

# MATP-4910 Final Project Notebook

DeFi

Jason Podgorski

2 December 2021

## Contents

<b>GitHub</b>	<b>1</b>
<b>Overview &amp; Problems Tackled</b>	<b>2</b>
<b>Data Description</b>	<b>2</b>
Transaction Data . . . . .	2
Rates Data . . . . .	2
<b>Results</b>	<b>3</b>
Problem 1 . . . . .	3
Problem 2 . . . . .	8
<b>Summary and Recommendations</b>	<b>12</b>
<b>References</b>	<b>13</b>
<b>Appendix</b>	<b>13</b>

## GitHub

GitHub ID: podgoj

Branch Name: dar-podgoj

Files on GitHub:

dar\_final\_podgoj\_2dec2021.Rmd

dar\_final\_podgoj\_2dec2021.pdf

dar\_final\_podgoj\_2dec2021.html

dar\_final\_podgoj\_2dec2021.ipynb

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 its 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. Additionally 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 2021. 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. The data incorporates 53 different reserves including both stable and non-stable coins. Stable coins are coins typically backed to other assets such as the US Dollar. They are reliable to have for long periods of time. Non-stable coins fluctuate and are used for making money. Stable and non-stable coins aren't distinguished in the data but knowledge of this difference is incorporated throughout this analysis.

### Transaction Data

The transaction data, `transactions_v2.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. For every transaction record in the data, it is classified as one of seven different transaction types: borrow, collateral, deposit, liquidation, redeem, repay, and swap. A borrow is when a user borrows coin from the coin's lending pool. A borrow must be overcollateralized (represented with the features ending with "Collateral"). There are two types of borrow modes, stable and variable. A stable borrow rate is a rate where the user generally takes a higher interest rate but it remains constant throughout the duration of the loan. Variable rates fluctuate constantly but are often lower than stable rates on a given day. Collateral is a record of the asset gained when successfully liquidating another user. A liquidation is the a record of a user having their assets taken due to a partial or total loss of their initial margin. A redeem is when the user removes deposited coins from the pool. Finally, a swap is when the user swaps one coin for another. Another aspect of this transaction dataset are user aliases. Each unique address in the "user" feature is matched up with a fake alias to better distinguish patterns between individual users.

### 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. The entries for each coin are collected roughly every 4-5 minutes each day.

# Results

## Problem 1

In this problem, I examined 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 its 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 time-series heatmaps by calculating the median stable and variable borrow rates for USDC, USDT, and DAI. This allowed me to visualize which days tended to be better or worse for borrowing. I then used the transaction data to find which users routinely made "good borrows" and "bad borrows." Finally, I summarized the statistics from these clusters of users to help tell a story in AAVE.

## Results

### USDC Stable Borrow Rates in 2021

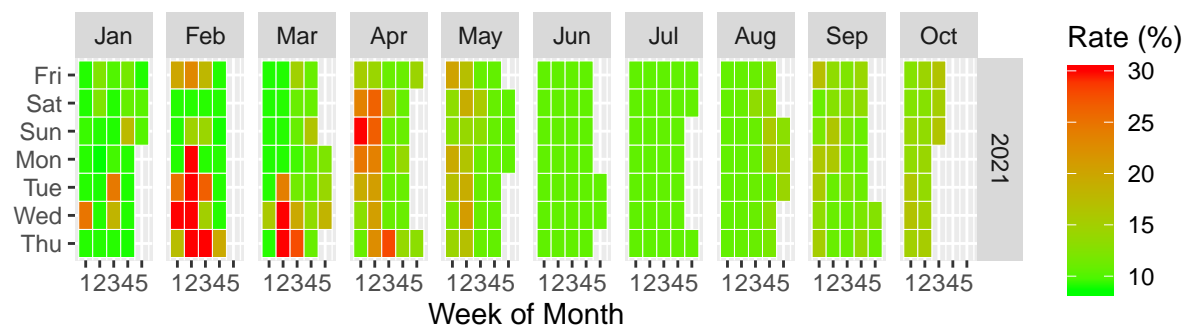


Figure 1.1.1

### USDC Variable Borrow Rates in 2021

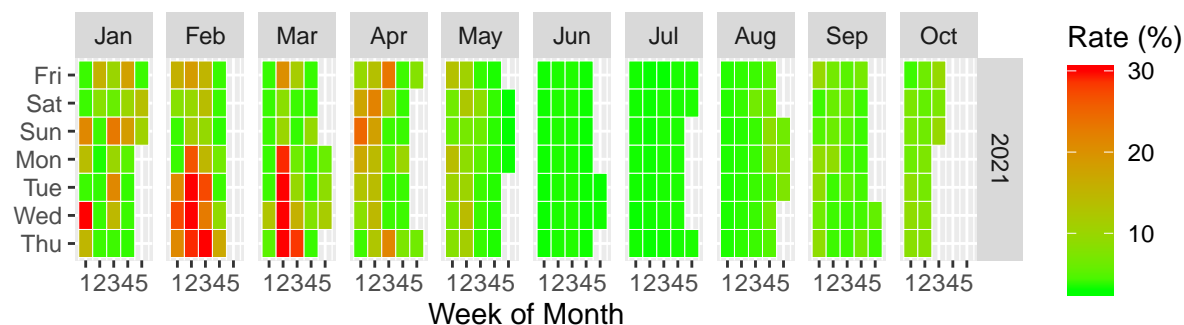


Figure 1.1.2

### USDT Stable Borrow Rates in 2021

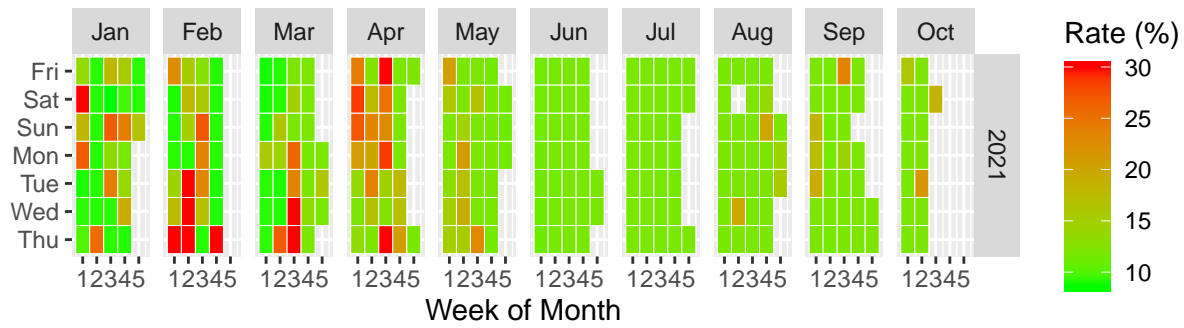


Figure 1.2.1

### USDT Variable Borrow Rates in 2021

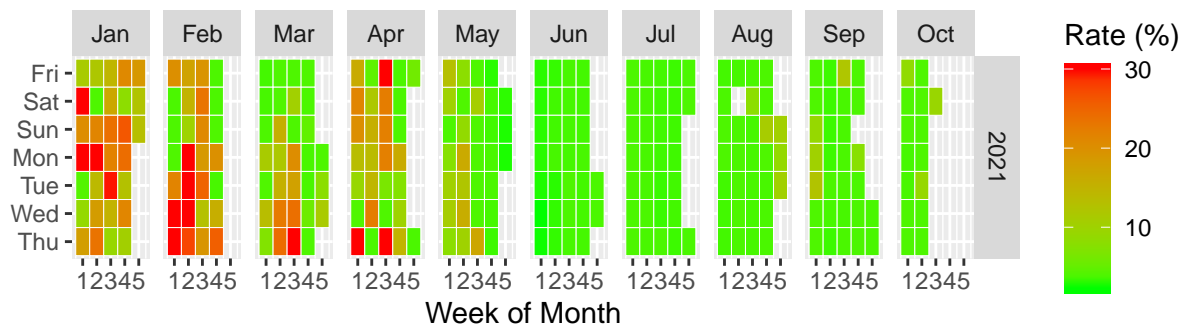


Figure 1.2.2

### DAI Stable Borrow Rates in 2021

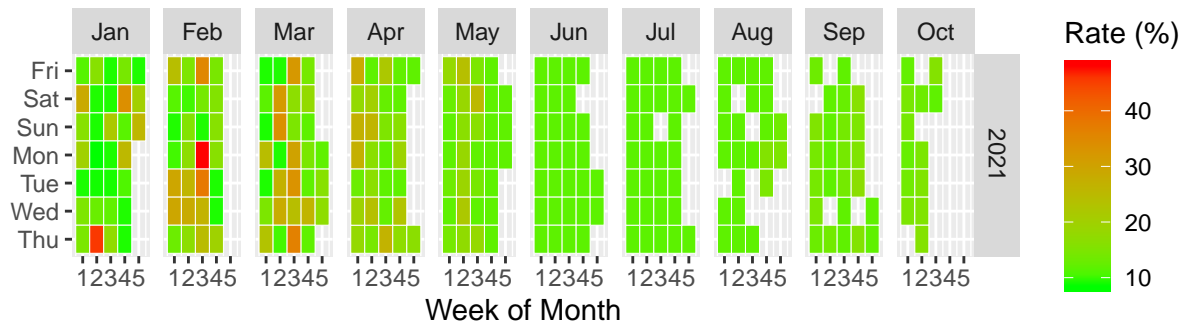


Figure 1.3.1

### DAI Variable Borrow Rates in 2021

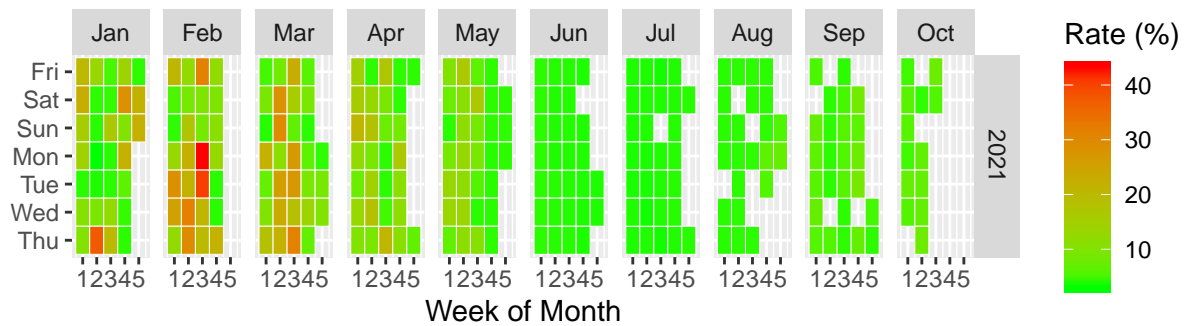


Figure 1.3.2

Figures 1.1.1 through 1.3.2 are time-series heatmaps that show the stable and variable borrow rates of USDC, USDT, and DAI. Days that are dark red are when borrow rates are at their highest while days that are light green represent lower rates. 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 became available in the pool (utilisation rate decreases, discussed more in problem 2).

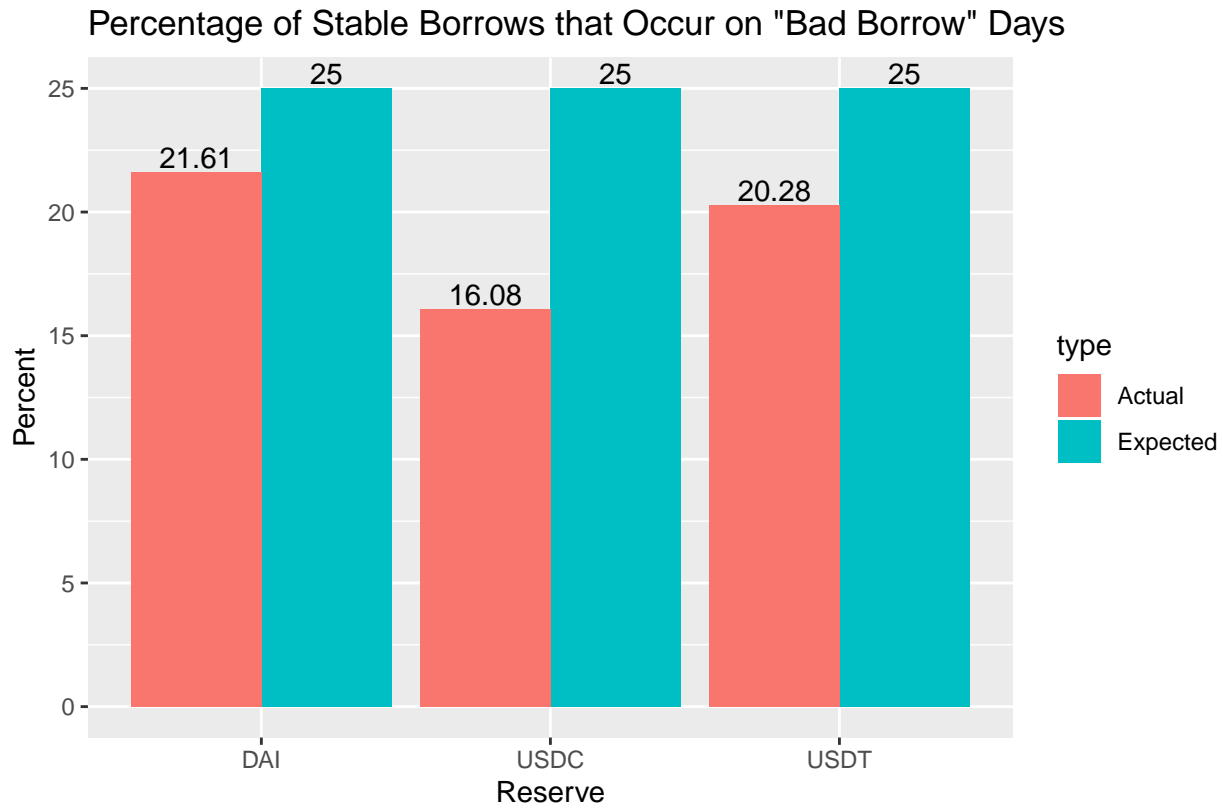


Figure 1.4.1

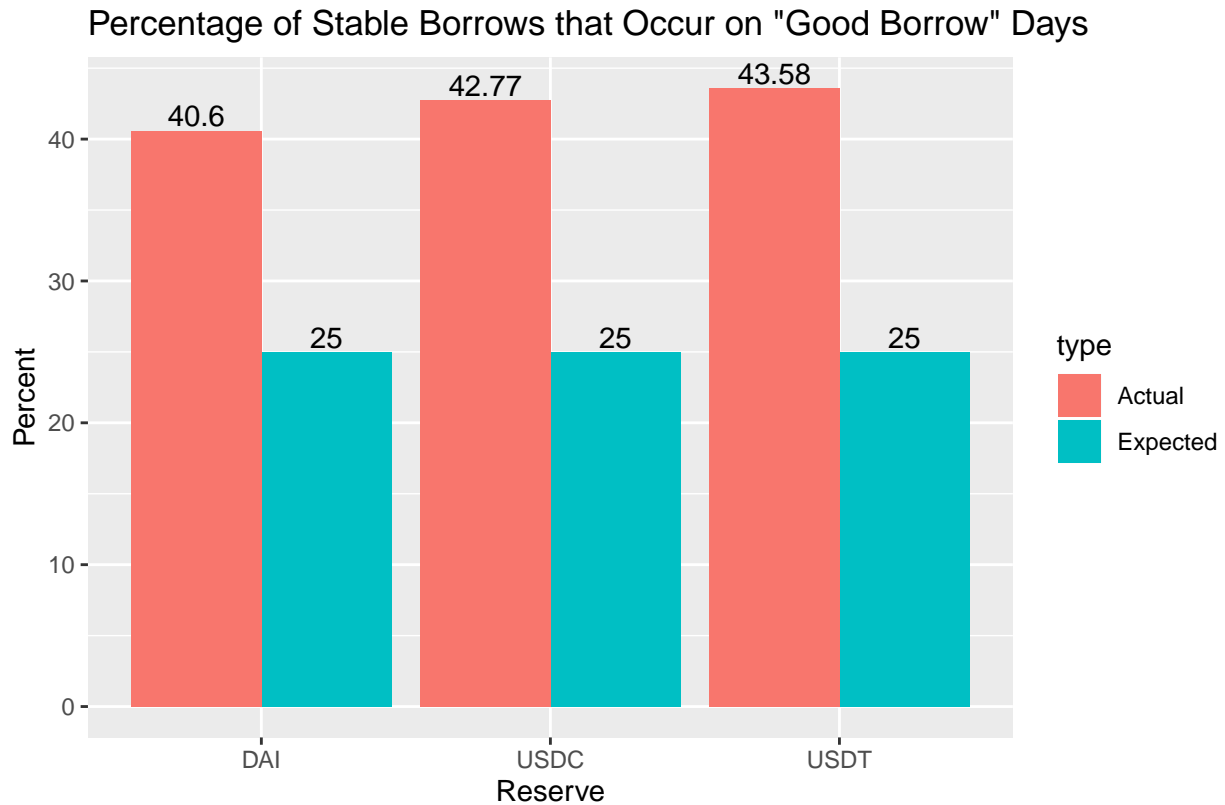


Figure 1.4.2

Figures 1.4.1 and 1.4.2 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 (first and fourth quartile). The results from the bar charts make sense. More people borrow at "good" rates than expected and less people borrow at "bad" rates than expected. Borrowers should limit borrowing at higher rates because it makes it tougher to pay back loans and can lead to liquidation.

## Proportion of Borrowed USD to Repayed USD by Bad Borrowers

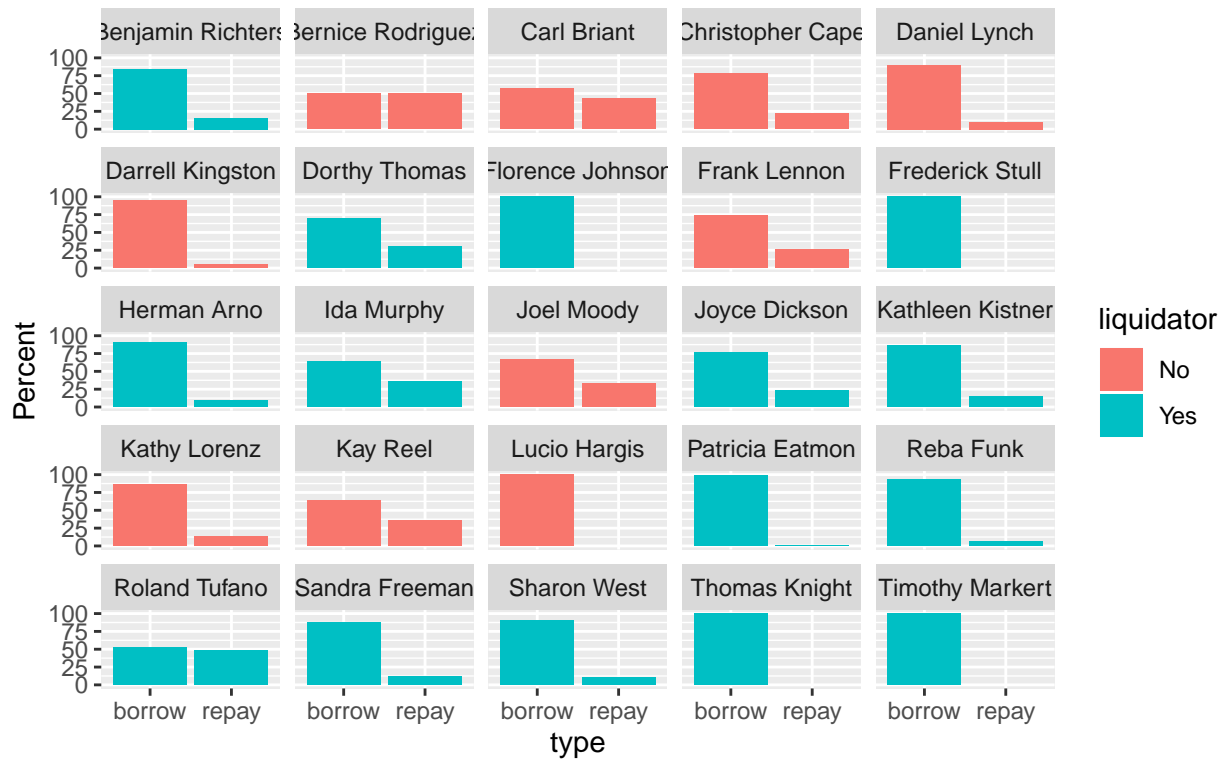


Figure 1.5.1

## Proportion of Borrowed USD to Repayed USD by Good Borrowers

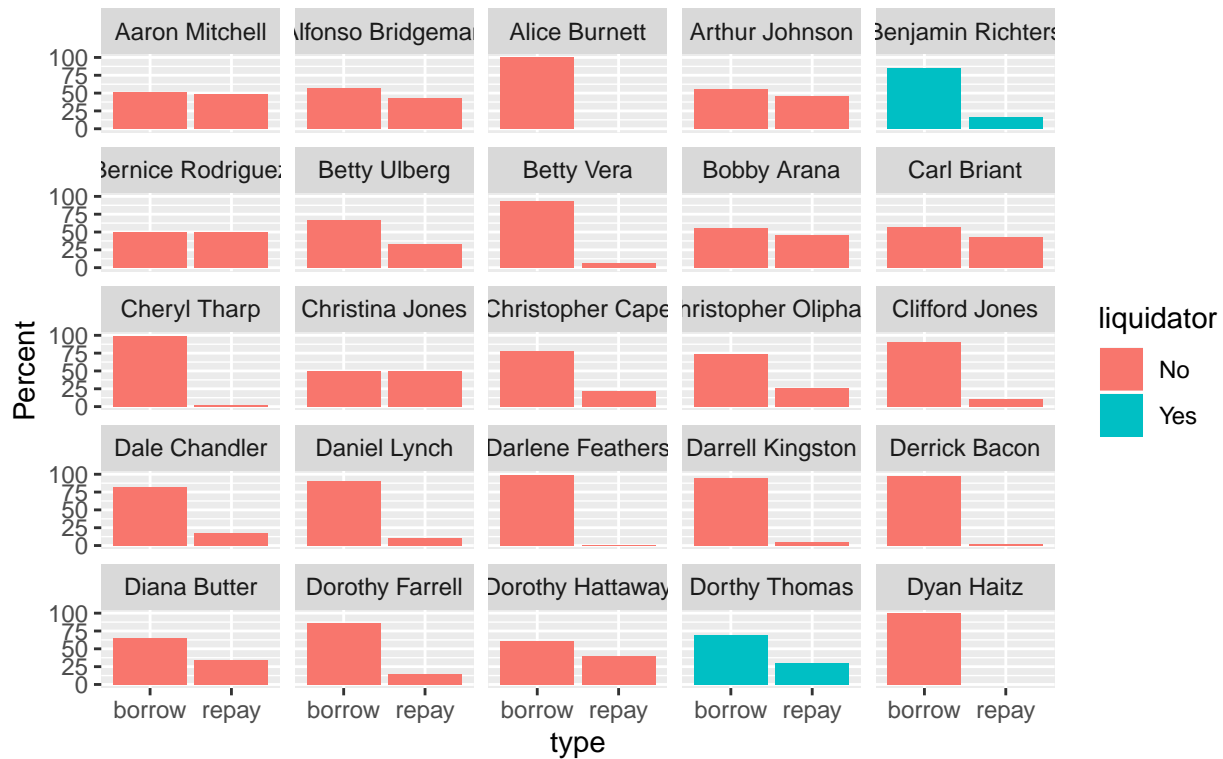


Figure 1.5.2

Figures 1.5.1 and 1.5.2 show the user aliases for the best and worst borrowers in AAVE. 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, their 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.

## Discussion

Overall, there are several takeaways gained from this analysis. To begin, figures 1.1.1 through 1.3.2 show a drastic difference between rates before China’s crackdown and after. There’s definitely some bias in this study because users who started borrowing after the crackdown had a much easier time making “good borrows.” Figures 1.4.1 through 1.5.2 demonstrate the importance of being patient and finding the optimal time to borrow assets. Users who don’t make optimal borrows are often forced to have their assets liquidated. This transitions into Problem 2 where we attempt to get a more precise outlook on borrow rate patterns in AAVE.

## 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 types of 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. Because I’m treating this model as a regression problem, the features used have to be values from the past that help in predicting the next day median. To generate those features, I thought of statistics used in the traditional stock market for predicting stock prices. Stats like opening, closing, minimum, and maximum price for a share are typically used to predict the next day opening price for example. I incorporated similar statistics with borrow rates for each day in AAVE. Since there’s no such thing as opening and closing prices with cryptocurrency, I used the first and last records for borrow rates on the given day. I also incorporated the median and mean rates since they are very relevant in predicting the next day median. In addition to using the previous day opening, closing, minimum, maximum, median, and mean borrow rates, I calculated rolling averages for these values for up to four days in the past. After running the model with all of these features, the features using mean and median were the most relevant for predicting the next day median borrow rate.

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 non-stable 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 the



testing data to September 8, 2021. For non-stable 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

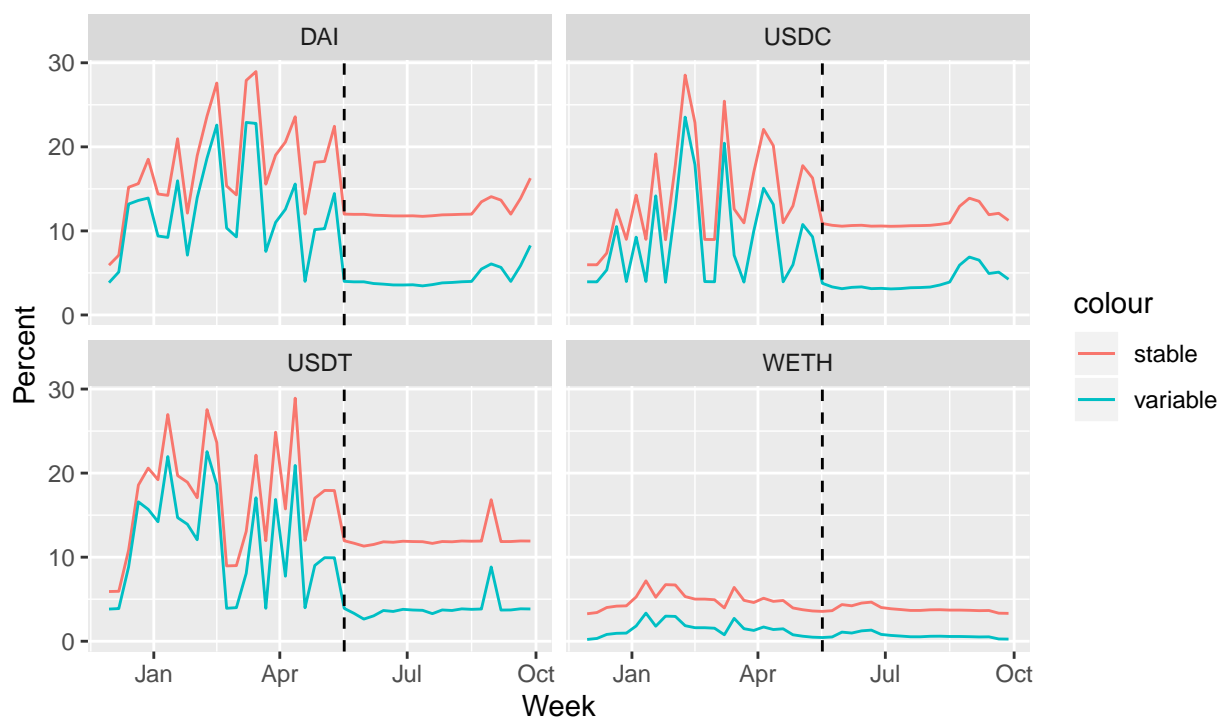


Figure 0

Figure 0 shows the variable borrow rates of the four coins used in the model. The dashed line represents the day China announced their cryptocurrency ban. This announcement had a big impact in AAVE and added a ton of new capital to the lending pool all at once which consequently drove down borrow rates. More on the China ban will be mentioned in the discussion.

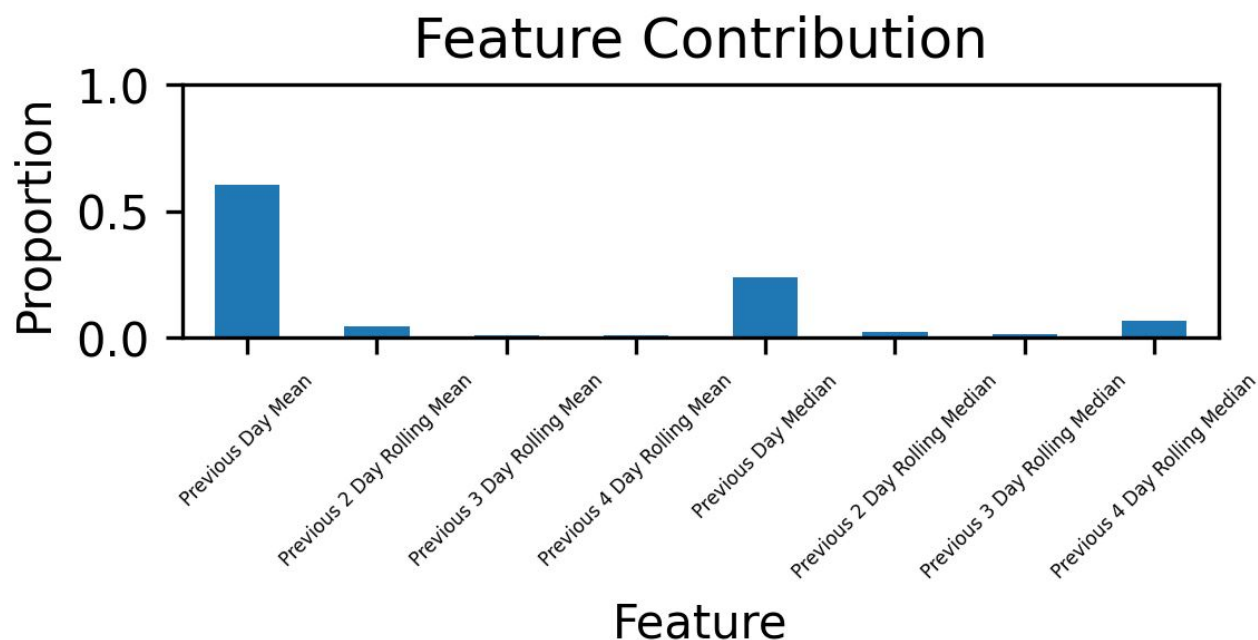


Figure 1: \_\_

Figure 1 breaks down how each feature is incorporated into the model. 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.

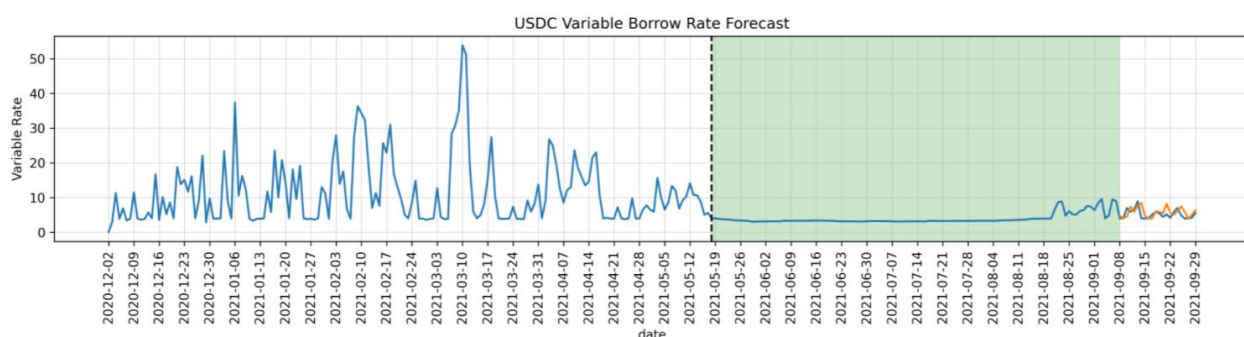


Figure 2: \_\_

Figure 2 shows the results of the forecast on USDC. The green area represents the dates used to train the model. As mentioned previously, I started the training data after China's ban because the rate patterns before and after behaved completely different. The orange line represents the predicted median variable borrow rate while the blue line is the actual value. The mean absolute error of USDC converged to 1.23%. This is an encouraging result given how much the China ban impacted the market and limited access to quality training data. Mean absolute error was used to judge how well the model performed. Mean absolute error represents the average distance between the actual and predicted value.

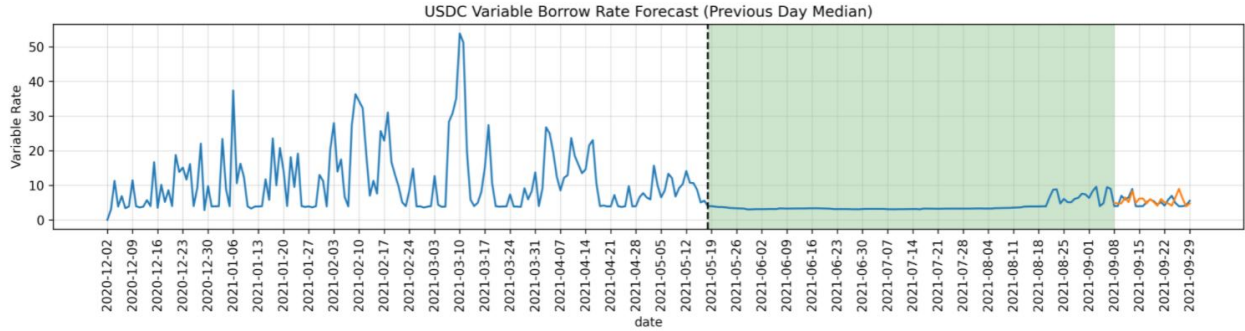


Figure 3: \_

Figure 3 is a comparison of the results of Figure 2 using the previous day median rate as the only feature. The mean absolute error for this forecast was 1.30%. While similar, I believe a greater difference will be seen as more data is incorporated into the model.

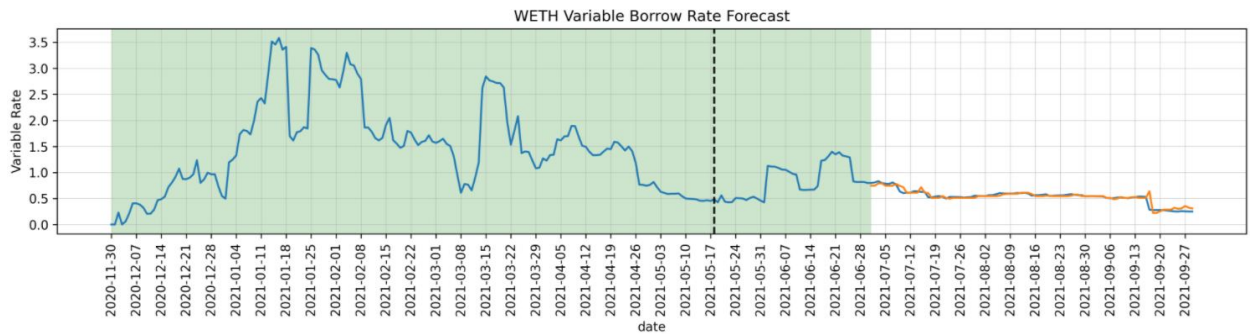


Figure 4: \_

Figure 4 shows the results of the forecasting model on WETH. The mean absolute error converged to 0.29% which is significantly better than the forecast for USDC. This was expected because non-stable coins didn't see an impact from China's ban.

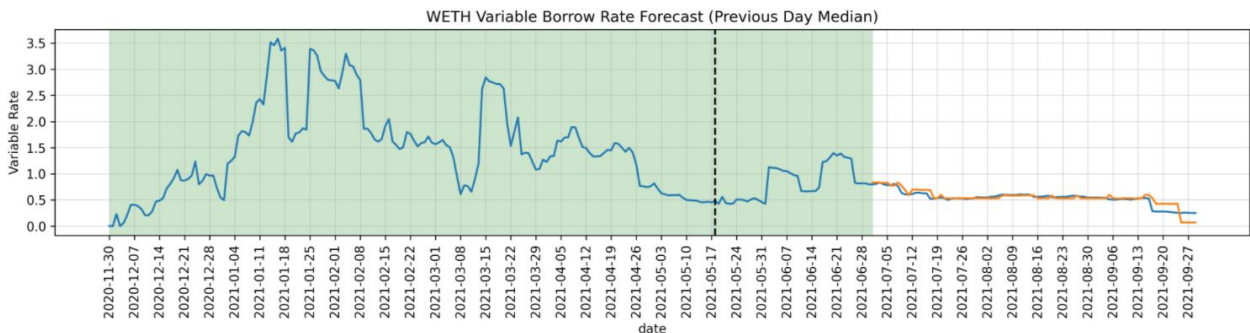


Figure 5: \_

Figure 5 is a comparison of the results of Figure 4 using the previous day median rate as the only feature. The mean absolute error for this forecast was 0.50%. By just using the previous day median rate as the only

feature, the model cannot predict abrupt large rises or falls until the day after which is exactly what is shown with the results of this forecast.

## Discussion

To preface this discussion, 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 a non-stable 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 non-stable 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. That being said, the duration of the dataset spans less than a year. I would anticipate this model to continually improve as time goes on.

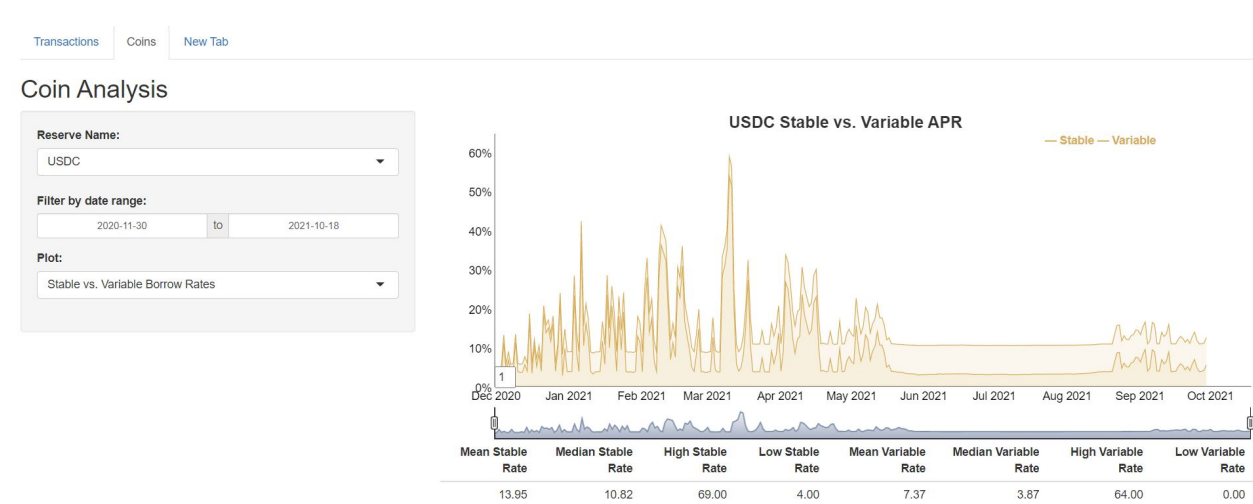
## Summary and Recommendations

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



app.R

As a team we started building an application that could be a source for coin, user, and survival data. With time running down in the semester, we were unable to finish a prototype. Above is a screenshot of the coin page that Cole and I began working on. The goal was to combine my borrow rate forecasting and analysis with his Nansen-inspired visualizations to give users a way of exploring how different coins behave in AAVE. When a future team picks up this project, I believe continuing to develop this app will be a good direction to go in.