Case Study
Loan Repayment Analysis

Lampros Lountzis

Data Scientist

May 2nd, 2024

cloudwalk

Agenda

Identifying the Challenge

Defining the Landscape

Key Insights

Loan Repayment Model and Analysis

Next Steps and Considerations

74%

of high-risk loans,
resulting in

R$ 6m

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

73%

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 by benchmarking their **default rate** to the **default rate of the category that they belong to (3rd quartile)**.

The categories are defined based on the principal amount and the interest rate.

Loans Resulting in Debt

… are considered immediately as **high-risk loans**.

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

Identifying our users' purchasing habits

Based on RFM modeling

8% Champions

Extremely active with moderate to high monetary value.

Very low transaction rejection rate (8%).

3% Big Spenders

Active customers with high monetary value.

Slight preference for online purchases.

Prefer to use mostly credit cards.

Opt frequently for installment plans.

High transaction rejection rate (27%).

35% Promising

Active customers with low to moderate monetary value.

Opt periodically for installment plans.

6% Recent

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

Prefer to use mostly credit cards.

Opt frequently for installment plans.

High transaction rejection rate (19%).

48% Inactive

Customers with extremely low activity.

Prefer to use mostly credit cards.

Opt frequently for installment plans.

High transaction rejection rate (16%).

Identifying our users' loan repayment habits

Based on RFM modeling

8% Champions

90% has defaulted at least once.

The typical default rate is around **2% to 5%.**

2% of their loans resulted in debt.

17% have fallen behind the due date.

2% to 6% of their repayments are overdue.

3% Big Spenders

81% has defaulted at least once.

The typical default rate is around **3% to 9%.**

2% of their loans resulted in debt.

10% have fallen behind the due date.

10% to 14% of their repayments are overdue.

35% Promising

77% has defaulted at least once.

The typical default rate is around **3% to 6%.**

11% of their loans resulted in debt.

30% have fallen behind the due date.

3% to 12% of their repayments are overdue.

6% Recent

62% has defaulted at least once.

The typical default rate is around **7% to 16%.**

21% of their loans resulted in debt.

41% have fallen behind the due date.

10% to 20% of their repayments are overdue.

48% Inactive

51% has defaulted at least once.

The typical default rate is around **4% to 18%.**

57% of their loans resulted in debt.

34% have fallen behind the due date.

7% to 16% of their repayments are overdue.

Identifying our users' loan repayment behavior

Based on RFM modeling

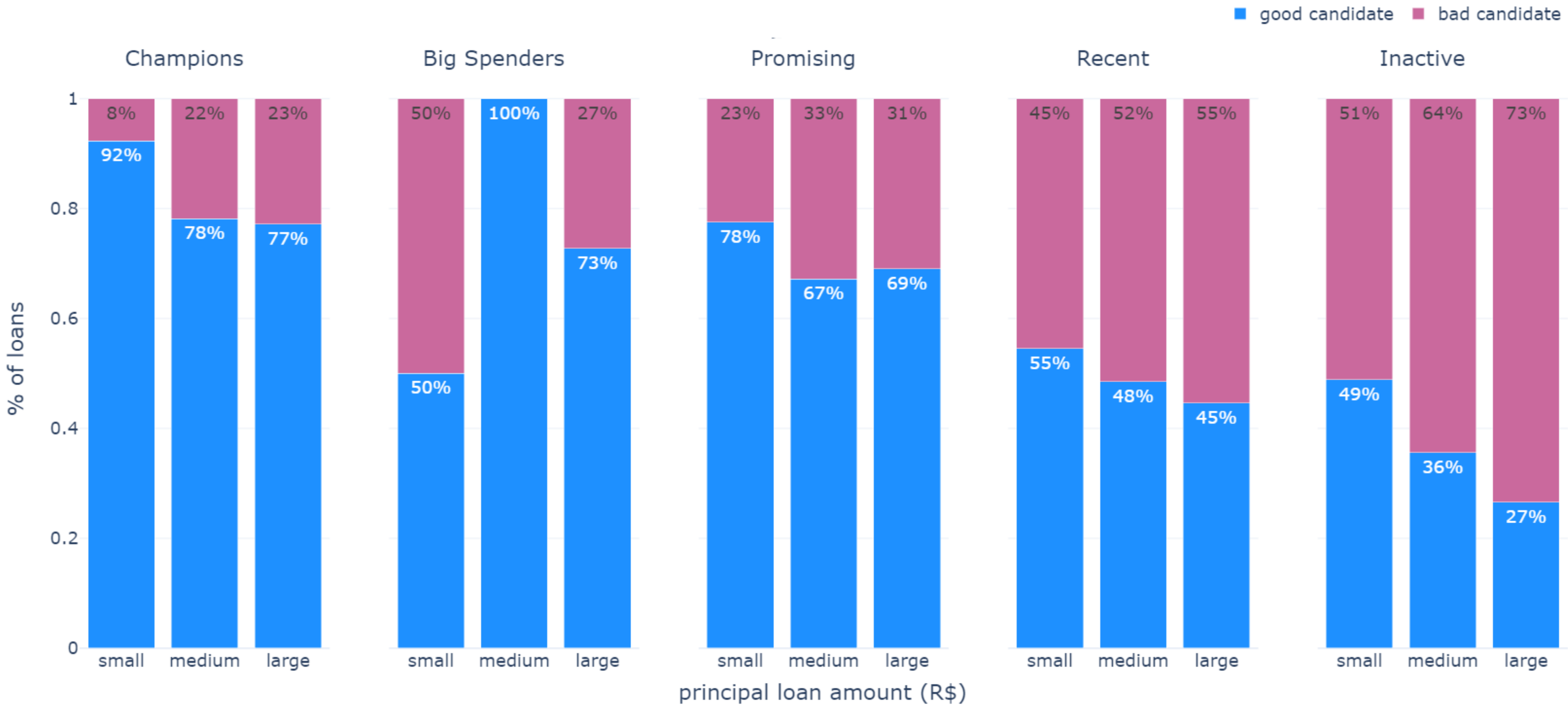
8% Champions	3% Big Spenders	35% Promising	6% Recent	48% Inactive
90% has defaulted at least once.	81% has defaulted at least once.	77% has defaulted at least once.	62% has defaulted at least once.	51% has defaulted at least once.
The typical default rate is around 2% to 5%.	The typical default rate is around 3% to 9%.	The typical default rate is around 3% to 6%.	The typical default rate is around 7% to 16%.	The typical default rate is around 4% to 18%.
2% of their loans resulted in debt.	2% of their loans resulted in debt.	11% of their loans resulted in debt.	21% of their loans resulted in debt.	57% of their loans resulted in debt.
17% have fallen behind the due date.	**10%** have fallen behind the due date.	**30%** have fallen behind the due date.	**41%** have fallen behind the due date.	**34%** have fallen behind the due date.
2% to 6% of their repayments are overdue.	**10% to 14%** of their repayments are overdue.	**3% to 12%** of their repayments are overdue.	**10% to 20%** of their repayments are overdue.	**7% to 16%** of their repayments are overdue.

Users with lower activity present a high-risk repayment behavior

Low activity is defined as low frequency and monetary values

Champions	Big Spenders	Promising	Recent	Inactive
22%	**27%**	**31%**	**54%**	**72%**
high-risk candidates	high-risk candidates	high-risk candidates	high-risk candidates	high-risk candidates

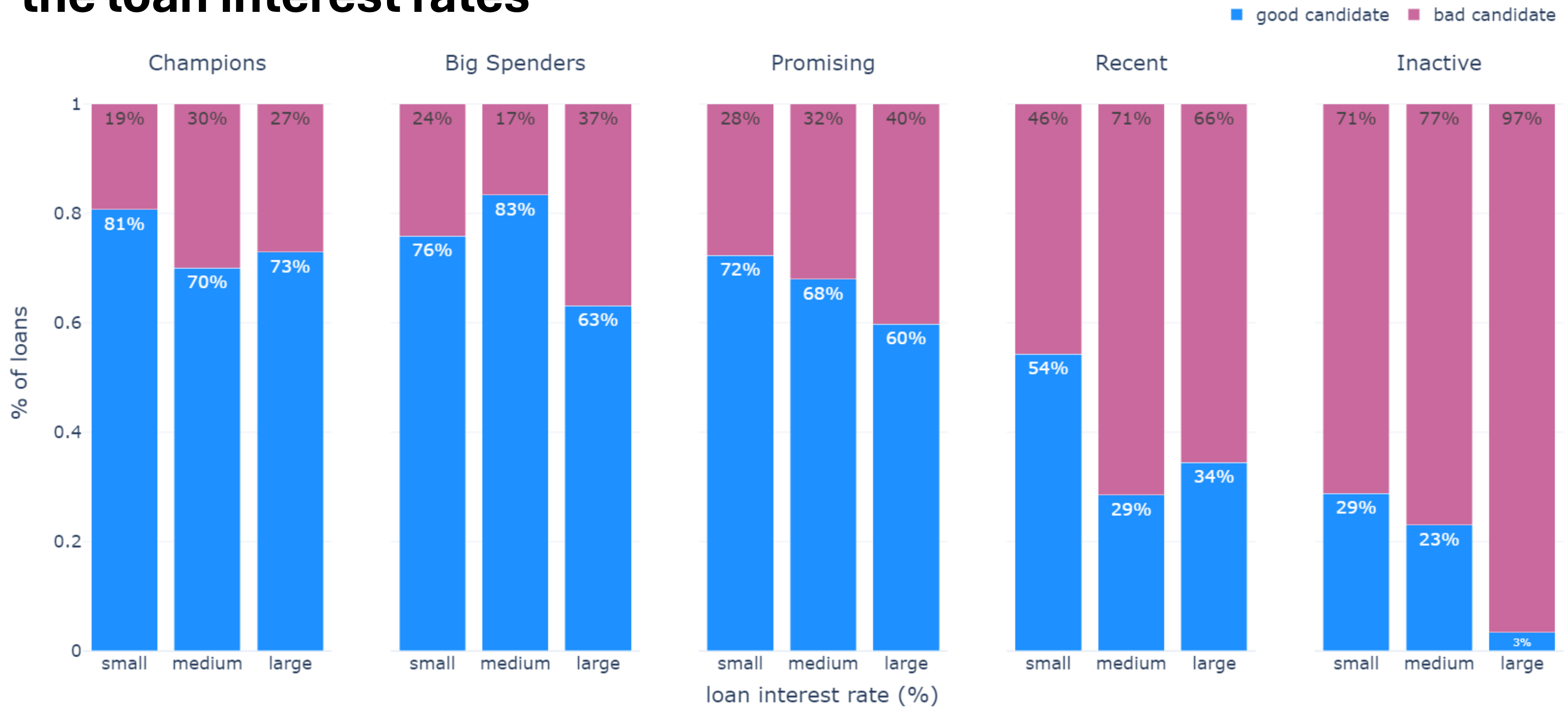
Risk **increases** with the loan amount, with *Recent* and *Inactive* users exhibiting a **high-risk repayment behavior**



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

Big Spenders have a limited number of small and medium sized loans

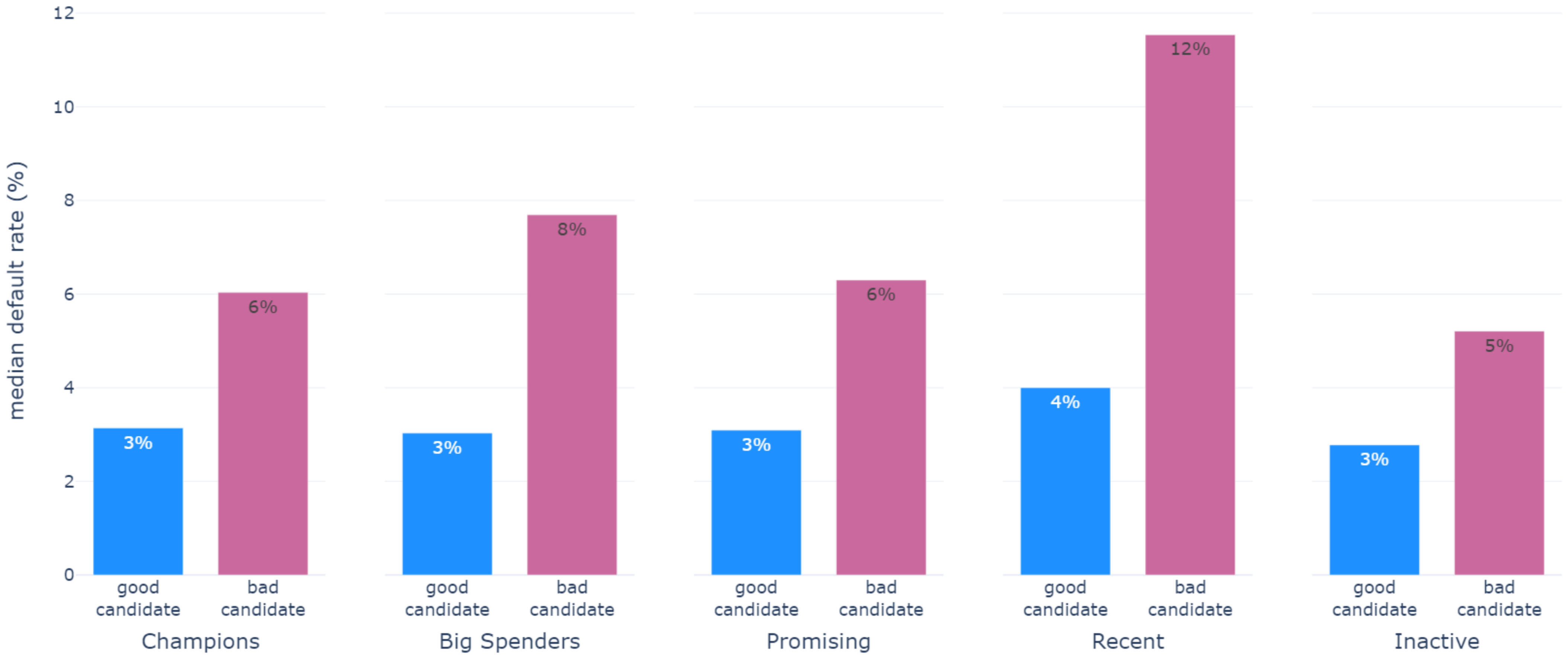
Similar risk can be observed in relation to the loan interest rates



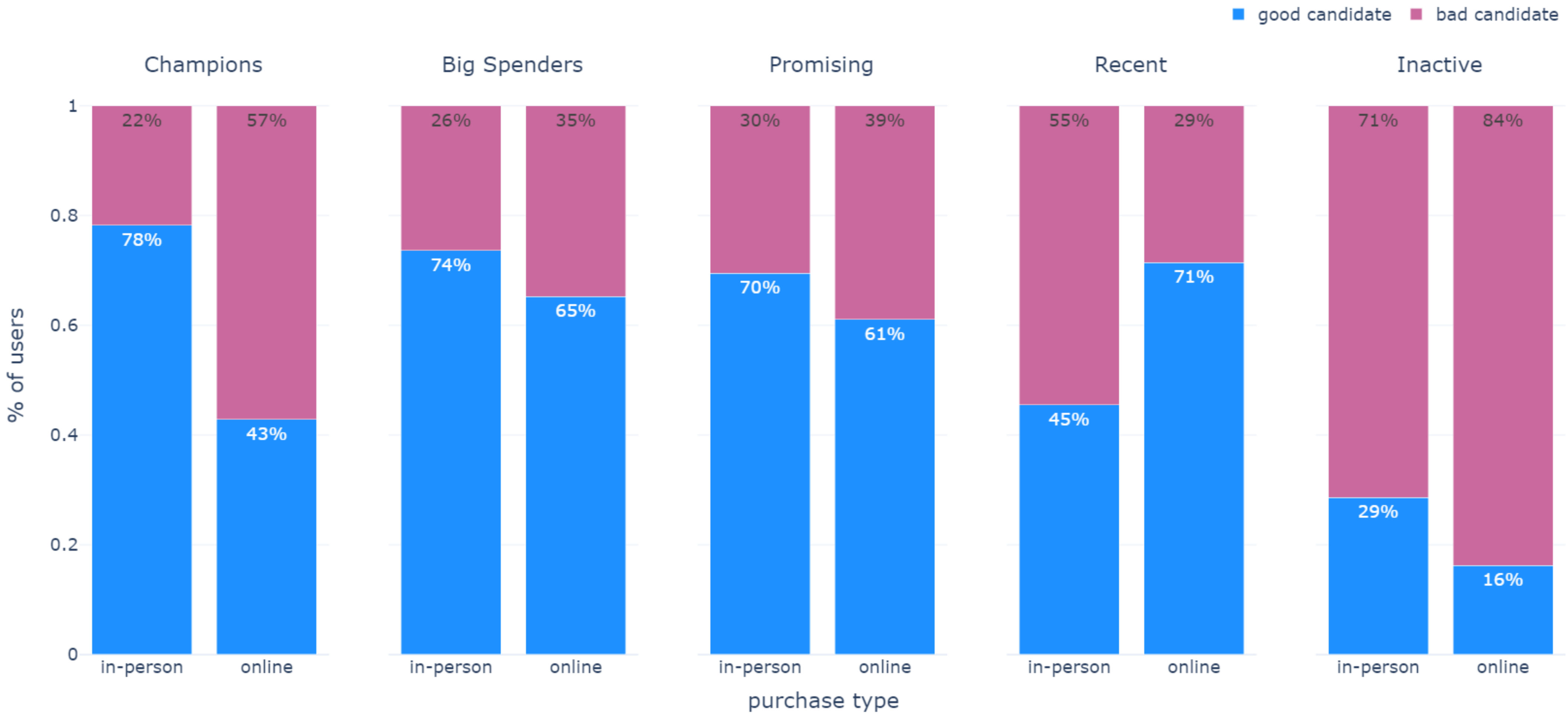
small: *less than 40%,* ***medium:*** *40% - 100% ,* ***large:*** *greater than 100%*

Inactive customers have a limited number of loans with large interest rates

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



Online spending habits can indicate high-risk repayment behavior, regardless of the user segment



Recent customers have a limited number of online transactions

The model identifies 56% of the risk loan candidates, however, with low confidence levels

Results on the test set

	Precision	Recall	F1 score
good candidate	69%	72%	70%
bad candidate	59%	56%	57%
overall	65%	65%	65%

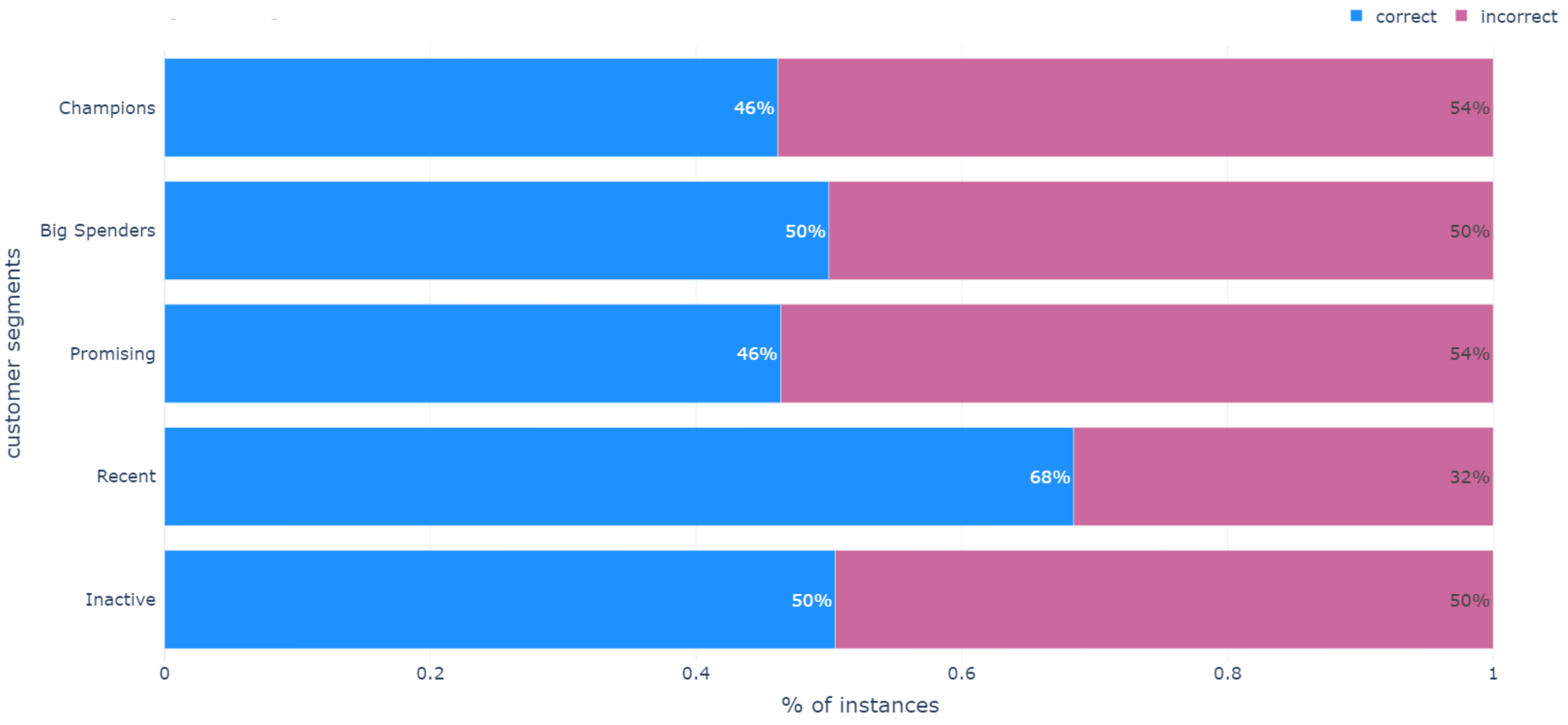
weighted average Precision, Recall, F1 score regarding LightGBM

Top predictive factors include

- **user purchasing behavior** (preference of credit over debit, transaction rejection rate, installment preference)
- **user repayment behavior** (default rate, preference of manual repayments, late repayments ratio, repaid loans ratio)

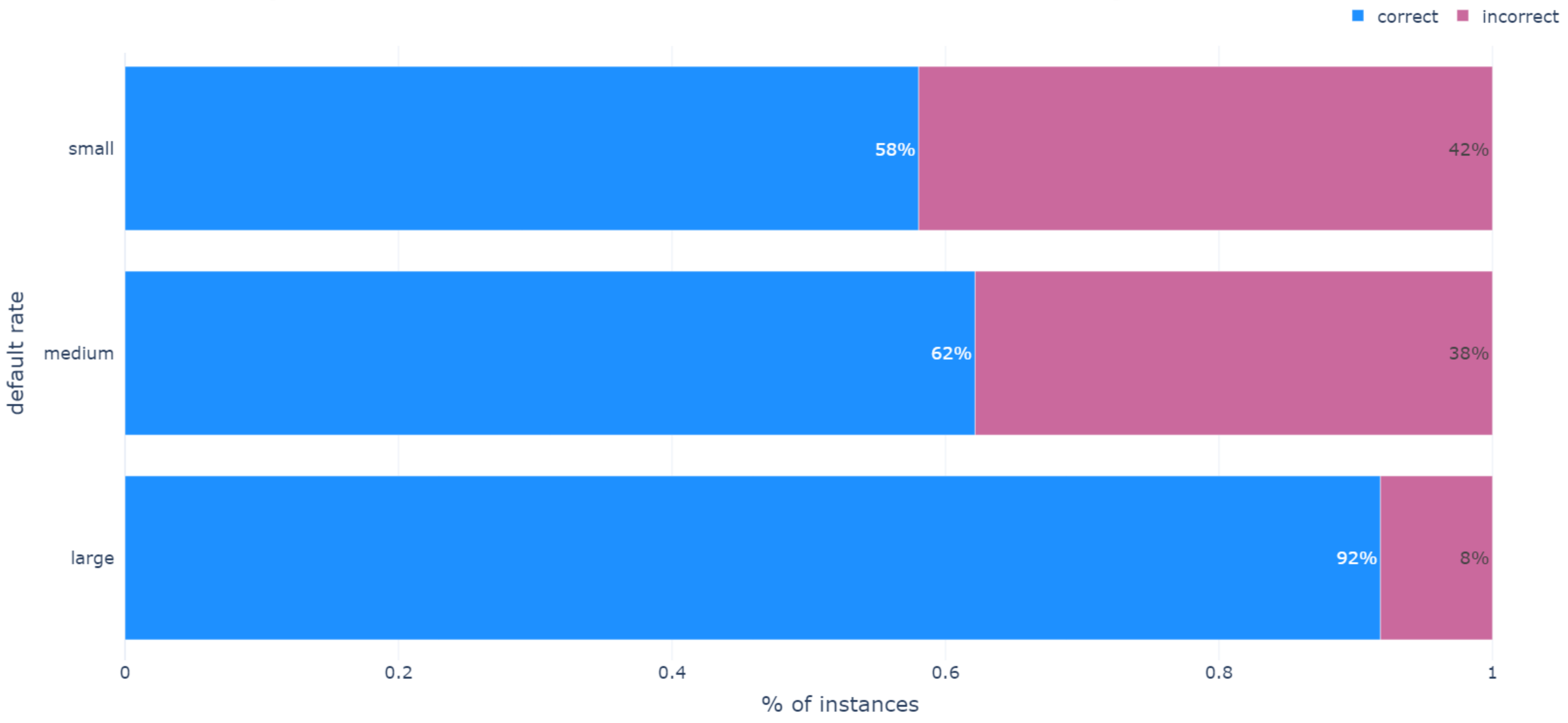
Inability to link the segments' purchasing behavior with repayment behavioral patterns

Bad candidates in the test set



The model cannot capture mainly the bad loan candidates with small to medium default rates

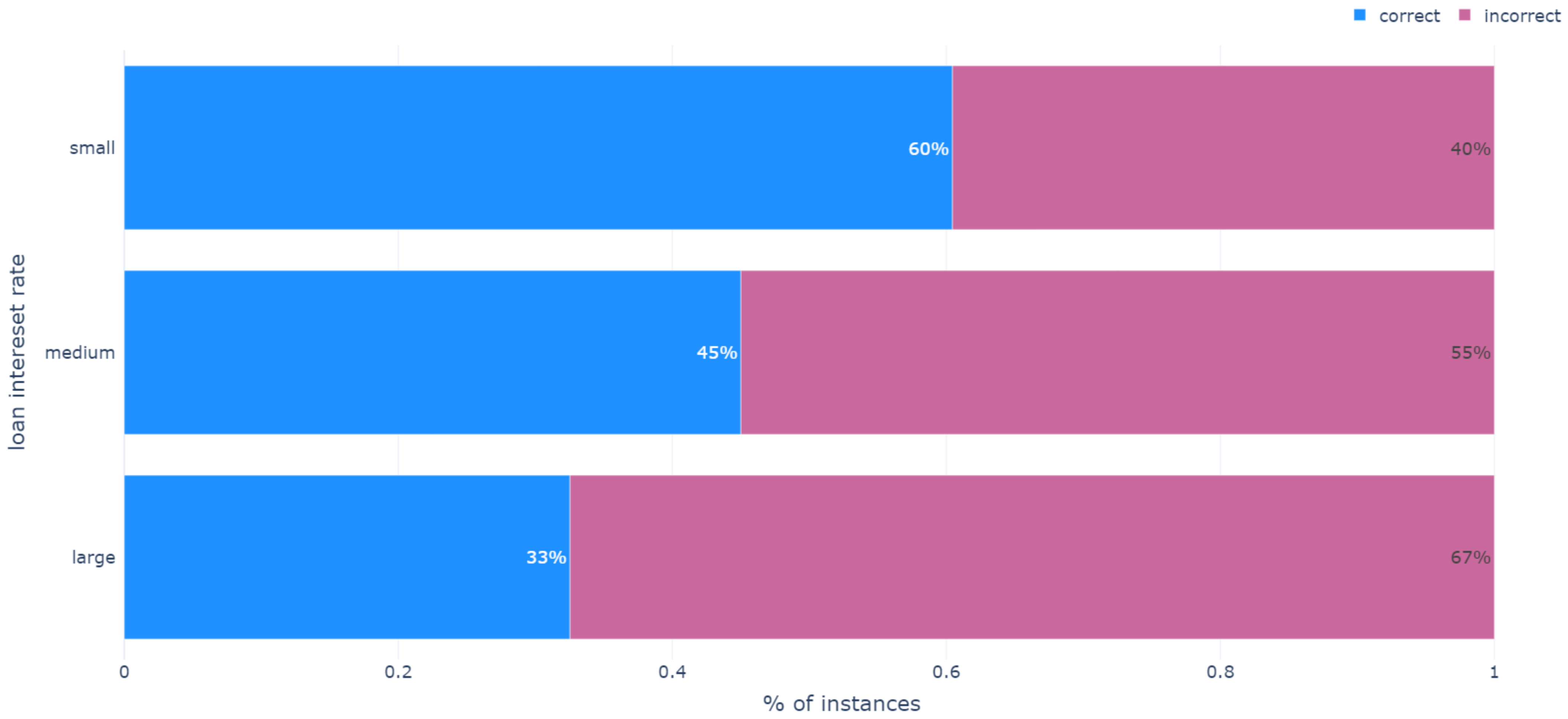
Bad candidates in the test set



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

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

Bad candidates in the test set



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

Next Steps and Considerations

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

- Transition from RFM customer analysis, which segments users based on their purchasing habits, to a more refined **clustering** method to model customer behavioral patterns.

- Inferring the new users' repayment behavior from the Recent customer segment, doesn't fully solve the **cold start problem**. **User embeddings** can be used for similarity-based inference and can be progressively learned.

- **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

Tel/WhatsApp: +30 6945261838
Email: lamproslntz@outlook.com
LinkedIn: https://www.linkedin.com/in/lamproslntz/

QnA Session

cloudwalk