**TASK-1**

* **PROBLEM STATEMENT:**

The categorization of customer turnover in a telecommunication dataset is the issue that the offered code attempts to solve. The goal variable in the dataset is "Churn," which indicates whether a customer has churned (1) or not (0). The dataset comprises a variety of information pertaining to customer behaviour and qualities. Developing a prediction model that correctly categorises consumers as prospective churners or non-churners is the aim.

* **INTERPRETATION & OBSERVATION**

The output of the provided code reveals a well-constructed and effective machine learning model for predicting customer churn in a telecommunications dataset. Through careful preprocessing, including data splitting and class-balanced upsampling, the Random Forest Classifier, embedded within a pipeline, achieves high accuracy (94.04%) and balanced performance in precision, recall, and F1-score for both churn and non-churn classes. Cross-validation and hyperparameter tuning further enhance the model's robustness and generalization capability. The confusion matrix illustrates the model's ability to correctly identify instances of churn and non-churn. The overall outcome signifies a reliable tool for businesses to proactively identify potential churners and implement targeted retention strategies, contributing to improved customer relationship management. The exportable pipeline is ready for deployment, allowing real-time predictions on new data, making it a valuable asset for practical applications.

**TASK-2**

* **PROBLEM STATEMENT:**

The code supplied addresses an issue that pertains to predicting customer churn in a dataset connected to telecommunications. Numerous parameters, including call failure, complaints, duration of subscription, charge amount, use seconds, usage frequency, SMS frequency, unique called numbers, age group, tariff plan, customer status, age, and customer value are all included in the dataset. "Churn," which indicates whether a client has churned (1) or not (0), is the desired variable.

* **INTERPRETATION & OBSERVATION**

The distribution of consumers who have churned and those who have not is visually represented by the scatter plot based on PCA components color-coded by the desired variable ("Churn").

According to the Elbow Method, two clusters is the ideal number.

The K-Means clustering scatter plot illustrates the partitioning of consumers into two groups. Cluster centres are shown by the red points.

The quality of the clustering is indicated by the silhouette score. Better-defined clusters are suggested by a higher silhouette score.

The agreement between the actual and anticipated cluster allocations is represented visually by the confusion matrix.