Template - Requirements Specifications Document

# **Introduction -**

## **Purpose** – The requirements are to enable the health insurance company to leverage customer data for creating targeted strategies, offering personalized insurance policies, and calculating customer royalties based on analysis of their behavior. This will also support enhanced revenue generation by making more informed decisions.

* 1. **Intended Audience and Use** –

**Business Analyst:** BA will gather the documents and understand the business requirement which help developers to code in ease.

**Developers**: Dev understand the technical requirement and implement it in system. Build and maintain the data pipelines and ensure the accurate flow of information from AWS S3 to AWS Redshift.

**Tester**: Tester will help to validate the data to check if it meets the requirements.

**Manager**: Manager will make sure if the data matches the business goal and whether it is delivered on time or not.

## **Product Scope** – The product will analyze the data sources to provide insight into customer behavior, helping the company offer targeted insurances policies. Enhance the revenue by offering customized insurance products.

## **Definitions and Acronyms** –

* **AWS S3**: Amazon Web Services Simple Storage Service, a cloud storage service.
* **AWS Redshift**: A cloud-based data warehousing service.
* **Databricks**: A unified analytics platform for big data and AI.
* **PySpark**: A Python API for Spark, used for big data processing.
* **ETL**: Extract, Transform, Load; a process for data migration and transformation.

# **Overall Description-**

## **User Needs** –

**Data Analysts**: They help to process and analysis the customer and competitor data.

**Business Analyst**: They will collect the requirements and help in customizing the insurances policy.

**Agent:** They will deal at ground level with customers and suggest the appropriate policies.

**Admins**: Work on the tools needed for data, storage and all workflows.

## **Assumptions and Dependencies** –

## The flow will be built in AWS s3 and Redshift with all other required tools.

* Data will be collected from different sources.
* The company will have access to all historical data from both customer and competitor data.

# **System Features and Requirements***.*

## **Functional Requirements** –

**Data Sources**: The process will pull customers and competitive data from different sources.

**Data Processing**: This includes handling the null, deleting duplicates and cleansing unnecessary data by using different transformations and loading clean data into Redshift

## **External Interface Requirements** -

### **User**: Creation of Reports or dashboard for business user to visualize the customer behavior, policies etc.

### **Hardware**: This system will interact with AWS (s3, Redshift) which doesn’t require users to interact directly.

### **Software :** Using pyspark for data cleaning and transformation which will be utilized later by AWS services.

### **Communications:**

## **System Features** –

This is pretty much a testing area where we can check if there is any mismatch in data, If any data is lost while applying transformation. Error occurred while processing data. Also ensure to handle the increasing data volume.

## **Nonfunctional Requirements**

### **Performance requirements:**

Processes should generate the required data on given timeframe and reports should generate.

### **Safety requirements:**

All data should be loaded, and secure access for secure data should be implemented if needed

### **Security requirements:**

Data security should be implemented in every place over the network and when it is stored in storage system or databases.

### **Usability requirements:**

The data should be visible for any non-technical users. Reports should be easy to use and customized to allow users to go deep into data.

### **Scalability requirements:**

The system should support the increasing data volume without losing data and impacting performance.

## 