Case Study
Loan Repayment Analysis

Lampros Lountzis

Data Scientist

April 12th, 2024

cloudwalk

Agenda

Identifying the Challenge

Factors Providing

Key Insights

Loan Repayment Model and Analysis

Next Steps

5%

of loan repayments are defaulted,
resulting in

58%

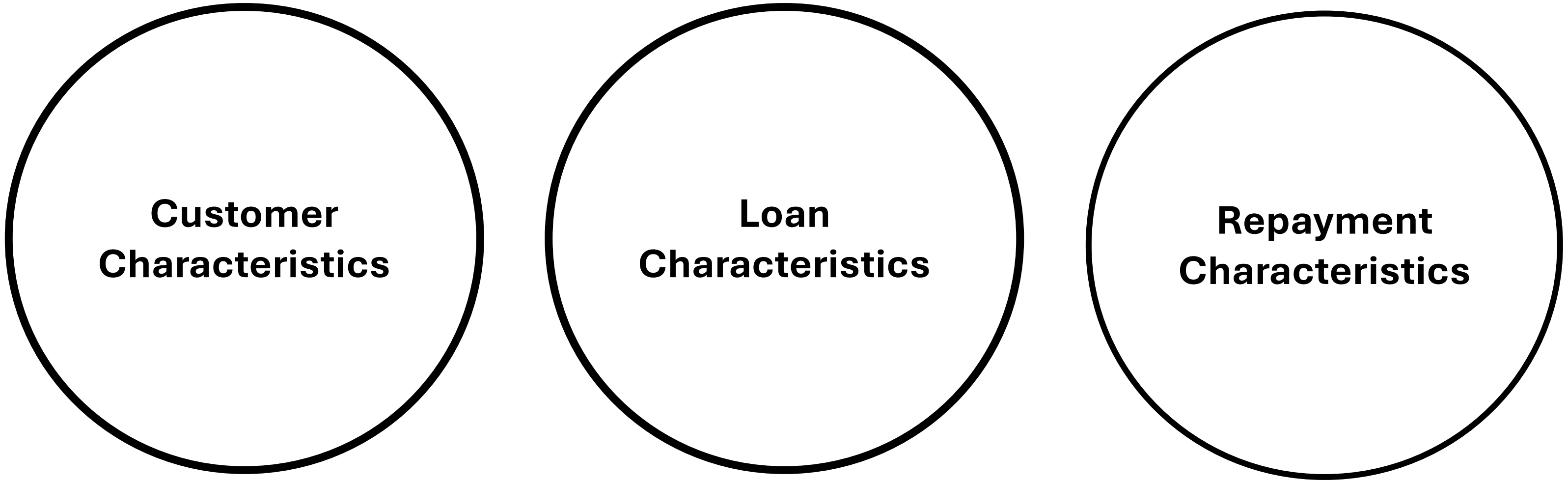
of loans with at least one defaulted repayment ,
causing lenders to incur financial loss and increased administrative burdens.

76%

of customers who have defaulted at least once,
causing damage to their credit score and additional fees and interest.

Based on 3,048 users who generated 6,598 loans totaling 172,445 loan repayments

Linking Loan and Customer behavior



Identifying our user base

Based on RFM model

8%
Champions

Extremely active with moderate to high monetary value.

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.

Higher transaction rejection rate (27%).

35%
Promising

Active customers with low to moderate monetary value.

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.

Higher transaction rejection rate (19%).

48%
Inactive

Customers with extremely low activity (monetary value isn't a factor here).

Prefer to use mostly credit cards.

Opt frequently for installment plans.

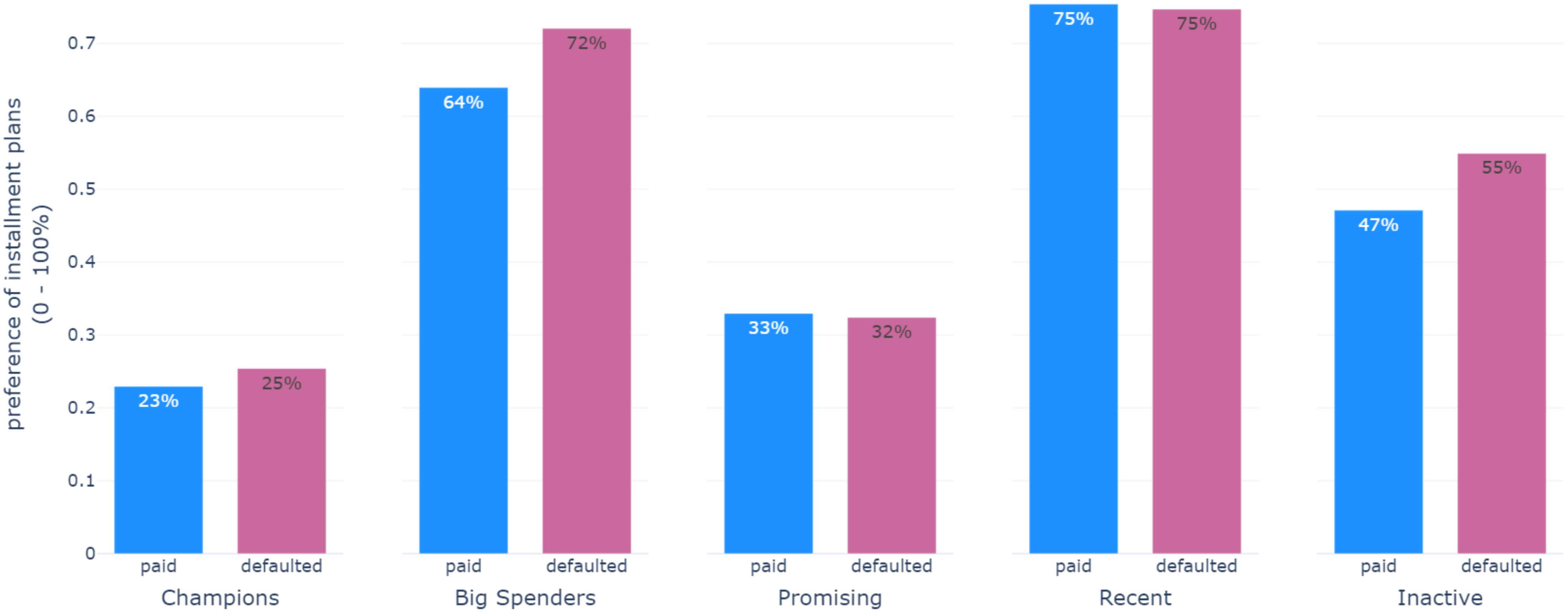
Higher transaction rejection rate (16%).

Users with lower activity present a higher risk of default

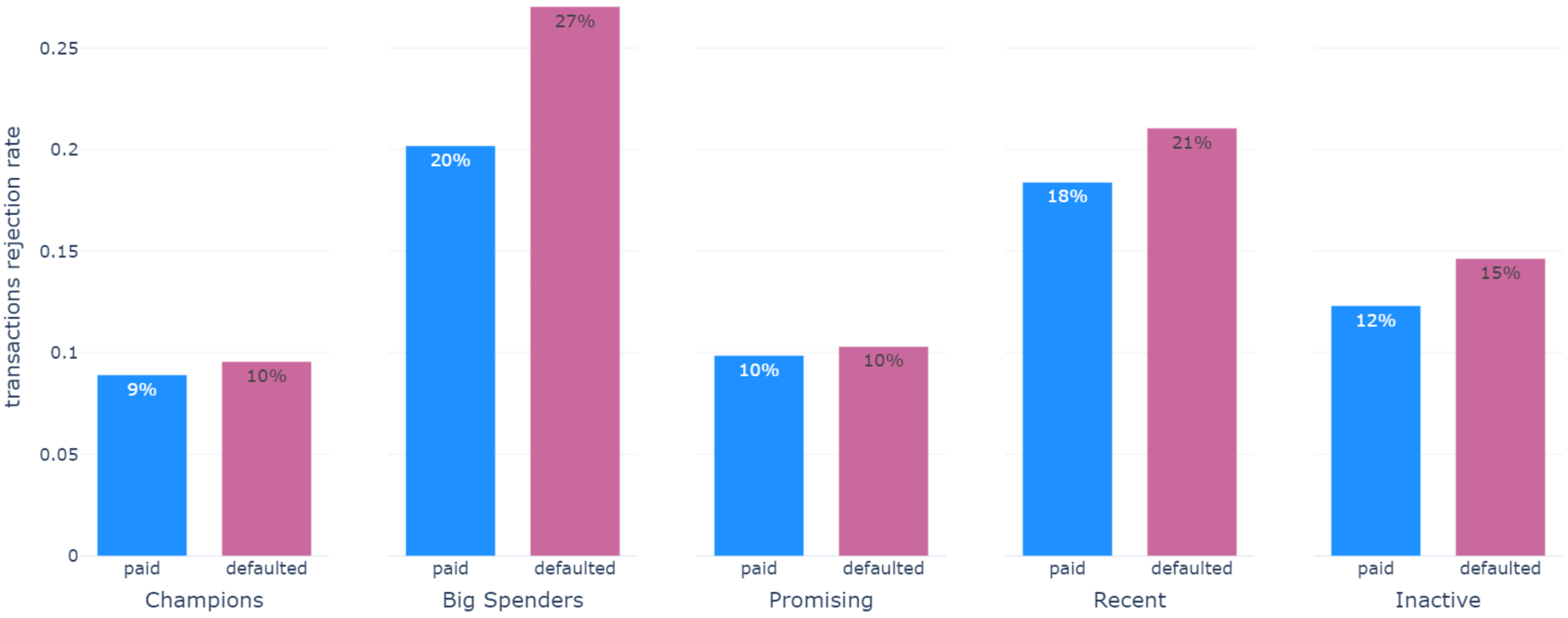
Low activity is defined as low frequency and monetary values

Champions	Big Spenders	Promising	Recent	Inactive
4%	5%	5%	10%	6%
defaulted repayments	defaulted repayments	defaulted repayments	defaulted repayments	defaulted repayments

Big Spenders and *Inactive* users who default, opt for installment plans more frequently

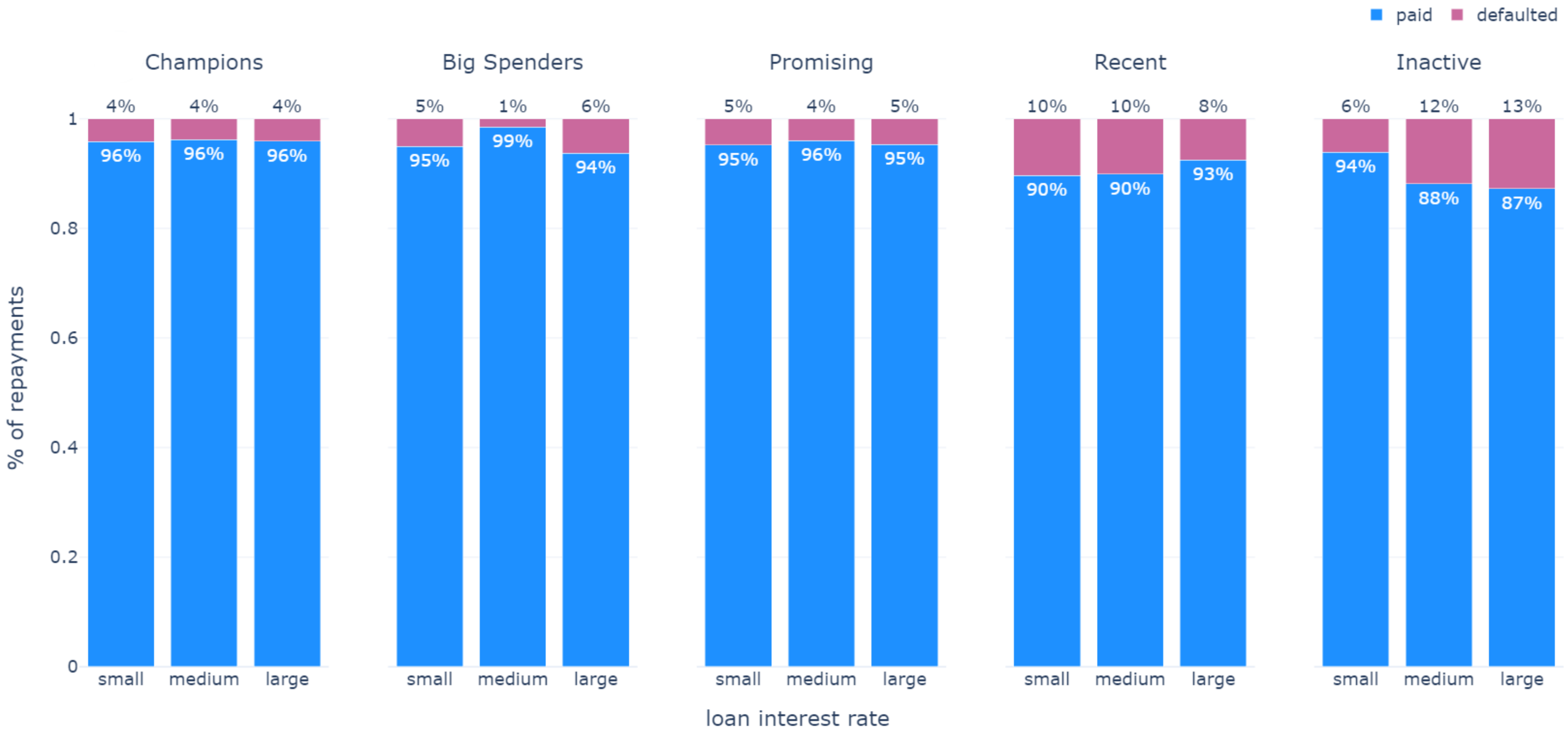


Big Spenders, *Recent* and *Inactive* users who default, have a higher transaction rejection rate than reliable payers



Inactive users are likely to default on loans with higher interest rate

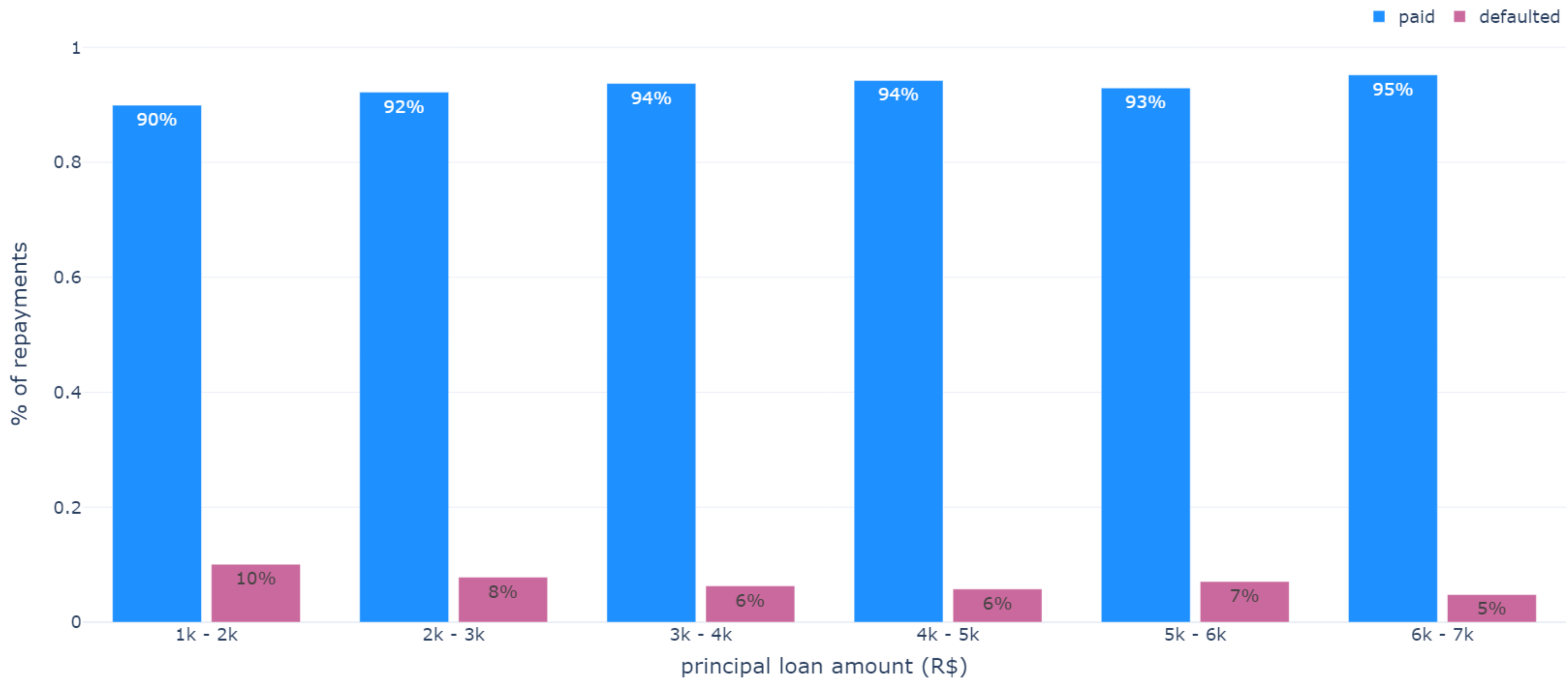
Recent users pose a high risk irregardless of the interest rate



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

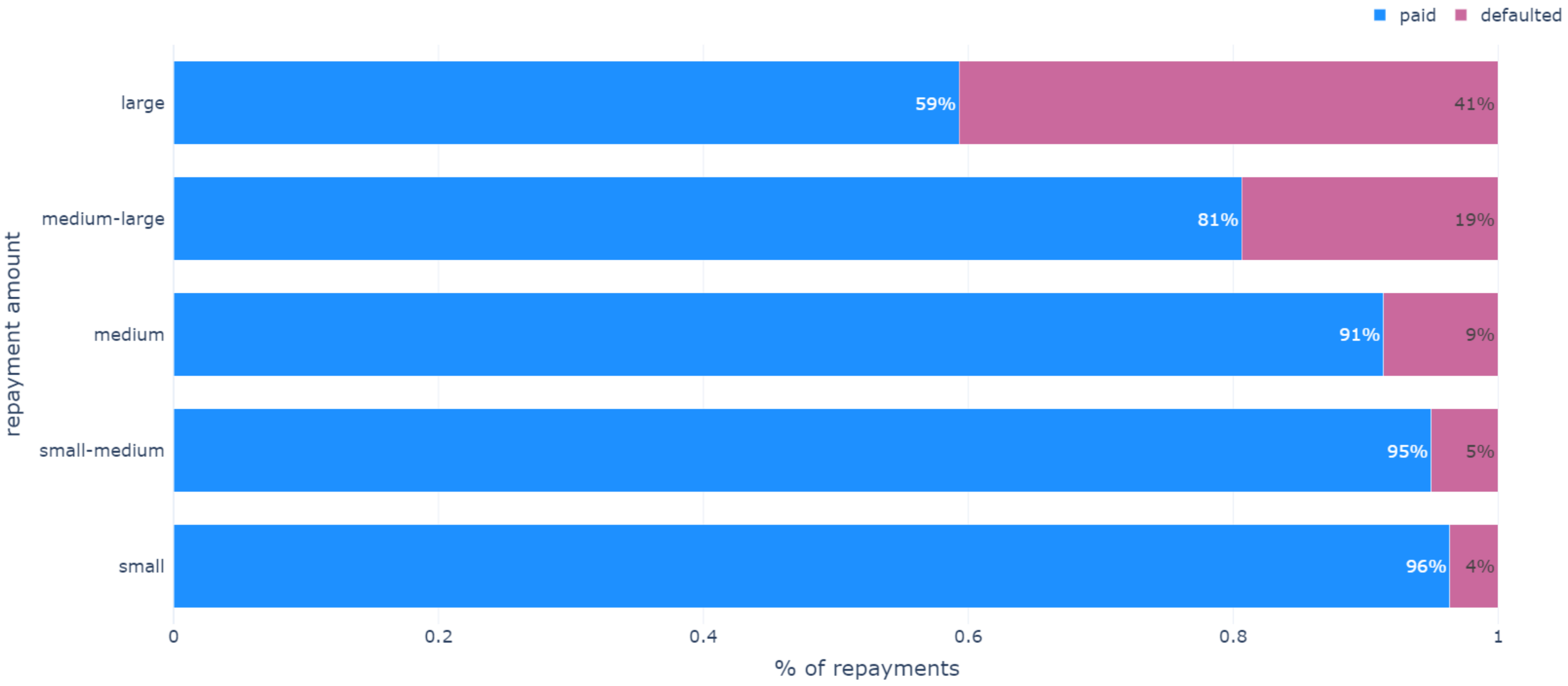
Big Spenders and *Recent* users pose the highest default risk when they surpass the loan due date

Smaller loans have higher default rates, attributed to riskier user segments



Recent and Inactive users tend to take out smaller loans compared to other segments

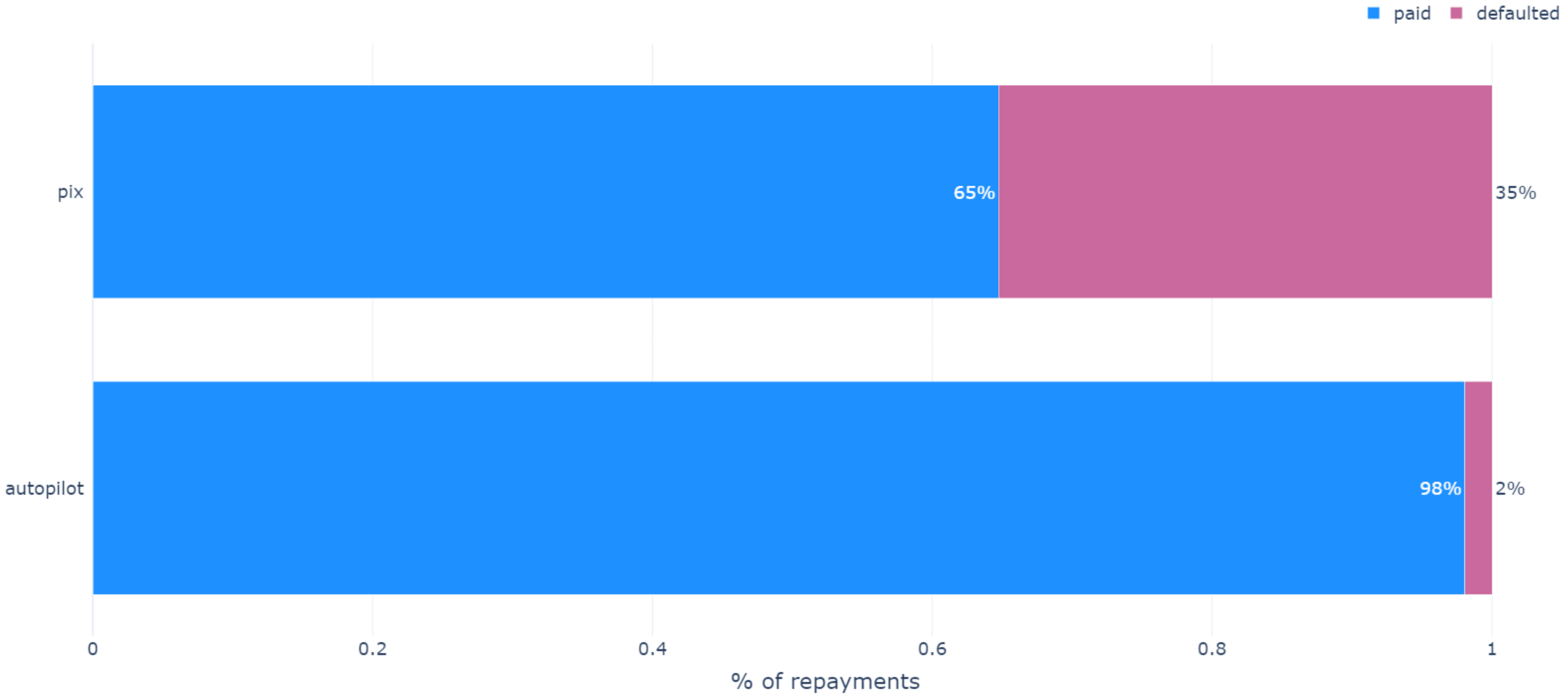
Larger loan repayment amounts have a higher potential of default



small: *less than R$250,* ***small-medium:*** *R$250 - 500,*
medium: *R$500 – 1,000* ***medium-large:*** *R$1,000 – 2,500* ***large:*** *greater than R$2,500*

Prior repayment defaults increase the risk of subsequent defaults

Users who opt for manual repayments are more likely to default



The model is certain when identifying defaulted repayments, but struggles with identifying all of them

Results of the test set

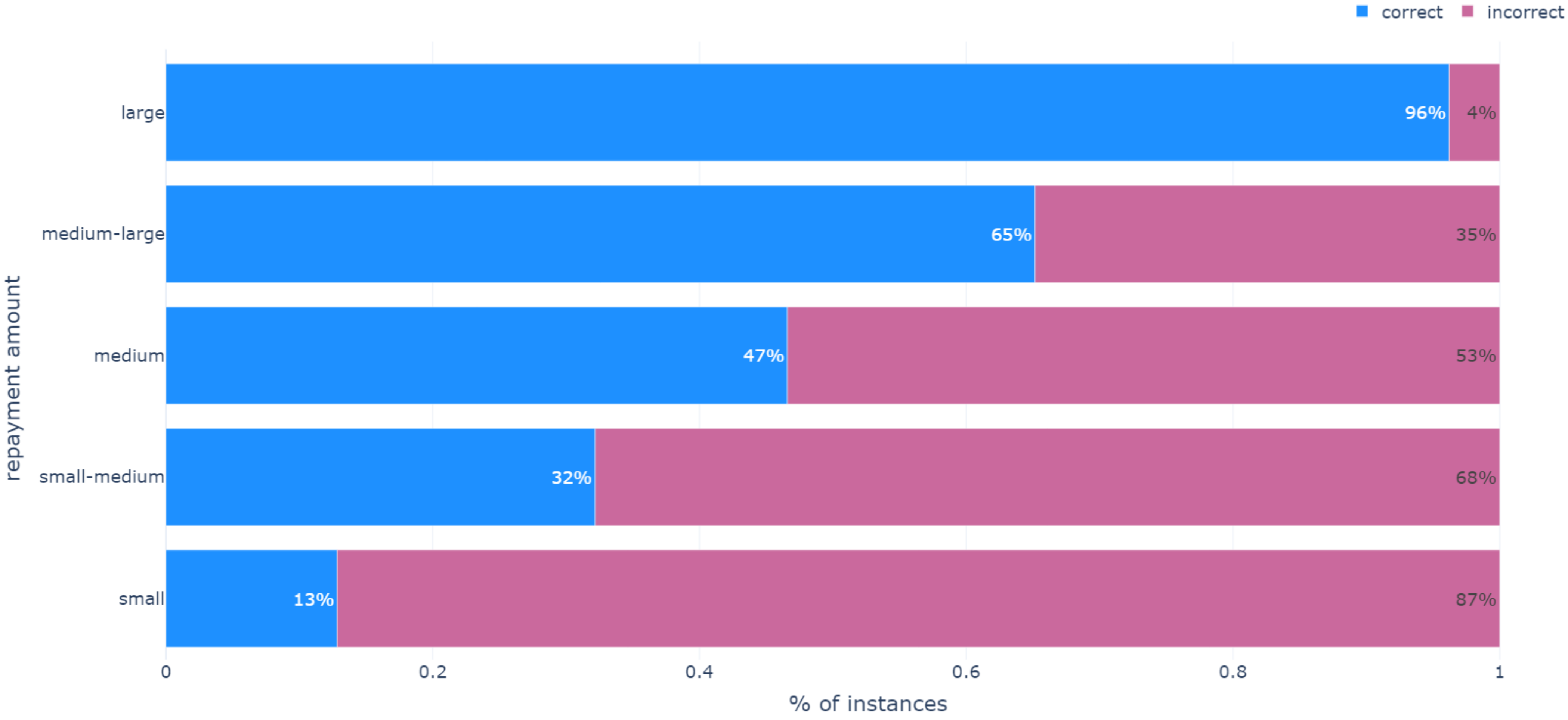
	Precision	Recall	F1 score
paid	96%	100%	98%
defaulted	88%	33%	47%
weighted avg	96%	96%	95%

Top predictive factors include

- **repayment characteristics** (current and prev. repayment amount, days passed since loan creation, pct. of loan repaid up to date)
- **user characteristics** (preference of credit over debit, transaction rejection rate)

The model is unable to distinguish the defaulted repayments for small to medium repayments

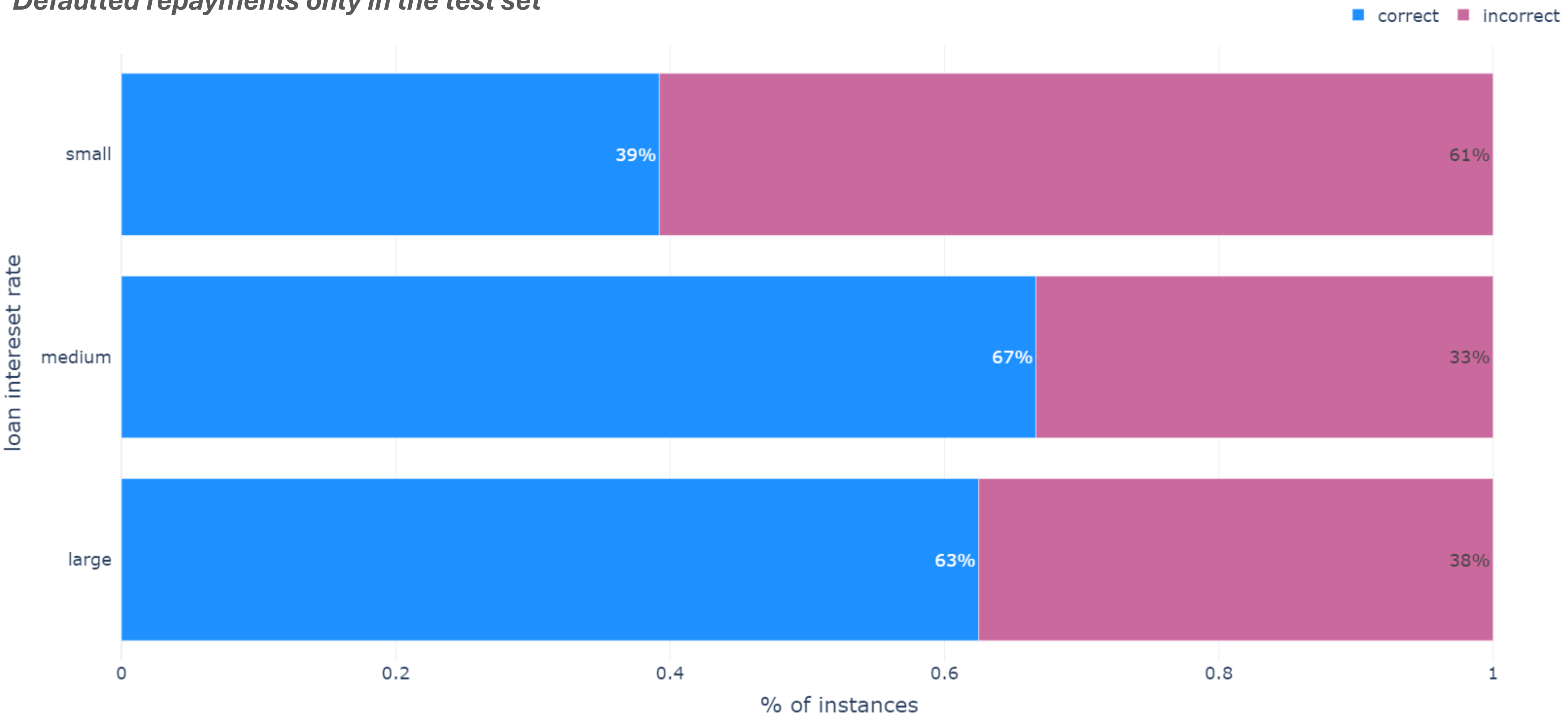
Defaulted repayments only in the test set



small: *less than R$250,* ***small-medium:*** *R$250 - 500,*
medium: *R$500 – 1,000* ***medium-large:*** *R$1,000 – 2,500* ***large:*** *greater than R$2,500*

The model cannot capture the defaulted repayments’ patterns mainly for small loan interest rates

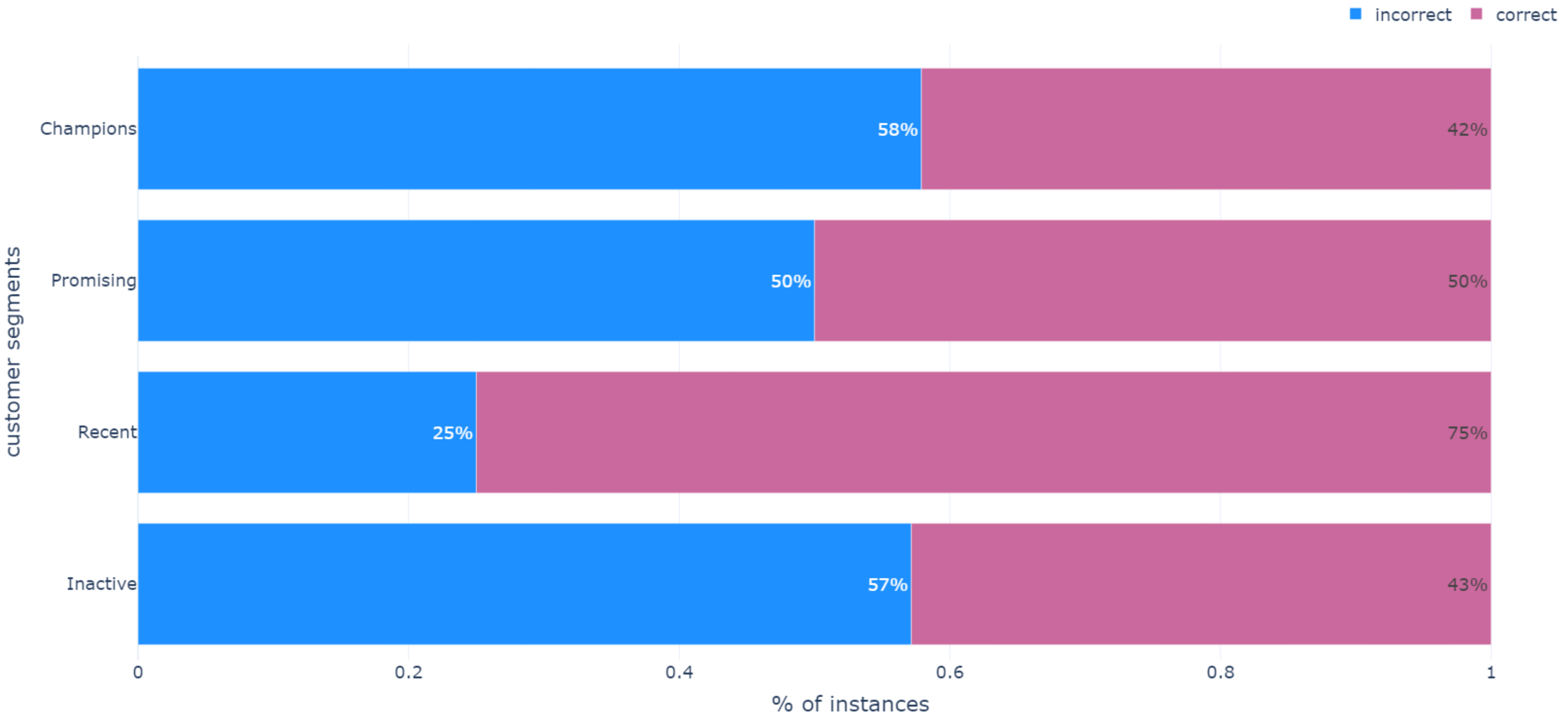
Defaulted repayments only in the test set



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

Inability to model the default patterns of the *Recent* users

Defaulted repayments only in the test set



Next Steps

- **Expand the set of features** to include user demographics (e.g., state, credit history, etc.) and other various loan characteristics (e.g., loan usage, loan type, etc.).

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

- **Oversampling** by creating synthetic loan repayment records that adhere to the patterns observed in authentic loan repayment data. The addition of synthetic data shouldn't skew or change the observed patterns.

Thank You!

Lampros Lountzis
Data Scientist

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

QnA Session

cloudwalk