Data Science

Take Home Assessment

Recruiting Season 2023

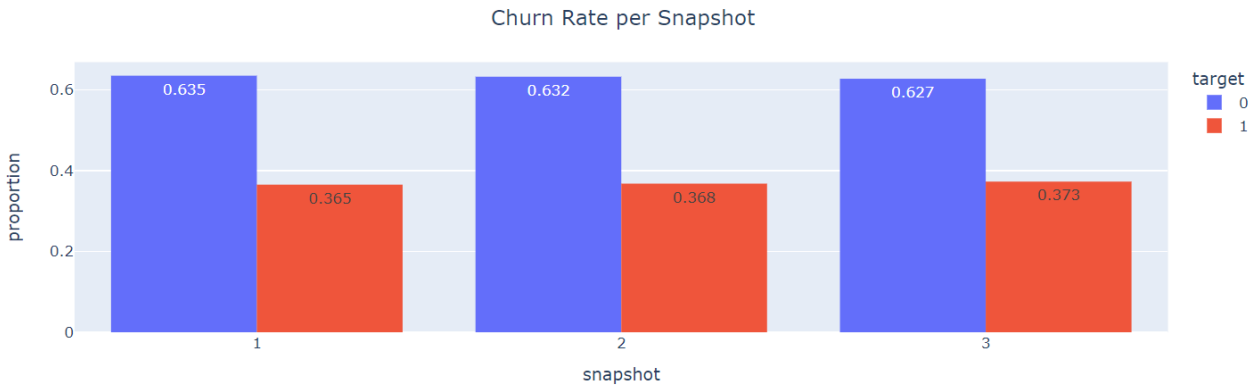
Project: Green Banana Company

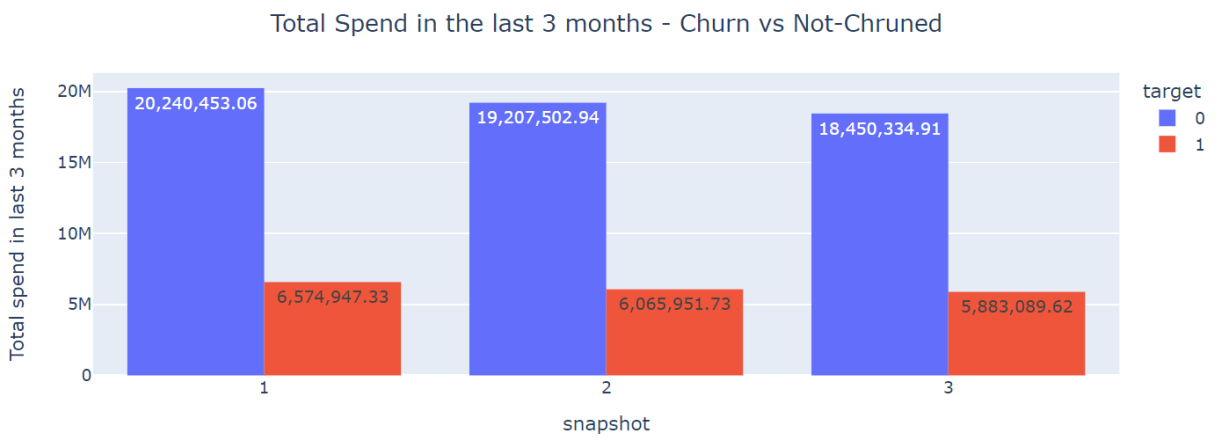
By: Marvin Lomo

Green Banana

High-Level Summary

- Green Bana Co. has a high churn rate problem. Approximately **36.8%** of customers churn from the available data.
- Churned customers are responsible for about 25% of total spending in the last 3 or 6 months, per the data.
- This project has three objectives:
  - ✓ Prove out the **business value** of investing in a churn model
  - ✓ Determine **features contributing** to Customer churn
  - ✓ **Develop a preliminary churn model**





Business Value

Spend Analysis

- On average, churned customers spend USD 37.83 per shipment.
- Given that the number of active shipments for churned customers is 348,476, the company loses **USD 13,182,847.08**

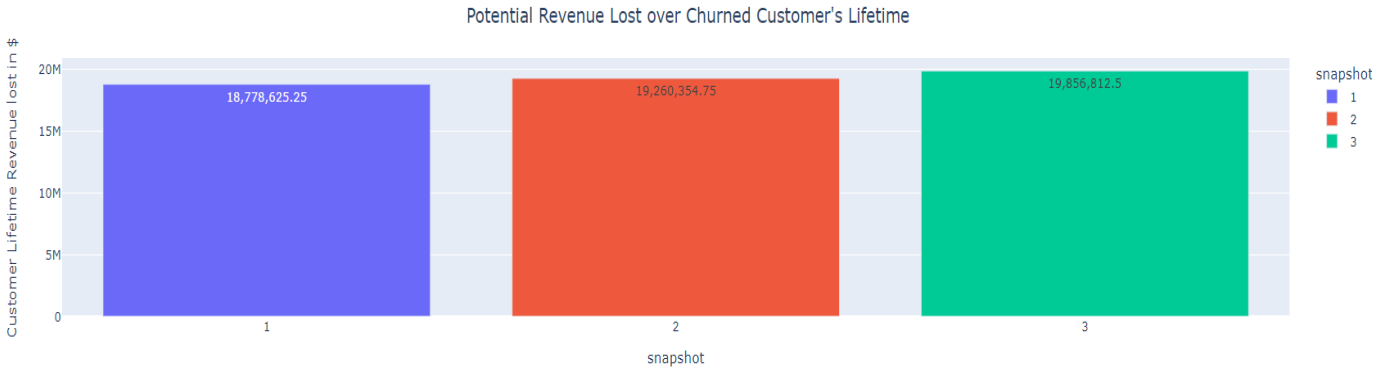
Salvage Perc	Savings
2.5%	$ 329,571.18
5%	$ 659,142.35
10%	$ 1,318,284.71
12.5%	$ 1,647,855.89
15%	$ 1,977,427.06

Customer Lifetime Value Analysis

- As it costs less to retain an existing customer than it does to retain a new one, I computed the CLV using:

$$AVG\ Purchase\ Freq * AVG\ Customer\ Value * AVG\ Customer\ Lifespan$$

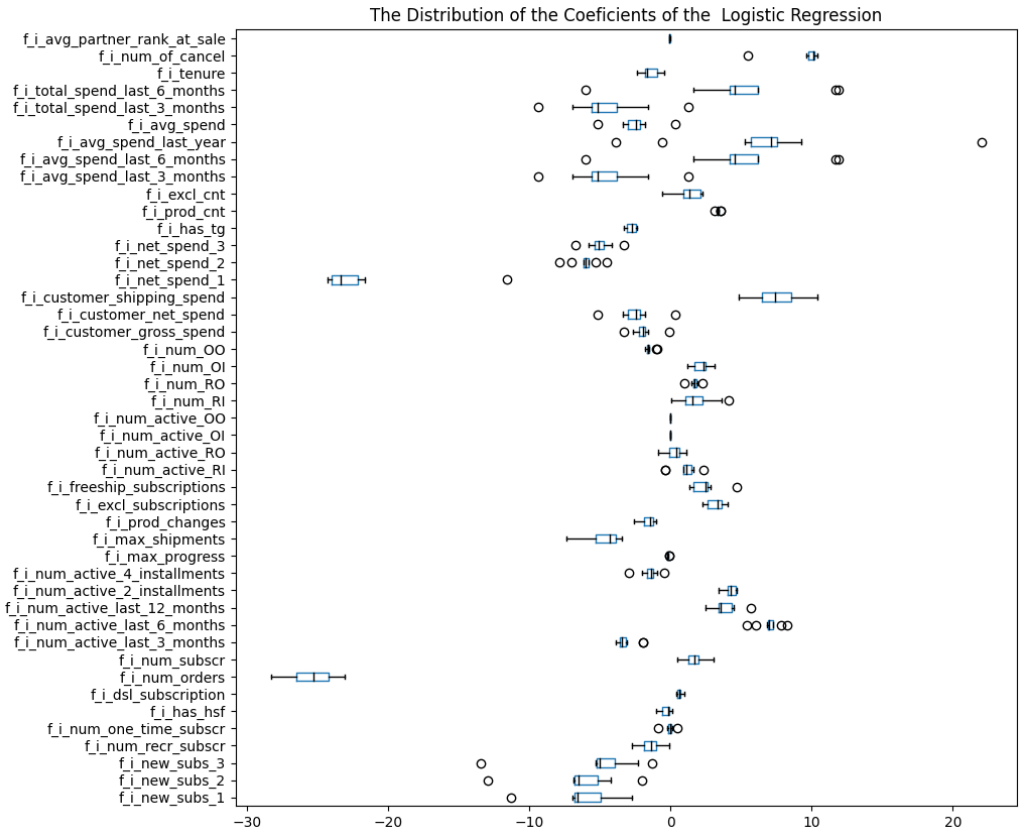
- From the above, the **average CLV** is **USD 215.25.** The chart below shows the potential revenue lost per snapshot over the lifetime of the churned customer.



Churn Model Solution

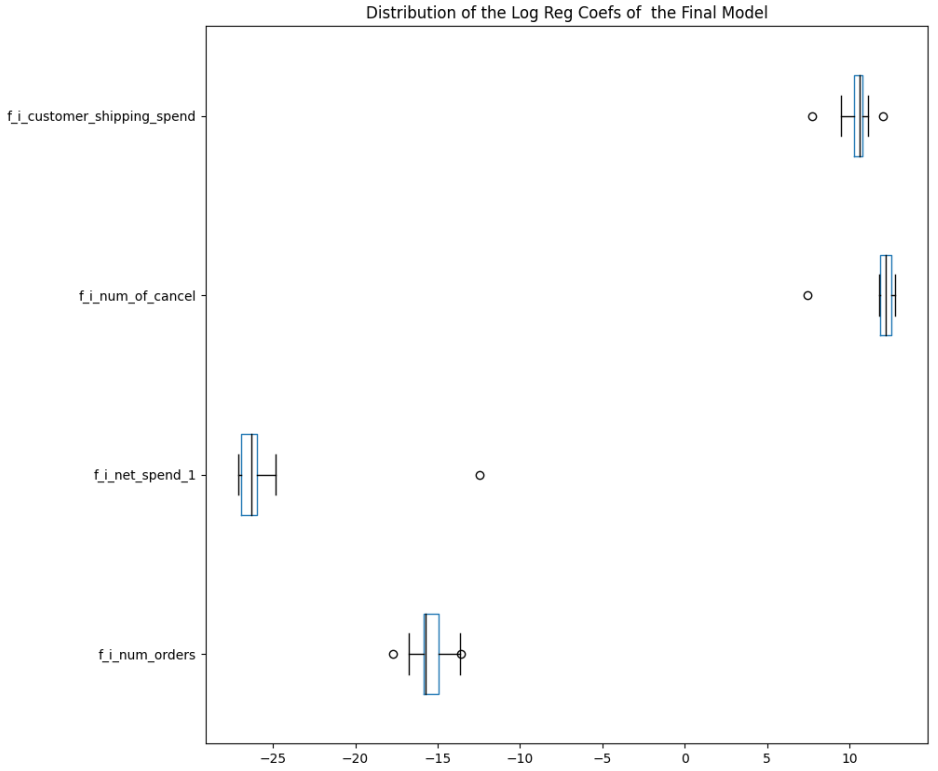
- Three major objectives underpinned model building process:
  - ✓ Determine Factors that contribute to churn
  - ✓ Build an MVP model
  - ✓ Build a business case for the model
- To build the model, the following processes were followed:
  - ✓ All variables were scaled
  - ✓ Missing features were removed
  - ✓ K-fold CV was used to ensure generalizable results

Model	Training Accuracy	Test Accuracy	Interpretation	Decision
Baseline (Dummy Classifier)	0.637 ± 0.00	0.637 ± 0.00	This is the baseline model	
Logistic Regression Classifier	0.714 ± 0.00658	0.637 ± 0.106	Model overfits the data. However, the std dev of test accuracy is high	Accept
Finetuned Logistic Regression Model	0.655 ± 0.0352	0.608 ± 0.0716	Model overfits the data with low test accuracy	Reject
Ridge Classification Model	0.694 ± 0.0069	0.633 ± 0.0907	Model overfits the data with low test accuracy	Reject
Finetuned Ridge Classification Model	0.701 ± 0.00585	0.626 ± 0.101	Model overfits the data with low test accuracy	Reject
Logistic Regression Model with Polynomial Features (degree=2)	0.732 ± 0.00665	0.624 ± 0.128	Model overfits the data. However, the standard deviation of the test accuracy is high	Reject



The Final Model and Recommendations

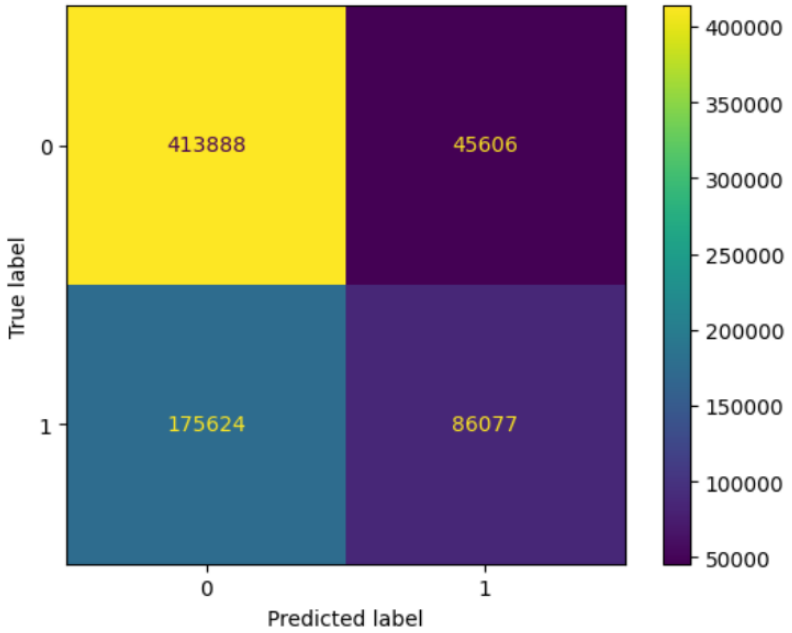
- From the analysis in the previous slide, the final model is a logistic regression classifier with the following features:
  - ✓ Number of orders
  - ✓ Net spend in the last 30-day window
  - ✓ Number of order cancellations
  - ✓ Customer total spend on shipping
- **Training Accuracy for this model is 0.695 ± 0.00622**
- **Testing Accuracy for this Model is 0.682 ± 0.0853**



- From the final model, my recommendations are:
  - ✓ The firm must critically examine the **customer spend on shipping**. Higher shipping costs contribute to churn (free shipping options).
  - ✓ The firm must institute policies to reduce cancelled orders (**promotions or free shipping**).
  - ✓ The firm can increase promotional strategies to increase net spend and total customer orders.

The Business Case for the Final Model

- The model's sensitivity, i.e. the ability of the model to correctly predict churned customers, **is 33%.**
- The table shows that for Green Banana Co to profit from the model, the strategies/recommendations outlined earlier should convert/retain at least **4% of potential churn candidates.**



Snapshot Analysis		
snapshot no:		3
Number of churned Customers		92,250
Customer Lifetime Value		215.25
Potential Revenue lost by firm	Customer Lifetime Value * Number of churned Customers	19,856,812.50
Sensitivity of the Churn Model		33.00%
Number of Churned Customer Correctly Predicted	Sensitivity of the Churn Model * Number of churned Customers	30443
salvaged after prediction with Churn Model	Correctly Predicted * Customer Lifetime Value	6,552,748.13

Sensitivity analysis		
Salvage Percentage	Potential Savings	Decision
1.00%	$ 65,527.48	Loss
1.50%	$ 98,291.22	
2.00%	$ 131,054.96	
2.50%	$ 163,818.70	
3.00%	$ 196,582.44	
3.50%	$ 229,346.18	
4.00%	$ 262,109.93	Profit
4.50%	$ 294,873.67	
5.00%	$ 327,637.41	
5.50%	$ 360,401.15	
6.00%	$ 393,164.89	
6.50%	$ 425,928.63	
7.00%	$ 458,692.37	

Thank you