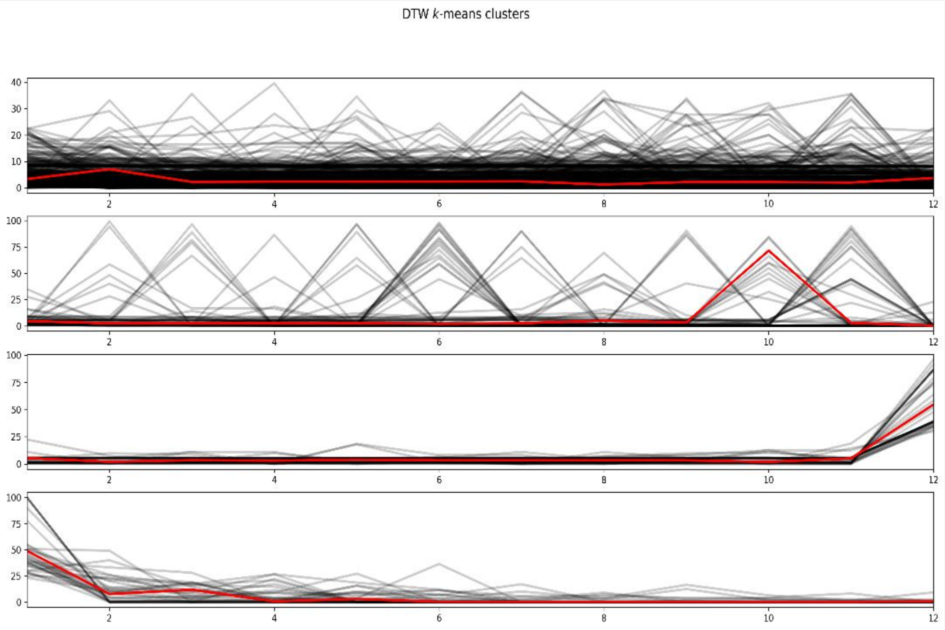
- **<u>Step 1</u>:** Guess the number of clusters. Hierarchical clustering on a small subset.
- **Assumption:** repayments are converted to relative (divided by contract sum and multiplied by 100%).

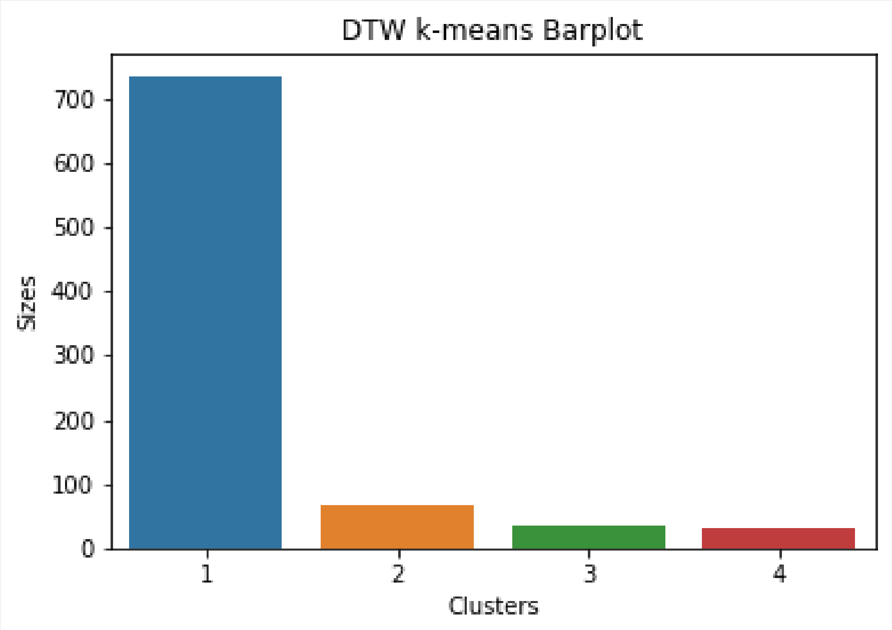


Visualization allows to make a guess a number of clusters as 3 to 5. The following behavior patters are visual: flat repayments, or single spike, or big initially then flat, or flat and big in the end.

Payment Time Series Clustering

- **<u>Step 2</u>:** Trying methods with pre-defined cluster amount. They are k-Medoids, k-Shape, DTW k-Means, and Kernel k-Means.





The most reasonable result is obtained by **DTW k-Means**. The following four patterns look the most realistic:

1. Flat repayment
2. Spiked repayment (flat, then fixing almost all debt, and then flatly concluding)
3. Flat and high payment at the end.
4. High payment initially and flat repayment until the end.

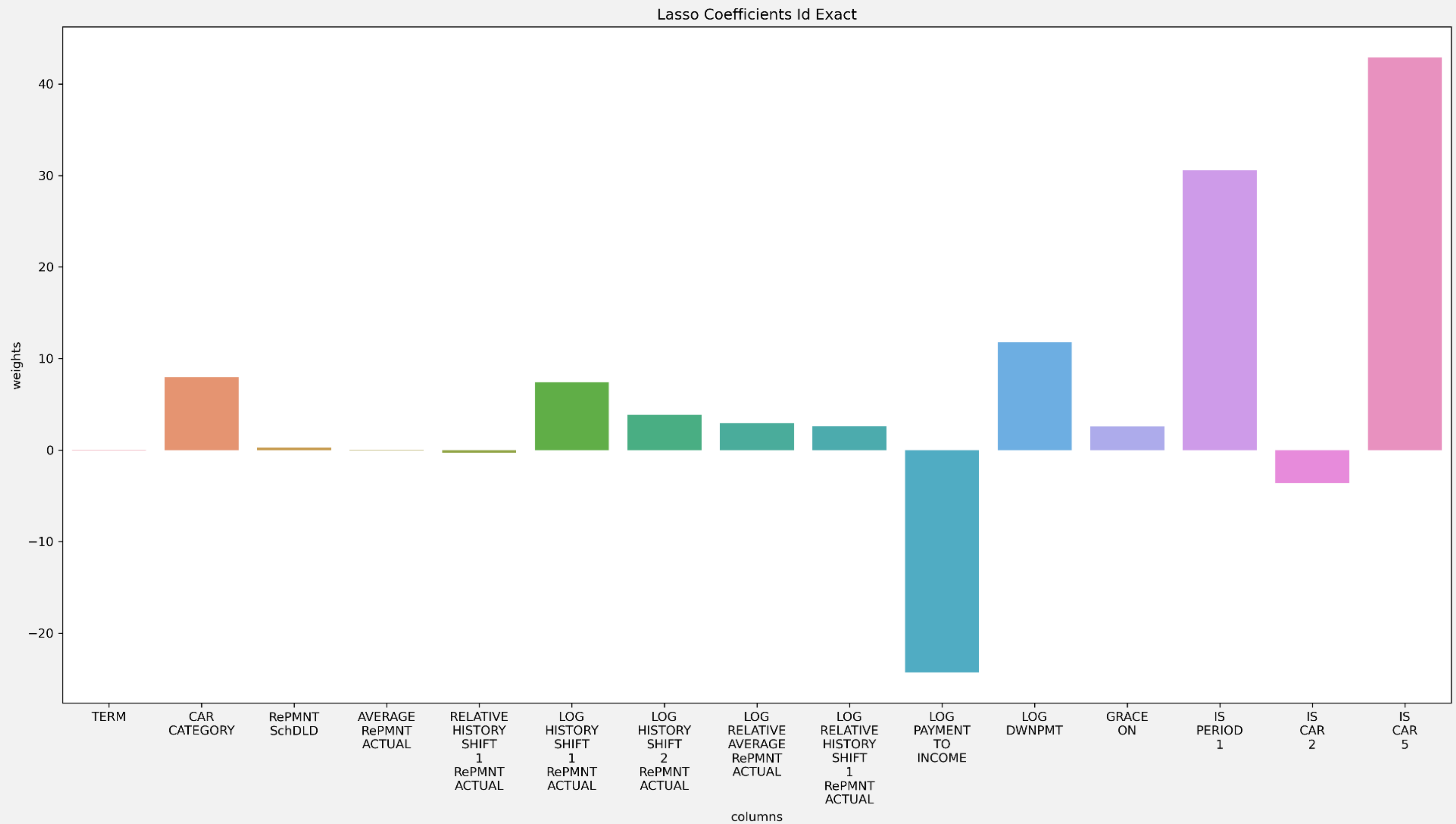
Next Month Repayment Regression

- To cluster all time series, we first need to complete them. Standard time series techniques may not work, because of context, behavior structure and limited information.
- It can be reasonable to predict repayment, log-repayment, relative repayment, and log-relative repayment. But let's describe Feature Engineering first.

Features	Description
CUMSUM_*	Cumulative repayments (actual and scheduled).
AVERAGE_*	Average repayment until this period (actual and scheduled).
RELATIVE_*, RELATIVE_CUMSUM_*	Relative repayments, periodic, average and cumulative, actual and scheduled.
LOG_*, LOG_RELATIVE_*, LOG_RELATIVE_CUMSUM_*, etc.	Log-transformation of features (cumulative, average, and relative), np.log1p.
HISTORY_*, HISTORY_LOG_*, HISTORY_LOG_RELATIVE_*, etc.	Information from two last periods, replaced by scheduled amounts if it is one of the first two periods.
IS_GRACE_CONSTANT_*	If payments during grace were constant.
RATIO_*	If LOAN_TO_INCOME is above some threshold (5, 10, 20, 30, 40).
IS_CAR_*	Indicator dummy on car category.
IS_PERIOD_*	Dummy for first periods 1-2-3.
IS_GRACE_ON_*	Indicator if grace is on.

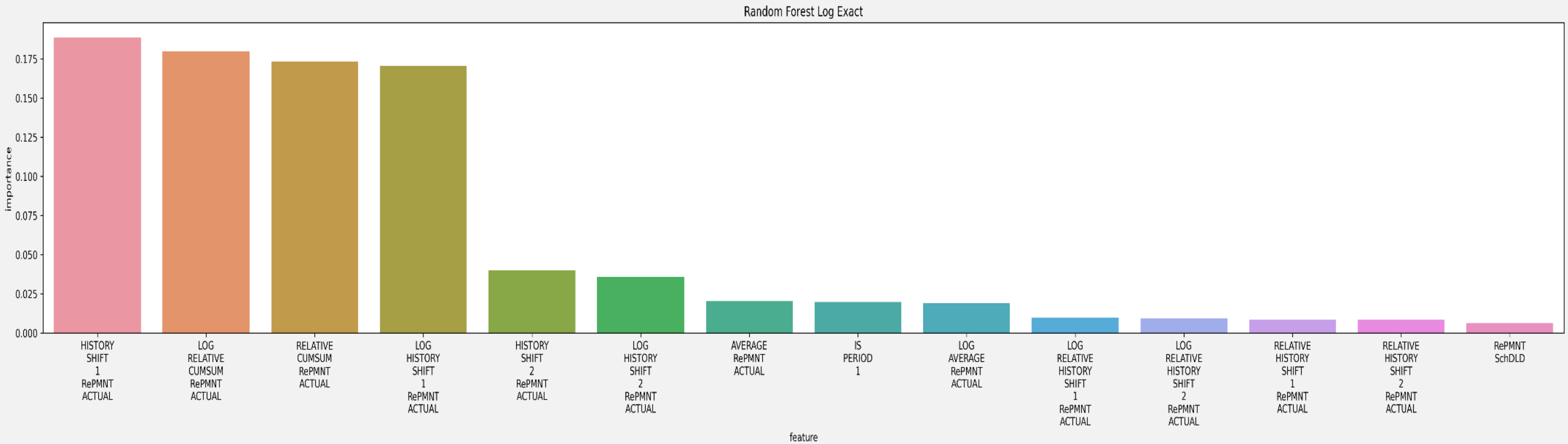
Next Month Repayment Regression

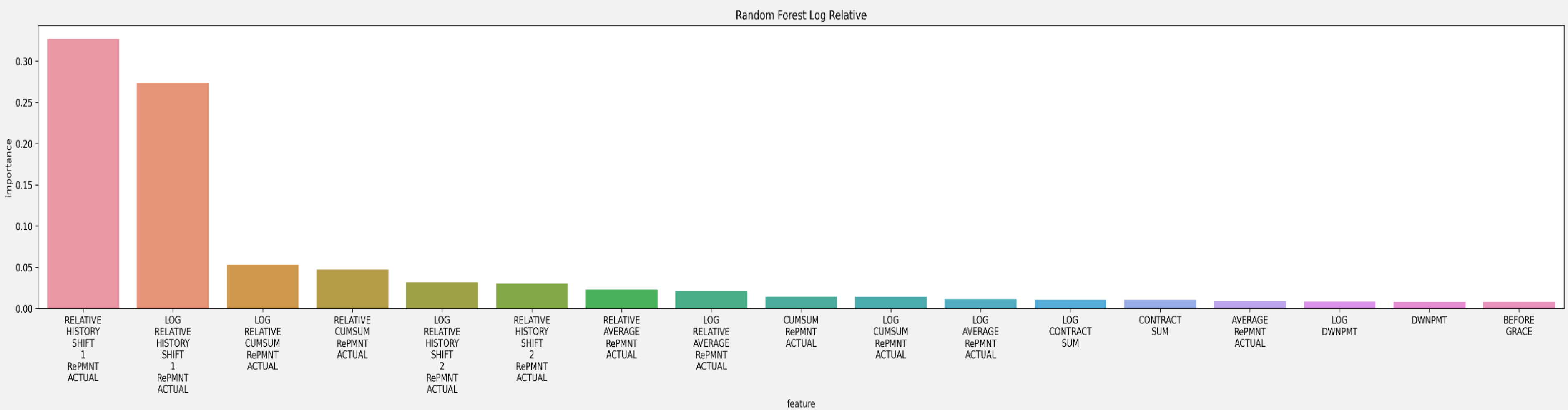
- As a side effect, we can estimate how features impact next month repayments.



Next Month Repayment Regression

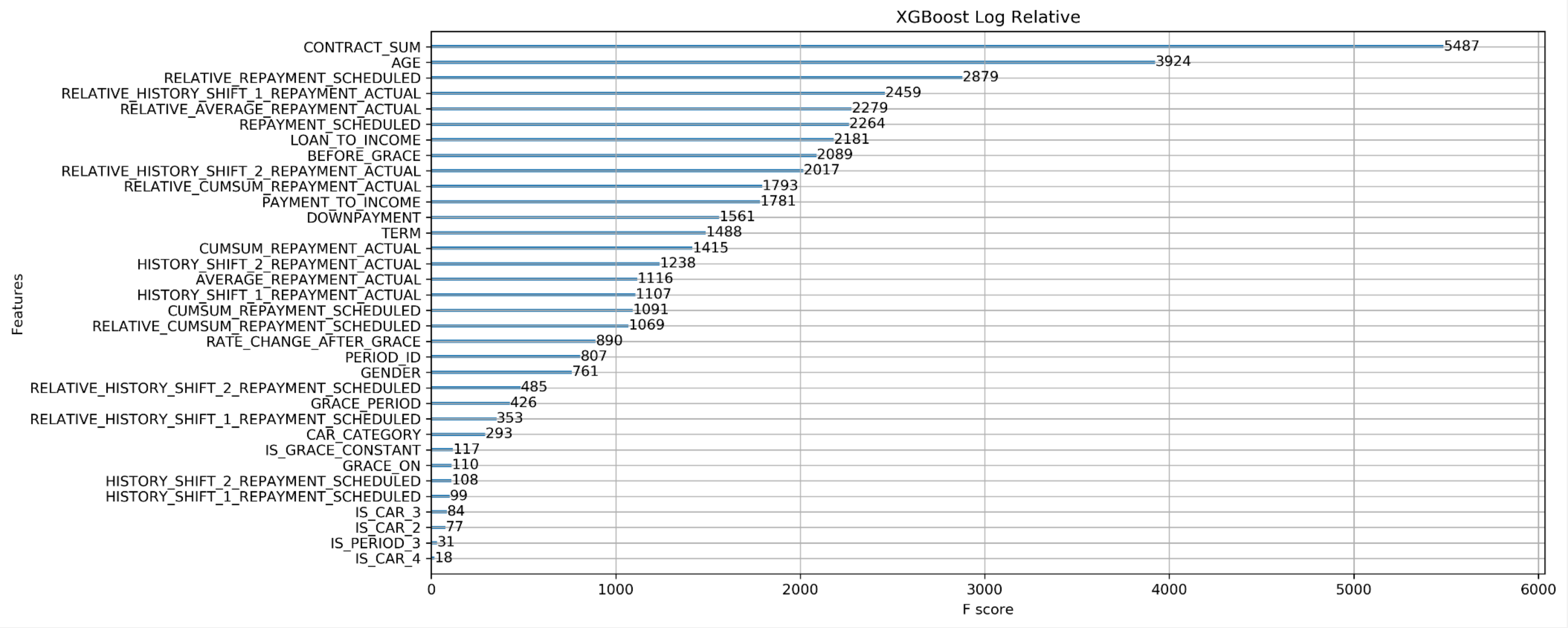
- As a side effect, we can estimate how features impact next month repayments.





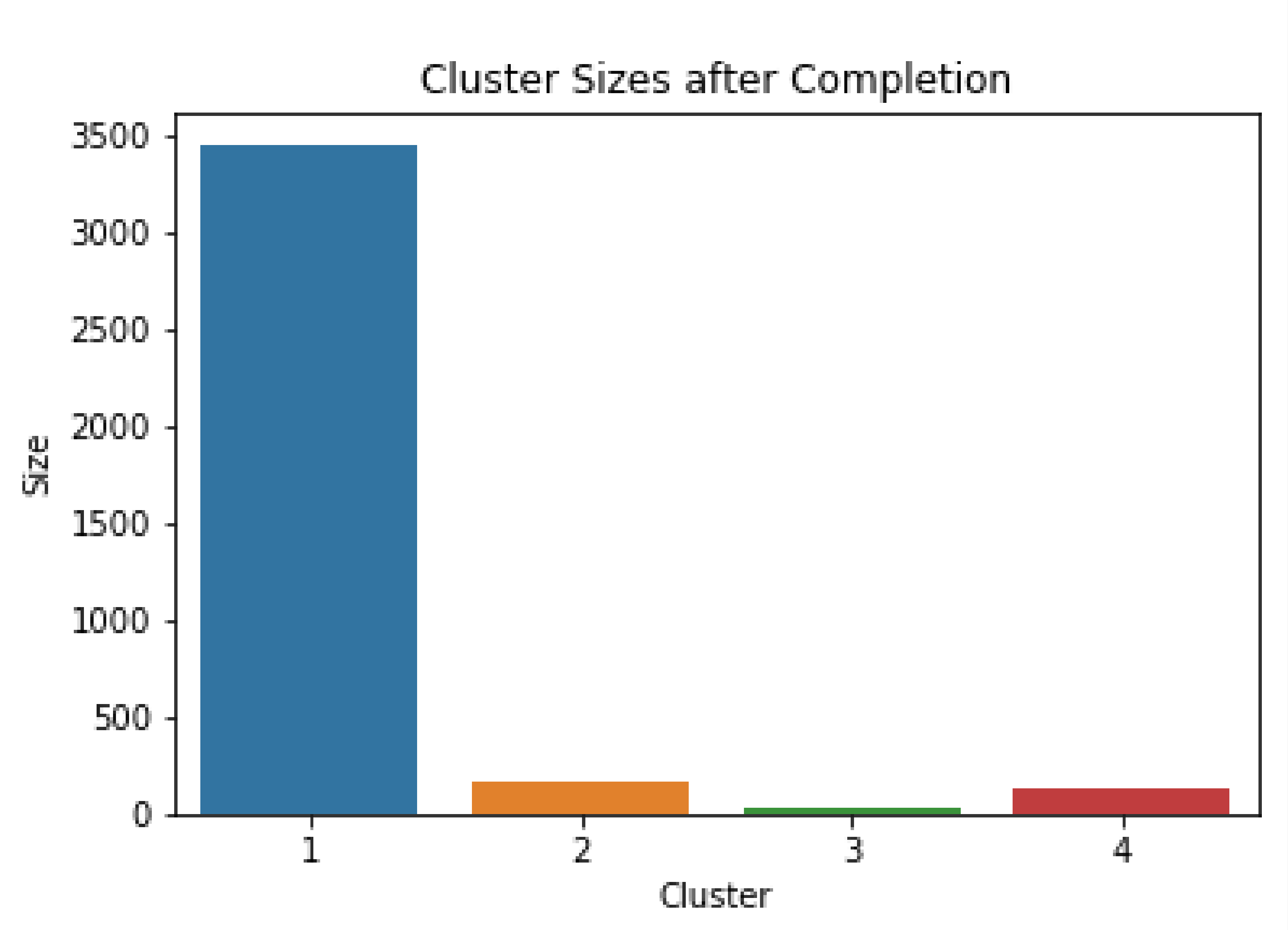
Next Month Repayment Regression

- As a side effect, we can estimate how features impact next month repayments.



XGBoost Feature Importances

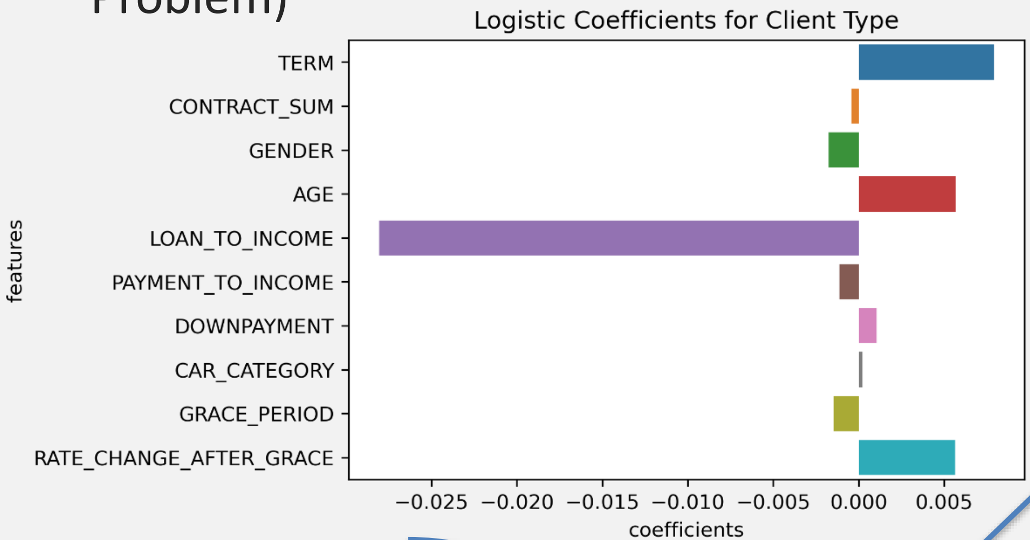
Clustering After Completion



Majority of clients are still very much linked to the schedule. However, we have found more clients with less standard behavior patterns.

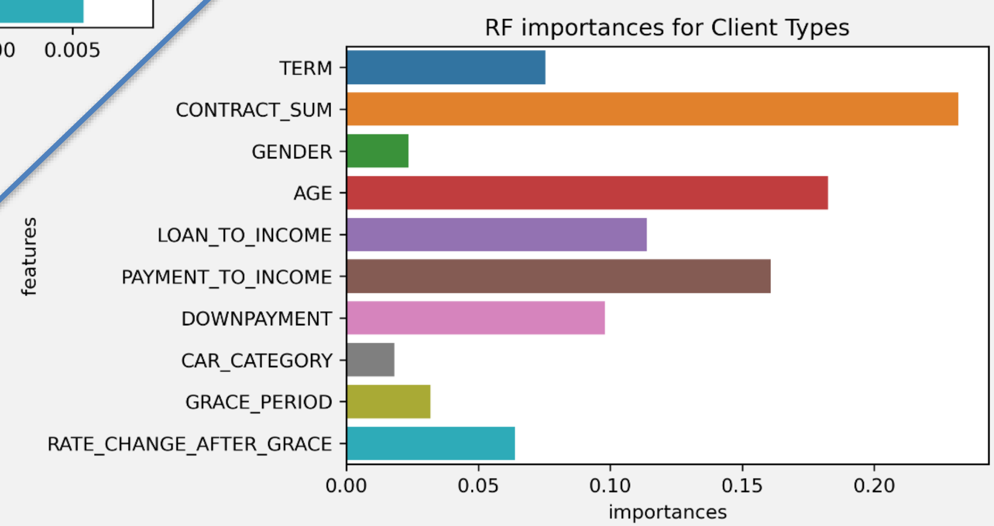
Client Reliability

- 3348 reliable, 444 unreliable in the training set (Imbalanced Classification Problem)



Reliable money can-be reinvested and improve Funds Transfer Pricing.

A client is **reliable** if the actual cumulative repayment sum is at least as big as the cumulative scheduled repayment sum at any period while the contract is not completed.



Algorithm	accuracy	precision	recall	f1-score	reliable precision	reliable recall	unreliable precision	unreliable recall
Logistic Regression	0.583388	0.81586	0.583388	0.655304	0.906976744	0.585585586	0.159817352	0.159817352
Random Forest	0.8029	0.808723	0.8029	0.805762	0.891584534	0.882882883	0.212121212	0.212121212

Summary and Acknowledgements

Results:

- Client repayment clustering into four behavior patterns.
- Interpretation for repayment drivers at the next period.
- Client reliability forecasting and interpretation.

Thank you for your attention!
Questions? Remarks, Suggestions, Comments.

Image Source: https://www.flickr.com/photos/juanelo242a/13632722623
Available on the Internet under the Creative Commons license.

Acknowledgement:

Author is grateful to some of his Datathon 2019 teammates which behaved reasonably during the competition. More precisely, it was only Artsem Shekh (https://www.linkedin.com/in/prestal/).