**Customer Churn Prediction**

**Problem Statement**

The task is to develop a machine learning model that predicts customer churn based on historical customer data. The provided dataset contains information about customer attributes, interactions, and whether they churned or not. Customer churn refers to customers discontinuing their business relationship with the company.

# ****Problem Solving Approach****

To solve this problem, I need to follow a typical machine learning project pipeline. Here's a step-by-step approach to tackle the tasks mentioned in the assignment:

**Data Preprocessing:** Load the provided dataset using a library like pandas and perform initial data exploration. This includes checking for the structure of the data, data types, and a summary of statistics.

**Handle missing data:** Identify and address missing values, which can be done by imputing values or removing rows/columns with missing data, depending on the situation.

**Outlier handling:** Identify and deal with outliers in the data using techniques such as z-score or IQR (Interquartile Range).

**Data encoding:** Convert categorical variables into numerical form, typically using one-hot encoding or label encoding. Split the dataset into training and testing sets. This allows you to train the model on one subset and evaluate its performance on another.

**Feature Engineering:** Generate relevant features from the dataset that can help improve the model's prediction accuracy. For example, you can calculate customer tenure, interaction frequency, or customer satisfaction scores. Apply feature scaling or normalization if necessary, especially if using models sensitive to feature scales (e.g., logistic regression, support vector machines).

**Model Building:** Choose appropriate machine learning algorithms. Common choices for customer churn prediction include logistic regression, random forest, decision trees, support vector machines, and neural networks. Train and validate the selected model on the training dataset. Use techniques like cross-validation to ensure robustness. Evaluate the model's performance using appropriate metrics, such as accuracy, precision, recall, F1-score, and the ROC curve.

**Model Optimization:** Fine-tune the model parameters to improve predictive performance. You can use techniques like grid search or random search for hyperparameter tuning. Explore cross-validation to ensure the model's generalization to unseen data.

**Model Deployment:** Deploy the model into a simulated production-like environment where it can take new customer data as input and provide churn predictions. Ensure the model is integrated with your application or systems for real-time predictions.

**Documentation and Reporting:** Document your code and the steps you've taken in a Python script. Provide a brief report summarizing your approach, including details on data preprocessing, feature engineering, model selection, and deployment. Include model performance metrics and visualizations in your report to demonstrate the effectiveness of the model.

Overall, the goal is to build a predictive model that helps Sunbase understand and potentially reduce customer churn. The evaluation will be based on your ability to handle data, engineer features, select and optimize models, and deploy them effectively.