# **ANALYTICS PLAN**

Submitted by: Pushkar Mishra, Student ID: 300980265.

# Analytics Startup Plan

**Synopsis: *This document provides a high-level walkthrough of the activities required to guide the completion of the analysis.***

|  |  |
| --- | --- |
| **Project** | ***Predict the “CHURN” of customers for a hypothetical telecommunication company.*** |
| **Requestor** | *Academic Capstone Project work from Centennial College, Progress Avenue Campus, Scarborough, Ontario, Canada.* |
| **Date of Request** | *17th July 2023.* |
| **Target Quarter for Delivery** | *Mid-3rd Quarter.* |
| **Epic Link(s)** | *Data sources and other links will be listed in the appendix of the final presentation since it is an ongoing project work.* |
| **Business Impact** | *Will be used for personal career growth.* |

## 1.0 Business Opportunity Brief

|  |  |
| --- | --- |
|  | Clearly articulated business statement of the Ask, opportunity, or problem you are trying to solve. An important step is to understand the nature of the business, system or process and the desired problems to be addressed. This will be communicated back to All stakeholders for alignment. |

“PROFIT” is the keyword for any business, which is maximized by the business’s customers’ retainership. “CHURN” is just the opposite of “RETAINERS.” Hence any business always has been historically inclined to keep its customers with them for as long as they possibly can to maximize the eventual profits, thereby giving its shareholders a better dividend, and increasing their stock prices.

However, the telecommunication industry over the course of the last 15 years in North America, since the advent of 3G services, and in a constant upgradation mode of the services, has faced a unique challenge to retain their customers because the customers are not just paying for services, but for an ever-evolving physical handset, getting more and more diverse with included and supported features. The newer, advanced handsets need compatible cellular services. In addition, the ever-increasing speed of the internet, and its multiverse associated functionality, need telecommunication companies worldwide to stay on top of the technology if they want to stay competitive. The third aspect is cable television and its changing delivery modes to the end user. These three services have become the core services for a telecommunication company, with the fourth avenue of home phones becoming more and more redundant with every passing day.

Internet-based Smart home monitoring has become the latest entry as far as services are concerned for a telco (Abbreviated for telecommunication company, and will be used in future references.)

**The specific ask:**

*The specific task for this data science project is to predict the churn for customers for a hypothetical telco, based on the features and observations provided via multiple files, capturing different aspects of a telco.*

## 1.1 Supporting Insights

|  |  |
| --- | --- |
|  | Define any supporting insights, trends and research findings. Where relevant, list key competitors in the market. What are their key messages, products & services? What is their share of market, nationally and regionally? |

## Churn Data as per Statista, obtained as a search from Bing, not directly from Statista (Being a Subscription based information) for the three Top cellular service providers in Canada is as follows, captured for 6 fiscal years, from 2015 to 2020, and is blended for both the “Prepaid” and “Postpaid” customers: (It’s assumed that the terms “Prepaid” and “Postpaid” are self-explanatory).

**ROGERS:**

|  |  |
| --- | --- |
| **Year** | **Blended Churn Percentage** |
| **FY2015** | **1.34%** |
| **FY2016** | **1.23%** |
| **FY2017** | **1.48%** |
| **FY2018** | **1.23%** |
| **FY2019** | **1.23%** |
| **FY2020** | **0.97%** |

**BELL:**

|  |  |
| --- | --- |
| **Year** | **Blended Churn Percentage** |
| **FY2015** | **1.4%** |
| **FY2016** | **1.4%** |
| **FY2017** | **1.4%** |
| **FY2018** | **1.3%** |
| **FY2019** | **1.3%** |
| **FY2020** | **1.2%** |

**TELUS:**

|  |  |
| --- | --- |
| **Year** | **Blended Churn Percentage** |
| **FY2015** | **0.94**% |
| **FY2016** | **0.93**% |
| **FY2017** | **0.99**% |
| **FY2018** | **0.90**% |
| **FY2019** | **0.89**% |
| **FY2020** | **0.86**% |

## 1.2 Project Gains

|  |  |
| --- | --- |
|  | *Describe any revenue gains, quality improvements, cost and time savings (as applicable). What will you do differently and why would our customers care. What are the implications if we do nothing? This section is particularly key for prioritization against company goals and KPI’s.* |

This data science project is an Academic Capstone Project work for Centennial College, Progress Avenue Campus, Scarborough, Ontario, Canada, and has no monetary value. Neither would it be used to present to anyone for any monetary gains, nor it has any outcome that could be assimilated into any ongoing project for any real company for quality improvements, cost, or time savings.

However, this entire project and all its components will be used for personal career gains and has the basic intent to be uploaded to various platforms on the internet, in addition to the personal resume. As a caveat, it could be stated that if a company is inclined to use it for its gains, then the project could be negotiated with that company.

## *Note: Completion of the following sections is possible only after a careful assessment and triage of the Ask. This is required to determine scope, resource, time, priority and data availability.*

## 2.0 Analytics Objective

|  |  |
| --- | --- |
|  | List the key questions, assumptions and define the hypotheses. Often the deliverable may not just be an analysis output, however a recommended operating model or blueprint for a pilot etc.  Note: Asking the right questions and truly understanding the problem will lead to the right data, right mathematics, and right techniques to be employed. |

## The core objective of this project is to identify factors for reducing churn rate, identify key churn drivers, and predict customer churn.

## 2.1 Other related questions and Assumptions:

|  |  |
| --- | --- |
|  | *List any assumptions that may affect the analysis* |

Since the dataset to be used is originally authored by a very renowned technology company, viz. IBM, and is downloaded from their website (Source will be updated in the final presentation in the Appendix Section), hence the authenticity of the entire dataset cannot be disputed and will be used in its entirety. Any modifications and deletions will be as and when required on an ad hoc basis, with no additions during the project to make the data corrupt. There could be some calculated fields, extracted from the given dataset, which once again would be as and when required.

## 2.2 Success measures/metrics

|  |
| --- |
| *What does success look like? Define the key performance indicators (success definition/indicators, drivers and key metrics) against which the objectives will be analyzed. These should be drawn from the interlock meeting with key stakeholders and will inform the approach and methodology for the analysis.* |

## If this project was to be implemented in a real-world scenario, then the success metrics would have been a decreased churn for the company, as compared to previous years, based on the dataset provided and assuming the data structure and observations hold true in the scenario for that telco.

## 2.3 Methodology and Approach

|  |  |
| --- | --- |
|  | *Now that you have a good understanding of the Ask and deliverable, detail the recommended approach/methodology.* |

**Possible Algorithmic Analysis to be deployed:**

1. Predictive Linear Regression,
2. Predictive Logistic Regression,
3. Some variants of Neural Networks with different hidden layers and directional propagations,
4. Principal Component Analysis,
5. Various Visualizations,
6. Dimension Reduction,
7. Performance Evaluation of various models,
8. Naïve Bayes,
9. Classification Trees,
10. Discriminant Analysis,
11. Various Neural Networks, &,
12. Certain Clustering Models, e.g., KNN.

**Methodology:**

1. Project Overview,
2. Data Collection and Understanding,
3. Exploratory Data Analysis (EDA),
4. Feature Engineering,
5. Data Splitting,
6. Modeling Approach,
7. Model Training and Tuning,
8. Model Evaluation,
9. Interpretability and Insights, &,
10. Communication and Reporting.

**Output:**

The expected output of this project is to identify factors for reducing churn rate, identify key churn drivers, and predict customer churn.

## 3.0 Population, Variable Selection, considerations

|  |  |
| --- | --- |
|  | Capture learning about the data available today location, structure, and reliability; this would include data in operational systems including dealer sourced, data warehouse and any CRM or email marketing systems available today. |

Being a hypothetical dataset, some of the following metrics may not be applicable. The applicable ones are not yet available to be described yet.

**Audience/population selection:**

**Observation window:**

**Inclusions:**

**Exclusions:**

**Data Sources:**

**Audience Level:**

**Variable Selection:**

**Derived Variables:**

**Assumptions and data limitations:**

## 4.0 Dependencies and Risks

|  |  |
| --- | --- |
|  | Identification of key factors that may influence the outcome of the project and likelihood of it happening: |

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Likelihood (based on historical data)** | **Delay (based on historical data)** | **Impact** |

The dataset is not gathered as a real-world scenario, hence there are no dependencies or any risk involved.

## 5.0 Deliverable Timelines

|  |  |
| --- | --- |
|  | List key dates and timelines as a work-back schedule. Activate line items based on complexity and line-of-sight required. Will set the stakeholder expectations for the process. |

|  |  |  |
| --- | --- | --- |
| **Item** | **Major Events / Milestones** | **Date** |
| 1. | Analysis Plan & Data Finalization | July 12 |
| 2. | Data Exploration | July 17 |
| 3. | Modeling | July 24 |
| 4. | Governance | August 7 |
| 5. | Documentation | August 12 |
| 6. | Peer Feedback | August 14 |
| 7. | Presentation | August 13 |
| 8. | Portfolio | August 14-16 |