Northeastern University
College of Engineering

CMS Hospital Management Data Warehouse

Group 15

Nikita Pai – 001347892
Praveen Jayasankar – 001058703
Vishal Baliga – 001027278
Yash Khokale – 001054321

Introduction

The Centres for Medicare and Medicaid Services (**CMS**) is a federal agency within the United States Department of Health and Human Services (HHS) to administer analysis over data produce research reports on hospital, physicians, cases etc.

With such broad motive of the organization, the shortlisting of the topics and clubbing them has been done in such a way that our scope revolves around the following topic:

- We try to narrow down how readily available was each state/city/zip code/FIPS when the whole country was suffering with the uncertainty caused by the pandemic.
- Top ranked hospital and their speciality
- Rating wise evaluation of the hospitals present and their type of organization
- Physicians and their primary speciality, and number of physicians per hospital/state/city/zip code/FIPS

Our project is based on the datasets provided by Centers for Medicare and Medicaid Services (CMS) as well as external data sources like Kaggle. The entire warehousing process and analysis will solely be concerned with the medical facilities, types of hospital, associated physicians available along with their speciality and whether the organization in the state have enough facilities and physician available for the service, especially during the crisis caused by covid-19. The population datasets are used for additional analysis.

Descriptions of the 5 considered datasets are given below -

The first two datasets below give us information about the hospitals and physicians who are licensed to work by CMS.

Dataset 1 - General Hospital

This dataset contains a list of all hospitals that have registered with Medicare across the US. It contains a total of 5315 rows. It also gives overall rating of the hospitals including comparison metric with regard to mortality rate, safety of care etc. in order to compare a particular hospital with the national average of the metric. Column information is briefly explained below:

- The FacilityID which is unique for each hospital
- The Name, Address, City, State, Phone number, Zip code and county name where the facility is located
- The Hospital types. Ex: Psychiatric, Acute care hospital etc. and also if it provides emergency services (binary)
- Type of ownership Ex: Government – State or federal, Proprietorship
- Hospital Overall Rating
- National Comparisons – Mortality, Safety of care, Readmission, Patient Experience, Effectiveness of Care, Timeliness of care and efficient use of medical imaging

Dataset 2 - Physician Dataset

The dataset is a .csv file containing information about the all the licensed CMS physicians in the United States. Each one of these physicians are assigned a unique NPI ID by CMS and one physician has the liberty to work in more than one facility. This dataset contains about 2.18 million rows which really gives a vast amount of information about the physicians, including their current/past information about education, address(s),

professional practice etc. We will be cleaning the data in such a way that just the current information remains distinct with data profiling.
The dataset contains hospital affiliation CCN (CMS certification number) and LBN (Legal Business Name) which technically gives us the FacilityID and Facility name where the physician practices. Column information is briefly explained below.

- Unique NPI (National Provider Identifier) ID
- PACID (Pecos Associate Control ID), Professional Enrolment ID
- Details about Name and gender
- Medical school, graduation year and primary specialty
- Group practice legal name, PACID, Address, City, State, Zip code and phone number
- Hospital Affiliation's CCN and LBN

Dataset 3
US Population by FIPS/STATE/Area Dataset

The US Population gives the number of population density for the constricted region according to FIPS as well as State along with Cumulative Birth and Death in 2019. This Dataset would help us given an idea about the overall population in that region. Along with that comparison between death in 2019 and death in 2020 can also be done thereby analysing the degree by which the mortality rate has increased because of COVID-19.

Dataset 4
COVID-19 Dataset Up to Nov 14th
This dataset contains information about the confirmed covid-19 cases county-wise as reported by Johns Hopkins University Centre for Systems science and Engineering. With the wake of the pandemic, it is quintessential for medical facilities to be brace the higher influx of patients. This data will help us evaluate the readiness of CMS medical facilities with the increasing covid-19 cases. Brief description of the dataset is given below-

- Update date and time
- County Name and FIPS code
- Population
- Confirmed cases (Cumulative) and confirmed per 100000
- Deaths (Cumulative) and deaths per 100000

Dataset 5
Lookup Dataset (State, City, FIPS, Zipcode)

The dataset has Zip code, County, FIPS Code, State. This data has been evaluated and conforms to quality dataset. We decided to use this data as a look up in such a way that all the matching record will be sent to data warehouse and the unmatched record will be sent to reject table.

Joining different Dataset:
The point commonality between the dataset are as follows:
General joining factors:

- State
- County
- Zip code
- FIPS code

Specific joining factors:

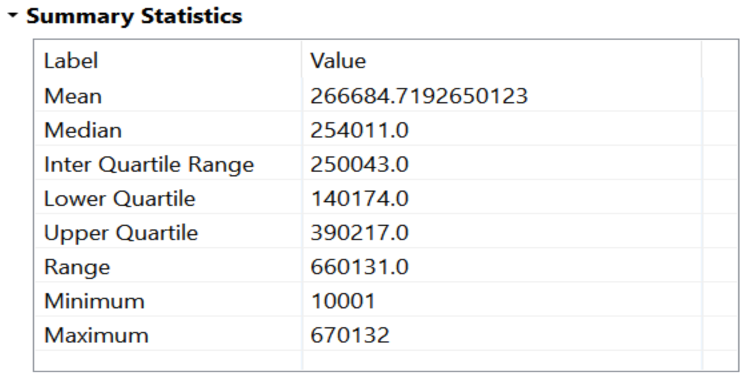
- Hospital Information Dataset (Facility ID) and Physician Dataset (Hospital affiliation CNN)

If you refer the Dimensional model, you will notice that Dim_Hospital and Dim_Physician is being joined by CCN1 and Facility ID.

Data Profiling

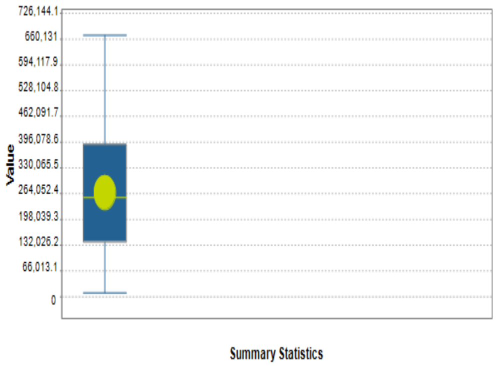
We are using Talend for column profiling that would provide statistical measurements associated with the frequency distribution of data values(patterns) within all the columns present in all the dataset.

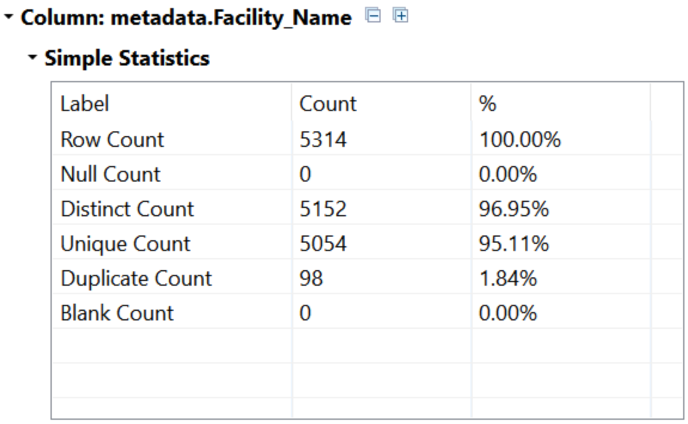
The columns comprise of Hospital dataset with the Column being highlighting the distinct counts, null values, and Duplicate values. We have implemented the analysis on all attributes mentioned to help us figure out how dirty the data is.



Summary Statistics

Label	Value
Mean	266684.7192650123
Median	254011.0
Inter Quartile Range	250043.0
Lower Quartile	140174.0
Upper Quartile	390217.0
Range	660131.0
Minimum	10001
Maximum	670132

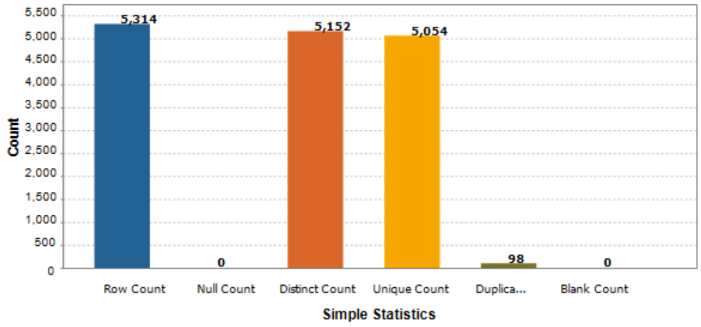


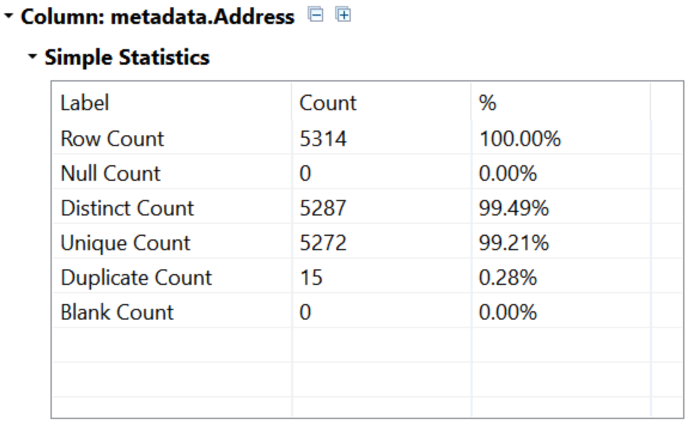


Column: metadata.Facility_Name

Simple Statistics

Label	Count	%
Row Count	5314	100.00%
Null Count	0	0.00%
Distinct Count	5152	96.95%
Unique Count	5054	95.11%
Duplicate Count	98	1.84%
Blank Count	0	0.00%

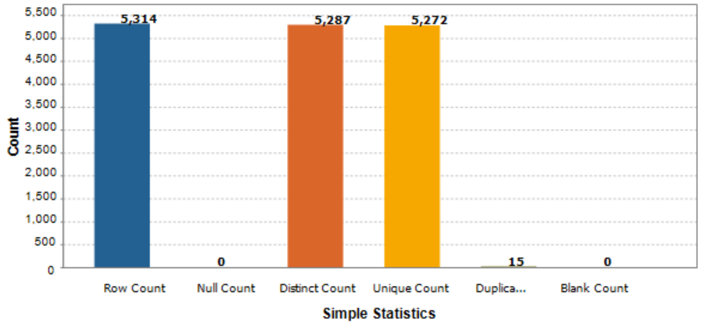


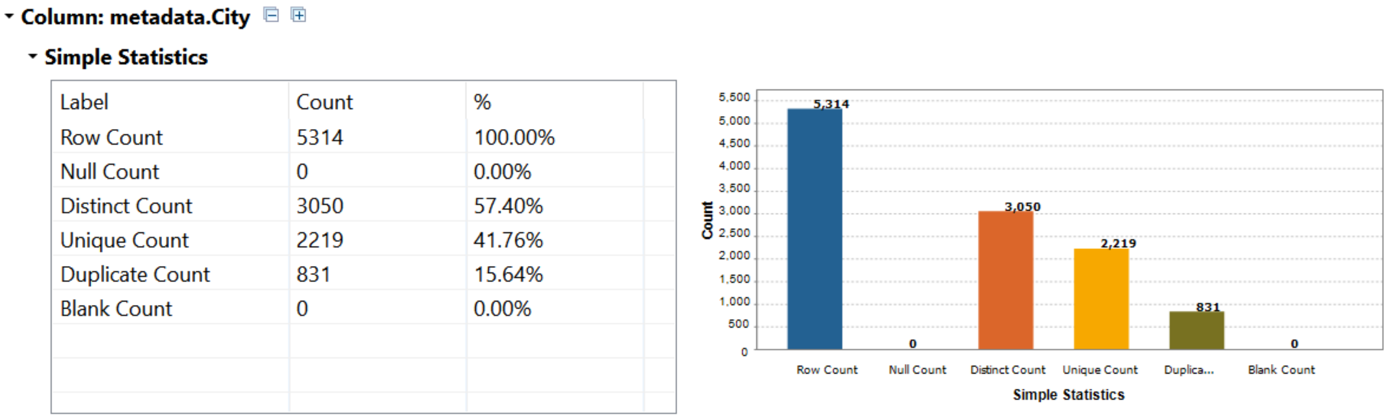


Column: metadata.Address

Simple Statistics

Label	Count	%
Row Count	5314	100.00%
Null Count	0	0.00%
Distinct Count	5287	99.49%
Unique Count	5272	99.21%
Duplicate Count	15	0.28%
Blank Count	0	0.00%

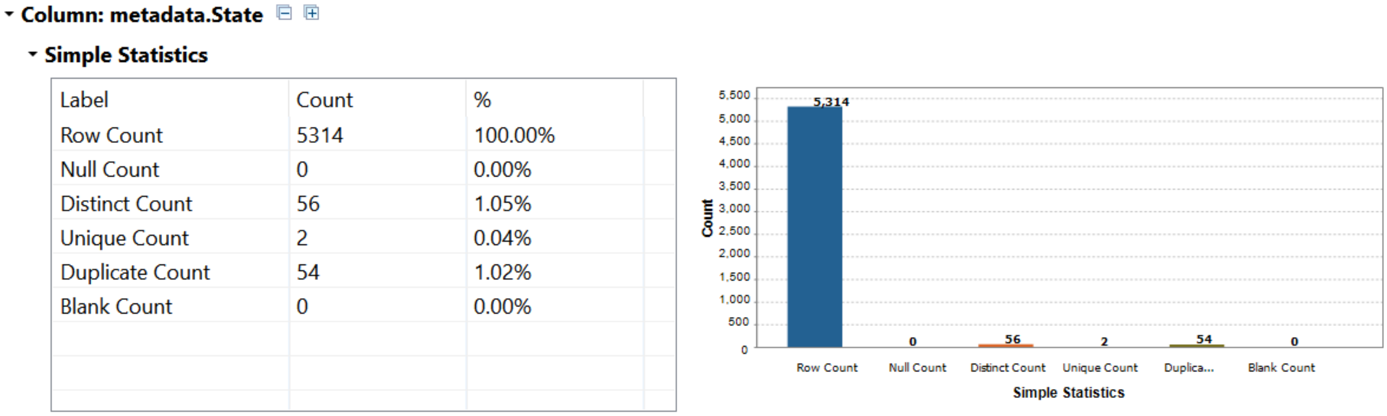




Column: metadata.City

Simple Statistics

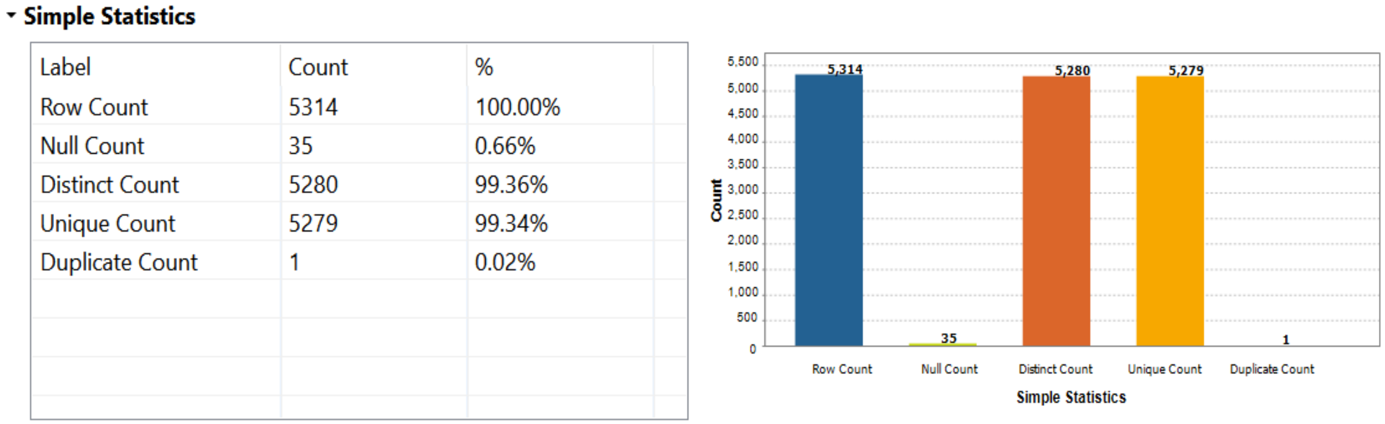
Label	Count	%
Row Count	5314	100.00%
Null Count	0	0.00%
Distinct Count	3050	57.40%
Unique Count	2219	41.76%
Duplicate Count	831	15.64%
Blank Count	0	0.00%



Column: metadata.State

Simple Statistics

Label	Count	%
Row Count	5314	100.00%
Null Count	0	0.00%
Distinct Count	56	1.05%
Unique Count	2	0.04%
Duplicate Count	54	1.02%
Blank Count	0	0.00%



Simple Statistics

Label	Count	%
Row Count	5314	100.00%
Null Count	35	0.66%
Distinct Count	5280	99.36%
Unique Count	5279	99.34%
Duplicate Count	1	0.02%

Data Transformation

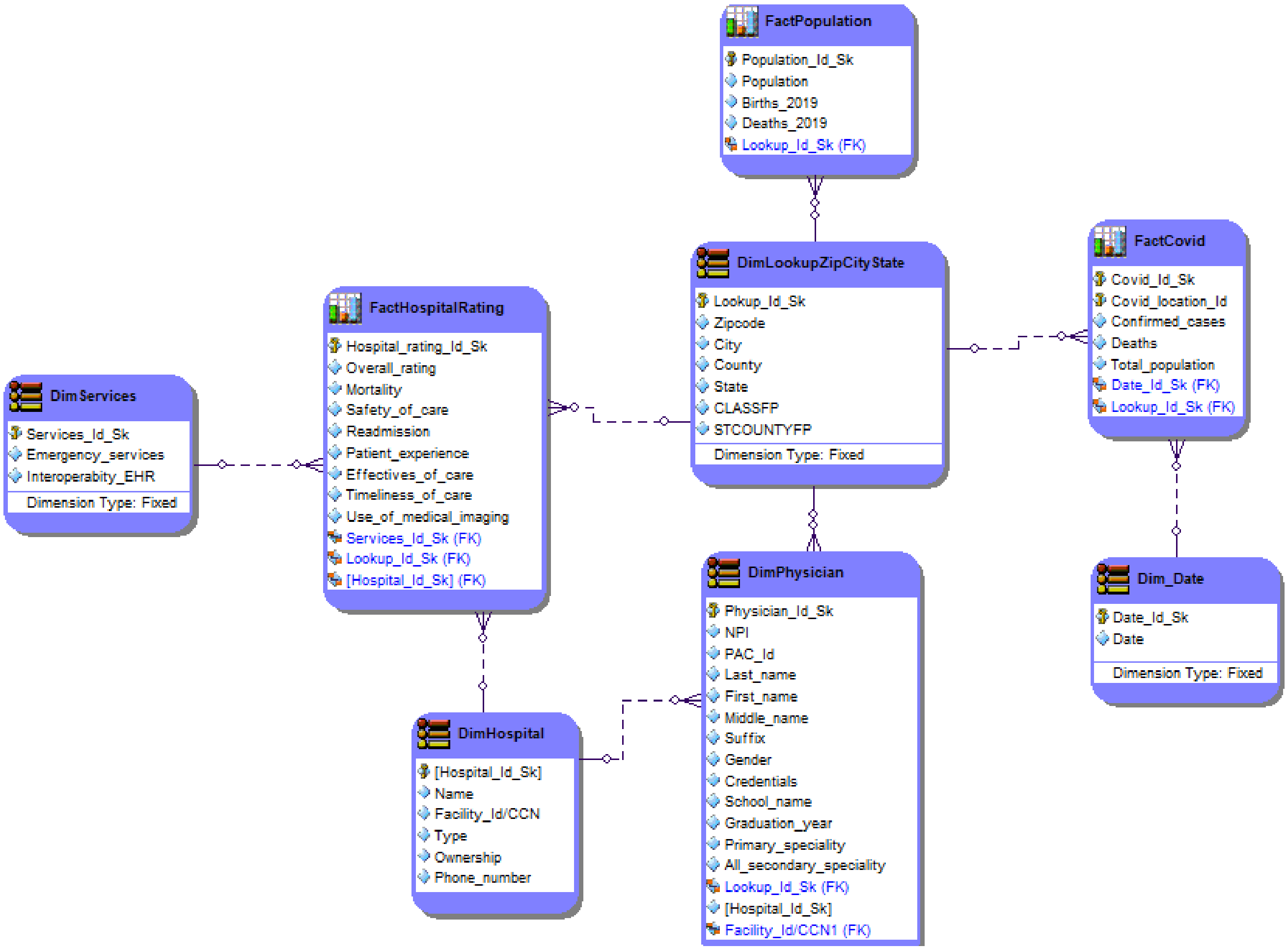
Data has been converted into the appropriate datatype after being loaded in the Staging. To avoid NULL values, all the datatype containing **int** converts the i/p into **–1** and then passes it on. All the attributes consisting of varchar datatype converts the NULL values to Unknown.

By scanning the dataset, we have seen that a lot of attributes have NULL values through Data Profiling in Talend (“Hospital overall rating,” “Mortality national comparison,” “Safety of care national comparison”, “Mortality national comparison,” “Safety of care national comparison,” “Readmission national comparison” etc.)

Dimensional Model

All the packages, DDL Statements, Dimensional Table and Datasets can be accessed here –

https://northeastern.sharepoint.com/:u:/s/Proposal-DWBI/Edm4-1cfSo9CvQ05c_ITLwgBiE2RlSlOTaC0Ecvc_yqCyQ?e=OALmOM



The Dimensional model has the following dimensions tables:

1. Dim_Date
2. Dim_Hospital
3. Dim_Lookup_Zip_City_State
4. Dim_Physician
5. Dim_Services

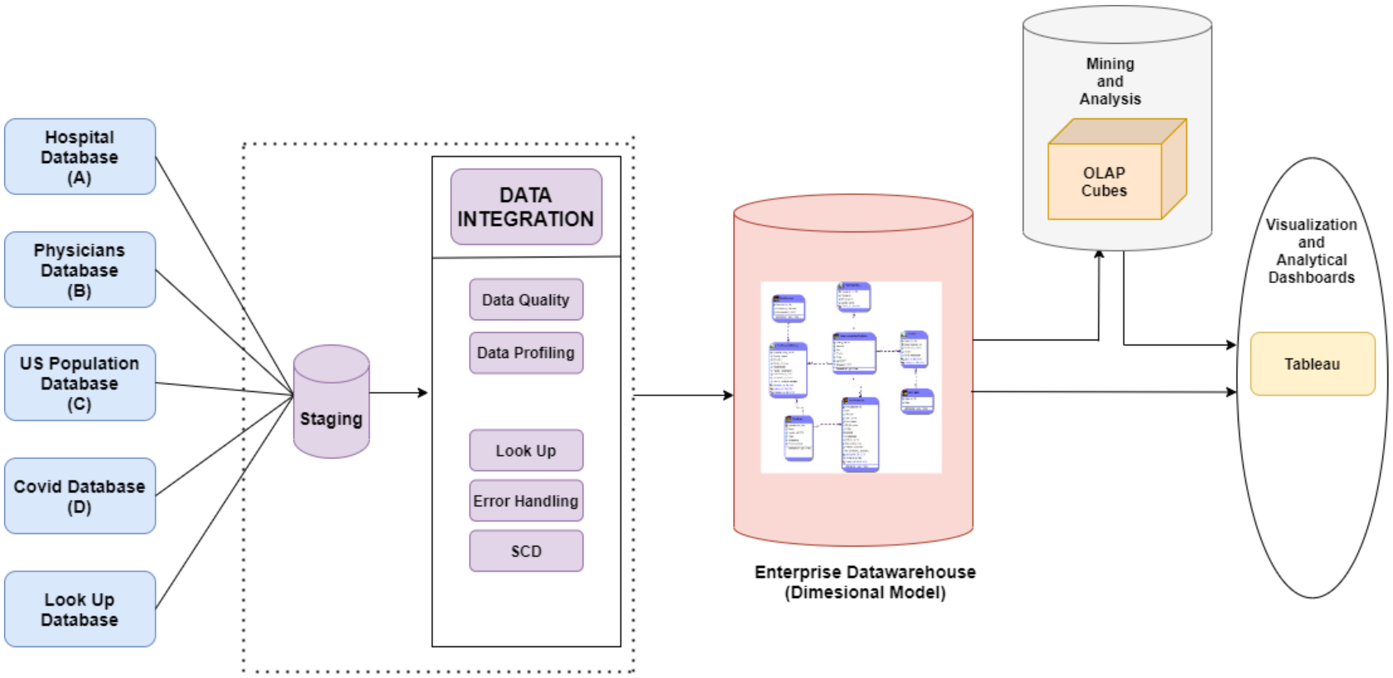
Fact tables:

1. Fact_Population
2. FactCovid
3. FactHospitalRating

Staging tables:

1. Stg_Covid
2. Stg_Hospital
3. Stg_Physician
4. Stg_Population
5. Stg_Zip_City_County_lookup

Diagram of the Dataflow



Workflow Steps:

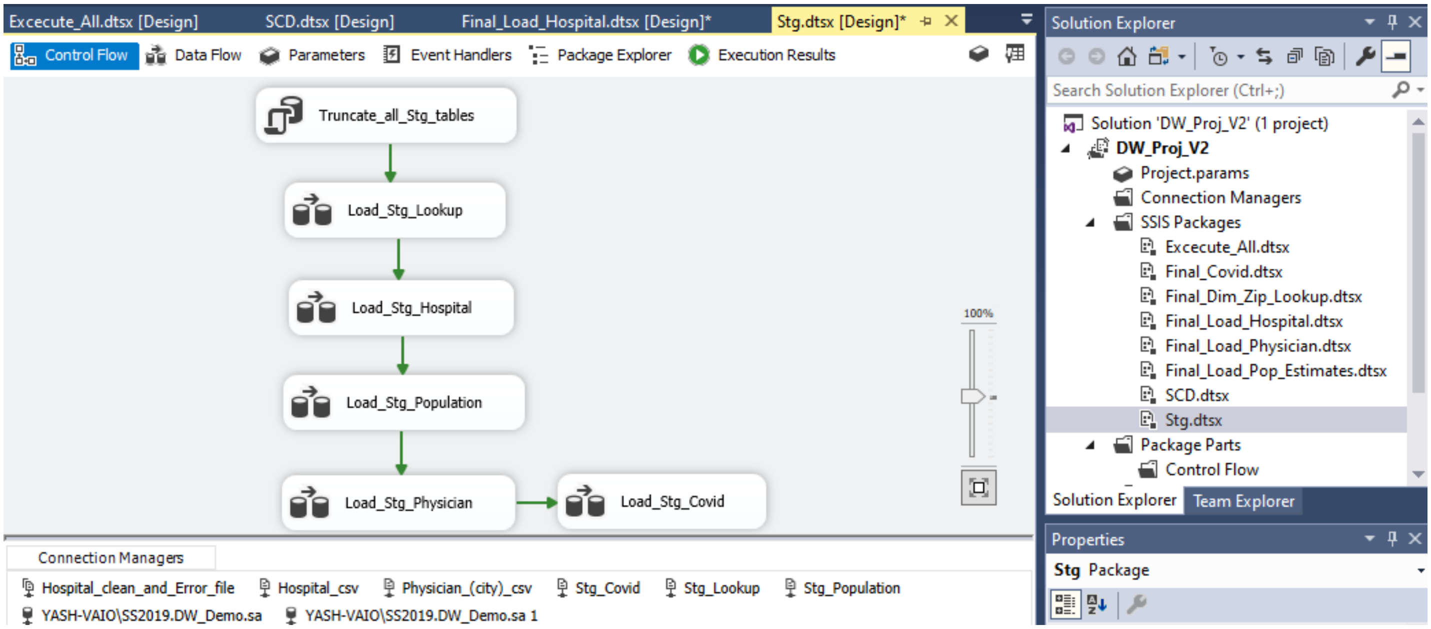
- The five datasets selected for the DW are loaded into the staging area initially with all attributes having varchar datatypes
- Data loaded into the staging tables undergo data integration tasks like data profiling using Talend software, data quality using SQL server data quality client, error handling.
- On completion of data integration process data is being loaded into the dimensional model using the staging tables to maintain referential integrity, following Kimball methodology.
- Further the OLAP cube is built for reporting using covid, population, Hospital and Zip/city/fips/state lookup datasets. This cube is built to answer the relation between the covid cases and various factors like number of hospitals, population in particular region.
- The final dimensional DW model is visualized using Tableau tool to justify the use of DW and answer various questions.

Aim: Enterprise-wide repository of disparate data sources (Hospital Database, Physicians Database, Population Database, Demographics and Cost, Covid cases)

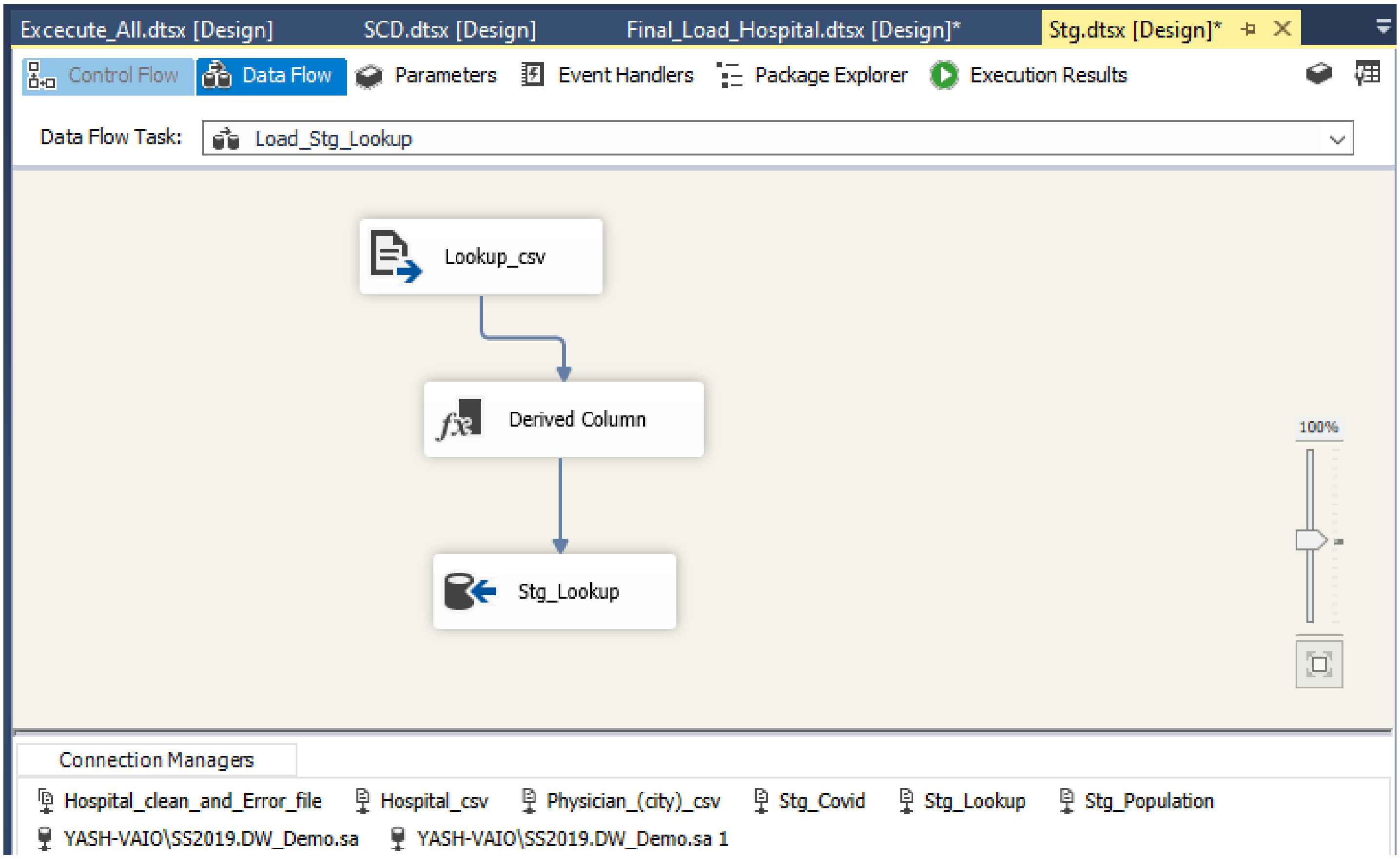
SSIS Package:

Loading the Stage tables:

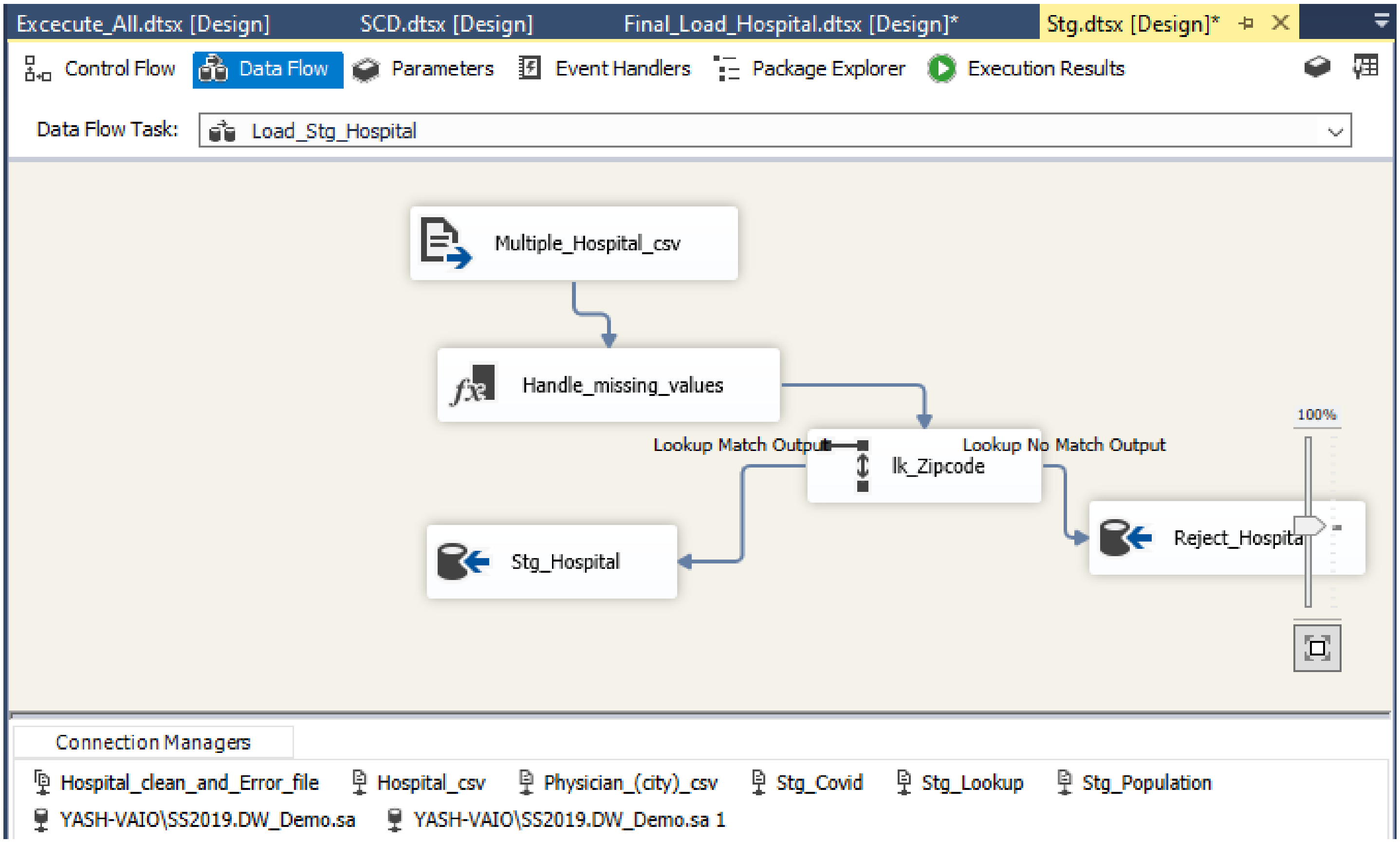
As shown in the picture below all the 5 datasets are loaded into their respective staging tables using varchar data type.



The staging table for Zip/city/fips/state lookup dataset is loaded initially so that it could used as a reference while loading other staging tables.



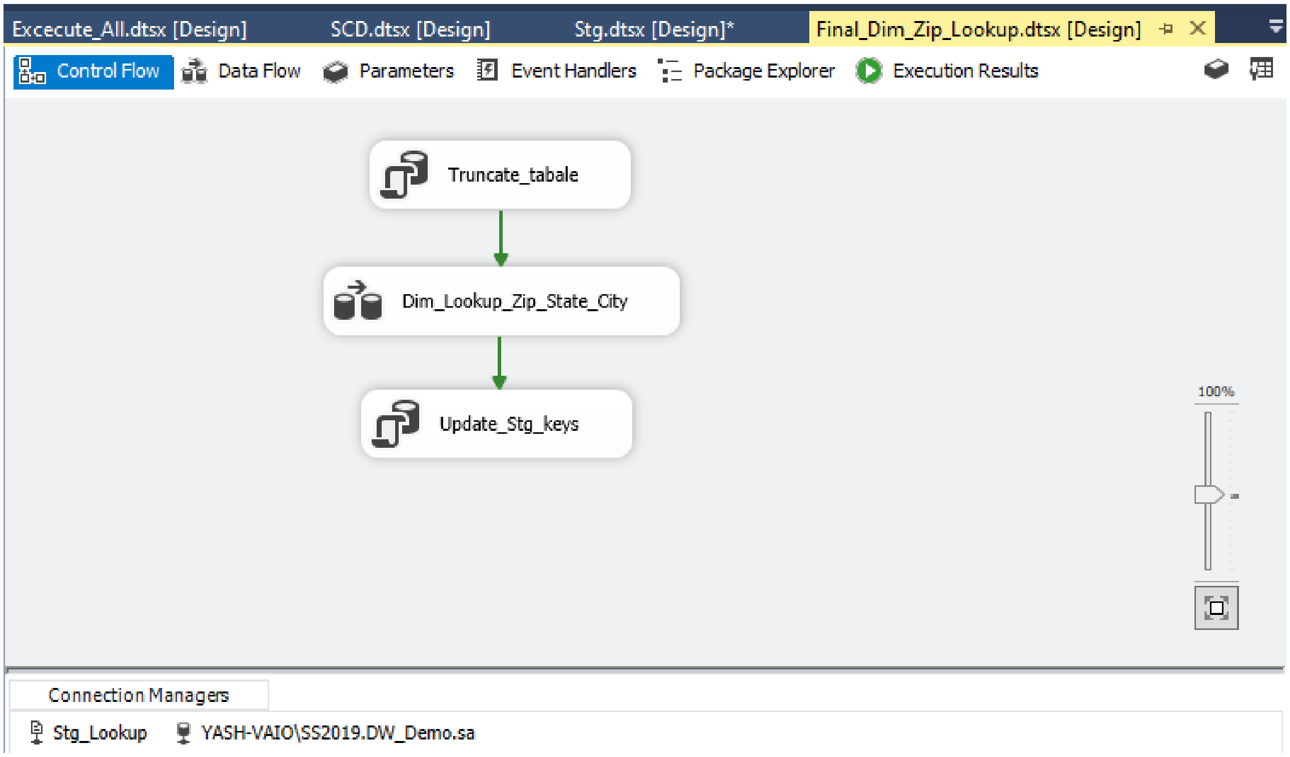
The staging table for Hospital dataset is loaded using previously loaded Stg_Lookup table as reference based on zip code i.e. if the zip code in the hospital dataset matches to zip code present in lookup table then it will be loaded into the staging table.



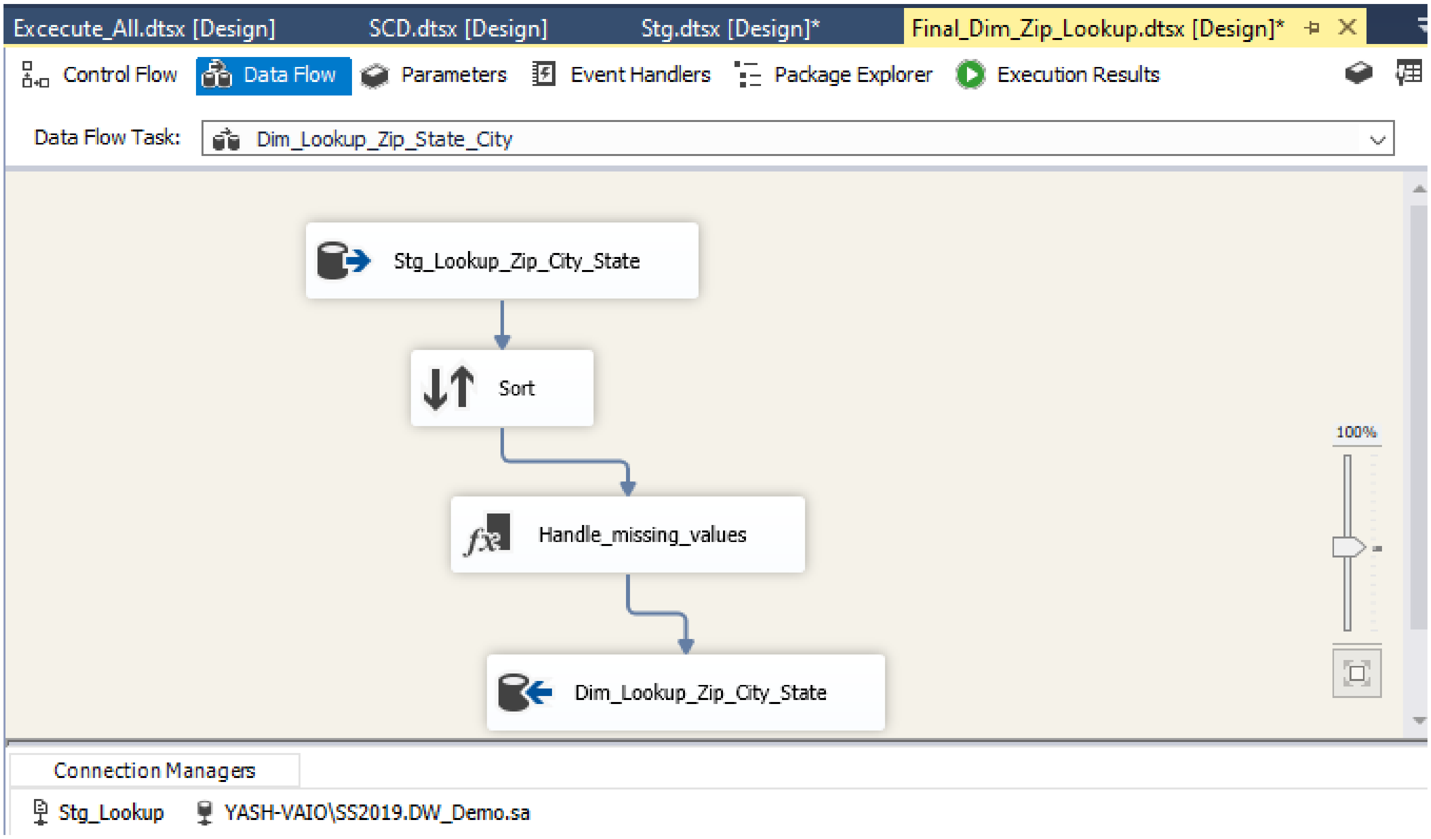
Similarly, all the staging tables are loaded using Stg_lookup table as reference.

Loading Zip/City/State/Fips lookup dimension:

Further, Dim_lookup dimension is loaded using its corresponding staging table.

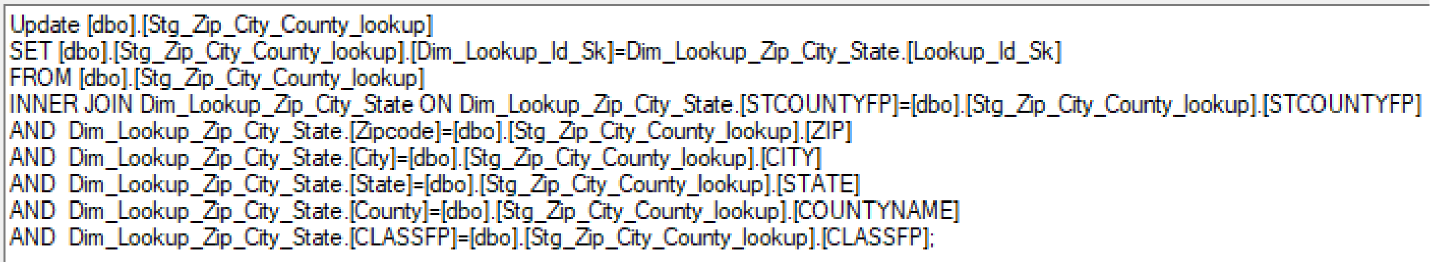


Before the data is loaded into the dimension all the null or missing values are replaced with “Unknown string” for varchar data type and “-1” for integer data type.



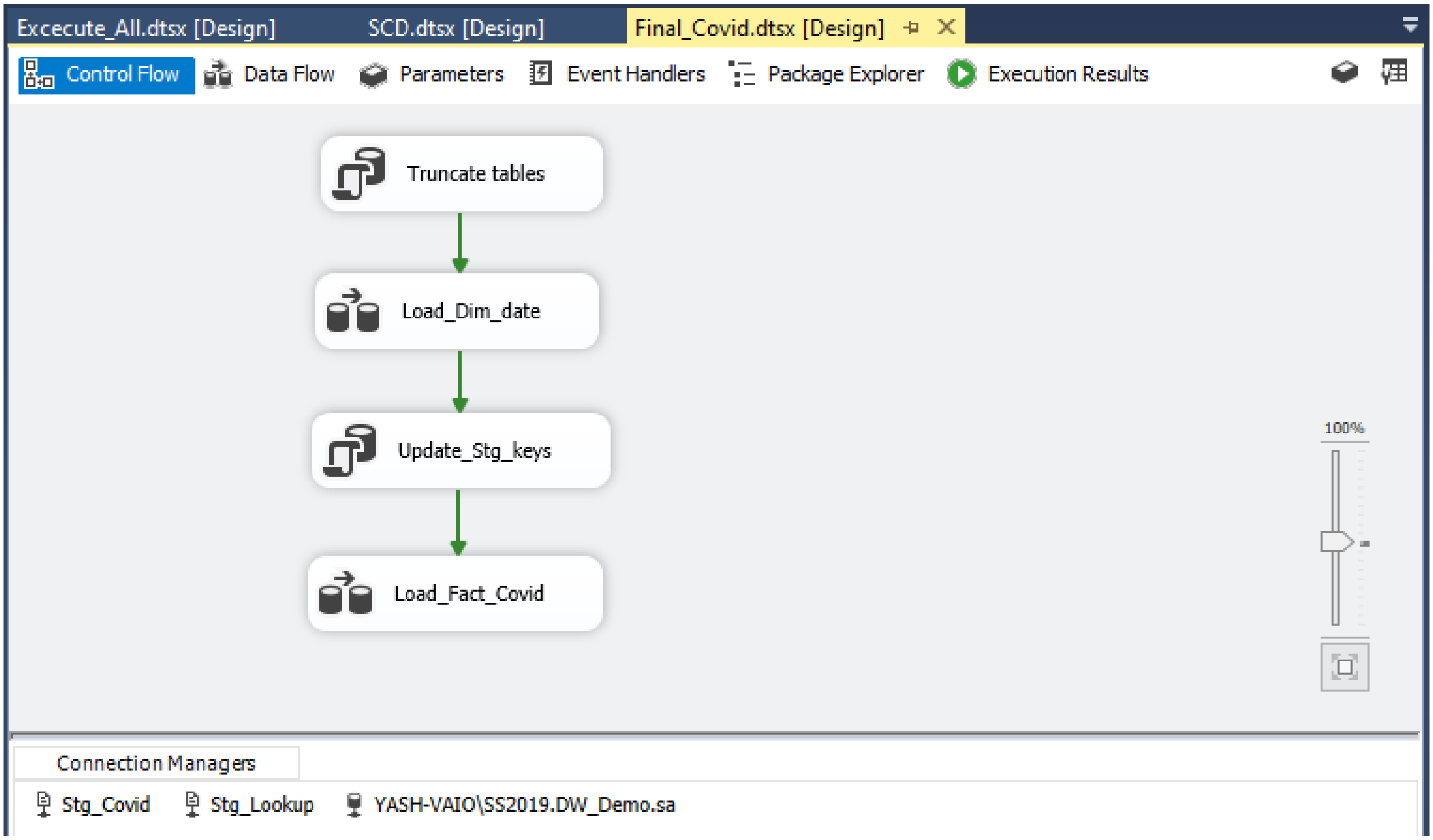
Data Flow Task: Dim_Lookup_Zip_State_City

Then the staging table is updated with foreign key Lookup_Id_Sk, which is the primary key of the lookup dimension.

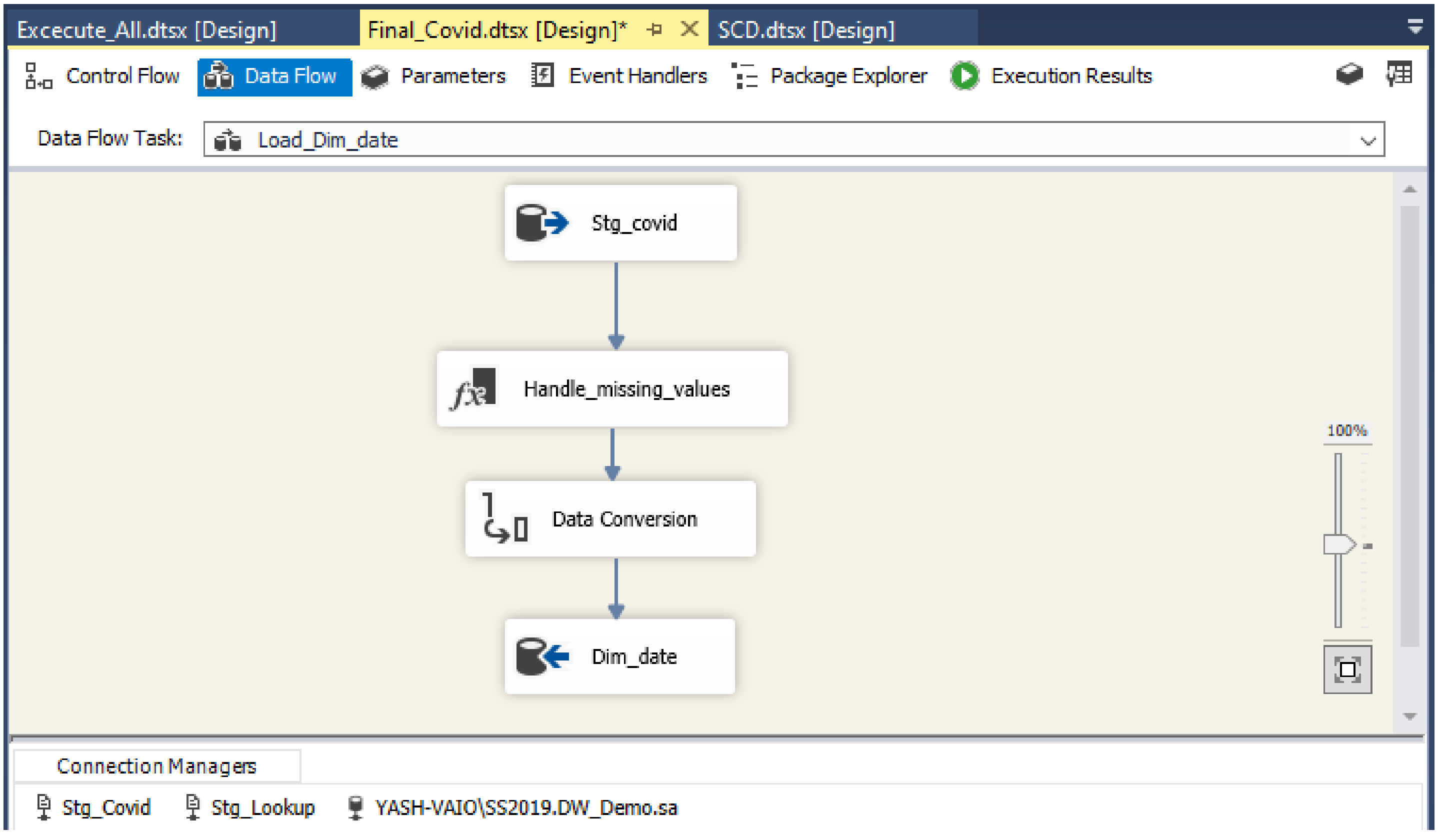


```
Update [dbo].[Stg_Zip_City_County_lookup]
SET [dbo].[Stg_Zip_City_County_lookup].[Dim_Lookup_Id_Sk]=Dim_Lookup_Zip_City_State.[Lookup_Id_Sk]
FROM [dbo].[Stg_Zip_City_County_lookup]
INNER JOIN Dim_Lookup_Zip_City_State ON Dim_Lookup_Zip_City_State.[STCOUNTYFP]=[dbo].[Stg_Zip_City_County_lookup].[STCOUNTYFP]
AND  Dim_Lookup_Zip_City_State.[Zipcode]=[dbo].[Stg_Zip_City_County_lookup].[ZIP]
AND  Dim_Lookup_Zip_City_State.[City]=[dbo].[Stg_Zip_City_County_lookup].[CITY]
AND  Dim_Lookup_Zip_City_State.[State]=[dbo].[Stg_Zip_City_County_lookup].[STATE]
AND  Dim_Lookup_Zip_City_State.[County]=[dbo].[Stg_Zip_City_County_lookup].[COUNTYNAME]
AND  Dim_Lookup_Zip_City_State.[CLASSFP]=[dbo].[Stg_Zip_City_County_lookup].[CLASSFP];
```

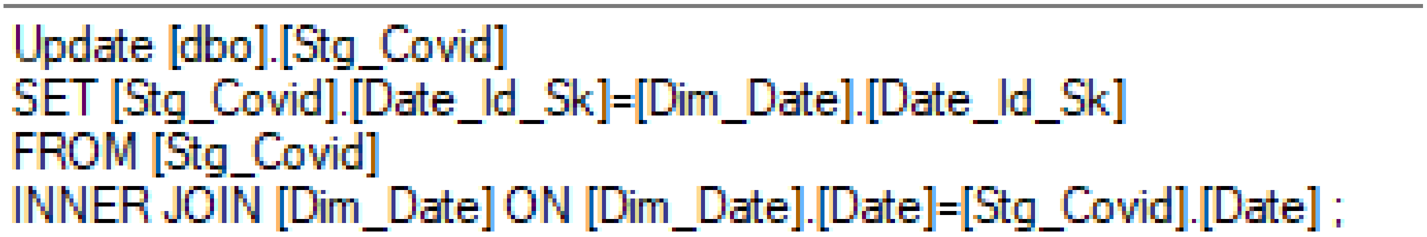
Loading covid data:
Dimension relating to covid data is loaded initially followed by its fact table, as shown below.



Dim_date dimension for covid data is loaded initially and null or missing dates are replaced with "2099/12/30" format.

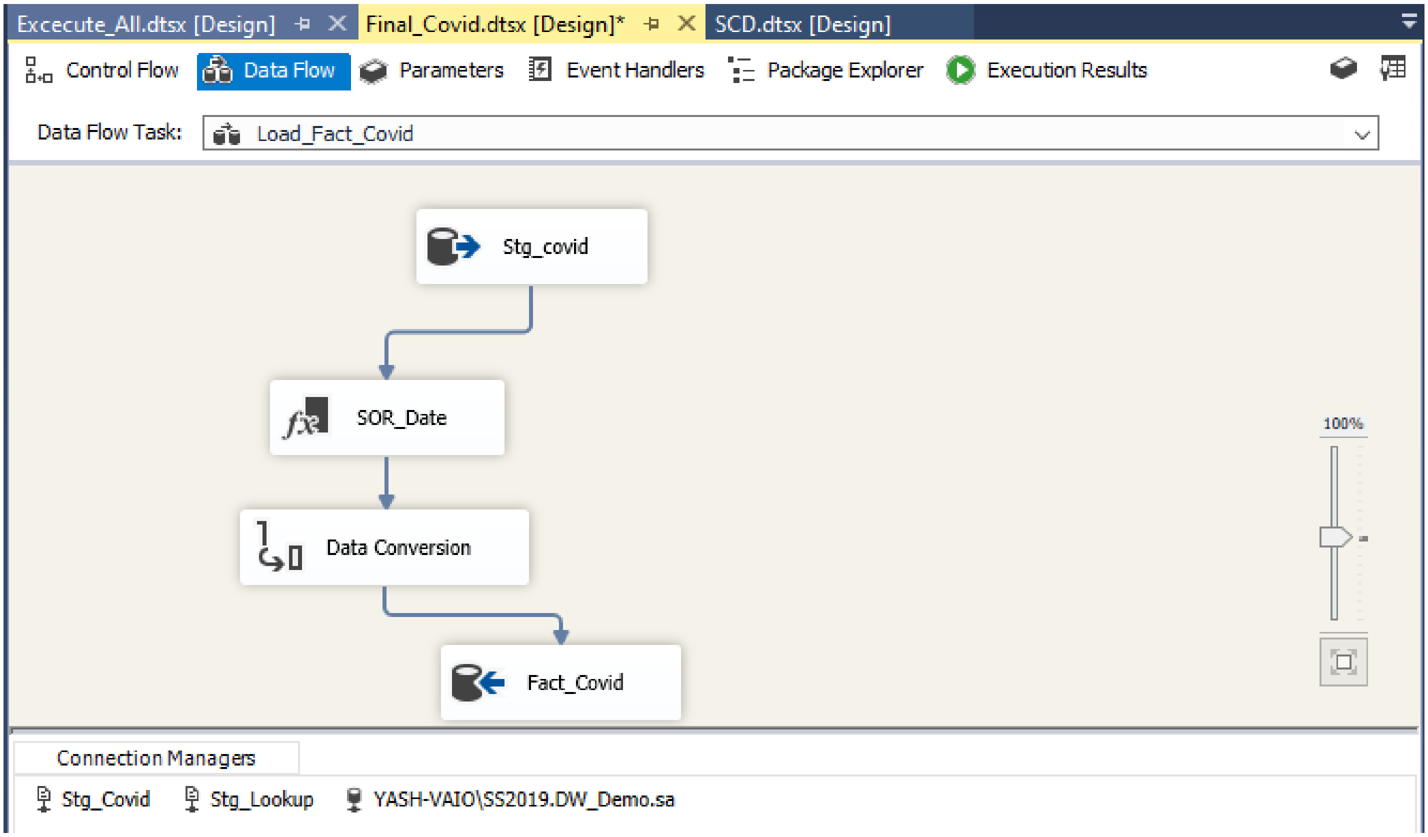


Staging table is updated for foreign key Date_id_Sk, which is primary key for the Dim_Date.



```
Update [dbo].[Stg_Covid]
SET [Stg_Covid].[Date_Id_Sk]=[Dim_Date].[Date_Id_Sk]
FROM [Stg_Covid]
INNER JOIN [Dim_Date] ON [Dim_Date].[Date]=[Stg_Covid].[Date] ;
```

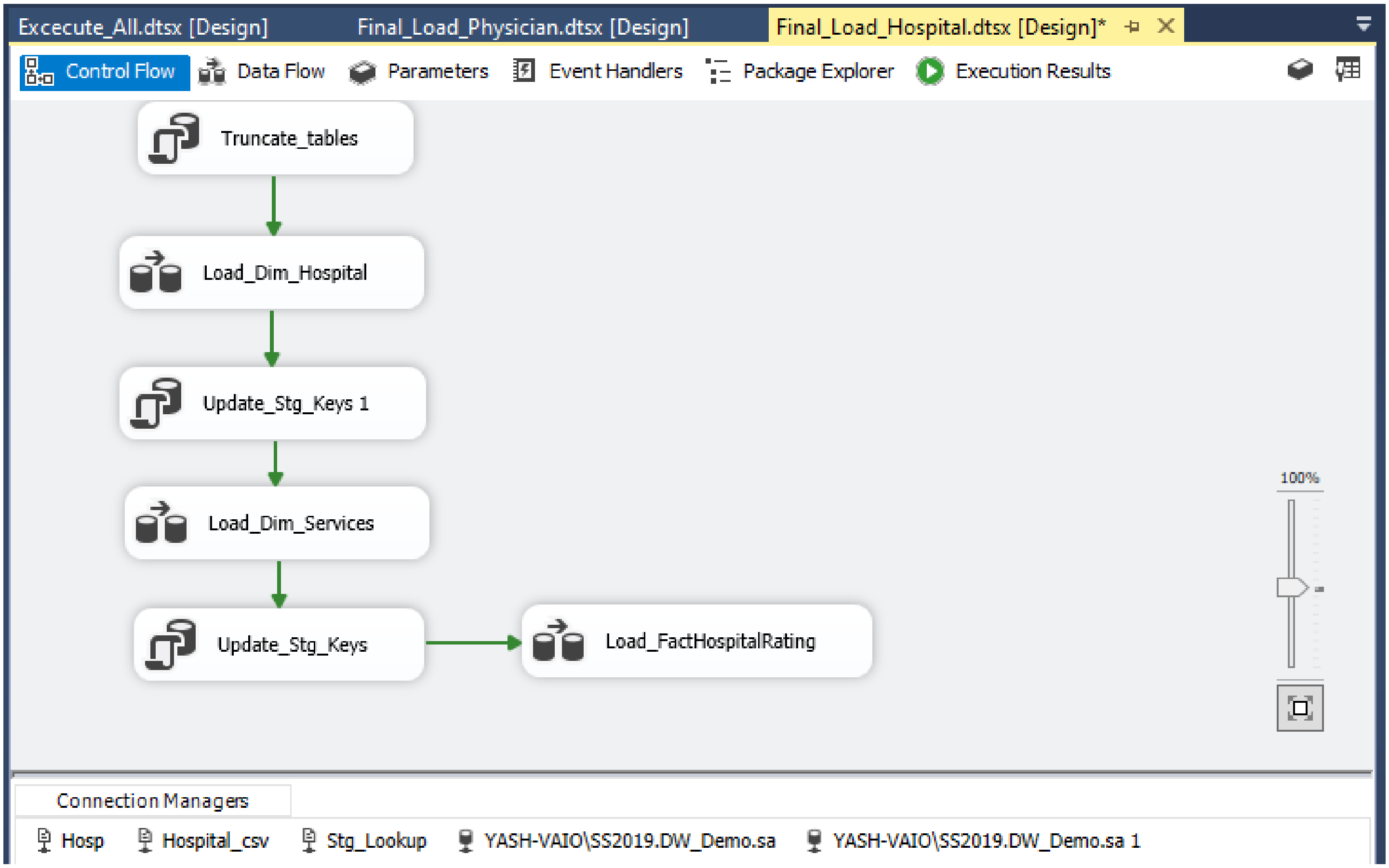
Then all the required numeric attributes with the foreign keys previously updated in the staging table are loaded into Fact_covid table. And in this manner the Dim_Date table is linked to Fact_covid table.



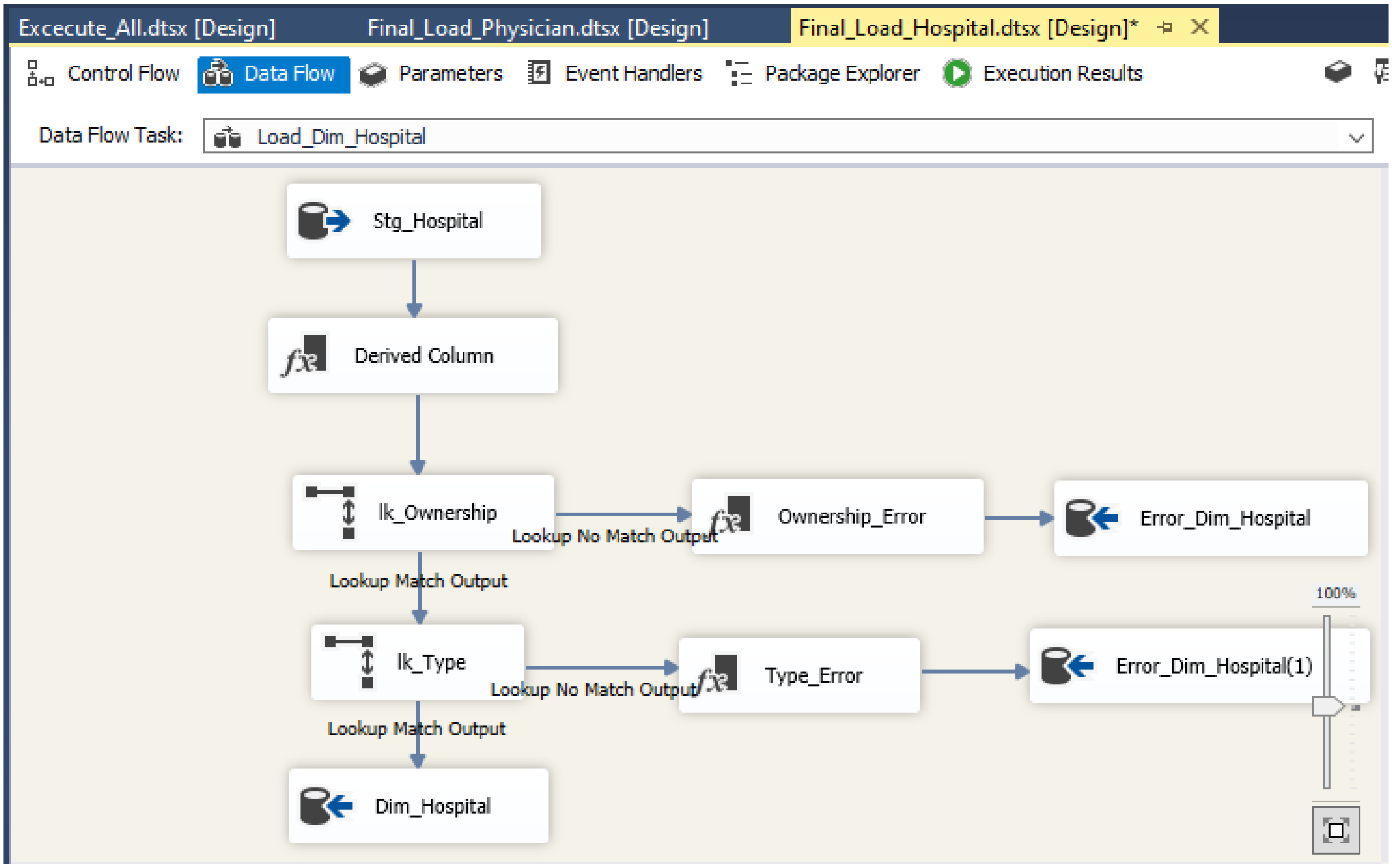
Data Flow Task: Load_Fact_Covid

Error handling:

The error handling is implemented for Hospital dimension, which loaded in similar way as the previous covid data.

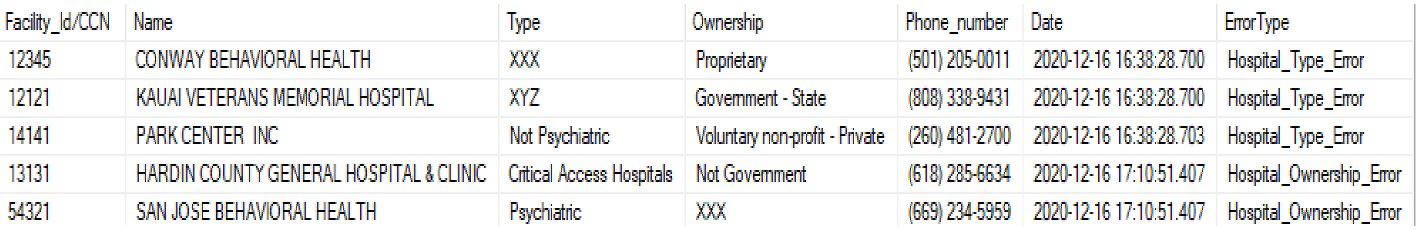


As shown below, while loading the hospital dimension series of error handling is performed. If the hospital ownership data does not match with ownership lookup table then data is inserted into Error table. Similarly, if there is mismatch with hospital Type data, the rows are directed to Error table.



Data Flow Task: Load_Dim_Hospital

Result of the error table hospital dimension. All the invalid hospital types and ownership data is loaded into error table with the Error Type shown below.



Facility_Id/CCN	Name	Type	Ownership	Phone_number	Date	ErrorType
12345	CONWAY BEHAVIORAL HEALTH	XXX	Proprietary	(501) 205-0011	2020-12-16 16:38:28.700	Hospital_Type_Error
12121	KAUAI VETERANS MEMORIAL HOSPITAL	XYZ	Government - State	(808) 338-9431	2020-12-16 16:38:28.700	Hospital_Type_Error
14141	PARK CENTER INC	Not Psychiatric	Voluntary non-profit - Private	(260) 481-2700	2020-12-16 16:38:28.703	Hospital_Type_Error
13131	HARDIN COUNTY GENERAL HOSPITAL & CLINIC	Critical Access Hospitals	Not Government	(618) 285-6634	2020-12-16 17:10:51.407	Hospital_Ownership_Error
54321	SAN JOSE BEHAVIORAL HEALTH	Psychiatric	XXX	(669) 234-5959	2020-12-16 17:10:51.407	Hospital_Ownership_Error

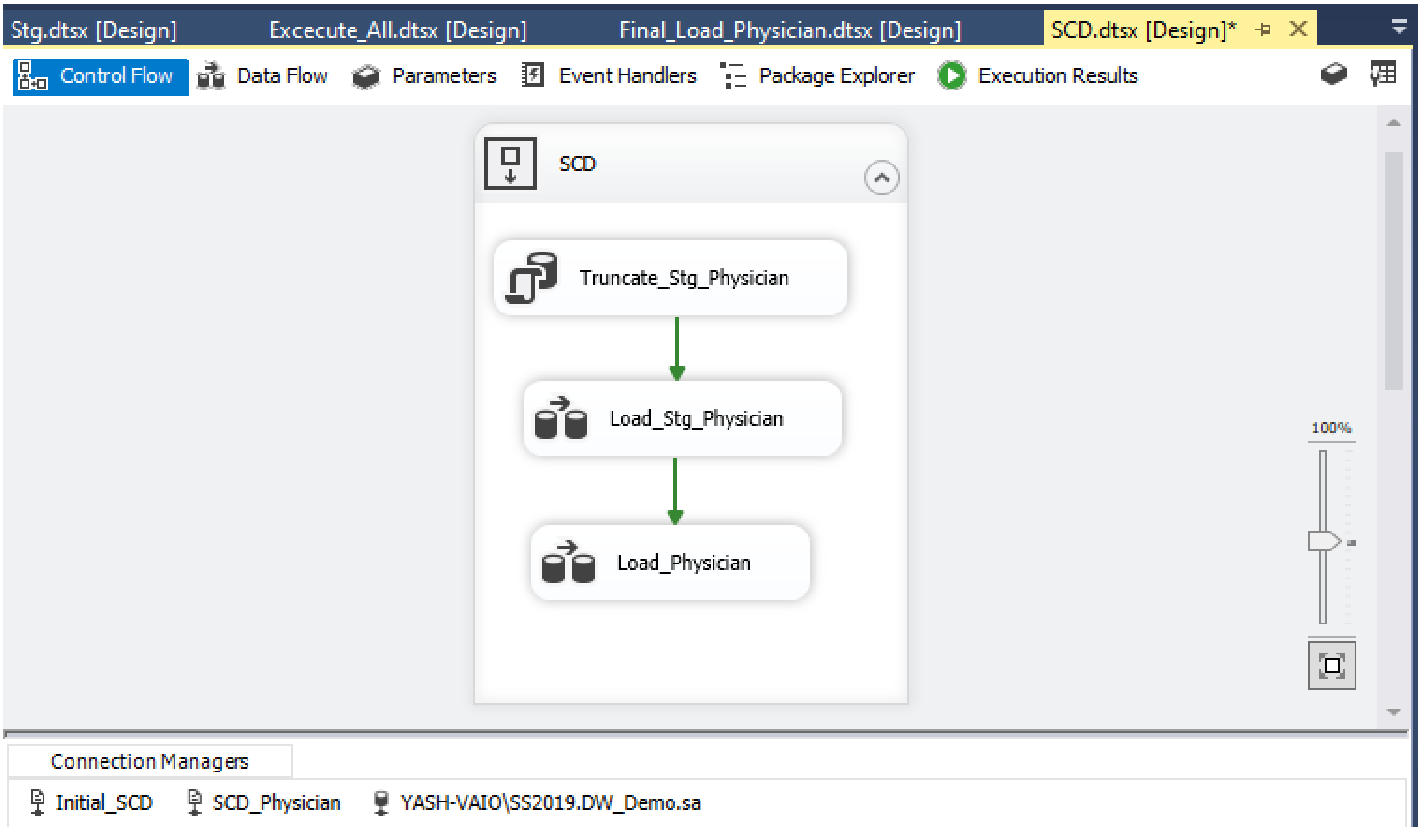
Errors like having string inputs in place of numeric values and having numeric values in place of string inputs are handled for all the fact and dimensional tables. All the null values are replaced by the string “Unknown” so that there no null values in the dataset.

For the Hospital dataset for the Rating column, where the ratings given between 1-5 are only accepted and records having any other numeric value can be discarded as error. And for the other columns with the National comparison, values other than Above Nation Average, Below National Average and Same as National Average are considered to be errors. For Population dataset any gender other Male or Female is labelled as error. For Physician dataset having strings for NPI and PAC ID are labelled as error.

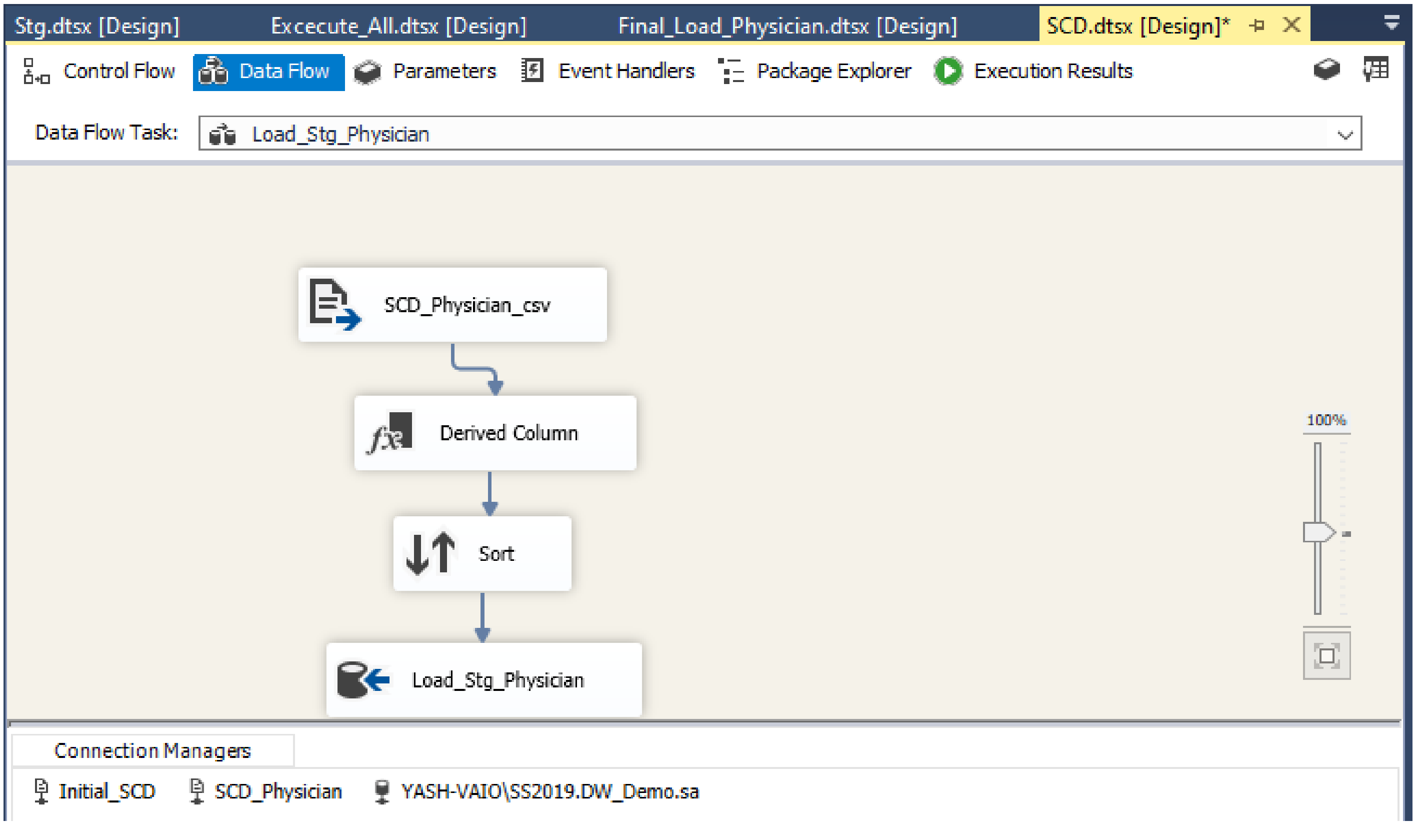
Similarly, all the rest of the fact and dimension tables are loaded using their respective staging.

SCD for Dim_Physician:

Once all the data is loaded into Dim_Physician dimension. To handle the next month's data, the staging table is truncated and loaded with next month's data as show below.



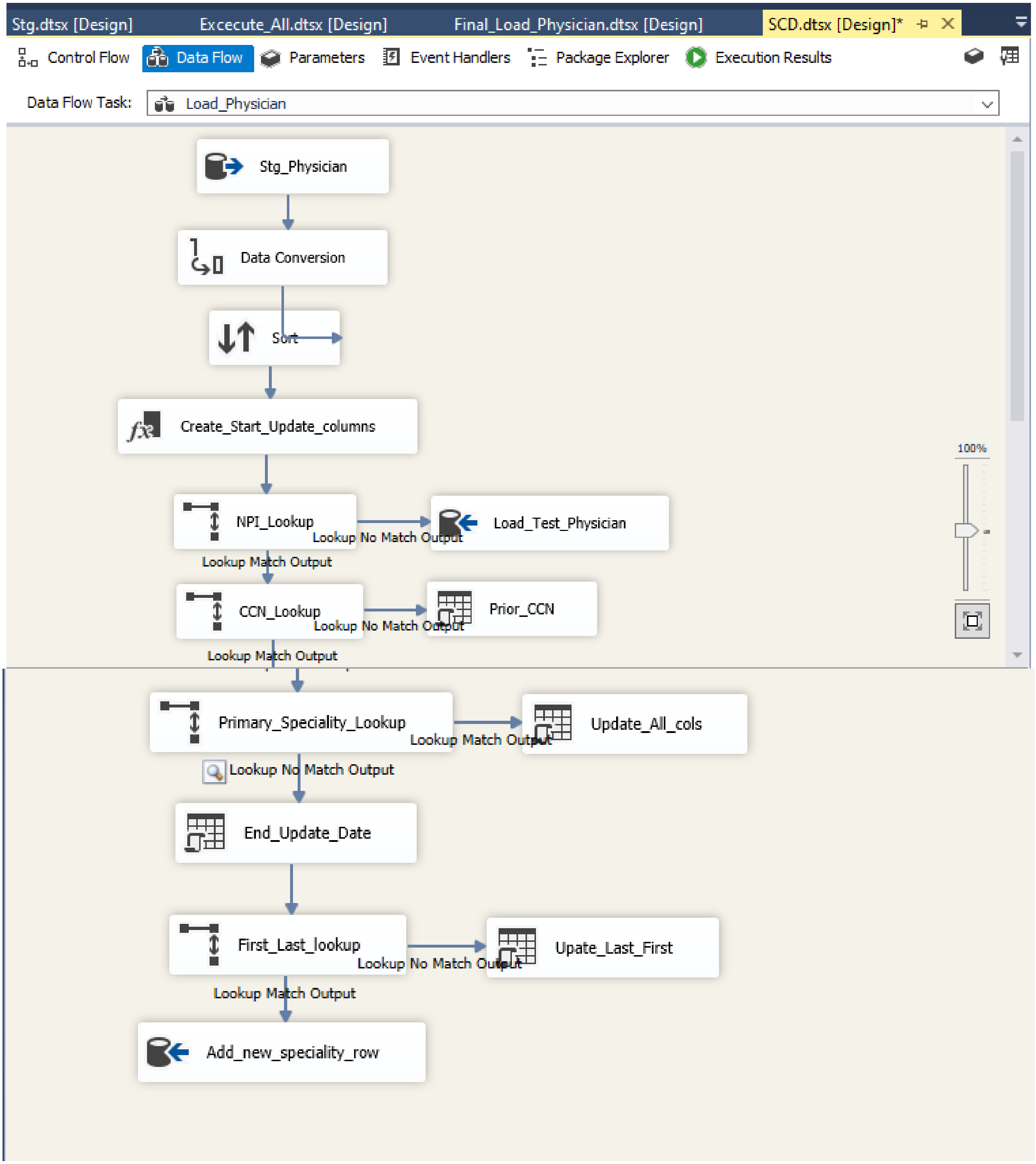
Next month's data is loaded into Stg_Physician from the csv file containing the updated data.



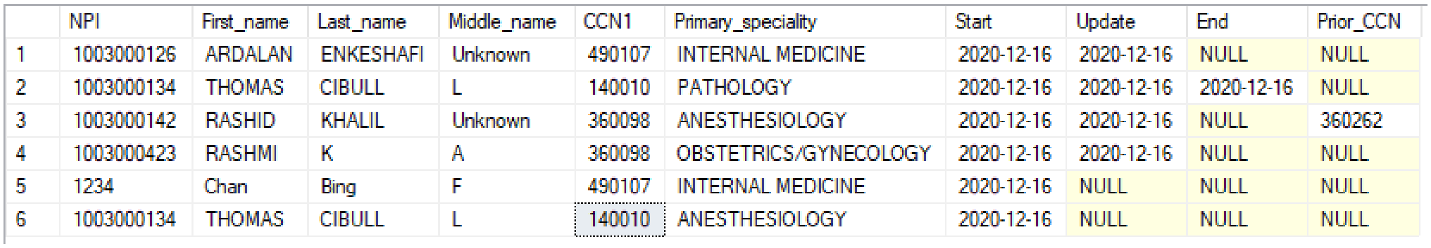
Data Flow Task: Load_Stg_Physician

The following workflow shows the implementation tasks like:

- Any change in the CCN (or FacilityID) will be update to the “Prior_CCN” (Type 3) and Update column.
- Change in the Primary Specialty column will be added to new row (Type 2) and the old row will be end dated.



The sample SCD implementation output is shown below:

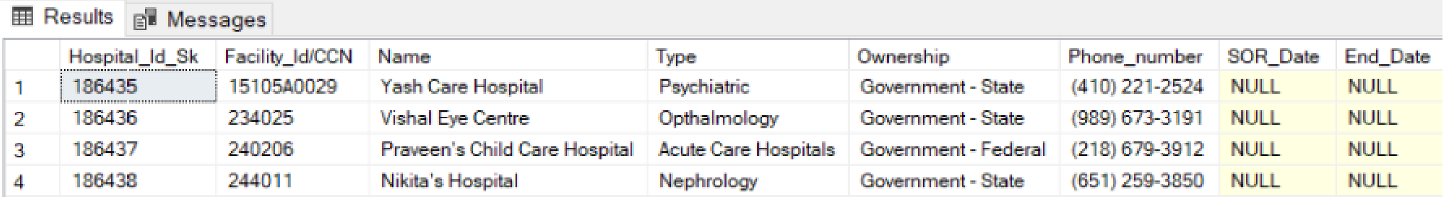


	NPI	First_name	Last_name	Middle_name	CCN1	Primary_speciality	Start	Update	End	Prior_CCN
1	1003000126	ARDALAN	ENKESHAFI	Unknown	490107	INTERNAL MEDICINE	2020-12-16	2020-12-16	NULL	NULL
2	1003000134	THOMAS	CIBULL	L	140010	PATHOLOGY	2020-12-16	2020-12-16	2020-12-16	NULL
3	1003000142	RASHID	KHALIL	Unknown	360098	ANESTHESIOLOGY	2020-12-16	2020-12-16	NULL	360262
4	1003000423	RASHMI	K	A	360098	OBSTETRICS/GYNECOLOGY	2020-12-16	2020-12-16	NULL	NULL
5	1234	Chan	Bing	F	490107	INTERNAL MEDICINE	2020-12-16	NULL	NULL	NULL
6	1003000134	THOMAS	CIBULL	L	140010	ANESTHESIOLOGY	2020-12-16	NULL	NULL	NULL

While loading the data for the first time, will create another column Start, Update and End column which will contain the 'DateTime' corresponding to that date and time. This will not only make understanding the age of the data easier but also help in loading new data into the data warehouse in the future.

The source data tables have to be updated in such a way that it contains datetime representing the corresponding update datetime. While the Nursing home dataset will need incremental loading every week, the rest of the datasets can be updated once a year.

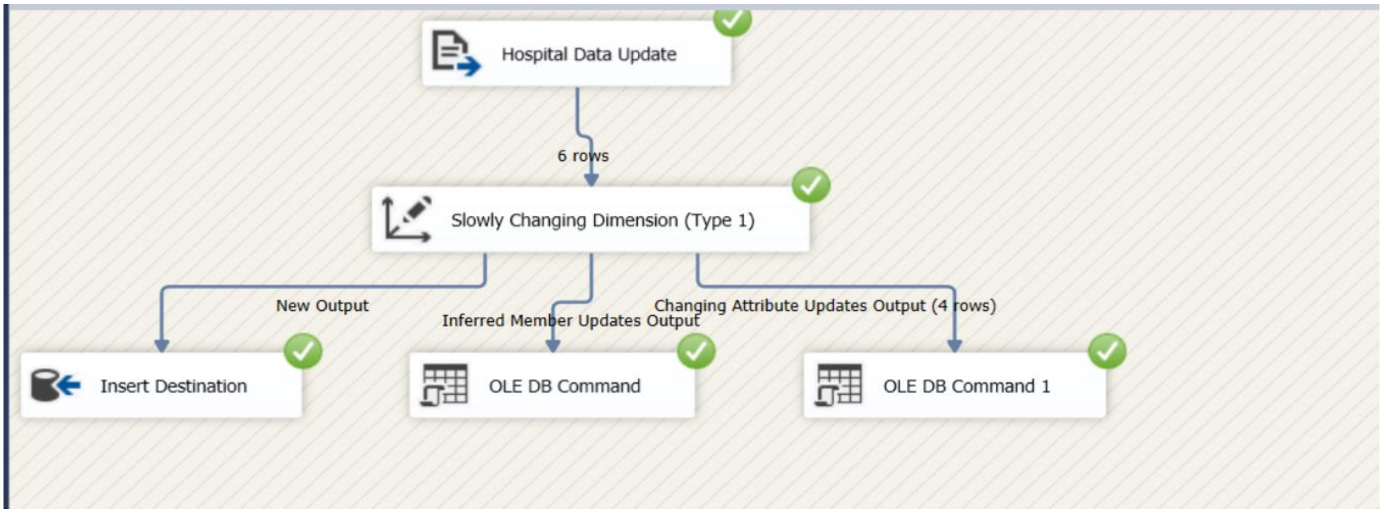
Type 1- We will be updating the rows with type 1 SCD in Hospital's Dataset. Case of consideration would be changes in the Name of hospital, Organization's Specialization and/or phone number. In this case, we will not be adding any historical attribute that gives the start and end date. The update with be modified in the row and not added.



Results
Messages

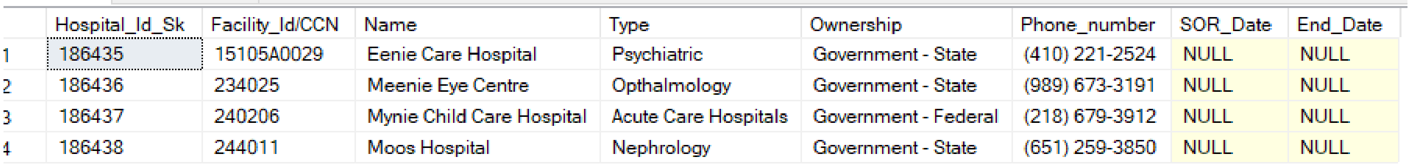
	Hospital_Id_Sk	Facility_Id/CCN	Name	Type	Ownership	Phone_number	SOR_Date	End_Date
1	186435	15105A0029	Yash Care Hospital	Psychiatric	Government - State	(410) 221-2524	NULL	NULL
2	186436	234025	Vishal Eye Centre	Opthalmology	Government - State	(989) 673-3191	NULL	NULL
3	186437	240206	Praveen's Child Care Hospital	Acute Care Hospitals	Government - Federal	(218) 679-3912	NULL	NULL
4	186438	244011	Nikita's Hospital	Nephrology	Government - State	(651) 259-3850	NULL	NULL

We perform the type 1 using SCD Wizard Package



Hospital Data Update
6 rows
Slowly Changing Dimension (Type 1)
New Output
Inferred Member Updates Output
Changing Attribute Updates Output (4 rows)
Insert Destination
OLE DB Command
OLE DB Command 1

The updated table looks as follows:



	Hospital_Id_Sk	Facility_Id/CCN	Name	Type	Ownership	Phone_number	SOR_Date	End_Date
1	186435	15105A0029	Eenie Care Hospital	Psychiatric	Government - State	(410) 221-2524	NULL	NULL
2	186436	234025	Meenie Eye Centre	Opthalmology	Government - State	(989) 673-3191	NULL	NULL
3	186437	240206	Mynie Child Care Hospital	Acute Care Hospitals	Government - Federal	(218) 679-3912	NULL	NULL
4	186438	244011	Moos Hospital	Nephrology	Government - State	(651) 259-3850	NULL	NULL

Logging process

Because datasets we have considered are large and somewhat inconsistent, there is a good chance we will hit a number of roadblocks during the ETL process. It is important for us to have a correct logging process to keep the entire ETL operation in a state of constant improvement, helping our team manage bugs and problems with data sources, format, transformations, destinations etc. We also require root cause analysis for any critical issues.

We have used the following SSIS logging methods for troubleshooting purposes –

- SSIS packages log providers
- Custom logging messages using scripts in the Execute SQL task
- The SSIS catalogue which provides the execution logs in the SSISDB database

Data validation

Data validation is the process of ensuring data has undergone data cleansing to ensure they have data quality, that is, that they are both correct and useful. When using SQL, data validation helps to have a consistent data. Before moving the data into the Datawarehouse we will load the data into the staging area where we would perform the data validation.

We have the following datasets –

- Dataset – 1 “Hospital_General_Information”
- Dataset – 2 “Physician_Compare_National”
- Dataset – 3 “population_by_zip_2010”
- Dataset – 4 “Zip/city/state/fips”

We are going to take smaller samples of the datasets to check validate if the data is loaded correctly. But if we sample the first 200 row in each dataset, it is going to be biased. Therefore, we are sampling data based on different criteria for our datasets.

For Dataset – 1 “Hospital_General_Information”, we sample the data based on each state. We take 4 random for each state so that they have different cities and zip codes, and this would give us around 200 records to validate from 5315 records.

For Dataset – 2 “Physician_Compare_National”, we again sample this data set based on state where we take 20 random for each state so that they have different cities and zip codes, and this would give us around 1000 records to validate from 2.19 million records.

Dataset – 3 “population_by_zip”, we can sample using zip codes where we take 10 random records for each unique zip code (33119 unique zip codes) which would give us 331190 records to validate against more than a million record records.

Dataset – 4 “Covid” we can sample based on different states. We take 4 random records for each state so that we have different counties, and this would give us around 200 records to validate from 3269 records.

We also perform visualizations on the sample data and sample data loaded in the DW. And when we visualize the sample data and the dimensionally modelled dataset the differences should be negligible to be valid and if the differences are drastic then we can conclude that the data loaded is invalid.

And after loading the data into the data warehouse we check the number of records, unique IDs and the source and target data fields. We are loading the data from different sources into SQL server for our project. And we have implemented data validation using check constraints, unique constraints, not null and primary constraints. We maintain referential integrity with the help of a lookup table, which would help us update and delete records based on pre-defined constraints. We also check for the number of columns and rows for each dataset at the source and the destination, so that if there are more number or less number of columns or rows than the source at the destination, we can be sure that the data was not loaded correctly, and we can investigate it.

Logging data validation

When the data is being loaded from the data source to the staging table, we are using a Lookup table in order to validate the data stored in the staging table. We are having a dataset with Zip codes, Cities, States and FIPS codes in the United States as our lookup table. For each staging table we have a reject table to store the rejected/invalid data. For the Hospital Dataset, we are comparing the Zip code in the Lookup Table and the Zip code in the csv file and if the Zip code is present the data is moved into the staging are and if not, the data is moved into the reject table where we log the invalid data using get date() function in a separate column. Similarly, we use the FIPS code to in the Lookup Table and compare it to the csv files of Population and Covid dataset to get the valid data into the staging table and the invalid data into the reject table which are logged by date and time.

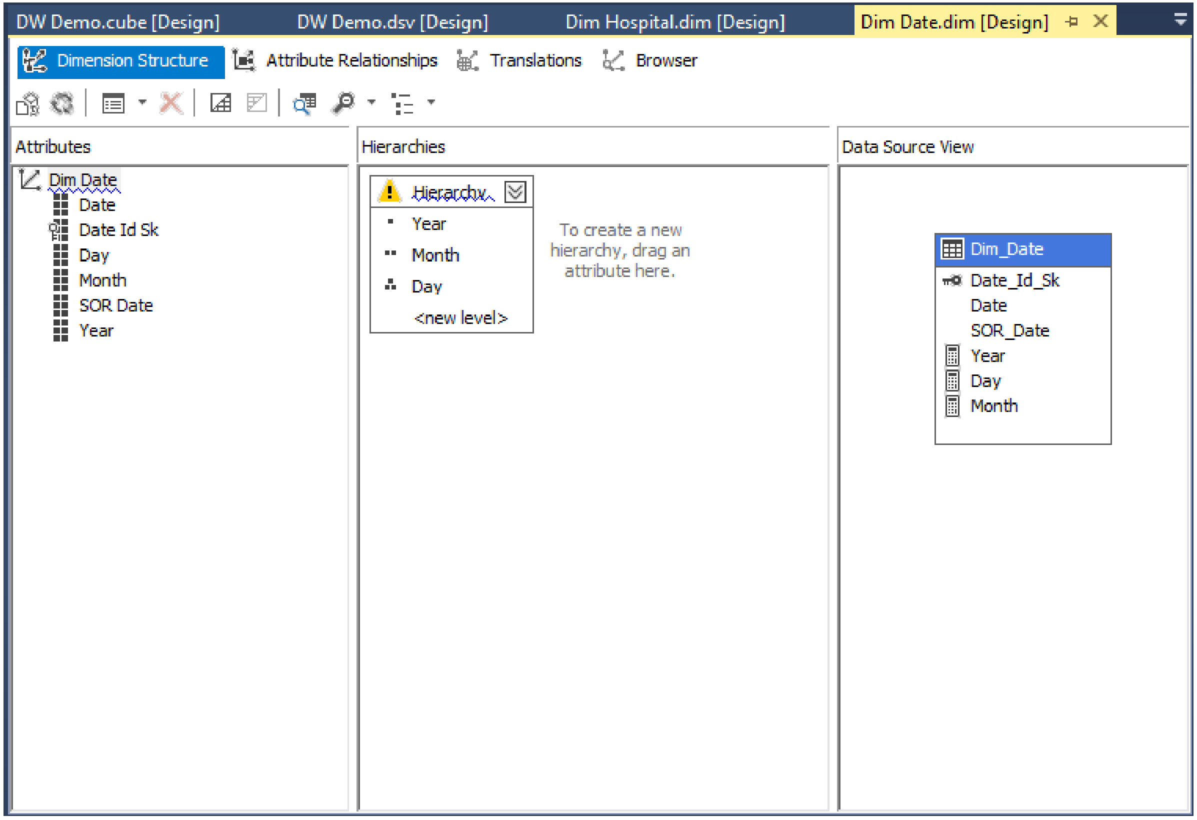
Analytics Design

a. OLAP cubes from star schema

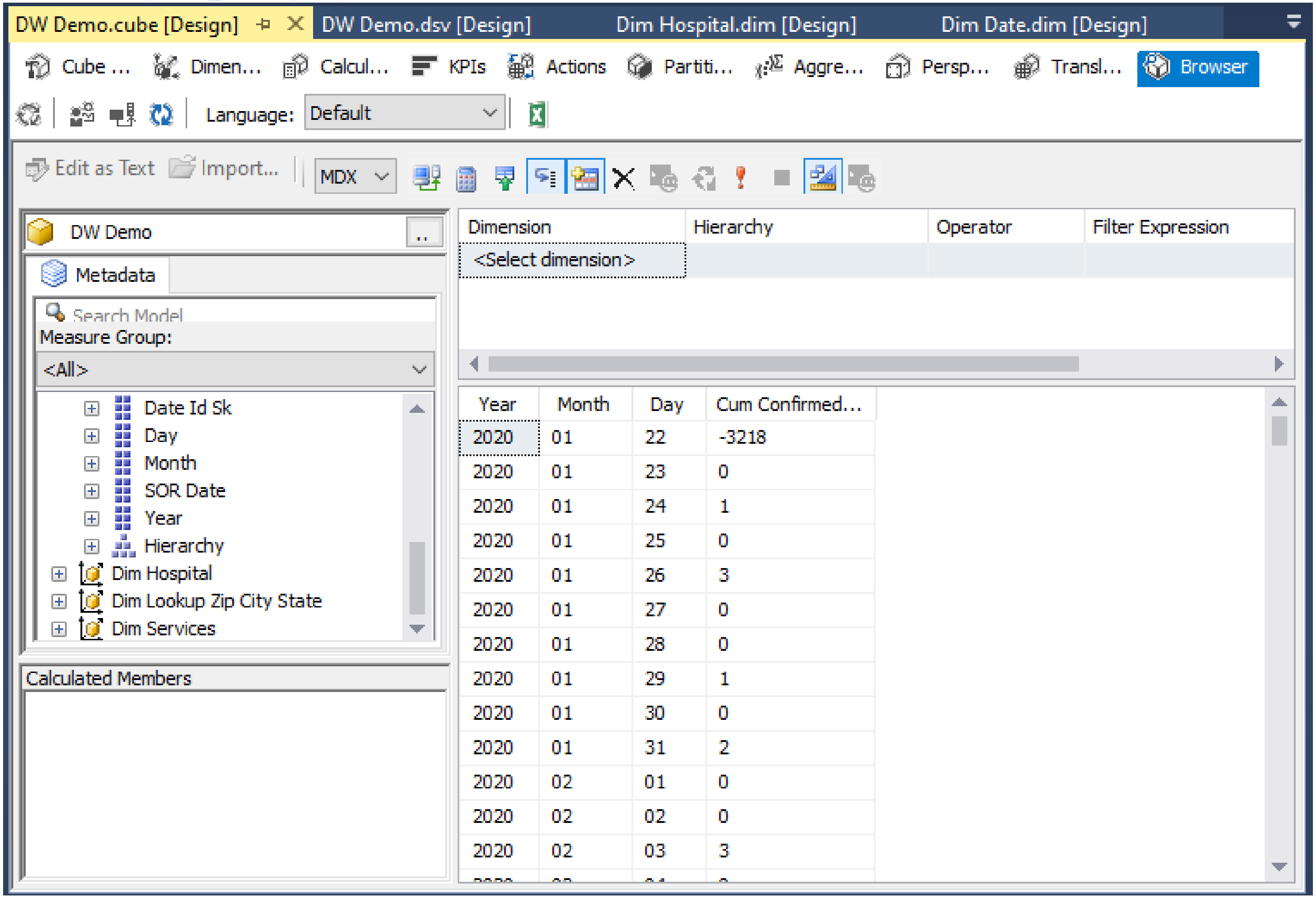
Significance of OLAP cube- Relational databases although are great for storing large databases, it's hard to generate management report from transactional data. For our project database having around 1 million records in each dataset, the relational database querying is time consuming and inefficient way. One such solution to deal with such problem is through OLAP. OLAP cube is implemented for four datasets which are Covid, Hospital, Population and Zip/City/State/Fips . Using this cube, we have increased the querying performance and reporting is achieved using tools like Excel.

Some of the analysis we've done but are not limited to are as follows:

1. Hierarchy for Dim_Date

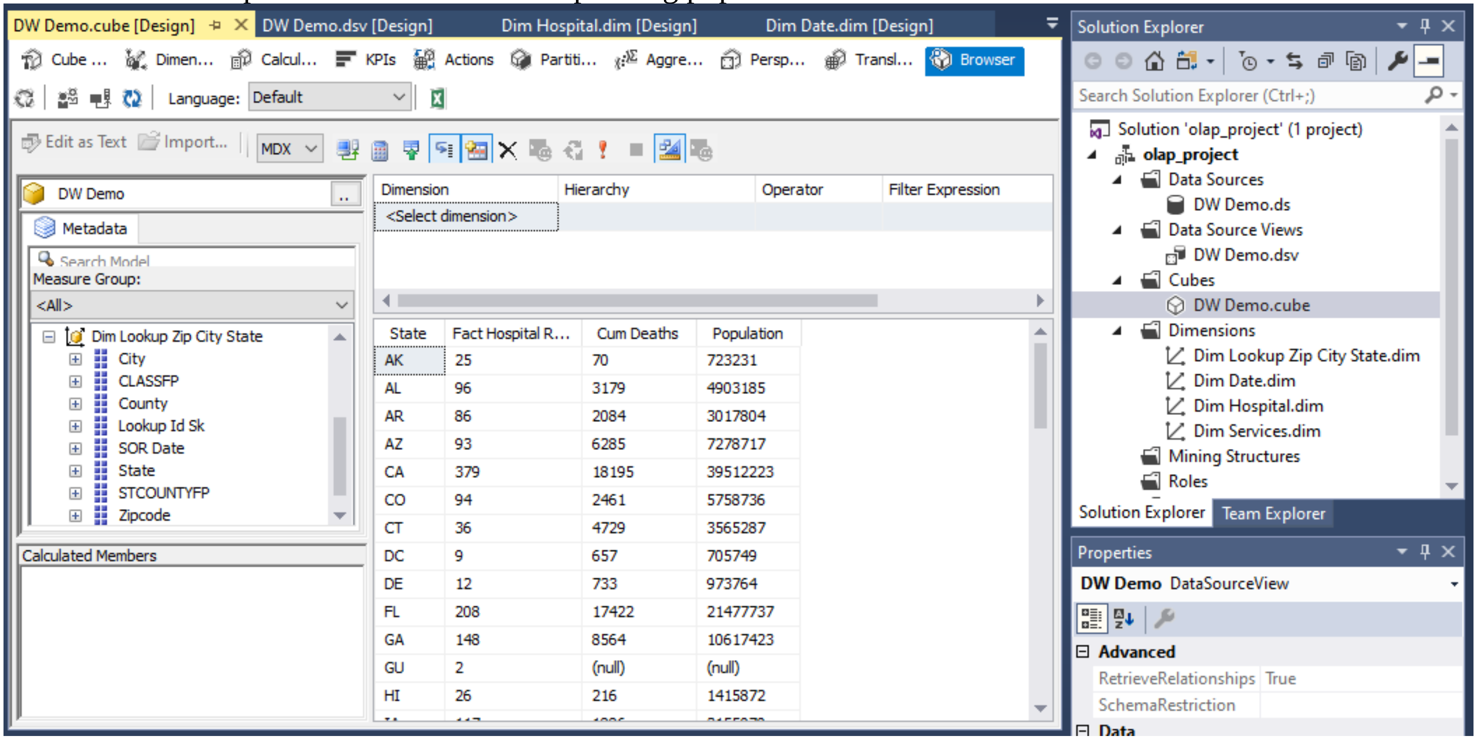


2. Hierchary Dim_Date query result vs covid cases



Year	Month	Day	Cum Confirmed...
2020	01	22	-3218
2020	01	23	0
2020	01	24	1
2020	01	25	0
2020	01	26	3
2020	01	27	0
2020	01	28	0
2020	01	29	1
2020	01	30	0
2020	01	31	2
2020	02	01	0
2020	02	02	0
2020	02	03	3

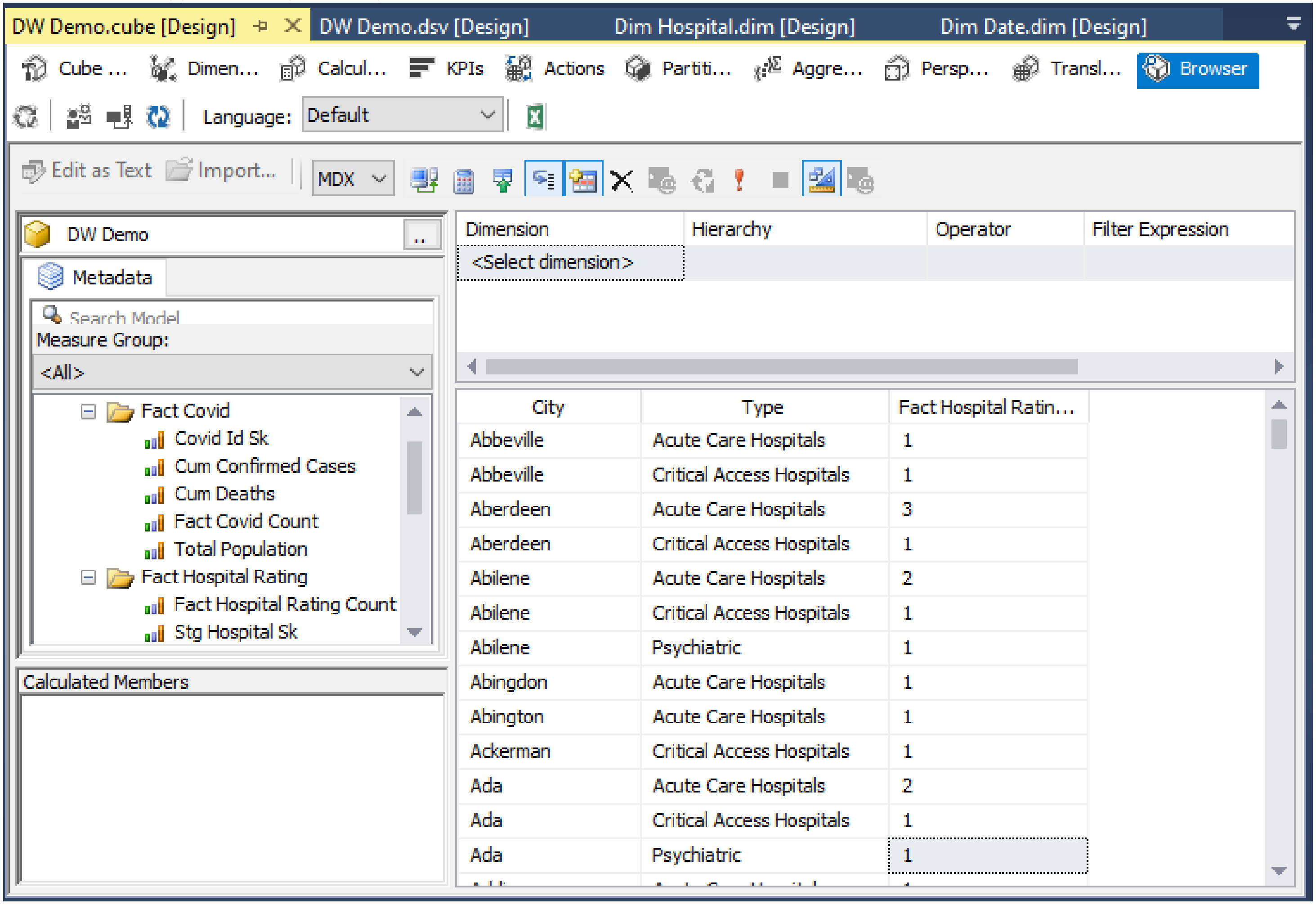
3. Number of hospital state wise and corresponding population and covid cases.



State	Fact Hospital R...	Cum Deaths	Population
AK	25	70	723231
AL	96	3179	4903185
AR	86	2084	3017804
AZ	93	6285	7278717
CA	379	18195	39512223
CO	94	2461	5758736
CT	36	4729	3565287
DC	9	657	705749
DE	12	733	973764
FL	208	17422	21477737
GA	148	8564	10617423
GU	2	(null)	(null)
HI	26	216	1415872

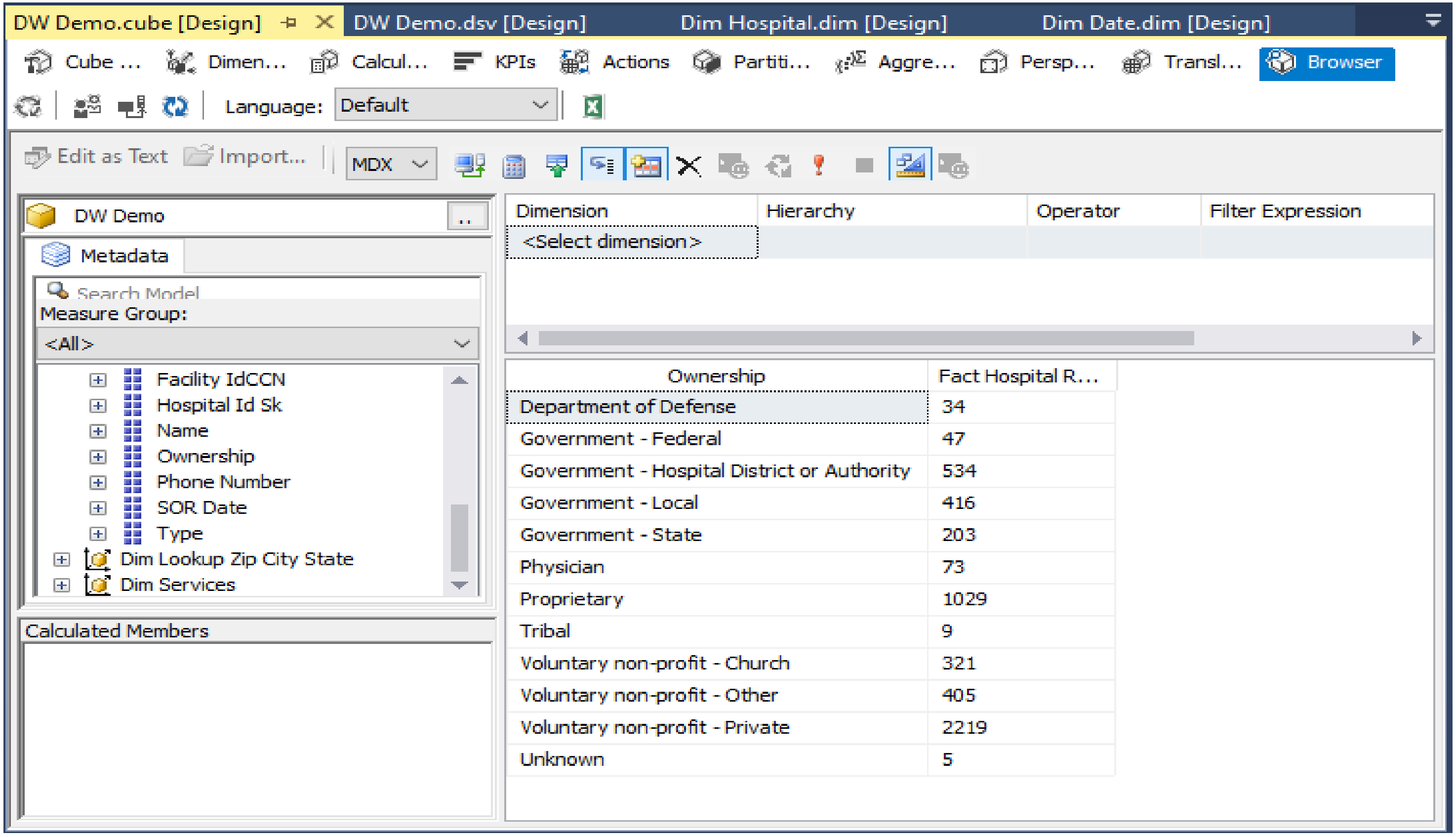
RetrieveRelationships True
SchemaRestriction

4. Different type of hospital city wise with their count



City	Type	Fact Hospital Ratin...
Abbeville	Acute Care Hospitals	1
Abbeville	Critical Access Hospitals	1
Aberdeen	Acute Care Hospitals	3
Aberdeen	Critical Access Hospitals	1
Abilene	Acute Care Hospitals	2
Abilene	Critical Access Hospitals	1
Abilene	Psychiatric	1
Abingdon	Acute Care Hospitals	1
Abington	Acute Care Hospitals	1
Ackerman	Critical Access Hospitals	1
Ada	Acute Care Hospitals	2
Ada	Critical Access Hospitals	1
Ada	Psychiatric	1

5. Number of different hospital ownerships with count



Ownership	Fact Hospital R...
Department of Defense	34
Government - Federal	47
Government - Hospital District or Authority	534
Government - Local	416
Government - State	203
Physician	73
Proprietary	1029
Tribal	9
Voluntary non-profit - Church	321
Voluntary non-profit - Other	405
Voluntary non-profit - Private	2219
Unknown	5

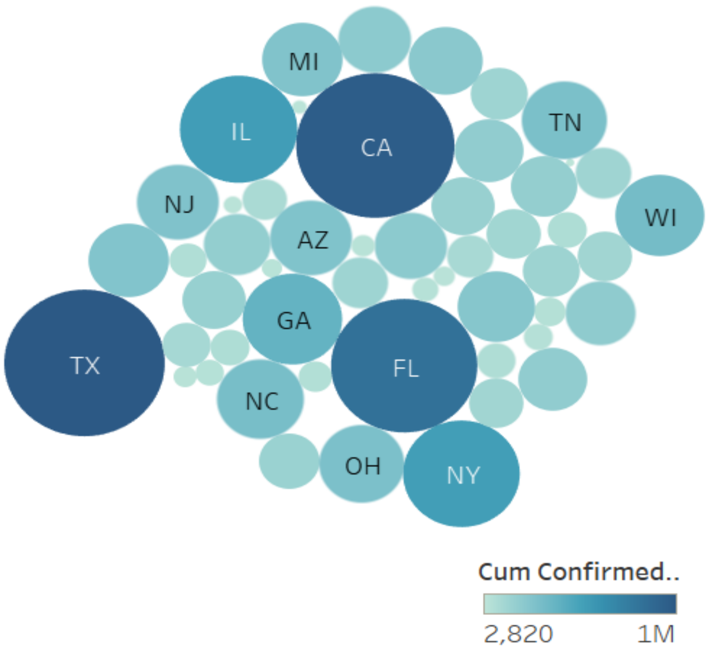
b. <u>Visualizations</u>

Visualization by Tableau:

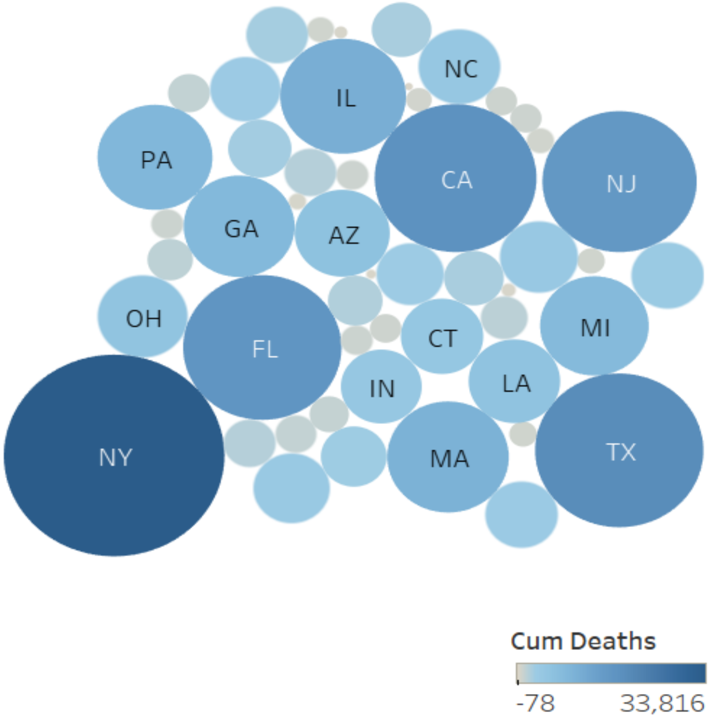
Reason to choose Tableau as Business Intelligence Tool:

1. Using the server connect, it can be easily connected to Microsoft sql server and extract the data for visualization.
2. In the connect pain, one can view the table through join and preview the outcome.
3. Gartner 2020 recognizes Tableau as a Leader for Analytics and Business Intelligence Platform
4. Tableau has scaled to meet the needs of the data driven enterprise

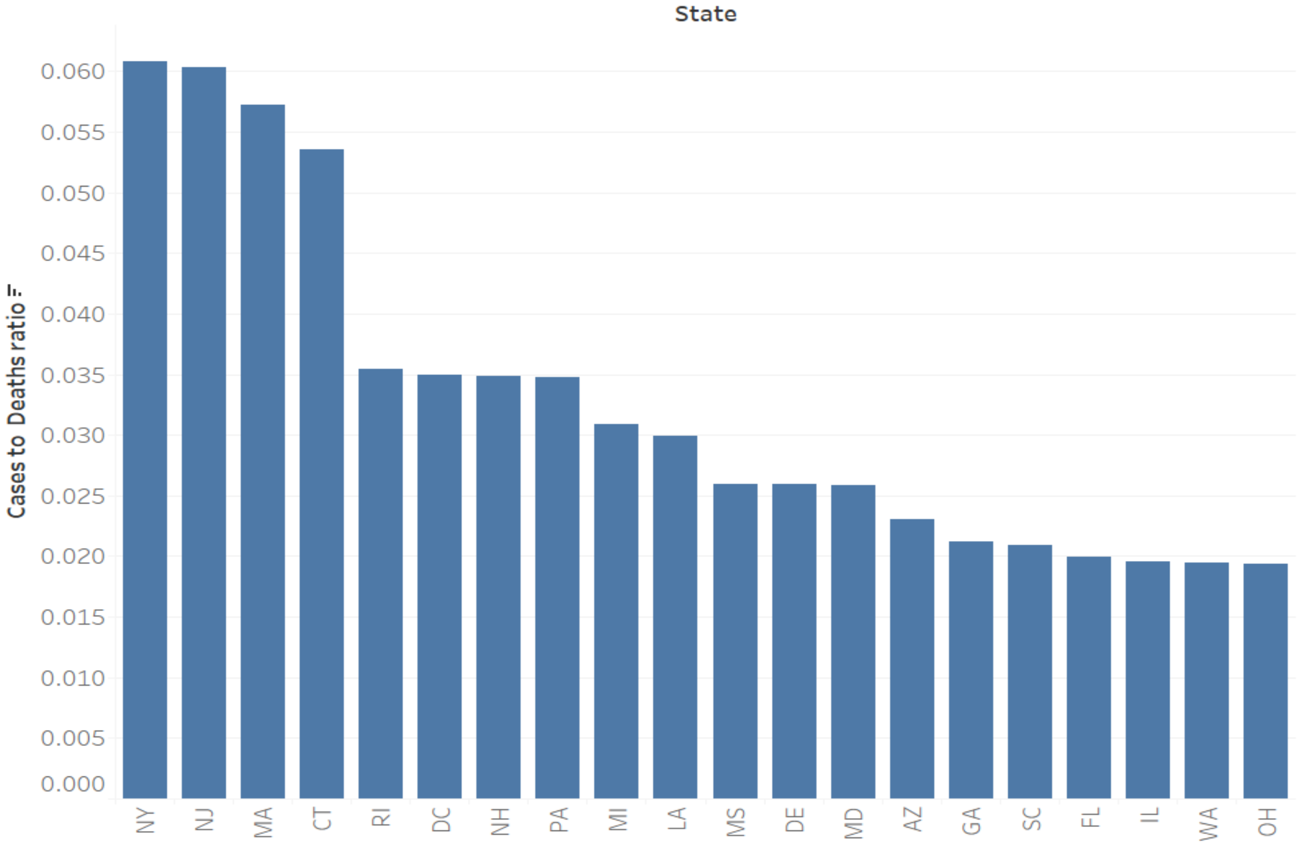
Statewise Covid-19 Confirmed Cases



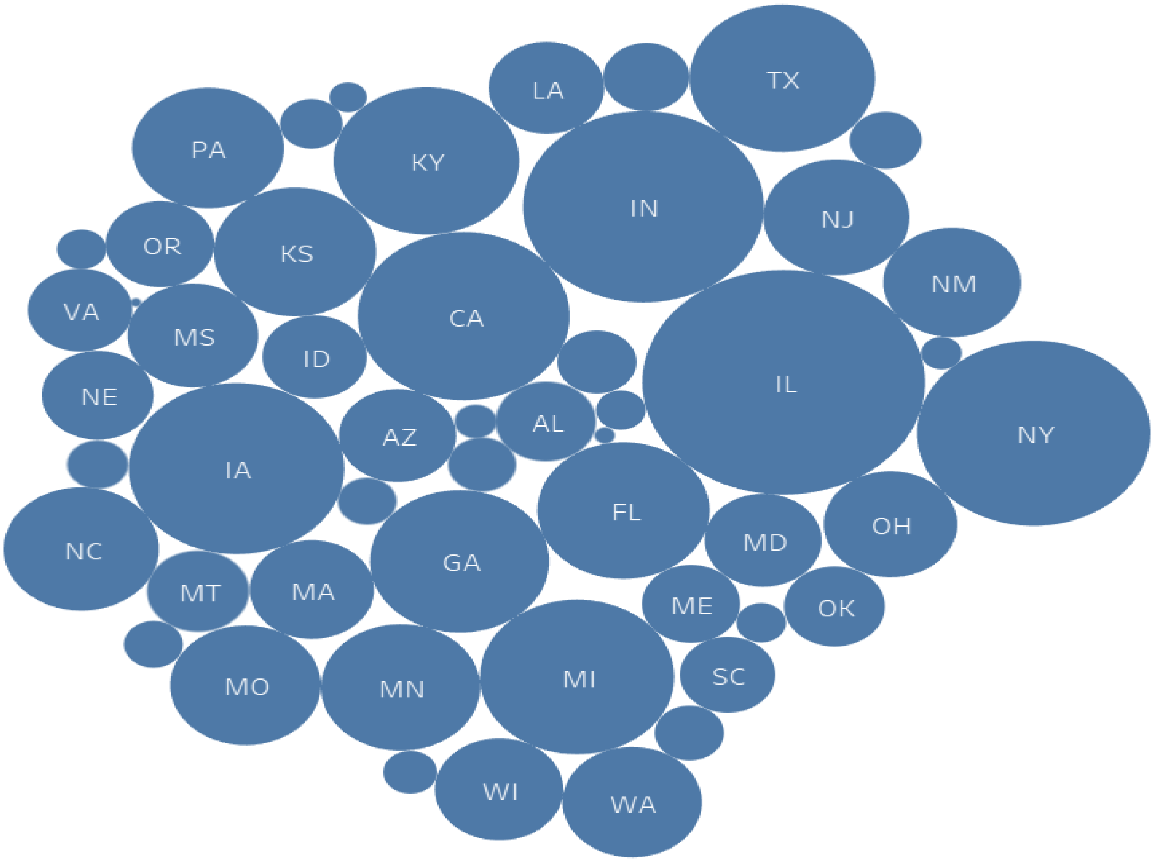
Covid-19 related Deaths



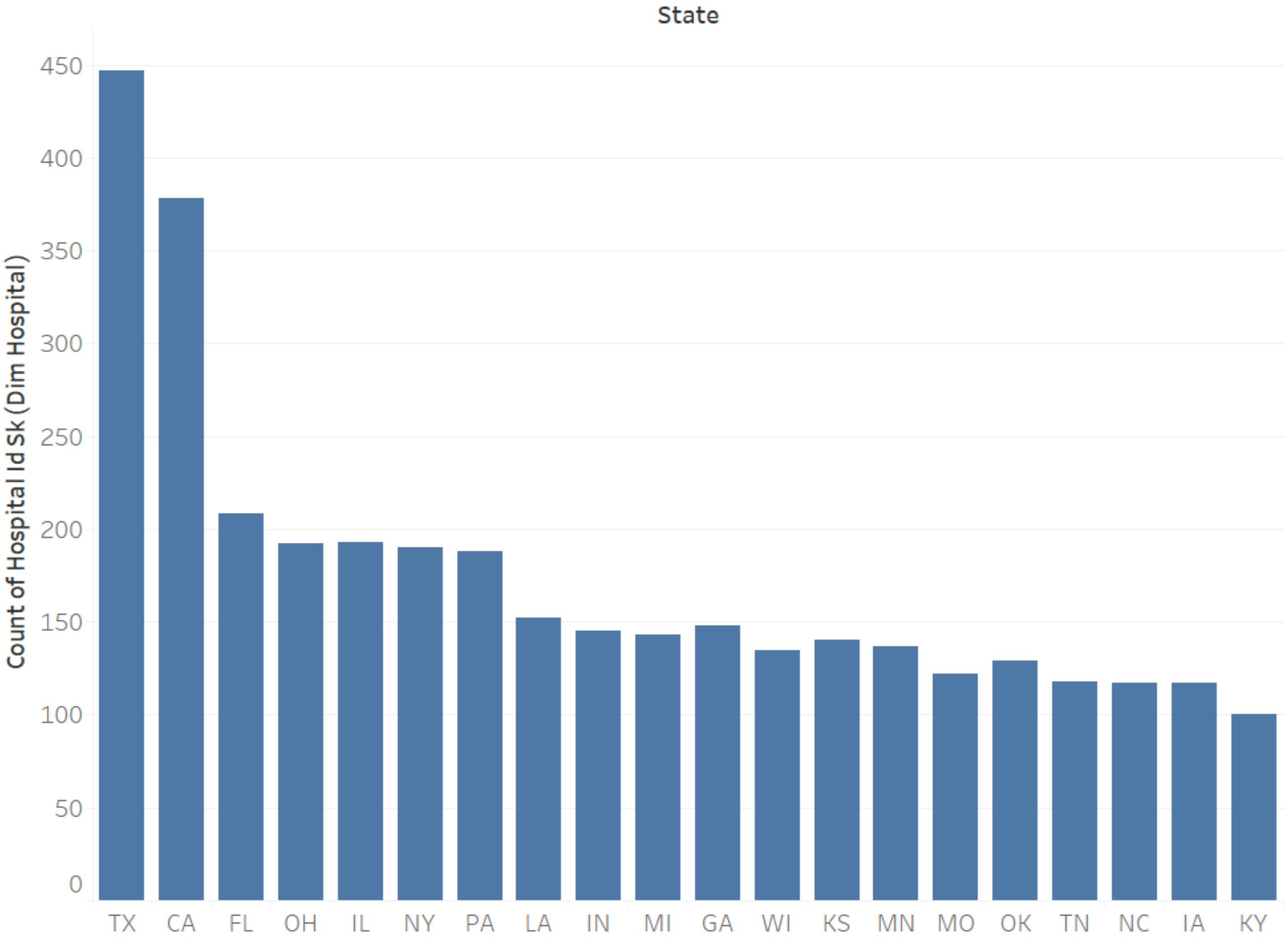
States with the highest Death rates per confirmed Covid-19 cases



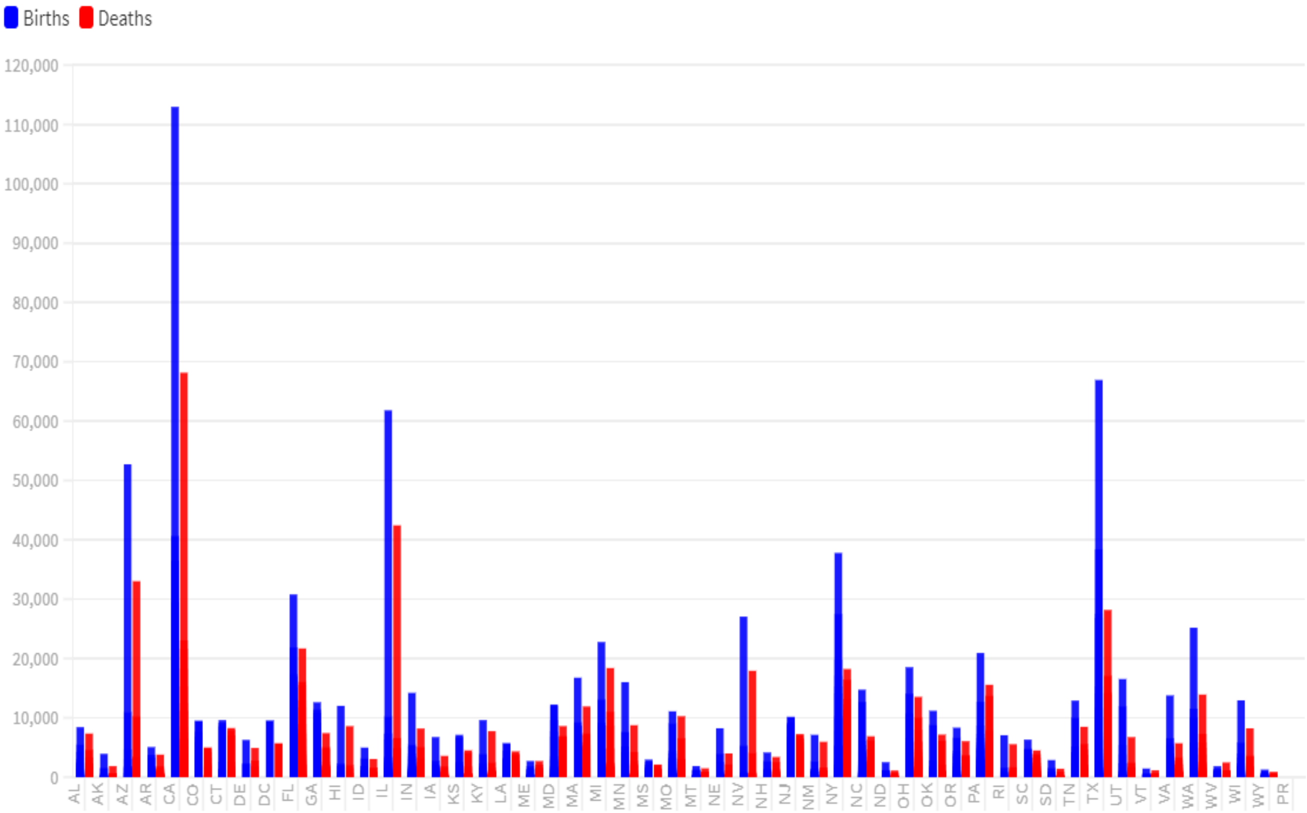
Total Physicians - Statewise



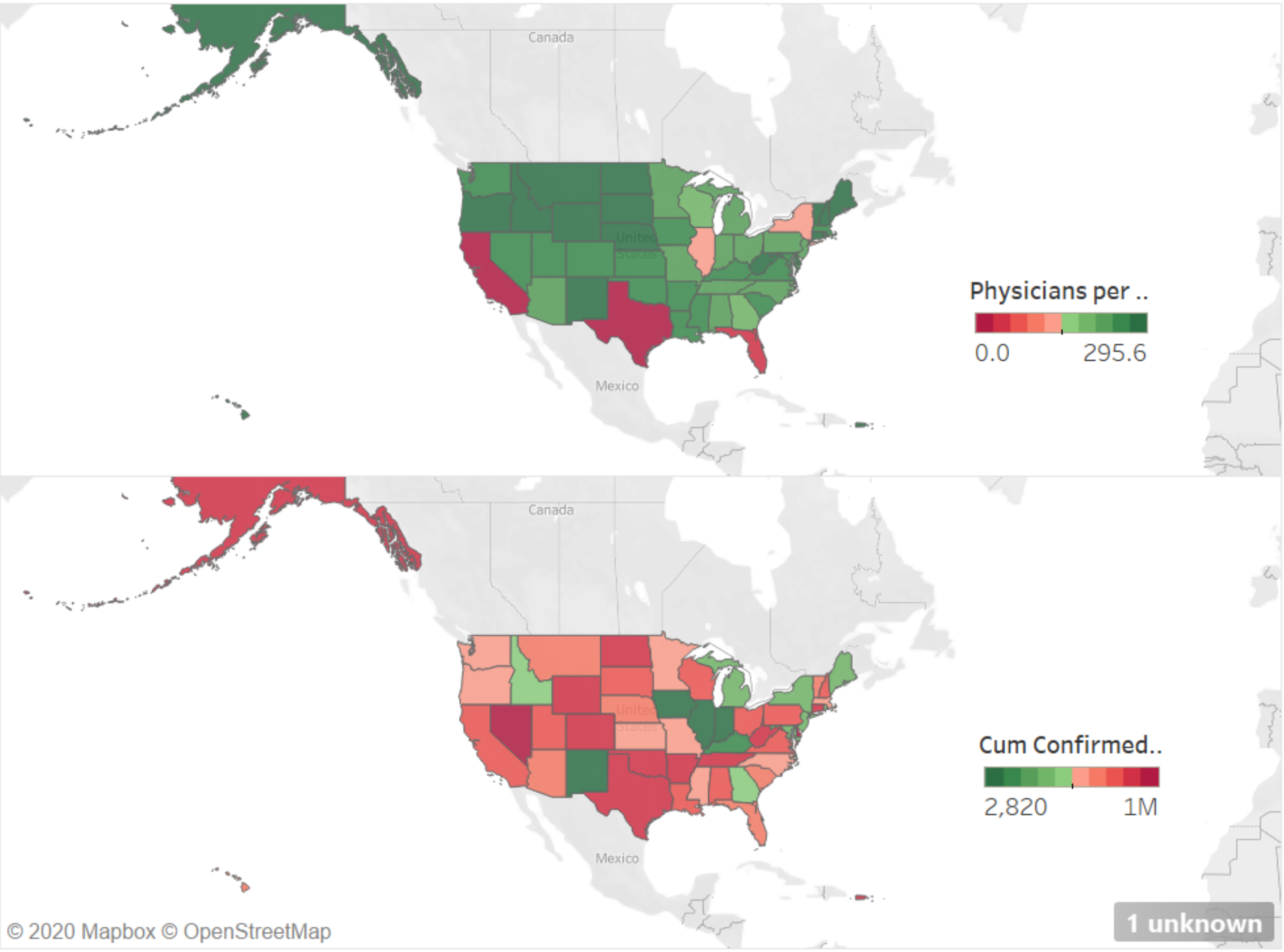
Total number of CMS facilities in the US



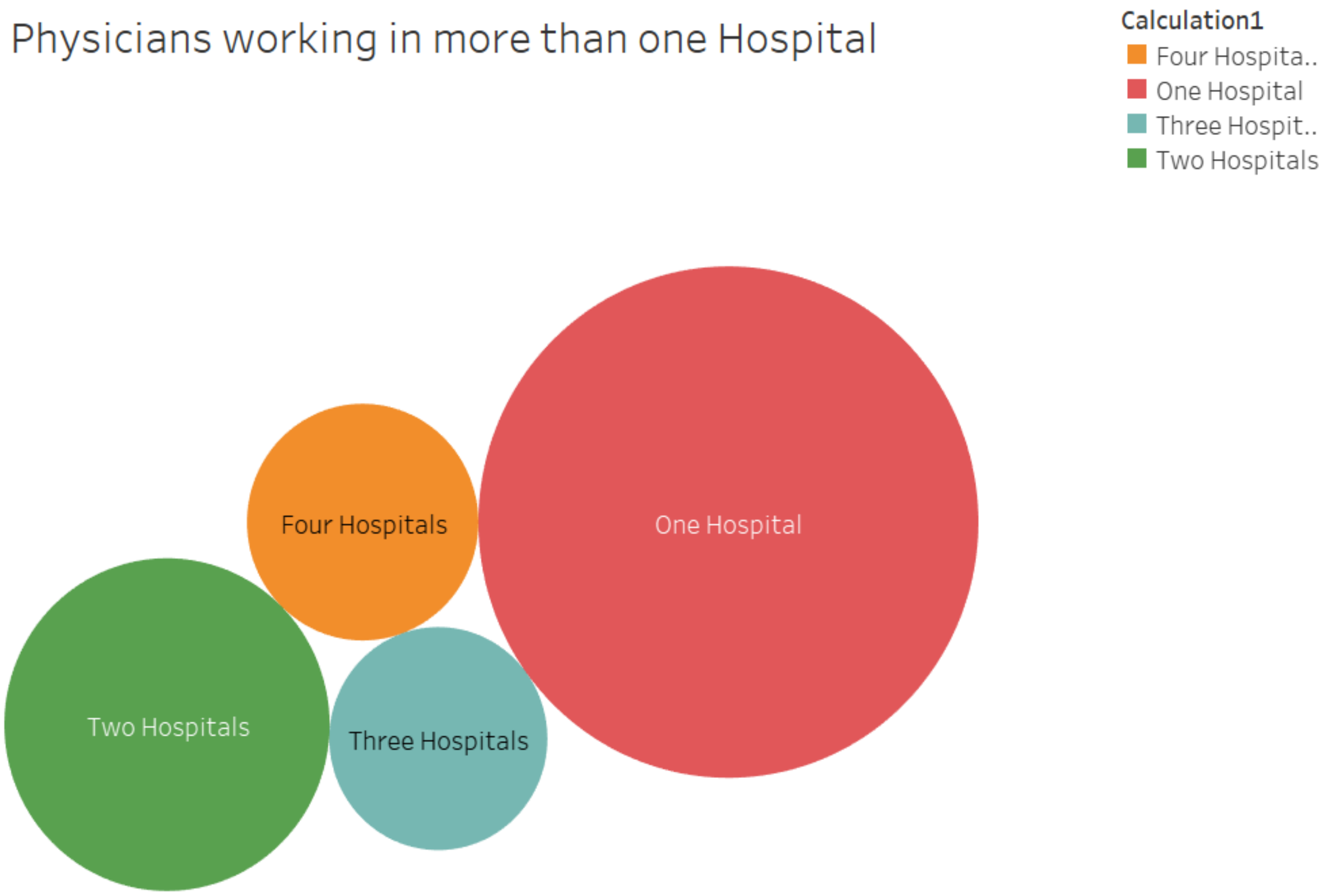
Births and Deaths in each state



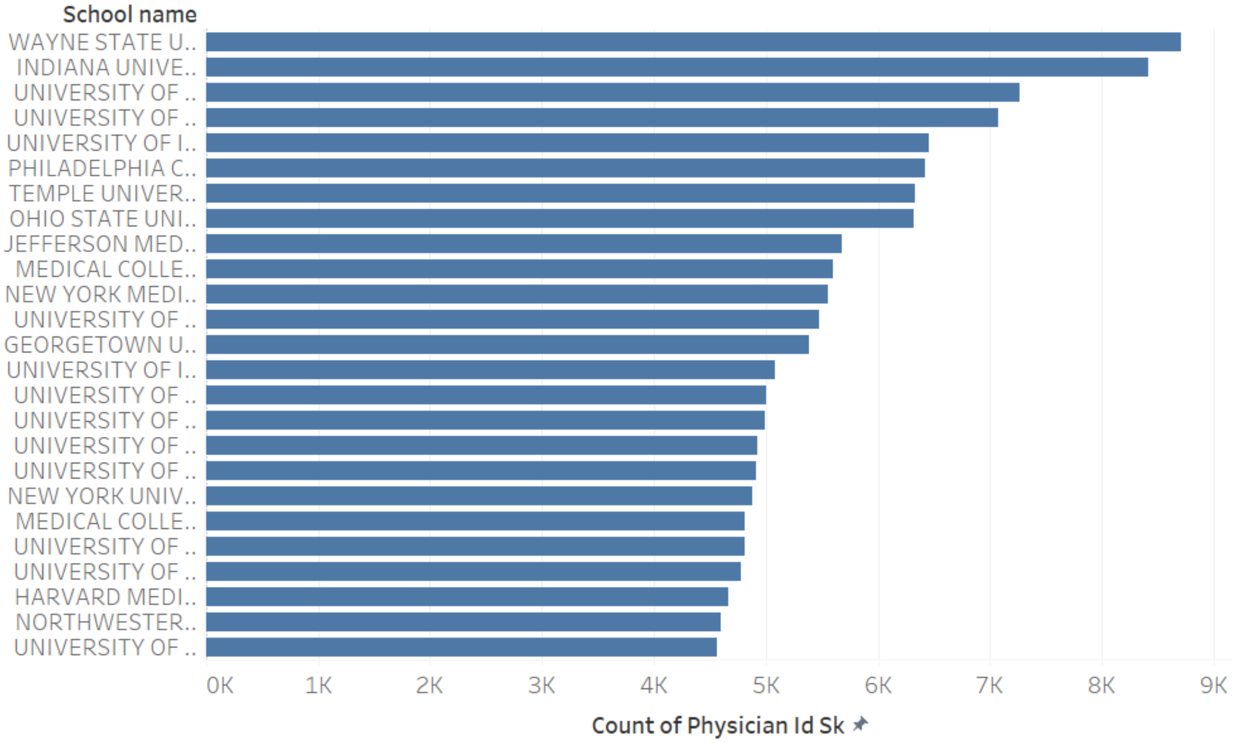
Total Covid-19 Cases (Map 1) and Physicians per Hospital (Map 2)



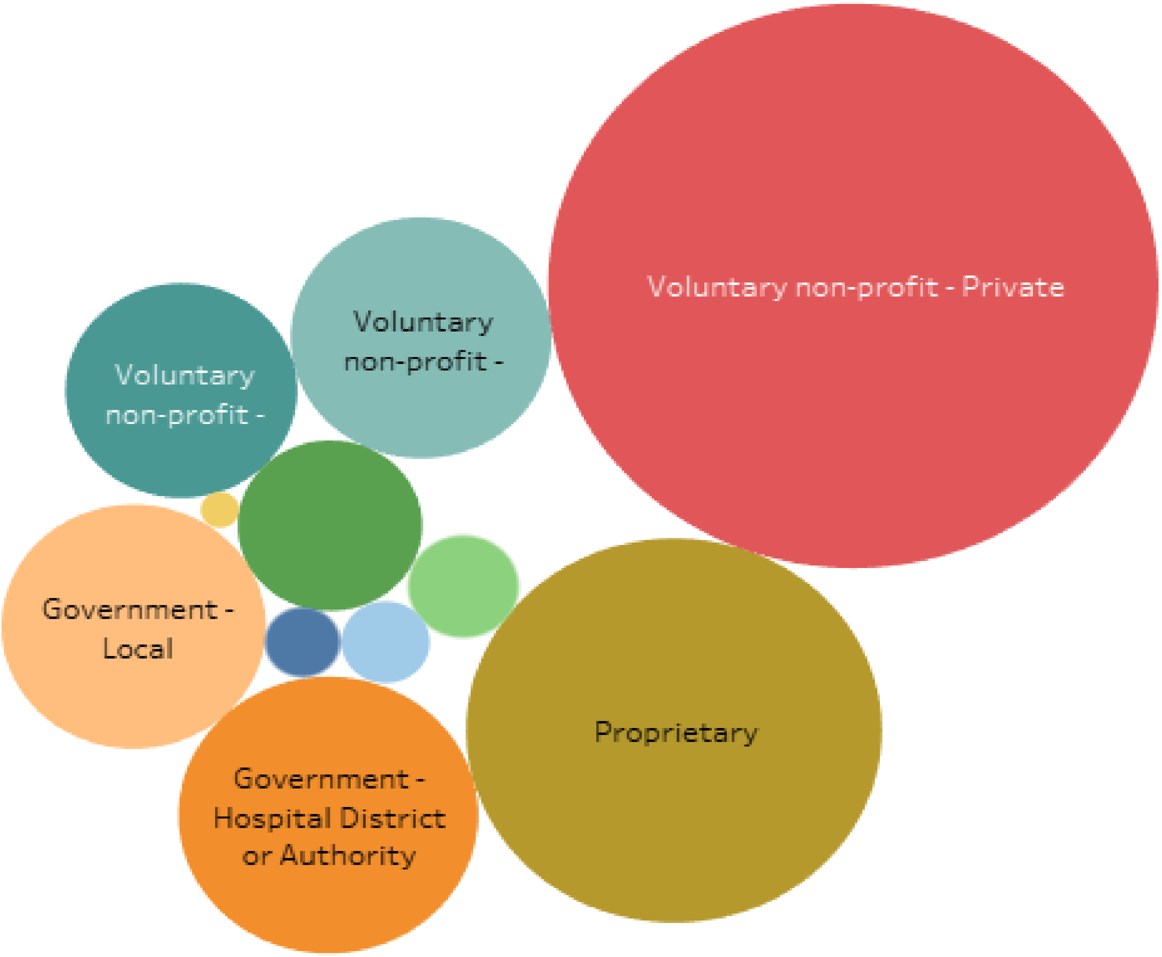
Physicians working in more than one Hospital



Top 25 Medical Schools Producing CMS physicians



Hospital Ownership types


Voluntary non-profit - Private
Voluntary non-profit -
Voluntary non-profit -
Government - Local
Proprietary
Government - Hospital District or Authority