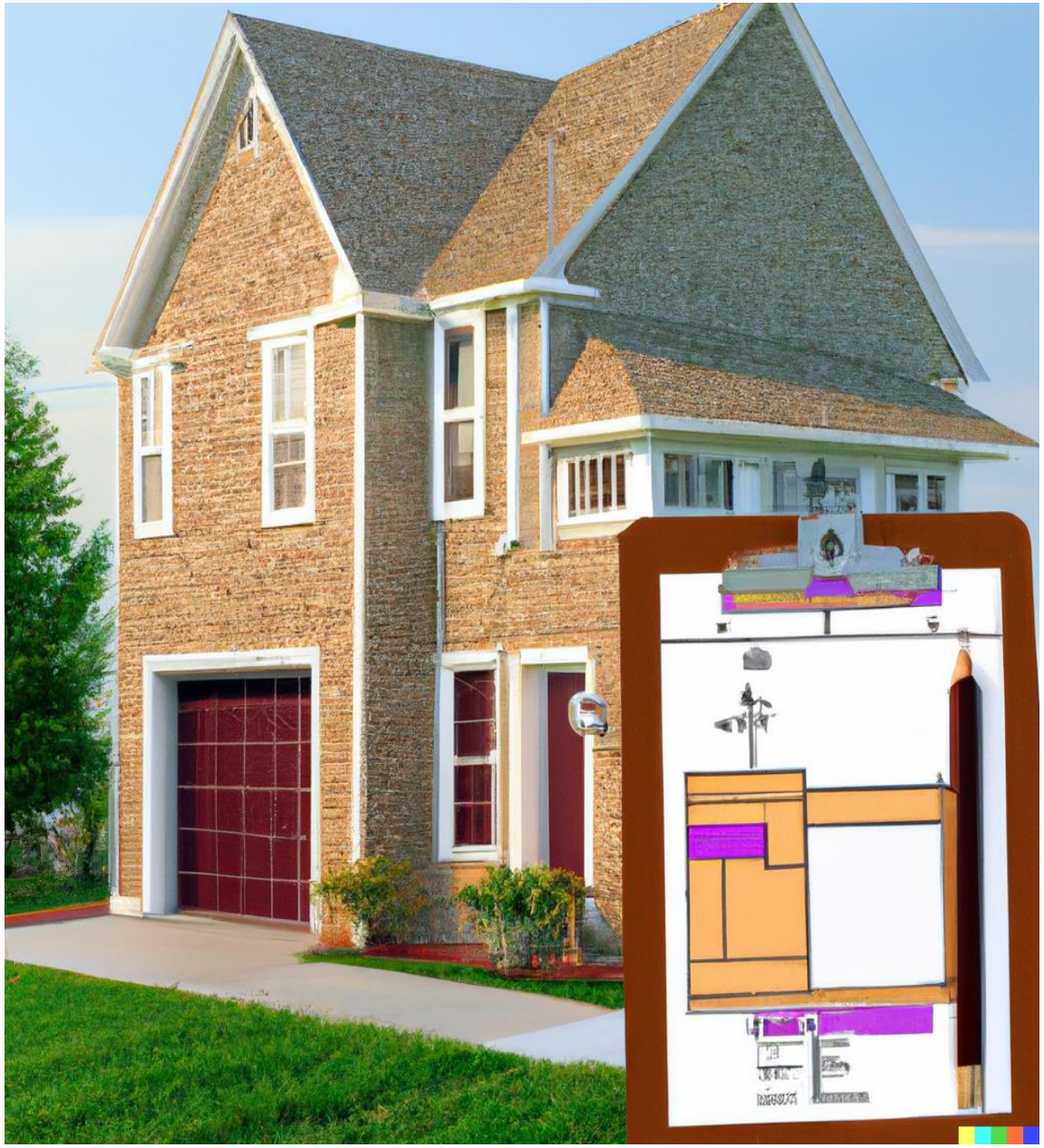
Direct To Home Customer Churn Prediction Study Final Submission

Charles Kilpatrick

September 15, 2023



Executive Summary

Key Takeaways:

- Accounts with a complaint in the last year are 2.91 times more likely to churn than accounts with no complaint.
- Accounts with a tenure of 6 months or less account for 81.28% of churn.
- The complaint frequency for accounts which churned with a tenure of 6 months or less is 54.60% and median customer care contact is 17.
- For accounts which churned within 6 months the proportion of Regular Plus accounts is 66.47%.
- Single status accounts are 2.34 times more likely to churn than married status.
- Females and Tier 2 city residents are underrepresented in the customer base.

Executive Summary (continued)

Key Next Steps:

- Perform root cause analyses of account complaints and implement mitigation.
- Study the account onboarding process to understand why 81.28% of the churn occurs for accounts with tenure less than 6 months.
- Complete a study to investigate relatively high churn for Regular Plus accounts.
- Perform study of why accounts with the primary account holder being of single marital status are 2.34 times more likely to churn than accounts with a primary customer of married status.
- Shifting advertising for increased targeting of females and Tier 2 city residents.

Problem and Solution Summary

Problem That Is Being Solved:

- This study addresses the issue of account churn for a direct to home provider that is facing significant competition and is challenged to retain accounts. Account churn negatively impacts the provider's market position and revenue.

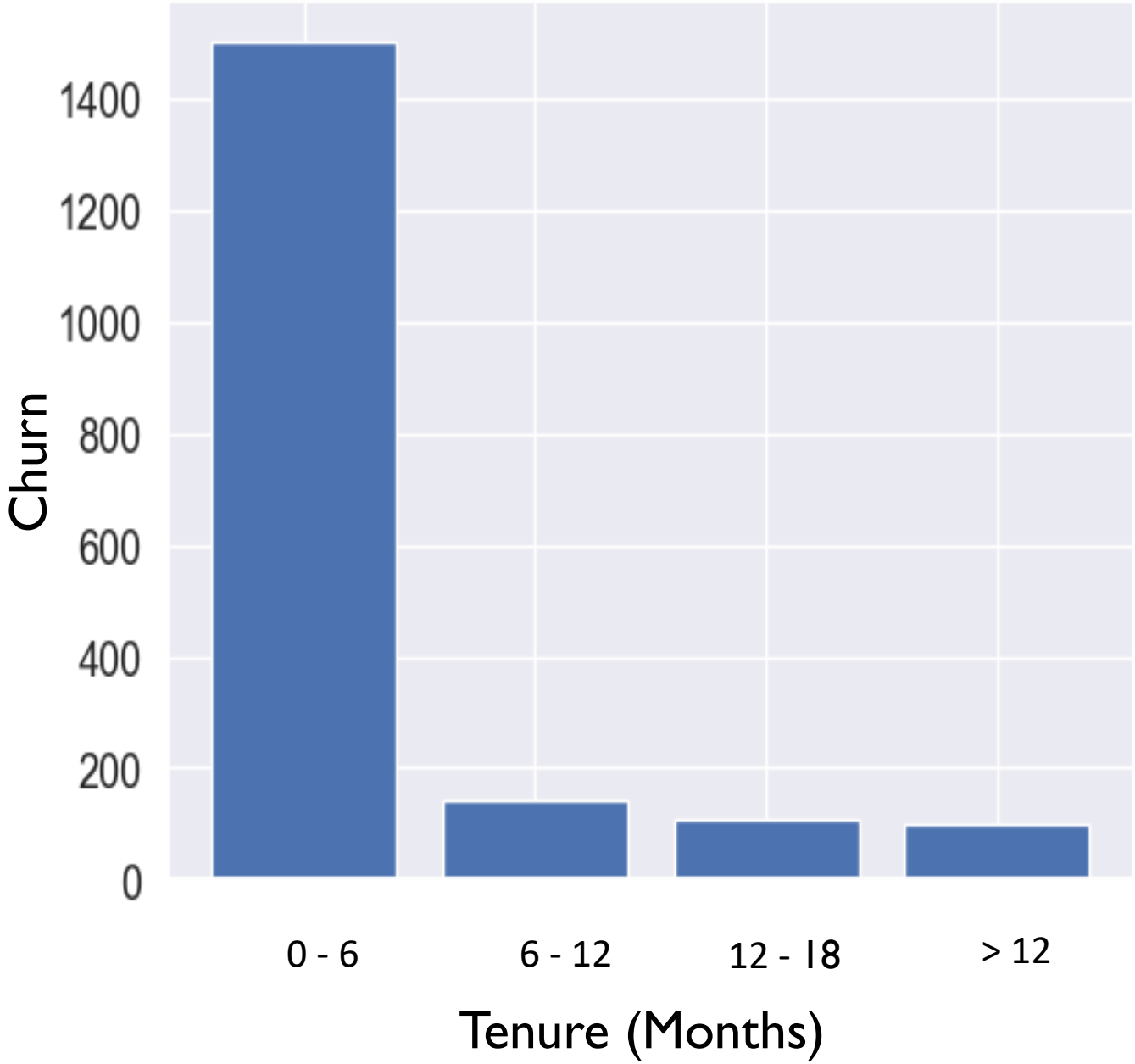
- Most of the account churn occurs within 6 months of accounts being established and the majority of the short tenure accounts have contacted customer service frequently and have registered a complaint.

- Churn for the Regular Plus account segment and for accounts with single marital status for the primary customer are high priority concerns.

Plot Of Churn wrt Tenure

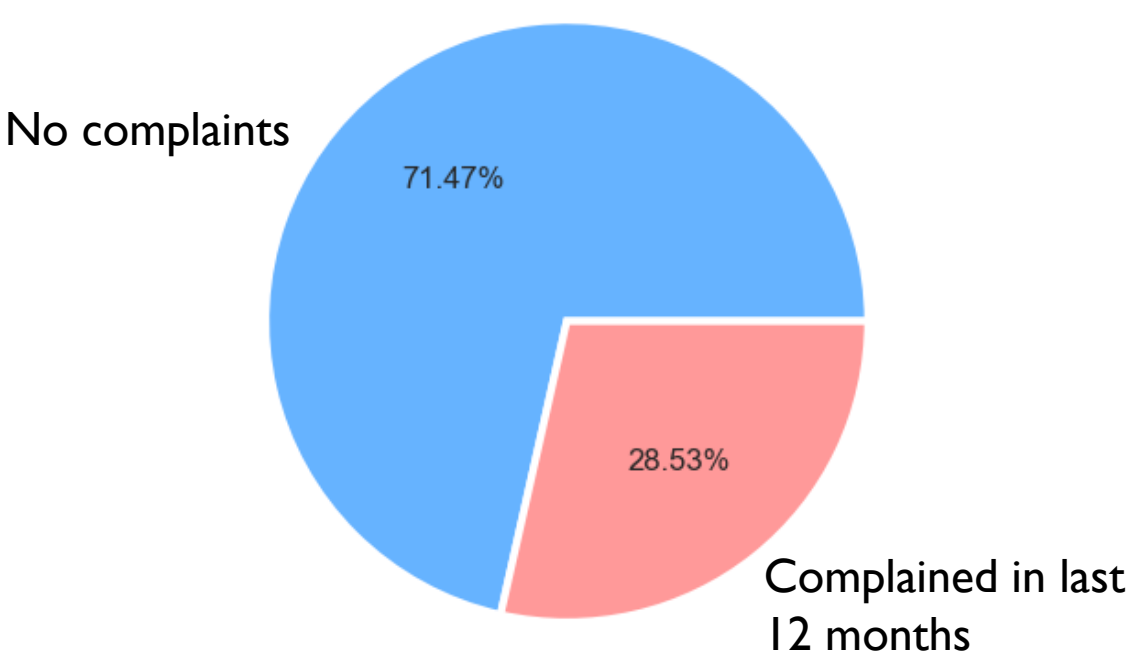
For accounts which churned, 81.28% of the churn occurred prior to completing 6 months of subscription. These notable observations apply only for the short duration (0 – 6 months) accounts:

- A complaint was registered for 54.60% of these accounts.
- 66.47% of the account segment was regular plus.
- 79.63% used a coupon for payment at least once.
- The median customer care contact was 17 times.

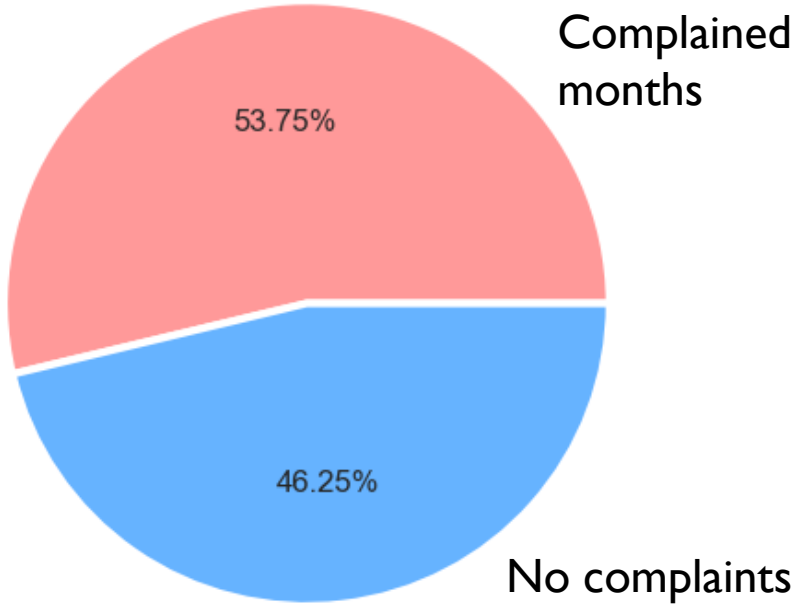


Customer Complaints In Last 12 Months wrt Churn

Overall Dataset



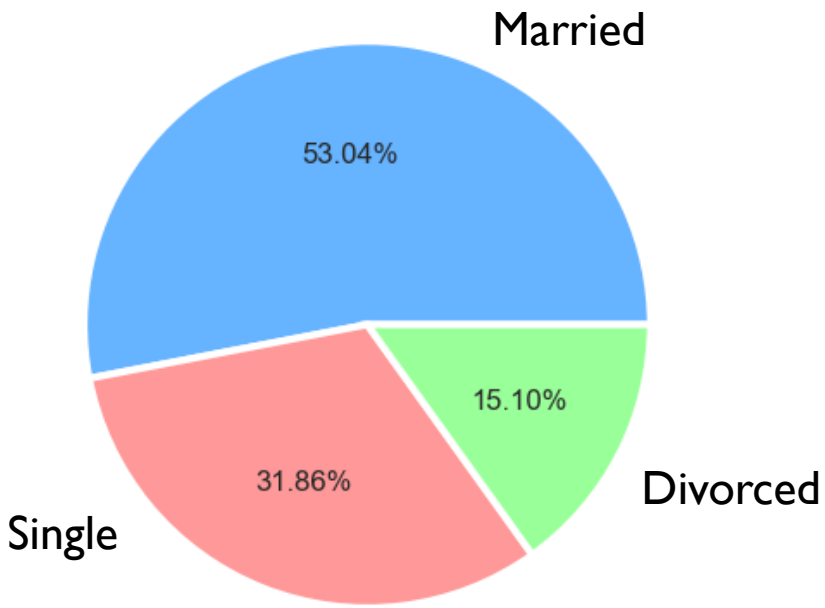
Accounts Which Churned



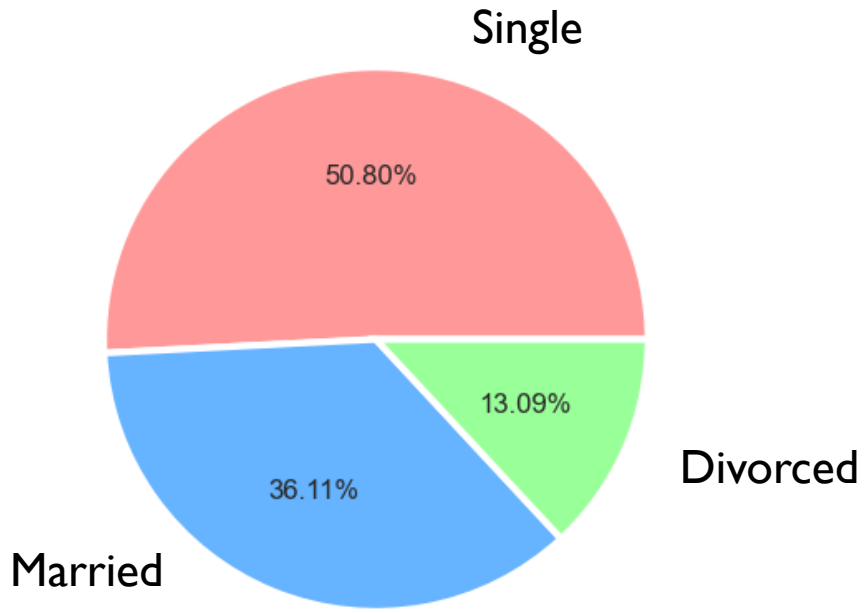
Accounts with a complaint in last 12 months are 2.91 times more likely to churn than accounts with no complaint.

Account Primary Customer Marital Status wrt Churn

Overall Dataset



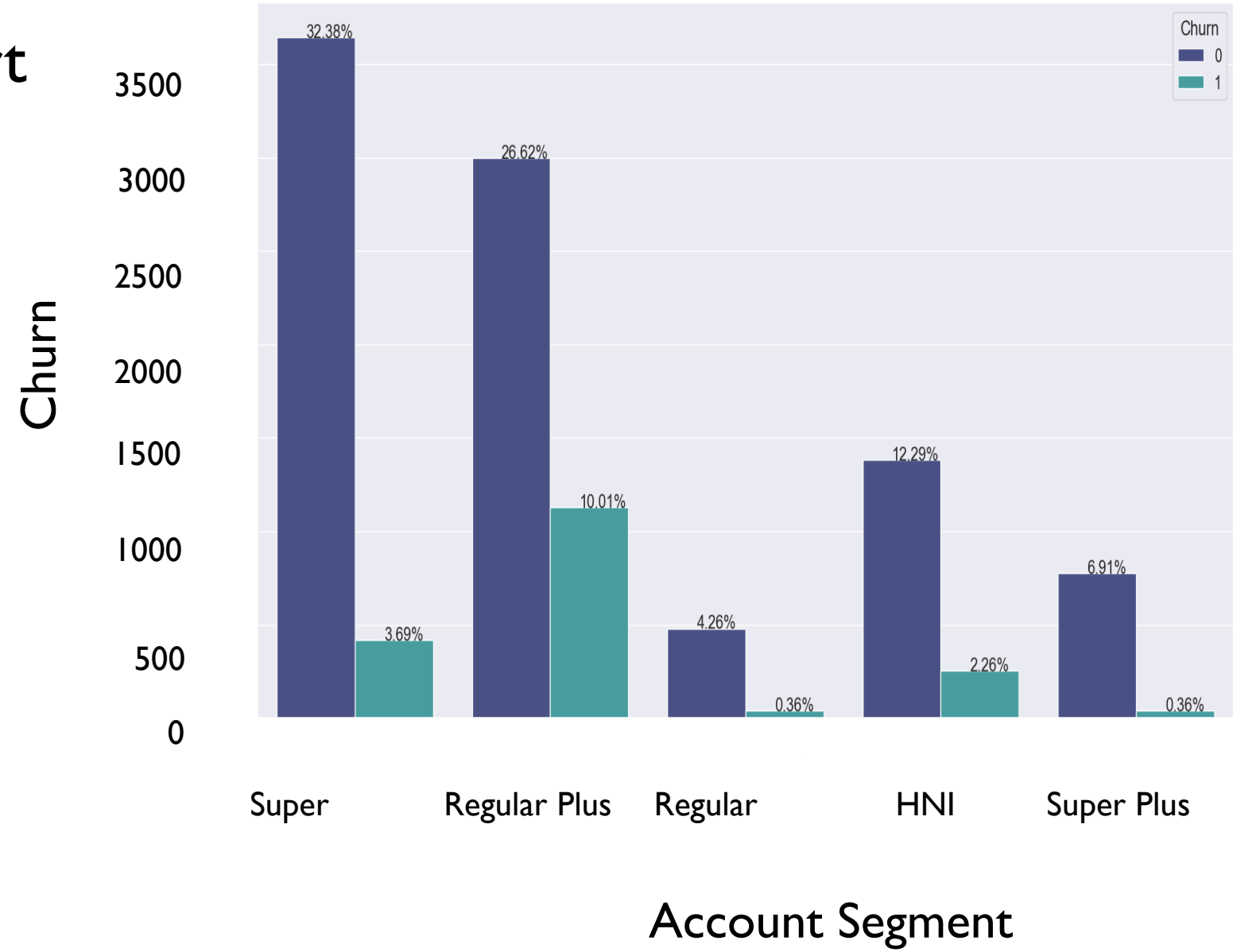
Accounts Which Churned



- Accounts with a single status primary customer are 2.34 times more likely to churn than accounts with the primary account customer with married status.

Account Segment wrt Churn

- Super Plus - 4.95% churned
- Regular - 7.79% churned
- Super - 10.23% churned
- HNI - 15.53% churned
- Regular Plus – 27.33% churned



Problem and Solution Summary

Key Points Describing the Final Proposed Solution Design:

- Improve account onboarding process to increase new customer satisfaction.
- Mitigate factors causing Regular Plus accounts to churn more frequently.
- Modify account terms to achieve better retention of single customers.
- Issue loyalty reward coupons for accounts with tenure greater than 6 months rather than giving payment coupons to new customers.
- Adjust advertising to target females, and Tier 2 cities.
- Consider enhanced cashback for enrollment in automated payment.

Problem and Solution Summary (Continued)

Assessment of the Proposed Solution Viability for Problem Resolution:

- Improving new customer satisfaction will reduce churn.
- Lowering Regular Plus churn to a closer to the mean level will reduce churn.
- Mitigating reasons for greater churn for accounts with single status primary customer is a viable measure to reduce churn.
- Improving customer care agent scores will reduce complaints and churn.
- Shifting advertising to targeting females and Tier 2 cities is easily achievable.
- Implementing increased cashback for automated payments is easily achievable.

Recommendations For Implementation

Key Actionable Insights and Expected Costs/Benefits:

- There is zero net cost for shifting loyalty rewards to accounts with greater tenure.
- The cost of improving the onboarding process and customer care agent scores will reduced future customer care costs.
- If changing Regular Plus terms is too costly, new customers can be steered to the Regular, Super, or Super Plus account segments which have lower churn.
- Consider increased cashback for accounts enrolling in automated payment as automated payments reduce churn.
- Shifting existing advertising funds to target females and Tier 2 cities will have zero net cost impact.

Recommendations For Implementation (Continued)

Key Risks and Challenges:

- New customers may push back on shifting from new account payment coupons to loyalty rewards such as enhanced cashback for accounts with tenure greater than 6 months.

- The Regular Plus account segment accounts for 34.34% of monthly income. Care should be taken to ensure any decrease in Regular Plus subscriptions is fully offset by increases in other account segments.

- Shifting advertising to target females and Tier 2 city customers may reduce other sales.

Recommendations For Implementation (Continued)

Other Problems To Be Explored In Order Of Priority:

- Implement a customer care agent score survey for customers to provide direct feedback on customer care agent service quality.
- Market to females who comprise only 38.89% of the primary customers for accounts.
- Market to Tier 2 city customers who comprise only 4.31% of customers.
- Audit the computer interface for user friendliness. Accounts using computer access are slightly more likely to churn than accounts using mobile access.

Appendix

Modeling Process (Validation and Interpretation)

- Split the dataset into train and test sets
- Impute missing values
- Add a calculated variable for average monthly cashback / revenue
- Build different models - Logistic Regression, Decision Tree, Random Forest, Gradient Boost, and XG Boost using base dataset (imbalanced)
- Calculate confusion matrix and plot feature importances
- Calculate training and testing Accuracy, Recall, Precision, F1-Score
- Compare all model performances on training and test data
- Recall is the key evaluation metric due to the need to minimize churn.
- Build corresponding models using oversampled data (balanced) and evaluate
- Select a final model to be used.

Reasoning For Model Selection

The Tuned XGBoost model with SMOTE oversampled data was selected for these reasons:

- Recall is the key metric as identifying all accounts liable to churn is key and Tuned XGBoost using SMOTE oversampled data was found to be superior:
- Testing Recall - 0.9420
- Testing Precision - 0.4149
- Testing Accuracy - 0.7664
- Testing F1 Score – 0.5760
- Total testing false negative count of 33

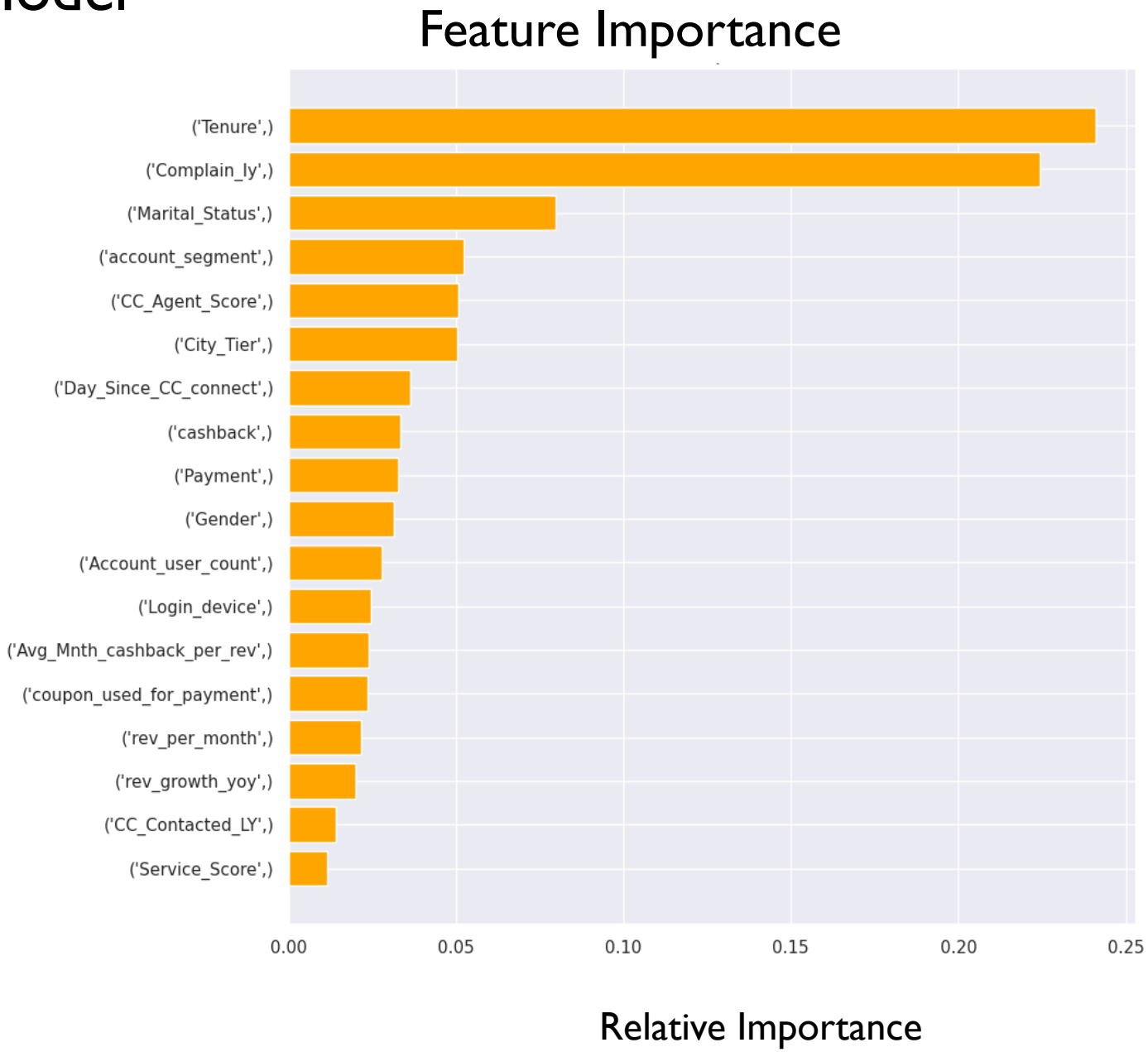
Interpretation From The Best Model

- Recall is the key evaluation metric.

- The superior model is XGBoost with hyperparameter tuning applied to SMOTE oversampled data.

- Testing Accuracy – 0.7664
- Testing Recall – 0.9420
- Testing Precision – 0.4149
- Testing F1 Score – 0.5760

- Tenure, Complain_ly, and Marital_Status are the three most important features.



Model Comparisons (Base Data and Hyperparameter Tuning)

	Decision Tree	Random Forest	Gradient Boosting	Logistic Regression	XG Boosting
Training Accuracy	0.8299	0.9580	0.8682	0.7891	0.9685
Training Recall	0.7784	0.9563	0.8500	0.8184	0.9894
Training Precision	0.4966	0.8230	0.5732	0.4332	0.8487
Training F1-Score	0.6064	0.8846	0.6847	0.5665	0.9137

	Decision Tree	Random Forest	Gradient Boosting	Logistic Regression	XG Boosting
Testing Accuracy	0.8286	0.9301	0.8606	0.7718	0.9405
Testing Recall	0.7592	0.8506	0.8155	0.7979	0.9051
Testing Precision	0.4943	0.7622	0.5590	0.4090	0.7779
Testing F1-Score	0.5988	0.8040	0.6633	0.5408	0.8367

Model Performance (Oversampled Data and Hyperparameter Tuning)

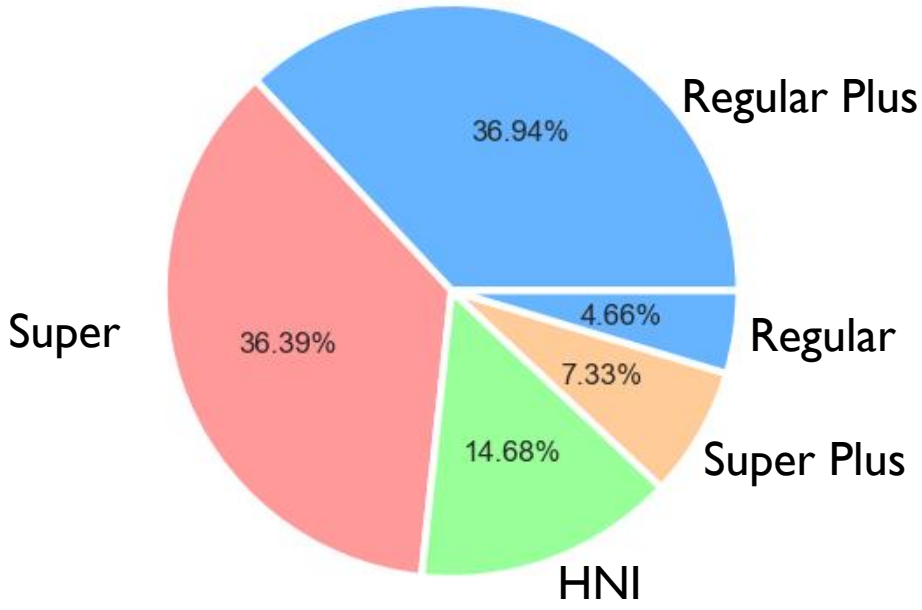
	Decision Tree	Random Forest	Gradient Boosting	Logistic Regression	XG Boosting
Training Accuracy	0.8577	0.9995	0.9177	0.8149	0.8843
Training Recall	0.8683	0.9991	0.9815	0.8433	0.9945
Training Precision	0.8502	0.9998	0.7818	0.8047	0.8027
Training F1-Score	0.8592	0.9995	0.8703	0.8235	0.8884

	Decision Tree	Random Forest	Gradient Boosting	Logistic Regression	XG Boosting
Testing Accuracy	0.8265	0.9615	0.7432	0.7812	0.7664
Testing Recall	0.7452	0.8313	0.9051	0.7909	0.9420
Testing Precision	0.4902	0.9184	0.3866	0.4206	0.4149
Testing F1-Score	0.5914	0.8727	0.5418	0.5491	0.5760

Account Segment Data

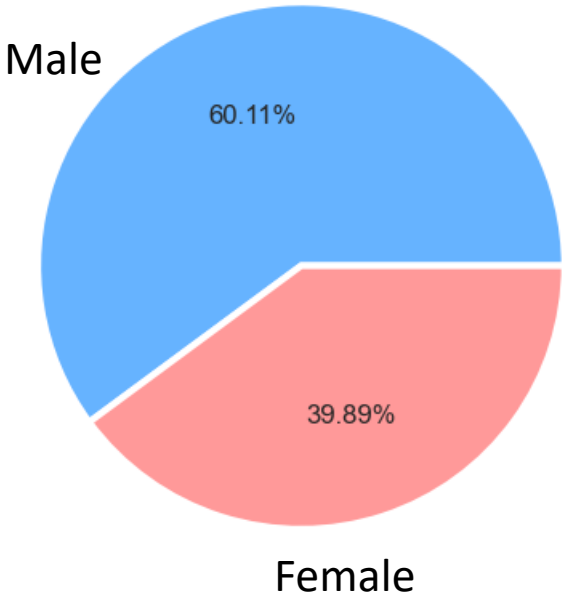
Account Segment Financial Impacts	Regular	Regular Plus	Super	Super Plus	HNI
% Monthly Income Contribution	5.10	34.24	36.66	8.15	15.86
% YOY Revenue Growth Contribution	4.34	37.19	36.28	7.39	14.72

Account Segment Population Distribution

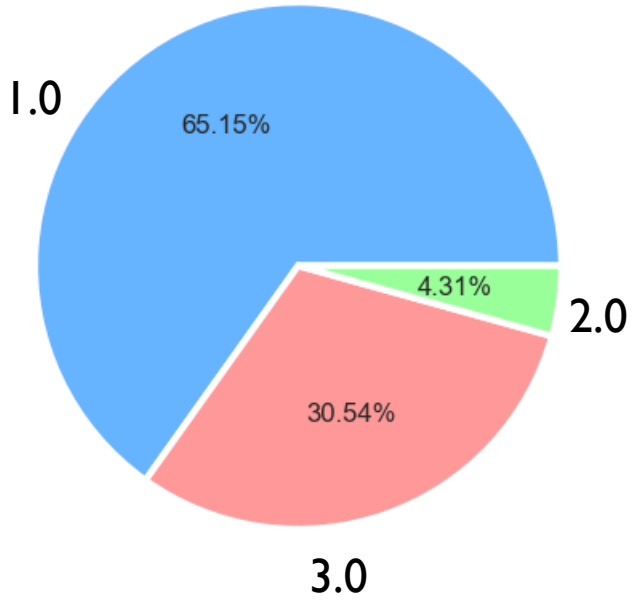


Gender, City Tier, and Login Device Demographics Charts

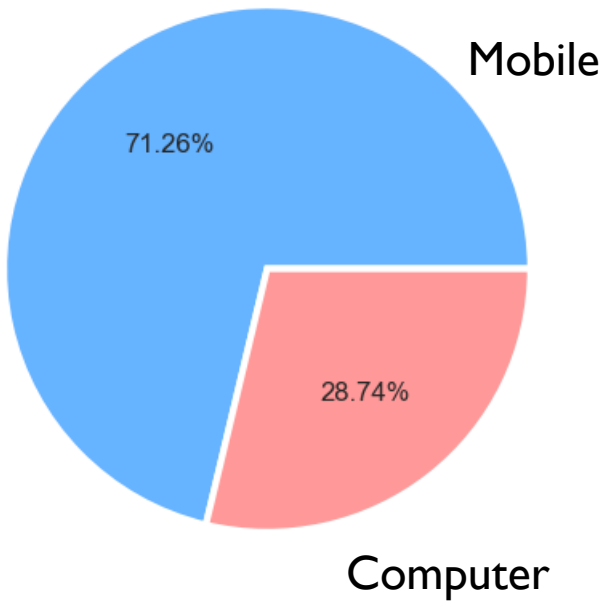
Gender



City Tier

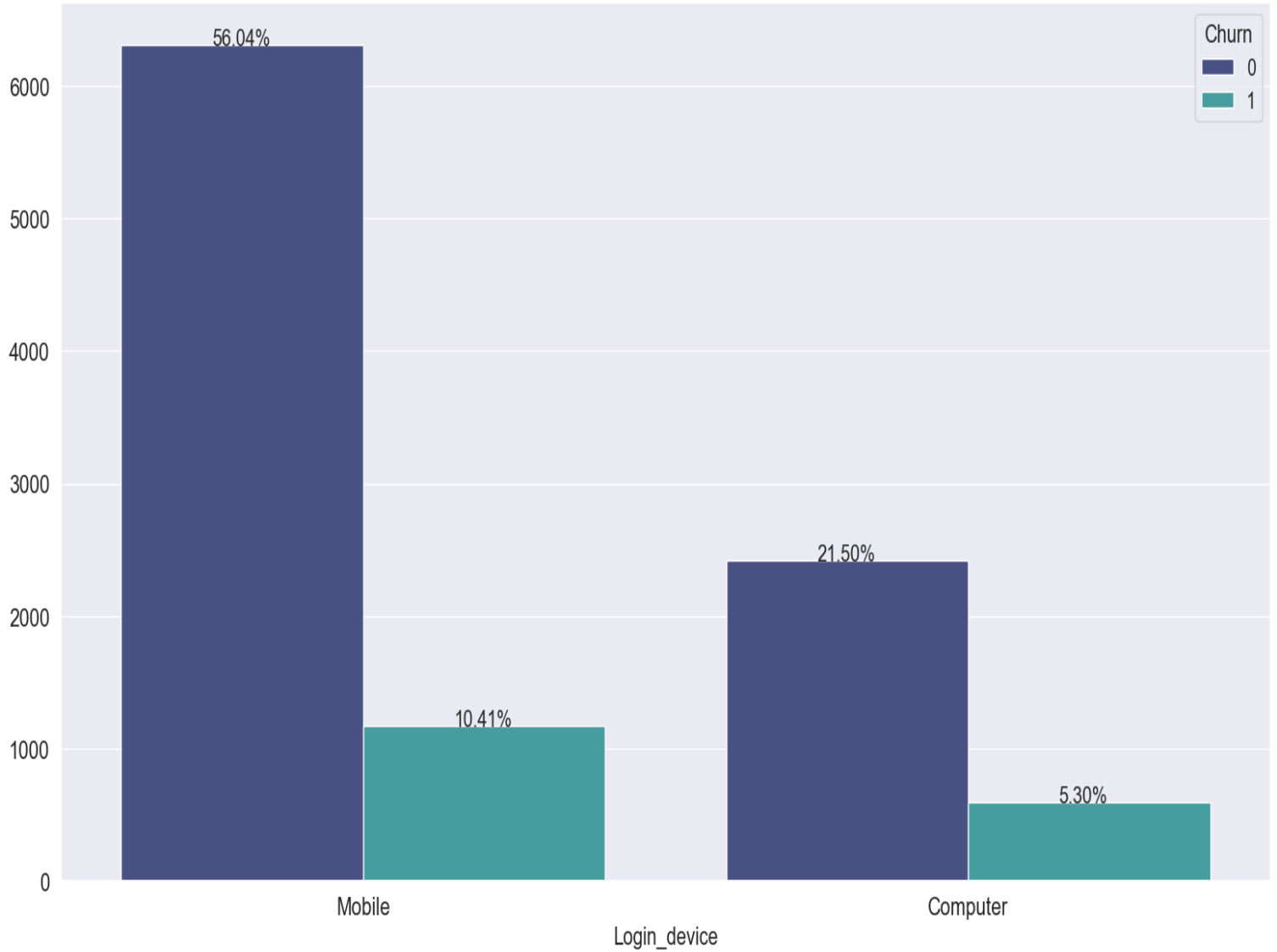


Login Device



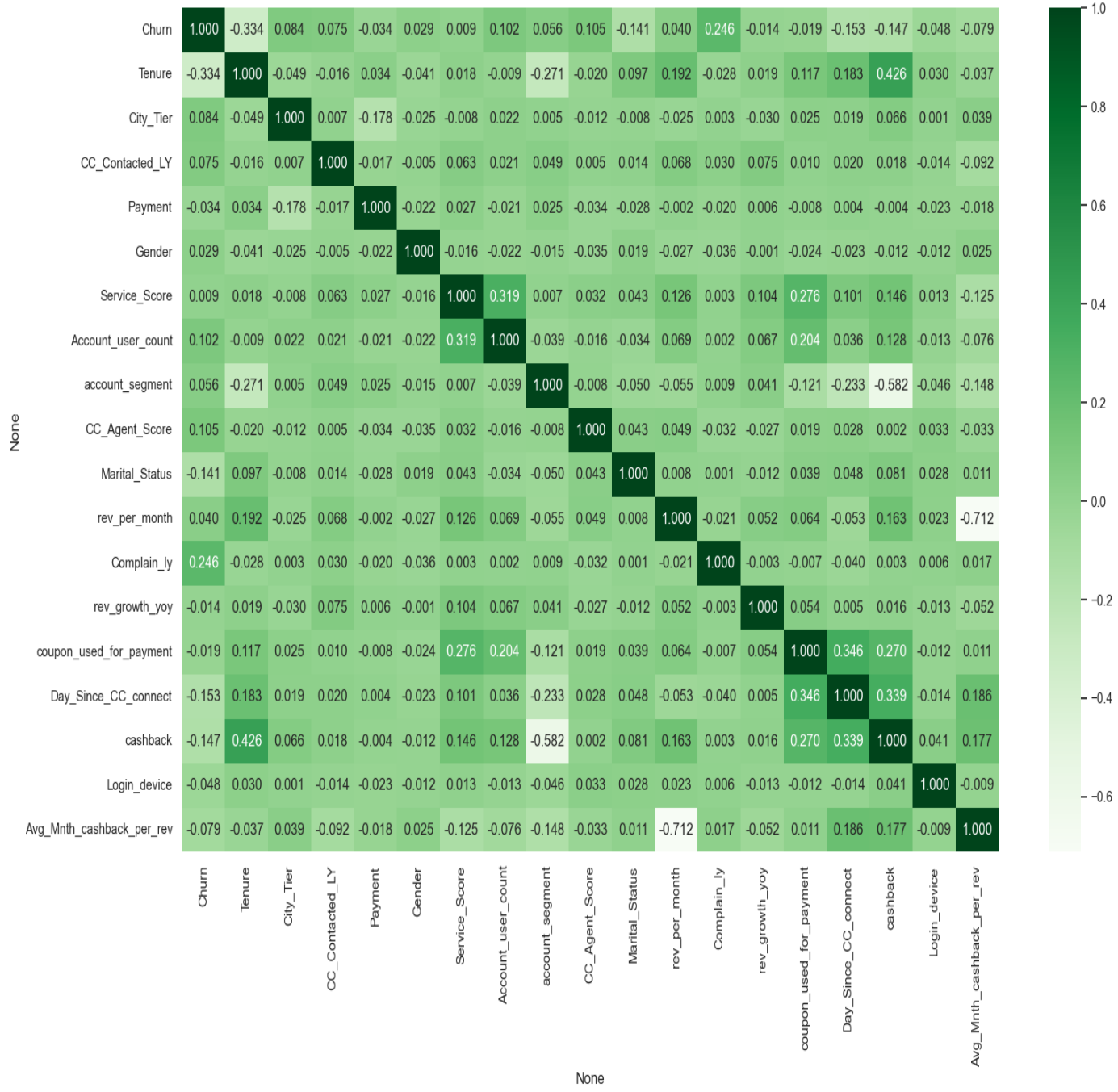
Relationship Between Churn and Login Device

- Mobile Login – 15.67% churned
- Computer Login – 19.78% churned



Correlation Plot

- Churn is negatively correlated with tenure.
- Churn is positively correlated with a complaint in the last year.
- Tenure and cashback are positively correlated.
- There is a strong negative correlation between monthly revenue and cashback amount per monthly revenue amount ratio.



Data Pre-Processing – Missing Variable Treatment

Variable Name	Unique Value Modifications	Total Values Imputed From NaN
Tenure	# changed to NaN	218
City_Tier	Not Applicable	112
CC_Contacted_LY	Not Applicable	102
Payment	Not Applicable	109
Gender	F and M changed to Female and Male	108
Service_Score	Not Applicable	98
Account_user_count	@ changed to NaN	444
Account_segment	+ changed to Plus	97
CC_Agent_Score	Not Applicable	116
Marital_Status	+ changed to NaN	212
rev_per_month	Not Applicable	791
Complain_ly	Not Applicable	357
Rev_growth_yoy	$ changed to NaN	3
Coupon_used_for_payment	#, $, * all changed to NaN	3
Day_Since_CC_connect	$ changed to NaN	358
cashback	Not Applicable	473
Login_device	&&&& changed to NaN	760

Data Pre-Processing – Outlier Treatment

Variable Name	Count Of High Outliers Limited To 75% Quartile + 1.5 * IQR	Count Of Low Outliers Limited To 25% Quartile - 1.5 * IQR
Tenure	139 Values > Right Whisker	Not Applicable
CC_Contacted_LY	42 Values > Right Whisker	Not Applicable
Service_Score	5 Values > Right Whisker	8 Values < Left Whisker
rev_per_month	185 Values > Right Whisker	Not Applicable
Coupon_used_for_payment	1380 Values > Right Whisker	Not Applicable
Day_Since_CC_connect	33 Values > Right Whisker	Not Applicable
cashback	859 Values > Right Whisker	Not Applicable