Template - Requirements Specifications Document

# Introduction: A Health Care insurance company is facing challenges in enhancing its revenue and understanding the customers so it wants to take help from Big Data Ecosystem to analyze the Competitors company data received from varieties of sources, namely through scrapping and third-party sources. This analysis will help them to track the behavior, condition of customers so that they can customize offers for them to buy insurance policies and also calculate royalties to those customers who buy policies in the past, this in turn will enhance their revenues.

* 1. Purpose:

The purpose of this project as follows:

* To analyze the competitor data to identify the customer behavior.
* To analyze the behavior and condition of customer to customize the insurance policy.
* Customize insurance policy offer based on customer profile.
* Increase revenue by implementing the appropriate business strategy based on data analysis.
  1. Intended Audience and Use*:*

The intended Audience and use are as follows:

* Project Manager: who will help to plan, monitor and track the project.
* Lead: who will help with issues and communication between the team related with project
* Developer team: who will help to create a pipeline to load, validate and write the data.
* Tester: the one who will help to implement functionality to meet the project requirement.
* Business Analyst: who can use data to analyze the business use cases.
  1. Product Scope:
* The product scope is to analyze competitors data to enhance the revenue of the insurance company.
* It will help to identify the trends of customer and their purchasing insurance policy behavior.
* It will help to make a efficient business strategy.
  1. Definitions and Acronyms:

The below are the definitions and Acronyms are use below:

* AWS s3: Amazon web service storage service.
* AWS Redshift: Amazon data warehouse service in the cloud.
* NA: Note Applicable.
* Jira: tool used for tracking and managing software development projects.

# Overall Description:

## User Needs -

The product is needed to help the Health Care insurance company analyze competitor data and understand customer behavior. The primary users of the product will be analysts and decision-makers within the company who will use the insights generated from the analysis to make informed business decisions. Additionally, developers and data engineers will use the product to implement and maintain the data pipelines and analysis processes.

## Assumptions and Dependencies -

* We are assuming that the current technology stack (AWS S3, AWS Redshift, Databricks, PySpark) will remain stable and available throughout the project.
* The project is based on a cloud-based architecture, specifically AWS, so we are assuming that there will be no major issues with the AWS services being used.
* We are assuming that the data provided for analysis is accurate and representative of the actual customer behavior.

# System Features and Requirements

# Functional Requirements -

* Identify the disease with the maximum number of claims.
* Identify subscribers under the age of 30 who have subscribed to any subgroup.
* Determine the group with the maximum number of subgroups.
* Identify the hospital that serves the most number of patients.
* Identify the subgroups that subscribe the most number of times.
* Determine the total number of claims that were rejected.
* Identify the city from which the most claims are coming.
* Determine whether subscribers mostly subscribe to government or private group policies.
* Calculate the average monthly premium paid by subscribers.
* Identify the group that is the most profitable.

## External Interface Requirements -

### User

* The system shall have a user interface for users to interact with and view the analysis results.
* The user interface shall be intuitive and easy to use, allowing users to view different analyses and generate reports.

### Hardware

* The system shall be compatible with standard hardware configurations, including servers and storage devices.
* It shall be able to run on the hardware available to the Health Care insurance company.

### Software

* The system shall be compatible with the specified software stack, including AWS S3, AWS Redshift, Databricks, PySpark, Jira, and GitHub. –
* It shall be able to integrate with these software tools seamlessly for data processing, storage, and project management.

### Communications

## System Features -

* The system shall have a communication interface for sending and receiving data from external sources.
* It shall be able to communicate with external systems and services to fetch data and perform analyses.

## Nonfunctional Requirements -

### Performance requirements: Efficiently process large volumes of data and generate analysis results with minimal latency.

### Safety requirements: Ensure the confidentiality and integrity of sensitive data and prevent unauthorized access.

### Security requirements: Comply with data protection regulations and have robust authentication and authorization mechanisms.

### Usability requirements: Provide a user-friendly interface with clear instructions for tasks and information access.

### Scalability requirements: Scale to accommodate increasing data volumes and user loads, handling concurrent requests and tasks without performance degradation.

## 