Template - Requirements Specifications Document

1. Introduction - Utilizing a Big Data Ecosystem to enhance revenue and customer understanding for a Health Care insurance company is the main requirement and this document provides a comprehensive guide for the development team to understand the project scope, objectives, and requirements.
    a. Purpose -The purpose of these requirements is to provide a clear understanding of the project goals, functionalities, and constraints involved in implementing the Big Data solution for revenue enhancement in the Health Care insurance sector.
    b. Intended Audience and Use - This document is intended for developers, testers, project managers, and other stakeholders involved in the development and implementation of the Big Data solution. It serves as a reference guide to ensure alignment with the project objectives and requirements throughout the development lifecycle.
    c. Product Scope - The scope of this project includes leveraging a Big Data Ecosystem to analyze competitors' data, track customer behavior, and customize insurance offers to enhance revenue generation for the Health Care insurance company. The primary goals are to improve customer insights, personalize offers, and optimize revenue streams through data-driven strategies.
    d. Definitions and Acronyms -
        i. SRS: Software Requirements Specification
        ii. Big Data Ecosystem: Denotes an assemblage of tools, technologies, and frameworks employed for processing and scrutinizing extensive datasets.
        iii. Revenue Augmentation: Signifies strategies aimed at amplifying revenue generation for the Health Care insurance company.
        iv. PySpark: A Python API for Apache Spark, employed for large-scale data processing and analysis.
        v. Redshift: Amazon Redshift, a cloud-based data warehousing service proffered by Amazon Web Services (AWS).
        vi. Databricks: A unified analytics platform furnishing a collaborative milieu for data science and engineering endeavors.

2. Overall Description - The project involves developing data pipelines and analytical capabilities using a Big Data Ecosystem to analyze competitors' data and customer behavior. The solution will enable the company to tailor insurance offers, track customer preferences, and optimize revenue streams.
    a. User Needs - Users of the product include data analysts, business strategists, and insurance company executives. They require a user-friendly interface to access analytics insights, visualize data trends, and make informed business decisions.
    b. Assumptions
        i. The availability and quality of data from competitors and third-party sources.
        ii. Adequate infrastructure and resources for implementing the Big Data solution.
    c. Dependencies:
        i. Availability of AWS services such as S3, Redshift, and EMR for data storage and processing.
        ii. Integration with Databricks for data analysis and visualization.
3. System Features and Requirements -
    a. Functional Requirements -
        i. Data Ingestion:
            1. Retrieve competitors' data from various sources, including web scraping and third-party providers.
            2. Store the data in AWS S3 for further processing.
        ii. Data Cleaning:
            1. Identify and handle missing values and outliers in the dataset.
            2. Remove duplicate records to ensure data integrity.
        iii. Data Analysis:
            1. Analyze disease claims data to identify patterns and trends.
            2. Segment subscribers based on age, gender, and policy preferences.
        iv. Revenue Optimization:
            1. Calculate royalties for customers based on their past policy purchases.
            2. Customize insurance offers based on customer behavior and preferences.

b. External Interface Requirements -

i. *User Interface:*
   1. *Provide a user-friendly interface for accessing analytics insights and visualization tools.*

ii. *Hardware Interface:*
   1. *Ensure compatibility with AWS infrastructure for data storage and processing.*

iii. *Software Interface:*
   1. *Integrate pyspark,AWS S3,Redshift and Databricks for data analysis and visualization.*

iv. *Communications Interface:*
   1. *Enable communication between different stakeholders with the use of Github,Jira.*

c. System Features -

i. Data Ingestion:
   1. Retrieve competitors' data from various sources, including web scraping and third-party providers.
   2. Store the data in AWS S3 for further processing.

ii. Data Cleaning:
   1. Identify and handle missing values and outliers in the dataset.
   2. Remove duplicate records to ensure data integrity.

iii. Data Analysis:
   1. Analyze disease claims data to identify patterns and trends.
   2. Segment subscribers based on age, gender, and policy preferences.

iv. Revenue Optimization:
   1. Calculate royalties for customers based on their past policy purchases.
   2. Customize insurance offers based on customer behavior and preferences.

d. Nonfunctional Requirements -

i. Performance Requirements:
    1. Ensure efficient data processing and analysis to meet business needs.

ii. Safety Requirements:
    1. Implement data security measures to protect sensitive information.

iii. Security Requirements:
    1. Ensure data privacy and compliance with regulatory standards.

iv. Usability Requirements:
    1. Provide intuitive user interfaces for data exploration and visualization.

v. Scalability Requirements:
    1. Design the solution to scale seamlessly with growing data volumes and user demands.

vi.