**Executive Summary**

1. **Review Existing Eligibility Criteria**
2. The top factors include VANTAGE\_SCORE, DAYS\_PAST\_DUE, CREDIT\_LIMIT, TOT\_NET\_REV
3. The values are VANTAGE\_SCORE >= 630, CREDIT\_LIMIT > 6500, DAYS\_PAST\_DUE = 0
4. We used Python to do a correlation analysis and a summary of statistics for relevant factors.

1. **Identify Performance Metrics**
2. The top factors include TOT\_NET\_REV, FUEL\_NET\_REV, CURRENT\_BALANCE, UNBILLED\_BALANCED, CLI\_AMOUNT, WO\_AMOUNT
3. The values vary, with the mean of each factor being 305.49, 4300, 444.49, 1925.17, 12441.91, 12910.08.
4. We used Python to get the statistics for these factors.

1. **Sensitivity Analysis**
2. The top factors include DAYS\_PAST\_DUE, CLI\_AMOUNT, CREDIT\_LIMIT.
3. To flag bad payments (BAD\_PAYMENT\_FLAG), DAYS\_PAST\_DUE > 30, CLI\_Percentage = CLI\_AMOUNT / CREDIT\_LIMIT, CLI\_BIN is cut by quartiles.
4. We used R to determine the values and visualize the differences between the two groups.

1. **Spend and loss difference between parent accounts**
2. The top factors include 'TOT\_NET\_REV', 'WO\_AMOUNT', 'DAYS\_PAST\_DUE', 'TOT\_SPEND', 'NONFUEL\_SPEND', 'FUEL\_SPEND'.
3. The values are
4. We used Python to get the statistics for these factors and used them to summarize our findings.

1. **Cross-Sell Performance dashboard**
2. To visualize performance data, variables used were TOT\_NET\_REV, FUEL\_NET\_REV, TOT\_SPEND, NONFUEL\_SPEND, WO\_AMOUNT. These factors were chosen as they are related to customer financial performance.
3. We used bar charts and line charts in Power BI to create this dashboard. Bar charts were the most effective way of displaying the data with different variables such as parent account and segment. Line charts were used to show trends over time.

1. **Re-evaluate Eligibility Variables**
2. The factors include 'VANTAGE\_SCORE', 'CREDIT\_LIMIT', 'NSF\_PMTS', 'DAYS\_PAST\_DUE', 'PAYDEX', 'LOCK\_DAYS', 'DUE\_DAYS', 'TERM\_DAYS', 'CLI\_AMOUNT', 'TOT\_SPEND', 'NONFUEL\_SPEND', 'FUEL\_SPEND', 'NSF\_AMT', 'NO\_OF\_PAYMENT', 'PAYMENT\_AMOUNT', 'BAL\_1\_30', 'BAL\_31\_60', 'BAL\_61\_90', 'BAL\_90\_PLUS', 'SEGMENT', 'CURRENT\_BALANCE', 'UNBILLED\_BALANCED'.
3. For the factors used, we found their coefficients and new eligibility criteria for some of them.
4. We used Python to create a linear regression model and CHAID analysis to evaluate factors that correlate with customer performance and make a list of variables to add in the next step.

1. **Apply revised eligibility model based on new cutoffs and variables (if any)**
2. The factors include 'TOT\_NET\_REV','CREDIT\_LIMIT', 'DAYS\_PAST\_DUE', 'PAYDEX', 'LOCK\_DAYS', 'DUE\_DAYS', 'TERM\_DAYS','NO\_OF\_PAYMENT', 'PAYMENT\_AMOUNT'
3. Identified 2 sets of cutoffs to detect “swap in” and “swap out” customers.
4. Used Python to determine whether customers were “swap in” or “swap out”

1. **Adjust Cutoff Values**
2. The factors include 'TOT\_NET\_REV','CREDIT\_LIMIT', 'DAYS\_PAST\_DUE', 'PAYDEX', 'LOCK\_DAYS', 'DUE\_DAYS', 'TERM\_DAYS','NO\_OF\_PAYMENT', 'PAYMENT\_AMOUNT'
3. DUE\_DAYS > 14, DAYS\_PAST\_DUE <= 0, TOT\_NET\_REV <= 637.32, WO\_AMOUNT <= 1557, CREDIT\_LIMIT <= 6500, LOCK\_DAYS <= 4
4. To readjust the cutoff values, we looked at the statistical summaries of each of the variables that were in the decision tree. We used Python to summarize the statistics.

1. **Recommendations**

From our analysis, we recommend that Fleetcor use combinations of variables to determine which customers to include as swap-in and swap-out. This will help make sure that customers are satisfied with the services they are being provided with, while maximizing profit for Fleetcor.

**Main Section**

Data Integration

To prepare the data for analysis, we took multiple steps. Our first step was to convert every Excel file into a csv file. This was because we used SQL Server Management Studio to create our database. Once the csv files were imported into the database, we created a primary key for each table. To model the relational database, we connected the Cross\_Sell\_Acct\_Info table to all the other tables using the ACCT\_CODE column. A similar process was done for the Non\_Cross\_Sell tables.

To get the data for analyses, we created 2 views; Cross\_Sell\_View and Non\_Cross\_Sell\_View. Our view has a total of 700,000 rows. We created a new dataset in Python that removed all NAs and ended up with 99,590 rows.

Deliverable 1 & 2

To review the existing eligibility criteria and identify performance metrics, we first looked at the data dictionary. We chose a few variables that we believed could impact performance. To find the values, we summarized the data as shown below:

From this, we can start to see some possible thresholds for the factors and eligibility criteria. To further analyze their impact on total revenue, we made a correlation matrix between these factors to determine which ones are most closely related to performance.

Deliverable 3

By creating a sensitivity analysis, we can see the bad rate differences between opt-in and non-opt-in customers. Using R, a new variable was created to flag whether each customer had a bad payment or not. If the value of DAYS\_PAST\_DUE was greater than 30, it was flagged with 1 as a bad payment. However, if DAYS\_PAST\_DUE was 30 or less, it was not flagged and a 0 was added. Another variable was created to categorize CREDIT\_LIMIT into four different bins (Low, Medium, High, Very High). This was calculated by quartiles. Both variables were created to both datasets. The values were then plotted to compare the difference between Opt-In and Non-Opt-In customers.

From the graphs, we can see some trends and patterns from both groups. For opt-in customers, the bad rate decreases as the credit limit increases, however there is a slight increase for the ‘Very High’ credit limit group. This suggests that opt-in customers with higher credit limits are usually more reliable when making payments, despite there being a slight increase at the very high credit limit.

Non-opt-in customers show us that there is less variation between the low, medium, and high credit limit. There is a significant jump to the very high credit limit bin. This suggests that there is a high bad rate for non-opt-in customers, especially as the credit limit increases.

Overall, these findings indicate that opt-in customers have better payment behaviors than non-opt-in customers as their credit limit increases, while non-opt-in customers show higher average bad rates, and increased delinquency rates as credit limit increases.

We also analyzed the bad rate difference by different CLI% for opt-in customers.

This graph suggests that as CLI% increases, the bad rate among opt-in customers decreases. Similarly, to the graph above, it suggests that opt-in customers are more likely to make payments as their credit limit increases.

To calculate these variables, CLI\_Bin was cut into quartiles using a new variable CLI\_Percentage. CLI\_Percentage was calculated by dividing CLI\_AMOUNT by CREDIT\_LIMIT.

Deliverable 4

To analyze and compare spend and loss between different parent accounts, we created two new data frames that separated rows by the column ‘FUEL\_ONLY\_PARENT\_ACCT’. We used the factors that related to spend and loss, which are: 'TOT\_NET\_REV', 'WO\_AMOUNT', 'DAYS\_PAST\_DUE', 'TOT\_SPEND', 'NONFUEL\_SPEND', 'FUEL\_SPEND'.

Universal card:

Fuel-only:

Summary of findings:

* Universal card has lower average total spend (UC = $4600 | FUEL = $$9829)
* Universal card has significantly lower fuel spend (UC = $894 | FUEL = $9829)
* Average write-off amount is higher for Universal card accounts (UC = $12969 | FUEL = $8779)
* Average days past due is lower for Universal card accounts (UC = 23.3 | FUEL = 28.3)
* Mean total net revenue is higher for fuel-only accounts (UC = $282 | FUEL = $1870)

Deliverable 5

For the Cross-Sell Performance Dashboard we used Power BI and used variables including write-off amount, total spend, total fuel spend, total revenue, total non-fuel spend, average days past due, count of customers, and opt-in date. This dashboard helps us visualize and give us an overview of the data we by their parent account information and their segment.

Deliverable 6

For linear regression we chose to do two versions. One version used ‘SEGMENT’ as the target variable, whereas the other used ‘TOT\_NET\_REV’. This would help us determine which variable should be used as our target variable and would help us evaluate the relationship between the predictor variables and the target variable.

We identified a set of potential variables that would be used for both versions. In the model where ‘TOT\_NET\_REV’ was the target variable, the RMSE was 520.05 and the RMSE with ‘SEGMENT’ as the target variable was 13.59. For decision tree analysis, we will use ‘SEGMENT’ as the target variable because of the lower RMSE value.

This is the output of our linear regression model, which shows predictors and their coefficients, as well as the RMSE on the first line.

The second model we chose was a decision tree analysis. It was chosen as it could help us identify key predictors and determine thresholds for re-evaluating the eligibility criteria and identifying swap-in and swap-out customers.

From the regression analysis, we decided to use ‘PAYDEX’, ‘NO\_OF\_PAYMENT’, ‘LOCK\_DAYS’, ‘DUE\_DAYS’, TERM\_DAYS’, ‘TOT\_NET\_REV’, ‘CREDIT\_LIMIT’ as the variables for the decision tree analysis.

Decision tree with depth of 4

Deliverable 7

From the decision tree with a depth of 4, we identified 94183 customers for “swap in” and 5407 for “swap out”.

These are the variables and cutoffs used to determine which accounts were swap-in and which were swap-out.

After applying the new eligibility criteria, we found the following information:

Deliverable 8

For both swap-in and swap-out customers, the expected revenue was negative, which resulted in an expected net loss for both groups. As we are trying to maximize profits, we will adjust the cutoffs. These are the new cutoffs that reduce the write-off amount for each customer.

After applying the revised cutoffs, these are the new expected revenue and write-off amounts.

Although the expected net value is a loss, it is significantly less when applying the new cutoff amount. These new amounts were determined by using the quartiles in the metrics identified in the first two deliverables.

Deliverable 9 – Recommendations

From the model and current dataset, the number of customers for swap-in is 5185 and the number of customers for swap-out is 94405. For both sets of customers, there is an expected net loss from revenue and write-off amount.

There are a number of risks that are related to using this data and model. These include market variability, customer behavior, changes in customer data that could make the model’s thresholds outdated, and ethical considerations. To mitigate these risks, we need to regularly monitor new data and changing external conditions and monitor the performance of the swap-in and swap-out customers and make necessary changes.

**Group Names**

* Lucia Placidi