# Solution – *The project will be help us analyse the historic data from our competitors and help us create better offering to attract new customers and retain existing one. The detailed steps on how we will achieve this are as follows:*

1. Store the raw data in AWS S3.
2. Use pyspark in Databricks to transform the data to help answer all of the use cases mentioned below.
3. Store the final tables into AWS redshift so, users can access it in format that is relevant to further analyse.

# Use Cases – *Below are all of the use cases :*

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
14. Database Design - List down all possible db(Redshift) tables here

## Tables Metadata Info with Pk/FK relationship: We will have 12 tables as we will be providing final data that will be readily usable and answer each of the thirteen use cases.

## ER diagram - *Optional*

# Technologies and Platforms to be used in this solution -*here are the list of tech we wil use :*

1. Databricks.
2. AWS redshift
3. AWS S3
4. Jupyter Notebook
5. Jira
6. Github

## 