**Syriatel Customer Churn Data Report**

1. **Overview**

Syriatel, a telecommunications company, aims to minimize revenue losses caused by customer churn. The company is focused on identifying the key factors contributing to customer attrition and understanding the reasons behind customers discontinuing their services.

1. **Problem Statement**
   1. **Problem**

The main task of this project is to identify the factors driving customer churn and develop actionable strategies to reduce it. This will help the company to take appropriate actions on time to avoid losing customers and thus revenue at the same time.

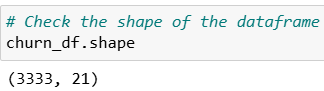
* 1. **Objectives**
* Develop a predictive model to identify customers at risk of churning and any characteristics that are indicative of churn.
* Focus retention efforts on the most at-risk segments to maximize return on investment in customer satisfaction programs.
* Explore patterns and behaviors and use the insights to implement targeted interventions, such as proactive customer support or offering better plans for high usage customers.
  1. **Outcomes**

At the end of the project, we should be able to come up with:

* A predictive model capable of identifying at-risk customers.
* Insights into key factors driving churn.
* A system for proactive customer retention strategies.

1. **Data Understanding**

The dataset contains 3,333 rows and 21 columns. This can be confirmed by checking the dimension of the dataset.



Here's a summary of its structure:

**Summary:**

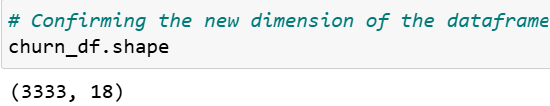
Columns: 21, including numerical, categorical, and boolean data types.

The respective columns are:

* state: Customer state (e.g., KS, OH).
* account length: Length of customer account in days.
* area code: Area code of the customer.
* phone number: Customer phone number.
* international plan: Whether the customer has an international calling plan (yes or no).
* voice mail plan: Whether the customer has a voicemail plan (yes or no).
* number vmail messages: Number of voicemail messages.
* total day minutes, total day calls, total day charge: Metrics for daytime usage.
* total eve minutes, total eve calls, total eve charge: Metrics for evening usage.
* total night minutes, total night calls, total night charge: Metrics for nighttime usage.
* total intl minutes, total intl calls, total intl charge: Metrics for international usage.
* customer service calls: Number of customer service calls made.
* churn: Whether the customer churned (True or False).
  1. **Exploratory Data Analysis (EDA)**

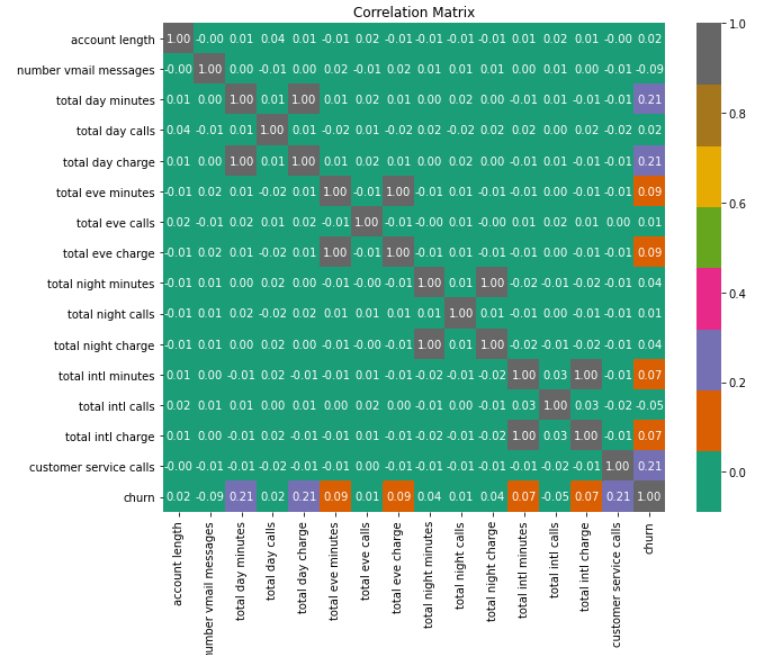
We now do EDA to understand the structure, relationships, and key insights of the data. Before that, drop columns that will not be useful for analysis such as phone number, area code and state.

After dropping the columns above, the dimension of the data frame will have changed as follows:



* 1. **Visualizations to show some key metrics about the data frame**

1. Correlation Matrix

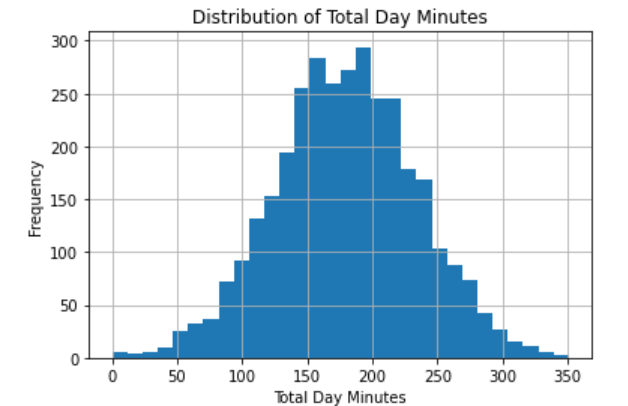
­ 

It is clear from the correlation matrix that the factors that have a high correlation with churn are:

* Total day minutes
* Total day charge
* Customer service calls

1. Histogram for Total Day Minutes

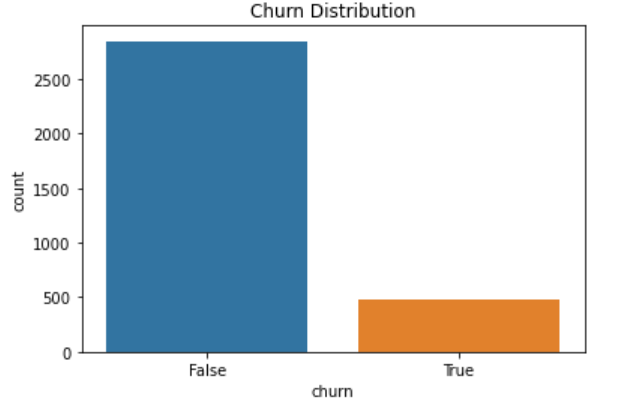
We also want to examine the distribution of minutes utilized by a customer during the day.



The histogram shows a somewhat normal distribution for the amount of minutes utilized by a customer during the day.

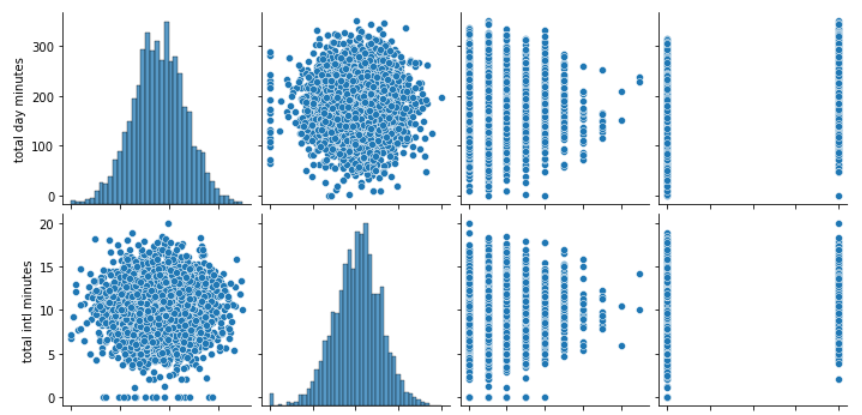
1. Countplot for churn distribution

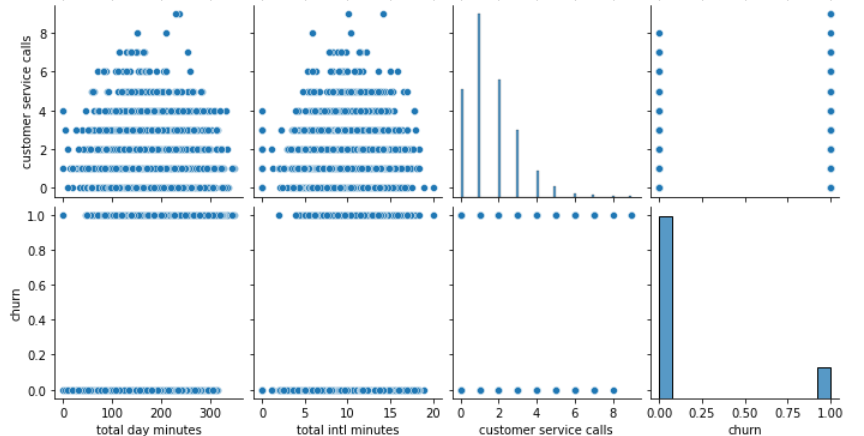
The plot shows the count of customers who have left the business and those who haven’t.



1. Pairplot for pairwise relationship between different variables.

The pairplot shows the relationship between different variables in the dataset and will come in handy in building a model that will help to predict churn.­





From the pairplots above, it is clearly visible that the relationship between most variables do have a somewhat normal distribution.