5-5-2024

Data Analysis Project for Preferred Credit Inc.

Emily Smith
College of Saint Benedict/Saint John's University

Greta Nesbit
College of Saint Benedict/Saint John's University

Jack Simonet
College of Saint Benedict/Saint John's University

Ignacio Sanchez-Romero
College of Saint Benedict/Saint John's University

Emily Smith

Greta Nesbit

Jack Simonet

Ignacio Sanchez Romero

Professor Bob Hesse

5/05/2024

Data Analysis Project for Preferred Credit Inc.

Our external partner for this semester-long project was Preferred Credit Incorporated (PCI). PCI was established in 1982 by Gene Windfeldt, who had previously worked with Kirby Vacuums. When he faced economic challenges, all of his company's financers withdrew their support which led Gene to launch his own financing company, which eventually evolved into PCI. PCI is a leader in direct sales financing, helping companies in different fields offer payment options to their customers. They essentially provide loan funding on behalf of the companies, acting similarly to a bank in this way.

For a brief overview of our process, we began our project by looking into the three large datasets that PCI gave us: The Accounts, Payments, and Phone Calls. We had a bit of a rocky start with this project because the data sets we were given were too large to open in excel so we were unable to view the data. We ended up opening the Accounts and Payments data sets using Python computer programing. Once we were able to open the data sets in Python, we used pandas to create data frames, clean the data, and eliminate null values when necessary. The first piece of information we looked at was the Overpayment column in the Payment data set. After learning all we could from the Overpayment category, we moved to Tableau so we could merge together the Accounts and Payments files. We decided to look into the top 5 brands which were Kirby, Rainbow, Saladmaster, Rainsoft, and Pr Home Products that account for a majority of the company and look into the loan amount and risk score within these 5 companies. Having a dataset of only the top 5 brands allowed us to open the file in Excel since it was small enough. Then we moved into looking at the distribution of loan amounts by delinquency periods for our top 5 brands. We then went back to python to clean the Accounts dataset of items in the ClosedStatus column that were null and calculated the percentage of our top 5 brand accounts that were written off. Finally, from there we looked into the loss rates for the delinquent data.

[-1]	Paid less than what was due
[0]	Paid exact amount due
[1]	Paid more than what was due

First, we looked into the overpayment data within the Payment data set. The Overpayment column contained Booleans of [-1], [1], and [0]. If there was a [-1] in the column, that would mean that the account did not pay the full amount that was due on time. If there was a [1] in the column, the account paid more than what was due at that time. If there was a [0] in the column, the account paid the exact amount that was due on time. We were interested in seeing if there was a trend in the number of underpayments, overpayments, and exact payments throughout each month. We wanted to see if there was a specific month or period of time where accounts were underpaying or overpaying. So, we created a count of each of the Booleans and created line graphs of that data. We then added regression lines to see if there were any significant trends to the graphs. In addition, the underpayment data did not contain a significant amount of data to draw any conclusions. The overpayment data had a significant amount of data but did not have a significant r-squared value. Due to the insignificant amount of data for the [-1] category and an insignificant r-squared value for the [1] category, we decided not to look into that data much further. The graph of the correct payment amounts per month had an r-squared value of 0.5331 so we decided to look at this category further. We then put the [0] data into Tableau in order to create the histogram below that contains a

Figure 1. Underpayment Regression Line [-1]

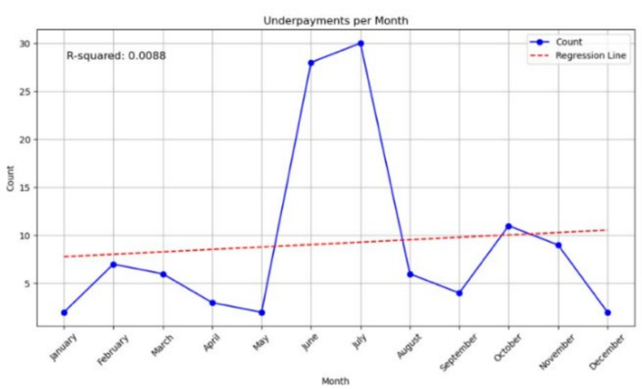


Figure 2. Overpayment Regression Line [1]

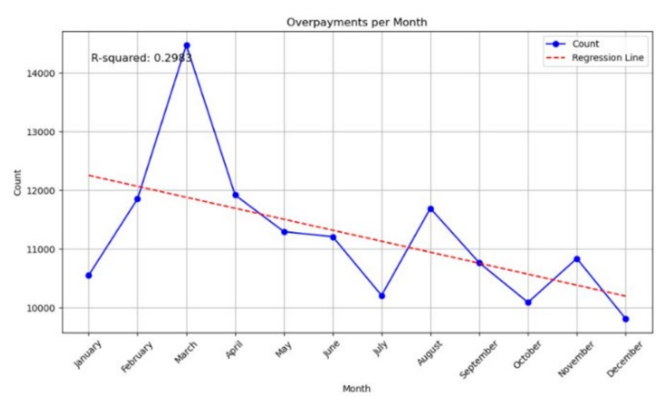
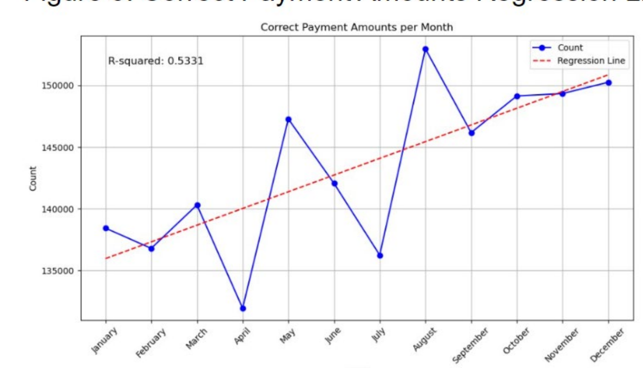


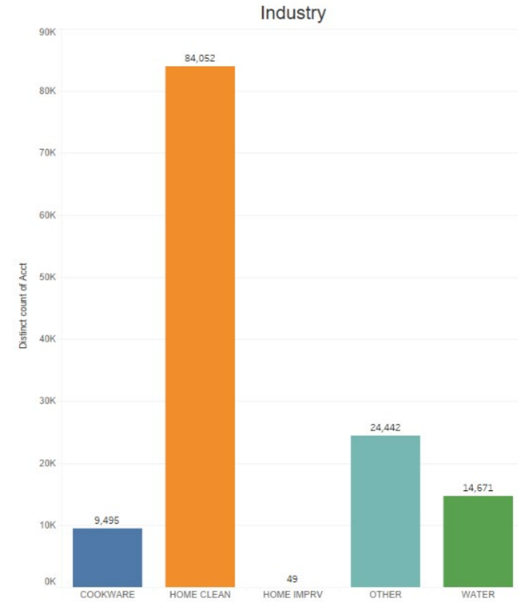
Figure 3. Correct Payment Amounts Regression Line [0]



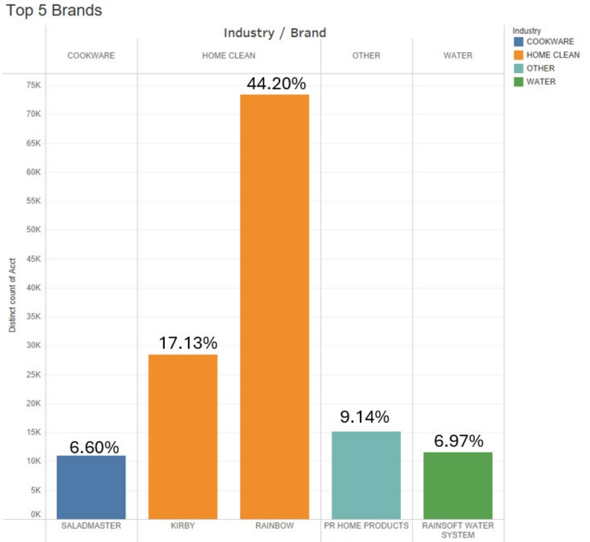
count of the data separated by industry so we could see which industry had the most correct payment amounts.

We decided to move into looking at the full Payment data set again but we needed to narrow down how much data we were working with because the file was too large to open in Excel. We decided to narrow down the data in python using pandas functions in order to determine the top brands that account for most of the company. When we looked at the overall account distribution by brand, there were five clear winners for the top brands: Kirby, Rainbow, Saladmaster, Rainsoft, and Pr Home Products. We categorized these brands by industry (cookware, home clean, water, and other) and identified them as significant based on their share of total accounts. These five brands account 84.04% of the company.

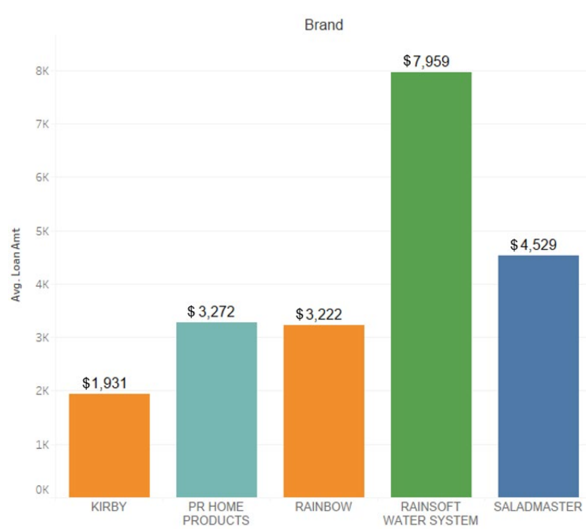
Figure 4. Correct Payments [0]



Top 5 Brands

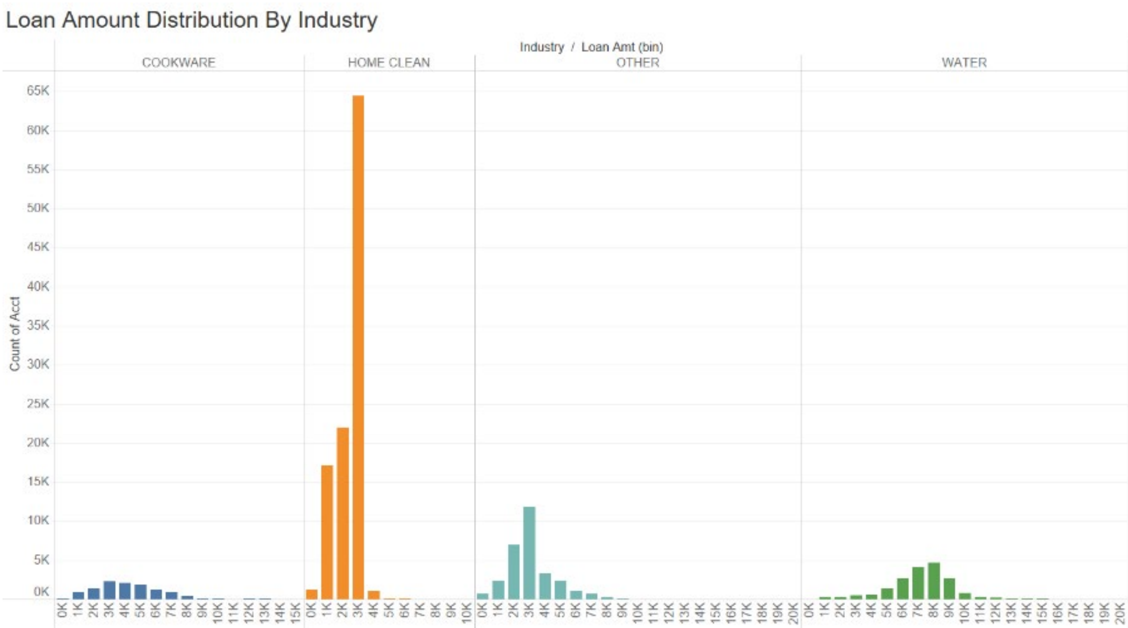


We dove further into the top five brand accounts by looking at the relationship between risk scores and loan amounts. A risk score for PCI is similar to a FICO score. It helps the company to determine whether or not they should give a loan to a company based on a variety of factors (fix definition of risk score). We wanted to see if there was any relationship between the loans given and risk score so we started by looking at the average loan amounts for each of the top five brands and their overall industry loan distribution. Using Tableau, we were able to calculate and display the average loan amounts, revealing Rainbow as the brand with the highest average loan amount of $7,959, followed by Saladmaster with $4,529, Pr Home Products with $3,272,

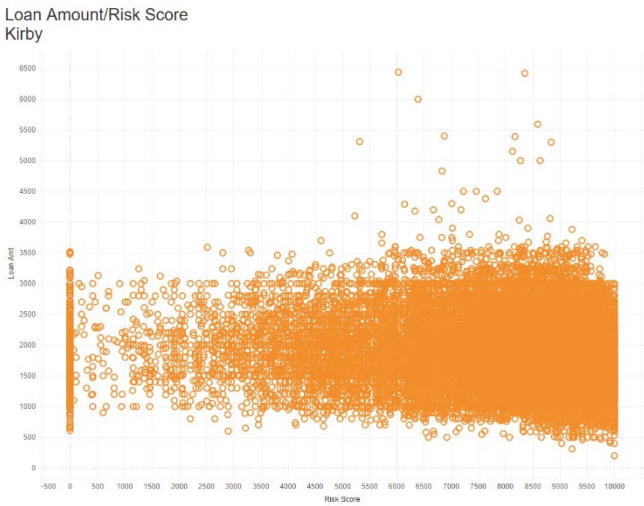


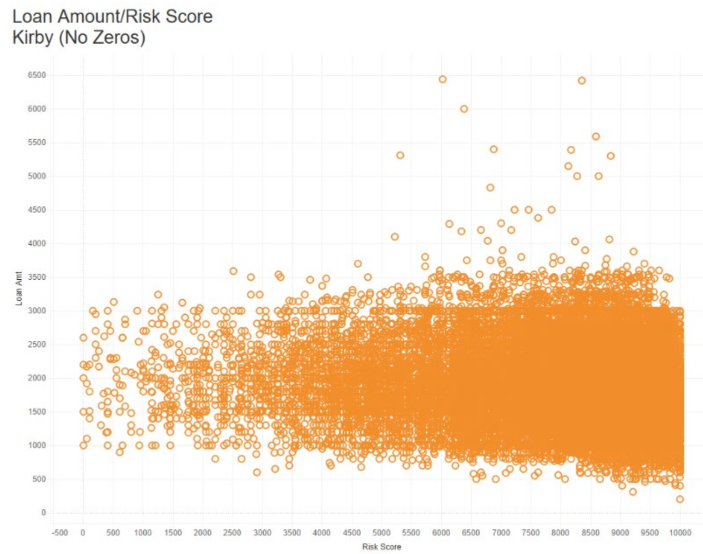
Rainbow with $3,222, and Kirby with $1,931. (Also talk about the risk score distribution by industry)

We also looked at the industry-specific loan distribution by using a histogram. With the exception of the Water industry, which had a peak around $8,000, Cookware, Home Clean, and Other showed peak loan amounts around $3,000. The higher loans with the Water industry is to be expected because the products sold at these companies are larger and more expensive purchases, like water softeners.

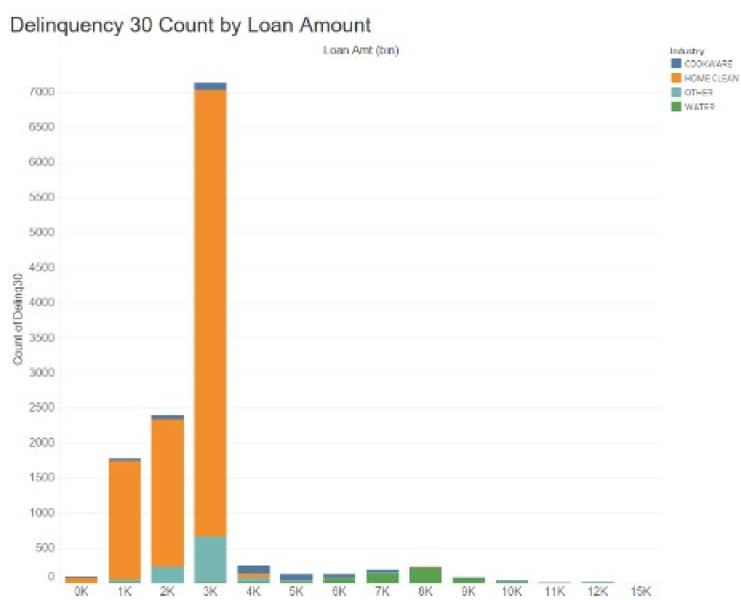


The analysis began by using a scatter plot to investigate the relationship between loan amounts and risk scores, recognizing the variability inherent in the dataset due to the variety of account numbers. Despite this variability, the scatter plot revealed useful insights, demonstrating a noticeable pattern in which larger loan amounts correlated positively with higher risk ratings. One interesting finding from the analysis was that there were cases where loans were approved even though the risk score was zero. Later conversations with PCI clarified the meaning of this score and compared it to a credit score, emphasizing the human element of risk evaluation. This human element is distinguished by the involvement of individuals in evaluating cases to determine creditworthiness, as illustrated by scenarios in which newcomers to the United States without a credit history may be rejected by computers but approved by human assessors based on additional context factors. The scatter plot on the left shows the distribution of Kirby's accounts that received loan funds, including those with a risk score of zero. The scatter plot on the right, on the other hand, eliminates any confusion and ensures a more focused analysis of creditworthy accounts by excluding the zero-risk score accounts and concentrating only on those that have a credit score.





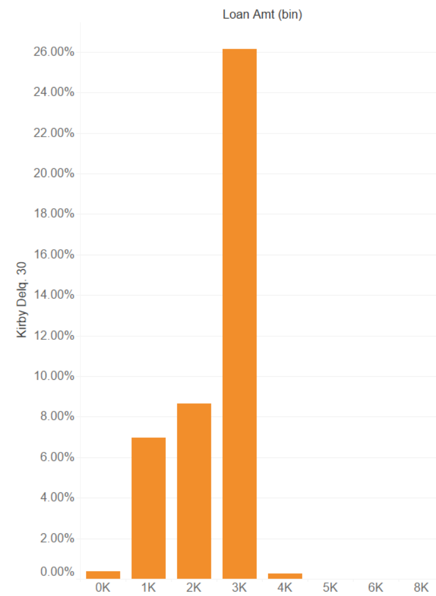
After many attempts to find a relationship between the Delinquency, Loan Amounts, and Risk Scores, we realized it was above our time limits and current capabilities. To find something else of interest, we narrowed our search to just comparing Delinquency and Loan Amounts. Delinquency is the period after an account fails to make a payment according to the agreed-upon terms. For the purpose of our project, we looked at periods of 30-, 60-, and 90-days delinquent, or passed due.



The first visual, shown on the right-hand side, depicts the distribution of loan amounts that reach a delinquency period of 30 days, looking just at our Top 5 Brands. This gave us an overarching idea of what we could expect to see from each and illuminated loan amount with the overall highest count of delinquent accounts from our Top 5 Brands, the $3,000 bin.

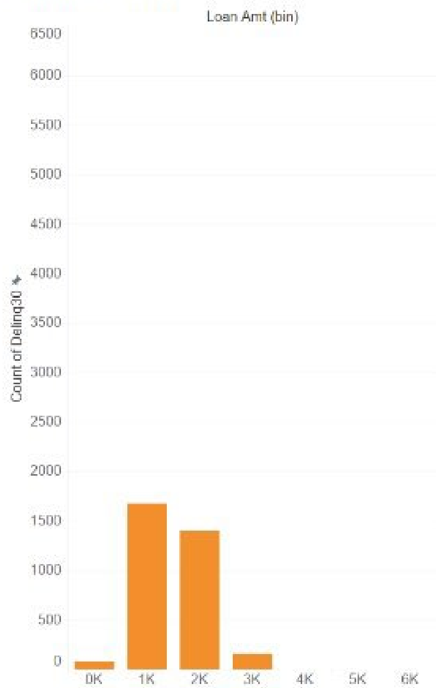
While the above figure gave the overall comparison, we knew we wanted to break it down by each of our Top 5 to get a better idea of what was going on at an individual level. The following graphs depict the percentage of delinquency 30 by loan amount. PR Home Products, Rainbow, Rainsoft Water Systems, and Saladmaster all showed expected trends, with peaks of less than 15% delinquency in any loan amount, and a frequency trend matching that of our average loan amounts. Kirby is the only one of our Top 5 Brands that really stood out, with 26% of their $3,000 loans going delinquent. With such a high percentage of delinquency, we knew we needed to dig deeper.

Percentage of Delinquency 30 by Loan Amount (Kirby)



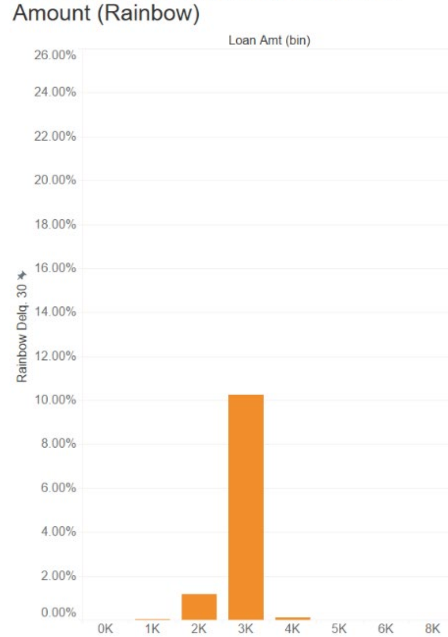
So, we contacted our partners at PCI, Kelli and Nate, and they were able to give a bit more context to our understanding, revealing that less than 4% of all of Kirby's loans fall in the $3,000 or above range. With that in mind, we created the visual on the left-hand side depicting the count of delinquent accounts by loan amount for just Kirby. This reiterated what Kelli and Nate had explained, showing that the actual count of Kirby's $3,000 loan was very low, with less than 300 total. In fact, Kirby's most frequently delinquent loans were in the $1,000 and $2,000 range. Now, looking at both visuals, a different story is told regarding the true risk of delinquency. While the 26% delinquency still demonstrates some level of risk, it is marketed as lower than what we had initially thought. This was an important moment of growth for us, as it taught us the value of being curious with data, and asking questions to reveal what is going on with the bigger picture.

Delinquency 30 Count by Loan Amount (Kirby)

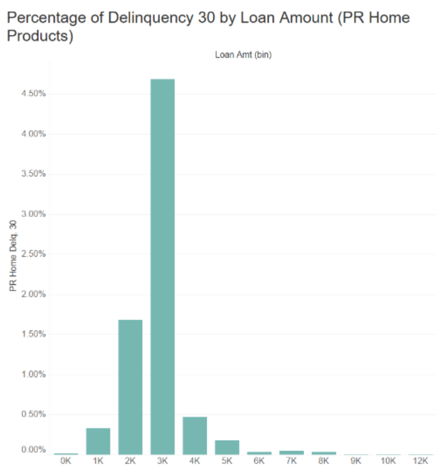


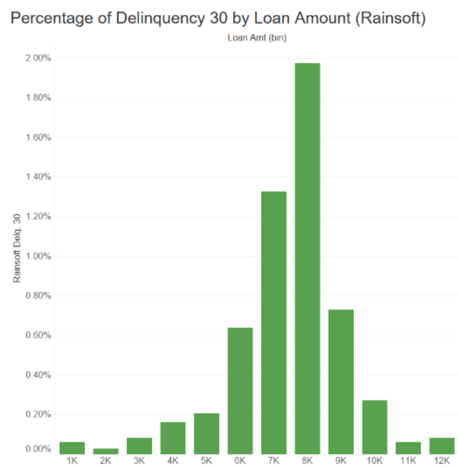
Again, the other four brands from our Top 5 showed very proportional visuals. To be sure, we also compared the percentage delinquency for Rainbow, another one of our Top 5 brands in the Home Clean industry, with its delinquency 30 counts by loan amount. The figure to the right depicts exactly what we

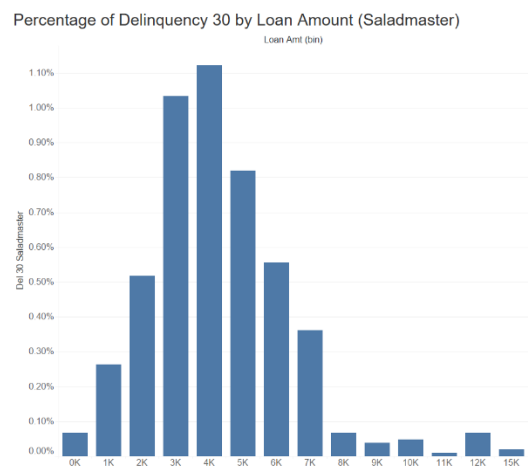
Percentage of Delinquency 30 by Loan Amount (Rainbow)



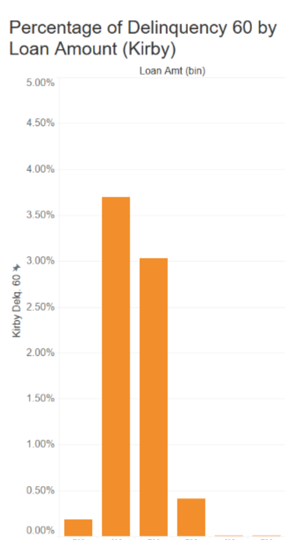
would have expected to see, with the two graphs being proportional and mimicking each other. Below are the graphs for the percentage of delinquency 30 for the remaining top 5 brands.

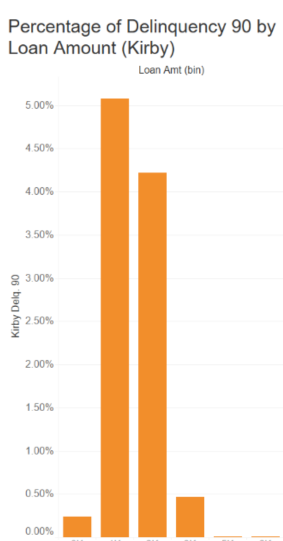


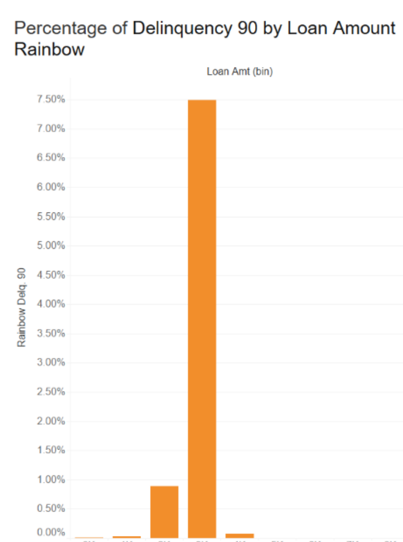


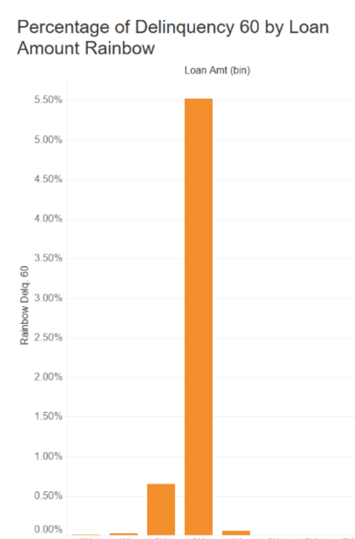


We then moved into looking at both the 60- and 90-day delinquency periods. These showed similar trends across all five of our top brands, with reduced percentages but the same general shape for each as seen in the 30-day delinquency period. Below are the graphs we made for both Kirby and Rainbow in the 60- and 90-day delinquency periods. The percentages are down to less than 5% delinquent, which follows what we expected to see, as there should be fewer accounts that make it to 60- and 90-day delinquent.









	Kirby	PR	Rainsoft	Rainbow	Saladmaster
# of Accounts	26,619	14,632	11,522	64,898	10,695
# of Accounts Written Off	2,276	599	436	4,321	447
% of Accounts Written-Off	**8.55%**	**4.09%**	**3.78%**	**6.66%**	**4.18%**
Total Value of All Loans to Brand	$ 49,600,579.10	$50,159,726.62	$ 92,756,187.84	$ 212,826,148.08	$ 52,412,106.32
Value of Loans Written Off	$ 4,077,277.08	$ 1,861,006.71	$ 3,476,173.46	$ 5,654,878.57	$ 357,406.96
Percent of Brand Loan Written off	**8.22%**	**3.71%**	**3.75%**	**2.66%**	**0.68%**

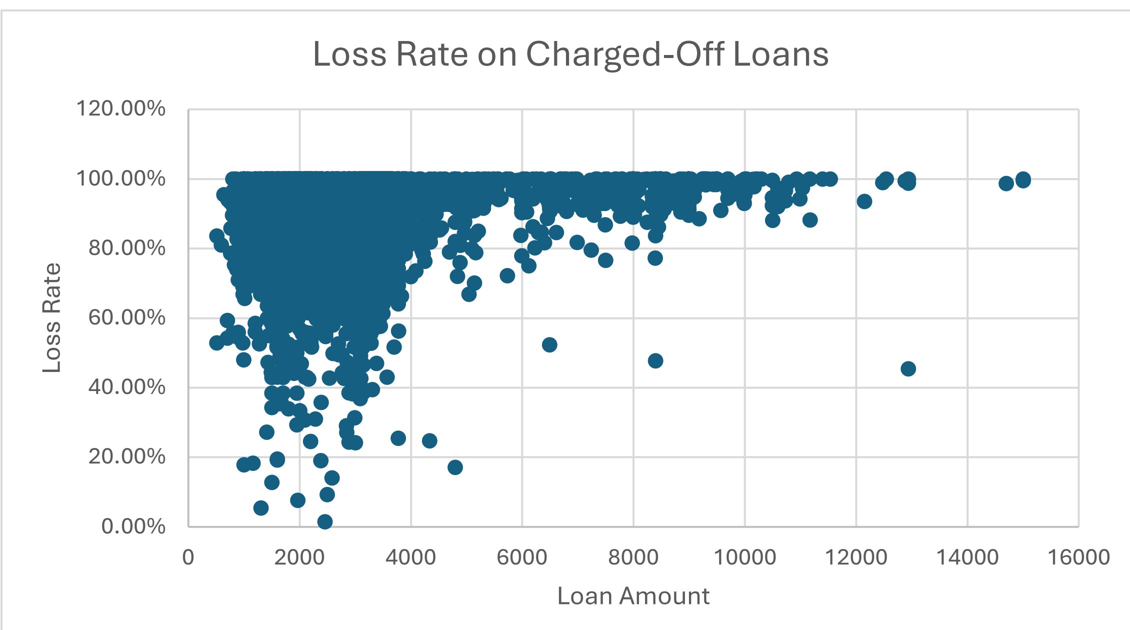
TOTAL Write-Off Worth	**$15,426,742.78**

While the delinquency information was interesting, it only really showed us the accounts that had failed to make payments in a timely manner. To see the accounts that lost PCI money, we transitioned into looking at the written-off accounts. An account is considered written off when PCI acknowledges that the customer will not be paying the loan back, and the debts are unrecoverable. We learned that for PCI, this is any time after the 120-day delinquency period. The table above shows our work with written-off accounts and how they compare between our Top 5 Brands. First off, we took the total number of accounts for each brand and divided it by the number of accounts that were written off for each to get the percentage of accounts that were written off. This gave us a comparable measure for all five brands and allowed us to see that Kirby had the highest percentage of written-off accounts.

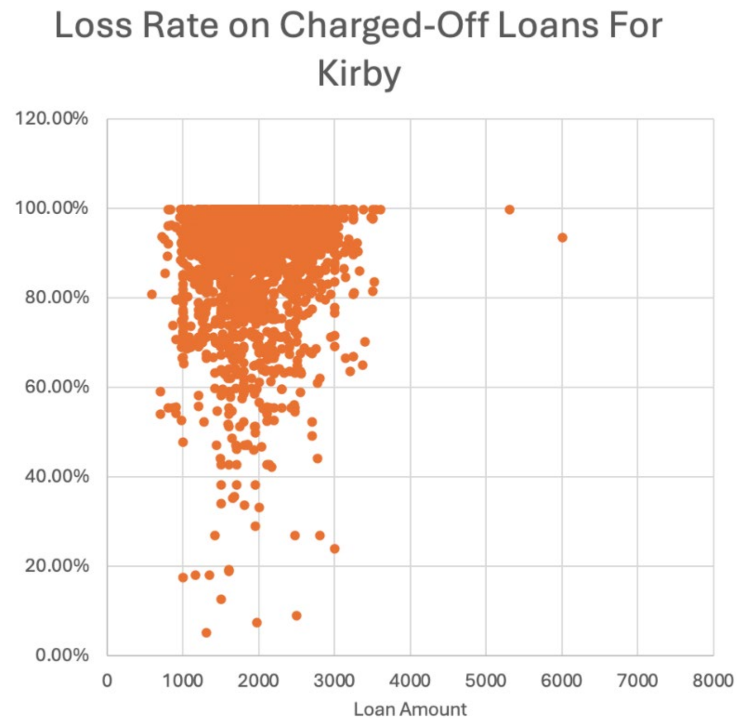
While this was helpful, it ultimately was not a great comparison yet, as the loan sizes for each brand greatly vary, and could have given us misleading results. To combat this and give our data some context, we looked at the actual value of the loans and divided them by the total values lost for the loan for each brand. Kirby, PR Home Products, and Rainsoft Water System all had similar results, with less than a 1% change for each. However, both Rainbow and Saladmaster had substantial changes with Rainbow having more than 6% of their accounts written off, but less than 3% of their actual loan worth written off. Similarly, Saladmaster had more than 4% of their accounts written off, but less than 1% of their total loan value written off. The distinction between the two metrics improved the quality of our comparisons.

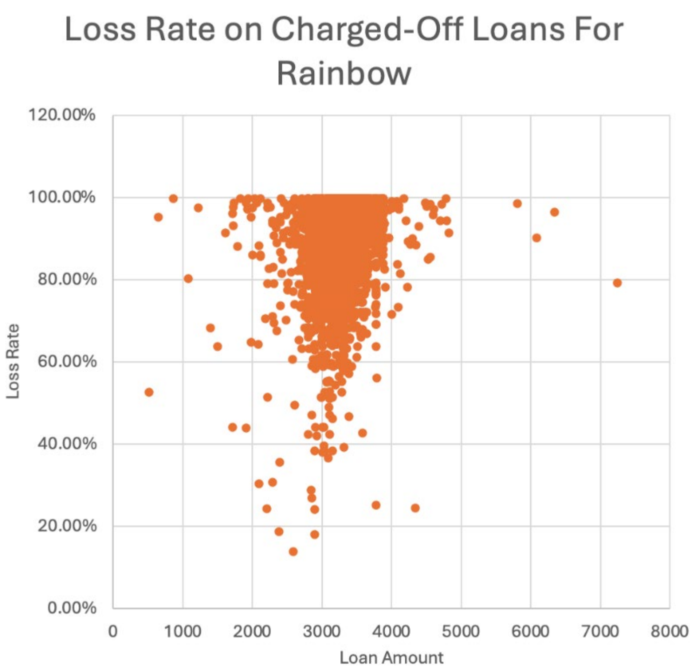
Lastly, we decided to look at the loss rates related to the written-off and charged-off accounts. For reference, both written off and charged off have the same definition meaning that PCI is not expecting to get any money back from the loan and will not be viewed as expected future income to the company. PCI uses both terms in their data sets, so for this analysis charged off will be used as the general term for these accounts. Using the CSV file that was cleaned of null values that we were able to source from Python, we were able to isolate the charged-off accounts which then allowed us to calculate the loss rates on loans related to these accounts. In order to calculate the loss rates on the loans we took the amount that was charged off and divided it by the initial loan amount. When we first did this calculation, some of the quotients came out to be over 100% which didn't make logical sense as PCI shouldn't be charging off amounts greater than what the initial loan amount was. After speaking to our partners at PCI, we were informed that loans could receive add-ons to the initial loan amount which would result in the charge-off amount being greater than the initial loan amount. In the end, we decided to exclude these values from our analysis for three reasons. The first is that it skewed the visualizations we made with this data in a way that made them illegible. Secondly, there were only nine loans out of thousands of data points that came out to be greater than 100%, thus classifying these loans as outliers. Lastly, after asking our partners at PCI, we were given approval to exclude these values from the data set which we did in order to provide cleaner images of the data.

Looking at the graphs themselves, we're able to see the patterns and points of interest that occur within these charged-off accounts. The first graph shown below shows the loss rates for the top five brands put together. By doing this, we're able to see the trends of the loss rates across all of PCI. Taking a closer look at this graph, a majority of the data takes place in the \$2000-\$4000 range. This lines up with the average loan amounts found earlier in the analysis. A majority of the data also lies near the higher loss rates as most charged-off loans end up losing most if not all of the initial loan amount due to delinquent customers. As the loan amount increases, it begins to stay above the 80% mark in terms of loss rate. This occurrence makes sense as loans with higher amounts are more risky for PCI to give out, and it is more likely at these loan amounts to see customers not pay at all if they show trends of delinquency since the creation of their loan.

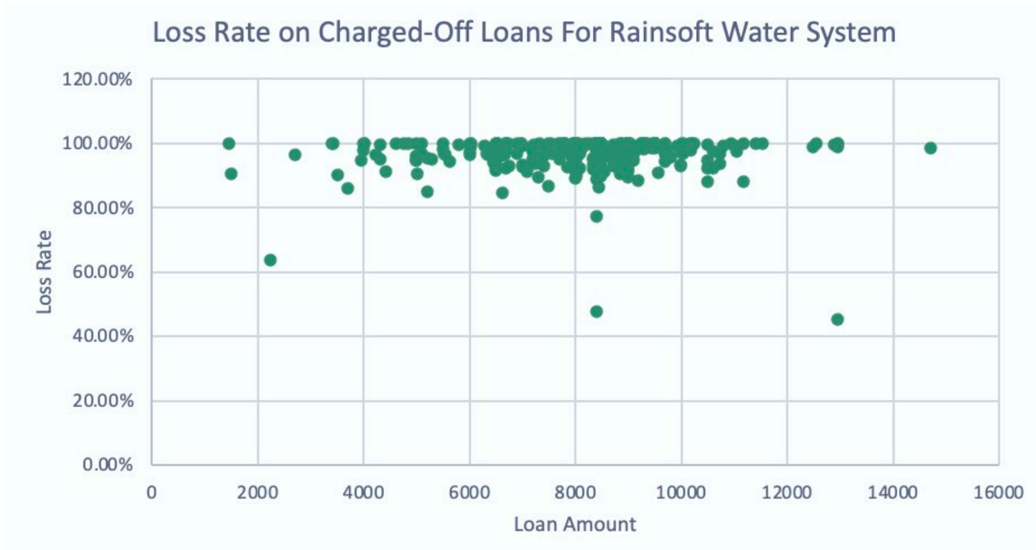


To give the graph above a little more context, we decided to create the same visualization for each of the top five brands which can be seen down below. For this analysis, Kirby and Rainbow can be looked at together due to the similar manner of their distribution. They both have a majority of their data around the average loan amount mentioned earlier as well as having a normal distribution throughout the loss rates. This means that there is nothing too out of the norm that warrants further looking into. PR Home Products had a similar distribution to these two companies, resulting in us coming to the same conclusions in its analysis as Kirby and Rainbow.

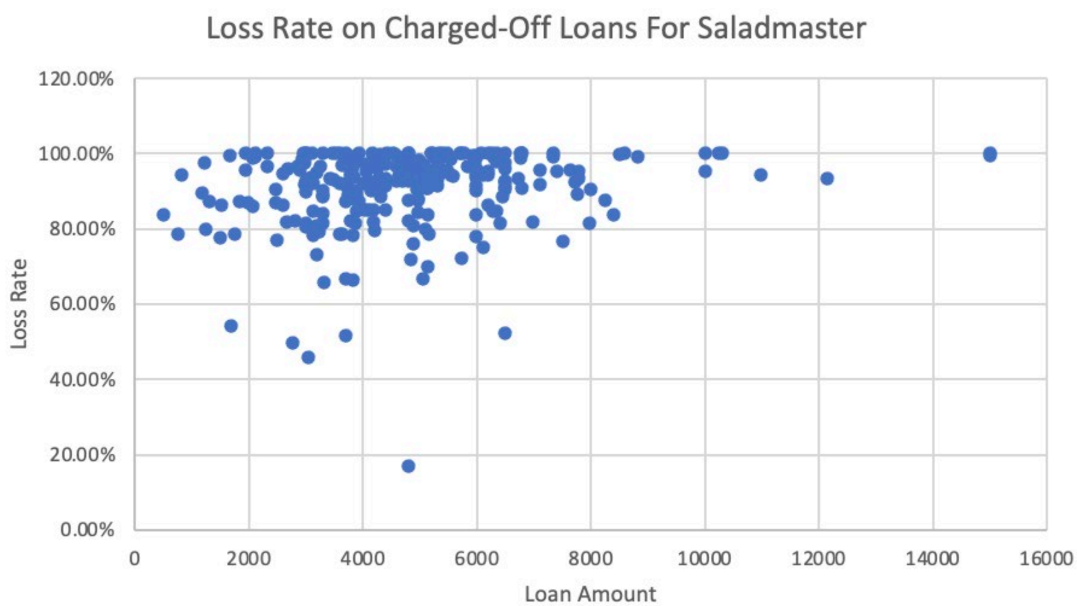




These last two graphs show the visualizations for Rainsoft Water System and Saladmaster. The distributions for these two companies look much different than the graphs above. A majority of the data sits above the 80% mark in terms of loss rate for both cases. The loan amounts for both graphs are higher compared to Kirby and Rainbow as well. Even though the distributions look different, we were not too concerned with these results. High loan amounts given out by PCI are riskier as they're more likely to lose a greater proportion of the loan if they're not paid back. This is reflected in the higher loss rates across these two graphs seen below.



Loss Rate on Charged-Off Loans For Rainsoft Water System
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
Loss Rate
0
2000
4000
6000
8000
10000
12000
14000
16000
Loan Amount



Loss Rate on Charged-Off Loans For Saladmaster
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
Loss Rate
0
2000
4000
6000
8000
10000
12000
14000
16000
Loan Amount