**Assignment 8 - Analysis of Telecom Customer Churn using Spark**

Customer churn is a significant concern for telecommunications companies, impacting their profitability and long-term customer retention. In this assignment, you will use Spark SQL to analyse a dataset containing detailed information about telecom customers, their subscription plans, and various service attributes. The dataset includes critical features such as contract types, monthly charges, tenure, and whether the customer has churned (left the service). This analysis of customer churn aims to provide actionable insights that can help reduce churn and improve customer satisfaction.

**Technology: Spark, Spark SQL**

**Dataset:** Telco\_Customer\_Churn.csv

Dataset Description

| **Column Name** | **Description** |
| --- | --- |
| customerID | Unique identifier for each customer |
| gender | Gender of the customer (Male, Female) |
| SeniorCitizen | Indicates if the customer is a senior citizen (0, 1) |
| Partner | Whether the customer has a partner (Yes, No) |
| Dependents | Whether the customer has dependents (Yes, No) |
| tenure | Number of months the customer has stayed with the company |
| PhoneService | Whether the customer has phone service (Yes, No) |
| MultipleLines | Whether the customer has multiple phone lines (Yes, No) |
| InternetService | Type of internet service (DSL, Fiber optic, No) |
| OnlineSecurity | Whether the customer has online security service (Yes, No) |
| OnlineBackup | Whether the customer has online backup service (Yes, No) |
| DeviceProtection | Whether the customer has device protection service (Yes, No) |
| TechSupport | Whether the customer has technical support service (Yes, No) |
| StreamingTV | Whether the customer has streaming TV service (Yes, No) |
| StreamingMovies | Whether the customer has streaming movies service (Yes, No) |
| Contract | Type of contract (Month-to-month, One year, Two year) |
| PaperlessBilling | Whether the customer has opted for paperless billing (Yes, No) |
| PaymentMethod | Payment method used by the customer (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)) |
| MonthlyCharges | Monthly charges billed to the customer |
| TotalCharges | Total charges billed to the customer over the tenure period |
| Churn | Whether the customer has churned (Yes, No) |

Write Spark SQL queries to show the following analysis with Visualization.

1. Analyze how customer retention varies based on how long the customer has stayed with the company (tenure).
2. Investigate the churn rate of customers who subscribe to streaming services like StreamingTV and StreamingMovies.
3. Write Spark SQL to group customers by their tenure (e.g., 0-12 months, 13-24 months, etc.) and analyze churn rates in these tenure groups.
4. Analyze the impact of contract types and payment methods on churn rates.
5. Explore the distribution of monthly charges for customers based on their type of internet service.
6. Identify the top 10 customers who have contributed the most revenue to the company, based on total charges.
7. Calculate the churn rate segmented by gender and whether the customer is a senior citizen.
8. Write a query to calculate Correlation between dependents and churn. Explore whether having dependents affects customer churn rates.
9. Predict potential churn rates by analyzing the relationship between monthly charges, contract types, and the churn rate.
10. Determine the churn rate for customers who have multiple services (Phone, Internet, and Streaming), which can help understand whether bundling services leads to higher or lower churn. Calculate churn rate for customers with multiple services.
11. Churn Impact by device protection and online backup services. Write a query to investigate whether having device protection or online backup services affects churn rates.
12. Explore churn rates among customers who do not have phone service and investigate if it influences customer retention.
13. Understand the relationship between payment methods and contract types on customer churn. This query will help you discover which combinations are most prone to churn.
14. Analyze how customer churn is affected by senior citizen status and whether the customer has dependents.
15. Explore whether subscribing to streaming services like Streaming TV and Streaming Movies influences the churn rate.
16. Understand how tenure and MonthlyCharges differ between churned and non-churned customers. This can provide insights into the behavior of long-term customers.
17. Compare monthly charges and churn rates between newer customers and long-time customers.
18. What is the correlation between senior citizen status and churn rate?
19. Partition customers based on whether they are senior citizens and divide them into 5 groups based on tenure. [Use NTILE.]
20. Use PERCENT\_RANK to identify the top 5% of customers by MonthlyCharges.
21. Find customers who fall within the top 5% of the distribution based on monthly charges. Compare total charges with the next customer in the same internet service type, based on monthly charges.
22. Find the top 5 customers with the highest MonthlyCharges within each Contract type.
23. Calculate the churn rate in each Contract type and rank the contracts by churn rate.
24. **Perform an in-depth analysis of customers using window functions to understand customer rankings, distribution, and trends in charges and tenure.**