



# Predicting Art Purchase and User Engagement

Converting users for an Art Startup

Paulo Arantes  
[paulo.espeschite@gmail.com](mailto:paulo.espeschite@gmail.com)  
[linkedin.com/in/espeschite](https://linkedin.com/in/espeschite)  
[github.com/pauloarantes/art-project](https://github.com/pauloarantes/art-project)

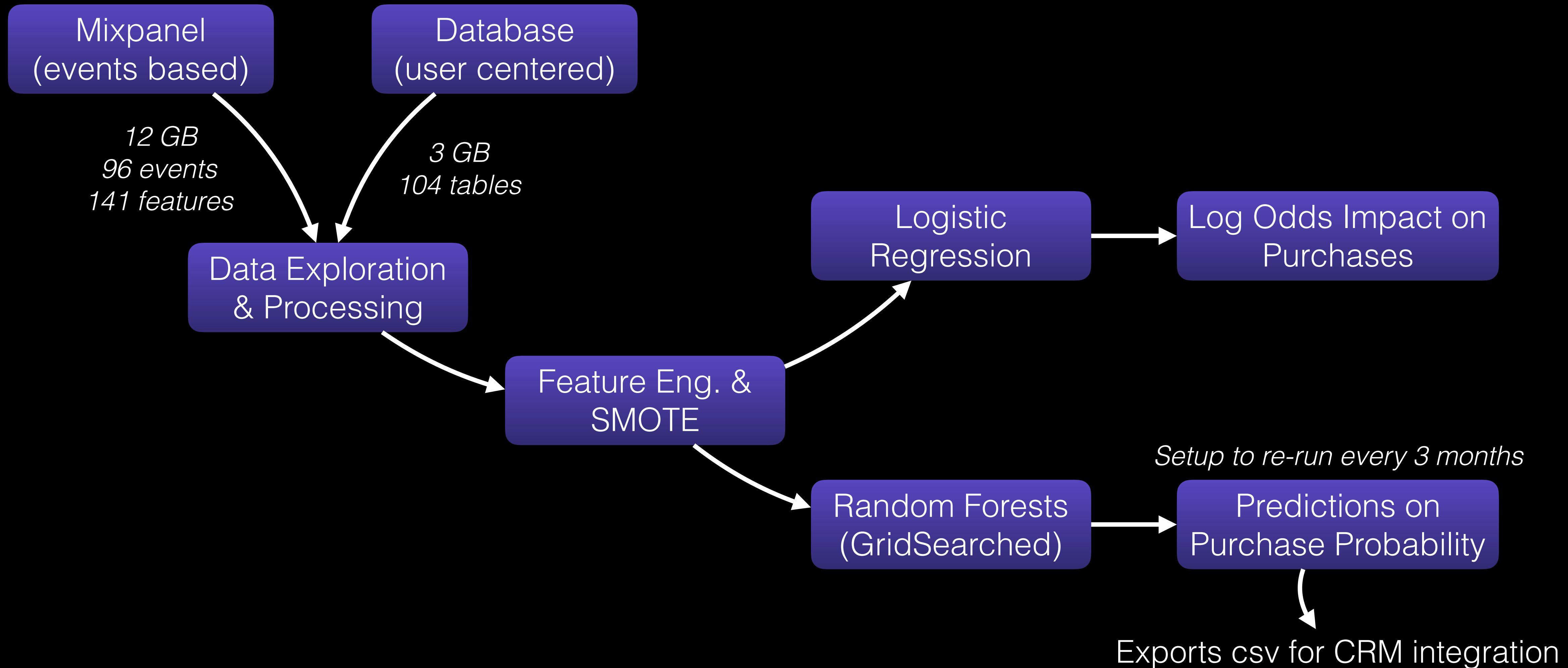
# Working with real startup data



# Challenges

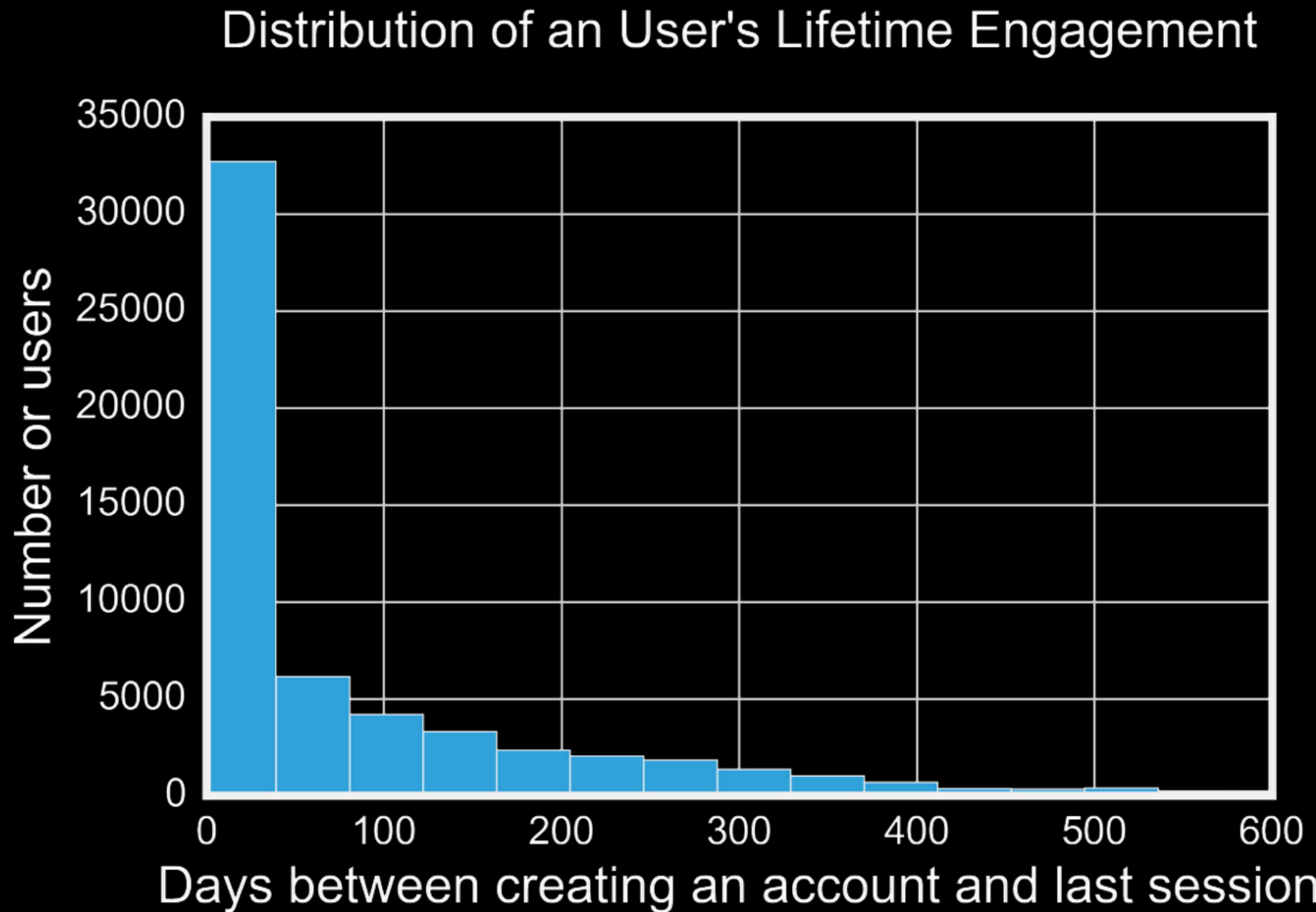
- When's the best timing to engage users?
- Can we predict such an abstract purchase?
- What are purchase habits like?
- How much money can we make?

# Pipeline



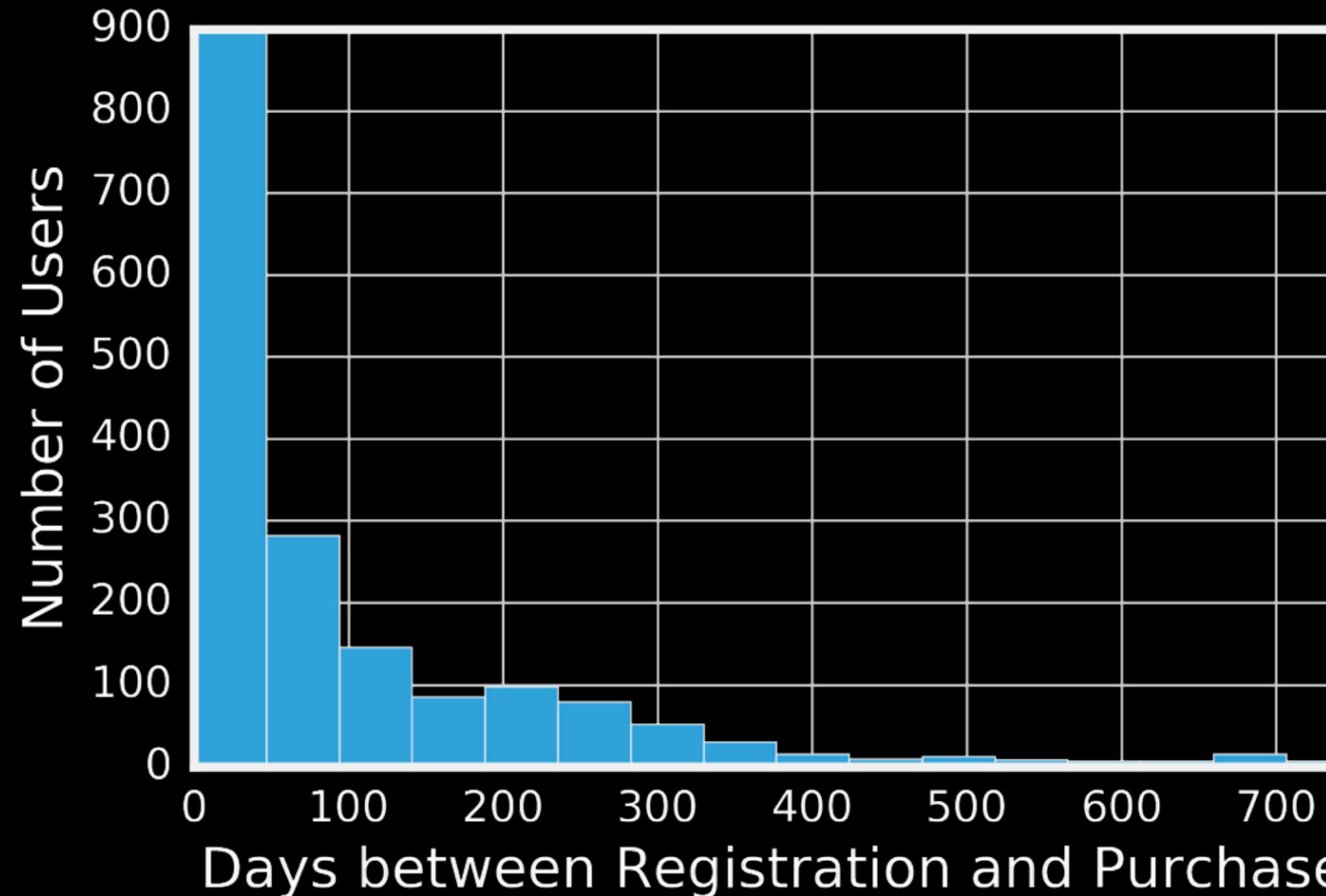
# Findings

# We have to act fast!



# Purchase cycle

Days since Registration to make First Purchase

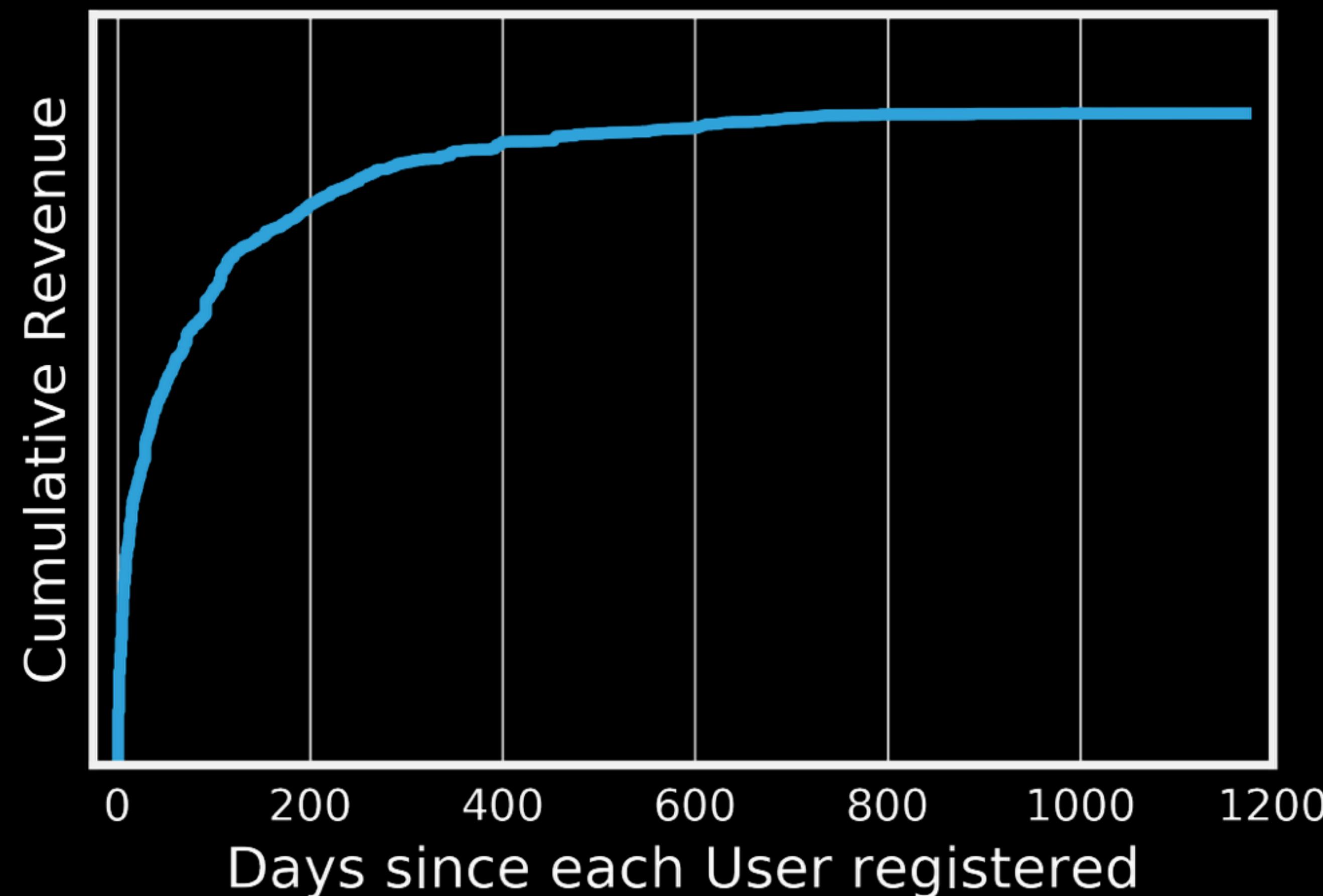


Average time since registration for an user to make the first purchase:  
**103 days**

**90%** of users only make **one** purchase

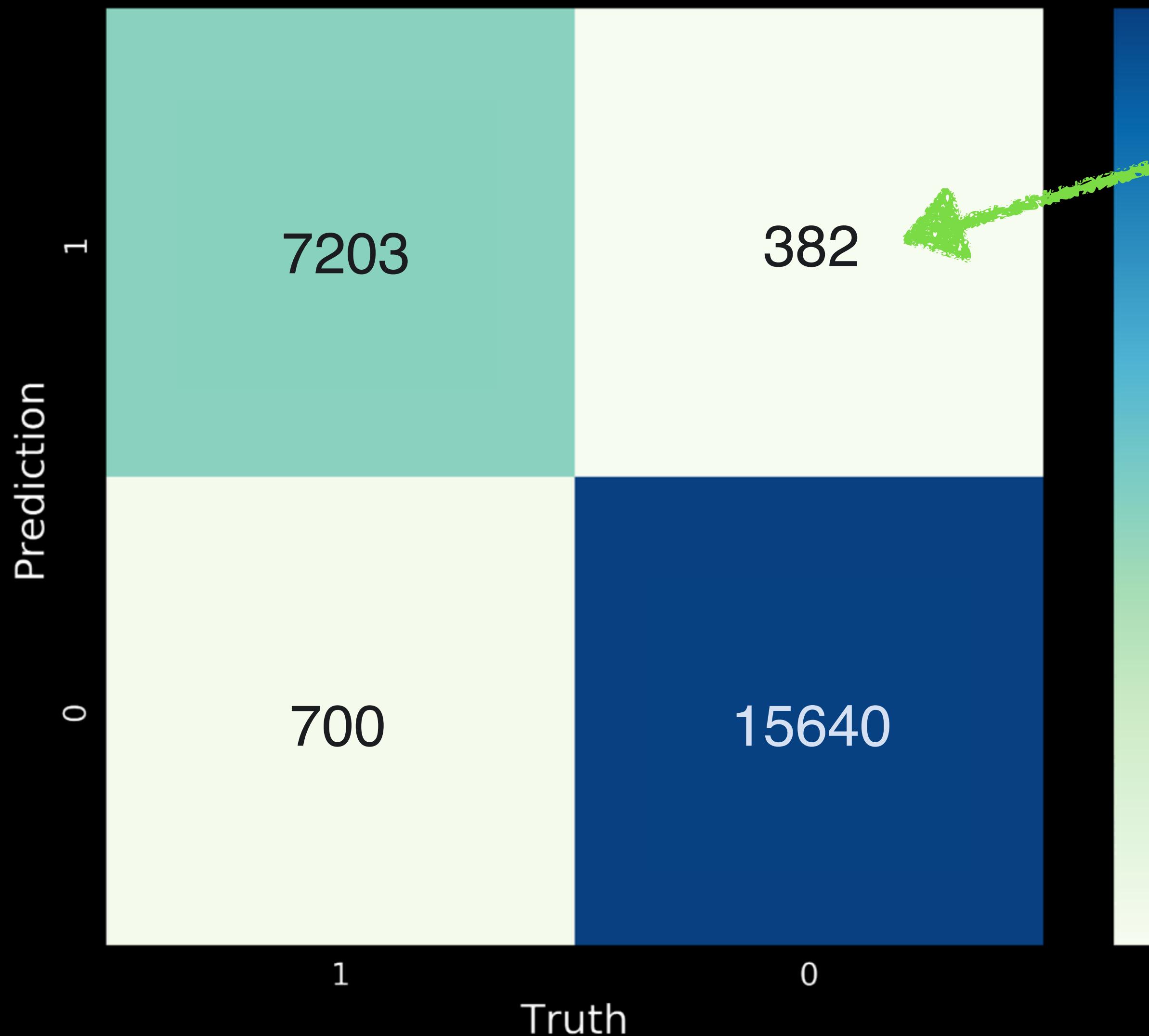
# First 120 days = 80% revenue

Cumulative Revenue centering all users on their Registration Day



# Model for Predicting Purchase Probability

Confusion Matrix - Random Forests



Minimize False Positives!

*It means we predicted purchase, spent money to convert, and they didn't*

*Cross Validation Scores*

accuracy: 0.954

precision: 0.949

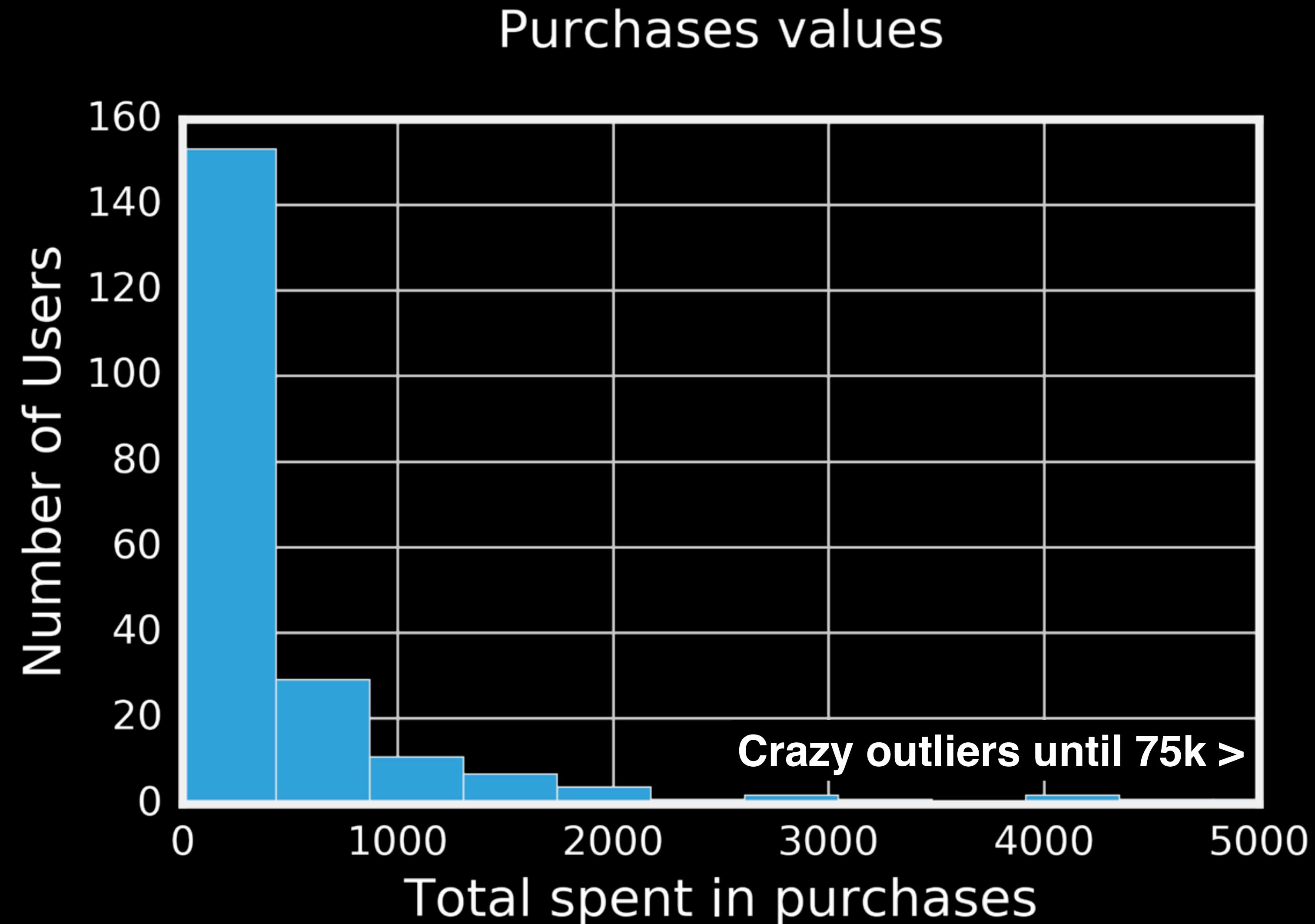
recall: 0.911

# How much do they spend?

Mean purchase value:  
\$ 515

↓

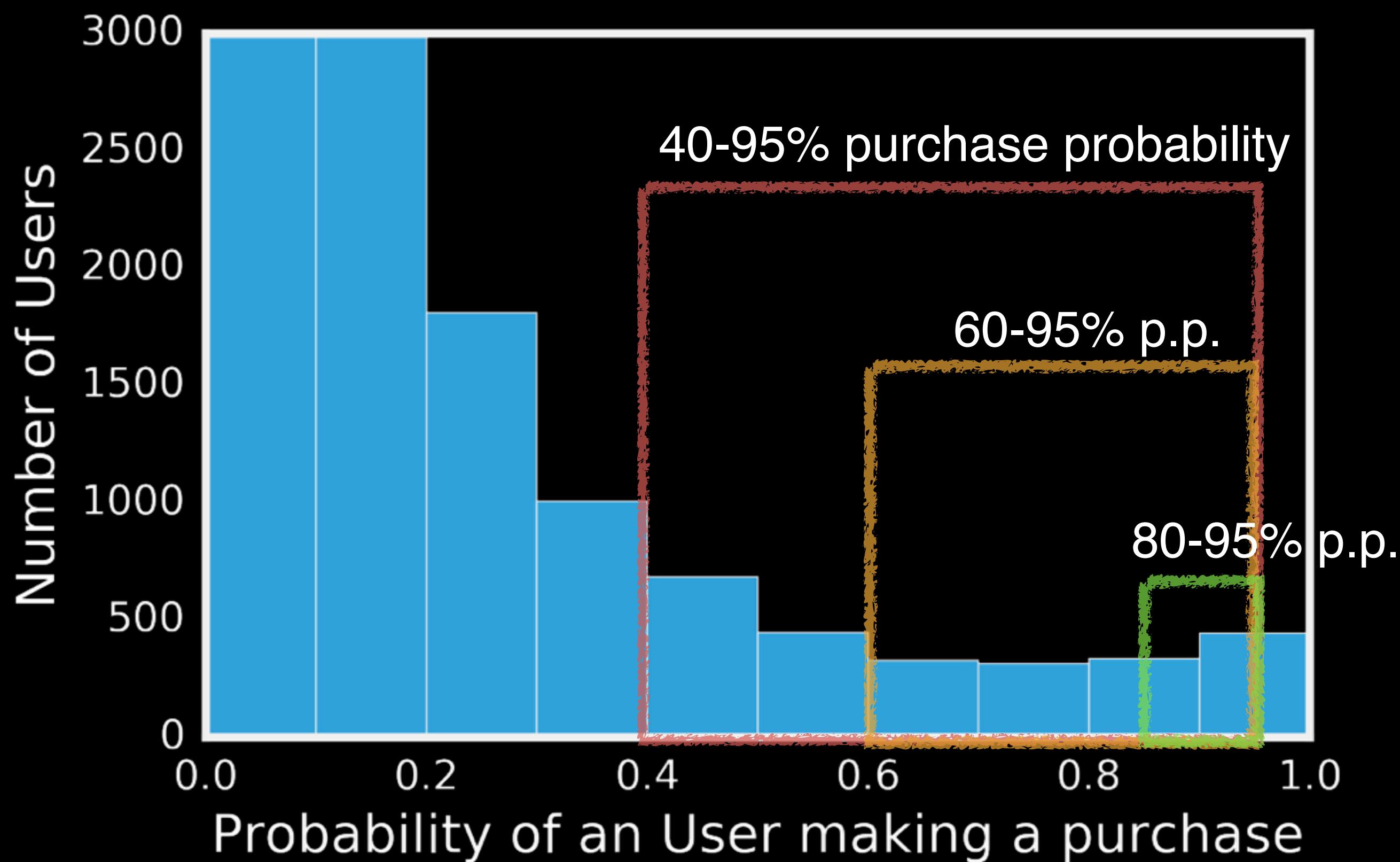
**Median purchase value:**  
\$ 250  
**(less sensitive to outliers)**



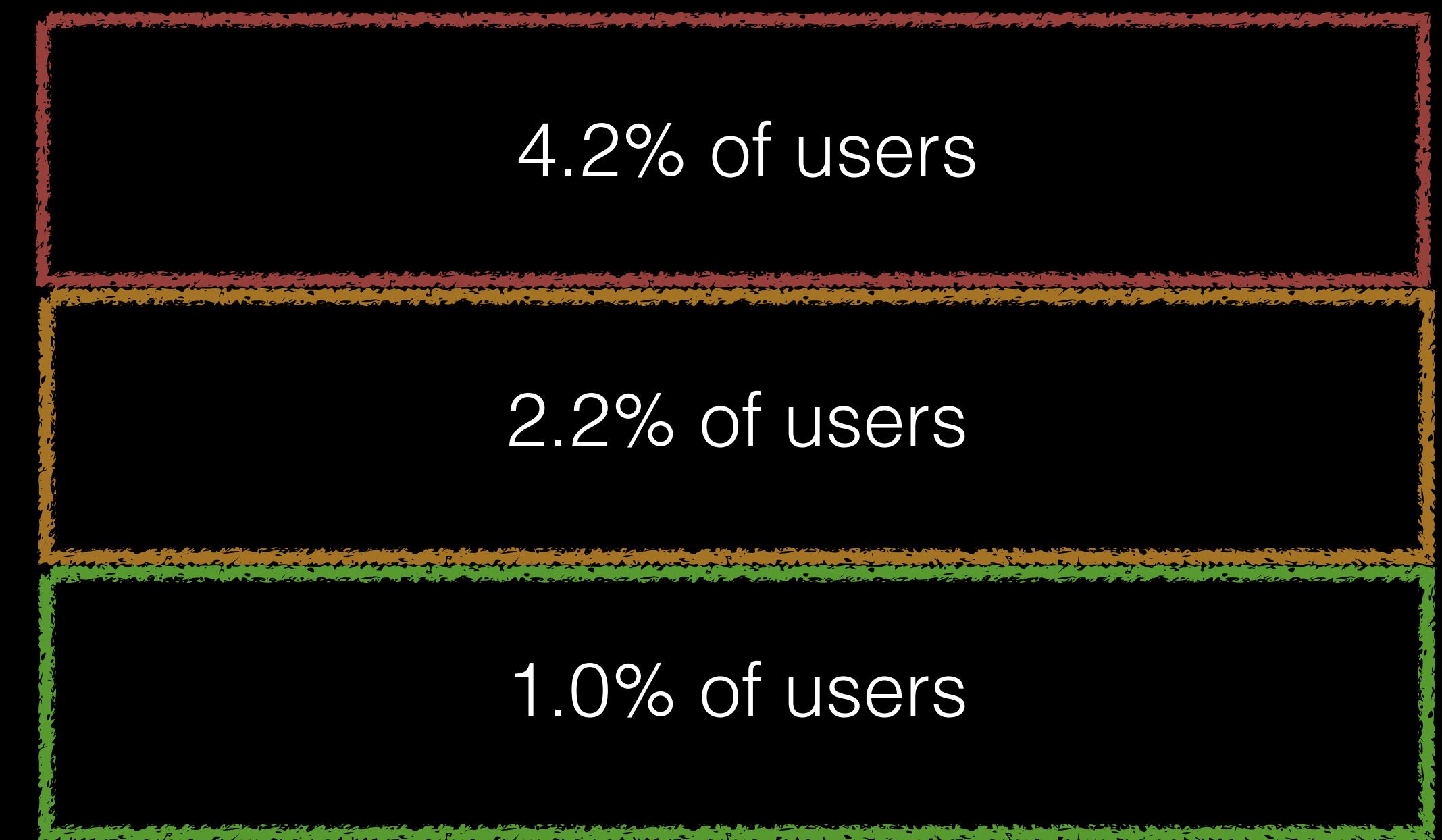
# Recommendations

# Who should we target?

Predictions distribution on original dataset  
[close-up]



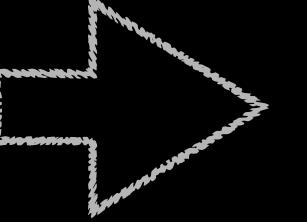
*Scenarios:*

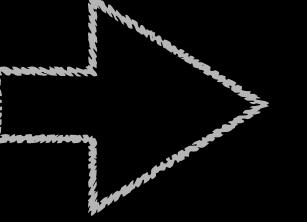


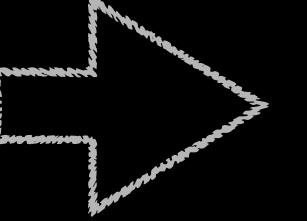
\* There will be more costs to convert users with lower probabilities, and the potential considers full conversion

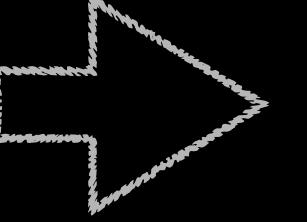
# What should we focus on?

Odds Ratio from Logistic Regression

Number of Sessions + 1            + 4.14% Purchase

Artists Followed + 1            + 11.43% Purchase

Artists Favorited + 1            + 0.95% Purchase

Artwork Shared + 1            + 8.65% Purchase

# What should we focus on?

To **double\*** the chance of purchase, an user should:

~~Have 142 more sessions~~

~~Favorites 149 artworks~~

Share 24 artworks 

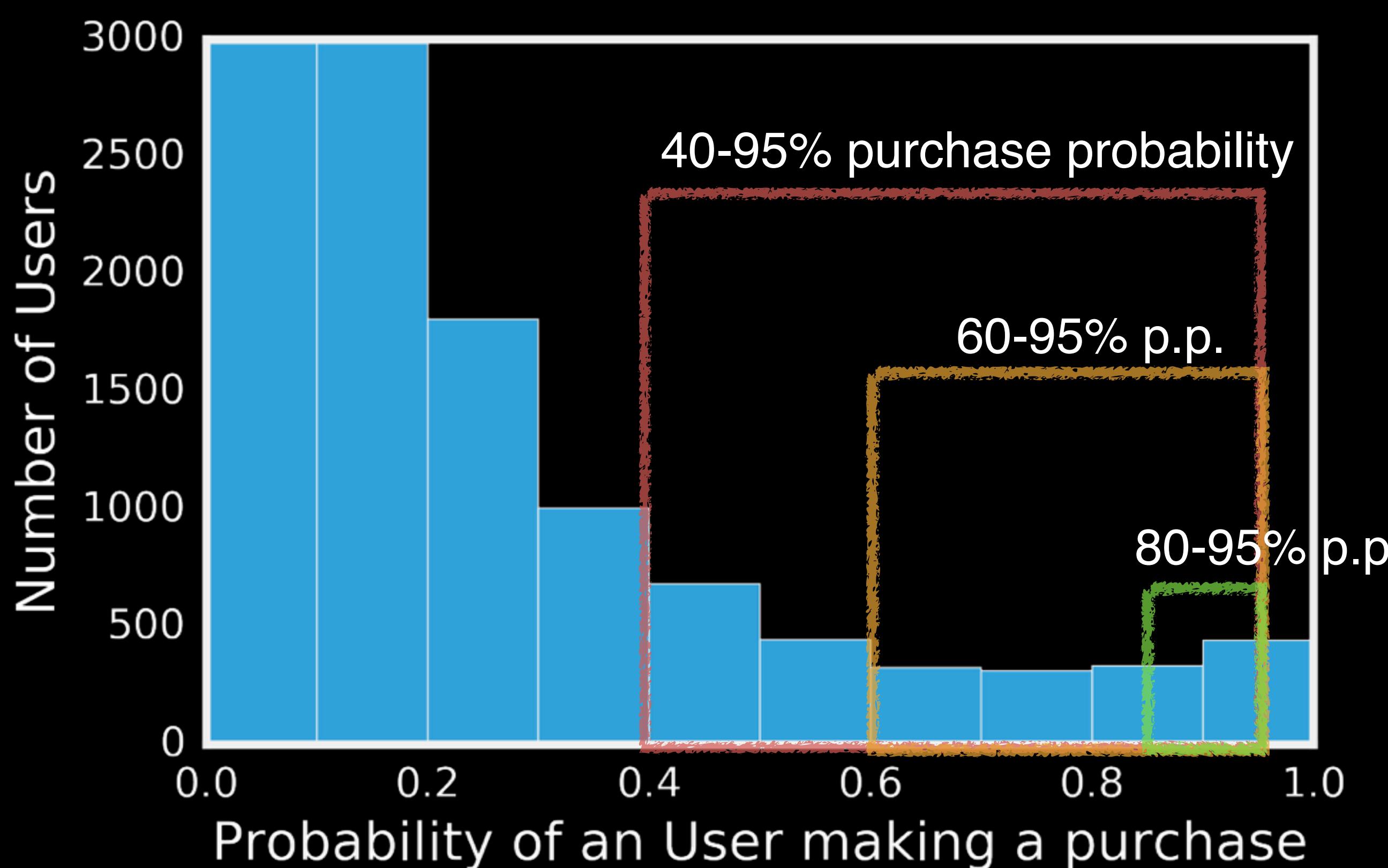
Follow 17 artists 

\* No causation relationship, but recommendations to be tested

# Simulations on Past Data

# Simulation!

Predictions distribution on original dataset  
[close-up]



*Scenarios:*

- 4.2% of users  
Potential of +61.3% in total revenue
- 2.2% of users  
Potential of +31.9% in total revenue
- 1.0% of users  
Potential of +14.3% in total revenue

\* There will be more costs to convert users with lower probabilities, and the potential considers full conversion

# Final Thoughts

When's the best timing? *Now.*

Can we predict such an abstract purchase? Yes.

What are purchase habits like? *Takes 103 days for first purchase.*

How much do they spend? *About \$250.*

How much money can we make? *Conservatively, +14.3% revenue!*

# Thank you!



Paulo Arantes

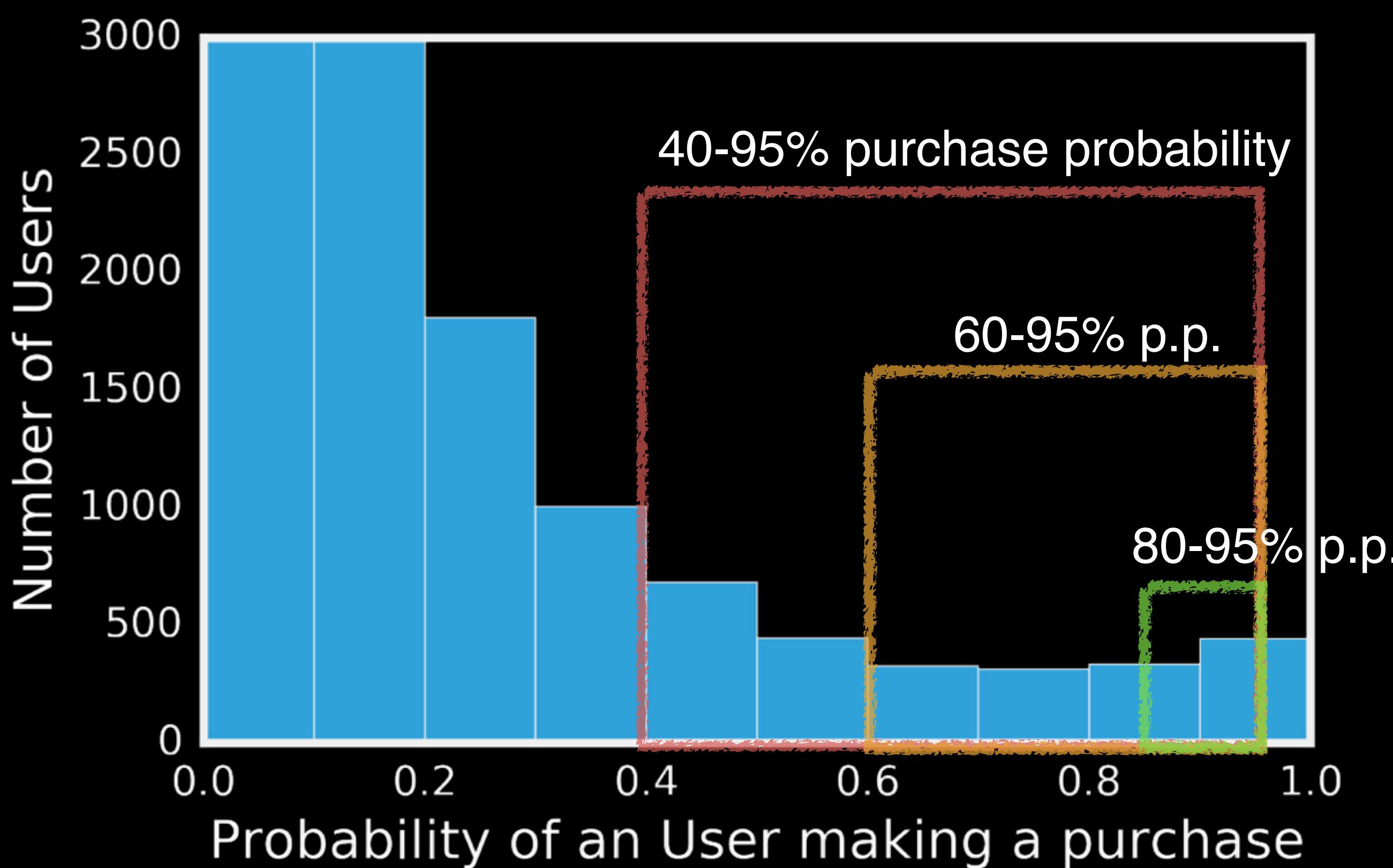
[paulo.espeschite@gmail.com](mailto:paulo.espeschite@gmail.com)

[linkedin.com/in/espeschite](https://linkedin.com/in/espeschite)

[github.com/pauloarantes/art-project](https://github.com/pauloarantes/art-project)

# Newest exported Model

Predictions distribution on original dataset  
[close-up]



*Scenarios:*

13.5% of users (7460)  
+189.7% = \$1,991,820

9.3% of users (5142)  
+130.8% = \$1,372,914

4.5% of users (2489)  
+63.3% = \$664,563

\* There will be more costs to convert users with lower probabilities, and the potential considers full conversion