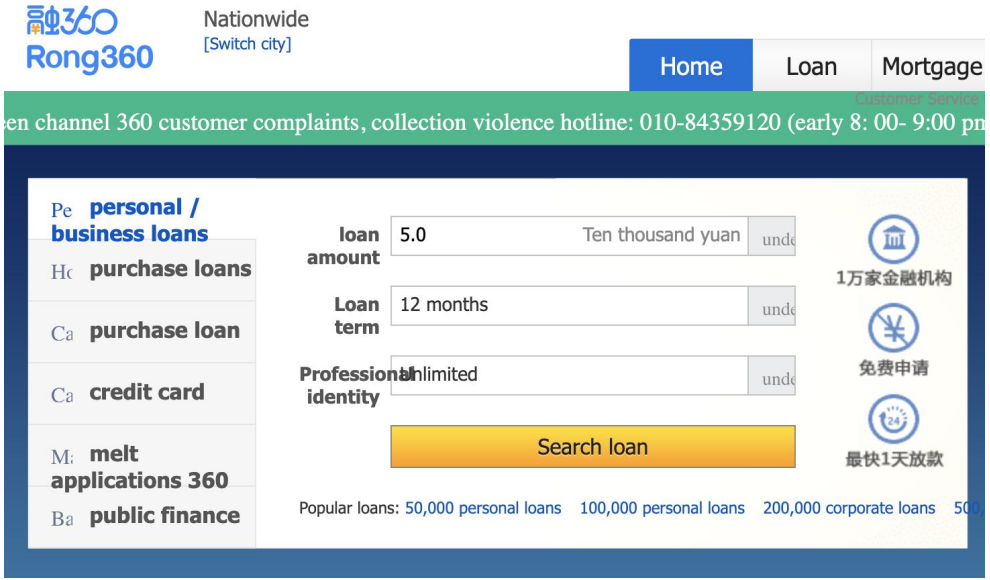
Are We Wrong360?

A User and Products Analysis by P3

Lilliana Nishihira, Annette Paciorek, Jennifer Ruddock PhD

What is Rong360?



- Chinese loan platform
- Streamlines process for loan applicants
- Provides access to wider customer base for banks

http://www.rong360.com/

Insights:

- Loan products can be segmented into describable levels: Gold, Silver and Bronze

- User platform activity data on Rong360 may reduce applicant risk and lead to higher approval rates.

Loan Levels

GOLD

- Largest Loans
- Most Information required
- Longest decision turnaround

SILVER

- Medium Sized Loans
- Lower Application Rate

BRONZE

- Smallest Loan
- Highest Application Rate
- Least Amount of Information Required for Application

About the Data

- **Loan product data**: All of the loans available on Rong360.com (48, 433) and 33 dimensions per data point, including the city, loan term, and which personal information is needed in order to apply

- **User application data**: Applications submitted by users, including users' personal information as well as loan information, like the type of loan and amount sought (1,506,097 x 89)

- **User web data**: Records of users' page views on Rong360 website (238958 x 12)

- **Order data**: Basic record of loan applications and approval result (143152, 6)

Data Cleaning

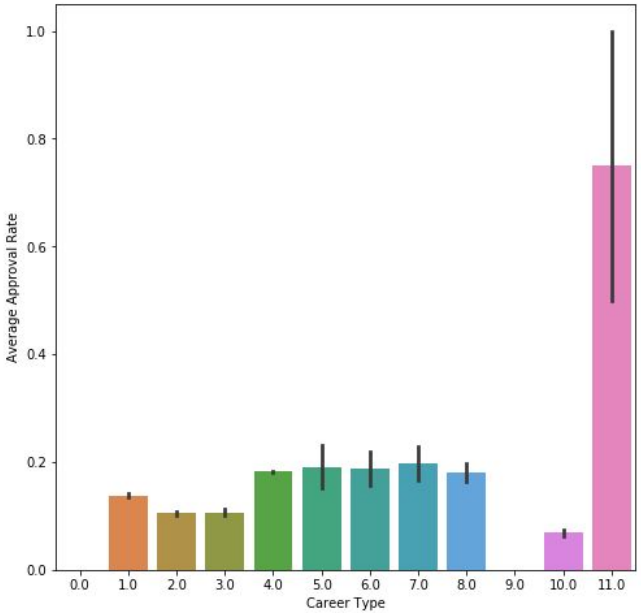
Pros:

- Relatively small data sets
- Segmented data allows for distribution of labor
- Several features were nearly uniform or mostly NA - allowed us to reduce dimensions

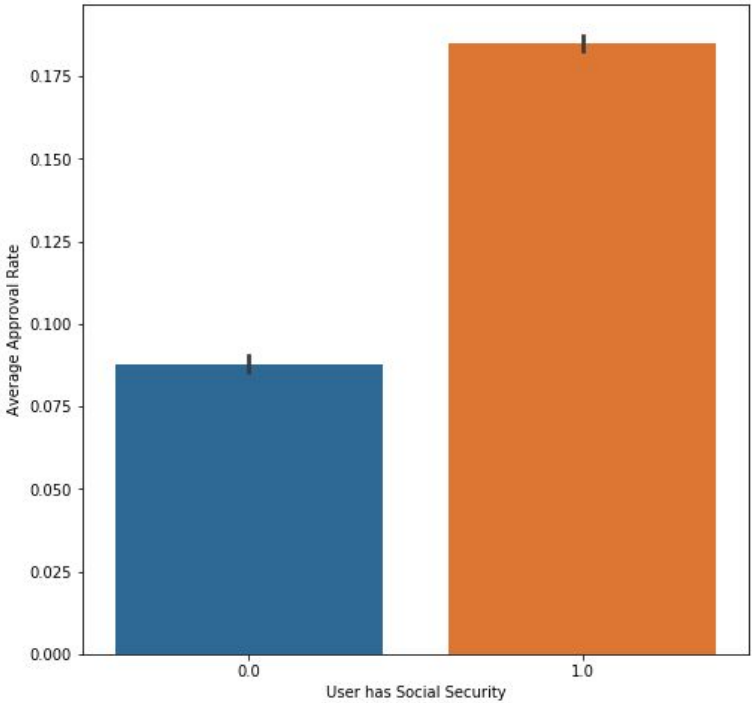
Cons:

- Information was encoded - meaning of categories was obfuscated and harder to orient ourselves in the data
- How to impute missing values
- Inconsistencies across various banks with respect which information they request and how it’s recorded

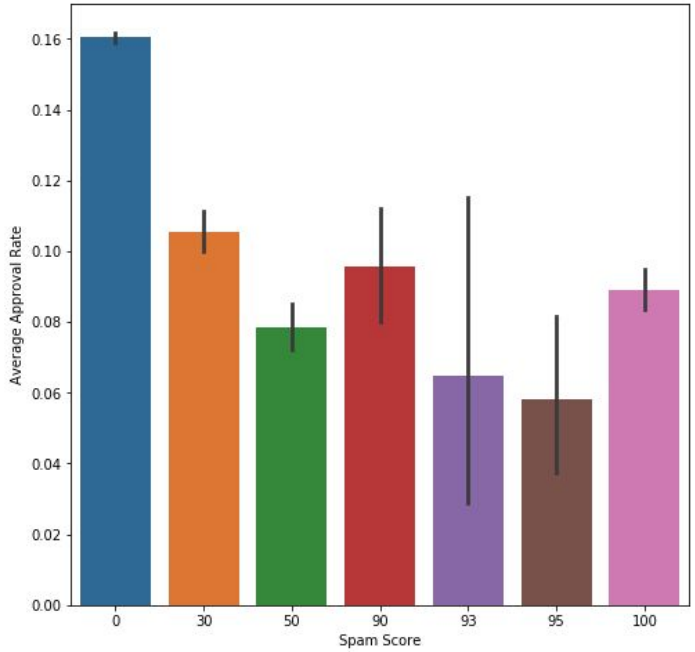
Data Analysis - Who is more likely to approved?



Users' careers matter



Whether users pay social security matters

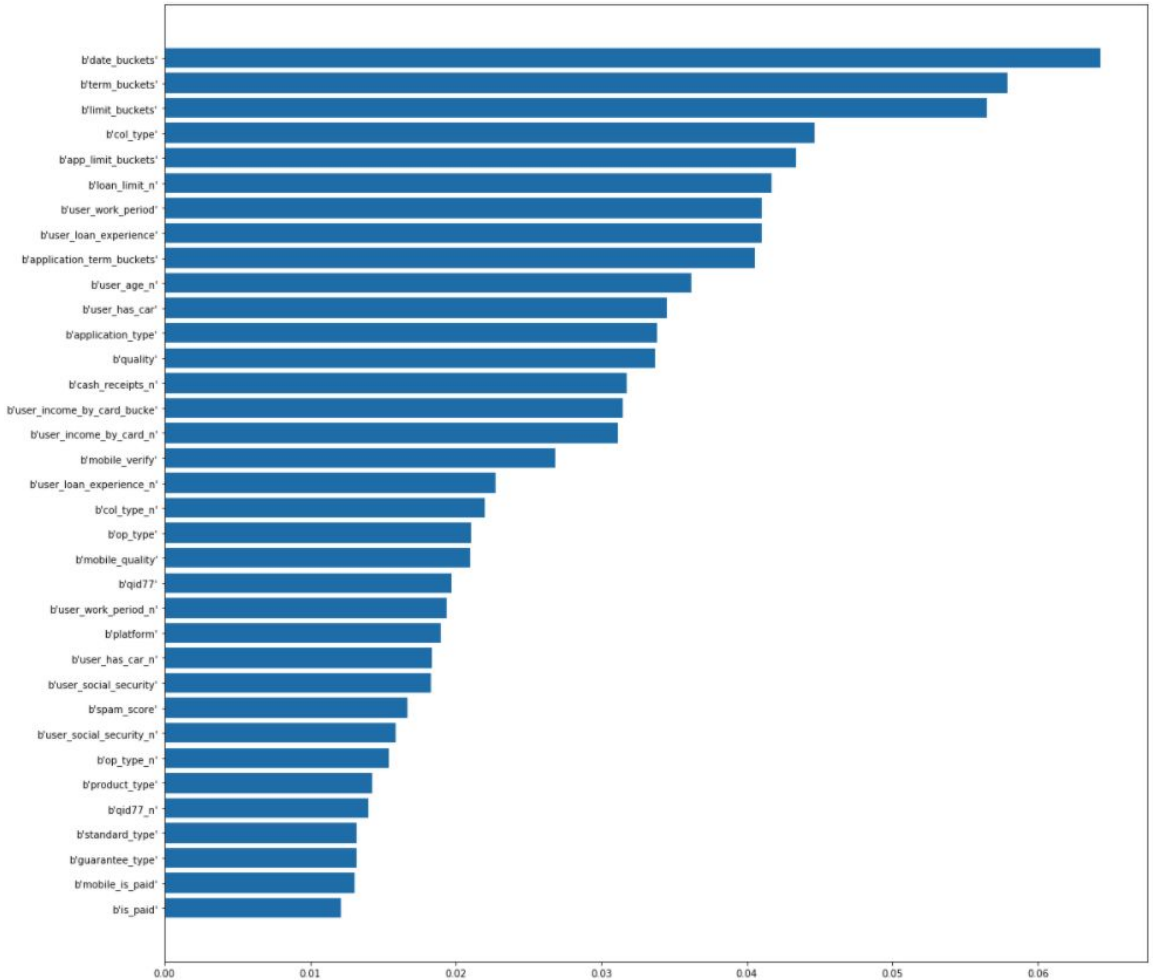


Whether Rong360 believes the applications are genuine matters

Other important characteristics

- Some important factors emerge through visualizations
- We use more rigorous methods to confirm what we've observed and reveal other determining characteristics
- Chi-squared tests failed to provide insight
- Random forest - generated feature importance was consistent with domain knowledge expectations and visualizations

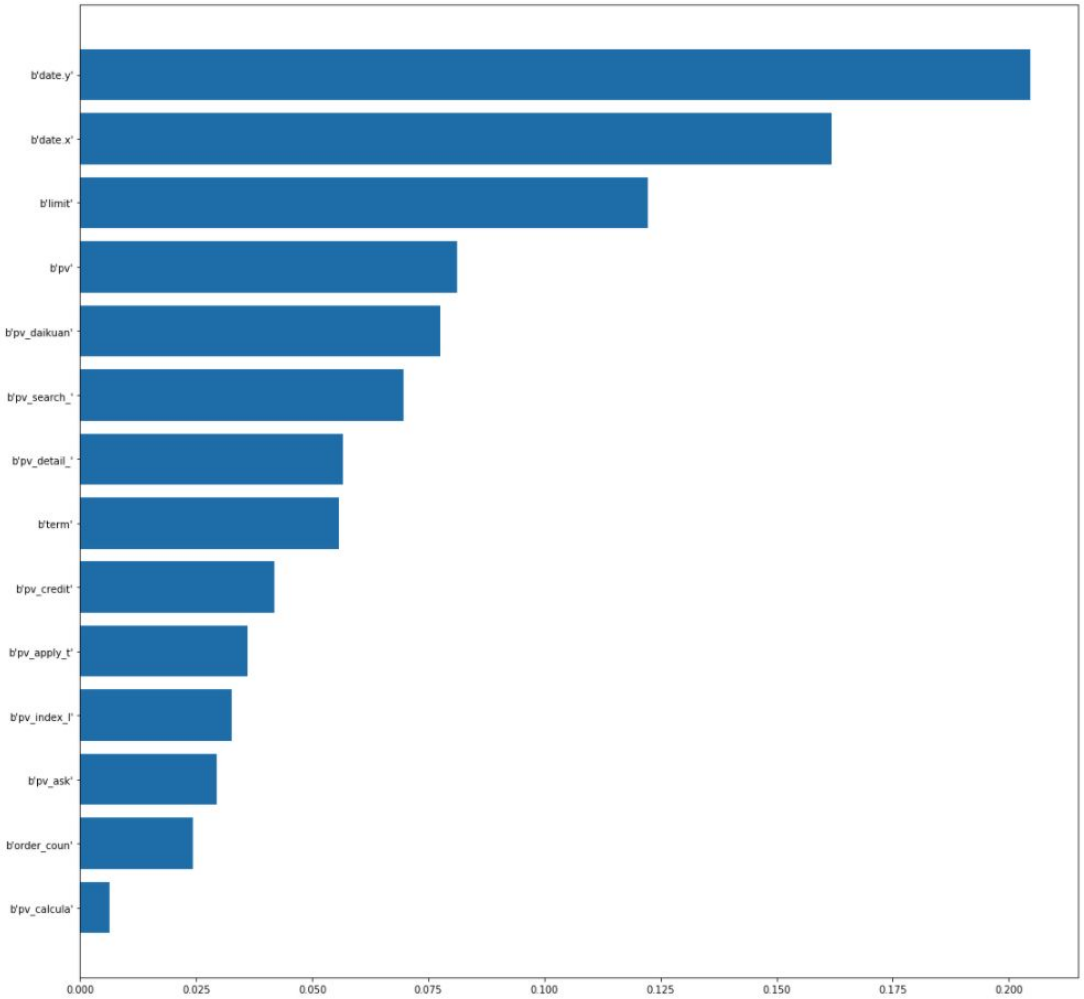
How user behavior on Rong360 site corresponds to loan approval



Random Forest* feature importance of application data

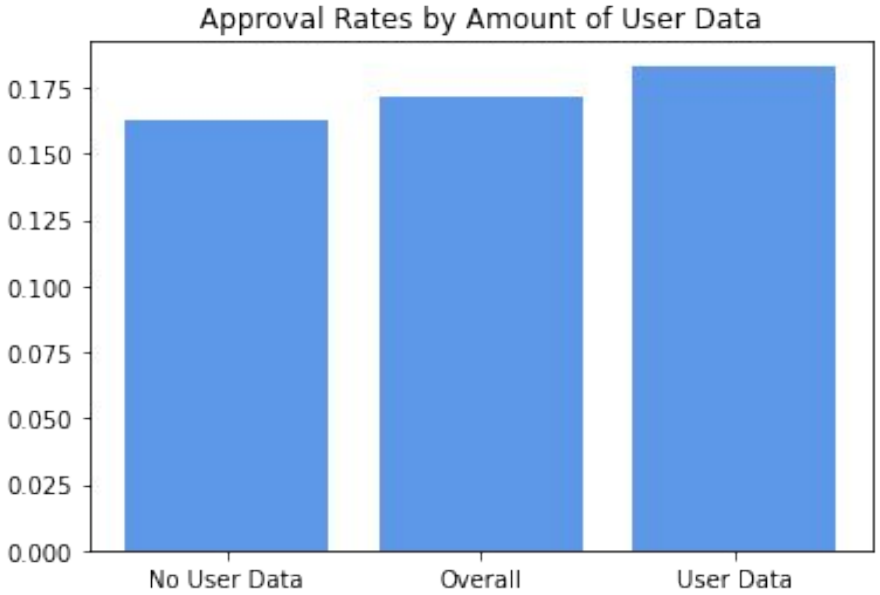
Notice loan term and loan limit are two of the three most important factors

*Train error: 0.06428
Test error: 0.14289



Most important user habits as determined by random forest*

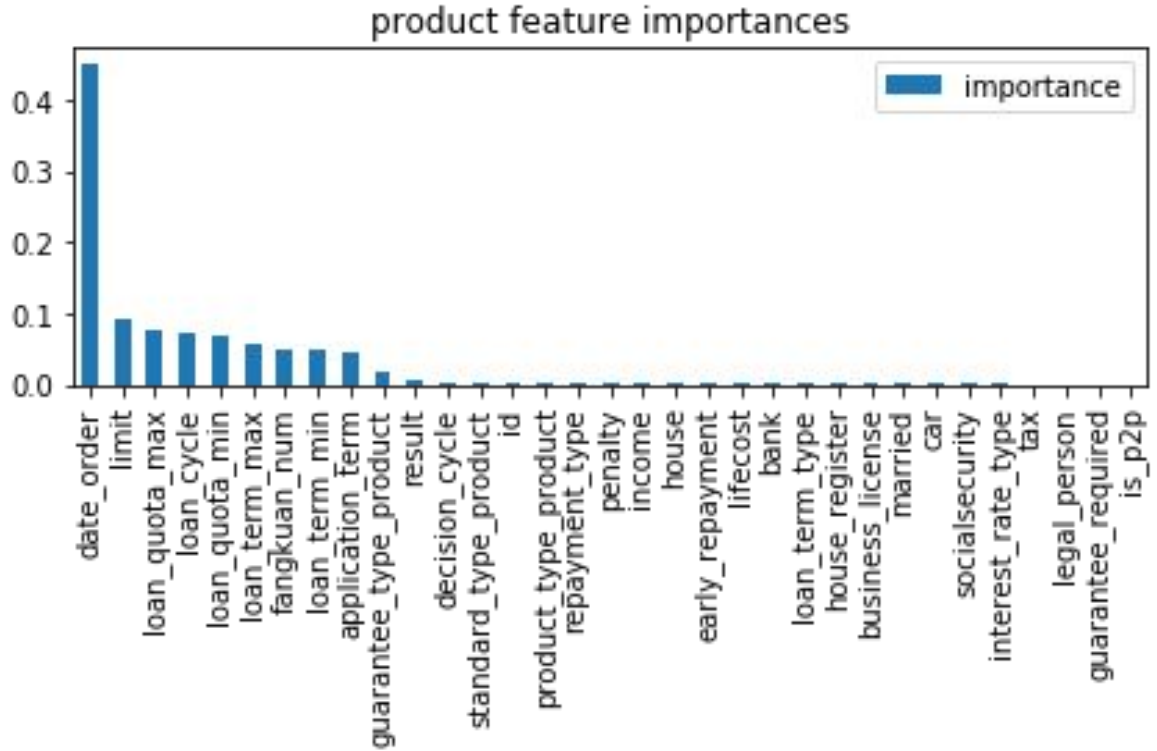
*Train error: 0.14149
Test error: 0.17155



When we have more user data (PV's etc) the general loan approval rate increases.

Understanding the loan landscape of Rong360

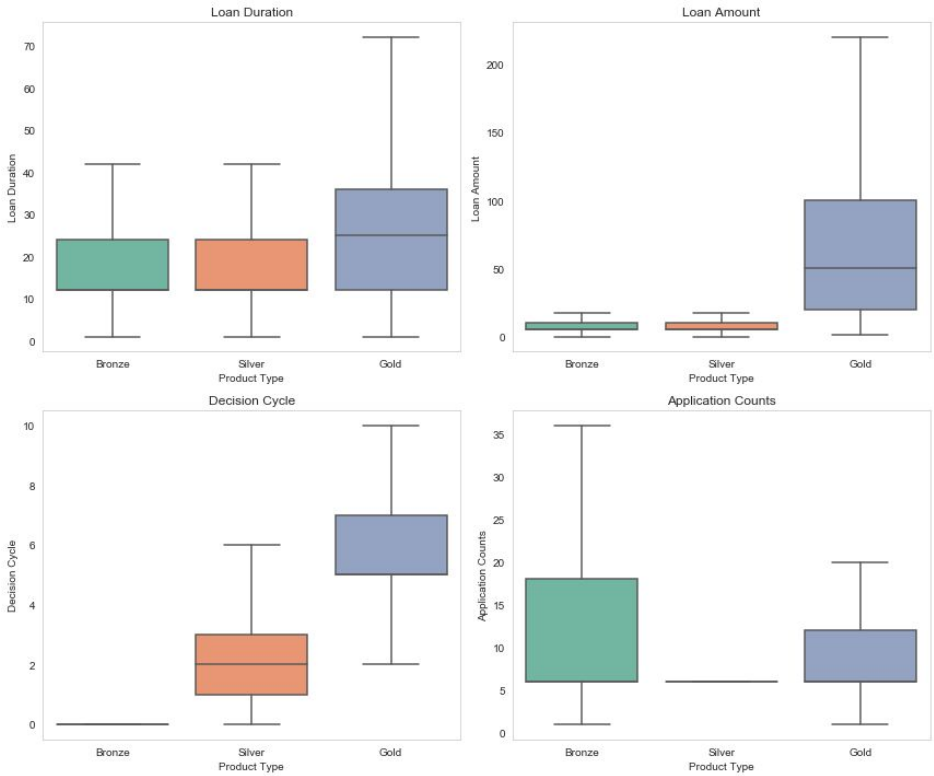
Which characteristics do the most sought-after loans have in common?

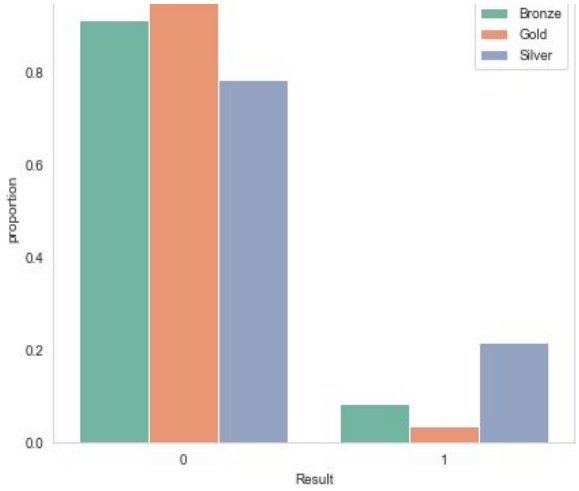


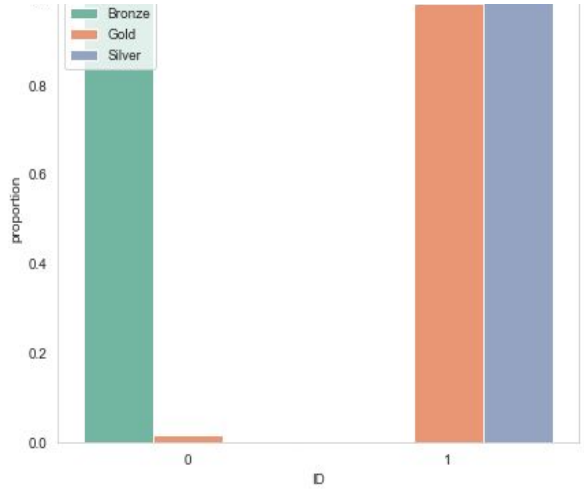
Which features determine the tier of a loan on Rong360?

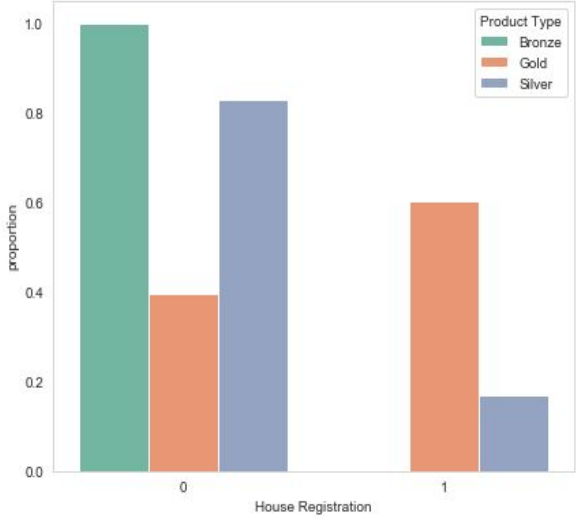
Three loan tiers

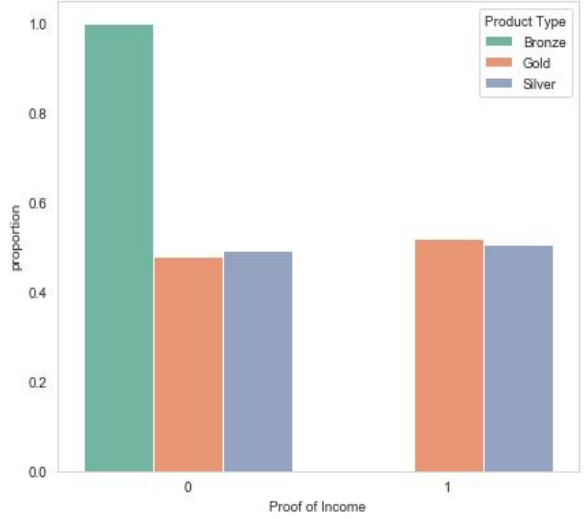
- Gold
- Silver
- Bronze

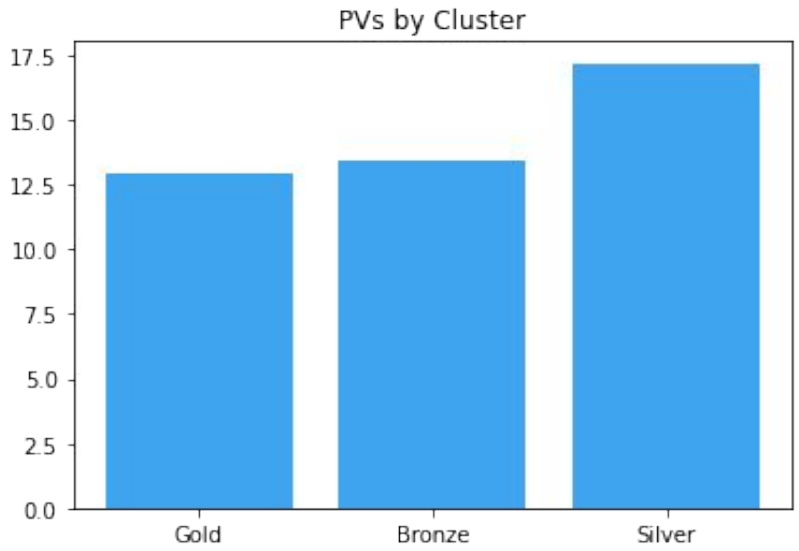


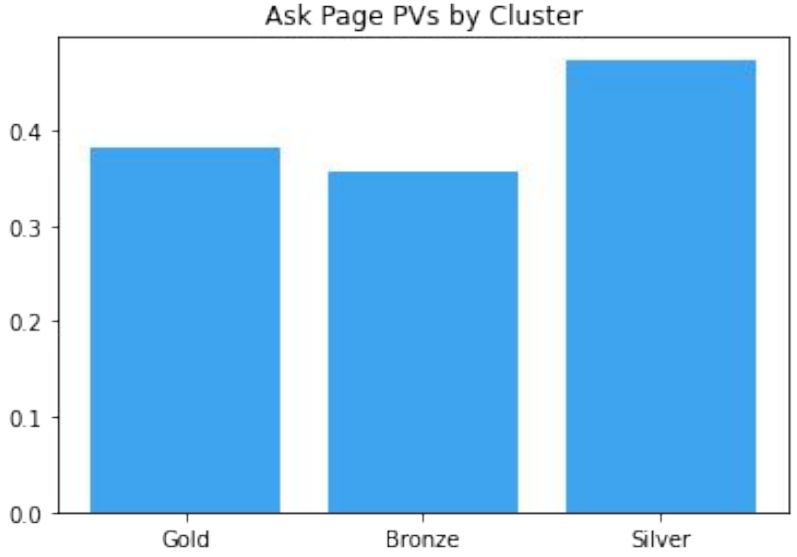

Bronze
Gold
Silver
proportion
0.0
0.2
0.4
0.6
0.8
0
1
Result


Bronze
Gold
Silver
proportion
0.0
0.2
0.4
0.6
0.8
0
1
ID

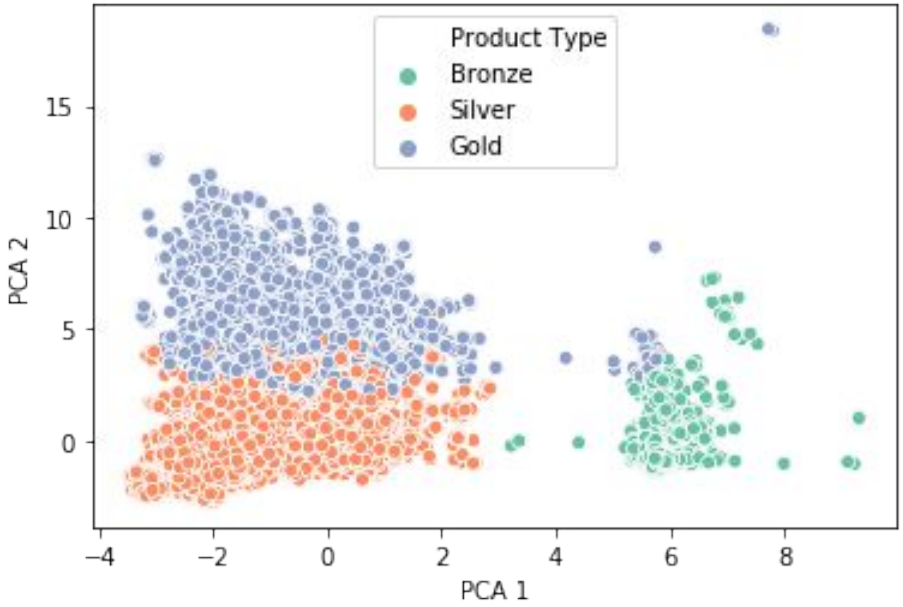

Product Type
Bronze
Gold
Silver
proportion
0.0
0.2
0.4
0.6
0.8
1.0
0
1
House Registration

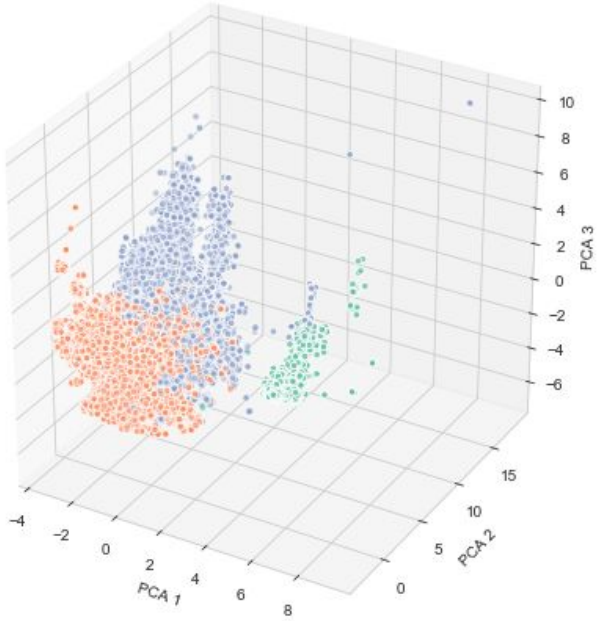

Product Type
Bronze
Gold
Silver
proportion
0.0
0.2
0.4
0.6
0.8
1.0
0
1
Proof of Income


PVs by Cluster
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Gold
Bronze
Silver


Ask Page PVs by Cluster
0.4
0.3
0.2
0.1
0.0
Gold
Bronze
Silver

PCA on loan characteristics





Recommendations

For Banks:

- Banks with products in the silver cluster can increase their numbers of applicants by loosening their requirements

For Consumers:

- Individuals and businesses that struggle to get a loan can approach banks through the Rong360 website to lower their risk profiles

Further Avenues of Research

- Digging deeper into the effect of location on loan approval
- Feature engineering a loan approval rate per bank, seeing how # of bank branches correlates with loan approval rate
- Digging deeper into a time series analysis
- Gaining more insight into the makeup of the clusters, use product types to further cluster users.

Talk about clustering

Elbow plots (in the appendix?)

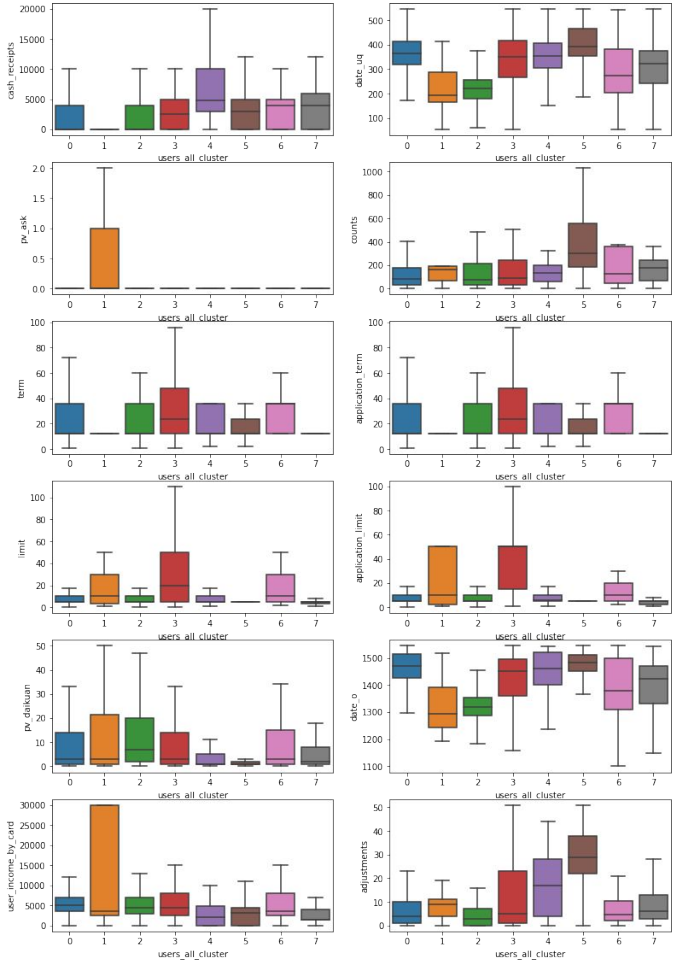
Talk about the more technical aspects of clustering (like normalization, parameter tuning, can put in the appendix)

Some more interpretation here about the clusters

Are these findings consistent with what we expected? Was anything surprising?

Banks can look at popular clusters and try to model themselves after them, likewise for users trying to increase their likelihood of approval

User clusters that could be more well defined

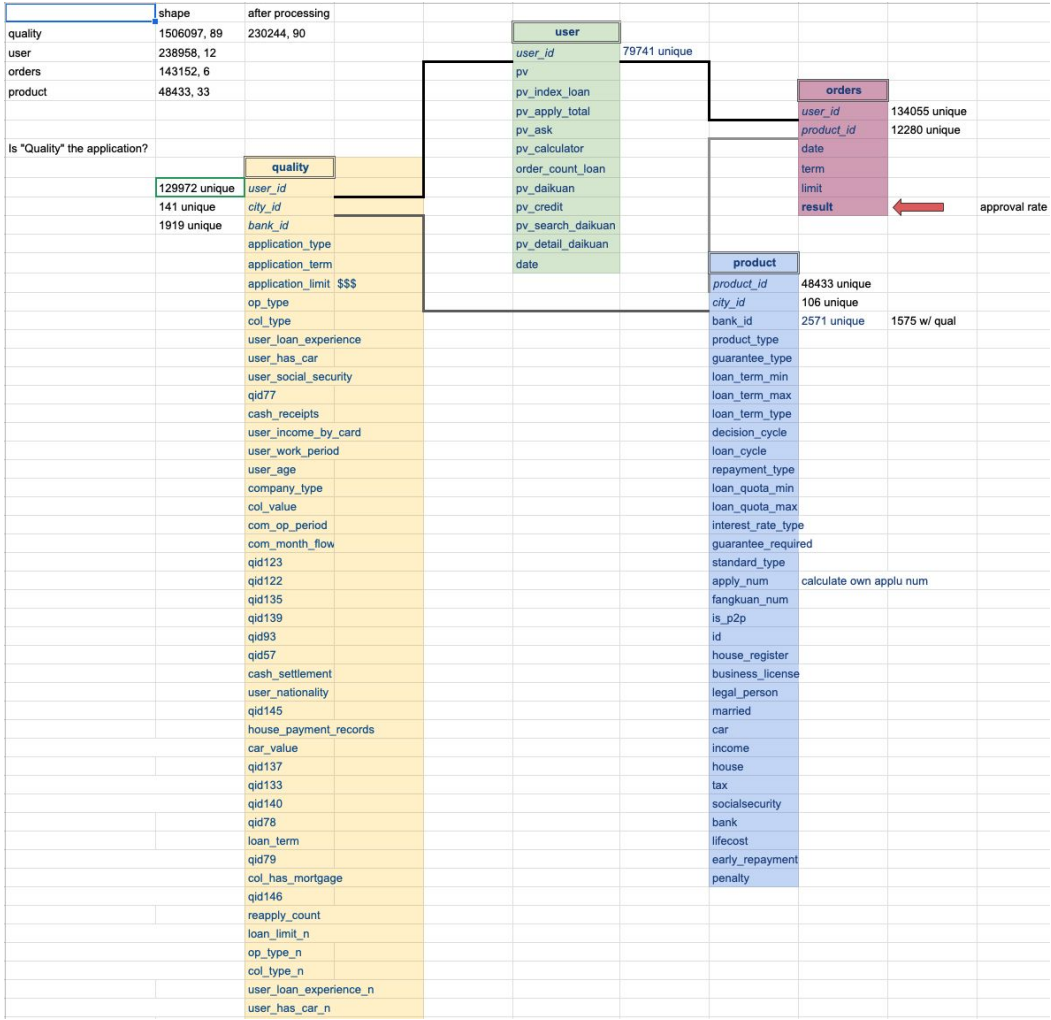


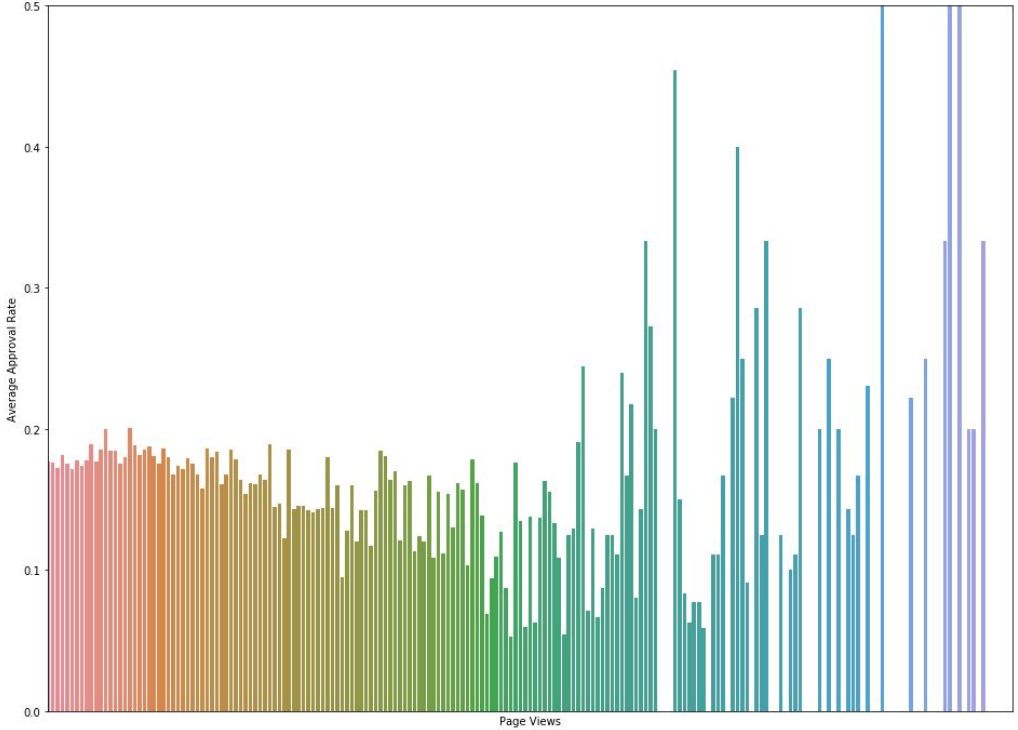
Product Levels:

Gold - Highest Limit, lowest approval rate and requires most “higher class” information. (Has car, has house, etc.)

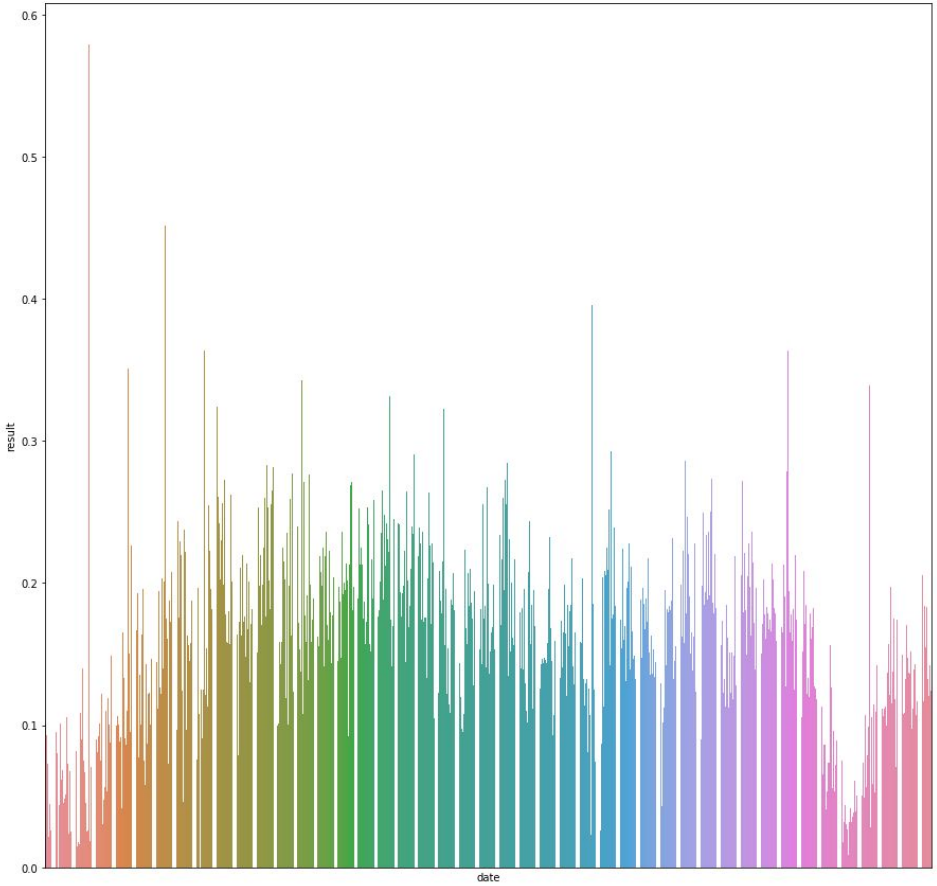
Silver - Lower limit, faster turnaround.

Bronze - Lowest limit, very little information required to apply.


shape
after processing
quality 1506097, 89 230244, 90
user 238958, 12
orders 143152, 6
product 48433, 33
Is "Quality" the application?
quality
129972 unique user_id
141 unique city_id
1919 unique bank_id
application_type
application_term
application_limit $$$
op_type
col_type
user_loan_experience
user_has_car
user_social_security
qid77
cash_receipts
user_income_by_card
user_work_period
user_age
company_type
col_value
com_op_period
com_month_flow
qid123
qid122
qid135
qid139
qid93
qid57
cash_settlement
user_nationality
qid145
house_payment_records
car_value
qid137
qid133
qid140
qid78
loan_term
qid79
col_has_mortgage
qid146
reapply_count
loan_limit_n
op_type_n
col_type_n
user_loan_experience_n
user_has_car_n
user
user_id 79741 unique
pv
pv_index_loan
pv_apply_total
pv_ask
pv_calculator
order_count_loan
pv_daikuan
pv_credit
pv_search_daikuan
pv_detail_daikuan
date
orders
user_id 134055 unique
product_id 12280 unique
date
term
limit
result
approval rate
product
product_id 48433 unique
city_id 106 unique
bank_id 2571 unique 1575 w/ qual
product_type
guarantee_type
loan_term_min
loan_term_max
loan_term_type
decision_cycle
loan_cycle
repayment_type
loan_quota_min
loan_quota_max
interest_rate_type
guarantee_required
standard_type
apply_num calculate own applu num
fangkuan_num
is_p2p
id
house_register
business_license
legal_person
married
car
income
house
tax
socialsecurity
bank
lifecost
early_repayment
penalty



Likelihood of loan approval drops as number of page views increases



A closer look at approval rates by date

Interesting avenue of further exploration

Partly because the dates were encoded, we chose to focus on understanding what makes loan products popular