Template - Requirements Specifications Document

# **Introduction** – A Health Care Insurance Company Requirements

## Purpose –To analyze the competitor's data received from varieties of sources(third-party sources).

## Intended Audience and Use - Developers, testers, project managers, data analysts, and data scientists.

## Product Scope - This analysis will help them to track the behavior, and condition of customers so that they can customize offers for them to buy insurance policies and calculate royalties to those customers who buy policies in the past, this in turn will enhance their revenues.

## Definitions and Acronyms –

1. **Overall Description** - The goal of the project is to create data pipelines for the Health Care insurance company which will make the company make appropriate business strategies to enhance their revenue by analyzing customers' behaviors and sending offers and royalties to customers respectively.
   1. We will upload the sample data in AWS S3.
   2. We will clean the datasets, especially duplicates and null values.
   3. Upload cleaned data corresponding to each data set into a redshift table.
   4. Create a separate Redshift table for each use case output in a Redshift schema
      * Schema = Project-Output
   5. User Needs – To track the behavior, and condition of customers so that they can customize offers for them to buy insurance policies and calculate royalties to those customers who buy policies in the past.
   6. Assumptions and Dependencies – We are assuming that there will be null values and duplicates. So, we need to drop the duplicates and replace the null values with NA.

# **System Features and Requirements** -

a. Functional Requirements –

Requirements :

* Which disease has the maximum number of claims.
* Find those Subscribers age less than 30 and they subscribe to any subgroup
* Find out which group has the maximum subgroups.
* Find out the hospital which serves most number of patients
* Find out which subgroups subscribe the most number of times
* Find out the total number of claims which were rejected
* From where most claims are coming (city)
* Which groups of policies do subscriber subscribe mostly Government or private
* Average monthly premium subscriber pay to the insurance company.
* Find out Which group is most profitable
* List all the patients below the age of 18 who admit for the scancer
* List patients who have cashless insurance and have total charges greater than or equal for Rs. 50,000.
* List female patients over the age of 40 that have undergone knee surgery in the past year

## External Interface Requirements –

## **Environment**

* AWS S3
* AWS Redshift
* Databricks
* AWS EMR Studio
* Pyspark
* Jira
* GitHub

**Two GitHub Repo**

* Requirement specification documents and Solution Design Documents Repo
* List of Jira user stories and tasks which you have created for individual use cases in a scrum-type project.
* List of Jira user stories and tasks created for test cases Or a single test cases xls sheet.
* A database schema design doc for final redshift tables with primary and foreign key combinations.
* PySpark Source Code Repo
* Data cleansing modules
* Result generations modules