**Data Collection Plan**

Professor: Dr. Mina Labib

Srilasya Garigipaty

Student ID 100822953

School of Business, IT and Management

DATA 1201: Data Collection

March 1st, 2021

**Table of Contents:**

ASSIGNMENT INTRODUCTION…………………………………………………………2

CONTENTS

1.Summary……………….……………….............................................................................. 4

2.Business Objective…………………………………………….............................................5

3.Data Collection Overview and Plan…………………………………………………………5-13

* Data Source and Purpose………………………………………………………………9-12
* What Data, and Type of Data………………………………………………………… 9-12
* Who and Where of the Data…………………………………………………………. 9-12
* Frequency and How to Display………………………………………………………. 9-12.

4.References………………………………………………………………………………….. 13

**Assignment Introduction:**

The purpose of the assignment is to create a **Data Collection Plan** for an organization of choice. **“Data”** is defined by the Meriam Webster English dictionary as facts of information used to **calculate, analyse, or plan** something**. Data** comes in various forms known as **Data Types. Data Types** can be classified based on the **characteristic** of the data, the **level of measurement**, **time components of the data**, and **the method used to obtain the data**. Data can be classified as **Quantitative** or **Qualitative.** **Quantitative** data has a numerical characteristic, and **Qualitative** data has an attributable characteristic. Data can also be classified as **Discrete** or **Continuous**. A **Discrete** data is data that is distinct or separate in nature whereas a **Continuous** data is interval in nature**. Data source** is the source of the data. It can be a file, a person, a live feed, or a database. There are two types of Data Sources**. Private data sources**, and **Public data sources**. **Private data** is defined as data that is proprietary to an organization and cannot be shared without explicit permission from the organization or individual. **Public data** is information that comes from the government or the public sector. This includes all data collected by public organizations at all levels. The **Data Collection Plan** is a thought-out approach that includes **where to collect the data, how to collect the data, when to collect the data, and who will collect the data.** The following figure illustrates the **Data Collection Plan roadmap**.

**Keywords:** Data Collection Plan, Data, Data types, Quantitative, Qualitative, Discrete, Continuous, Private Data, Public Data, Data Source, Data Collection roadmap



Figure 1.1: Data Collection Roadmap: Six Sigma



**Summary of Business Problem and Objective:**

* **The organization name**: **Tangerine Bank**
* **Tangerine Bank** is a subsidiary of Scotia Bank, and it is an online bank.
* Tangerine, launched in Canada in April 1997, and introduced the country’s first no fee, high interest savings account.
* It was the first bank in Canada to offer the direct banking business model, and the **motivation** for Tangerine Bank was to offer more favorable rate compared to competing banks by avoiding the costs associated with running a network of branches.

**Business Problem Organization is trying to solve:**

* Tangerine Bank has a growing customer base. It offers services such as savings accounts, chequing accounts, credit card services, mutual funds and mortgages. Majority of the customers at the bank are liability customers with deposits of varying sizes.

**Problem Definition:** The **number of customers who are borrowers (asset-based customers) is small**, and the bank is **interested in expanding** **its asset-based customers** to bring in **more loan business, and in the process, earn more revenue for the bank through interest on loans.**

* The retail marketing departmenthas decided to devise campaigns to better target their marketing and **increase the success ratio of loan borrowers and retain them as bank customers.**
* The department wants to build a model that will help identify the potential customers who have a higher probability of purchasing a loan. This will increase success ratio and **reduce the costs of the marketing campaign by only targeting those with more chance of success.**

**Business Objective:** **Service quality** is the top reason customers switch banks. Banks have a lot of customer data, but effective and personalized service is lacking because there is a lack of big data analysis to **learn about customers and their expectations**. Tangerine Bank needs **to employ data analysts** and professionals in the marketing and finance field to **gain insights** from the large amount of data the bank collects. Tangerine Bank plans to **achieve greater customer loyalty** through **personalized offers.** The bank will **utilize the power of big data** to get an understanding of each customer. Through analytics process, Tangerine Bank **will reduce business risk by using predictive analytics and saving costs**.

**Data Collection Sources and Purpose** The data that is required for the collection plan will be gathered from internal and external data sources. The data sources for this project include **the bank’s internal data, social media blogs and company website, and customer service department voice and transcript records.**

The data must be classified into **private data sources** and **public data sources** in order to make informed decisions around the business problem.

**Internal Data (Private Data Sources)**

* Customer details data
* Customer Satisfaction Surveys
* Customer financial records
* Customer Service phone call transcripts
* Marketing Campaign number of contacts with customer, duration of last contact, outcome, and days passed since last contact.

The **unstructured data** that is generated through customer phone transcripts can be used to detect the concerns of the customers, and what sentiments are the most common about issues faced with the bank.

**Process Capability-Shows how closely a process is able to produce the output to its overall specifications.** The internal data selected will provide the analytics team with different variables to determine loan increase strategies as well as customer retention strategies. This can be done by **developing data mining machine learning models** **using Python and R.** **Queries can be done on the internal databases using SQL** to modify, sort, delete, or create relational databases for model building. In addition, **key insights and visualizations can be made through Tableau to be shown to those involved in business marketing campaign goal**.

**Stability**- **Means the process is being steady or in other words that means the process is producing stable output at all times.** The **core banking system is a stable and centralized online real time environment**, that uses centralized databases which allows the bank to generate retrieve data quickly. **The core banking system is constantly assessed, managed, and updated.** The **servers should be highly consistent, and available with partition tolerance to allow for the ability to collect data more often and with stability.**

**Predicting Outcomes-** **The Reliability or Consistency of the model to reproduce. Making sure that the predictive model created will increase marketing campaign outcomes.** The data collection plan and the data collected will help inform the bank’ s marketing team of the customer’s concerns, which customer base to target to increase loan borrowers and retain them and determine relationships or trends between different variables. **The data collection plan will provide an accurate analysis for the bank.** After the data is collected, processed, cleaned, and stored the decision makers will analyze and visualize the results in a presentation and have a discussion of findings.

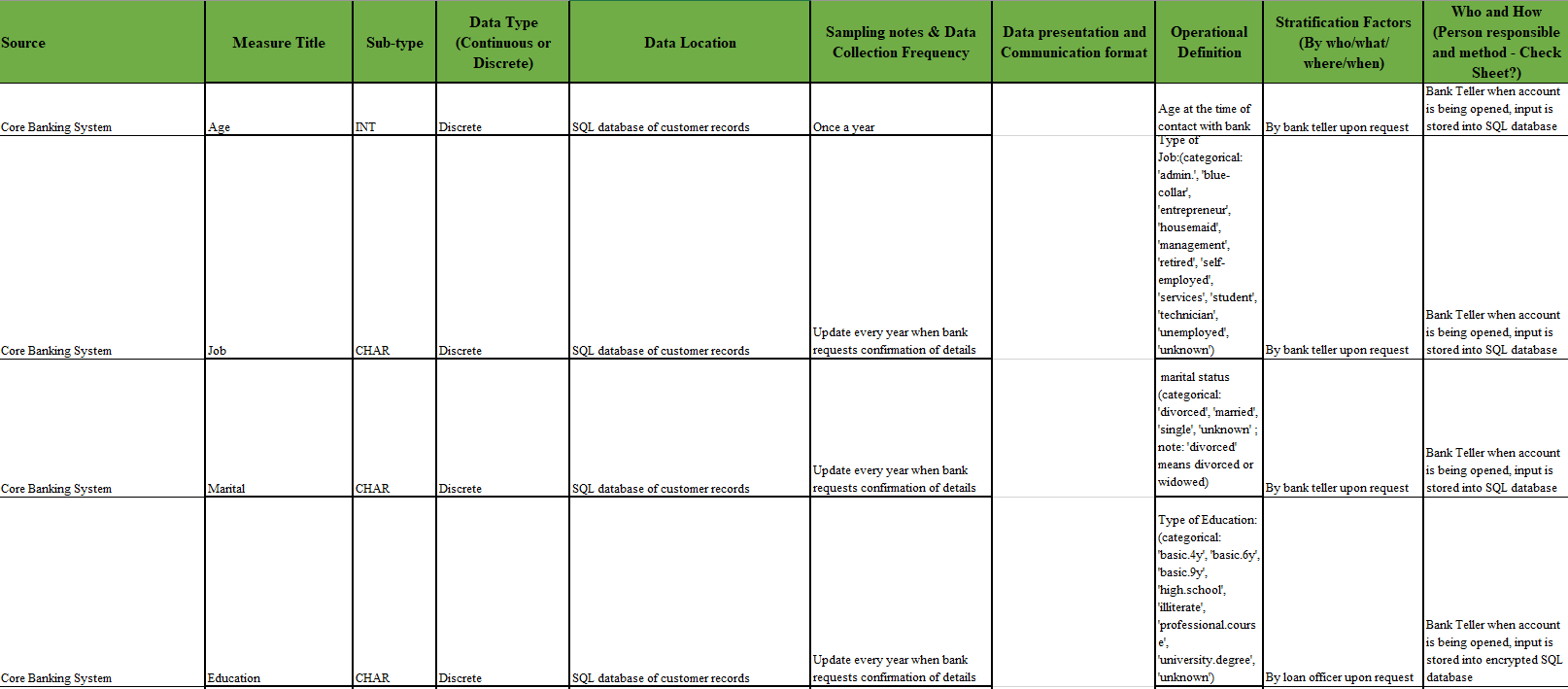
**External Data (Public Data Sources)**

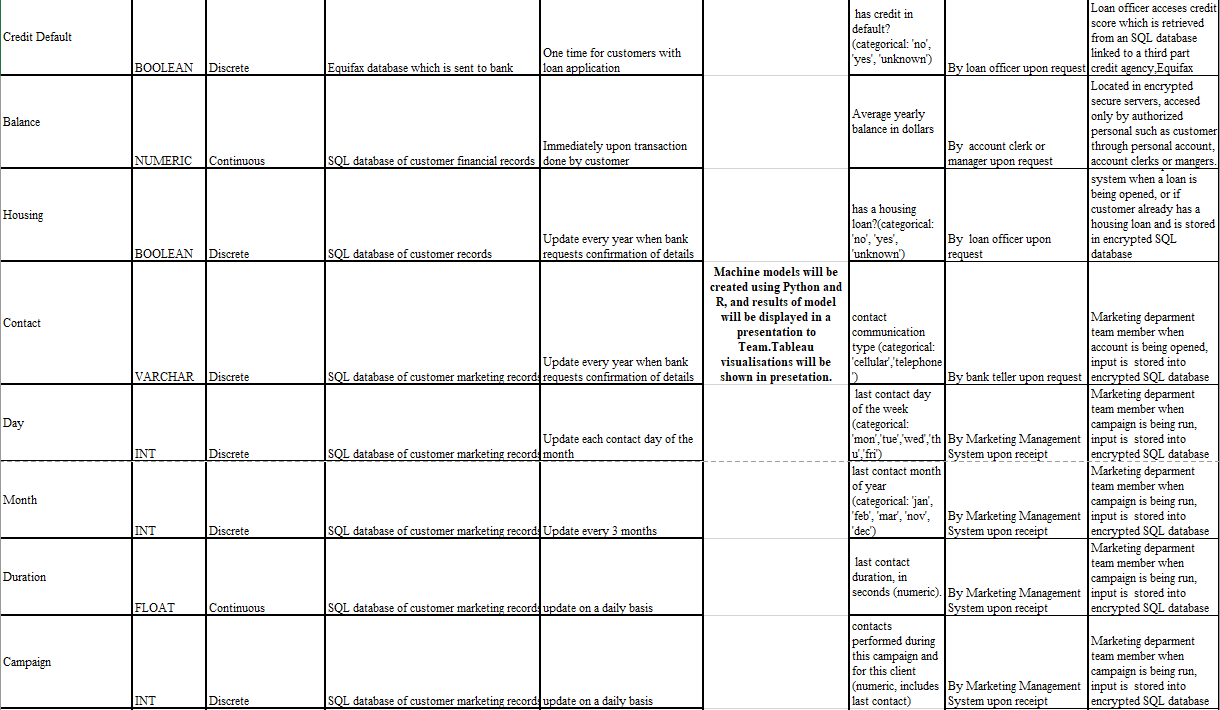
* Article/Blog Engagement from Bank Website, and online bank related blogs such as Tangerine Bank blog, Reddit, Quora, etc. can be used to learn about the feedback, comments, bank- related topics of interest, and customer satisfaction in different areas of service the bank office.

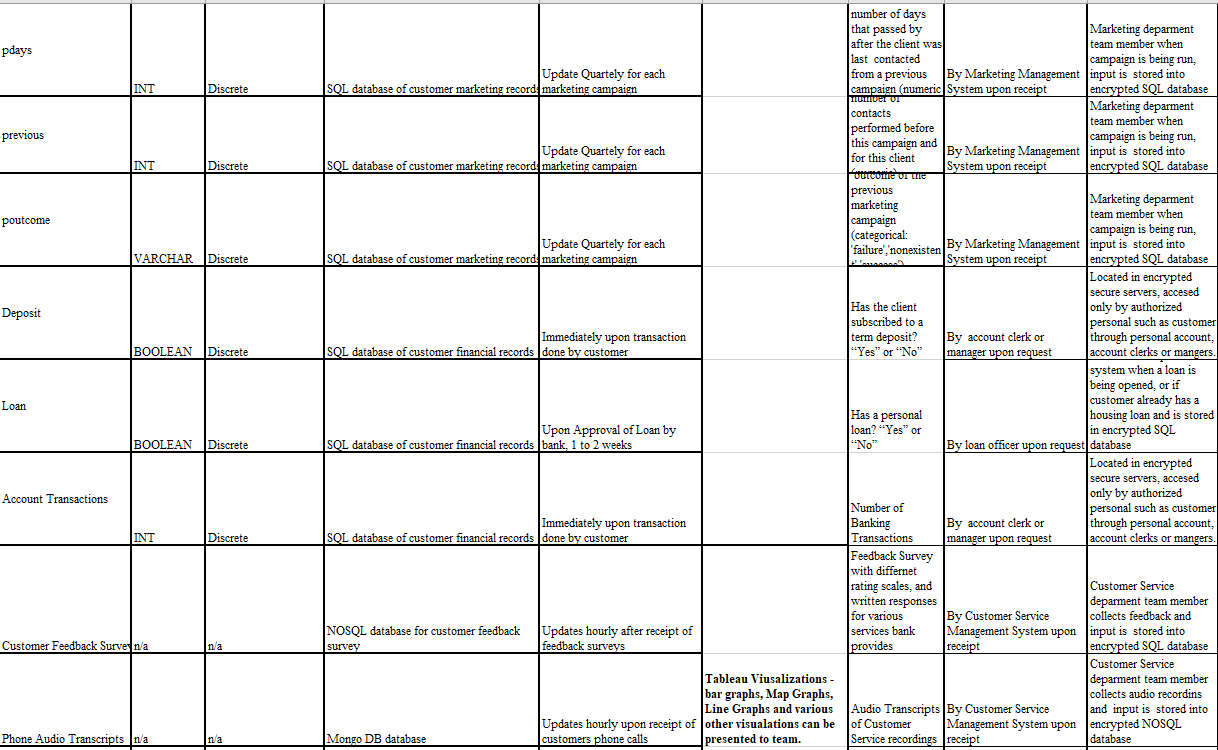
**Process Capability-Shows how closely a process is able to produce the output to its overall specifications.** The external data selected will be used to determine consumer banking industry trends and concerns. The external data derived from the Tangerine bank blog can help determine key consumer trends, and what bank customers are primarily interested in through their article engagement. Large blogging sites such as Reddit and Quora can provide public opinions about the bank, as well as people’s concerns and experiences to determine areas for improvement.

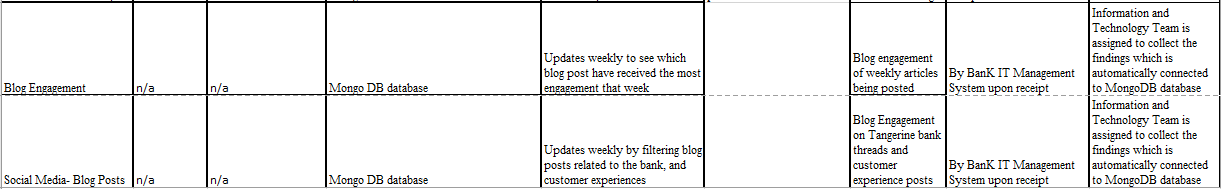
**Stability**- **means the process is being steady or in other words that means the process is producing stable output at all times.** The data collected from the external sources will include both structured and unstructured data. The Structured and Unstructured data will be stored in non-relational databases (NOSQL) and MongoDB.

**Predicting Outcomes-** **the Reliability or Consistency of the model to reproduce.** The use of this data will provide consumer trends and help provide insight into why the bank has had difficulty retaining their loan customer base, as well as how to improve customer retention overall. The data can be used to identify areas of improvement for the marketing team, and what improvements can be made to the bank services.

**Data Collection Map:** 







**References:**

DATA 1201- Data Collection Class Lectures

Tangerine Bank Website& Blog Posts. Quora and Reddit posts on customer opinions of Tangerine Bank.

José María Cavanillas, Curry, E., & Wolfgang Wahlster. (2016). *New Horizons for a Data-Driven Economy*. Cham Springer International Publishing.

Miller, M. (2013). From Data to Decisions: A Value Chain for Big Data. IT Professional, 15(1), 57–59. <https://doi.org/10.1109/MITP.2013.11>

Moro, Sérgio & Cortez, Paulo & Laureano, Raul. (2011). Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. Proceedings of the European Simulation and Modelling Conference.

Parlar, T., & ACARAVCI, S. K. Using Data Mining Techniques for Detecting the Important Features of the Bank Direct Marketing Data. International Journal of Economics and Financial Issues, 7(2), 692-696, 2017.

Sedkaoui, S., n.d. *Data analytics and big data*.

Tuga Mauritsius et al., International Journal of Advanced Trends in Computer Science and Engineering, 8(5),September - October 2019, 2322- 2329