Project Requirements Document: Google Fiber

## **Client/Sponsor:** Hiring Manager

## **Purpose:** The purpose of this project is to evaluate the technical ability of a potential fictional Data Science team member. The project requires the fictional candidate to; find a dataset; analyze the dataset with SQL to identify the correct metrics/features; investigate, select, and implement tools for presenting the dataset metrics to stakeholders. No statistical analysis should be used. The fictional candidate is interviewing for a BI Professional position, however the requirements for the fictional candidate appears to exceed the abilities for a BI Professional therefore Data Science team member is used instead of BI Professional; BI Professionals are likely not responsible for writing code that automates data ingestion of the found dataset/s to database systems, nor write automated client-server side Javascript/HTML code to dynamically connect databases to dashboard interfaces.

## As part of the interview process scenario, the Fiber customer service team has asked for a dashboard using fictional call center data based on the data they use regularly on the job to gain insights about repeat callers. The team’s ultimate goal is to communicate with the customers to reduce the call volume and increase customer satisfaction and improve operational optimization. The goal of the dashboard is to inform stakeholders about insights about repeat caller volumes in different markets and the types of problems they represent.

**Key dependencies:**  
No datasets were given for the fictional scenario, but one was tasked to find fictionalized versions of the actual data. Stakeholders have data access to all datasets on the Google Cloud BigQuery platform for the project, so they can explore the steps taken.

**Stakeholder requirements:**

Below is a list of the established stakeholder requirements, based on the Stakeholder Requirements Document, prioritizing the requirements as: R - required, D - desired, or N - nice to have. Quantitative project success is measured by the fulfillment of these requirements.

In order to continuously improve customer satisfaction, the dashboard must help Google Fiber decision-makers understand how often customers are having to repeatedly call and what problem types or other factors might be influencing those calls.

* A chart or table measuring repeat calls by their first contact date R
* A chart or table exploring repeat calls by market and problem type R
* Charts showcasing repeat calls by week, month, and quarter D
* Provide insights into the types of customer issues that seem to generate more repeat calls D
* Explore repeat caller trends in the three different market cities R
* Design charts so that stakeholders can view trends by week, month, quarter, and year. R

**Success criteria:**

**Specific:** BI insights must clearly identify the specific characteristics of a repeat calls, including how often customers are repeating calls. **Measurable:** Calls should be evaluated using measurable metrics, including frequency and volume. For example, do customers call with a specific problem more often than others? Which market city experiences the most call? How many customers are calling more than once? **Action-oriented:** These outcomes must quantify the number of repeat callers under different circumstances to provide the Google Fiber team with insights into customer satisfaction. **Relevant:** All metrics must support the primary question: How often are customers repeatedly contacting the customer service team? **Time-bound:** Analyze data that spans at least one year to understand how repeat callers change over time. Exploring data that spans multiple months will capture peaks and valleys in usage.

**User journeys:**   
The team’s ultimate goal is to communicate with the customers to reduce the call volume and increase customer satisfaction and improve operational optimization. The dashboard should show relationships between: number of calls, contract type, Internet Service/market type, and problem type. Understanding relationships between these four attributes will help stakeholders understand which caller demographics calls the most for a specific concern, thus allowing them to target specific help to call demographics and reduce calls and churn.

**Assumptions:**   
In order to anonymize and fictionalize the data, the datasets the columns market\_1=Fiber optic, market\_2=DSL, and market\_3=No to indicate three different city service areas the data represents.

The data also lists two problem types:

* Type\_1 is account management
* Type\_2 is technician troubleshooting

Additionally, the best found fictional dataset appears to have been partitioned by the CustomerId and the calls per problem type were SUM aggregated, causing the original DateTime information to be lost. The removal of the DateTime information added a restriction for completing the requirements.

**Compliance and privacy:**

The datasets are fictionalized versions of the actual data this team works with. Because of this, the data is already anonymized and approved. However, you will need to make sure that stakeholders have data access to all datasets so they can explore the steps you’ve taken.

**Accessibility:**   
The dashboards should offer text alternatives including large print and text-to-speech.

**Roll-out plan:**   
The stakeholders have requested a completed BI tool in six weeks.