Template - Requirements Specifications Document

# Introduction - *This documentation is the SRS of the Healthcare Insurance Project*

## Purpose -*The goal of the Project is to create data pipelines for the Health Care Insurance Company which will make the company make appropriate business decisions*

## Intended Audience and Use - *This SRS is accessible to the business stakeholders who are responsible for enhancing its revenue and making decisions in the company. This is also accessible to the data engineering team which consists of developers, testers, and project managers*

## Product Scope - *Benefits: increase the revenues and attract more customers by providing them customized offers, Objective: We should analyze the data such that we can provide answers for the requirement that stakeholders can make business decisions, Goals: Increase Revenues by acquiring new customers, Increase the retention rate of the existing customers*

## Definitions and Acronyms -

# Overall Description *We will be creating a ETL Pipeline.*

1. We will first ingest the data from different CSV files to AWS S3
2. For each file we will do data cleaning which involves
   1. Checking for null values
   2. Count the total Null Values for each columns
   3. Replace the Null values by NA
   4. Deduplication
   5. Check the labels
   6. Join different tables to get the corresponding datasets
3. After the cleaning, we will upload each data set into redshift table
4. We should create Schema design for target tables
5. We should create separate redshift table for each use case output in a redshift schema

This a completely new product developed to understand the customer behavior so that we can make a decision to increase company’s revenue by acquiring more customers.

The final usecase of the product will be viewed by business stakeholders to make data-driven decisions

## User Needs - The business stakeholders use the end result of the project which is to get information as described in System Features to make different decisions

## Assumptions and Dependencies -

# System Features and Requirements -*.*

## Functional Requirements –

* + 1. Which disease has a maximum number of claims.
    2. Find those Subscribers having age less than 30 and they subscribe any subgroup
    3. Find out which group has maximum subgroups.
    4. Find out hospital which serve most number of patients
    5. Find out which subgroups subscribe most number of times
    6. Find out total number of claims which were rejected
    7. From where most claims are coming (city)
    8. Which groups of policies subscriber subscribe mostly Government or private
    9. Average monthly premium subscriber pay to insurance company.
    10. Find out Which group is most profitable
    11. List all the patients below age of 18 who admit for cancer
    12. List patients who have cashless insurance and have total charges greater than or equal for Rs. 50,000.
    13. List female patients over the age of 40 that have undergone knee surgery in the past year

## External Interface Requirements

### User

### Hardware

Cloud Computing (AWS Serverless Compute), Storage (Amazon S3)

### Software

AWS S3 => Databricks, PySpark, AWS EMR Studio => AWS Redshift

Using S3 for staging Area and/or permanent storage for the CSV files we will be working with

We will be using PySpark in AWS EMR Studio to run the transformations, which includes data cleaning, and processing the data

We will be testing the code running in AWS EMR Studio in Databricks to save costs.

We will create tables in AWS Redshift from the each cleaned data source.

### Communications

Jira for User stories and tasks

Github for collaboration and version control

## System Features - *System features are a type of functional requirements. These are features that are required in order for a system to function.*

## Nonfunctional Requirements - *Nonfunctional requirements, which help ensure that a product will work the way users and other stakeholders expect it to, can be just as important as functional ones. These may include:*

### Performance requirements

### Safety requirements

### Security requirements

### Usability requirements

### Scalability requirements

## 