SQL Portfolio Project Report

Background

Customer churn, or attrition, is a critical concern for businesses, particularly in the telecommunications industry, where retaining subscribers directly impacts revenue and growth potential. High churn rates can lead to financial instability, prompting companies to invest heavily in marketing efforts to attract new customers. Understanding the factors contributing to churn is essential for developing effective retention strategies. By analyzing customer behavior and satisfaction levels, companies can identify areas for improvement, such as service quality and customer support, ultimately enhancing customer loyalty and reducing attrition.

Objective

The primary objective of this analysis is to explore the **IBM Telco Customer Churn** dataset to identify patterns and insights related to customer churn. By examining demographic information, service usage, and churn status, the analysis aims to uncover actionable insights that can inform targeted retention strategies. Additionally, the project seeks to answer hypothetical questions regarding the factors influencing churn, allowing the telco company to enhance its customer retention efforts.

Dataset Overview

The dataset used for this analysis is derived from the IBM Telco Customer Churn dataset, which provides a comprehensive view of customer interactions and behaviors. It consists of five tables: **telco_demographics**, which captures customer information such as age and gender; **telco_location**, detailing customer geographical data; **telco_population**, containing population statistics relevant to customer locations; **telco_services**, which includes information on the services subscribed to by customers; and**telco_status**, indicating the churn status of each customer. This structured dataset allows for a thorough examination of the relationships between customer demographics, service usage, and churn behavior.

Query 1:

Considering the top 5 groups with the highest average monthly charges among churned customers, how can personalized offers be tailored based on age, gender, and contract type to potentially improve customer retention rates?

```
SELECT d.Age, d.Gender, s.Contract, AVG(s.`Monthly Charge`) AS Avg_Monthly_Charge
FROM Telco_demographics AS d
JOIN Telco_services AS s
ON d.`Customer ID` = s.`Customer ID`
JOIN Telco_status AS st
ON d.`Customer ID` = st.`Customer ID`
WHERE st.`Churn Label` = 'Yes'
GROUP BY d.Age, d.Gender, s.Contract
ORDER BY Avg_Monthly_Charge DESC
LIMIT 5;
```

Explanation:

The query uses inner joins to connect customer demographic data from the **Telco_demographics** table with service information from the **Telco_services** table using the **Customer ID**. This ensures that only records for the same customer appear in the results. It then performs another inner join with the **Telco_status** table to bring in the churn status, filtering the dataset to include only those customers who have churned, indicated by a **Churn Label** of **Yes**. After joining the necessary data, the **WHERE** clause filters out the non-churned customers. In contrast, the **GROUP BY** clause organizes the results by age, gender, and contract type to calculate the average monthly charge for each group. Finally, the results are sorted in descending order of average monthly charge, and the **LIMIT 5** clause ensures that only the top five groups with the highest charges are displayed.

Result:

	Age	Gender	Contract	Avg_Monthly_Charge
▶	48	Male	Two Year	113.15
	66	Male	One Year	111.4
	69	Male	One Year	110.9
	43	Male	One Year	110.15
	25	Male	Two Year	110.1

Query 2:

What are the feedback or complaints from churned customers?

```
SELECT st.`Churn Reason`, COUNT(st.`Customer ID`) AS Number_of_Churns
FROM Telco_status AS st
WHERE st.`Churn Label` = 'Yes'
GROUP BY st.`Churn Reason`
ORDER BY Number_of_Churns DESC;
```

Explanation:

This query retrieves the reasons behind customer churn and counts how many customers left for each reason, focusing on the **Telco_status** table. It filters the data to include only churned customers (those with a **Churn Label** of **Yes**) and groups them by **Churn Reason**. The **COUNT** function calculates the number of churned customers associated with each reason, and the results are ordered in descending order, showing the most common churn reasons first. This allows the company to identify key drivers of churn and take targeted actions, such as addressing specific issues like billing or service dissatisfaction, to improve customer retention.

Result:

	Churn Reason	Number_of_Churns
	Attitude of support person	220
	Don't know	130
	Competitor offered more data	117
	Competitor offered higher download speeds	100
	Attitude of service provider	94
	Price too high	78
	Product dissatisfaction	77
	Network reliability	72
	Long distance charges	64
	Service dissatisfaction	63
	Moved	46
	Extra data charges	39
	Limited range of services	37
	Poor expertise of online support	31
	Lack of affordable download/upload speed	30
	Lack of self-service on Website	29
	Poor expertise of phone support	12
	Deceased	6

Query 3:

How does the payment method influence churn behavior?

```
SELECT s.`Payment Method`, COUNT(st.`Customer ID`) AS Total_Customers,
    SUM(CASE WHEN st.`Churn Label` = 'Yes' THEN 1 ELSE 0 END) AS `Churned Customers`,
    (SUM(CASE WHEN st.`Churn Label` = 'Yes' THEN 1 ELSE 0 END) / COUNT(st.`Customer ID`)) * 100 AS
`Churn Rate`
FROM Telco_services AS s
JOIN Telco_status AS st ON s.`Customer ID` = st.`Customer ID`
GROUP BY s.`Payment Method`
ORDER BY `Churn Rate` DESC;
```

Explanation:

This query is designed to analyze how different payment methods impact customer churn. It joins the **Telco_services** and **Telco_status** tables using the **Customer ID**, linking service information with churn status. For each payment method, it counts the total number of customers and calculates the number of churned customers using a **CASE** statement, which assigns a value of 1 for churned customers (**Churn Label** = **Yes**) and 0 otherwise. The query also computes the churn rate by dividing the number of churned customers by the total number of customers for each payment method and multiplying it by 100 to express the rate as a percentage. Finally, it groups the results by payment method and orders them in descending order by churn rate, allowing the company to identify which payment methods are associated with higher churn rates and potentially investigate why certain methods correlate with more customer losses.

Result:

Payment Method	Total_Customers	Churned Customers	Churn Rate
Mailed Check	385	142	36.8831
Bank Withdrawal	3909	1329	33.9985
Credit Card	2749	398	14.4780

Medium blog link: https://medium.com/@wasifmaqsood1999/understanding-customer-churn-insights-and-strategies-from-the-ibm-telco-dataset-5ef84ef60415

GitHub Link: https://github.com/wasifmaqsood/IBM-Telco-Customer-Churn-Project-MySql