Telecom Churn Data Driven Storyteller

Analytics Application Project

Prescriptive Analytics

Abadila, Frances Mari

Bayani, Gabriel

Español, Mazi Love

Malilay, Christopher Andre

S.Y. 2024-2025

Table of Contents

1.1 Project Statement

The Telecom Churn Data-Driven Storyteller project aims to leverage data analytics and visualization methods learned from previous courses offered by the CICS Information Systems to transform customer churn data into actionable insights for the telecom sector. This project seeks to enhance understanding of the key variables influencing voluntary customer attrition, identify critical trends and patterns, and provide telecom firms with the knowledge necessary to implement targeted retention strategies based on customer profiles.

In addition, the project will focus on the early detection of potential voluntary churners through predictive modeling, enabling businesses to take preemptive actions to retain customers. By analyzing customer behavior and key indicators, telecom companies will be better equipped to arrest rising voluntary churn rates, allowing for the implementation of proactive solutions to reduce attrition.

This data-driven approach will empower telecommunication providers to make informed decisions that foster business growth, minimize potential losses, and strengthen customer loyalty.

1.2 Scope and Objectives

This section will explain what the Telecom Churn Data Driven Storyteller project aims to achieve.

1.2.1 Scope

The scope of the Telecom Churn Data Driven Storyteller project encompasses the following key aspects:

1. Data Analysis: Examining customer data to identify trends and patterns related to churn.
2. Churn Factors: Discovering key reasons why customers decide to leave telecom services.
3. Insight Generation: Creating clear and actionable insights from the data to help understand and address churn.
4. Visualization: Developing easy-to-understand visualizations and stories to communicate findings effectively to stakeholders.
5. Recommendations: Offering practical recommendations for telecom companies to identify which customers are more likely to churn based on data analysis.

1.2.2 General Objective

The Telecom Churn Data Driven Storyteller project aims to use customer data to understand why people leave telecom services. It will help telecom companies identify the primary reasons why customers leave, take action to keep their customers longer, and overall improve their services.

1.2.3 Specific Objectives

1. **Analysis of Historical Data**: To detect common patterns and trends associated with customer churn by analyzing historical customer data.
2. **Identification of Key Factors**: To pinpoint the main factors and behaviors leading to customer churn through data storytelling.

3. **Explanation of Churn Reasons**: To explain why customers leave and provide actionable insights to address these issues.
4. **Creation of Visualizations**: To design user-friendly visualizations and narratives that effectively communicate findings to stakeholders.
5. **Providing Recommendations**: To offer practical strategies and recommendations to help telecom companies improve customer retention efforts based on the analysis.

1.3 Project Plan

The table below presents the project plan for creating a data-driven storytelling model focused on Telecom Churn. It highlights the essential milestones and tasks needed to complete the project, covering everything from defining the scope to delivering the final narrative.

Table 1.3.1 Project Plan

Task	Priority	Owner	Start date	End date	Duration	Milestone
Create Project Statement	P0	CHRISTOP...	8/26/2024	8/31/2024	5	Complete and approve the Project Statement document.
Define Scope and Objectives	P0	CHRISTOP...	8/26/2024	8/31/2024	5	Finalize and obtain sign-off on the project scope and objectives.
Project Plan	P0	FRANCES ...	8/26/2024	8/31/2024	5	Develop, review, and approve the detailed Project Plan.
Data Collection	P1	GABRIEL B...	9/1/2024	9/4/2024	3	Complete data collection from all identified

						sources.
Data Cleansing	P1	MAZI LOV…	9/1/2024	9/4/2024	3	Finish cleaning and preprocessing the collected data.
Data Analysis	P2	FRANCES …	9/5/2024	9/10/2024	5	Complete initial data analysis and identification of key insights
Data Visualization	P2	CHRISTOP…	9/5/2024	9/10/2024	5	Develop and validate initial data visualizations.
Develop a Narrative	P2	GABRIEL B…	9/5/2024	9/10/2024	5	Draft and finalize the data narrative integrating insights and visualizations.
Key Stakeholders	P0	MAZI LOV…	8/26/2024	8/31/2024	5	Identify, engage, and document all key stakeholders.
Relevant Key Performance Indicators	P0	GABRIEL B…	8/26/2024	8/31/2024	5	Define and agree on KPIs for evaluating project success.
Defined Criteria for Success	P0	GABRIEL B…	8/26/2024	8/31/2024	5	Establish and formalize criteria for project success.
Data Warehousing	P2	FRANCES …	9/5/2024	9/10/2024	5	Implement data warehousing solutions and integrate data.
Data Mining	P2	CHRISTOP…	9/5/2024	9/10/2024	5	Complete data mining to extract relevant patterns and insights.
Build the Model	P4	GABRIEL B…	9/10/2024	10/1/2024	20	Develop and validate the data storytelling model.

Build the Model	P4	FRANCES ...	9/10/2024	10/1/2024	20	Develop and validate the data storytelling model.
Build the Model	P4	CHRISTOP...	9/10/2024	10/1/2024	20	Develop and validate the data storytelling model.
Visualize Data	P4	MAZI LOV...	10/2/2024	10/9/2024	7	Finalize and approve all data visualizations.
Review and Refine	P4	GABRIEL B...	10/2/2024	10/9/2024	7	Complete review and refinement of the model, narrative, and visualizations.
Deliver the Story	P4	CHRISTOP...	10/2/2024	10/9/2024	7	Present and deliver the final data story in all required formats.
Evaluate Impact	P4	FRANCES ...	10/2/2024	10/9/2024	7	Assess the impact and effectiveness of the storytelling against KPIs and success criteria.
Detailed Data Dictionary	P0	GABRIEL B...	8/26/2024	8/31/2024	5	Complete and approve the Data Dictionary for reference and consistency.
Comprehensive List of Data Sources	P0	MAZI LOV...	8/26/2024	8/31/2024	5	Compile and validate the Comprehensive List of Data Sources.
Utilized Tools	P0	FRANCES ...	8/26/2024	8/31/2024	5	Document and review all tools utilized throughout the project.

1.4 Stakeholders

The table below lists the stakeholders for the Telecom Churn data-driven storytelling model, detailing their type, title, role, and whether they are internal or external. This summary helps ensure clear communication and effective engagement with all key participants.

Table 1.4.1 Stakeholder Identification

Stakeholder/ Group	Type	Stakeholder Title	Role	External/ Internal
Executive Officers	Group	Officers	Functional, Approval	Internal
Board of Directors	Group	Directors	Functional, Approval	Internal
Project Manager	Individual	Manager	Functional	Internal
Team Members	Group	Members	Functional	Internal
Marketing Team	Group	Team	Informational	Internal
Data Analysts	Individual	Analysts	Functional	Internal
IT Department	Group	Team	Functional	Internal
Customers	Group	Customer Stakeholder	Informational	External
Auditor	Group	Auditor	Functional	External

1.5 Key Performance Indicator

This section serves as a guide for the proponents to achieve the project output.

1.5.1 Project Output

The proponents intend to produce a data pipeline that contains the necessary procedures and models. The data pipeline to be produced is for Churn analysis however it must be emphasized that it is "data ambiguous" or general meaning it is not specific for the company or used data set only.

With the data pipeline, the proponents will be able to provide the necessary insights to the company about the retention or churning of customers using their services. It will also allow the proponents to provide prescriptions on what things to lessen or focus on to reduce the leaving of customers and attract or retain customers.

1.5.2 Criteria for Success

The proponents created the following criteria to support the objectives of the project in sections 1.2.3 and 1.2.4. The criteria below will serve as a guide to ensure that the project will produce the necessary results that will make it relevant to Churn Analysis even after the project.

Table 1.5.2.1 Success Criteria Table

Criteria	Description
Straightforward Data Pipeline	- The Data Pipeline contains the necessary methods starting from Data Gathering to Data Storytelling - Methods or models used are further explained by supporting statements - All of the methods have been properly utilized

Flexible Data Pipeline	- The Data Pipeline is not exclusive to a specific dataset - Models are retrainable to cater to newer datasets
Robust Data Warehouse	- Data Warehouse has the necessary ETL functions - Data mapping matches relevant data - A consolidated and complete fact table is produced - Data is preserved all throughout the process
Accurate Model Results	- The test of Accuracies done to models is within the acceptable range of 70% and above
Identifiable Causes of Churn	- Proponents and stakeholders can identify the reason why customers churn through data-driven analytics. - Coherent and sensible data storytelling is made possible
Relevant Retention Prescriptions	- Proponents can propose retention strategies in line with the data. - Stakeholders can make sense of the prescriptions given the information provided,

1.6 List of Data Sources

The project primarily utilizes an External Data Source namely from Kaggle. The data set used is publicly available and provided by Shi Long Zhuang which contains customer churn data from a telecommunications company based in California. The data acquired is from a Non-Government Publication given that it is not taken from a government source,

Churn Telecom businesses prioritize acquiring new customers atop sustaining existing clients. Prior to that, understanding customer behavior post-acquisition is essential for effectively managing churn and improving retention strategies.

Customers provided very detailed information may be a useful tool for training customer churn models and providing insightful data.

Understanding customer behavior is crucial. Monitoring Payment Timeliness through metrics like Monthly Charge and Total Charges can reveal whether customers are paying on time or experiencing delays, which may signal an increased churn risk. To enhance churn prediction, Value-Based Segmentation is crucial by analyzing Total Revenue, Contract, and Monthly Charge, high-value customers can be identified and assess their risk of churn. Computer learning is essential for absorbing information from the past and having crucial knowledge at hand to enhance future encounters.

- Utilizing Micro Surveys
- Operating Databox, Spreadsheets, Churnly, and such.

1.7 Data Collection Report

1.7.1 Data Warehousing

1.7.2 Data Mining

1.7.3 Data Collection

1.7.4 Data Cleansing

1.8 Data Dictionary

The Data Dictionary provides context on the collected data by describing the fields present.

Table 1.8.1: Data Dictionary

Field	Description
CustomerID	Unique identification of customer
Gender	Male/Female
Age	Current Customer age in years when Q2 2022 Ended
Married	Marriage indication (Yes/No)
Number of Dependents	Number of dependents that live with the customer
City	The customer's Primary residence is in California
Zip Code	Primary residence Zip Code
Latitude	Latitudinal coordinates of primary residence
Longitude	Longitudinal coordinates of Primary Residence
Number of Referrals	Number of times customers were referred to by an external entity
Tenure in Months	Total months the customer has been subscribing to the company until the end of Q2
Offer	Last marketing offer accepted by the customer (None, A,B,C)
Phone Service	Indicates subscription to a phone service (Yes/No)
Avg Monthly Long Distance Charges	Average long-distance charges of the customer calculated until the end of the quarter (0 if not subscribed to phone service)

Multiple Lines	Indicates subscription to multiple telephone lines with the company (Yes/No. If not subscribed to phone service then-No)
Internet Service	Indicate subscription to internet service with company (Yes/No)
Internet Type	Type of internet connection (None, DSL, Fiber Optic, Cable)
Avg Monthly GB Download	Average download volume of the customer in Gigabytes calculated until the end of the quarter (0 if not subscribed to internet service)
Online Security	Indicates if the customer subscribes to additional online security service (Yes/No. If not subscribed to internet service then-No)
Online Backup	Indicates if the customer subscribes to additional backup service (Yes/No. If not subscribed to internet service then-No)
Device Protection Plan	Indicates if the customer subscribes to additional device protection service for their equipment (Yes/No. If not subscribed to internet service then-No)
Premium Tech Support	Indicates if the customer subscribes to a premium technical support plan (Yes/No. If not subscribed to internet service then-No)
Streaming TV	Indicates if the customer streams TV programs from 3rd party provider with no additional fee while using internet service (Yes/No. If not subscribed to internet service then-No)
Streaming Movies	Indicates if the customer streams TV programs from 3rd party provider with no additional fee

	while using internet service (Yes/No. If not subscribed to internet service then-No)
Streaming Music	Indicates if the customer streams music from 3rd party provider with no additional fee while using internet service (Yes/No. If not subscribed to internet service then-No)
Unlimited Data	Indicates if the customer pays an additional fee for unlimited downloads/uploads (Yes/No. If not subscribed to internet service then-No)
Contract	Current contract type of customer (Month-to-month, 1 year, 2 year and related to the service availed whether Internet or Telephone)
Paperless Billing	Paperless billing was chosen (Yes/no)
Payment Method	How bills are being paid (Bank withdrawal, credit card, mailed check)
Monthly Charge	The current total monthly charge for all services
Total Charges	Total charges calculated until the end of Q2
Total Refunds	Total refunds calculated until the end of Q2
Total Extra Data Charges	Total charges for extra downloads above the specified plan by the end of Q2
Total Long Distance Charges	Total charges for long distances above the specified plan by the end of Q2
Total Revenue	Total revenue from customer calculator to the end of Q2 (Total Charges - Total Refurnds + Total Extra Data Charges + Total Lond Distance Charges)

Customer Status	Status of a customer at the end of the Quarter (Churned, Stayed or Joined)
Churn Category	High-level category reason for churning- why the customer left the company and is directly related to Churn Reason (Attitude, Competitor, Dissatisfaction, Other, Price)
Churn Reason	The specific reason why a customer left the company (Related to Churn Category, Qualitatively stated)
Zip Code	Primary residence zip code
Population	Population estimate for zip code area

1.9 Utilized Tools

1. Visual Studio Code

   Visual Studio Code allows interactive development and execution of Phyton scripts which are commonly used for data analysis and visualization. It enables users to write and manage code for data analytics, statistical modelling, and database queries.

2. Rapid Miner

   Rapid Miner is a data science platform that enables to transform, build, and deploy of data and machine learning pipelines. It provides various data mining and machine learning procedures.

3. DBeaver

   DBeaver is a free and open-source universal database tool. It accommodates various database types, such as SQL, NoSQL, and cloud databases, offering a versatile solution for database administrators and developers.

4. Phyton

   Python is a flexible, high-level programming language known for its readability and extensive use in various fields such as web development, data analysis, and scientific computing. It will be used for preprocessing analysis and the development of machine learning models.

https://www.questionpro.com/blog/sources-of-data/