**📄 Report: Understanding Variables Influencing Churn & Creating Churn Risk Scores**

**1. Understanding the Variables that Influence Churn**

The primary objective of this step is to uncover which customer behaviors or attributes most significantly contribute to customer churn in No-Churn Telecom.

📊 **Approach Used**:

* **Exploratory Data Analysis (EDA)** was conducted on the original telecom dataset, identifying trends and patterns in features such as:
  + Total usage (minutes and charges across day, evening, night, and international)
  + Number of customer service calls
  + Subscription to international or voicemail plans
* **Correlation Analysis** and **Chi-Square tests** were applied to quantify relationships between variables and the churn outcome.
* **Model-based Feature Importance** was extracted using machine learning models (like Random Forest & XGBoost) to determine which features best predict churn.

📌 **Top Predictive Features**:

| **Feature** | **Description** |
| --- | --- |
| Customer Service Calls | High number of calls signals dissatisfaction |
| International Plan | Correlated with higher churn |
| Day Charges | Higher day usage may indicate heavy users |
| VMail Message | Lack of voicemail use linked to churn |
| International Calls | High international use also flags churn risk |

These insights help target at-risk users and personalize customer experience.

**2. Creating Churn Risk Scores for Retention Campaigns**

The second goal was to create a **Churn Risk Score** to quantify how likely a customer is to leave the service provider.

🧠 **Methodology**:

* Used supervised machine learning models trained on historical churn data.
* Algorithms used included:
  + Logistic Regression
  + Decision Tree
  + Random Forest
  + Gradient Boosting
  + XGBoost
* Model was trained on features identified as important in Step 1.

📈 **Risk Score Generation**:

* model.predict\_proba() was used to output a probability value between **0 and 1** representing churn likelihood.
* These values were stored in a new column: **Churn\_Risk\_Score**
* The scores were then binarized with a defined threshold (e.g., ≥ 0.5) to assign a **CHURN-FLAG** (YES/NO)

✅ **Benefits**:

* Helps prioritize retention efforts for high-risk customers.
* Enables tailored retention campaigns based on risk level.
* Quantifies customer churn vulnerability using a repeatable, data-driven method.