# Solution –

Created S3 folders and Stored all the different format files into S3. Created Crawler and added the S3 data source and created tables. After the data creation for visualization purposes created ER diagram with different tables and column name for input metadata, then created another table accordingly with the use case requirements into order to generate desired output metadata.

Cleaned the data, transformed the cleaned data into redshift.

# Use Cases –

* Which disease has a maximum number of claims.
* Find those Subscribers having age less than 30 and they subscribe any subgroup
* From where most claims are coming (city)
* Average monthly premium subscriber pay to insurance company.

1. Database Design –

Anysubgroup

Diseasemaxclaim

Mostclaims

premiumaverage

## Tables Metadata Info with Pk/FK relationship -

## ER diagram – A screenshot of a computer Description automatically generated

# Technologies and Platforms to be used in this solution –

● AWS S3

● AWS Redshift

● Databricks

● Pyspark

● Jira

● GitHub