Case Study:
Loan Repayment Analysis

Lampros Lountzis

Data Scientist

Agenda

Identifying the Challenge

Users’ Purchasing Habits

Users’ Repayment Behavior

User Loan Eligibility

Loan Repayment Model and Analysis

Next Steps and Considerations

1. 2. 3. 4. 5. 6.

Identifying the Challenge

29%

of high-risk loans,
resulting in

R$ 5.5m

of accumulated loan debt,
causing lenders to incur financial loss and increased administrative burdens.

58%

of customers with high-risk repayment behavior,
causing damage to their credit score and additional fees and interest.

Based on 3,046 users who generated 6,588 loans

Key Concept: High-Risk Loans and Candidates

Loans Resulting in Repayment

... are assessed based on their **default rate**. **High-risk loans are considered those with a default rate that is an outlier compared to the category they belong to**.

The categories are defined based on the principal amount.
Outliers = Q3 + 1.5 x IQR

Loans Resulting in Debt

... are considered **high-risk loans**.

A user's risk is defined by linking their **purchasing and loan repayment behavior** with the **characteristics and quality of the loan**.

1. **2.** 3. 4. 5. 6.

Users’ Purchasing Habits

Identifying our users' purchasing habits

Based on RFM modeling

8%
Champions

Extremely active with *moderate to high* monetary value.

Prefer to use *both* credit and debit cards.

Opt for installment plans that span an *avg. of 3 installments*.

Low transaction rejection rate (avg. 10%).

3%
Big Spenders

Active customers with *high* monetary value.

Slight preference for online purchases.

Opt *frequently* for installment plans, typically spanning an *avg. of 6 installments*.

High transaction rejection rate (avg. 30%).

35%
Promising

Active customers with *low to moderate* monetary value.

Opt *periodically* for installment plans, typically spanning an *avg. of 3 installments*.

6%
Recent

Customers who entered our base *recently* with *low to moderate* monetary value.

Opt *frequently* for installment plans, typically spanning an *avg. of 5 installments*.

Moderate transaction rejection rate (avg. 21%).

48%
Inactive

Customers with *extremely low* activity.

Opt *frequently* for installment plans, typically spanning an *avg. of 5 installments*.

Moderate transaction rejection rate (avg. 20%).

1. 2. 3. 4. 5. 6.

Users’ Repayment Behavior

Identifying our users’ loan portfolio

Champions

Mainly large loans (90%).

Mostly loans with small interest and large interest rates (58% and 38% respectively).

Prefer to breakdown the repayment into **multiple small** ones **(39 of R$265 each on average).**

Big Spenders

Exclusive preference for large loans (97%).

Mostly loans with small interest and large interest rates (70% and 27% respectively).

Prefer to breakdown the repayment into **fewer larger** ones **(14 of R$835 each on average).**

Promising

Mainly large loans (89%).

A good mixture of loans with different interest rates (72% small, 7% medium and 21% large).

Prefer to breakdown the repayment into **multiple small** ones **(40 of R$257 each on average**).

Recent

Mainly large loans (79%).

A good mixture of loans with different interest rates (55% small, 12% medium and 33% large).

Prefer to breakdown the repayment into **fewer larger** ones **(14 of R$663 each on average).**

Prefer manual repayments over an automated plan.

Inactive

Mainly large loans (89%).

Tend to avoid loans of medium to large interest rates (only 2%).

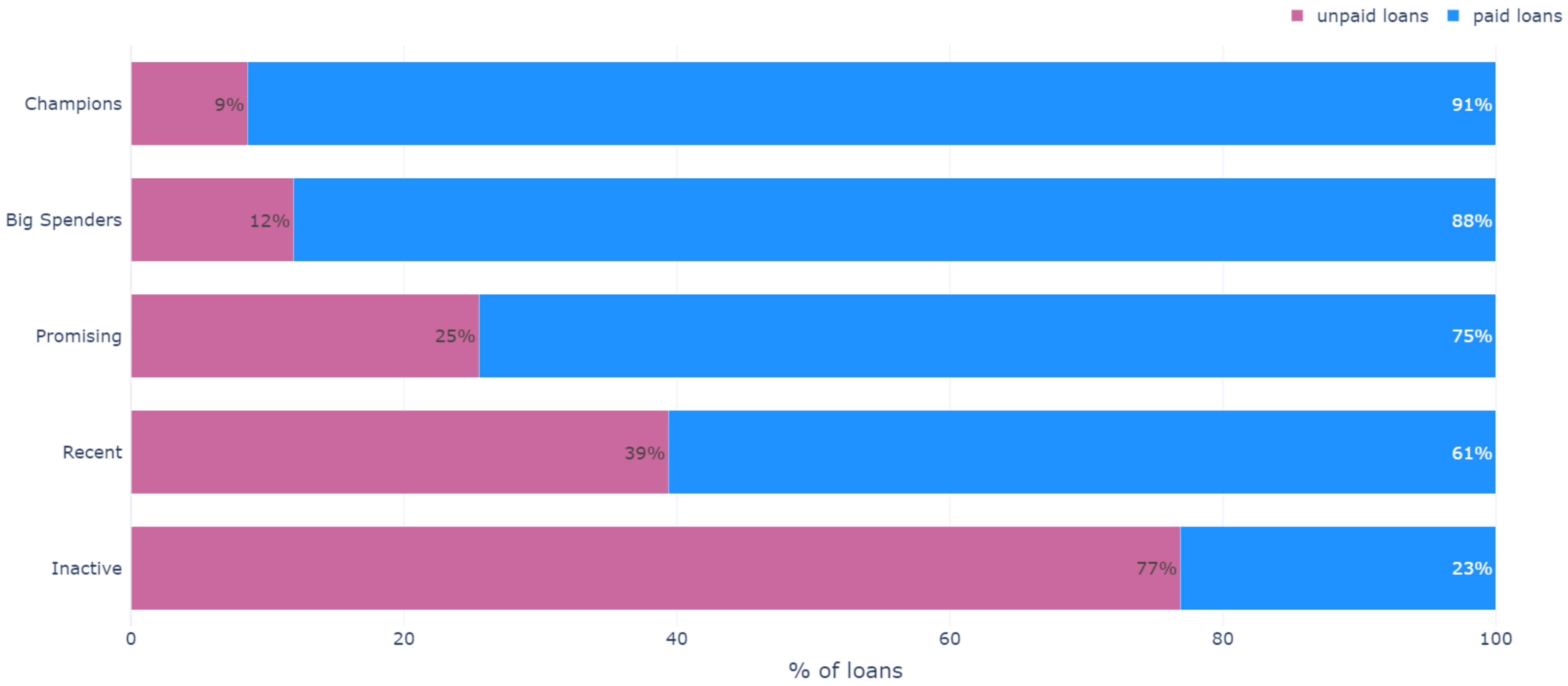
Prefer to breakdown the repayment into **fewer larger** ones **(21 of R$464 each on average)**.

Prefer manual repayments over an automated plan.

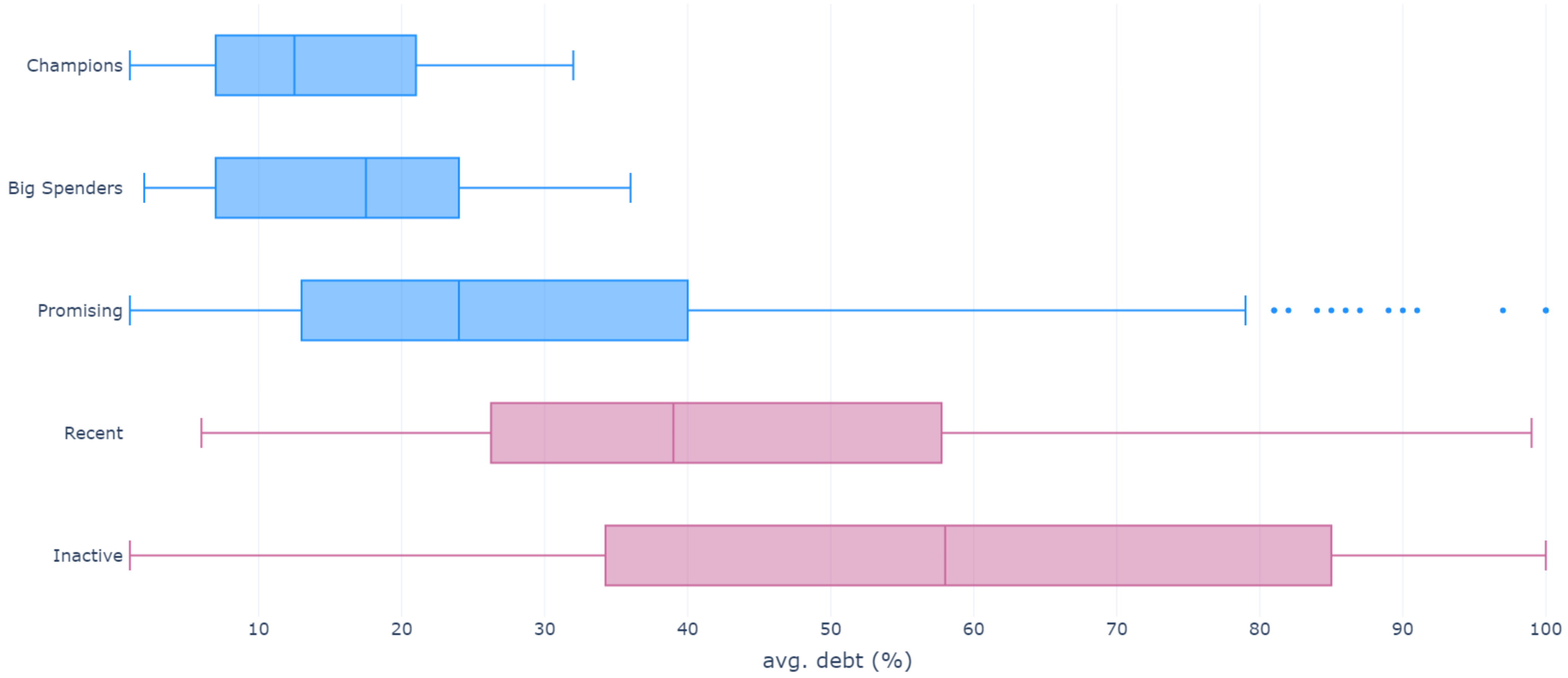
Amount: small: *less than R$ 3k,* ***medium:*** *R$ 3k – 5k,* ***large:*** *more than R$ 5k*

Interest Rate: small: *less than 40%,* ***medium:*** *40% - 100% ,* ***large:*** *greater than 100%*

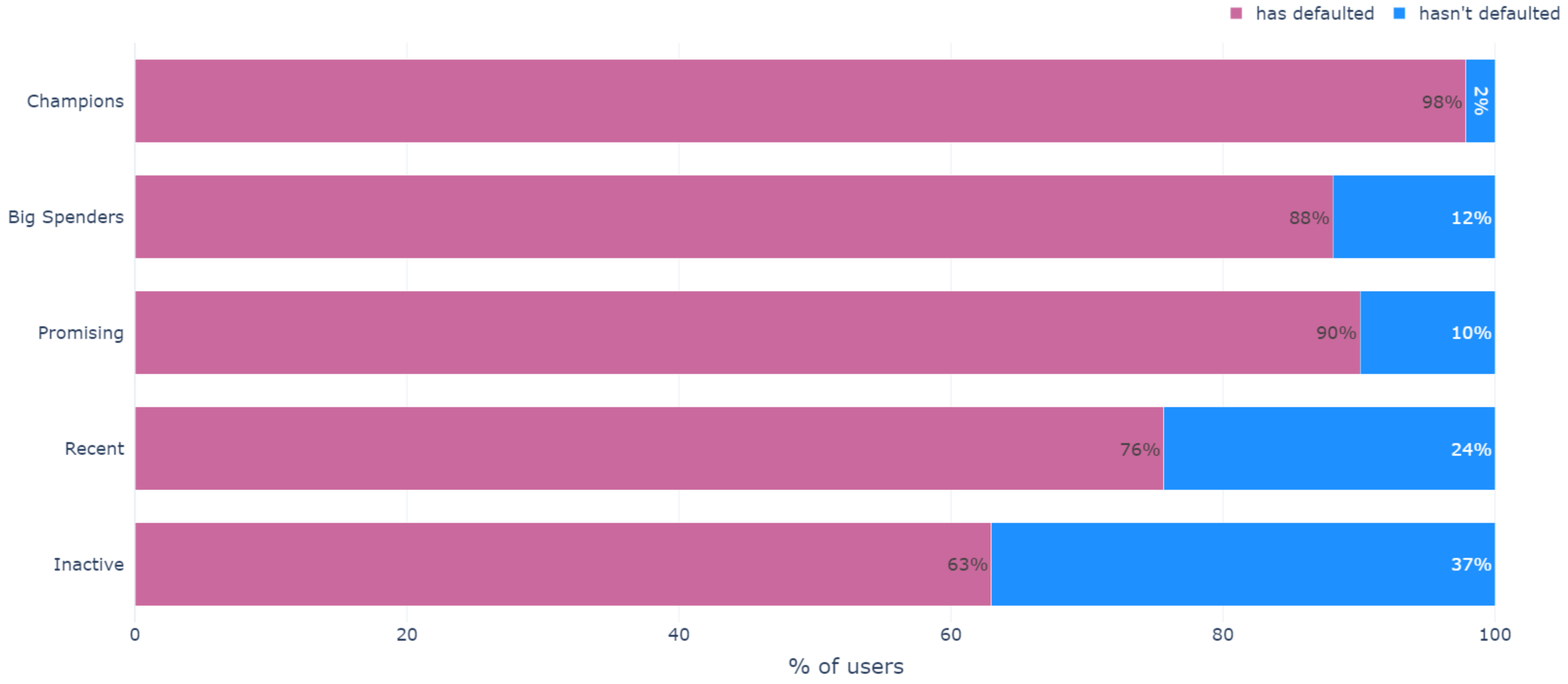
Recent and Inactive users are more prone incurring loan debt ...



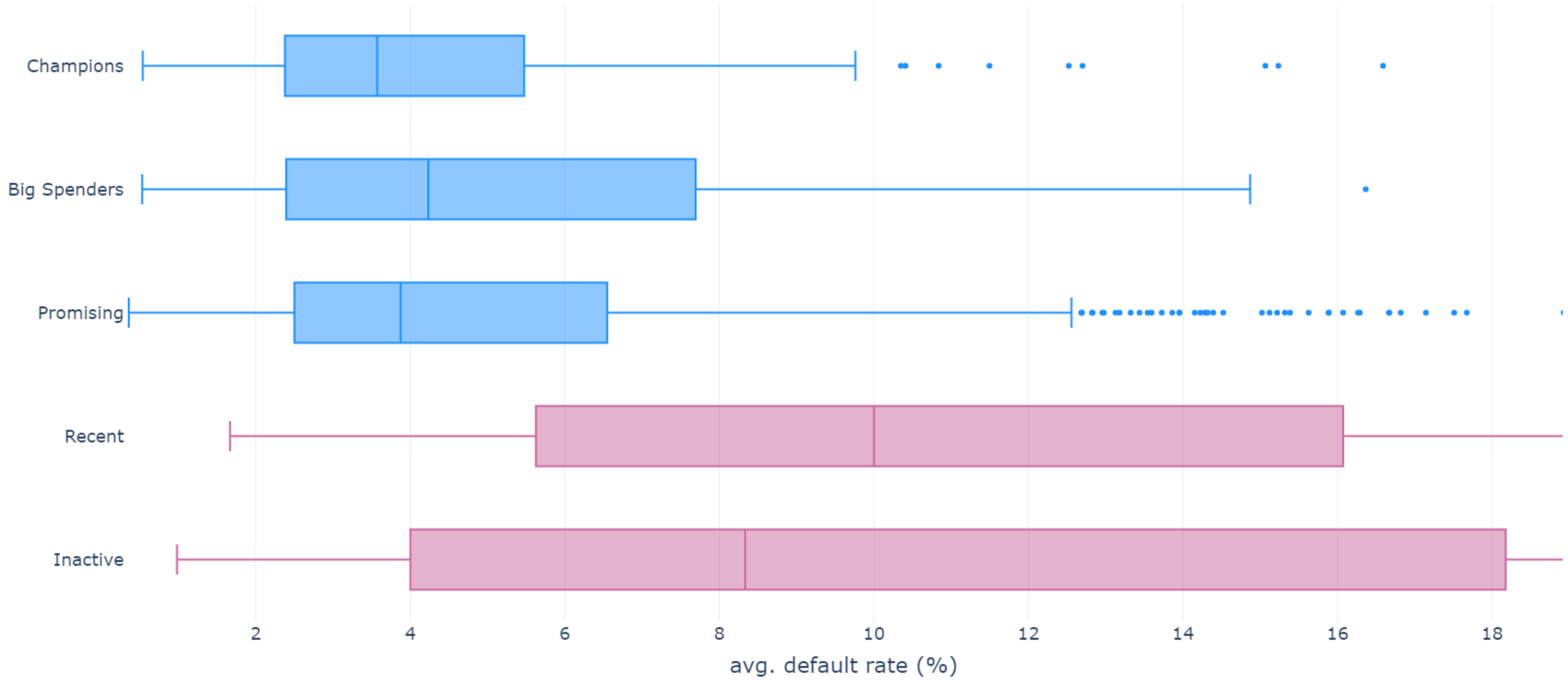
… and can potentially owe more than 50% of the loan amount



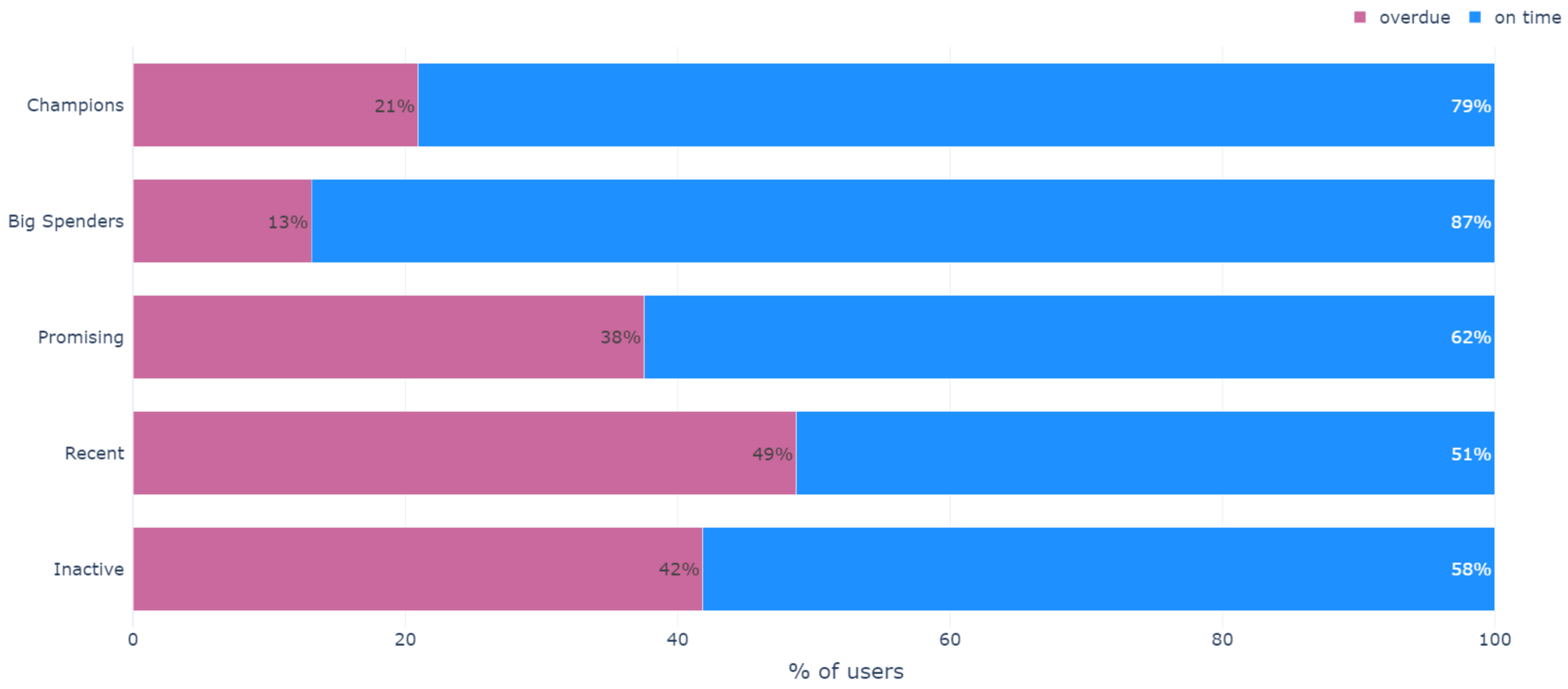
Frequent users exhibit a higher tendency towards defaulting ...



… however, infrequent users show greater repayment uncertainty



Users with low to moderate spending habits are likely to fall behind the due date



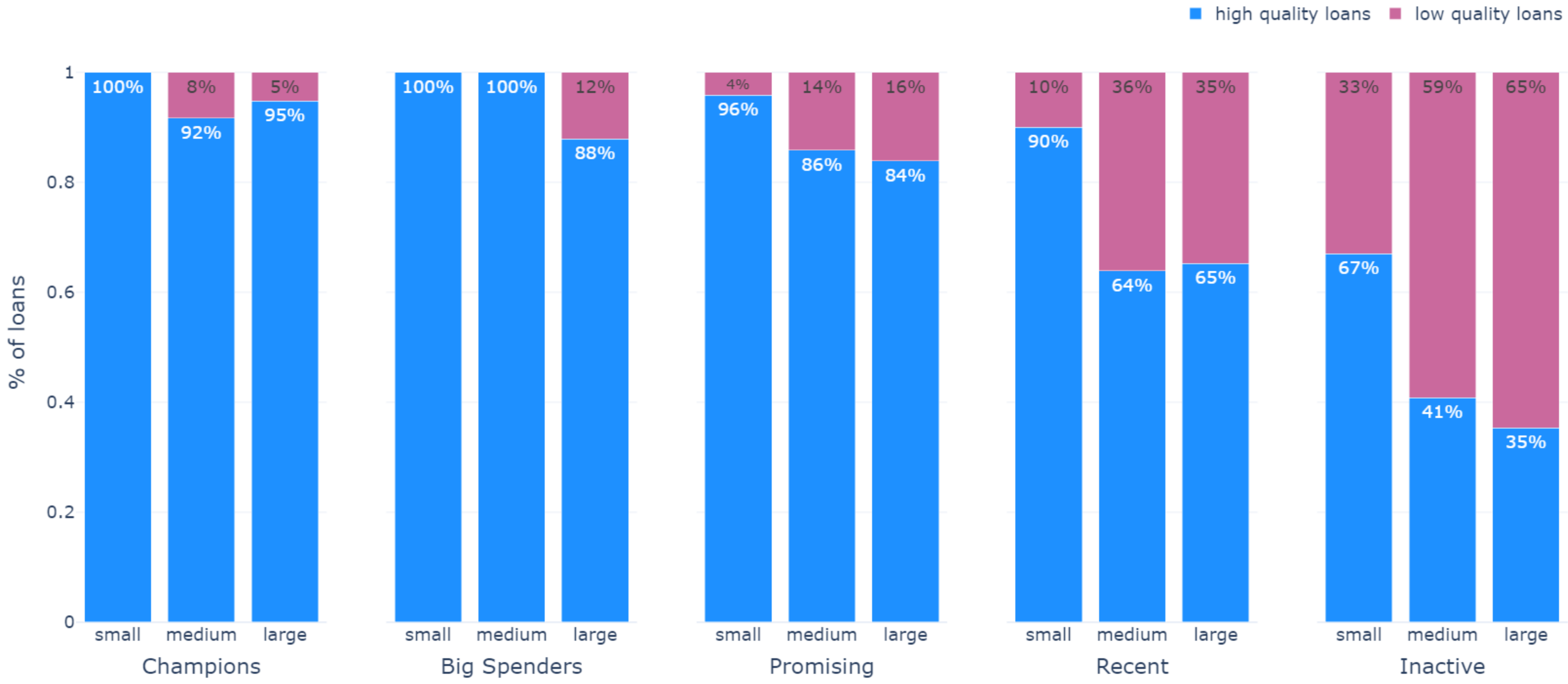
1. 2. 3. **4.** 5. 6.

User Loan Eligibility

Infrequent users present a high-risk repayment behavior

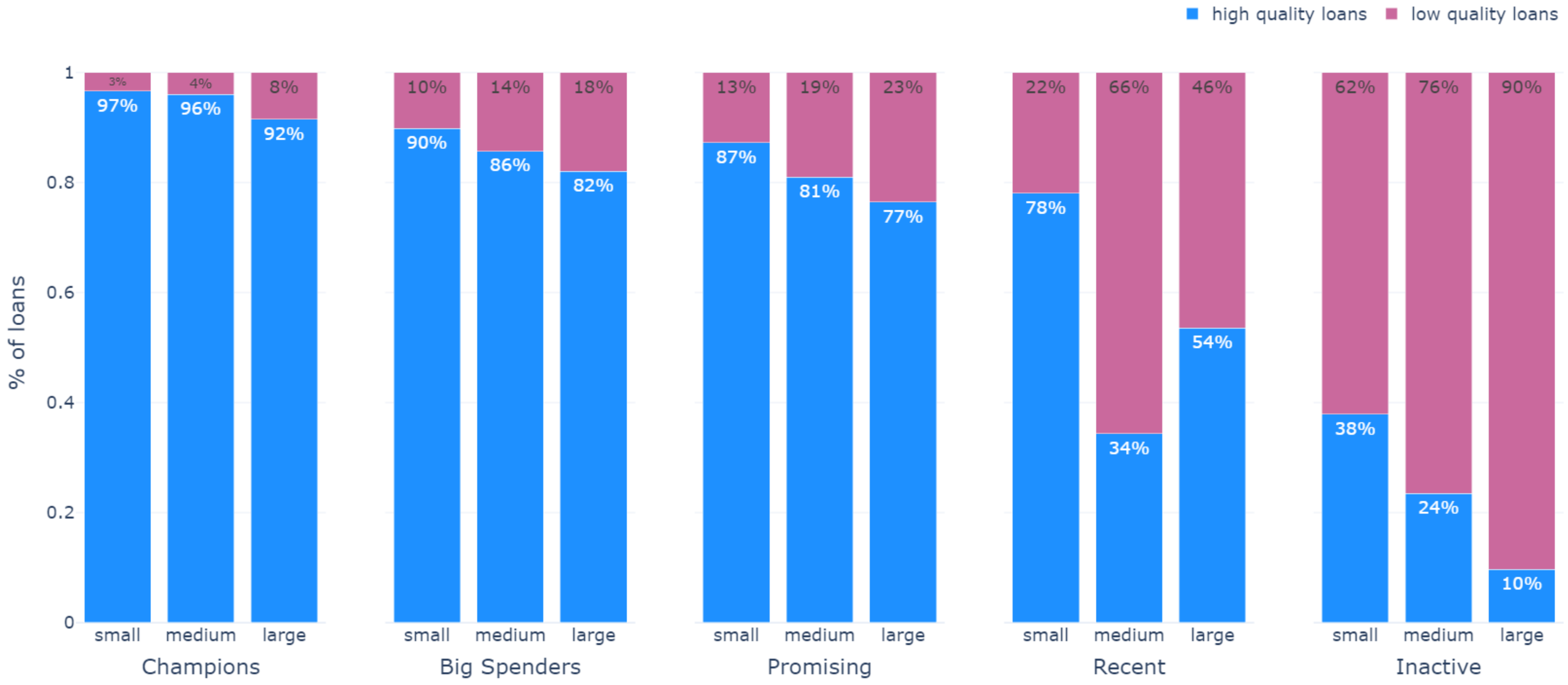
Champions	Big Spenders	Promising	Recent	Inactive
5%	12%	15%	32%	63%
high-risk loans	high-risk loans	high-risk loans	high-risk loans	high-risk loans

Risk increases with the loan amount, with *Recent* and *Inactive* users exhibiting a high-risk repayment behavior



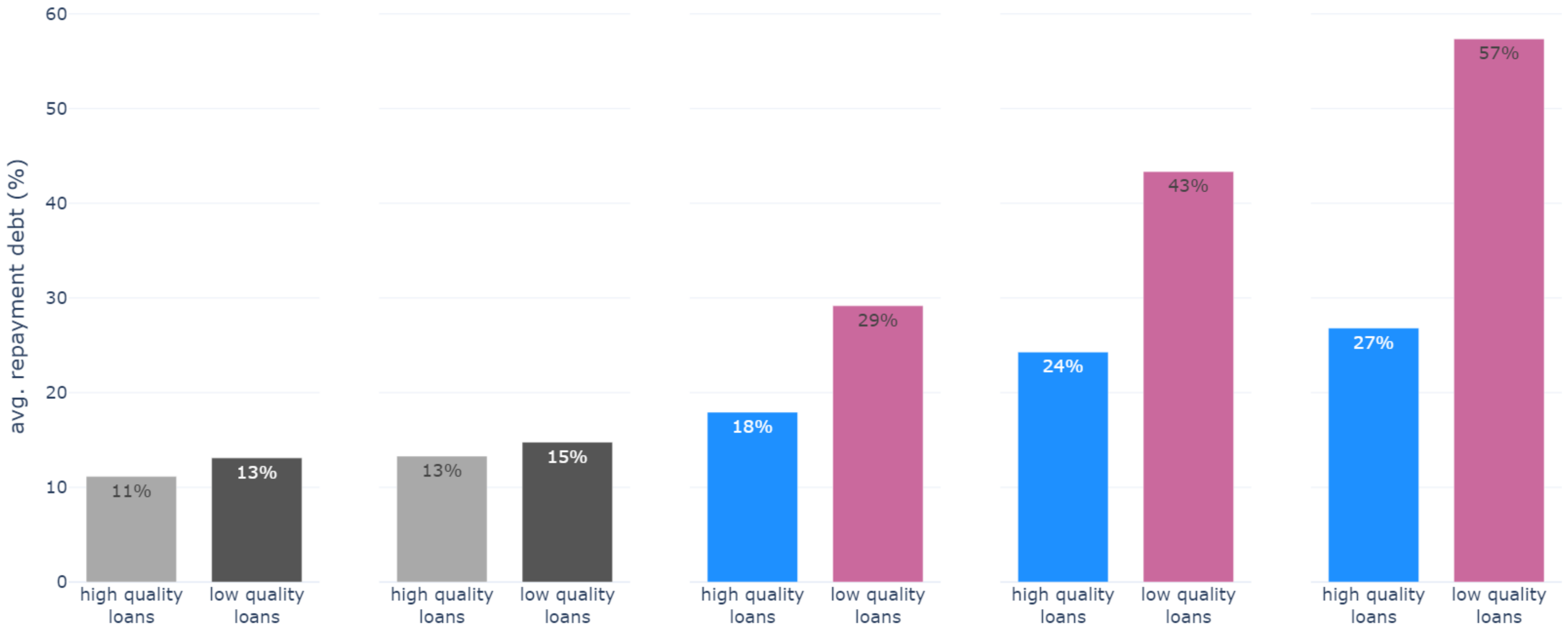
small: *less than R$ 3k,* ***medium:*** *R$ 3k – 5k,* ***large:*** *more than R$ 5k*

Similar risk can be observed in relation to the loan interest rates

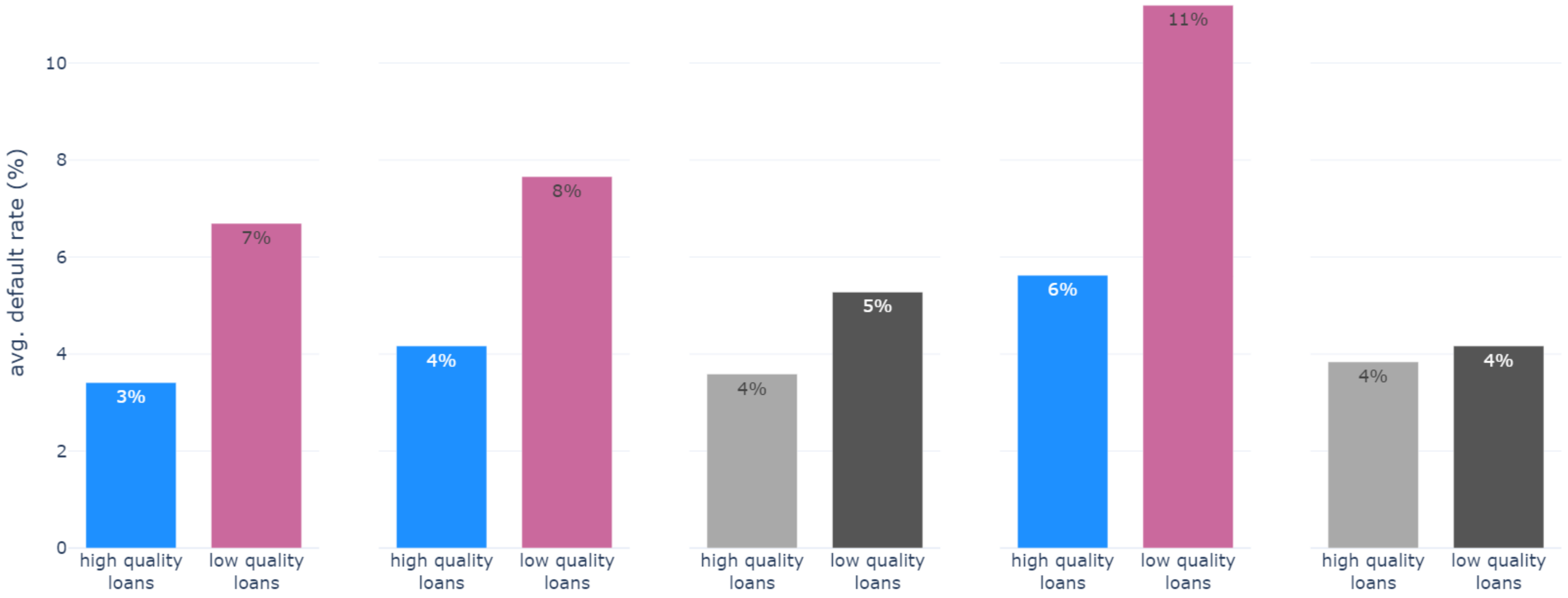


small: *less than 40%,* ***medium:*** *40% - 100% ,* ***large:*** *greater than 100%*

Users with a significant repayment debt, particularly *Recent* and *Inactive* users, exhibit poor repayment behavior

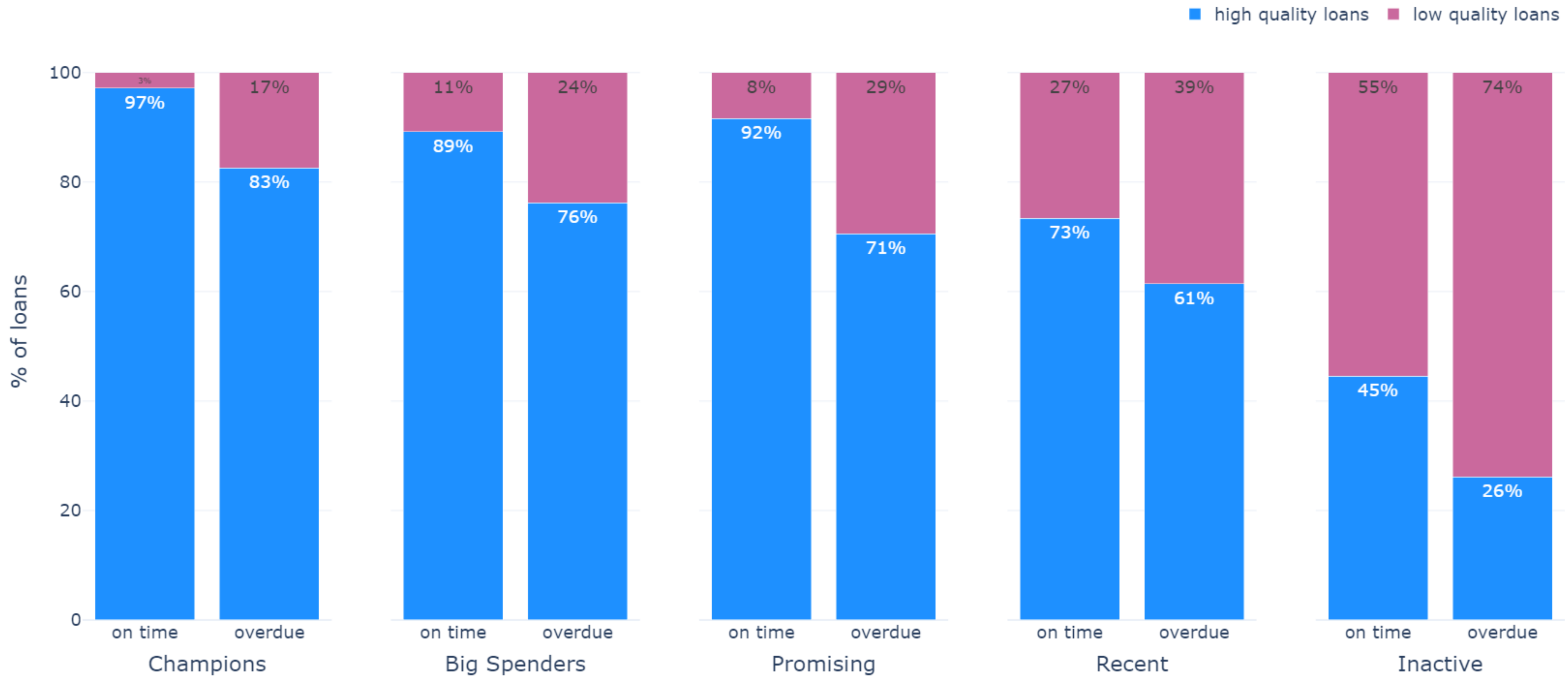


Higher default rates within the user segments are indicators of high-risk repayment behavior

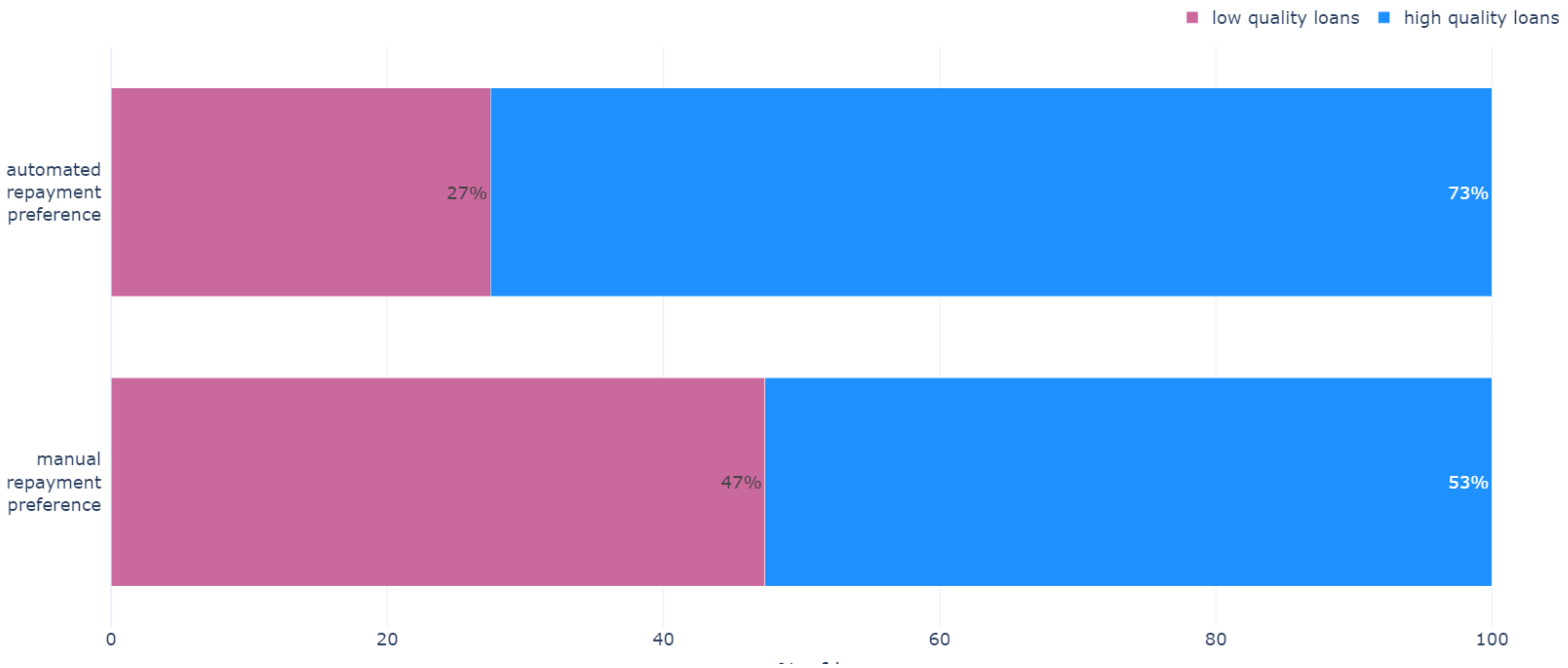


Users who exceed the repayment due dates are typically associated with riskier repayment behaviors

Recent and Inactive users have shown a higher propensity to miss their due dates



Users who don’t prefer automated repayment schedules are more likely to exhibit riskier repayment behavior



1. 2. 3. 4. **5.** 6.

Loan Repayment Model and Analysis

The model identifies 71% of the high-risk loan candidates, however, with low confidence levels

Results on the test set

	Precision	Recall	F1 score
low-risk candidate	88%	83%	85%
high-risk candidate	62%	71%	66%
overall	80%	80%	80%

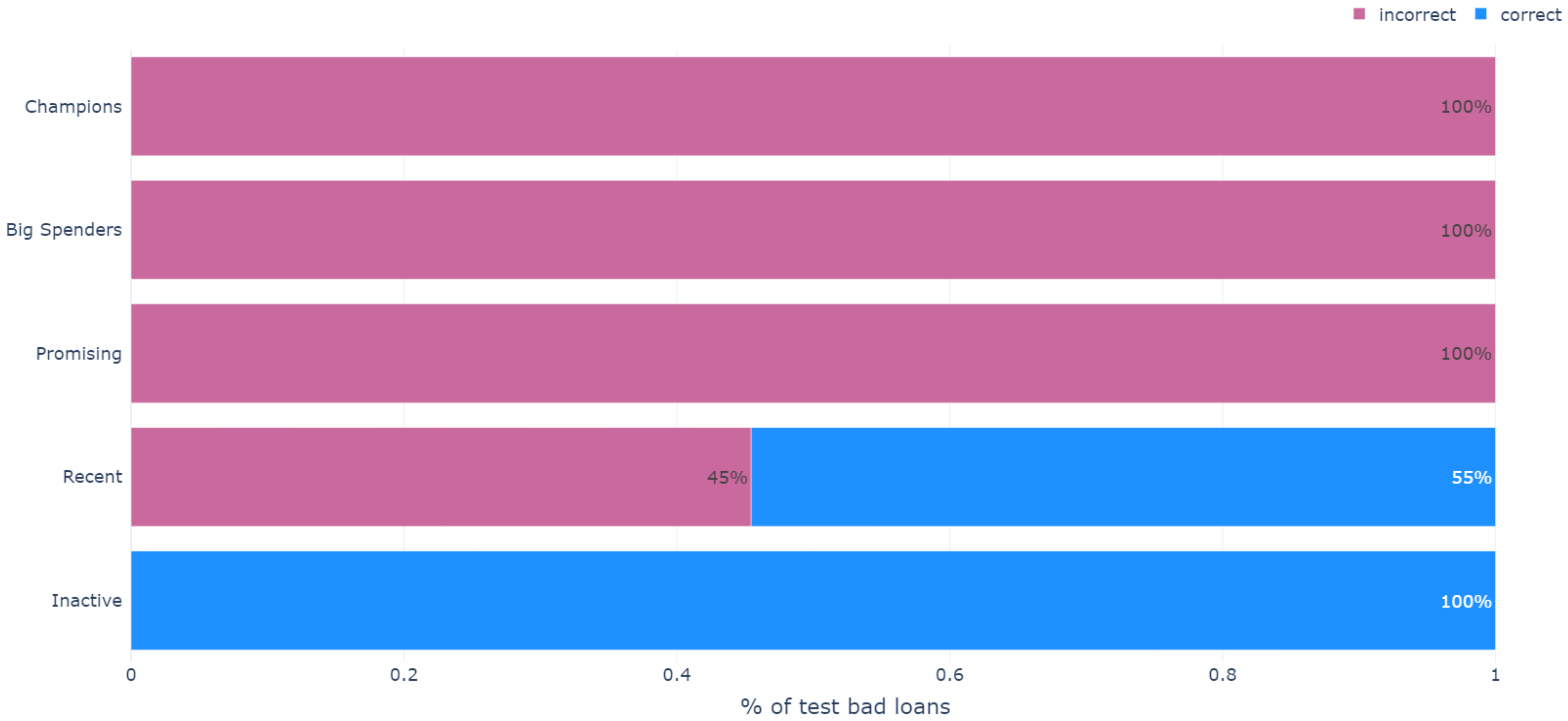
weighted avg. scores for Random Forest

Top predictive factors include

- **purchasing behavior** (RFM segments, transaction rejection rate, installment preference, preference of purchasing online)
- **repayment behavior** (preference of manual repayments, avg. repayment amount, users' portfolio of amount and interest rate sizes)
- **loan characteristics**

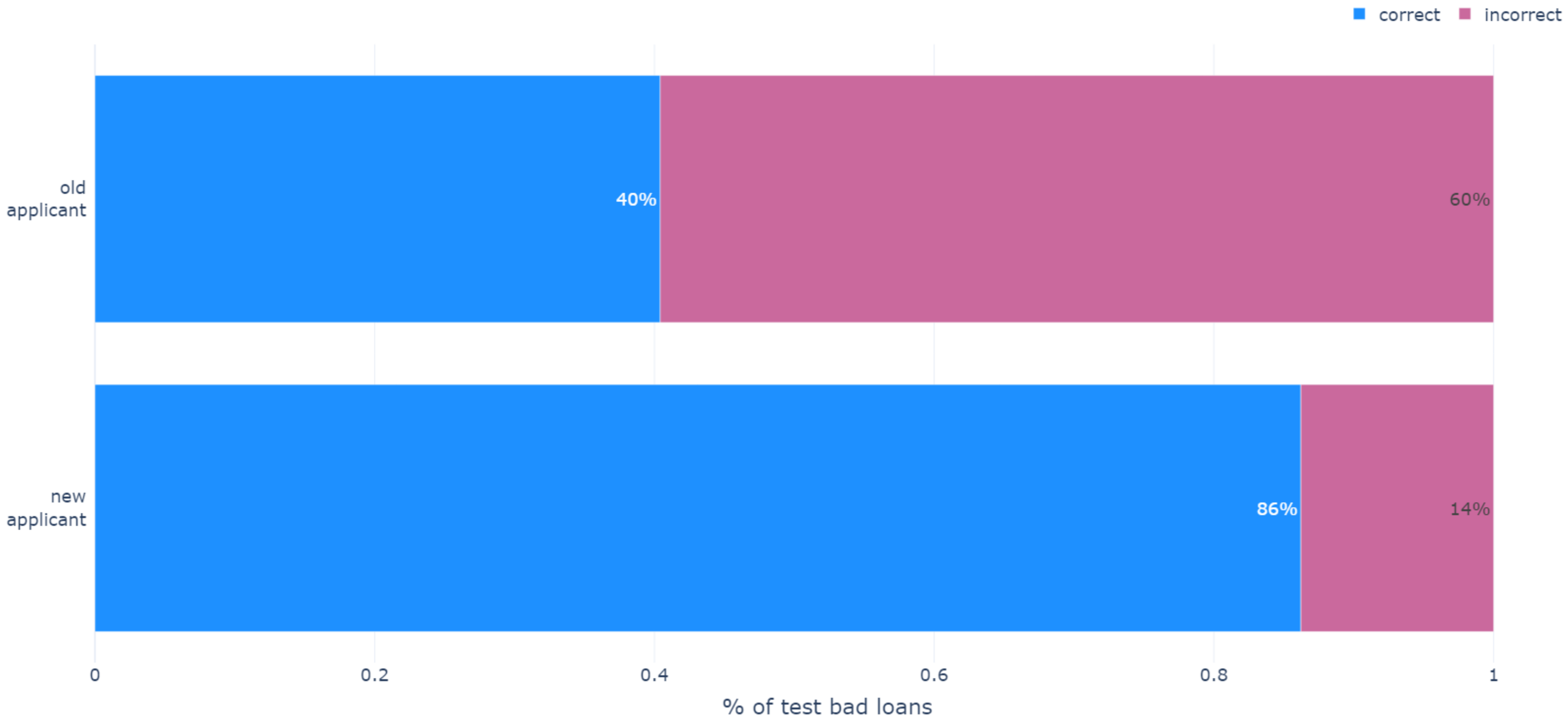
Challenge in predicting when “healthy” users incur high-risk loans

High-risk candidates in the test set



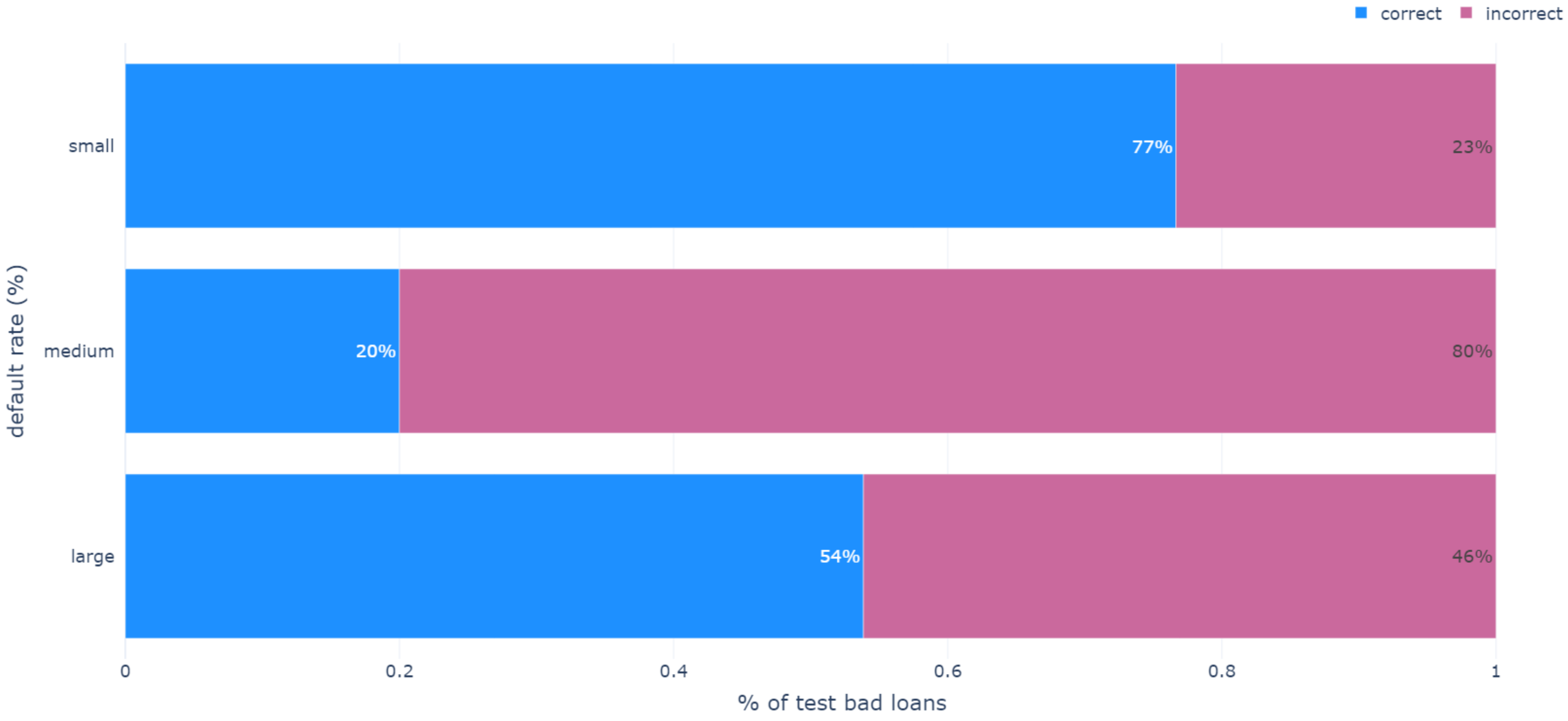
Although the model is accurate when assessing new applicants, its performance declines for old applicants

High-risk candidates in the test set



Difficulty to capture bad loan candidates with medium to large default rates

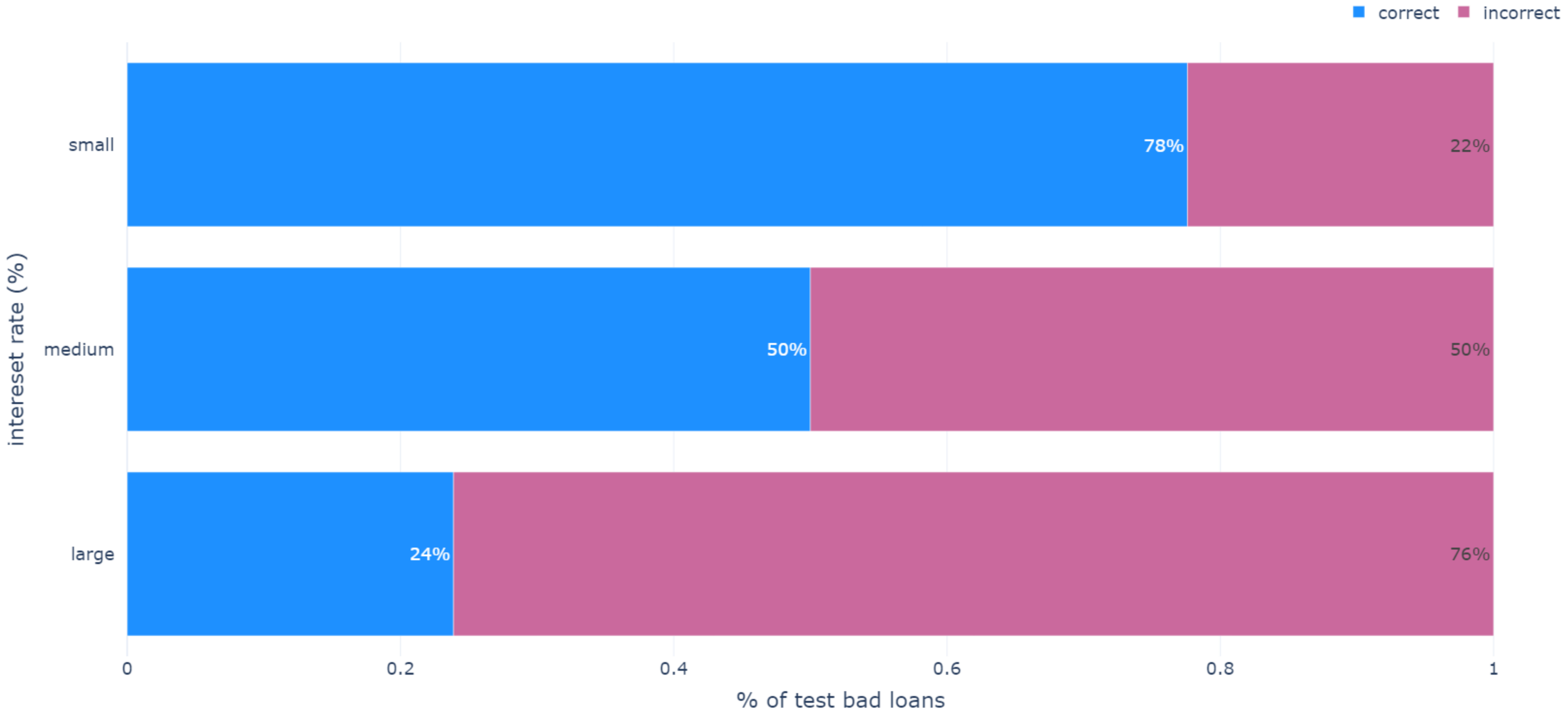
High-risk candidates in the test set



small: *less than 5%,* ***medium:*** *5% - 10% ,* ***large:*** *greater than 10%*

The model cannot capture bad repayment patterns mainly for loans with medium to large interest rates

High-risk candidates in the test set



***small:** less than 40%, **medium:** 40% - 100% , **large:** greater than 100%*

1. 2. 3. 4. 5. **6.**

Next Steps and Considerations

Next Steps and Considerations

- **Expand the set of features** to include user demographics (e.g., state, age, occupation, etc.) and other various loan characteristics (e.g., loan purpose, etc.).

- Transition from RFM customer analysis to a more refined **clustering** method to model customer purchasing behavioral patterns.

- **User embeddings** to capture complex and nuanced information about the users' behaviors, attributes, and interactions.

- **Signal features** about the user, e.g., when the user is taking loans, changes in lending amounts, etc.. Withing the context of Lending, *different signals* have *different strengths* and can quantify when the user is **switching context**.

Thank You!

Lampros Lountzis
Data Scientist

QnA Session