**Customer Churn Prediction Analysis**

**Objective**

The objective of this customer churn prediction analysis is to leverage data-driven insights to identify and understand the factors influencing customer churn within our organization. By predicting churn, we aim to reduce customer attrition, enhance customer retention strategies, and ultimately improve our business's overall performance.

**Design Thinking Process**

**Empathize**

In this initial phase, we aim to understand our customers' needs, concerns, and preferences. We will:

* Gather information on customer demographics and behaviors.
* Conduct customer surveys and interviews to gain qualitative insights.
* Explore historical churn data to identify common patterns.

**Define**

During this phase, we will define the specific problem and goals of the churn prediction analysis. We will:

* Clearly define what constitutes "churn" for our business.
* Set clear objectives and key performance indicators (KPIs) for the analysis.
* Create user personas to represent different customer segments.

**Ideate**

The ideation phase involves brainstorming and generating ideas for addressing the churn problem. We will:

* Brainstorm potential features and data sources that can be used for analysis.
* Explore different machine learning algorithms and modeling approaches.
* Consider external factors (e.g., market trends) that might influence churn.

**Prototype**

We will create a prototype or plan for the churn prediction analysis. This involves:

* Selecting the data sources and datasets to be used.
* Designing the data preprocessing and feature engineering pipeline.
* Outlining the structure of predictive models to be developed.

**Development Phases**

**Data Collection and Preprocessing**

In this phase, we will collect and preprocess the data needed for the analysis. Steps include:

* Gathering historical customer data, including demographic, transactional, and behavioural information.
* Cleaning and preparing the data by handling missing values and outliers.
* Feature engineering to create relevant predictors for churn prediction.

**Model Development**

The heart of the analysis lies in model development. This involves:

* Selecting appropriate machine learning algorithms (e.g., logistic regression, random forests, neural networks).
* Splitting the dataset into training and testing sets for model evaluation.
* Training and fine-tuning the chosen models using the training data.

**Model Evaluation**

To ensure the effectiveness of our churn prediction models, we will:

* Evaluate model performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score).
* Conduct cross-validation to assess model generalizability.
* Interpret the results and identify key features influencing churn.

**Deployment and Continuous Improvement**

Once the models are ready, we will deploy them into our operational environment. This phase includes:

* Integrating the model into our customer relationship management (CRM) system.
* Monitoring model performance in real-time and updating it as needed.
* Continuously improving the analysis by incorporating new data and insights.

**Communication and Reporting**

Finally, we will communicate the findings and results to relevant stakeholders, such as:

* Creating detailed reports with insights and recommendations.
* Presenting findings to the management team and other relevant departments.
* Collaborating with marketing and customer service teams to implement strategies based on the analysis.

This customer churn prediction analysis aims to reduce churn, enhance customer retention, and contribute to the overall success of our organization. By following the design thinking process and carefully executing the development phases, we can make data-driven decisions to tackle the challenge of customer churn effectively.

# Analysis Objectives

The primary objectives of this analysis are as follows:

1. **Churn Prediction:** To develop predictive models that can identify customers at risk of churning (leaving our services) in advance. This allows us to take proactive measures to retain these customers.
2. **Customer Insights:** To gain a deeper understanding of customer behavior and preferences by analyzing historical data. This will help in creating targeted retention strategies.
3. **Data-Driven Decision-Making:** To enable data-driven decision-making within the organization by providing actionable insights to various departments, including marketing and customer service.

# Data Collection Process

To achieve the analysis objectives, the data collection process involves the following steps:

1. **Data Sources:** Gather data from various sources, including transaction records, customer profiles, customer service interactions, and marketing campaigns.
2. **Data Cleaning:** Clean the data to handle missing values, outliers, and inconsistencies. Ensure data quality and integrity.
3. **Data Integration:** Integrate data from different sources into a unified dataset for analysis.
4. **Feature Engineering:** Create relevant features or variables that capture customer behavior and characteristics, such as frequency of interactions, purchase history, and customer demographics.
5. **Data Split:** Divide the dataset into training and testing sets for predictive modeling.

# Predictive Modeling

Predictive modeling is a crucial component of this analysis, aimed at building models to predict customer churn. Here are the key steps involved:

1. **Data Preprocessing:** Prepare the data for modeling by encoding categorical variables, scaling features, and addressing any class imbalance issues.
2. **Model Selection:** Choose appropriate machine learning algorithms, such as logistic regression, decision trees, random forests, or neural networks, for churn prediction.
3. **Model Training:** Use the training dataset to train the selected models, fine-tuning hyperparameters for optimal performance.
4. **Model Evaluation:** Evaluate the models using relevant metrics like accuracy, precision, recall, and F1-score. Utilize techniques like cross-validation to assess their generalization capabilities.
5. **Feature Importance Analysis:** Determine which features have the most significant impact on churn prediction. Visualize this using IBM Cognos to aid in decision-making.
6. **Deployment:** Deploy the trained model into the operational environment to predict customer churn in real-time.
7. **Continuous Improvement:** Continuously monitor model performance and update it as needed to adapt to changing customer behavior.

# Visualizations using IBM cognos:

### customerID by Churn

35

30

25

20

customerID (Count distinct)

15

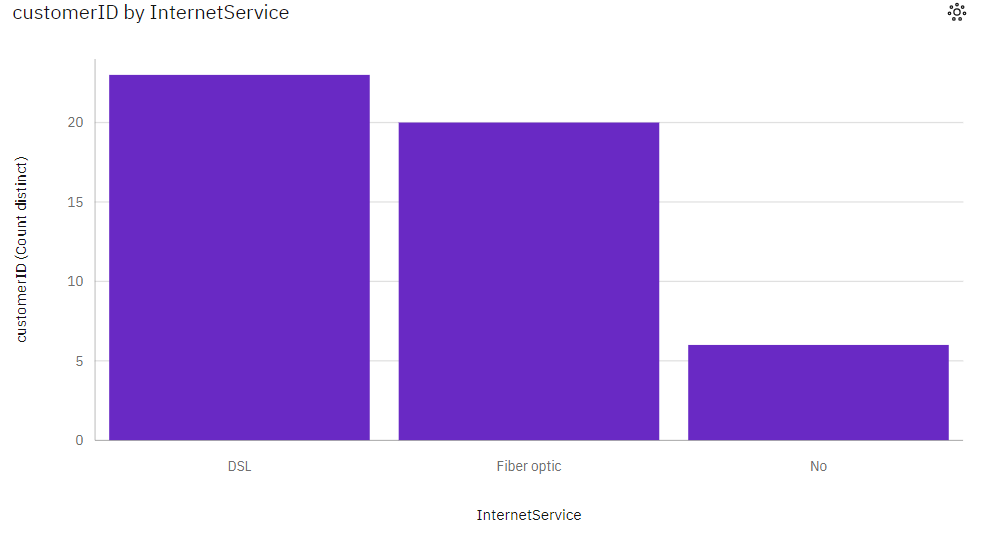
10

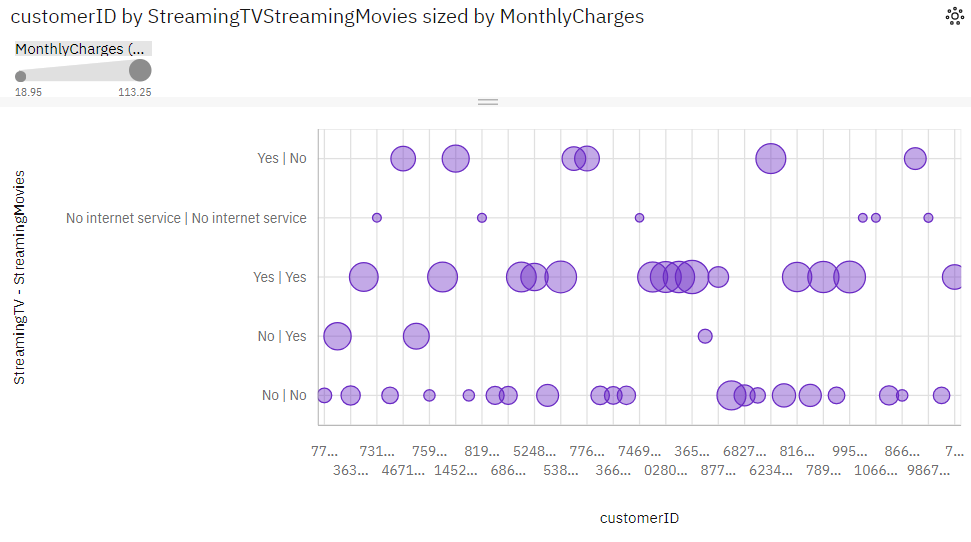
5

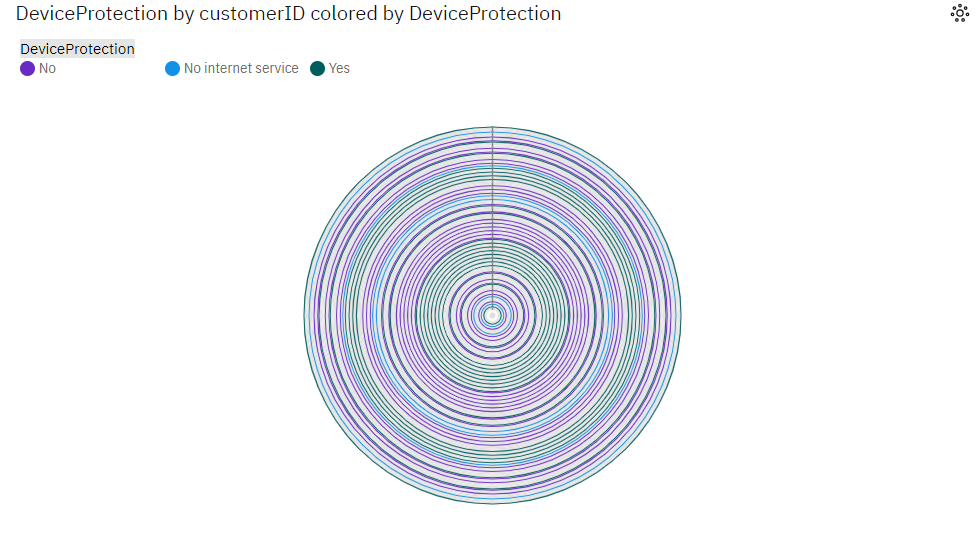
0

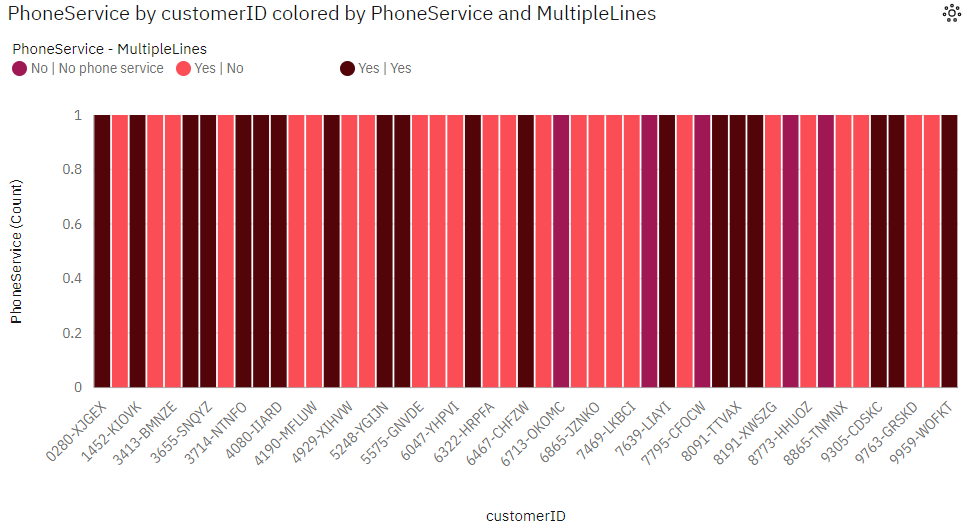
No Yes

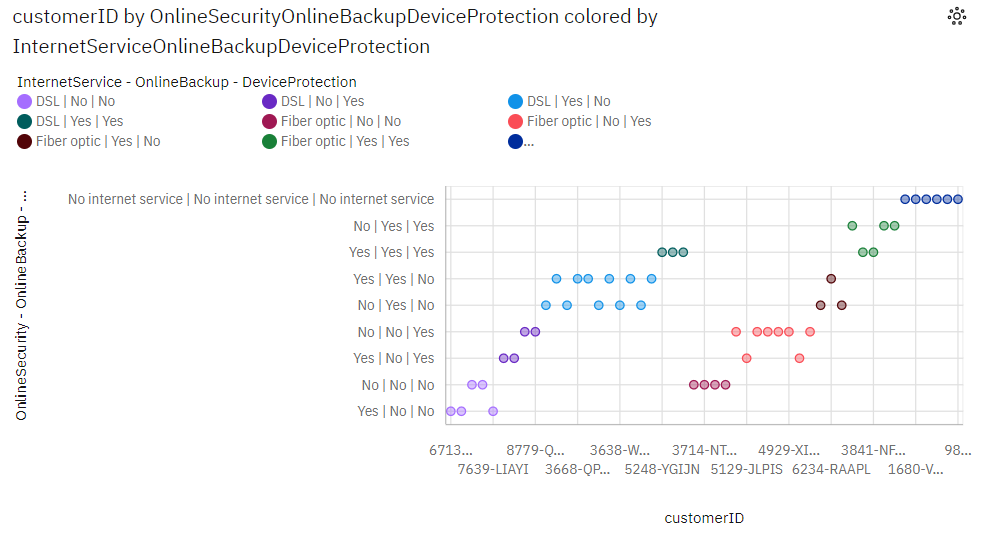
Churn

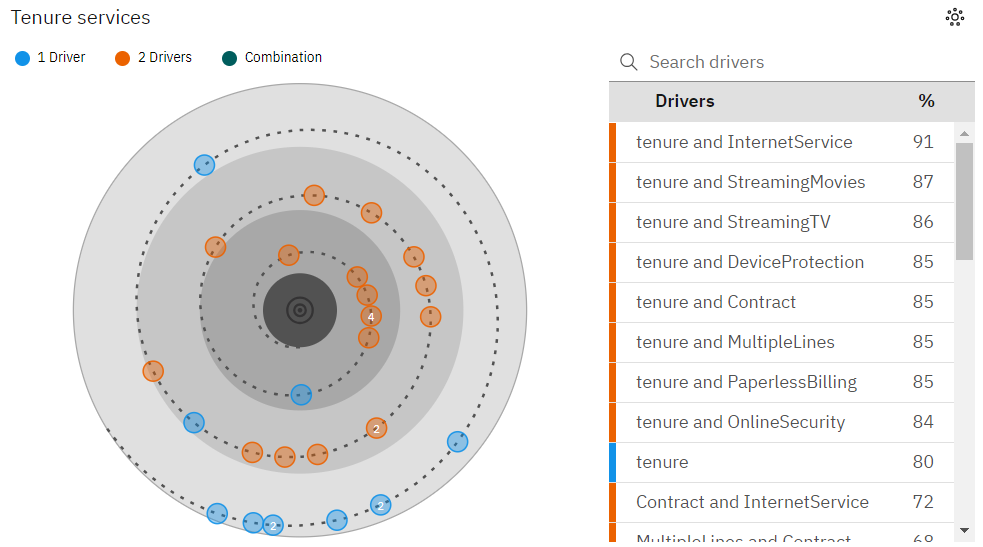














**Customer Churn Prediction Analysis Insight:**

1. **Early Identification of Churn Risk:**
   * Prediction models can identify customers who are at risk of churning well before they actually leave. This early warning allows businesses to take proactive measures.
2. **Personalized Retention Strategies:**
   * Insights derived from the analysis can reveal the specific reasons why customers churn. Businesses can then tailor retention strategies to address these specific issues for different customer segments.
3. **Optimizing Customer Engagement:**
   * The analysis can provide insights into which customer touchpoints are most effective in retaining customers. Businesses can allocate resources to the channels and actions that have the greatest impact.
4. **Product and Service Improvement:**
   * By understanding customer behavior and preferences, businesses can make data-informed decisions to improve their products or services. Addressing customer pain points can enhance satisfaction and reduce churn.
5. **Segmentation and Targeting:**
   * Customer segmentation based on churn risk or behavior can enable businesses to target specific groups with relevant offers and promotions. This increases the chances of retaining those most likely to churn.
6. **Feedback Loop for Marketing:**
   * By analyzing the results of marketing campaigns, businesses can fine-tune their strategies. Understanding which campaigns are effective at retaining customers allows for better allocation of marketing resources.
7. **Resource Allocation Efficiency:**
   * Knowing which customers are more likely to churn helps in allocating resources efficiently. It allows businesses to focus on high-risk customers while not overspending on low-risk ones.
8. **Continuous Improvement:**
   * Insights and predictive models are not static. Businesses can continuously improve their retention efforts by monitoring and adapting to changing customer behavior and market conditions.
9. **Cost Reduction:**
   * Acquiring new customers is often more expensive than retaining existing ones. By reducing churn, businesses can save on customer acquisition costs and potentially increase profitability.
10. **Enhancing Customer Experience:**
    * By addressing the pain points and concerns that lead to churn, businesses can enhance the overall customer experience. Satisfied customers are more likely to stay loyal.
11. **Competitive Advantage:**
    * Businesses that effectively reduce churn gain a competitive advantage. They build a reputation for customer satisfaction, which can attract new customers and strengthen their position in the market.
12. **Long-term Growth:**
    * By reducing churn and retaining customers over the long term, businesses can achieve sustainable growth. A loyal customer base can provide a steady source of revenue and referrals.

# Algorithm:

**Step 1: Import Necessary Libraries**

* Import essential Python libraries, including NumPy, pandas, Matplotlib, Seaborn, Missingno, and various components from Scikit-Learn.

**Step 2: Set Configurations**

* Suppress warnings using **warnings.filterwarnings("ignore")**.
* Set an option to display all columns in DataFrames using **pd.options.display.max\_columns = None**.
* Read a CSV file into a DataFrame called **data**.

**Step 3: Data Preprocessing**

* Drop the 'customerID' column from the DataFrame using **data.drop('customerID', axis=1, inplace=True)**.
* Replace empty strings in the 'TotalCharges' column with NaN.
* Convert the 'TotalCharges' column to a float data type.
* Convert all column names to lowercase using **data.columns = data.columns.str.lower()**.
* Identify and separate numerical columns and other categorical columns.
* Handle missing values using iterative imputation.

**Step 4: Data Visualization**

* Create various plots to visualize the data distribution, relationships between numerical features, and the distribution of categorical features. This includes histograms, rug plots, box plots, and count plots.

**Step 5: Encoding Categorical Features**

* Label-encode some ordinal categorical columns using a custom **lencoder** function.
* Label-encode the target variable ('churn').
* One-hot-encode other categorical columns.

**Step 6: Machine Learning with PyCaret**

* Install the **pycaret** library.
* Set up your machine learning experiment using **setup** with the target variable 'churn' and other configurations.
* Compare different machine learning models based on the AUC (Area Under the Curve) using **compare\_models**.
* Select the best model based on the AUC score and plot its features and AUC.
* Calculate custom profit-based metrics using a custom **calculate\_profit** function and select the best model based on profit.
* Save the best model as 'churn-predict' using **save\_model**.

**Conclusion**

The dataset is now ready for further analysis and model development. The preprocessing steps undertaken have enhanced the dataset's quality and will contribute to the accuracy of the predictive model for customer churn. The next phase involves model selection, training, evaluation, and eventually utilizing the predictive model to assist the business in reducing customer churn and improving overall performance