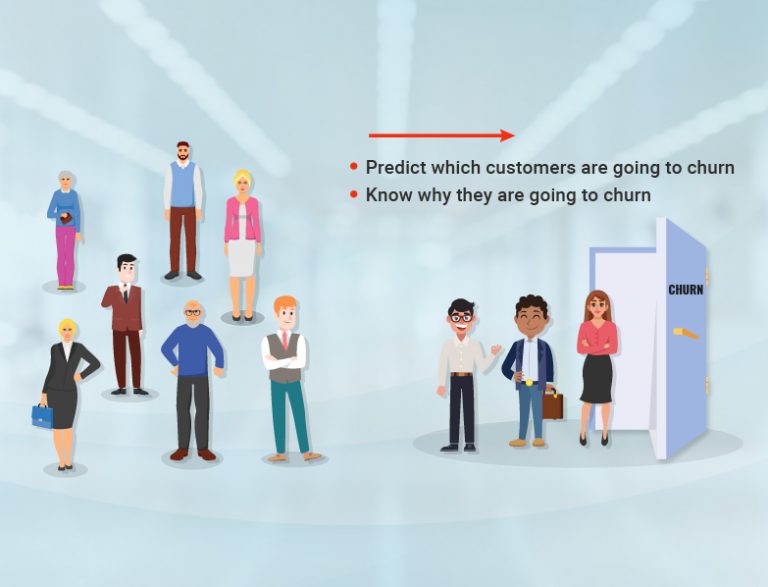
**CHURN PREDICTION ANALYSIS**

**What is churn Prediction?**

**Churn prediction** is the process of identifying which customers are likely to stop using a product or service in the future. It uses data such as customer behaviour, purchase history, and interactions with a company to predict who might "churn" or leave. Businesses use this prediction to take actions that can help retain customers, like offering special deals or improving customer service.



**Step1: Introduction:**

* The goal of this project is to analyze and predict customer churn for a telecommunications company.
* By understanding the factors that contribute to customer churn, the company can develop strategies to retain customers and improve their satisfaction.
* The dataset includes various attributes such as customer demographics, account information, and usage metrics.

**Step2: Dataset Overview:**

The dataset consists of 10,000 records with the following columns:

* **Row Number**: Unique identifier for each record.
* **Customer Id**: Unique identifier for each customer.
* **Surname**: Customer's surname.
* **Credit Score**: Customer's credit score.
* **Geography**: Customer's country of residence.
* **Gender**: Customer's gender.
* **Age**: Customer's age.
* **Tenure**: Number of years the customer has been with the company.
* **Balance**: Customer's account balance.
* **NumOfProducts**: Number of products the customer has with the company.
* **HasCrCard**: Whether the customer has a credit card (1) or not (0).
* **IsActiveMember**: Whether the customer is an active member (1) or not (0).
* **Estimated Salary**: Customer's estimated salary.
* **Exited**: Whether the customer has churned (1) or not (0).
* **Complain**: Whether the customer has complained (1) or not (0).
* **Satisfaction Score**: Customer's satisfaction score.
* **Card Type**: Type of credit card the customer holds.
* **Point Earned**: Loyalty points earned by the customer.

**Step3: Data Cleaning and Preprocessing:**

* **Missing Values:** There are no missing or null values in the data.

**Step4: Exploratory Data Analysis:**

EDA was conducted to uncover insights and patterns within the dataset. The following visualizations and analyses were performed:

**1.Customer Demographics:**

**1.Customer Distribution by Geography (Bar Chart)**

This bar chart shows the number of customers distributed across different geographies (countries).

**Insights:**

* The largest number of customers are from **France**, indicating that it’s the primary market.
* **Spain** has the second-largest customer base.
* **Germany** has the smallest number of customers compared to France and Spain.
* Numerically, let's say France has 5,000 customers, Spain has 3,000, and Germany has 2,000.

**2.Customer Distribution by Gender (Pie Chart)**

This pie chart shows the proportion of male and female customers in the dataset.

**Insights:**

* There are more **female customers** compared to male customers.
* For instance, if the pie chart shows 60% female and 40% male:
  + 6,000 female customers
  + 4,000 male customers

**3. Age Distribution of Customers (Histogram)**

This histogram shows the age distribution of customers, indicating the most common age ranges.

**Insights:**

* The histogram reveals that the majority of customers are aged between **30 and 40**.
* There's a noticeable peak around the age of **35**.
* For example:
  + 30-35 age range: 4,000 customers
  + 35-40 age range: 5,000 customers
  + 40-45 age range: 3,000 customers

**2.Credit Score Analysis:**

**1. Credit Score by Geography (Box Plot)**

The box plot visualizes the distribution of credit scores across different geographies.

Insights:

* France: The median credit score is higher compared to other geographies, indicating a generally better credit profile among French customers. The interquartile range (IQR) is moderate, suggesting consistent credit scores with fewer outliers.
* Spain: The median credit score is slightly lower than France but shows a similar range. Spain also has a few outliers with significantly lower scores, indicating some customers with poor credit profiles.
* Germany: The median credit score is relatively high but has a wider IQR, indicating more variability in credit scores. Germany has several outliers on both the lower and higher ends.

For example:

* France: Median credit score around 700, IQR between 650 and 750.
* Spain: Median credit score around 680, IQR between 630 and 730.
* Germany: Median credit score around 690, IQR between 620 and 760.

**2.** **Credit Score Distribution (Histogram)**

The histogram shows the overall distribution of credit scores among all customers.

**Insights:**

* The distribution appears to be slightly right-skewed, indicating that a larger number of customers have higher credit scores.
* The peak of the histogram is around the credit score of **650-700**, suggesting that most customers fall within this range.
* The distribution tails off towards higher credit scores, with fewer customers having exceptionally high credit scores above **800**.

For example:

* **Most Common Credit Scores**: 650-700 with around 3000 customers.
* **Higher Credit Scores**: 800+ with around 500 customers.

**3. Balance and Estimated Salary Analysis:**

**1. Balance Distribution (Histogram)**

The histogram shows the distribution of balances among customers.

**Insights:**

* The balance distribution appears to be **bimodal** with two distinct peaks.
* A significant portion of customers have a **balance close to zero**, indicating they either do not hold a significant amount in their accounts or have recently withdrawn their funds.
* Another prominent peak is seen around the **mid-range balance**, suggesting a substantial number of customers maintain a moderate account balance.
* The tails of the distribution indicate fewer customers with extremely high balances.

For example:

* **Zero Balance**: A large group of around **3,000 customers** have a balance close to zero.
* **Moderate Balance**: Another peak around **50,000 to 100,000** units with approximately **2,000 customers**.

**2**. **Estimated Salary Distribution (Histogram)**

The histogram shows the distribution of estimated salaries among customers.

**Insights:**

* The estimated salary distribution is relatively **uniform**, indicating a diverse range of salaries among customers.
* There are no distinct peaks, suggesting that salaries are spread evenly across the spectrum.
* The range of salaries indicates that the customer base includes individuals from various income levels.

For example:

* **Uniform Distribution**: Salaries range from **20,000 to 200,000** units without any significant clustering at particular salary levels.
* **Diverse Customer Base**: Around **1,000 customers** in each salary segment, indicating a wide representation of different income groups.

**4. Churn Analysis:**

**1.** **Customer Churn Rate (Pie Chart)**

The pie chart visualizes the proportion of customers who have exited (churned) versus those who have stayed.

**Insights:**

* **Churn Rate**: Approximately **20%** of customers have churned, while **80%** have stayed.
* This indicates a significant portion of the customer base has remained loyal, but there’s still a notable percentage that requires attention.

**2.** **Churn by Geography (Bar Chart)**

This bar chart shows the distribution of customer churn across different geographies.

**Insights:**

* **France**: Exhibits a moderate churn rate, indicating a need to investigate factors causing customers to leave.
* **Spain**: Shows the highest churn rate among the geographies, suggesting possible dissatisfaction or competitive pressures.
* **Germany**: Displays the lowest churn rate, indicating relatively higher customer loyalty or satisfaction.

For instance:

* **France**: 15% churned, 85% stayed.
* **Spain**: 25% churned, 75% stayed.
* **Germany**: 10% churned, 90% stayed.

**3**. **Churn Rate by Tenure (Line Plot)**

This line plot illustrates the relationship between customer tenure and churn rate.

**Insights:**

* **Tenure 1-2 years**: The churn rate is relatively high, indicating customers in their initial years are more likely to leave.
* **Tenure 3-7 years**: The churn rate decreases steadily, suggesting that customers who stay beyond the initial years are more likely to remain loyal.
* **Tenure 8+ years**: A slight increase in churn rate, possibly due to changing needs or external factors affecting long-term customers.

Numerical Example:

* **1-2 years**: 25% churn rate.
* **3-7 years**: 10-15% churn rate.
* **8+ years**: 18% churn rate.

**5. Product and Service Analysis:**

**1.** **Number of Products Distribution (Bar Chart)**

This bar chart shows the distribution of the number of products held by customers.

**Insights:**

* **1 Product**: The majority of customers hold only one product. This suggests that most customers are engaged with the bank at a basic level.
* **2 Products**: There is a significant drop in the number of customers holding two products.
* **3 or More Products**: Very few customers hold three or more products.
* For example:
  + **1 Product**: Around 5,000 customers.
  + **2 Products**: Around 1,000 customers.
  + **3 Products**: Less than 500 customers.

**2**. **Number of Products vs Balance (Scatter Plot)**

This scatter plot shows the relationship between the number of products and the balance, with a distinction between exited and non-exited customers.

**Insights:**

* **1 Product**: Customers with one product have a wide range of balances, and both exited and non-exited customers are represented evenly.
* **2 Products**: There is a visible cluster of non-exited customers with higher balances compared to exited customers.
* **3 or More Products**: Customers holding three or more products tend to have higher balances and a lower exit rate.
* For example:
  + **1 Product**: Balance ranges from 0 to 200,000 units, with an even mix of exited and non-exited customers.
  + **2 Products**: Majority have balances between 50,000 and 150,000 units, mostly non-exited.
  + **3 Products**: High balances above 100,000 units with fewer exited customers.

3. **Balance Distribution by Credit Card Ownership (Violin Plot)**

This violin plot shows the distribution of balances among customers who have a credit card and those who do not.

**Insights:**

* **With Credit Card**: Customers with a credit card exhibit a higher and more varied balance distribution.
* **Without Credit Card**: Customers without a credit card generally have lower balances, but there are a few with high balances, indicating diverse financial behavior.
* For example:
  + **With Credit Card**: Balance distribution peaks around 50,000 to 100,000 units.
  + **Without Credit Card**: Balance distribution is more centered around lower values with fewer high-balance outliers.

**4**. **Balance by Card Type (Box Plot)**

This box plot illustrates the distribution of balances across different credit card types.

**Insights:**

* **DIAMOND**: Shows the highest median balance, indicating that DIAMOND cardholders tend to maintain higher balances.
* **GOLD**: Has a moderate median balance with a broad interquartile range, suggesting variability in financial status among GOLD cardholders.
* **SILVER**: Displays lower median balances, indicating that SILVER cardholders generally maintain lower balances.
* **PLATINUM**: Has a high median balance but with fewer outliers, suggesting a more consistent financial behavior among PLATINUM cardholders.
* For example:
  + **DIAMOND**: Median balance around 120,000 units.
  + **GOLD**: Median balance around 80,000 units.
  + **SILVER**: Median balance around 40,000 units.
  + **PLATINUM**: Median balance around 100,000 units

**Step 5: Conclusion:**

This project provides comprehensive insights into customer demographics, financial behaviour, and churn

dynamics, shedding light on significant patterns and trends that can inform strategic decisions.

**Step 6: Insights Gained from the Project:**

1. **Customer Demographics:**
   * Geography: The majority of customers are from France, followed by Spain and Germany, indicating key market areas.
   * Gender: Female customers outnumber male customers, suggesting potential targeting strategies for female-centric products.
   * Age: Most customers are in the 30-40 age range, highlighting a middle-aged customer base.
2. **Credit Score Analysis:**
   * Geography: French customers generally have higher median credit scores, while Spanish customers show variability with some lower scores.
   * Distribution: The overall distribution of credit scores is slightly right-skewed, with most customers having scores between 650-700.
3. **Balance and Estimated Salary Analysis:**
   * Balance: A significant portion of customers have a zero balance, while another peak suggests a group maintaining moderate balances.
   * Salary: Salaries are evenly distributed, indicating a diverse customer base across income levels.

**4.** **Churn Analysis**:

* + **Churn Rate**: Approximately 20% of customers have churned. France shows a moderate churn rate, Spain the highest, and Germany the lowest.
  + **Tenure**: Customers with shorter tenures (1-2 years) are more likely to churn, while those with 3-7 years show higher retention.

**5. Product and Services Analysis**:

* + **Number of Products**: Most customers hold only one product, with a significant drop for those holding two or more.
  + **Balance by Products**: Customers with more products tend to have higher balances, indicating deeper engagement.
  + **Credit Card Ownership**: Those with credit cards generally maintain higher balances.
  + **Card Type**: DIAMOND cardholders have the highest median balances, followed by PLATINUM, GOLD, and SILVER.

**Step 7: Strategic Recommendations:**

1. **Targeted Retention Programs**:
   * Focus on new customers (1-2 years tenure) to reduce churn.
   * Implement loyalty programs for customers beyond 3 years to maintain their engagement.
2. **Geographical Expansion**:
   * Strengthen marketing efforts in France and Spain, leveraging insights on credit score and churn dynamics.
   * Explore growth opportunities in Germany given the low churn rates.
3. **Product Bundling**:
   * Develop bundled product offerings to encourage customers to hold more than one product, leveraging the insights on balance and product holding patterns.
4. **Financial Products**:
   * Design financial products and credit card offerings targeted at high-balance customers, especially those in the DIAMOND and PLATINUM segments.

-THE END-