Template - Requirements Specifications Document

**1. Introduction**

**a. Purpose**

- The purpose of this Requirements Specifications Document is to outline the requirements for developing a data analysis system for enhancing revenue and customer understanding in a Health Care insurance company using Big Data Analysis.

**b. Intended Audience and Use**

- This document is intended for a diverse range of stakeholders within the Health Care insurance company and its operations, including executives, marketing teams, data analysts, technical teams, project managers, and others. It serves as a reference for understanding the project's goals, scope, and specific requirements.

**c. Product Scope**

- The product scope entails the creation of a data analysis system that addresses challenges faced by the Health Care insurance company. The system aims to enhance revenue generation and customer understanding through data-driven insights. The scope includes benefits, objectives, and goals such as enhanced revenue generation, data-driven decision making, customer satisfaction, competitive advantage, and efficient operations.

**d. Definitions and Acronyms**

- Definitions:

1. Health Care Insurance Company: The company providing insurance coverage for medical expenses and healthcare services.

2. Big Data Ecosystem: A collection of technologies and tools used for processing and analyzing large and complex datasets.

3. Competitors' Data: Information about other insurance companies' policies, customer demographics, claims data, etc.

4. Data Scraping: Extracting data from websites using automated methods.

5. Third-Party Sources: External data providers offering information and datasets.

6. Royalties: Payments to customers as rewards for purchasing insurance policies.

- Acronyms:

1. EMR: Elastic MapReduce

2. S3: Simple Storage Service

3. AWS: Amazon Web Services

**2. Overall Description**

**a. User Needs**

- Executives and Decision Makers: Require high-level insights for informed business decisions, accessed through dashboards and reports.

- Marketing and Sales Teams: Need customer segmentation and personalized insights for effective campaigns.

- Product Development Teams: Require customer preferences and subscription data for creating relevant insurance products.

**b. Assumptions and Dependencies**

- Data Analysts/Businesss analyst for generating reports in Tableau/Power(visualizing tools).

- Data Quality and Availability: Assuming provided data is accurate and relevant for analysis.

- Data Privacy and Compliance: Assuming data used is compliant with regulations.

- Data Integrity – Assuming there are no duplicates, nulls, consistent and accurate.

**3. System Features and Requirements**

**a. Functional Requirements**

- Identify the disease with the highest number of claims.

- Locate subscribers under 30 who have joined any subgroup.

- Determine the group with the most subgroups.

- Find out which group is generating profits

- Claims that were rejected

- Age group of people with cancer

**b. External Interface Requirements**

- AWS S3

- AWS Redshift

- Databricks

- AWS EMR Studio

- Pyspark

-Jira

- GitHub

**c. System Features**

1. Data Ingestion and Integration:

- Ability to gather and integrate data from diverse sources.

2. Data Cleaning and Preprocessing:

- Modules for handling missing values, duplicates, and inconsistencies.

**d. Nonfunctional Requirements**

- Performance Requirements: Query response times within 3 seconds.

- Safety Requirements: Handling sensitive healthcare data in compliance with regulations.

- Security Requirements: User authentication and authorization mechanisms.