Churn Prediction Project Report

Overview

This report provides a comprehensive overview of the churn prediction project undertaken for a B2B SaaS company. The project aims to identify customers at risk of churning and provide actionable insights to reduce churn rates, thereby improving customer retention and revenue growth.

Objective

The primary objective of this project is to develop a predictive model that identifies customers likely to churn. This will enable the company to take proactive measures to retain these customers and minimize revenue loss.

Scope

The scope of this project includes data collection, feature engineering, model development, and visualization of insights. The project covers the following key areas:
- Data creation and preprocessing
- Feature engineering using SQL in Snowflake
- Model training and evaluation in Jupyter Notebook
- Visualization and reporting in Power BI

Data Sources

The data for this project was generated using Python in Jupyter Notebook. It includes the following:
- Customer account details (e.g., CUSTOMER_ID, SEGMENT, INDUSTRY, COUNTRY, PLAN)
- Subscription metrics (e.g., MRR, SUBSCRIPTION_DURATION)
- Product usage indicators (e.g., AVG_MONTHLY_LOGINS, AVG_TICKETS_RAISED, AVG_FEATURE_USAGE)
- Churn labels and customer status

Step-by-Step Breakdown

Data Creation in Jupyter

The dataset was created using Python libraries such as pandas, numpy, and random. The data includes realistic customer details, subscription metrics, and usage indicators. The generated data was saved as CSV files for further processing.

SQL Querying in Snowflake

The data was uploaded to Snowflake for feature engineering. SQL queries were used to calculate key metrics such as:
- Subscription duration
- Average monthly logins
- Revenue at risk

The processed data was exported for modeling.

Modeling in Jupyter

The processed data was used to train and evaluate multiple machine learning models, including Logistic Regression, Random Forest, and XGBoost. The Random Forest model was determined to be the best-performing model based on accuracy, precision, recall, and F1-score.

Top 10 Predictor Variables

The top 10 predictor variables identified by the Random Forest model are:
1. AVG_FEATURE_USAGE
2. MRR (Monthly Recurring Revenue)
3. AVG_TICKETS_RAISED
4. AVG_MONTHLY_LOGINS
5. SEGMENT_SMB
6. SEGMENT_Mid-Market
7. COUNTRY_UK
8. COUNTRY_USA
9. INDUSTRY_Financial Services/Fintech
10. INDUSTRY_Healthcare

These variables provide insights into customer behavior and churn risk.

Measures to Address Churn

Based on the analysis, the following measures are recommended to reduce churn:
- Increase feature usage by providing training and support
- Address customer issues promptly to reduce ticket volume
- Focus on retaining high-MRR customers
- Tailor strategies for SMB and Mid-Market segments

Database Saving and Power BI Import

The final dataset, including predictions and key metrics, was saved as a CSV file and uploaded to Snowflake. The data was then imported into Power BI for visualization and reporting.

Visualizations Created

1. Churn Distribution

- **Analysis**:
  - The **donut chart** reveals that **3.28% of customers are predicted to churn**, while **96.72% are likely to stay**. Although the churn percentage appears small, it represents a significant revenue risk, especially for high-value customers.
  - This highlights the importance of focusing on the small subset of customers at risk to prevent revenue loss.
- **Recommendations**:

  - **Proactive Retention Campaigns**: Launch targeted campaigns for the 3.28% of customers predicted to churn. Use personalized offers, discounts, or enhanced support to retain them.
  - **Customer Segmentation**: Further segment the churned customers by industry, geography, and plan to identify patterns and tailor interventions.
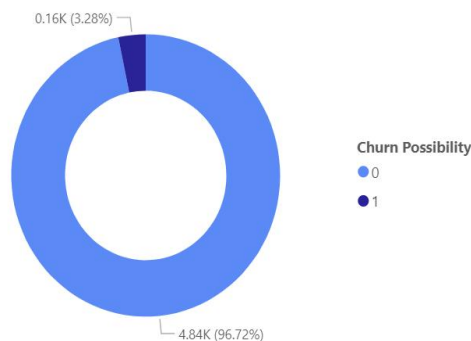 - **Monitor Trends**: Regularly track churn rates to identify any upward trends and act swiftly.

2. Churn by Plan

- **Analysis**:

  - The **bar chart** shows that the **Growth plan** has the highest number of churned customers, followed by the **Pro** and **Enterprise** plans.
  - This suggests that customers on the Growth plan may not be finding sufficient value or may be outgrowing the plan.

- **Recommendations**:

  - **Plan Optimization**: Evaluate the Growth plan's features and pricing. Consider adding value through additional features or better support.
  - **Upsell Opportunities**: Identify customers on the Growth plan who are at risk of churn and offer them tailored upsell opportunities to the Pro or Enterprise plans.
  - **Customer Feedback**: Conduct surveys or interviews with Growth plan customers to understand their pain points and address them.
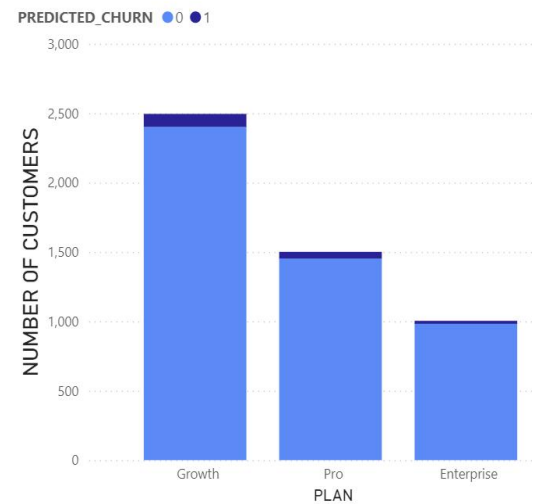
3. Churn by Segment

- **Analysis**:

  - The **SMB (Small and Medium Businesses)** segment has the highest churn rate, followed by **Mid-Market** and **Enterprise**.
  - SMBs are often more price-sensitive and may churn due to cost or lack of perceived value.

- **Recommendations**:

  - **Cost-Effective Solutions**: Offer SMBs cost-effective solutions or flexible payment plans to reduce churn.
  - **Value Demonstration**: Provide SMBs with case studies, ROI calculators, or success stories to demonstrate the value of the product.
  - **Dedicated Support**: Assign account managers or provide enhanced support to SMBs to build stronger relationships.
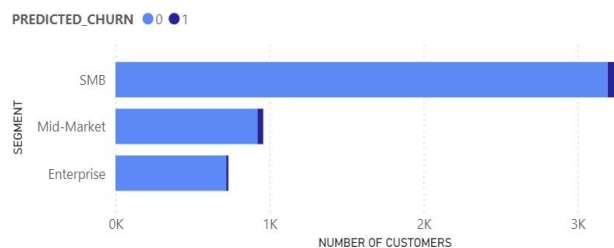
CHURN DISTRIBUTION

0.16K (3.28%)

Churn Possibility
- 0
- 1

4.84K (96.72%)

CHURN BY PLAN

PREDICTED_CHURN ● 0 ● 1

NUMBER OF CUSTOMERS

3,000
2,500
2,000
1,500
1,000
500
0

Growth Pro Enterprise

PLAN

CHURN BY SEGMENT

PREDICTED_CHURN ● 0 ● 1

SEGMENT

SMB
Mid-Market
Enterprise

0K 1K 2K 3K

NUMBER OF CUSTOMERS

4. MRR at Risk by Industry

- **Analysis**:

  - o The **Retail** and **Professional Services** industries have the highest MRR at risk, followed by **Logistics** and **Real Estate**.
  - o This indicates that these industries may face unique challenges or may not be fully utilizing the product.

- **Recommendations**:

  - o **Industry-Specific Solutions**: Develop tailored solutions or features for high-risk industries like Retail and Professional Services.
  - o **Customer Success Programs**: Assign industry-specific customer success managers to help these customers maximize the value of the product.
  - o **Partnerships**: Collaborate with industry associations or thought leaders to build trust and credibility.

5. Total MRR at Risk by Geographies

- **Analysis**:

  - o The **Treemap** highlights that the **USA** has the highest MRR at risk, followed by the **UK**, **Brazil**, and **India**.
  - o This suggests that these geographies should be prioritized for churn prevention efforts.

- **Recommendations**:

  - o **Localized Strategies**: Develop region-specific retention strategies, considering cultural and economic factors.
  - o **Regional Teams**: Strengthen regional customer success teams to provide localized support and build stronger relationships.
  - o **Market Research**: Conduct market research in high-risk geographies to understand customer needs and address them effectively.

6. MRR at Risk Globally

- **Analysis**:

  - The **global map visualization** shows that North America and Europe have the highest concentration of MRR at risk.
  - This aligns with the Treemap insights and emphasizes the need for a global yet localized approach.

- **Recommendations**:

  - **Global Retention Task Force**: Establish a task force to focus on high-risk regions and coordinate efforts across teams.
  - **Data-Driven Insights**: Use predictive analytics to identify early warning signs of churn in these regions and act proactively.
  - **Customer Advocacy Programs**: Build customer advocacy programs in these regions to strengthen loyalty and reduce churn.
  -

MRR AT RISK BY INDUSTRY

MRR AT RISK GLOBALLY

TOTAL MRR AT RISK BY GEOGRAPHIES
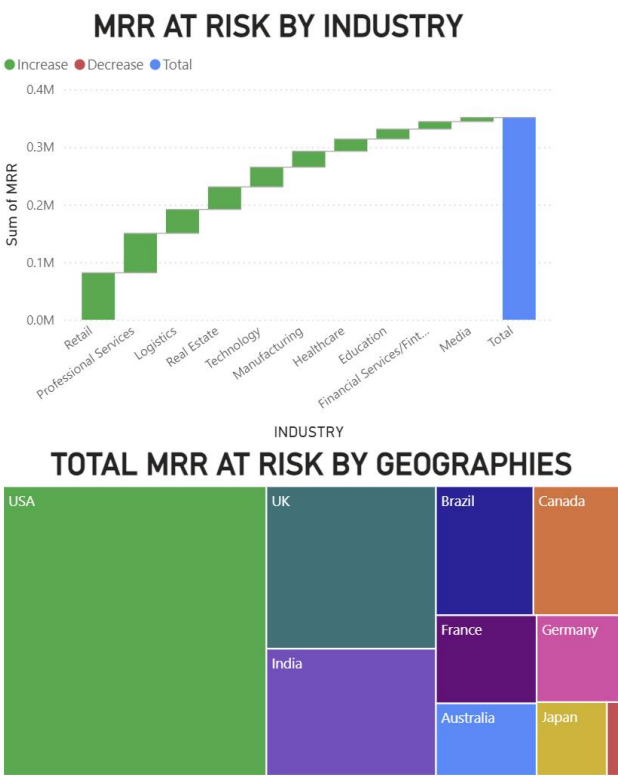
Plan of Action

1. **Immediate Interventions**:

   o Launch targeted retention campaigns for high-risk customers in the Growth plan, SMB segment, and Retail industry.
   o Assign dedicated account managers to high-risk geographies like the USA and UK.

2. **Mid-Term Strategies**:

   o Develop industry-specific solutions and features for Retail, Professional Services, and Logistics.
   o Conduct customer feedback sessions to identify pain points and address them.

3. **Long-Term Goals**:

   o Invest in predictive analytics to continuously monitor churn risk and act proactively.
   o Build a robust customer success framework with regional teams and industry-specific expertise.

4. **Monitoring and Reporting**:

   o Regularly update the Power BI dashboard to track progress and share insights with stakeholders.
   o Set up automated alerts for significant changes in churn rates or MRR at risk.