# **Requirements Specification Document**

## **Template: Requirements Specification Document**

### **Introduction**

In a competitive market, the ability to enhance revenue streams through a deep understanding of customer behavior is crucial. This project leverages the integration of big data analytics with diverse data sources like competitor insights, market trends, and third-party data. Through this integration, the company aims to personalize offers for potential customers and provide value-added services to existing customers, ultimately boosting customer loyalty and increasing revenue.

### **Purpose**

This project is designed to establish efficient data pipelines for the Health Care Insurance Company, enabling it to harness data to make strategic business decisions. The focus areas include:

1. Analyzing customer behaviors in-depth.
2. Evaluating market trends and their impact on customer preferences.
3. Determining the value of loyalty incentives for existing customers.
4. Enhancing overall customer engagement and profitability.

### **Intended Audience and Use**

* **Stakeholders and Project Managers:** To monitor project alignment with company goals and oversee progress.
* **Business Analysts:** To align the derived insights with strategic business initiatives.

### **Product Scope**

* **Benefits:** Enhance revenue and attract new customers through targeted offers.
* **Objective:** Provide robust data-driven insights for effective business strategies.
* **Goals:** Expand customer base and boost retention rates.

#### **Strategic Insights for Decision Making**

* Modify policies and pricing for conditions with frequent claims.
* Develop targeted policies for different demographic groups based on their health conditions and needs.
* Strengthen collaborations with hospitals that serve a large patient base to enhance service delivery.

### **Definitions and Acronyms**

* **Subscriber:** An individual who holds an active insurance policy.

### **Overall Description**

The objective is to build a comprehensive data analytics framework capable of processing extensive datasets from various sources, including competitor data and client databases. This infrastructure aims to refine policy offerings and increase profitability through strategic insights, ultimately enhancing client satisfaction and operational efficiency.

### **User Needs**

* Enhanced insights into customer demographics and health profiles.
* Capability to tailor insurance offerings based on detailed data analyses.
* Knowledge of competitor actions and industry trends.
* Increased financial returns through strategic decisions.

### **Assumptions and Dependencies**

* Data is sourced legally and is readily accessible.
* The existing AWS infrastructure supports project demands.
* Compatibility between business analytics tools and current systems.
* Prompt approval and resource allocation for the project.

## **System Features and Requirements**

### **Functional Requirements**

* Identify prevalent diseases by claim frequency.
* Profile subscribers under 30 engaged in specific subgroups.
* Ascertain the subgroup with the highest number of affiliates.
* Highlight the hospital with the highest patient intake.
* Analyze frequent subscription trends among subgroups.
* Quantify the number of rejected claims.
* Pinpoint the city with the highest claim origin.
* Compare government versus private policy subscriptions.
* Determine the average monthly premiums paid.
* Identify the most financially advantageous group.
* Catalog patients under 18 treated for cancer.
* Detail patients with high-cost cashless insurance claims.
* Identify women over 40 who underwent knee surgery within the last year.

### **External Interface Requirements**

#### **User**

* Enable stakeholder and analyst access to AWS Redshift.
* Implement role-based access controls.

#### **Hardware**

* Utilize AWS EC2 for running EMR clusters.
* Employ AWS S3 for extensive data storage.
* Use AWS Redshift for advanced data warehousing.

#### **Software**

* Deploy Databricks for comprehensive data analysis.
* Use PySpark for complex data transformations.
* Integrate AWS EMR as a developmental platform.
* Manage projects through Jira.
* Maintain code and collaborate via GitHub.

#### **Communications**

* Coordinate project activities and documentation through Jira.
* Collaborate and manage code updates through GitHub.
* Ensure seamless data integration via API interfaces.

### **System Features**

* Develop data pipelines for efficient data ingestion and processing.
* Implement automated systems for reporting and alerts.
* Provide comprehensive analysis capabilities through Redshift.

### **Nonfunctional Requirements**

* **Performance:** Ensure system supports simultaneous use by multiple users and processes large data volumes daily.
* **Safety:** Implement consistent data backup and disaster recovery strategies.
* **Security:** Enforce multi-factor authentication and conduct regular security reviews.
* **Scalability:** Scale system capacity to meet growing data demands and user engagement.