# Datawarehouse and Business Intelligence for Healthcare Provider Data

# Objective/ Scope of Document

Medicare is the most important sector for any government to gauge their performance. However, it can be tedious for a government official to see all the metrics at a glance like cost, revenue, expenses etc. Our objective for this project would be to create a tool which will enable the user to slice and dice this data as per convenience. This includes creating dashboards and a cube(yet to decide based on further scope of project). Upon completion of the project, following objectives are expected to be achieved:

* To analyze the Provider’s covered charges, Total payments to all Providers and Medicare payments at hospital level and also based on the type of clinical condition (diagnosis) and the procedure furnished by the hospital during the stay of the inpatients
* To analyze the Provider’s submitted charges, Medicare and Beneficiaries’ payments to the provider at hospital level and also based on Medicare’s APC (Ambulatory Payment Classification) description for outpatients
* To connect the hospital dataset with census data based on City and State level and finding if there is sufficient availability of Hospitals for the people living in that area
* As an individual interested in knowing the cost of setting up a hospital. I want to know the cost break-up to set-up a new hospital in a given location in the US.
* As an investor I want to know if there is a difference in the cost and the break-up of cost for a given provider. I would be interested in knowing the difference in cost for the different healthcare provided by a given hopital.
* An exhaustive data warehouse, easily available, clean and understood
* Dashboard, with easily interpretable information and interactive
* An OLAP cube, with slicing and dicing by geography, type of hospitals, time, titles etc.
* Create a sentiment on hospital financials on their performance state wise to gauge performances

**Stakeholders:**

Anyone who wants to understand the Healthcare performance at a glance and make decisions regarding the Hospitals.

# Tools

1. SQL Server Integration Services (SSIS)
2. SQL Server Management Studio (SSMS)
3. Azure SQL Server
4. Azure SQL Database
5. Azure DevOps
6. Power BI Desktop

# Overview of the data model

# 

# 

# ER diagram

# 

# 

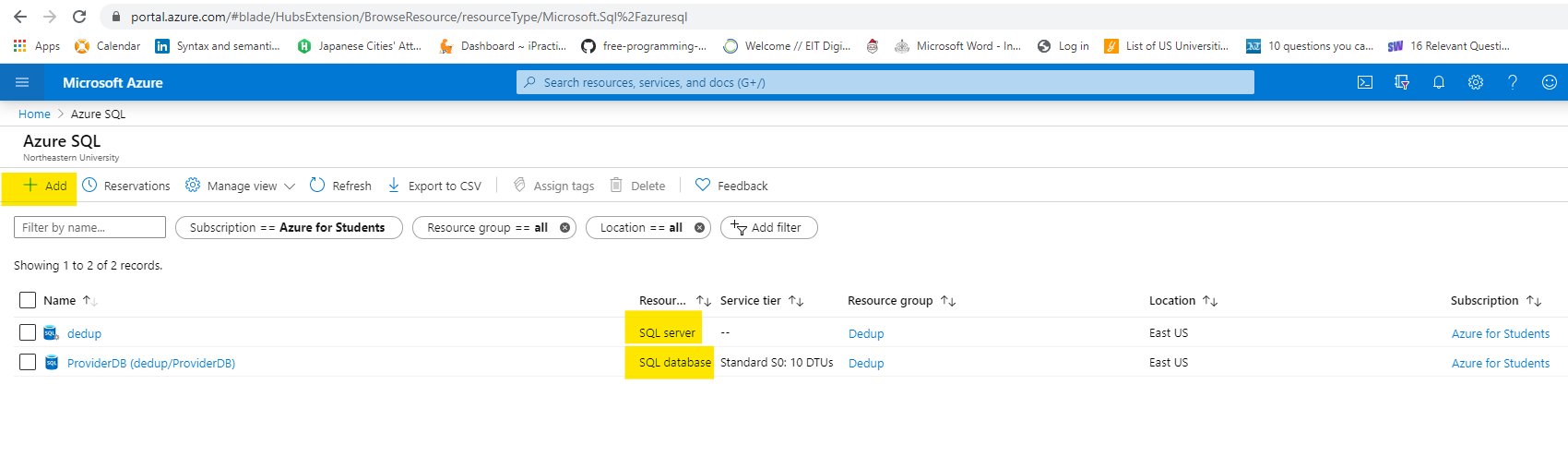
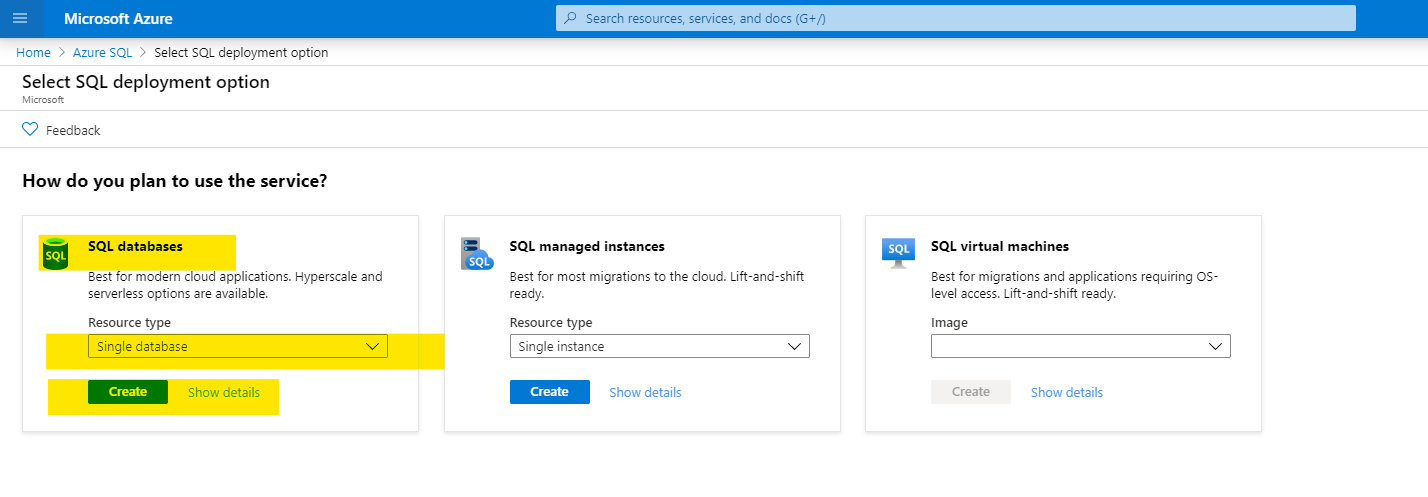
# 

# Implementation

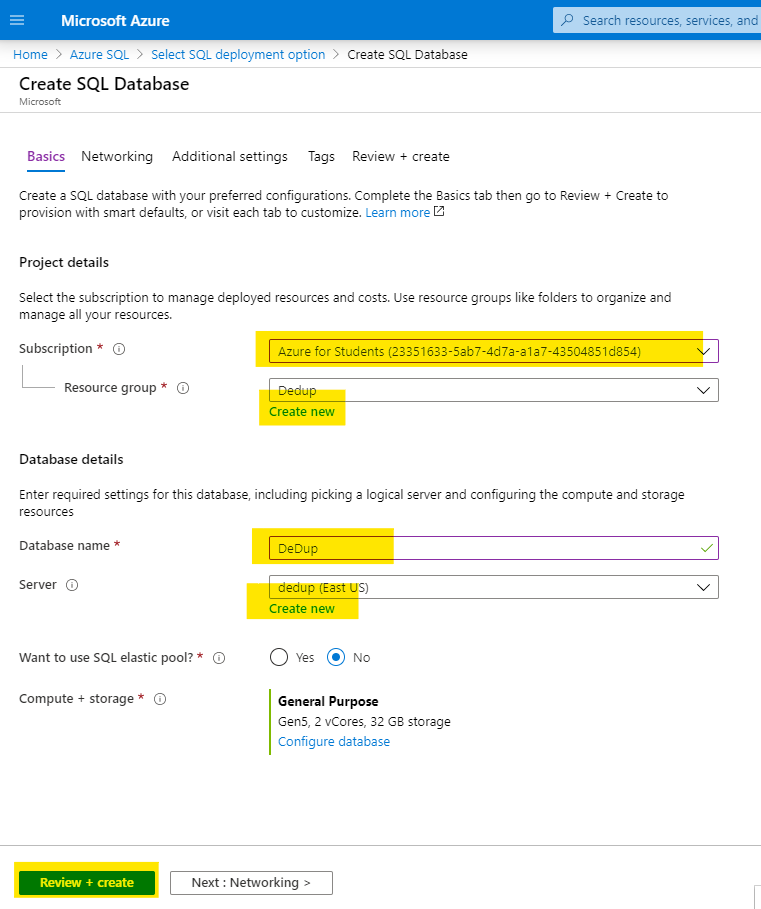
Azure Setup Instructions:

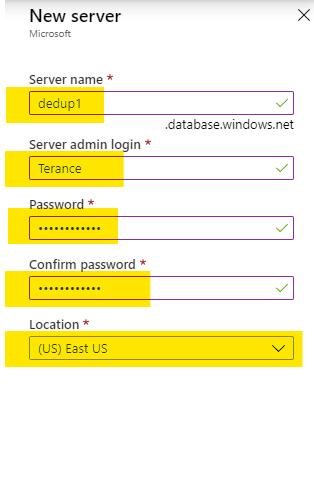
1. Create the Azure account.   
   Reference link: <https://azure.microsoft.com/en-us/free/students/>

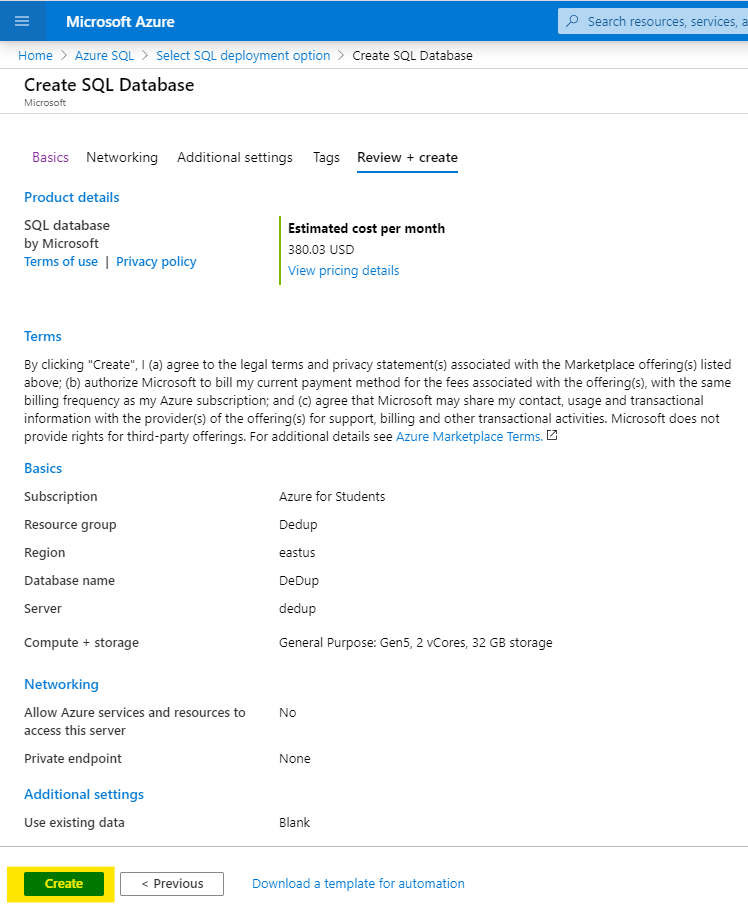
1. Go to <https://portal.azure.com/> to access the resource creation.
2. To create the Azure Database

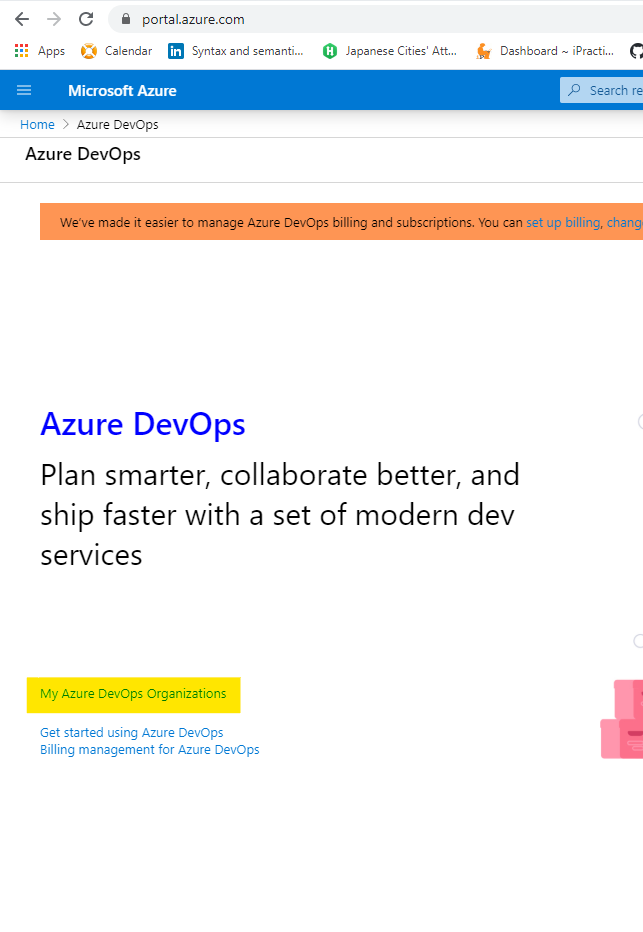
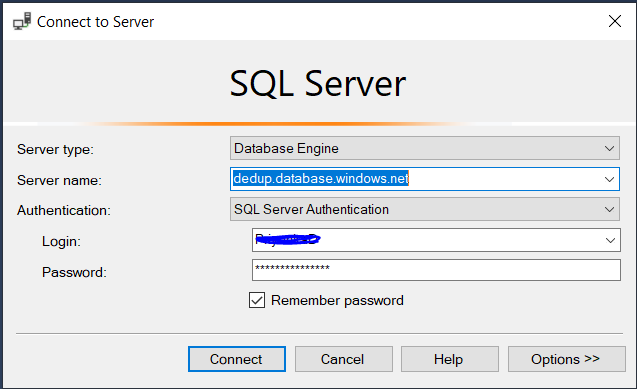
Go to Azure SQL server . Click Add  
  
Create a SQL database: Single database  
  
Enter the required details to create the database:

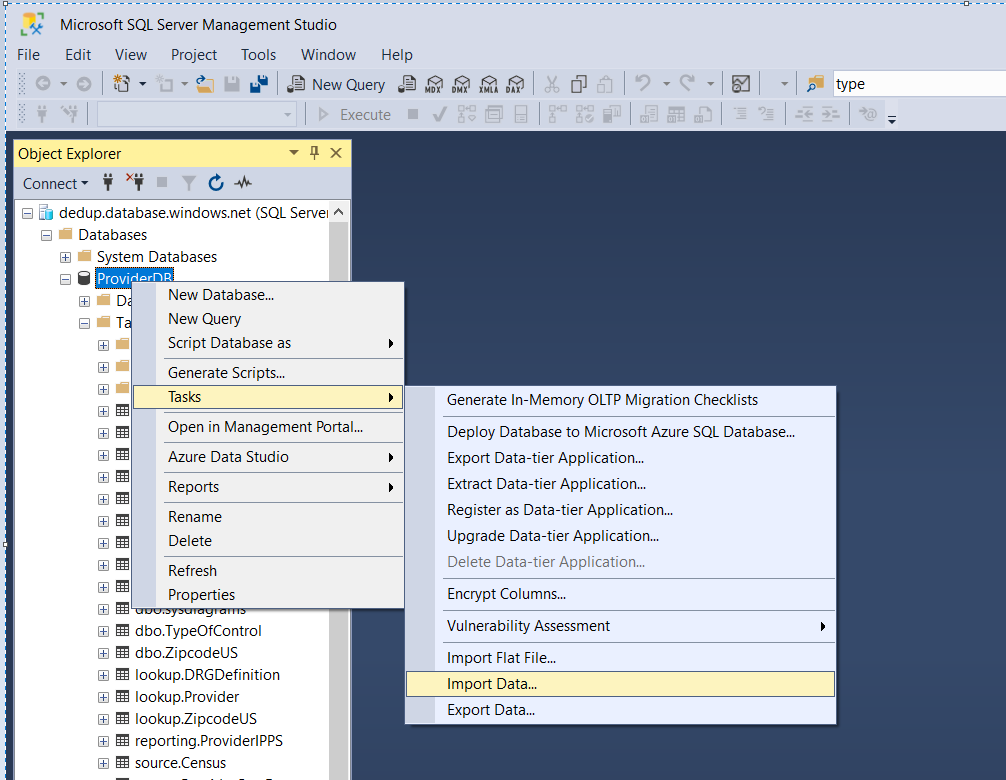
Create the Server if it isn’t created:





Check the resources and additional resources and create the server.

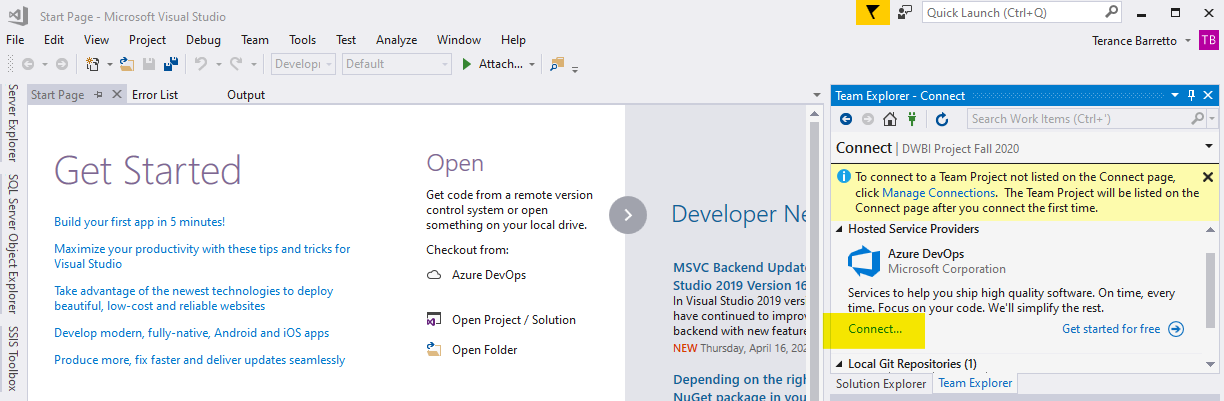
1. Create Azure DevOps  
   
2. Download the files
3. Connect to Azure SQL server database
4. Import the file in SSMS into the Storage Schema



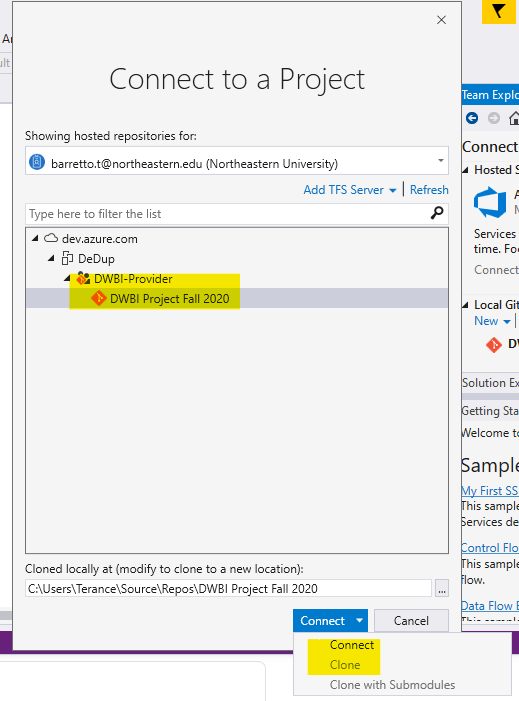
1. Connect to Azure SSIS
2. Make a connection to the Git Repo created via AzureDevOps



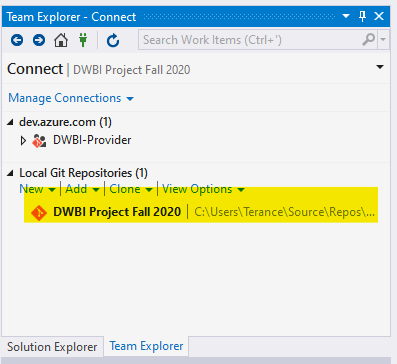
1. Click on manage connection to make a connection to the repo



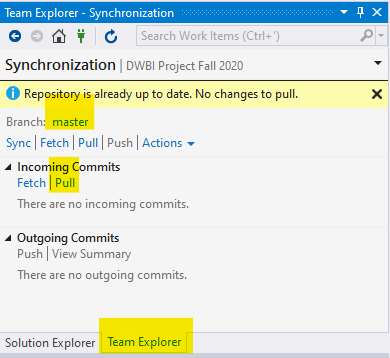
1. Connect to your microsoft account and clone the project repo



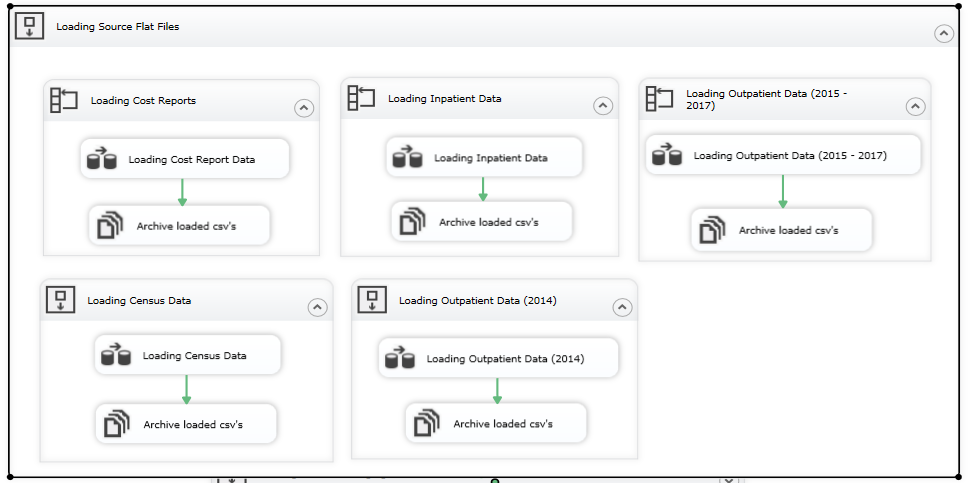
1. Go to the cloned repo location, and open the project file.



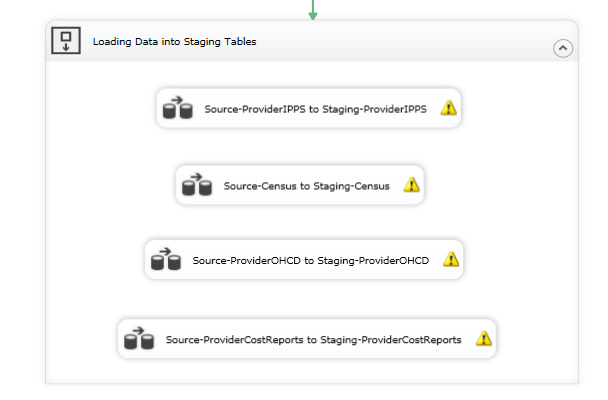
1. Use the “Pull” command to sycn the online repo with the copy on your local machine.



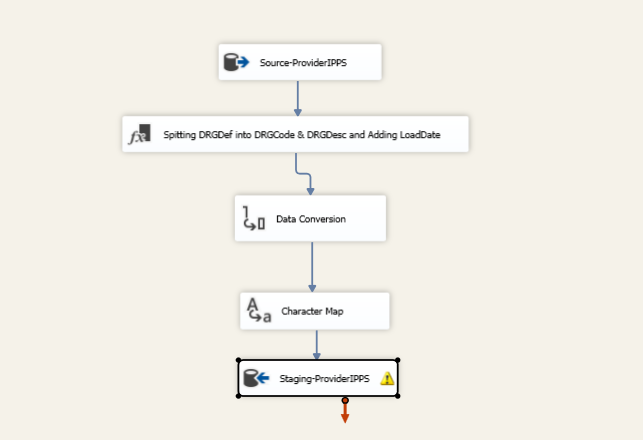
1. Run the container “Load Source Flat Files” in SSIS to load it to Source Table



1. Run the container “Loading Data into Staging Tables” in SSIS to load it to Staging table



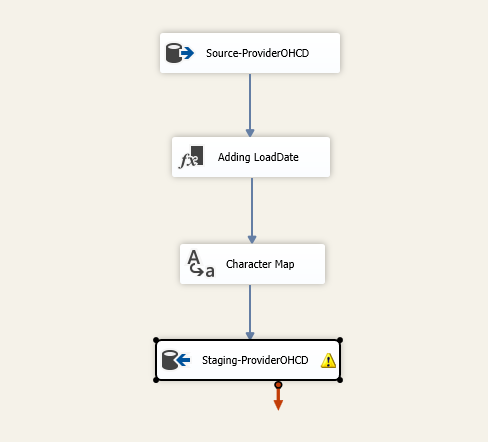
* Provider IPPS (Source to Staging)



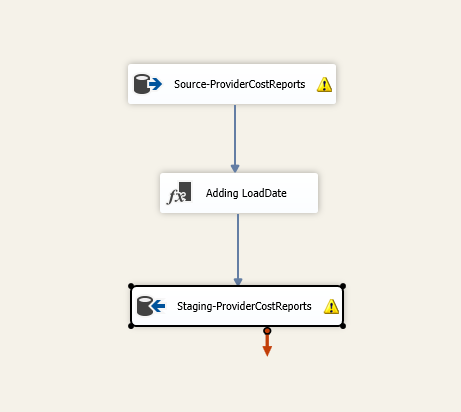
* Census (Source to Staging)



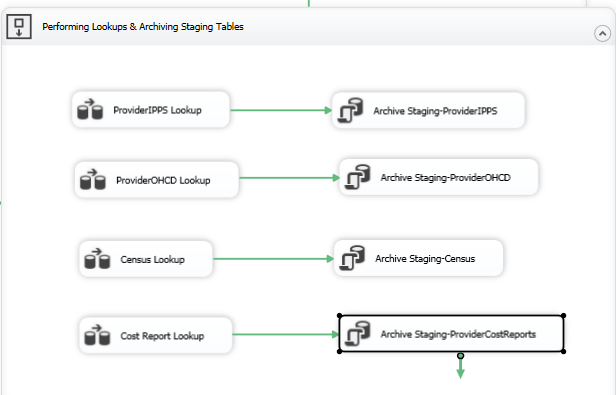
* Provider OHCD (Source to Staging)



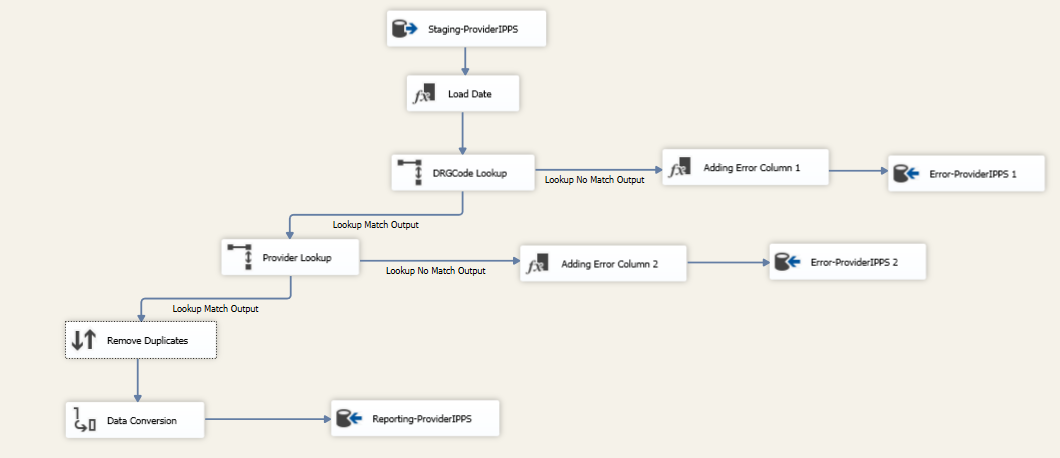
* Provider Cost Reports (Source to Staging)



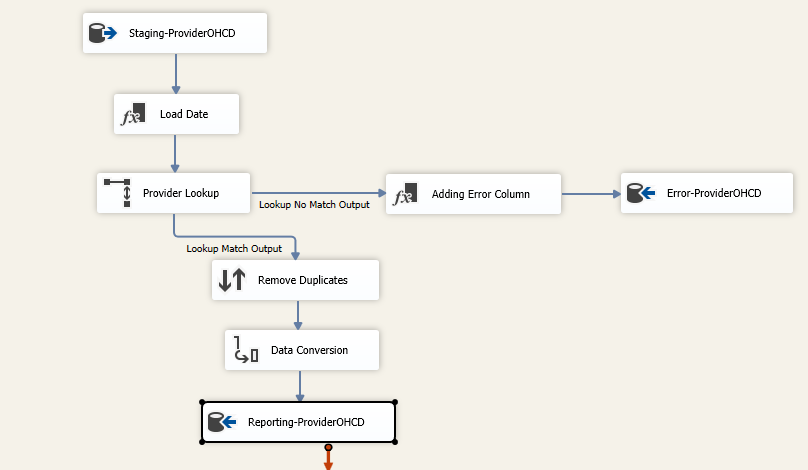
1. Lookup



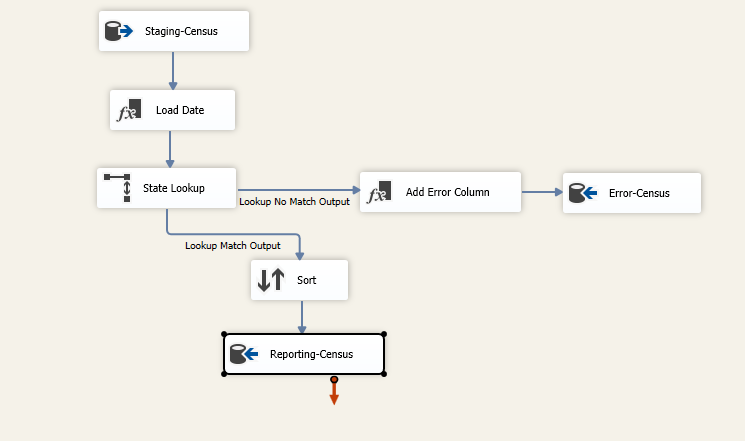
* Provider IPPS Lookup



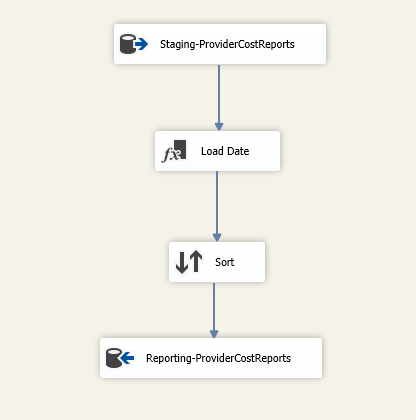
* Provider OHCD Lookup



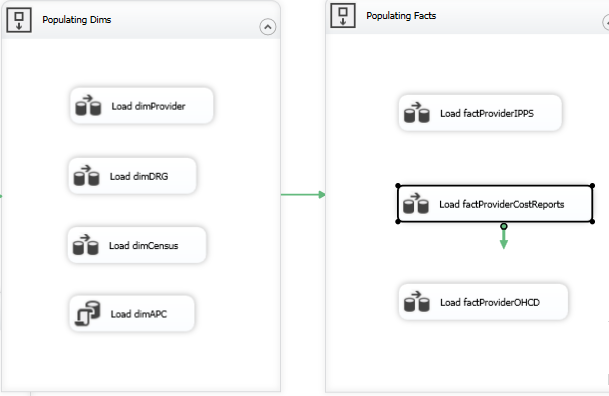
* Provider OHCD Lookup



* Provider OHCD Lookup



1. Populating Dims and Facts

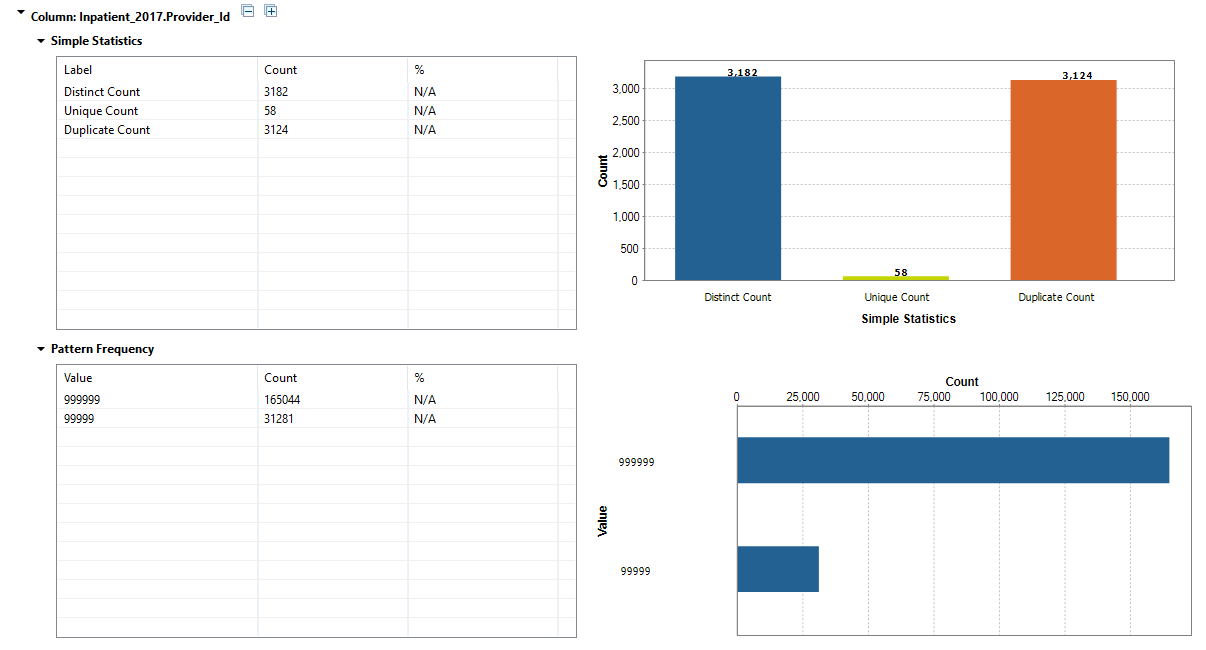


# Observations

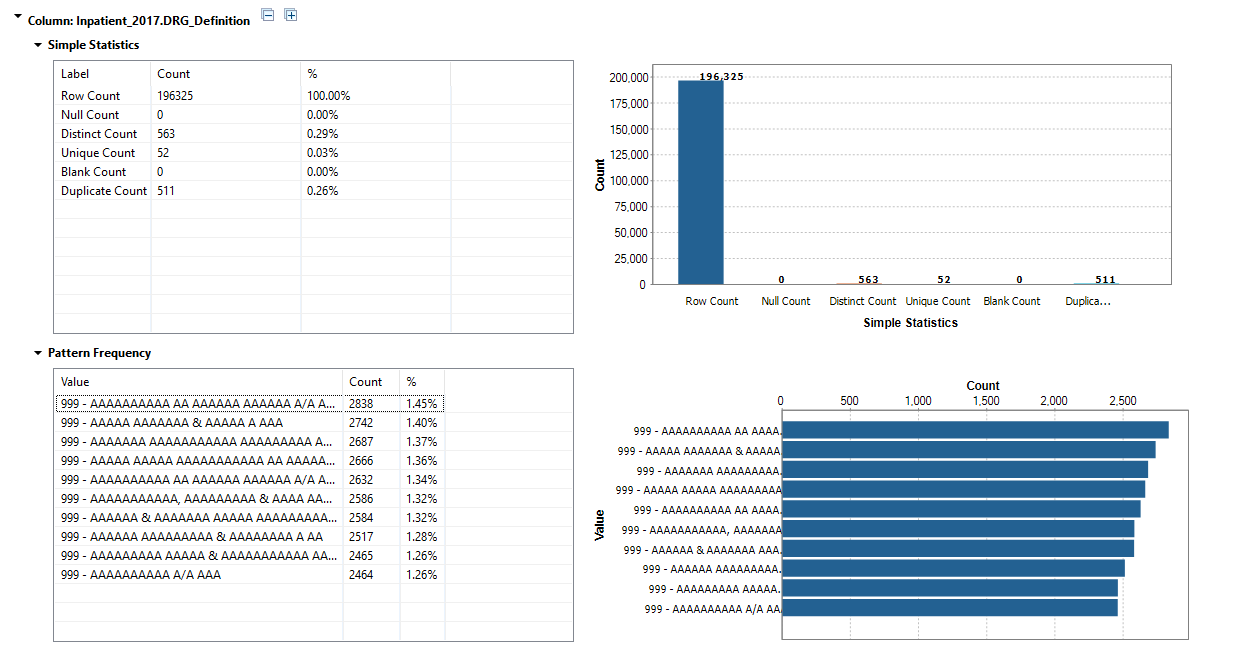
Data Cleaning

For data cleaning,we used Talend for data profiling to examine the available data. The key fields from each data source were examined using the Pattern Frequency and Simple Statistics feature of Talend.

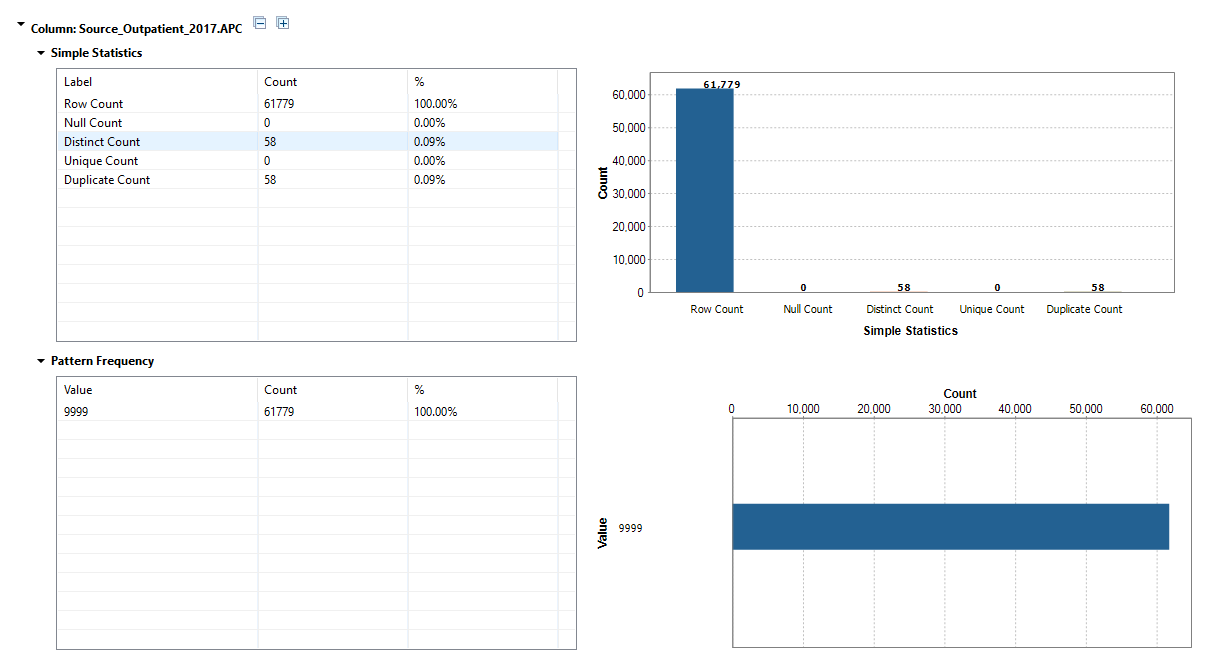
* Provider ID



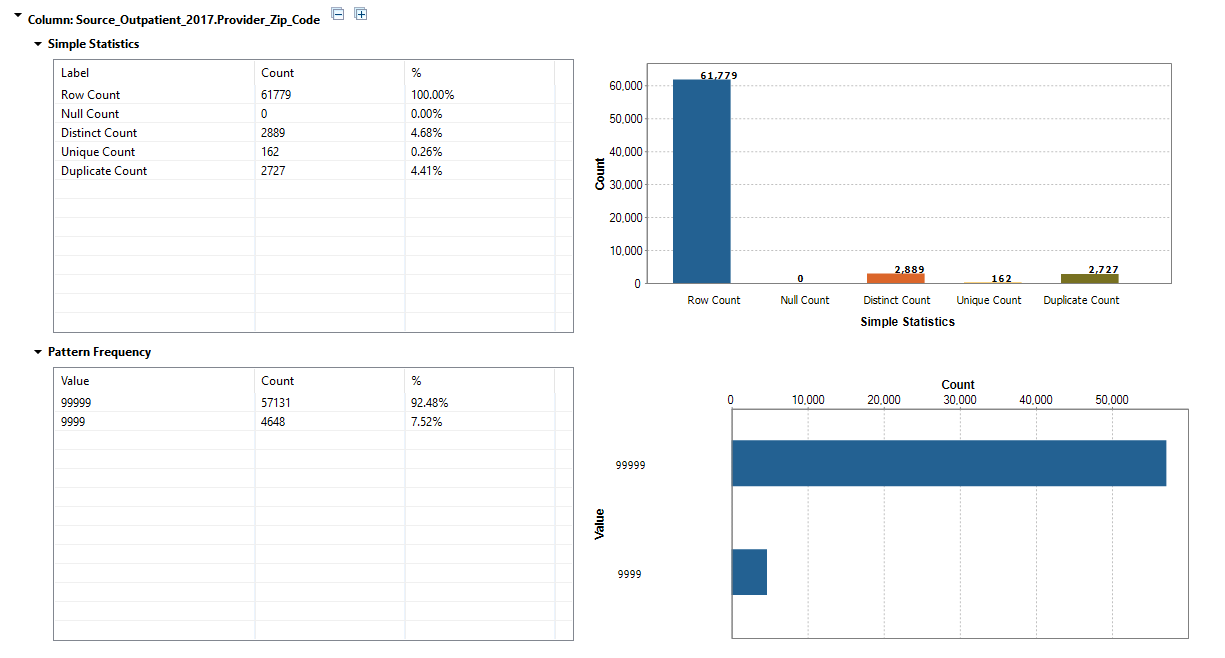
* DRG Definition



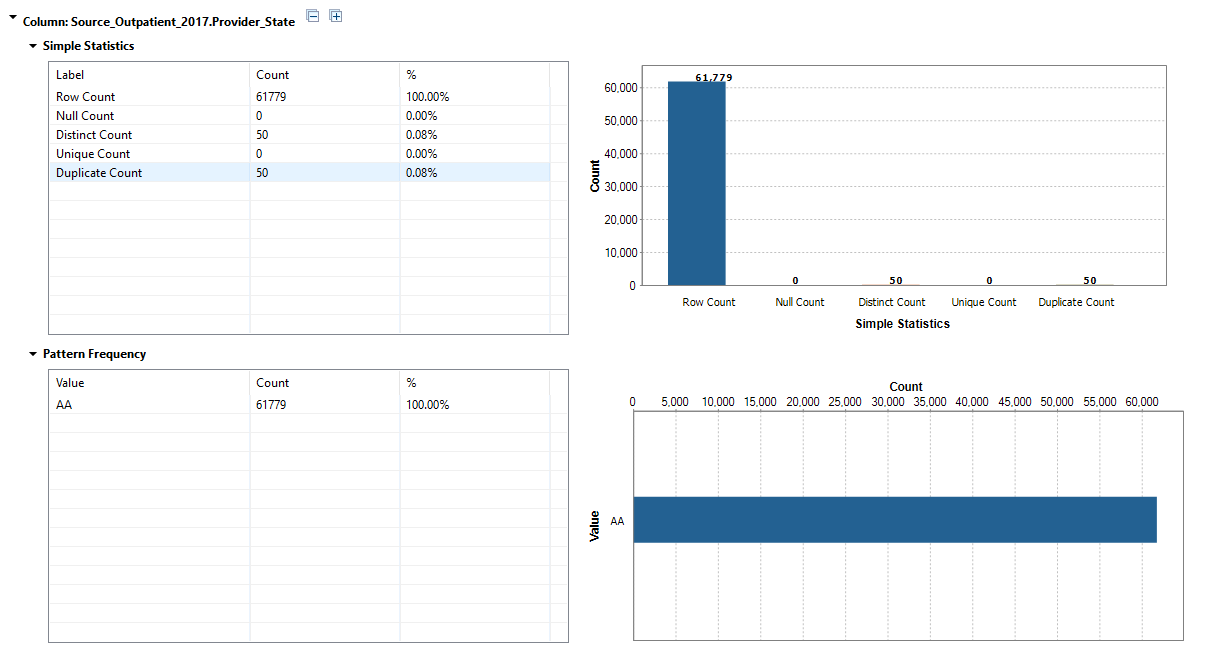
* APC Code



* Zip Code

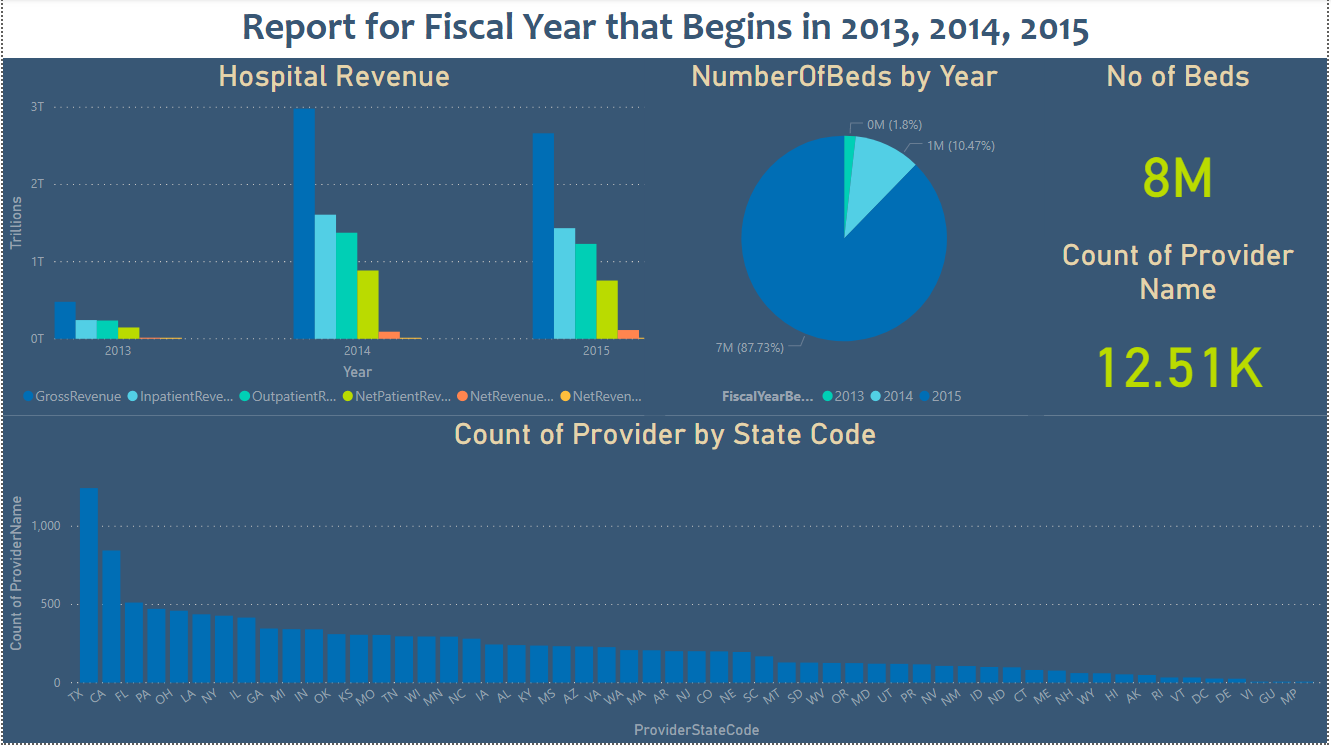


* State Code

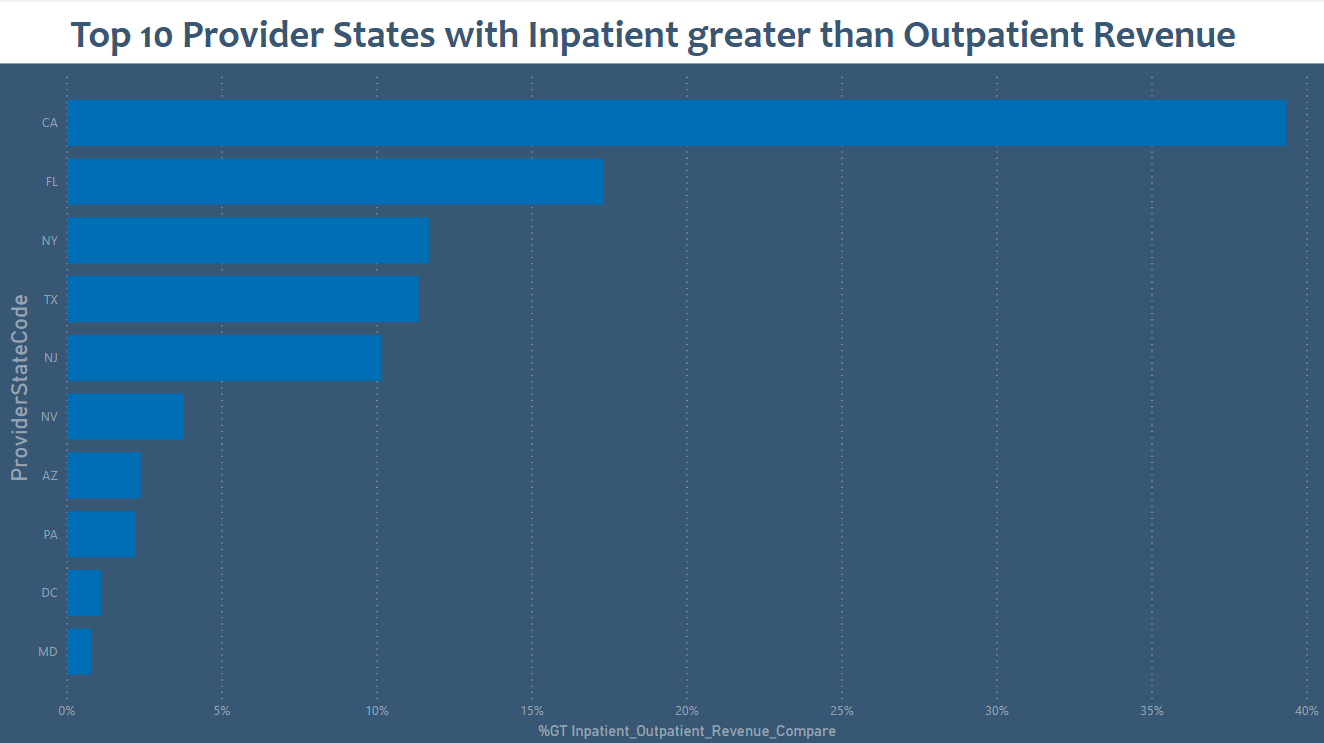


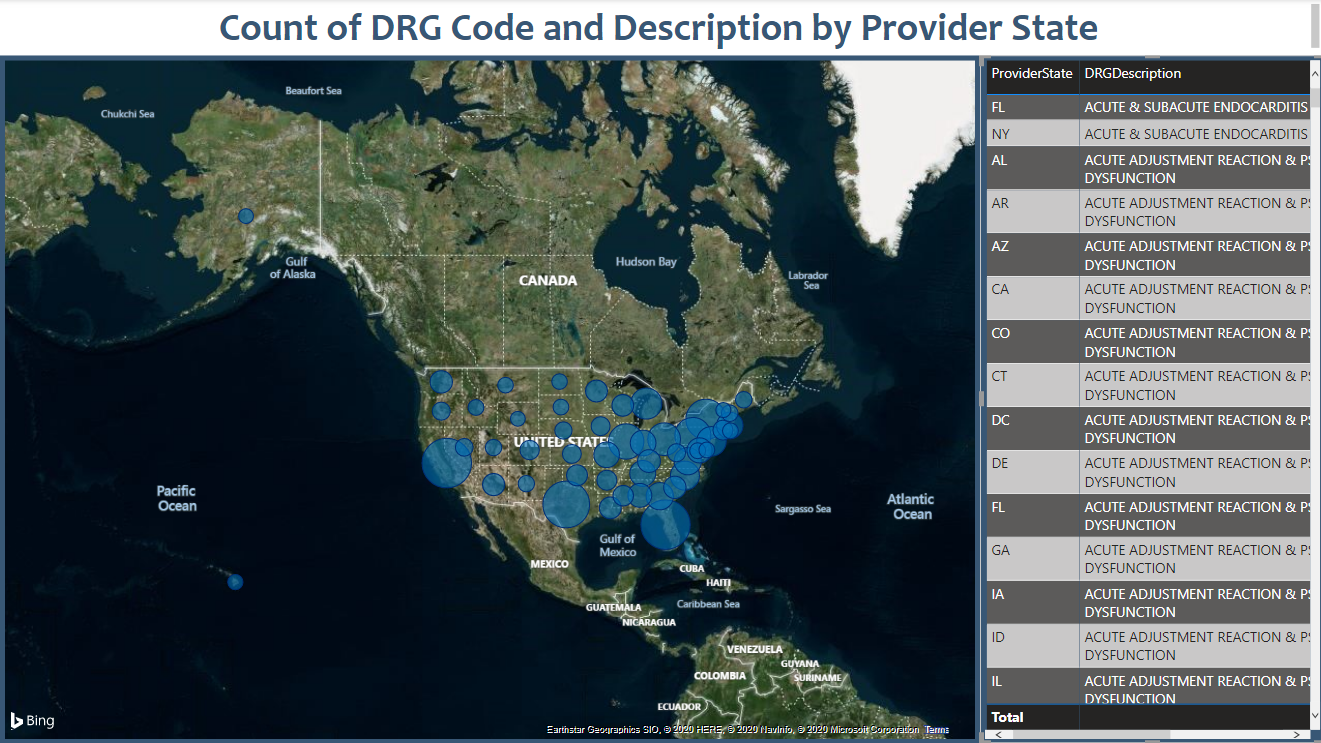
* Apart from this, fields in Provider Cost report required to exclude certain characters such as “$” and “,” to make the data uniform.
* All the character fields were converted to UpperCase
* CamelCase was used as naming convention
* Zip Code, Provider CCN, APCCode and DRGCodes were kept as varchar to maintain data integrity of the codes.

# Visualizations











Future Scope

* Move flat file loading completely to cloud to implement all cloud based infrastructure
* Improve dimensional model by adding dimTime and dimGeography