Project on telecom customer's churn

Abstract:

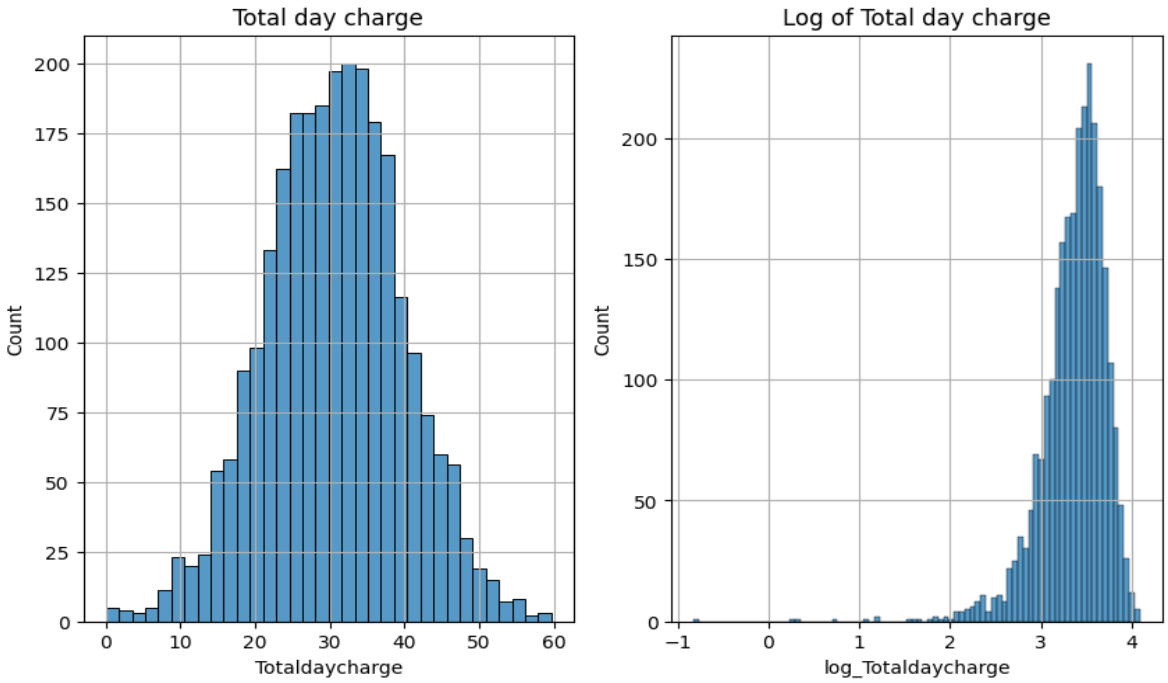
This study aims to analyze a dataset containing information on charges incurred during different times in a day and their distribution internationally. The dataset comprises diverse transactions from various sources, capturing both temporal and geographical aspects of the charges. The primary objective is to examine the temporal distribution of charges, discerning patterns and trends throughout a 24-hour period. Additionally, an investigation into the international distribution of charges will be conducted to identify any variations in expenditure across different countries or regions.

To achieve this, advanced data analysis techniques will be applied, including statistical methods and data visualization tools. The temporal analysis will involve aggregating charges by time intervals to reveal peak periods of activity and uncover potential diurnal spending habits. Meanwhile, the international distribution will entail mapping charges to countries or regions and exploring variations in spending patterns.

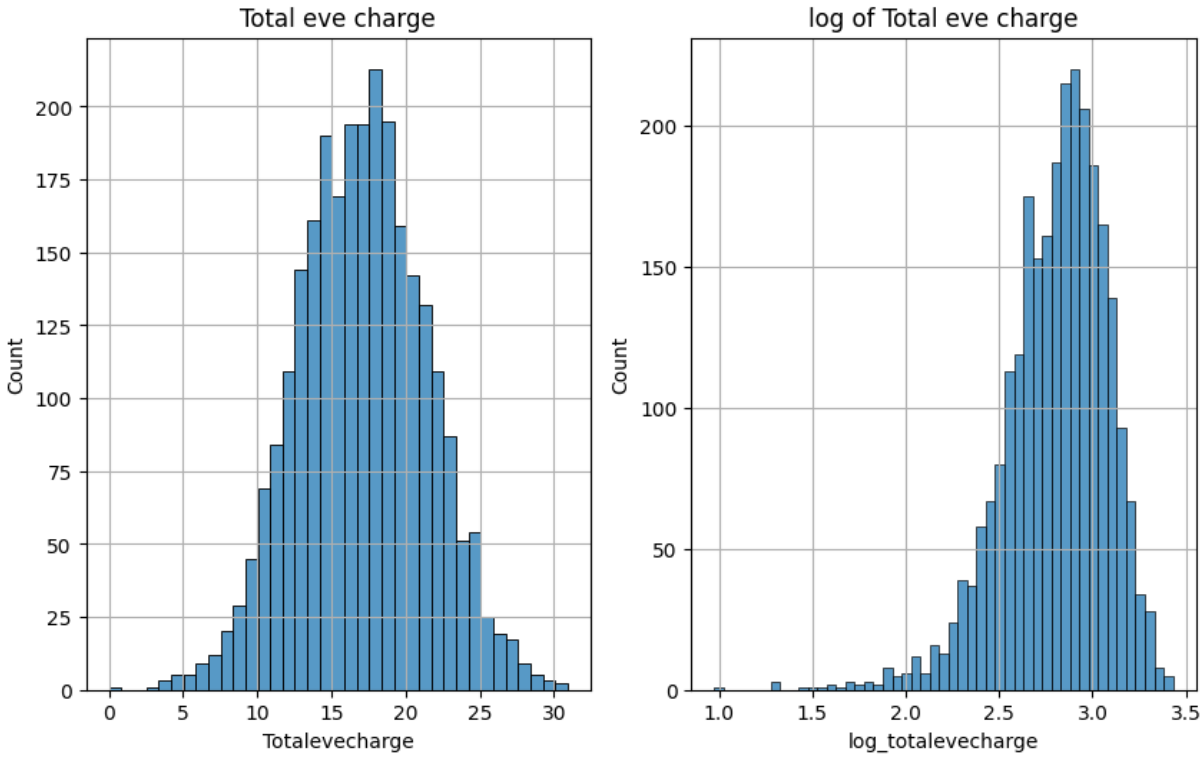
The findings from this research will provide valuable insights into consumer behaviour, temporal spending preferences, and international spending patterns. Such knowledge could have significant implications for businesses, policymakers, and financial institutions, enabling them to optimize their strategies, tailor services, and enhance decision-making processes. Additionally, the methodology and insights generated from this study could serve as a foundation for further research in related fields, contributing to the advancement of data analysis and its applications.

Problem Statement:

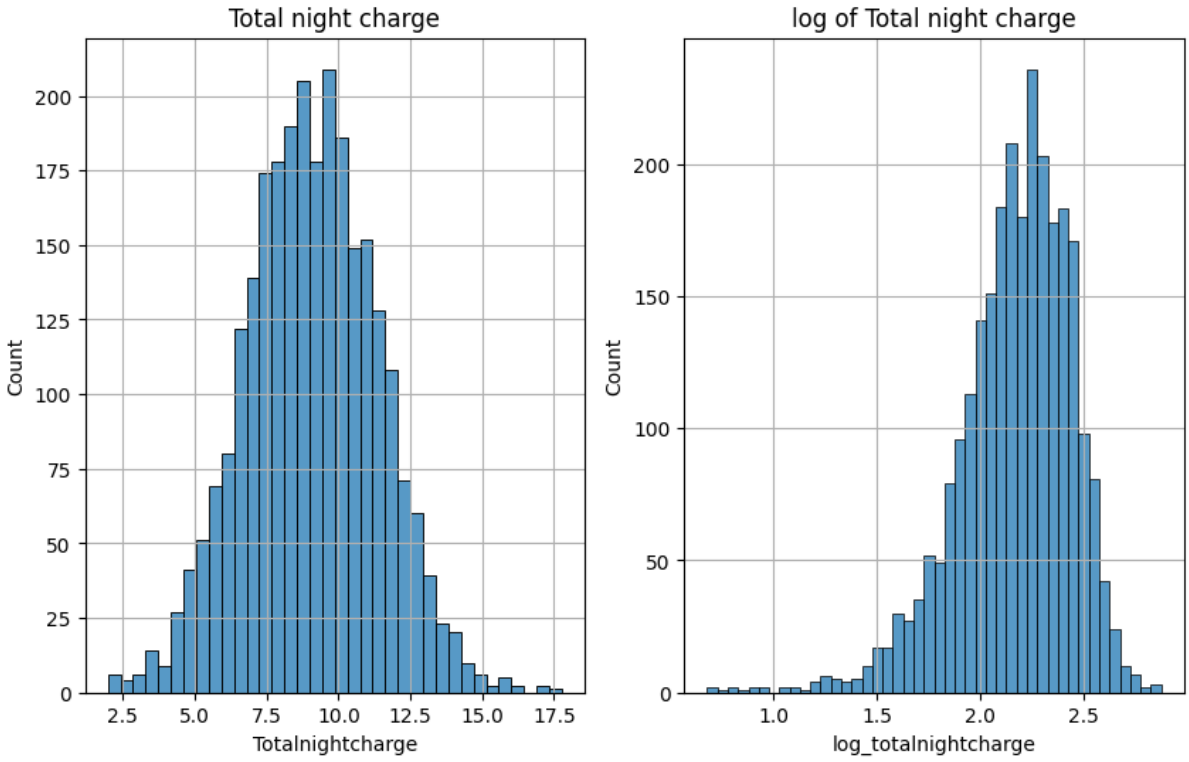
1. **What is the Distribution of the charges in all the given points in a day for the 1st table?**
   The distribution at each point in the day is a normal distribution with no outliers (skewness) in the histogram. The difference in the charges in each point of the day is shown by the decreasing value of the x-axis (charges) from the 1st table used (telecom). The following graphs show the accurate representation and their logarithmic counters (left skewed data).



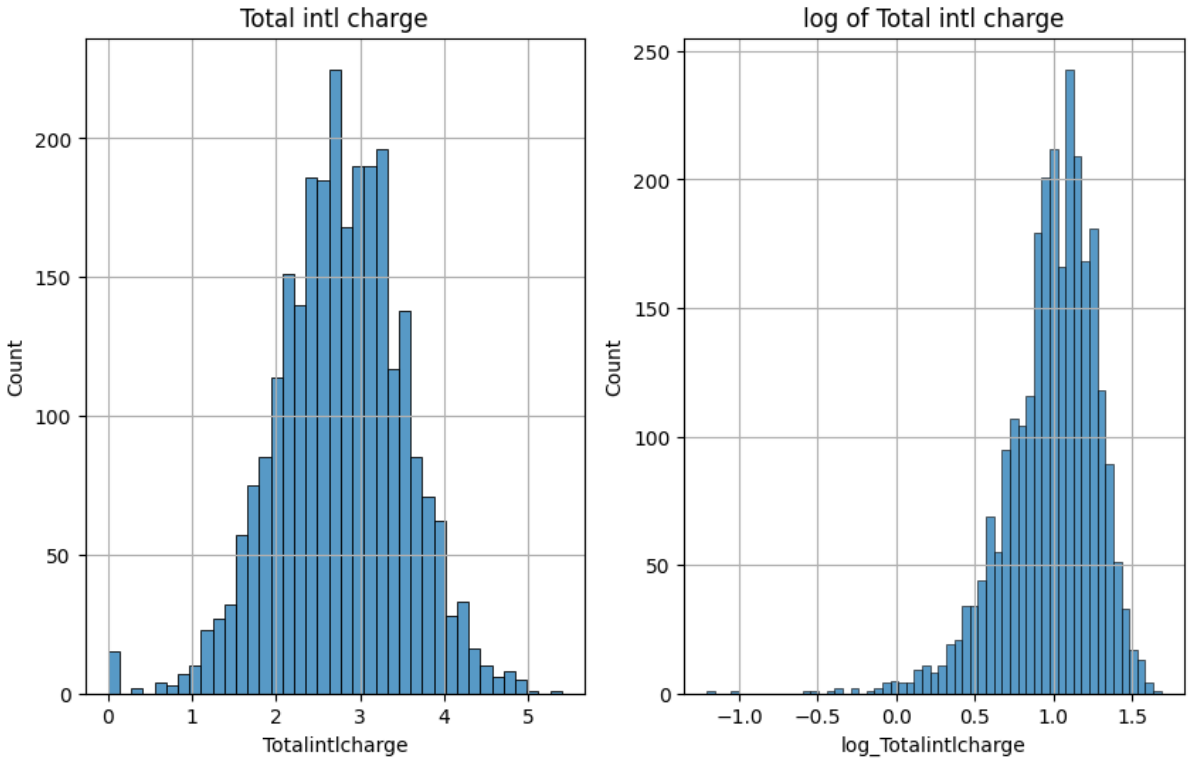
Representation of Day charges (Uni-variate)



Representation of evening charge (Uni-variate)



Representation of night charges (Uni-variate)



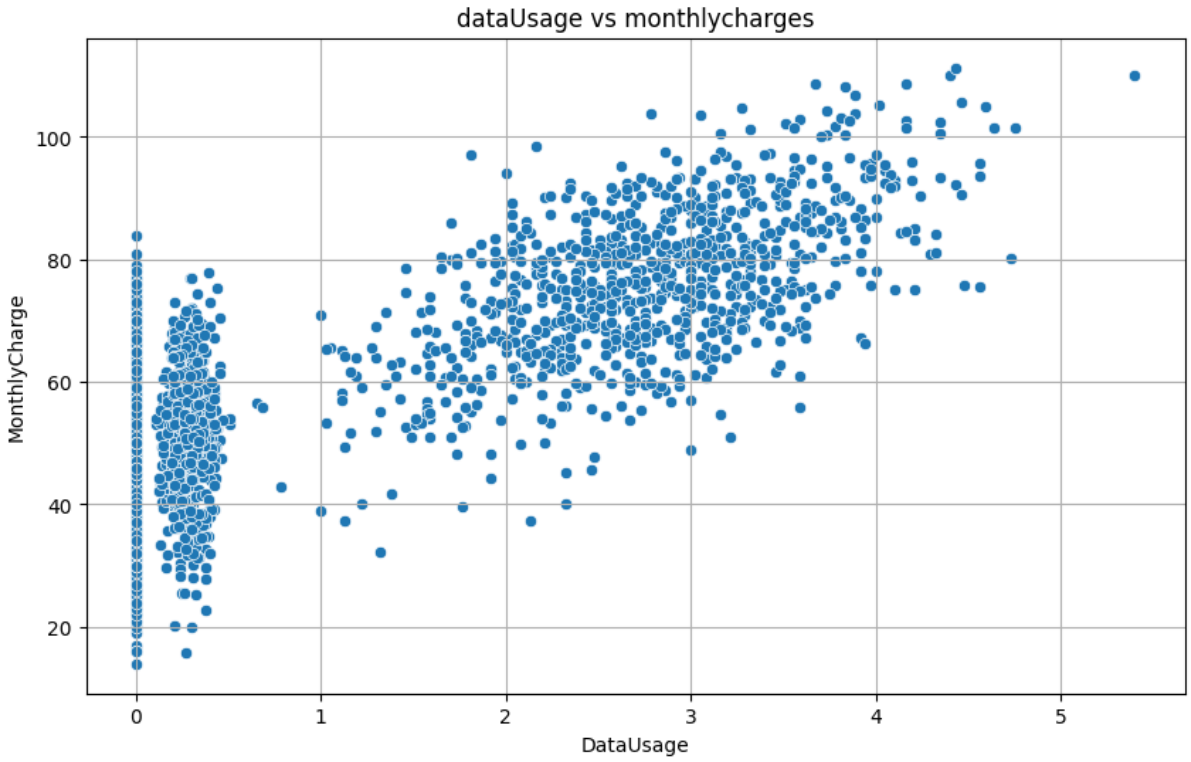
Representation of International charges (Uni-variate)

2. **What is the interpretation from the 2nd table for data usage vs the monthly charges?**

We can see that the 1st region

- (DataUsage = 0), we can see charges being incurred even if the usage is 0.
- In the region DataUsage>0 but DataUsage<1, we can see a cluster of Charges, which means that a medium portion of customers use less than 1 GB data Per month and are charged between 20-80$ a month..
- In the region DataUsage>1, we can see that a large portion of users are in the 2-3GB usage region, which means that many users use 2-3 GB per month and Their charges vary from 60-100$ a month.

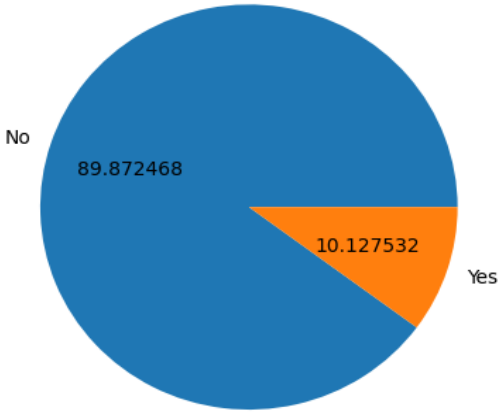
The below graph shows the scatter plot for the above explanation:



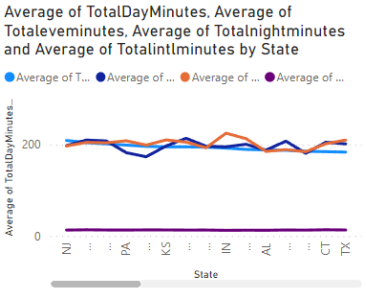
Scattered representation of Bi-variate data (DataUsage vs MonthlyCharges)

3. **What were the other data observed from the given datasets?**

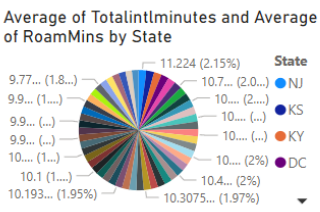
From both the datasets, we have studied the customer service call comparisons, Account active status comparison, Churn to contract renewal comparison for the 2nd dataset, Data usage based on roaming minutes used, churn comparison for each point of the day for the 1st dataset, Call minutes average for each point of time on a given day etc.



Pie representation of international plan users for the 1st dataset



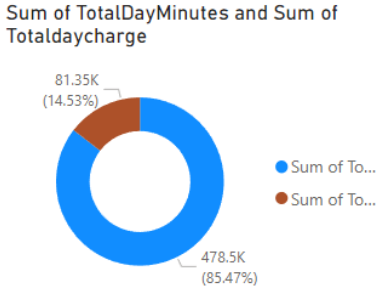
At 209.23, NJ had the highest Average of TotalDayMinutes and was 44.86% higher than WI, which had the lowest Average of TotalDayMinutes at 144.43. Across all 51 State, Average of TotalDayMinutes ranged from 144.43 to 209.23, Average of Totaleveminutes ranged from 173.52 to 227.54, and Average of Totalnightminutes ranged from 183.77 to 225.14.

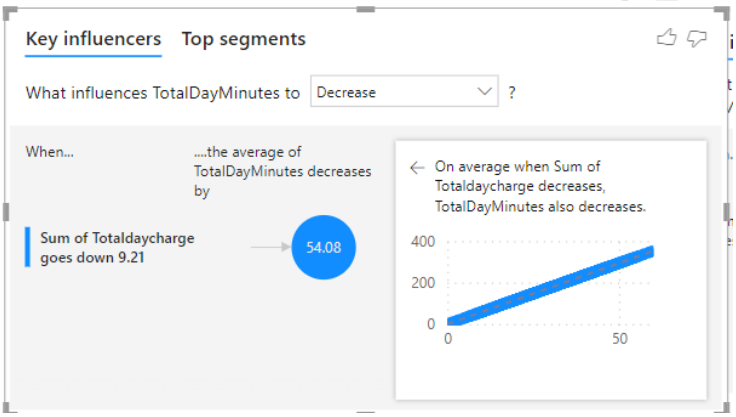


At 209.23, NJ had the highest Average of TotalDayMinutes and was 44.86% higher than WI, which had the lowest Average of TotalDayMinutes at 144.43.

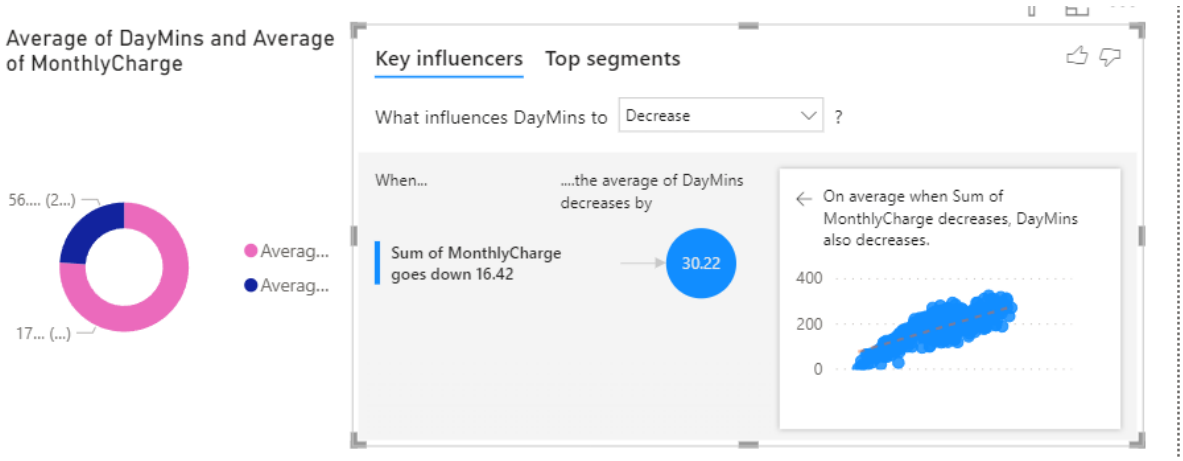
Across all 51 State, Average of TotalDayMinutes ranged from 144.43 to 209.23, Average of Totaleveminutes ranged from 173.52 to 227.54, and Average of Totalnightminutes ranged from 183.77 to 225.14.

The above 2 data representation shows the average of call minutes in each point of the day for the respective states for the 1st Dataset.

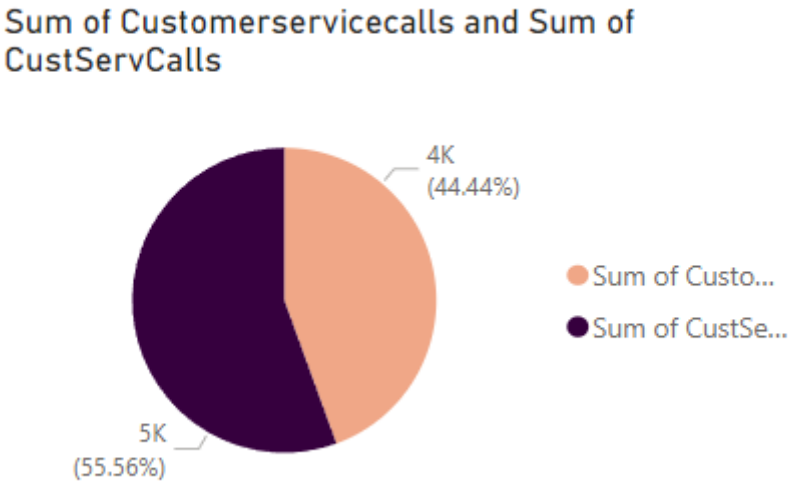




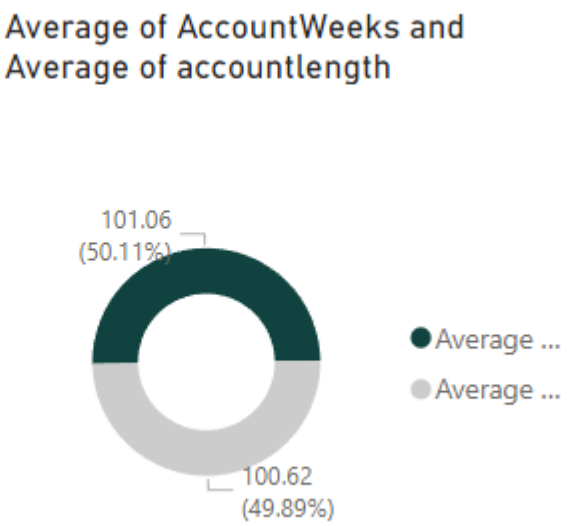
Comparison of Total Day minutes and the charges.



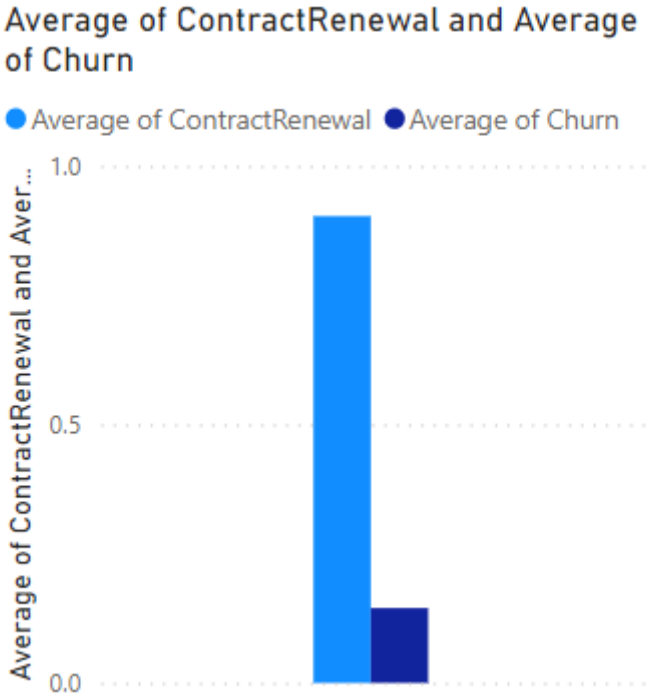
Average of DayMins to the monthly charge (Charges according to the minutes used)



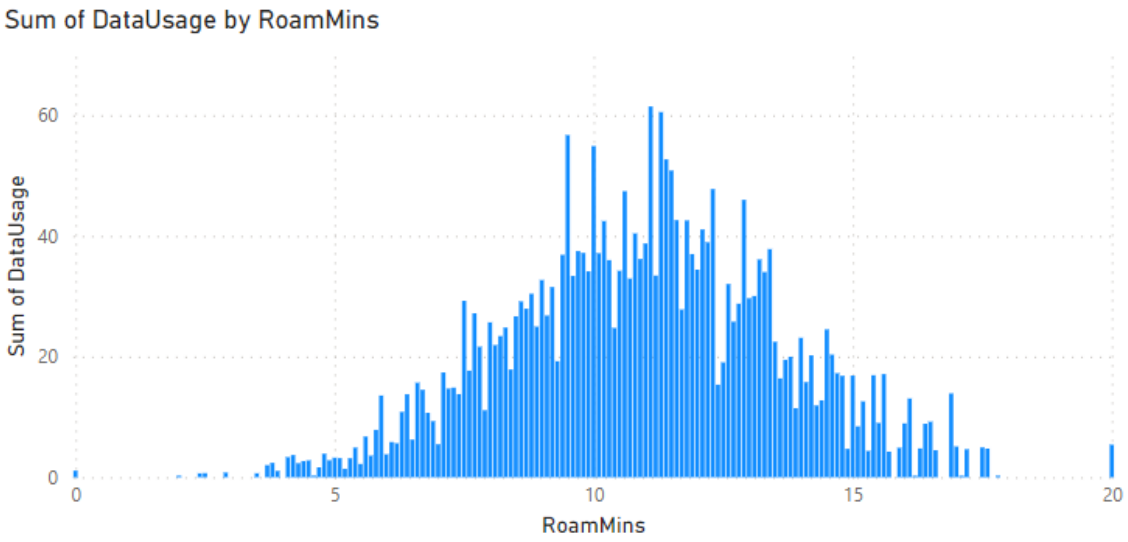
Customer service calls comparison between the two telecom table data used.



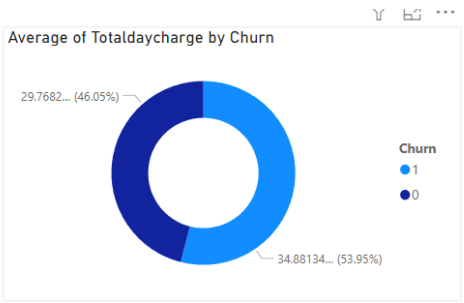
Account active status comparison between the two telecom table data used.

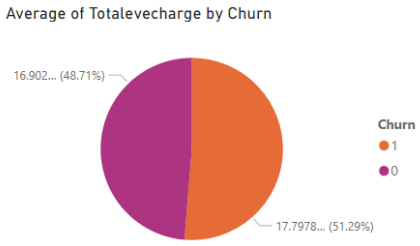


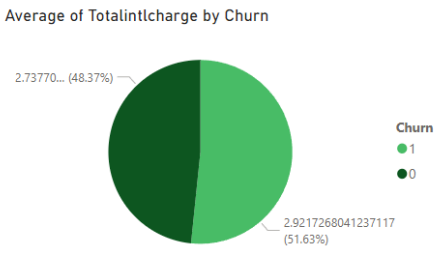
comparison of the renewal (vs) the churn for the 2nd table used (telecom3)

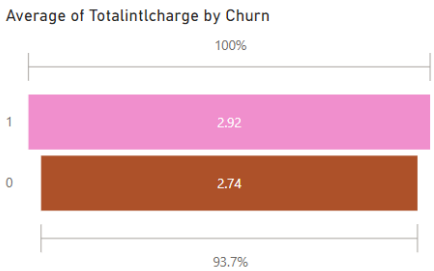


Data used by the minutes spent on roaming (international) calls for the 1st Dataset (Bi-variate)

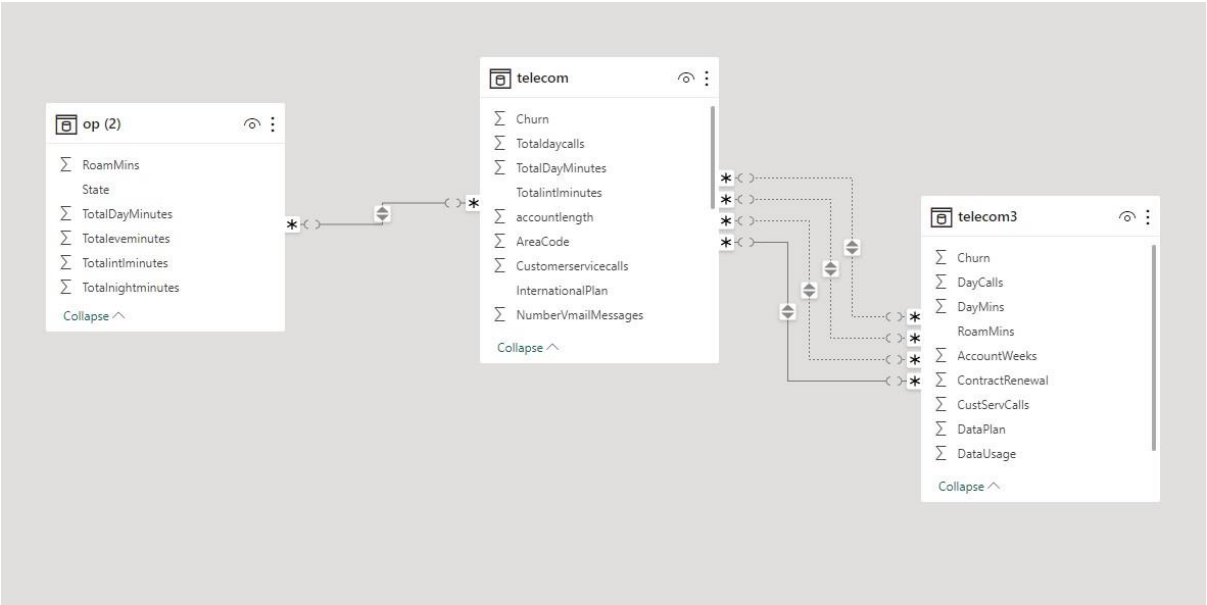








Churn data at all the given time in a day for the 1st table (telecom)



Model view of the relationships between tables (op(2)-telecom=> RoamMins – TotalIntlMinutes, telecom-telecom3 => TotalDayCalls-DayCalls, TotalDayMinutes-DayMins, Churn-Churn

Process undertaken to achieve this:

The provided datasets, namely "telecom" and "telecom3," underwent a comprehensive analysis and data cleansing process using the Python programming language. During this phase, various techniques and algorithms were employed to prepare the data for further analysis. Additionally, a new table was created by executing SQL queries, incorporating specific data elements from the original tables.

Once the Python data cleaning and SQL table creation tasks were completed, the entire dataset, comprising the original tables and the newly generated SQL table, was integrated into Power BI for advanced visualization and reporting purposes. Power BI's robust capabilities were harnessed to create informative and insightful representations of the data. The visualizations showcased essential patterns, trends, and relationships within the datasets, enabling stakeholders to make data-driven

decisions and gain valuable business insights. The iterative process of data transformation, SQL query execution, and Power BI visualization fostered a thorough understanding of the data's nuances and uncovered meaningful information that contributed to a comprehensive project documentation.

What were the conclusions derived from observing both the datasets?

The data shows that they are of the of the same distribution type and all the data acts same once log transformation is applied.
The charges, however, are not like one another, they differ depending on their own values.
Only 10% (exactly 10.127532%) users across all states have international plans, the rest do not.
The observation ultimately leads us to believe that most of the money is made from day calls.
The churn data for the various given points in a show that customer churn rate is more than retention rate. Hence, the company must work on increasing their customer count.

What are actionable insights that have been gathered from this?

Charge Disparity: To address the disparity in charges, further investigate the factors influencing these variations. Perform a detailed analysis to identify potential reasons for the differences in charges, such as usage patterns, customer segments, or pricing structures. Tailor marketing and pricing strategies to accommodate diverse customer needs.

International Plans: Given that only about 10% of users have international plans, explore opportunities to promote and incentivize international plan subscriptions. Analyse the reasons behind the low adoption rate, such as pricing, benefits, or marketing reach, and devise strategies to increase its popularity among customers.

Revenue from Day Calls: If day calls are indeed generating the most revenue, focus on optimizing and enhancing day call services. Consider offering attractive packages or features to encourage more customers to engage in day calls, thus further boosting revenue generation.

Churn and Retention: Examine the reasons behind the higher churn rate compared to the retention rate. Identify common factors leading to customer attrition and design targeted retention strategies. These could include personalized offers, improved customer service, loyalty programs, or enhanced service quality.

Customer Acquisition: While increasing customer count is crucial, a balanced approach should be adopted to ensure that efforts are also directed towards retaining existing customers. Acquiring new customers can be expensive, so investing in customer retention strategies can lead to long-term profitability.

Market Segmentation: Consider segmenting customers based on their usage patterns, preferences, and needs. This will help in tailoring marketing strategies and service offerings to specific customer segments, resulting in higher customer satisfaction and reduced churn.

Feedback and Surveys: Regularly collect feedback from customers to understand their experiences, pain points, and expectations. Conducting surveys and analysing customer feedback can provide valuable insights for continuous improvement and decision-making.

Competitor Analysis: Monitor the actions and offerings of competitors in the market. Understanding the competitive landscape can help in identifying gaps in the company's services and formulating effective strategies to stay ahead.

In conclusion, a comprehensive solution involves a data-driven approach to understand the underlying factors influencing charges, churn rates, and customer behaviour. By implementing the suggested strategies and continuously analysing data, the company can make informed decisions to improve customer retention, increase revenue, and enhance overall business performance.