Churn propensity modeling

Nutritional supplement Company

Developed and deployed a churn propensity model to identify at-risk customers at various stages of customer life cycle on a weekly basis through an automated report to enable retention efforts on an ongoing basis

Nutrition supplements company needs “food for thought” for customer retention

Picture this…

You’re looking to reduce the high customer churn with more than 60% of the customers discontinuing subscriptions within 3 months of acquisition, by using tailored marketing efforts.

You turn to Accordion.

We partner with your team to deploy the predictive model to identify customers with highest probability of churn on a weekly basis through an automated report to enable retention efforts on an ongoing basis.
We partner with you team to -

1) Developing a customized advanced analytics model to identify customers at high risk of churn at 2nd order, 3rd order, and 4th order
2) Leveraging transactional behavior along with demographic characteristics of customers to further improve the accuracy and effectiveness of the churn propensity model
3) Automating the model output along with its reporting layer to adapt based on new patterns emerging from customer behavior. This would drive retention efforts such as Outbound calling, email campaigns etc. on an ongoing basis based on latest customer data

Your value is enhanced.

You have predicted high-risk customers after 2nd order and 3rd order with more than 90 percent accuracy. You have improved 2nd order and 3rd order retention rate by 8 pps and 5 pps respectively, which could lead to a potential revenue uplift of ~$2.5 million through an ~7 pps uplift in reorders

CHURN PROPENSITY MODELING

KEY RESULT

- >90% accuracy in prediction
- ~$2.5 Mn increase in potential revenue

VALUE LEVERS PULLED

- Churn propensity model

Methodology/ Approach

01 **Data Acquisition** – Transformed the transactional level dataset at customer level for each order (1,2,3,3+). Utilized additional factors in the analysis that could potentially impact the churn rates like **Questionnaire responses (during acquisition), order delay, discount etc., controlled by the company and external factors like covid cases.**

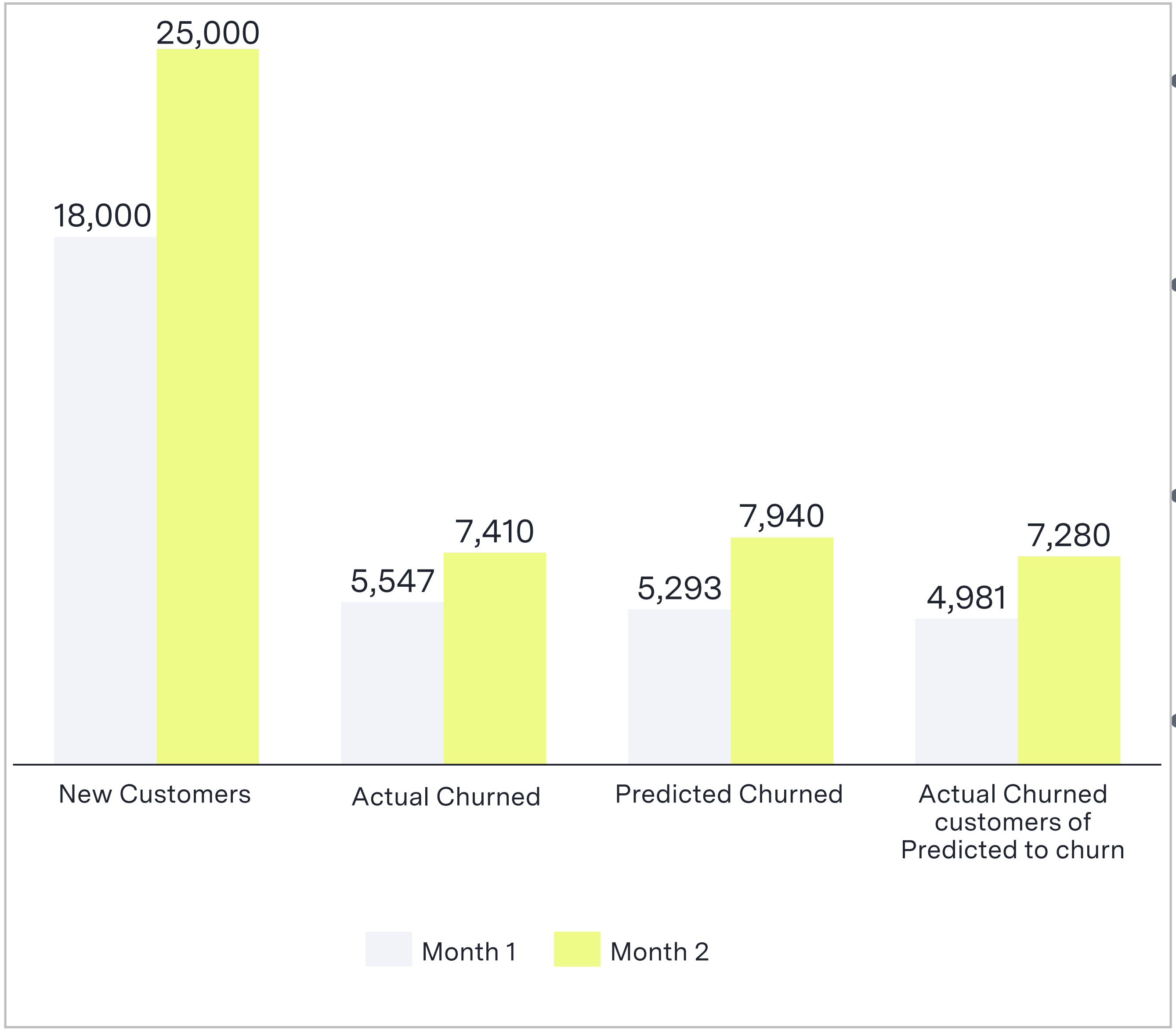
02 **Feature engineering** (Shortlisting of appropriate variables) – **Correlation analysis for the above variables** with customer churn to ensure appropriate variables are selected and create combined variables to ensure appropriate features for modelling

03 **Model selection** – Select the best-performing model based on iterative modelling across different techniques such as **Random forests, XGBoost, Cat-boost, Decision tree etc.** Identify the key features that drive customer churn. Create separate models for 2nd order churn, 3rd order churn and 3+ order churn.

04 **Model iteration** – Iterate the model to further fine tune to enhance overall accuracy (balance between precision and recall metrics)

05 **Model deployment** – The best performing model was **deployed in Azure Databricks notebook** and was scheduled to run on a **weekly cadence using Azure Data Factory**. The Databricks notebook interacts with the Azure database to extract relevant data and stores the results which is leveraged by the future iterations of the model as a learning. Additionally, an **automated reporting layer** was built to provide the list of customers having high risk of churn on a weekly basis which was leveraged by the Marketing team on retention initiatives.

Model evaluation



Accuracy = (True Positive + True Negative)/Total number of examples)
Month 1 – 95% Month 2 – 97%

Precision = (True Positive / (True Positive + False Positive))
Month 1 – 94% Month 2 – 92%

Recall = (True Positive / (True Positive + False Negative)
Month 1 – 90% Month 2 – 98%

F1 Score = (2 *(Precision * Recall))/ (Precision + Recall)
Month 1 – 0.92 Month 2 – 0.95

Model evaluation

1 Month-1 Prediction Confusion Matrix

Predicted Class \ Actual Class	Churned	Retained
Churned	4,981 (True Positive)	312 (False Positive)
Retained	566 (False Negative)	12,141 (True Negative)

Total Customers in month-1: 18,000

2 Month-2 Prediction Confusion Matrix

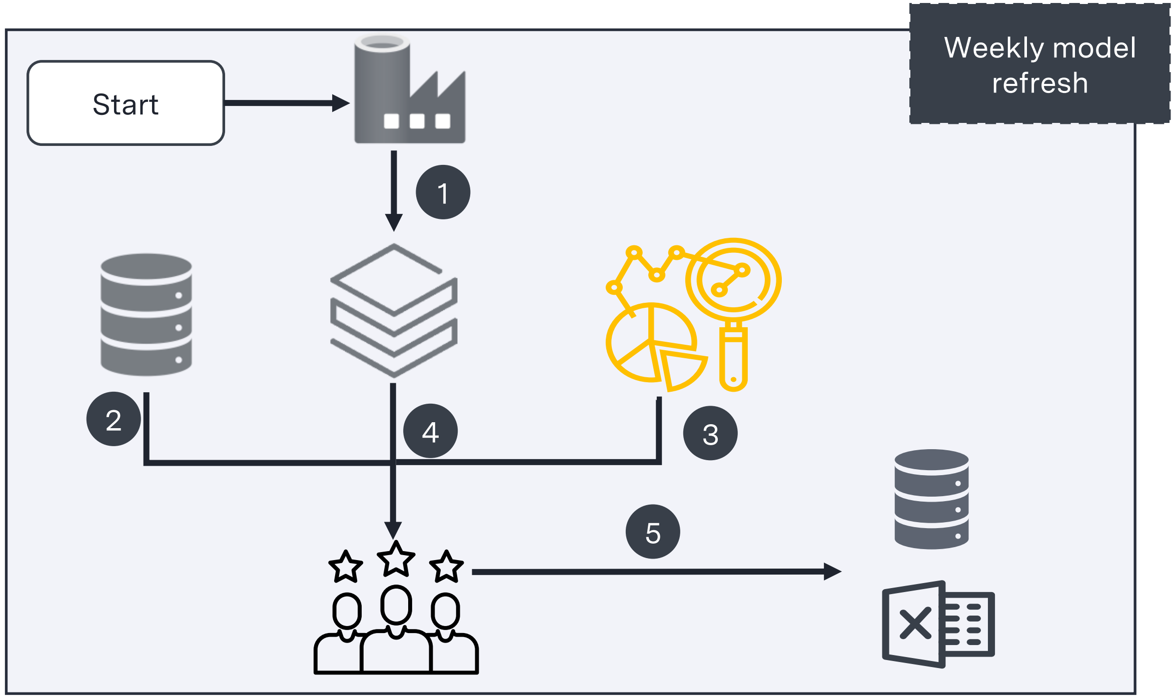
Predicted Class \ Actual Class	Churned	Retained
Churned	7,280 (True Positive)	660 (False Positive)
Retained	130 (False Negative)	16,930 (True Negative)

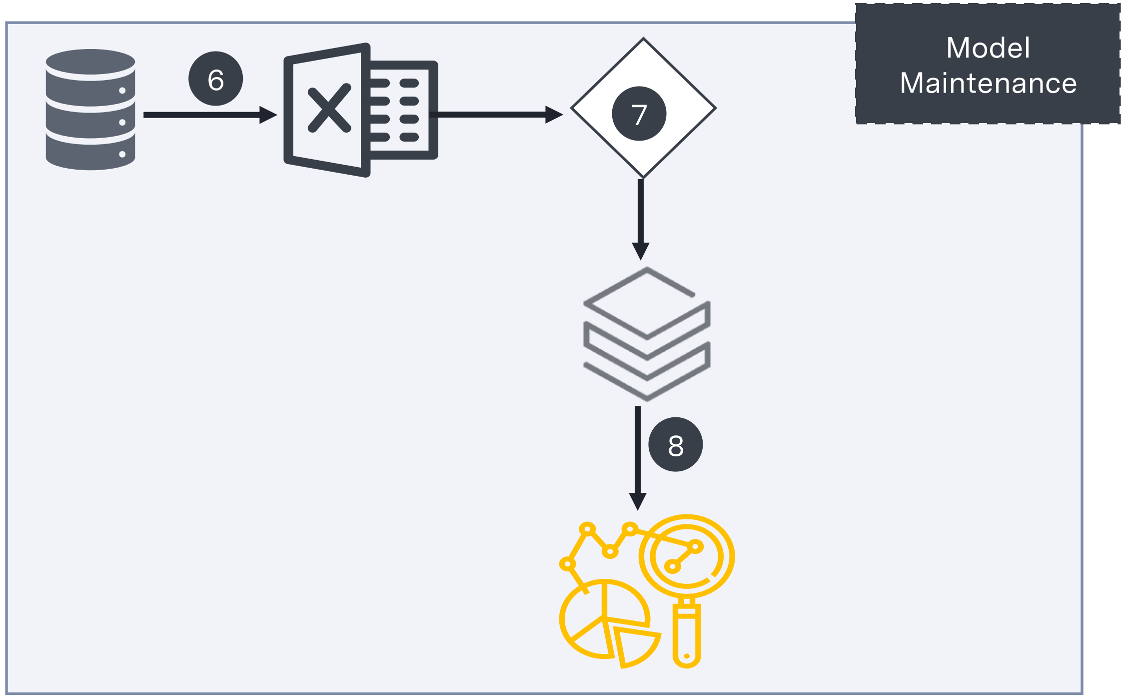
Total Customers in month-2: 25,000

Model Evaluation Metrics

Metric	Formulae	Month-1	Month-2
Accuracy	(True Positive + True Negative)/Total number of examples)	95%	97%
Precision	(True Positive / (True Positive + False Positive))	94%	92%
Recall	(True Positive / (True Positive + False Negative)	90%	98%
F1-score	(2 *(Precision * Recall))/ (Precision + Recall)	0.92	0.95

Deployment architecture in azure ecosystem





1. Weekly Automated trigger using Azure data factory
2. Extract demographics and transactional data till 3rd order for active customers with less than 3 orders
3. Read trained models stored in Databricks File System
4. Predict the probability to churn for these customers before order 2,3 and 4 respectively
5. Store the final output in database for future reference and generate an excel based report
6. Generate model performance report for customers with actuals monthly
7. Retrain the model if significant decrease in accuracy is observed
8. Store the re-trained model in Data bricks file system