Customer Base Analysis

MITIGATING CHURN RISK

CHARLES CORBITT

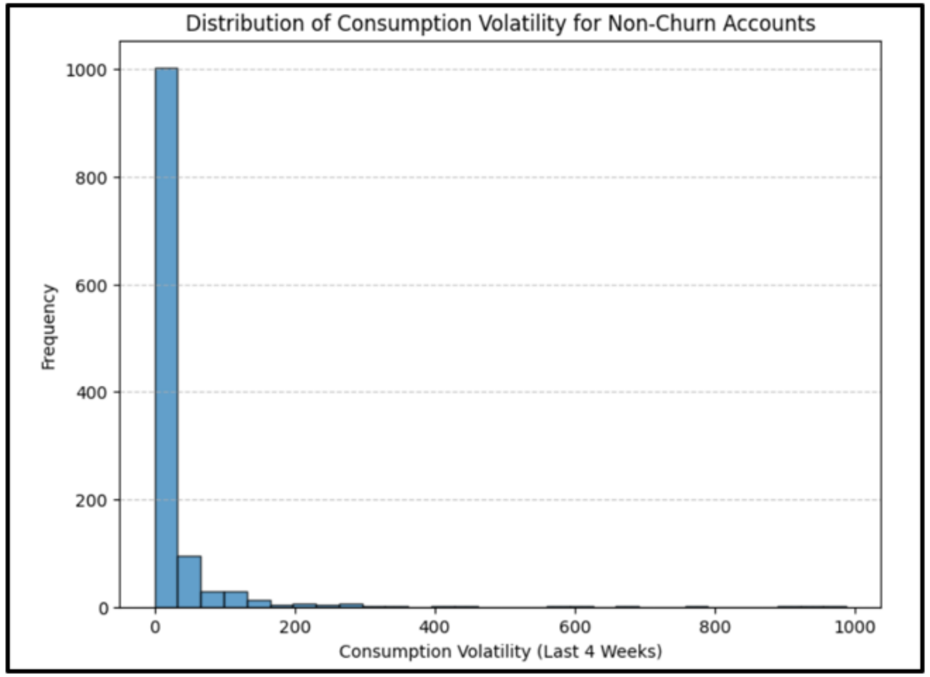
Contents

Background

- [redacted company]'s platform helps companies monitor, troubleshoot, and optimize their software systems in real-time.
- The management team is concerned about varying performance across the customer base.
- Some accounts are expanding, while others are at risk of churning.
- Goal of this analysis is to predict accounts at risk of churning

Key Takeaway and Impact

- Consumption volatility (high variation in week-to-week consumption) is a strong predictor of future churn
- 10% of currently non-churned accounts have very high volatility, making them the most at-risk
- Average weekly revenue for these accounts is $12,268, which is 34% of overall weekly revenue
- Immediate intervention may help prevent a significant loss in weekly revenue



Volatility_Bin	
Low	602
Moderate	301
High	180
Very High	121

Other Findings

- Issues with Data Quality
  - 41% of revenue data is missing (mostly in 2022)
- Logistic Regression
  - Model needs fine tuning; initial version has moderate overall accuracy of 67%
  - Consumption volatility was the top predictor
- Volatility
  - Top 5% most volatile accounts represent 10% of all accounts

Classification Report:

	precision	recall	f1-score	support
0.0	0.65	0.99	0.78	592
1.0	0.69	0.05	0.10	332
accuracy			0.65	924
macro avg	0.67	0.52	0.44	924
weighted avg	0.67	0.65	0.54	924

ROC-AUC Score: 0.632179054054054

Confusion Matrix:

```
[[584   8]
 [314  18]]
```

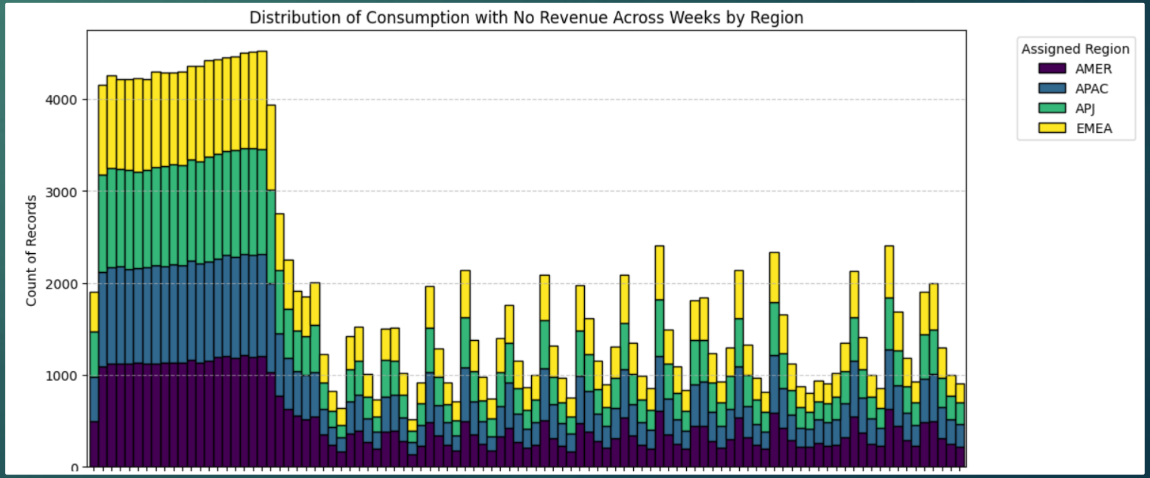
	feature	coefficient
3	consumption_volatility	0.766046
2	consumption_max	-0.751166
4	log_consumption_slope	-0.411443
5	log_consumption_volatility	-0.177336
8	ACCOUNT_TIER_Enterprise	-0.169382
1	consumption_mean	0.125703
0	consumption_sum	0.125703
6	TOTAL_USERS	0.116253
7	ACCOUNT_TIER_Basic	0.107571
9	ACCOUNT_TIER_Pro	-0.084568

Methodology

- Exploratory Data Analysis (EDA)
  - Deduped both input files
  - Visualized distributions of all fields
  - Removed records with negative revenue or consumption

Methodology

- Data Preprocessing
  - Merged files together on account_id and date
  - Removed first 2 weeks of data that were missing consumption data
  - Flagged 2022 as missing large portion of revenue
  - Rolled data up to weekly level to help eliminate noise



Methodology

- Defining Churn Risk
  - Required 6-week baseline for accounts to establish themselves
  - Created rolling 4-week window to measure churn
  - Defined churn as 4 consecutive weeks with no consumption

Methodology

- Prepping Data for Modeling
  - Log-transformed consumption and calculated baseline mean/sum/max
  - Calculated slope of week-to-week consumption to model gradual increases/decreases
  - Created consumption volatility (standard deviation week-to-week)
  - Encoded categorical variables
  - Split data 50/50 into train/test and scaled the predictor set

Methodology

- Modeling, Validation and Interpretation
  - Ran logistic regression and output scoring criteria
  - Calculated coefficients of predictor set to view impact of individual features in the model
  - Used consumption volatility to flag non-churned accounts at-risk of churning in future
  - Identified top 5% of accounts in terms of volatility for potential intervention

Recommendations and Next Steps

- Primary Recommendation
  - Intervention of high-risk accounts – offer discounted services, or additional product lines for free to try to keep accounts from churning
- Missing data
  - A significant portion of accounts with consumption are missing revenue (41%), remedying this would allow use of revenue features in data model
- Industry Field
  - Many unique values but frequency distribution heavily weighted to few categories
  - Consider bucketing some of the lower frequency industries together
  - Reduce frequency of ‘Other’ if possible since it’s 2nd category; it may be cannibalizing the other categories if it’s placed too high in a dropdown

Recommendations and Next Steps

- Further refine the model
  - Product line and industry were not included in this initial model, but may help predict churn if used for segmentation
  - Add more calculated features derived from revenue, such as gradual or sharp declines, similar to the consumption features
  - Experiment with different definitions for churn, varying the baseline and/or churn periods per industry benchmarks
  - Try more advanced ensemble models such as XGBoost that may better account for complex interactions

Thank You!