MATP-4910 Final Project Notebook

DeFi

Jason Podgorski

2 December 2021

Contents

GitHub

GitHub ID: podgoj

Branch Name: dar-podgoj

Files on GitHub:

podgoj_final_notebook.Rmd

podgoj_final_notebook.pdf

podgoj_final_notebook.html

app.R

Issues:

#89 Does dygraph work in shiny (Closed)

#90 Develop Coin rate graphs for borrow and deposit and put it in shiny if possible (Closed)

#91 Create plots that have rates and amounts on the same plot from time (Closed)

#96 Coin View Page (In Progress)

Overview & Problems Tackled

The problems tackled in this paper deal with borrow rates in AAVE. AAVE is a decentralized lending protocol that allows users to borrow and lend crypto assets. The focus of this paper is to do a deep dive into how borrowing works in AAVE, who the good borrowers are, and if it's possible to forecast future borrow rates. To accomplish this, I defined two statistics to help analyze patterns in AAVE. When analyzing how a user borrows, a "good borrow" is when they make a borrow when the stable interest rate is between it's minimum and first quartile for the range of the dataset. Similarly, I define a "bad borrow" when a user makes a stable borrow with the interest rate being between the third quartile and maximum. Problem 1 focuses on stable borrows because variable rates can fluctuate drastically over the duration of the loan. Adding on this in problem 2, I attempt to predict variable interest rates using gradient boosting to help inform users when they can make "good borrows."

Data Description

The datasets used in this analysis were scraped using AAVE's API. The data starts with the beginning of AAVE V2 and spans until late October. The first dataset, primarily used in Problem 1, contains transaction data. The other dataset, used in both problems, contains variable and stable borrow rates through time of each coin. Collectively, the data was manipulated in order to draw the conclusions discussed later in this analysis.

Transaction Data

The transaction data, transactionsv2.rds, contains 33 features representing every transaction in AAVE V2. Some features include amount, reserve, transaction type, borrow rate, borrow mode, timestamp, etc.. There are 745,612 transactions containing over 50,000 unique users in the dataset.

Rates Data

The rates data, rates.csv, contains 5 features representing the borrow and liquidity rates through time. The features included are liquidity rate, variable borrow rate, stable borrow rate, reserve, and timestamp. There are 545,441 entries in the dataset.

Results

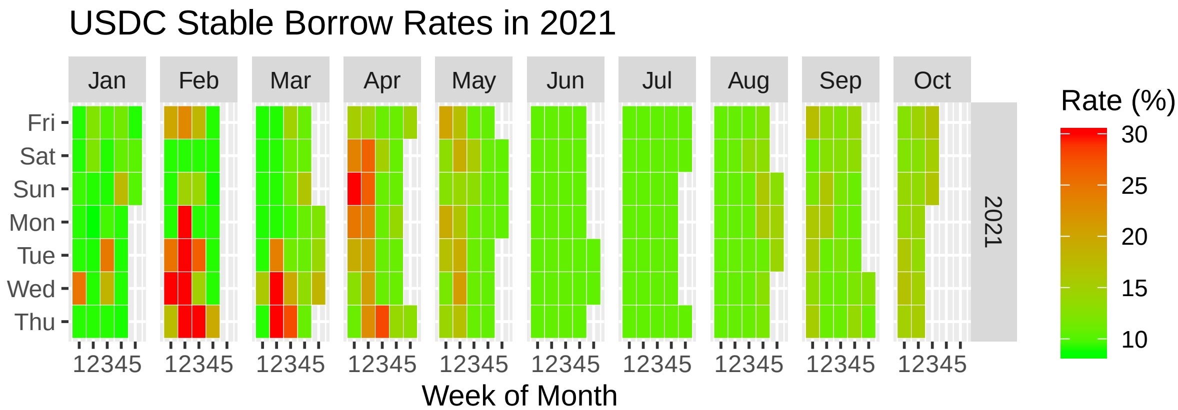
Problem 1

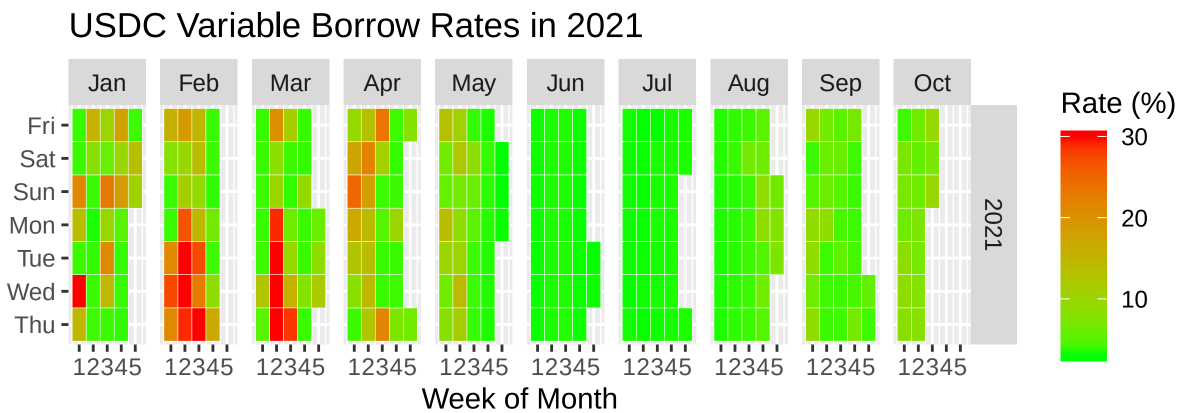
In this problem, I examine user's borrow patterns using my created definition of "good borrows" and "bad borrows." As described in the overview section, a "good borrow" is when a user makes a borrow when the stable interest rate is between it's minimum and first quartile for the range of the dataset. Similarly, a "bad borrow" is when a user makes a stable borrow with the interest rate being between the third quartile and maximum. This analysis focuses on stable borrows because variable rates can fluctuate drastically over the duration of the loan. The coins analyzed in this problem are USDC, USDT, and DAI because they are the most borrowed.

Methods

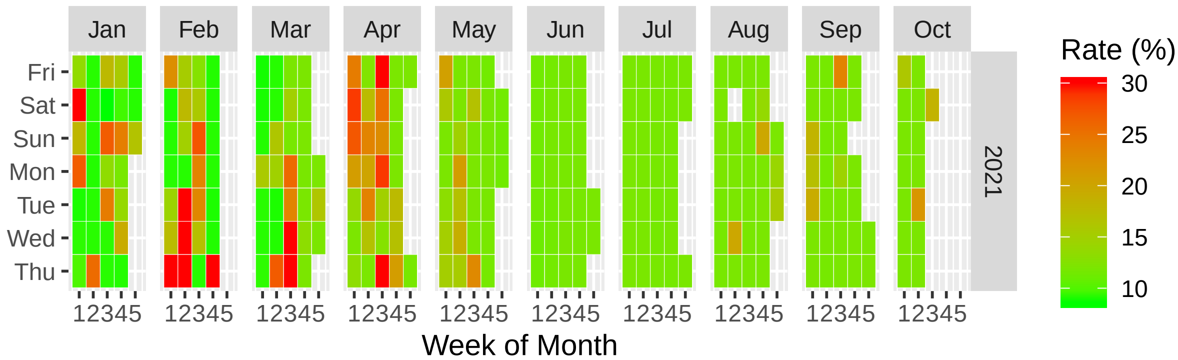
For this problem, I used a combination of the rates and transaction data. I created the time-series heatmaps by calculating the median stable and variable borrow rates for USDC, USDT, and DAI. I then used the transaction data to find which users borrowed on "good borrow" and "bad borrow" days. Finally, I summarized the statistics from these clusters of users to help tell a story in AAVE.

Results

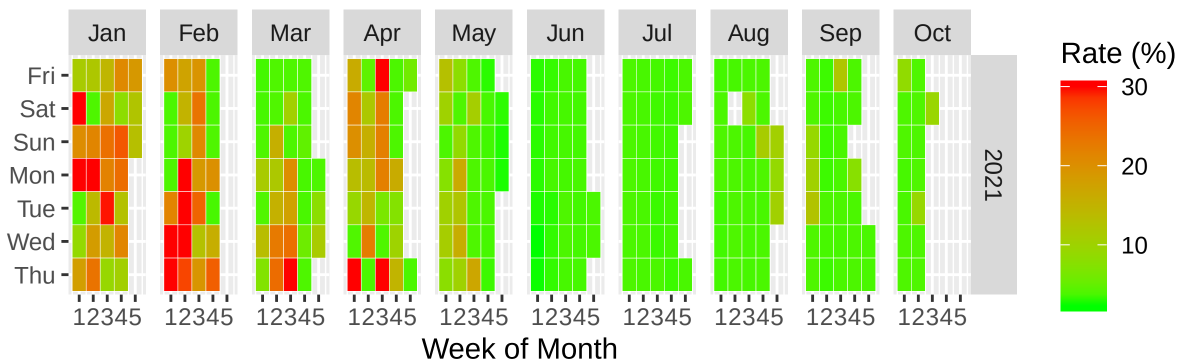

USDC Stable Borrow Rates in 2021
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Fri
Sat
Sun
Mon
Tue
Wed
Thu
12345
2021
Week of Month
Rate (%)
30
25
20
15
10


USDC Variable Borrow Rates in 2021
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Fri
Sat
Sun
Mon
Tue
Wed
Thu
12345
2021
Week of Month
Rate (%)
30
20
10

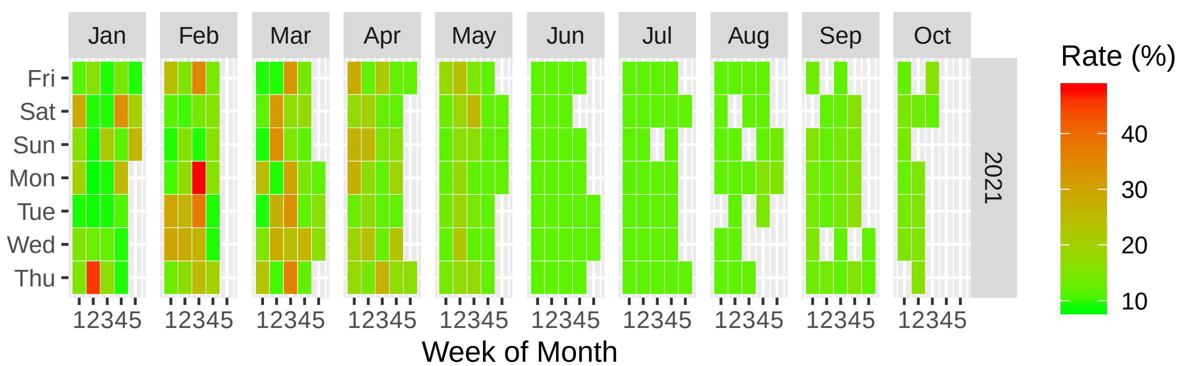
USDT Stable Borrow Rates in 2021



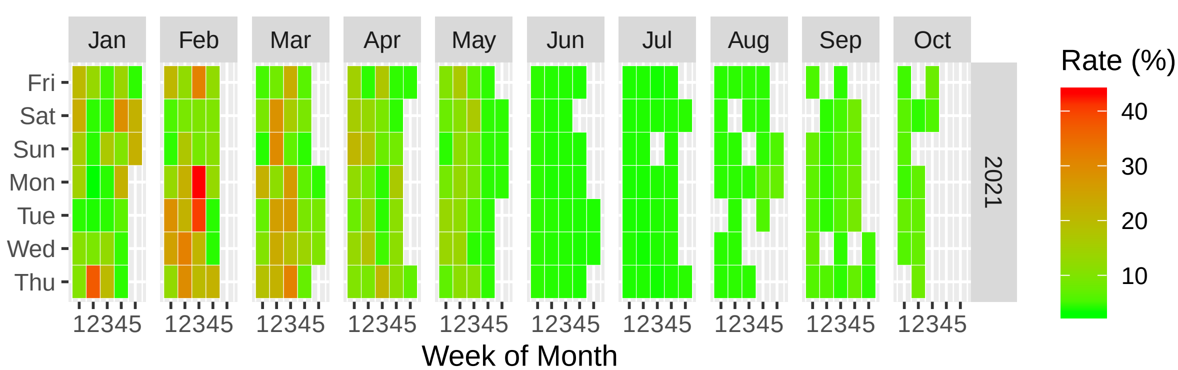
USDT Variable Borrow Rates in 2021

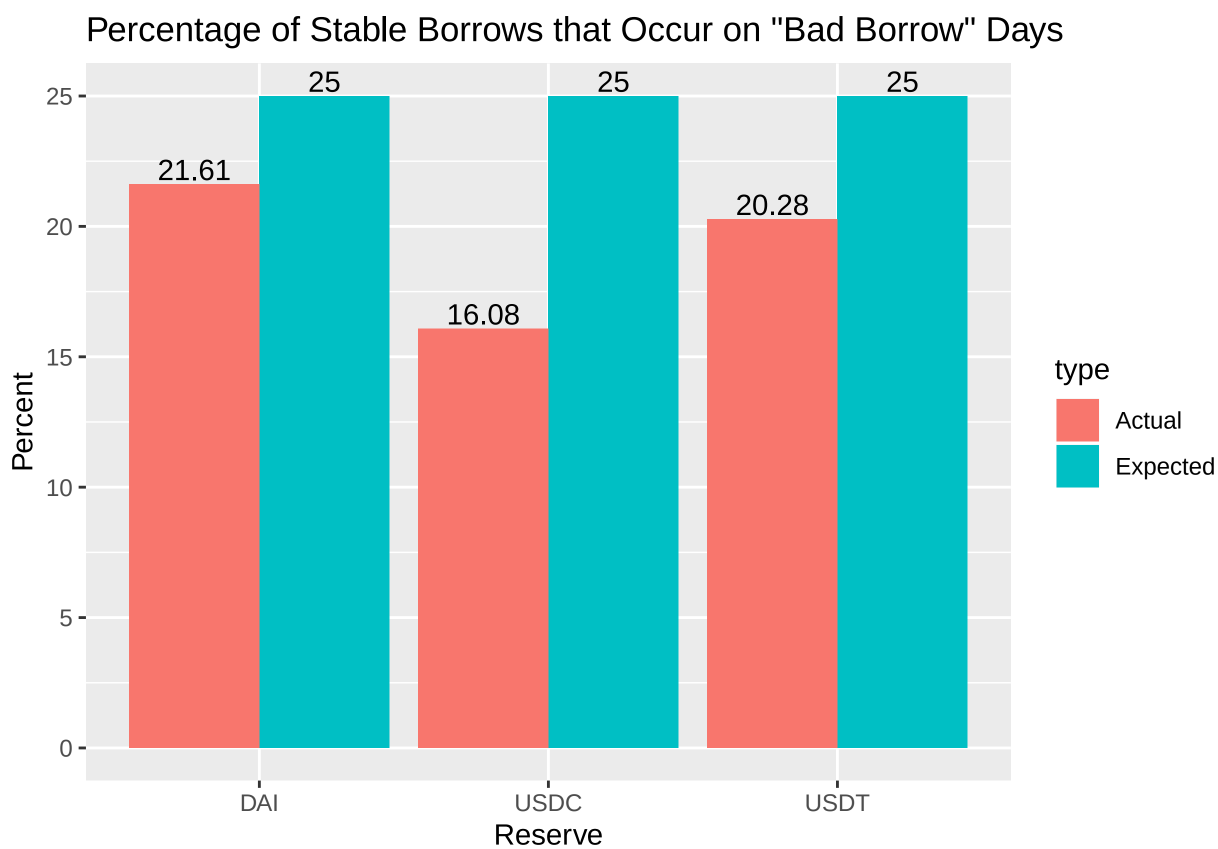


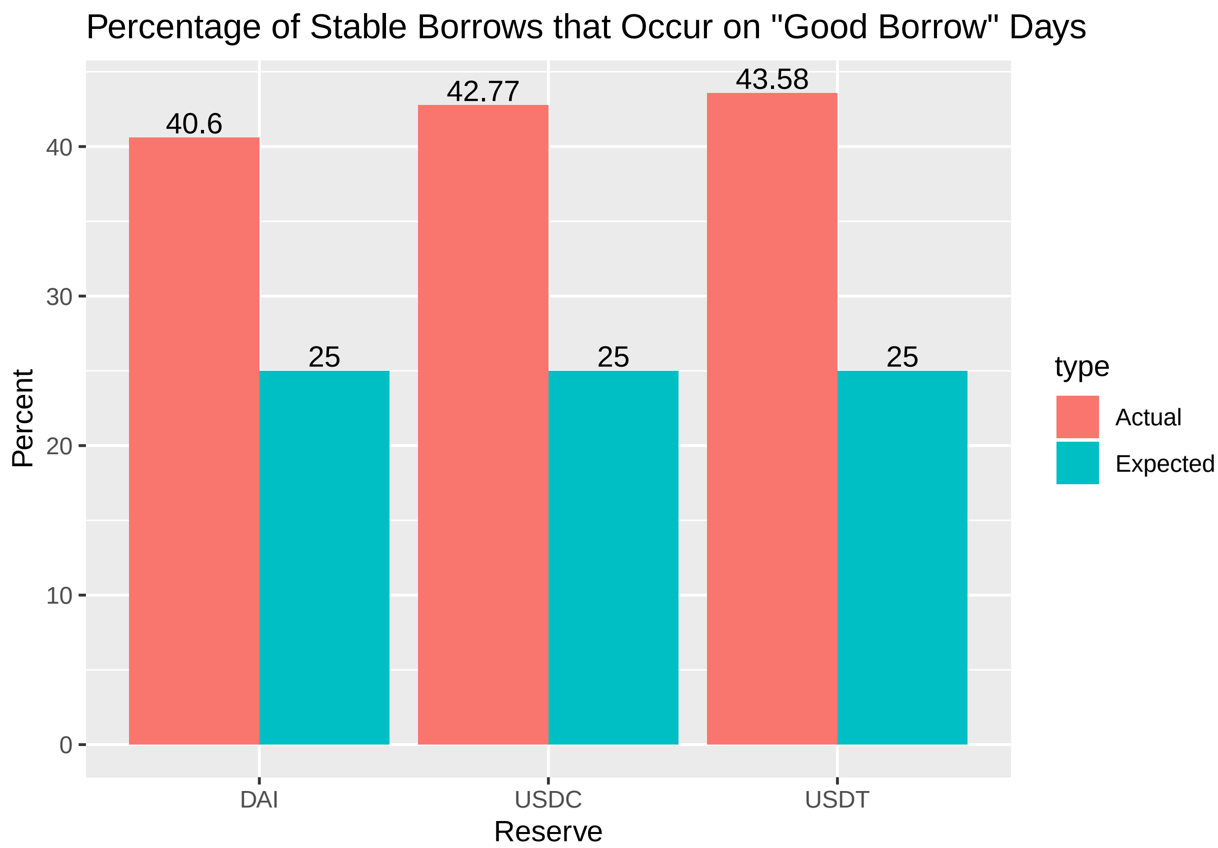
DAI Stable Borrow Rates in 2021

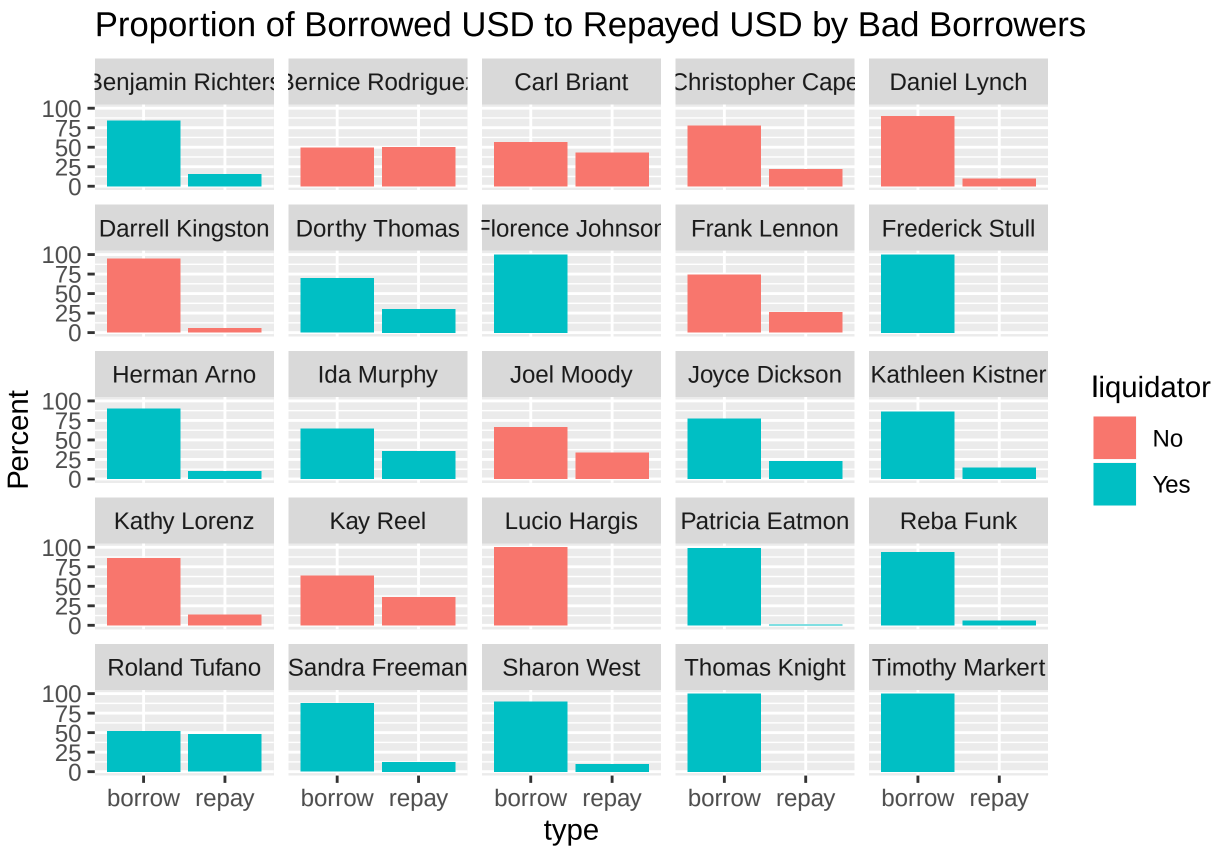


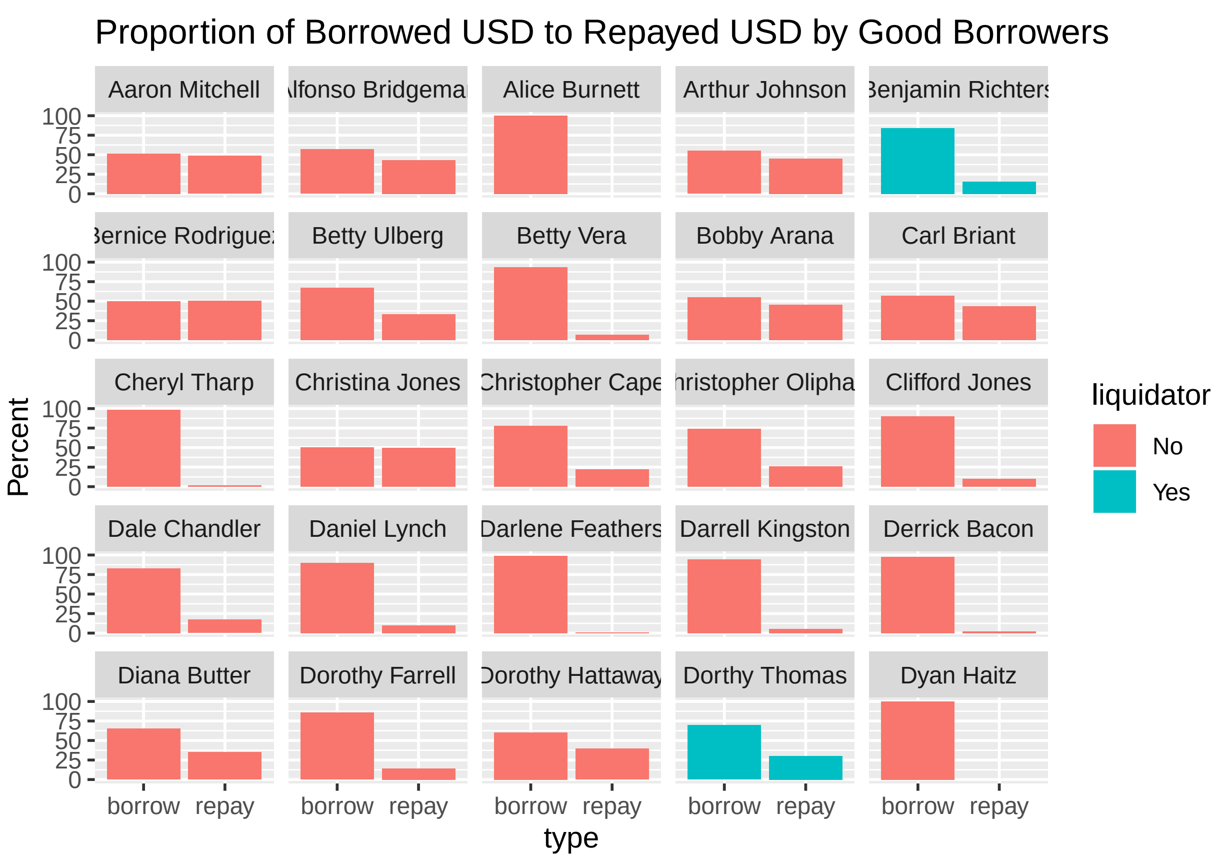
DAI Variable Borrow Rates in 2021




Percentage of Stable Borrows that Occur on "Bad Borrow" Days
Percent
25
20
15
10
5
0
21.61
25
16.08
25
20.28
25
DAI
USDC
USDT
Reserve
type
Actual
Expected


Percentage of Stable Borrows that Occur on "Good Borrow" Days
Percent
40
30
20
10
0
40.6
25
42.77
25
43.58
25
DAI
USDC
USDT
Reserve
type
Actual
Expected


Proportion of Borrowed USD to Repayed USD by Bad Borrowers
Benjamin Richters
Bernice Rodrigue
Carl Briant
Christopher Cape
Daniel Lynch
Darrell Kingston
Dorthy Thomas
Florence Johnsor
Frank Lennon
Frederick Stull
Herman Arno
Ida Murphy
Joel Moody
Joyce Dickson
Kathleen Kistner
Kathy Lorenz
Kay Reel
Lucio Hargis
Patricia Eatmon
Reba Funk
Roland Tufano
Sandra Freeman
Sharon West
Thomas Knight
Timothy Markert
Percent
100
75
50
25
0
borrow
repay
type
liquidator
No
Yes


Proportion of Borrowed USD to Repayed USD by Good Borrowers
Aaron Mitchell
lfonso Bridgema
Alice Burnett
Arthur Johnson
Benjamin Richters
Bernice Rodrigue
Betty Ulberg
Betty Vera
Bobby Arana
Carl Briant
Cheryl Tharp
Christina Jones
Christopher Cape
hristopher Olipha
Clifford Jones
Dale Chandler
Daniel Lynch
Darlene Feathers
Darrell Kingston
Derrick Bacon
Diana Butter
Dorothy Farrell
Dorothy Hattaway
Dorthy Thomas
Dyan Haitz
Percent
100
75
50
25
0
borrow
repay
type
liquidator
No
Yes

Discussion

From the time-series heatmaps, most of the worst days to borrow occur early in the year in January, February, March, and April. This could be due to AAVE V2 just being launched in December. As discussed more in problem 2, China announced their ban on cryptocurrency in mid-May. As a result, all borrow rates plummeted because an influx of capital is available in the pool (utilisation rate decreases, discussed more in problem 2).

In the next set of visualizations, I examine the percentage of good and bad borrows based on what's expected. By definition, good borrows and bad borrows should make up 25% each of all borrows. The results from the bar charts make sense. More people borrow on "good" days than expected and less people borrow on "bad" days than expected. Borrowers should limit borrowing on bad days because borrowing at higher rates makes it tougher to pay back and can lead to liquidation.

The next set of visualizations incorporates the best and worst borrowers. To be grouped with the best borrowers, a user must have at least 10 "good borrows" from at least one of the three coins and vice versa for the worst borrowers. Of the 25 worst borrowers, 15 have been liquidated in their history (60%). There were 69 users in the best borrower category (25 made the visual), 19 of them have been liquidated in their history (27.5%). To go further, the visuals represent the total amount of USD borrowed versus repayed for both groups. For users who haven't liquidated, there borrow and repay totals should be similar. Otherwise, they may have made the borrow recently. All in all, this supports the notion that users who make smarter borrows with more favorable rates have less issues with liquidation in the future.

Problem 2

Diving deeper into borrowing, problem 2 is a time-series forecasting analysis that predicts the median borrow rate of several coins for the next day. This analysis is particularly useful for a user who is deciding what type of borrow rate would be more optimal depending on their estimated length of pay back.

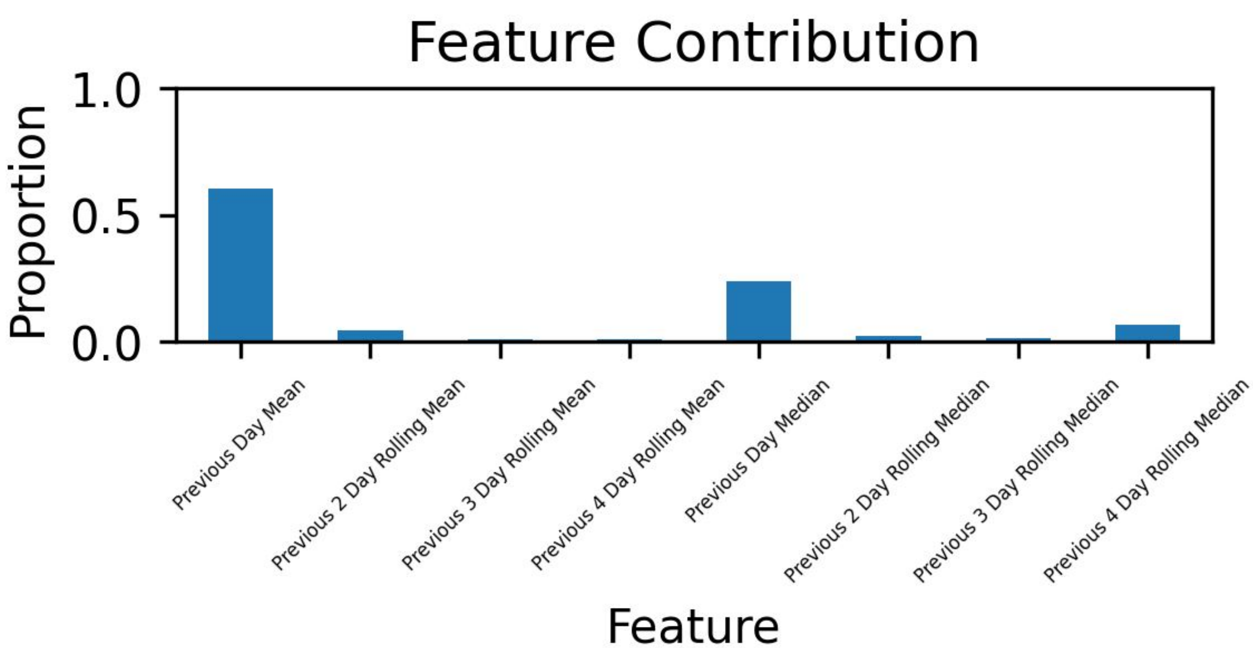
Methods

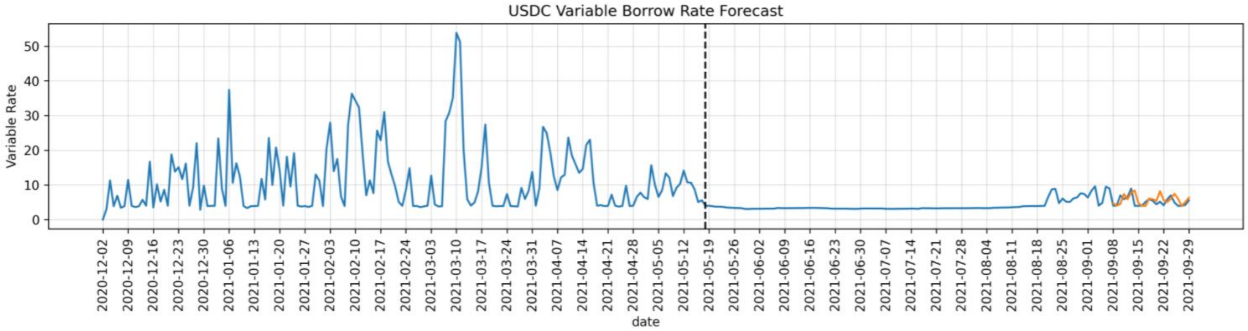
The first step in this analysis is to chose a model. There are two different directions that are typical for solving time-series forecasting problems. The two types of models are traditional time-series models and machine learning models. Traditional time-series models are recursive (can make predictions easily for any time in the future), tougher to get right, and can't have regressors added. Machine learning models, on the other hand, have to be trained to predict values a set period of time in the future. Because of this, they are easier to get right and regressors can be added to predict future values. After testing both models, the gradient boosting machine learning model, XGBoost, in Python was used. The problem is framed as a regression problem. The model will be trained to fit the curvature that makes up the time series.

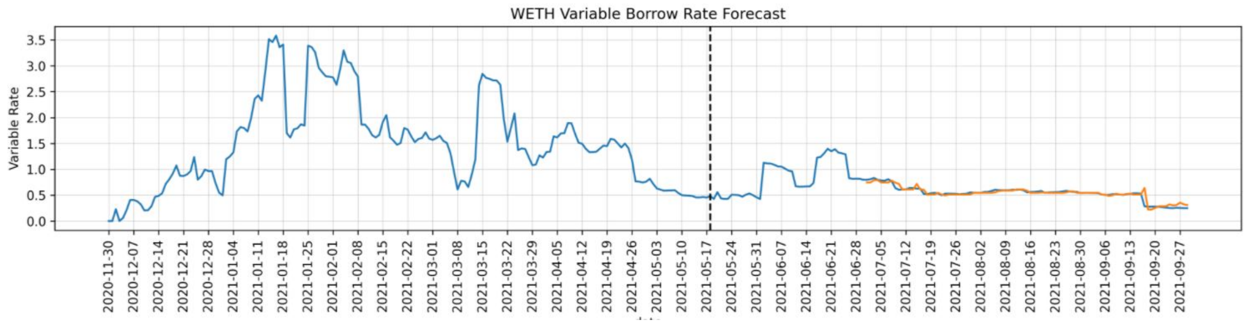
The features initially selected to train the model were the previous day median borrow rate, previous day mean borrow rate, and rolling averages for the means and medians over the last two, three, and four days in the past. To get this data, I used rates.csv and calculated the mean and median variable borrow rates for each coin per day. The selected coins to test the model were USDC, USDT, DAI, and WETH. They have the most borrows in AAVE by a wide margin. This analysis includes just USDC and WETH to demonstrate my results because USDC, USDT, and DAI behave similarly. WETH is the only non-stable coin of the bunch and behaved quite differently than the other three (shown in the results section). To account for this difference, I had to treat predicting stable and nonstable coins as two separate problems. I started the training data for the stable coins from May 18, 2021, the day that China announced their cryptocurrency ban and delayed the start of testing data to September 8, 2021. For nonstable coins like WETH, I started the training data the earliest date in the dataset and started the testing set on July 1, 2021. The reason for this difference is that the interest rates behaved entirely different after China's ban and that using those dates to train the model would be meaningless. Because of this, I am forced to use less test data in order to sufficiently train the model.

Results


Interest Rates for Most Borrowed Coins
Dashed Line Represents the China Crypto Ban
DAI
USDC
USDT
WETH
Percent
0
10
20
30
Jan
Apr
Jul
Oct
Week
colour
stable
variable


Feature Contribution
Proportion
0.0
0.5
1.0
Previous Day Mean
Previous 2 Day Rolling Mean
Previous 3 Day Rolling Mean
Previous 4 Day Rolling Mean
Previous Day Median
Previous 2 Day Rolling Median
Previous 3 Day Rolling Median
Previous 4 Day Rolling Median
Feature





Discussion

To preface these results, AAVE borrow rates are dependent on the utilisation rate. When a large proportion of a reserve is borrowed, borrow rates increase to discourage more borrowing. Borrow rates are decreased when a small proportion of a reserve is borrowed to encourage borrowing. Utilisation data isn't in the dataset I had access too. Using utilisation rate data is another way to go about the problem but a time-series model would be better suited for that data.

When analyzing the stable and variable borrow rates through time, we notice similar patterns between USDC, USDT, and DAI. The borrow rates collapse the same day China announces their cryptocurrency ban. There is uncertainty why there isn't the same behavior demonstrated with WETH. The only difference is that WETH is an unstable coin and the other three are stable. The rates drop significantly because there is more capital in the pool (utilisation rate decreases) so interest rates decrease to encourage borrowing. There isn't a bounce back until a few months later. Major events like the China ban are pitfalls that make forecasting future rates a challenging problem. No model can ever account for these types of market shakeups. Another huge difference between the borrow rates of stable and nonstable coins are their volatility. It's difficult to be precise with USDC, USDT, and DAI because the rate can change by thirty percent in a given day.

The feature distribution shown in the bar chart is the percentage of impact it has in the gradient boosting model. As shown, the previous day mean makes up half of the model while some other features appear to have little to no value. This set of features was reduced from the full feature list to optimize the model. While it appears that some of the kept features aren't useful, they are important in minimizing overfitting. This was concluded by calculating the mean absolute error of the testing data of the converged models for the different coins.

After examining the forecasts for USDC and WETH, as expected, forecasting USDC (and the other stable coins) was less accurate than WETH. The mean absolute error of USDC converged to 1.23% and converged to 0.29% for WETH. Mean absolute error is the average distance between the actual and predicted value.

Summary and Recommendations

In summary, I used the first problem to creatively find clusters of users within AAVE. While creating this analysis, I did find certain patterns interesting and worth talking about it. In particular, the many users who willingly borrow at obscenely high stable rates. For long-term research, one could probably incorporate "good borrow" and "bad borrow" statistics to bolster a paper but not have it be the main topic. For the second problem, I believe there's a lot of promise with my model. It will only get better as time goes on and more data is collected. This type of work would be best in a paper. Nonetheless, time-series forecasting is an incredibly powerful tool and can be applied to many different aspects of AAVE.

Additionally, the borrow rate visuals and forecasts will be incorporated into the app to help showcase how different coins behave in AAVE.

References

Appendix