

Medicare Provider Utilization And Payment Data

Database Warehouse And Business Intelligence

Prepared By:
Shaishav Shah



Objective

Analyze cost and treatment KPIs of different types of health service provider and also, analyze the nursing home facilities condition and Inpatient healthcare conditions.





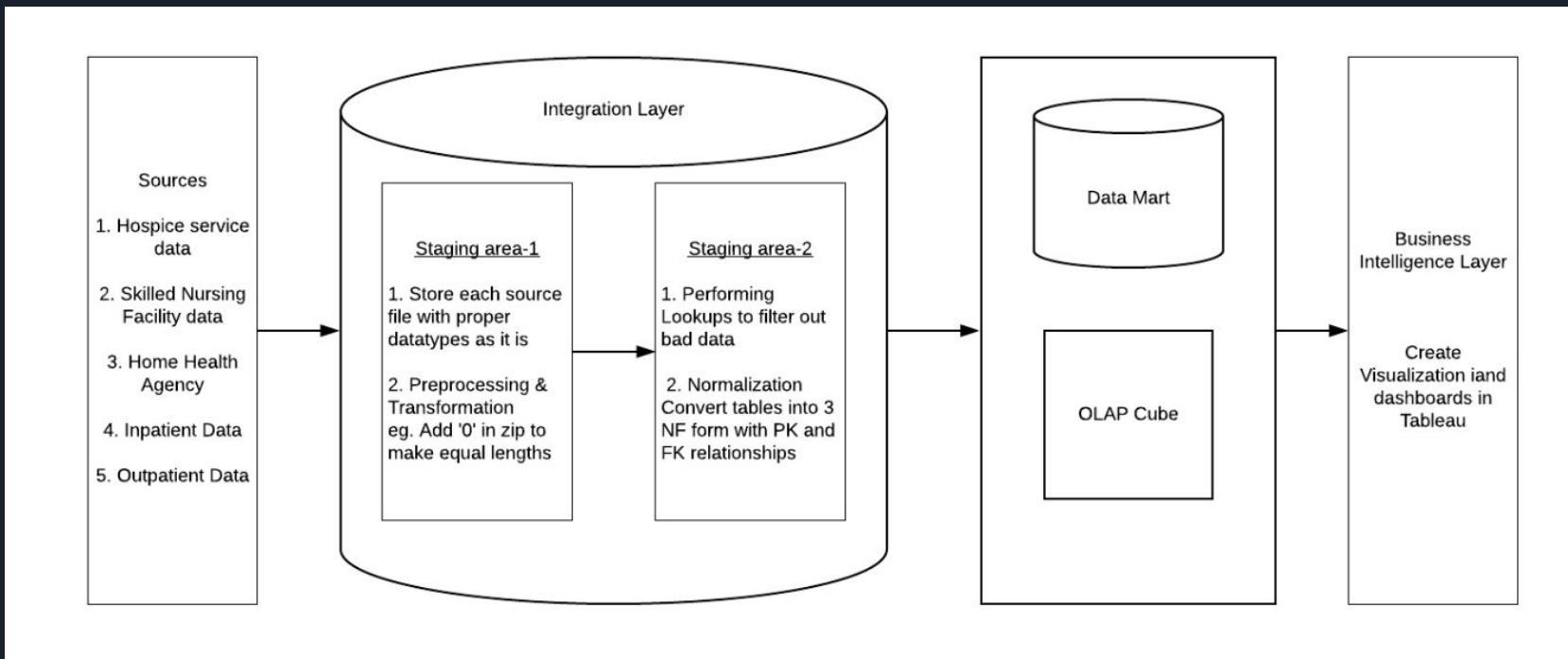
Data Source

Data was collected from CMS website. The data set consist of five (5) files which were linked with each other based on their Provider ID over 2015 and 2016 year.

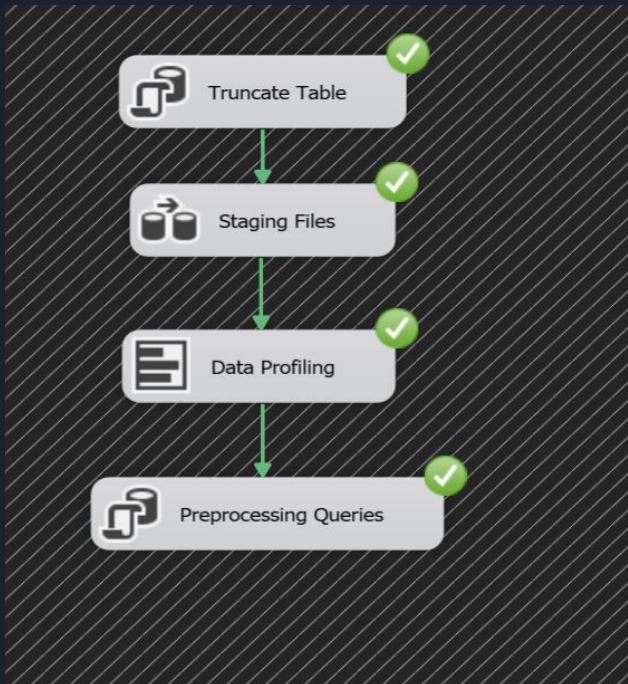
1. Medicare Provider Utilization and Payment Data: Skilled Nursing Facilities
2. Medicare Provider Utilization and Payment Data: Hospice Providers
3. Medicare Provider Utilization and Payment Data: Home Health Agencies
4. Medicare Provider Utilization and Payment Data: Outpatient
5. Medicare Provider Utilization and Payment Data: Inpatient

For all datasets key column is : Provider ID

Data and Process Flow



Data Profiling and Preprocessing



- Automated Data Profiling using in-built tool and SSIS Package
- Customised Pre-processing queries according to the report generated

Data Profiling and Preprocessing

Data Profile Viewer -

Open Refresh

Profiles (Table View)

- [dbo].[Stag_HHA_2015]
 - Candidate Key Profiles
 - Column Length Distribution Profiles
 - Column Null Ratio Profiles
 - Column Pattern Profiles
 - Column Statistics Profiles
 - Column Value Distribution Profiles
- [dbo].[Stag_HHA_2016]
 - Candidate Key Profiles
 - Column Length Distribution Profiles
 - Column Null Ratio Profiles
 - Column Pattern Profiles
 - Column Statistics Profiles
 - Column Value Distribution Profiles
- [dbo].[Stag_HOS_2015]
 - Candidate Key Profiles
 - Column Length Distribution Profiles
 - Column Null Ratio Profiles
 - Column Pattern Profiles
 - Column Statistics Profiles
 - Column Value Distribution Profiles
- [dbo].[Stag_HOS_2016]
 - Candidate Key Profiles
 - Column Length Distribution Profiles
 - Column Null Ratio Profiles
 - Column Pattern Profiles
 - Column Statistics Profiles
 - Column Value Distribution Profiles
- [dbo].[Stag_Imp_2015]
 - Column Length Distribution Profiles
 - Column Null Ratio Profiles
 - Column Pattern Profiles
 - Column Statistics Profiles
 - Column Value Distribution Profiles
- [dbo].[Stag_Imp_2016]
 - Column Length Distribution Profiles
 - Column Null Ratio Profiles
 - Column Pattern Profiles
 - Column Statistics Profiles
 - Column Value Distribution Profiles
- [dbo].[Stag_Nurse_2015]
 - Candidate Key Profiles
 - Column Length Distribution Profiles
 - Column Null Ratio Profiles
 - Column Pattern Profiles
 - Column Statistics Profiles
 - Column Value Distribution Profiles

Column Null Ratio Profiles - [dbo].[Stag_HHA_2015]

Column	Null Count	Null Percentage
Agency Name	0	0.0000 %
American Indian or Alaska...	3123	29.6694 %
American Indian or Alaska...	3123	29.6694 %
Asian Pacific Islander Bene...	5027	47.7579 %
Asian Pacific Islander Bene...	5027	47.7579 %
Average Age	0	0.0000 %
Average Age	0	0.0000 %
Average HCC Score	0	0.0000 %
Average HCC Score	0	0.0000 %
Average Number of ST Visi...	0	0.0000 %
Average Number of ST Visi...	0	0.0000 %
Average Number of Home...	0	0.0000 %
Average Number of Home...	0	0.0000 %
Average Number of Medica...	0	0.0000 %
Average Number of Medica...	0	0.0000 %
Average Number of OT Visi...	0	0.0000 %
Average Number of OT Visi...	0	0.0000 %
Average Number of PT Visit...	0	0.0000 %
Average Number of PT Visit...	0	0.0000 %
Average Number of Skilled...	0	0.0000 %
Average Number of Skilled...	0	0.0000 %

Successfully loaded data profile from ...

Message



Summary of Data Profiling

Some of our observations were:

1. Zip Code were of 4 digits and 3 digits instead of 5 digits
Example: 02215 becomes 2215
2. According to the definition, Provider ID was an unique 6 digit number. However, there were many Provider IDs which having only 5 digits.
3. There were few NULL values in the important columns like Zip Code, State Etc.
4. Data source was not consistent with Name. For example, few Provider Names were in Upper Case while Few were in lower case.

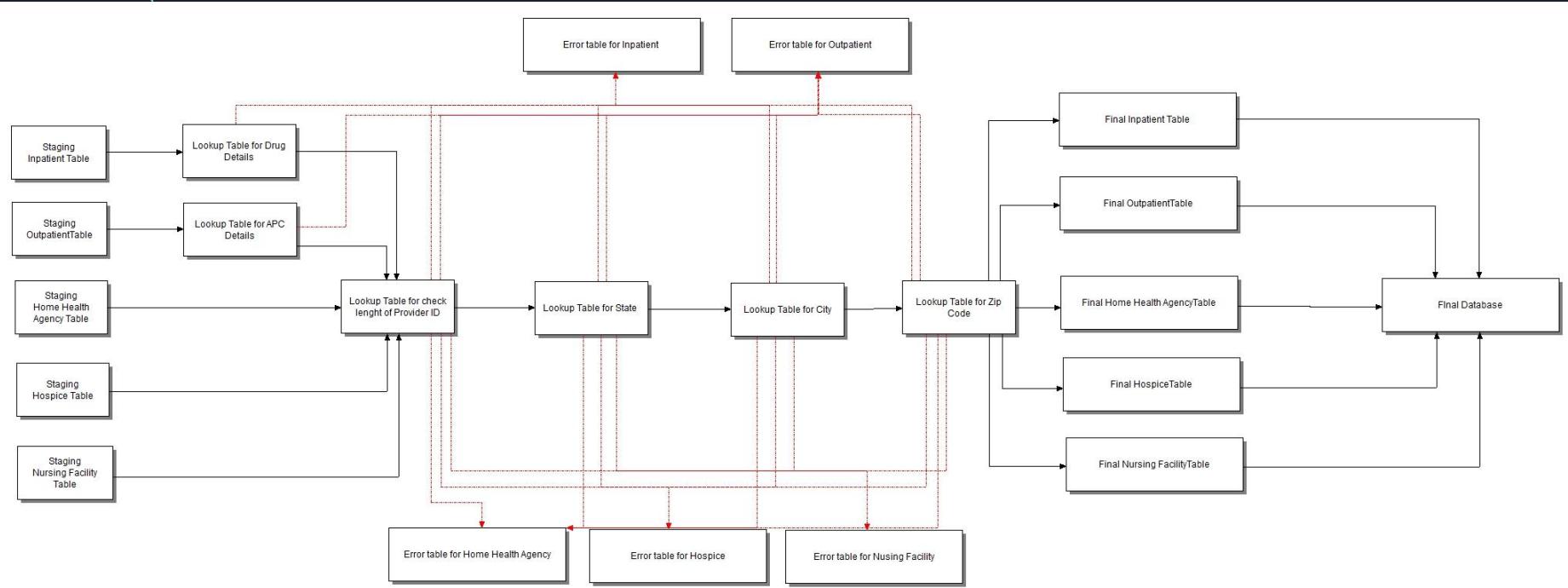


Summary of Data Preprocessing

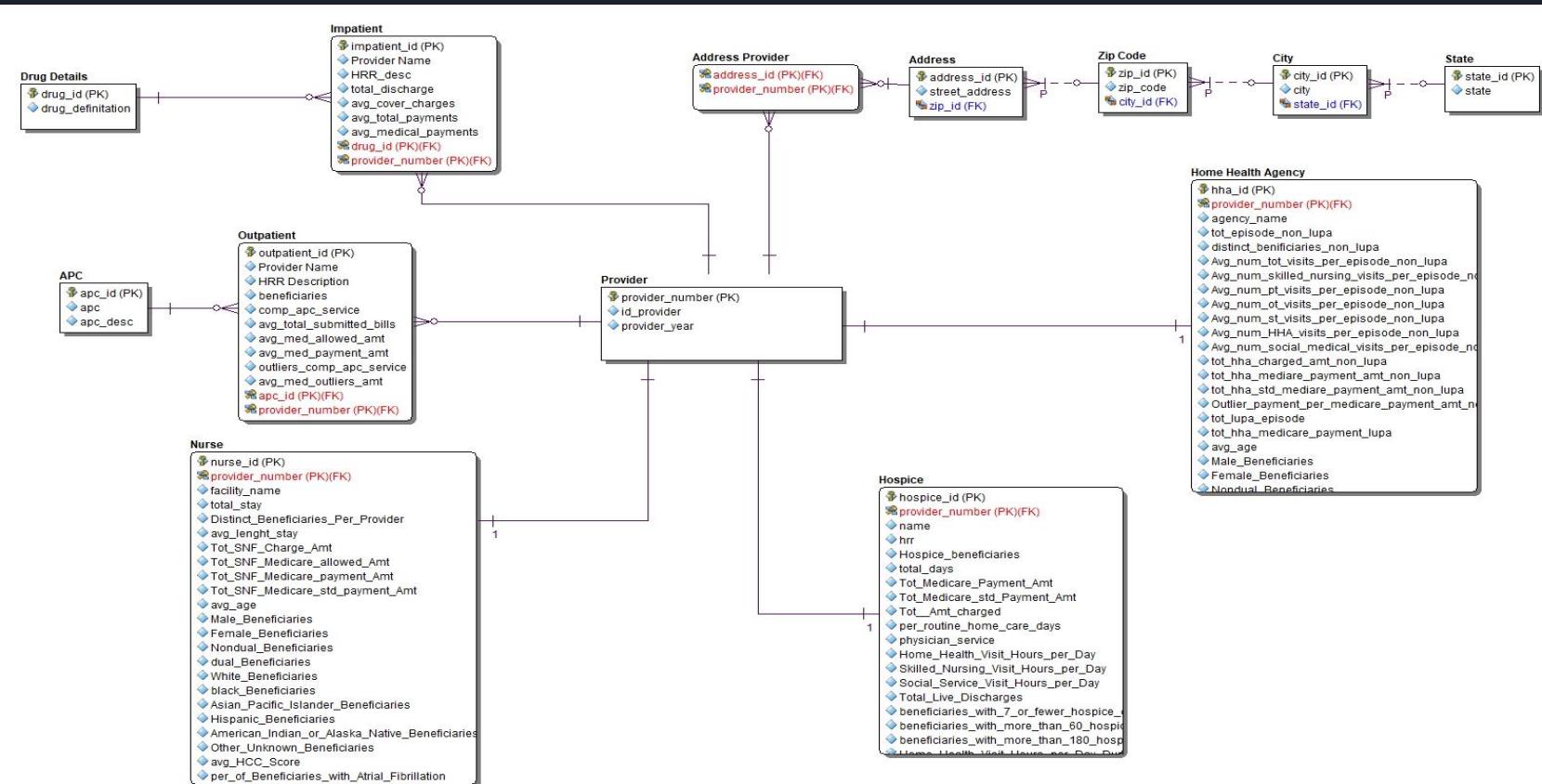
Data Preprocessing is important for data consistency. In order to make data pure and consistent, few of the techniques where used:

1. Since the Provider ID and Zip Code were of data type integers in the data source, the initial “0” were removed automatically in the data. We changed the data type to “varchar” and using the concatenating technique appended “0” in the required field
2. We removed Null values from the fields like Provider ID, State, City, APC Details, and Drug Details.
3. In order to maintain the naming consistency, we made all the rows of Provider Name, Street Address, and City lowercase and State Upper Case
4. We trimmed out white spaces and unwanted characters from important columns like Drug Definition which was leading to duplicate values.

Error Handling - Lookup Tables (Staging-2)



Data Schema - Physical Model (ERD Model)





Unique ID Creation

- We created an unique Id field using provider key and year of the data.
E.g. For provider id '100010' and year 2015, the unique id is 1000102015.
- This unique key is used as an primary key as the primary key in the destination table.



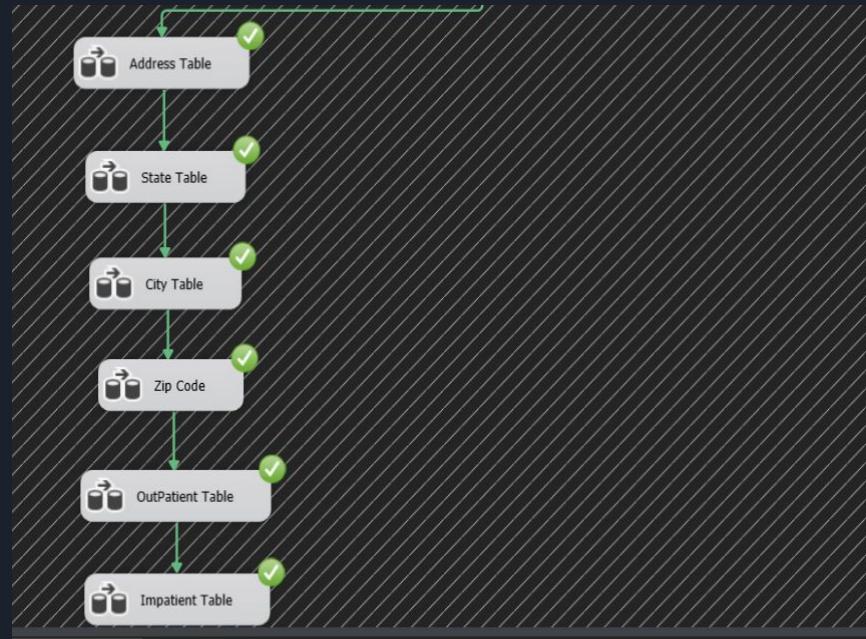
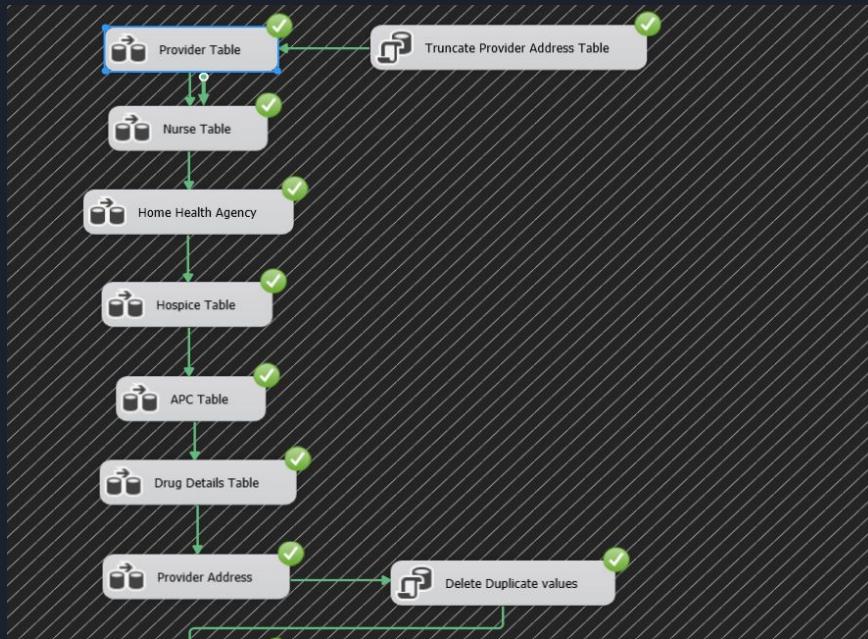
Data Loading for ERD

Data Flow Diagram was loading the data from source into our Data Schema.

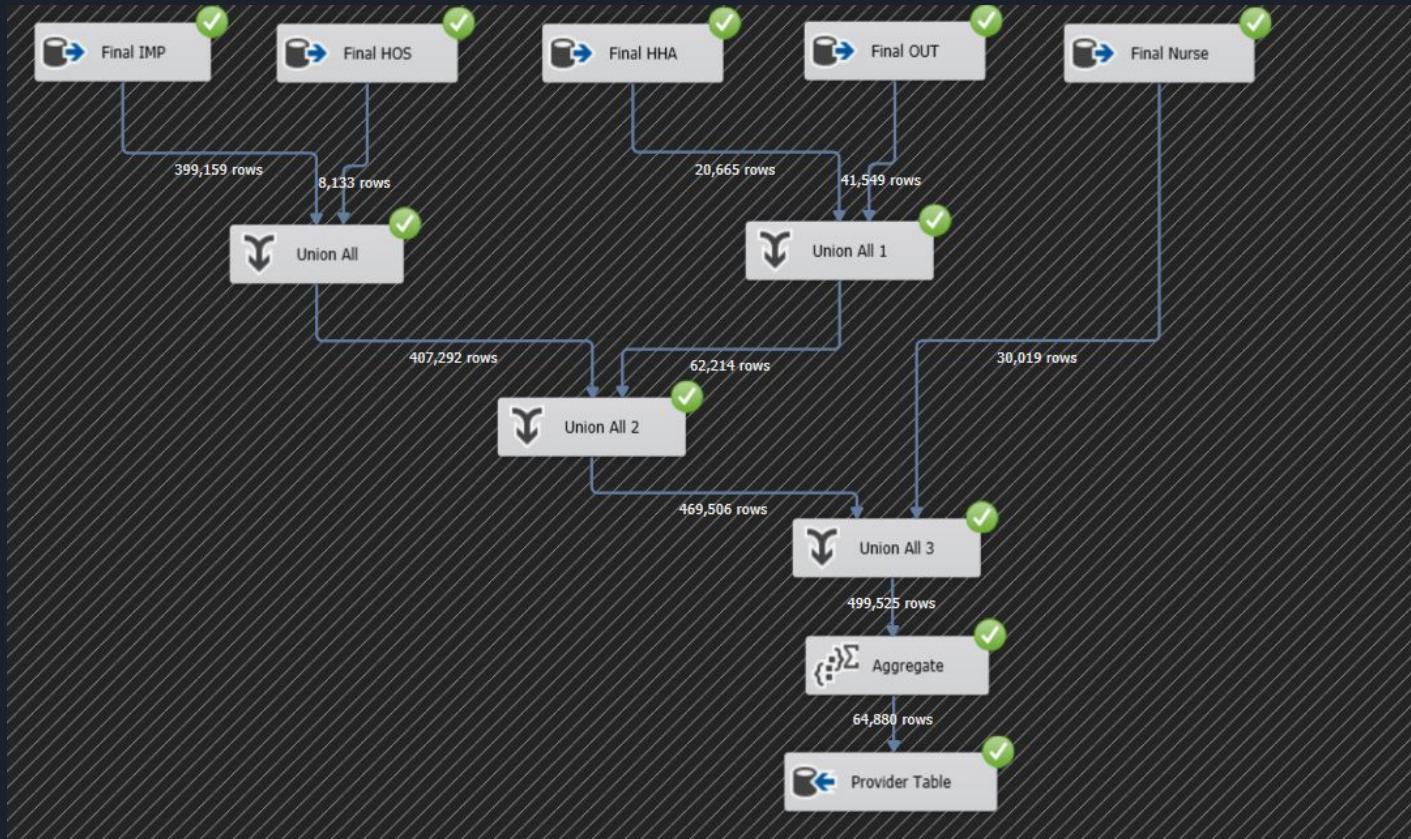
Data Schema was designed with the help of ERStudios.

We used Visual Studio Integration Package (SSDT) to data integration.

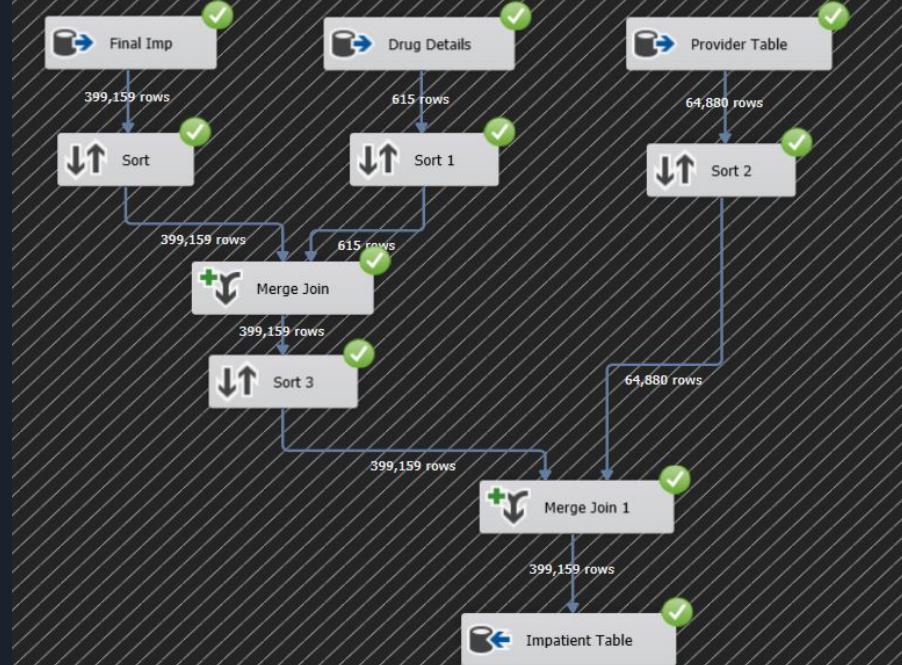
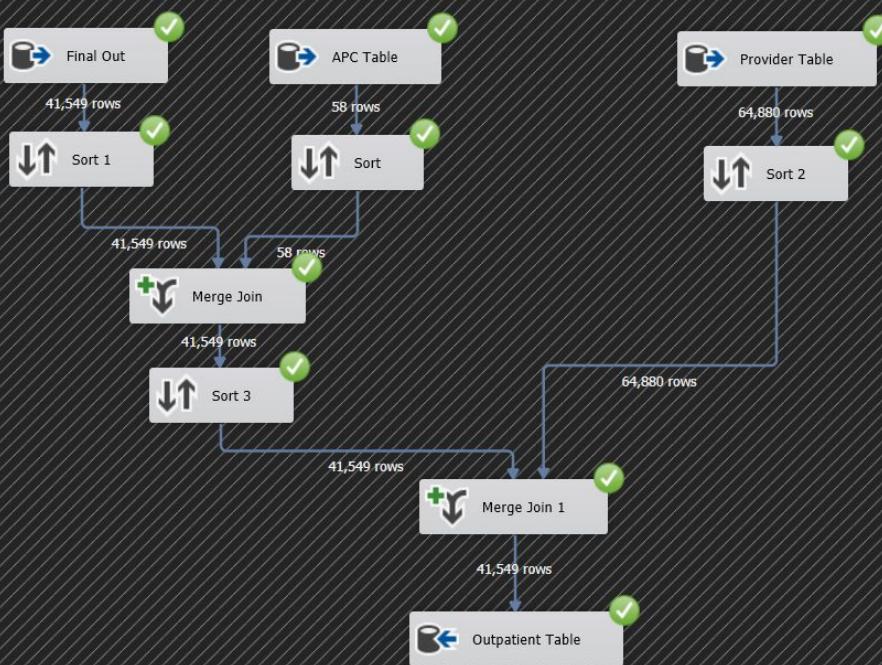
Visual Studio's SSIS View



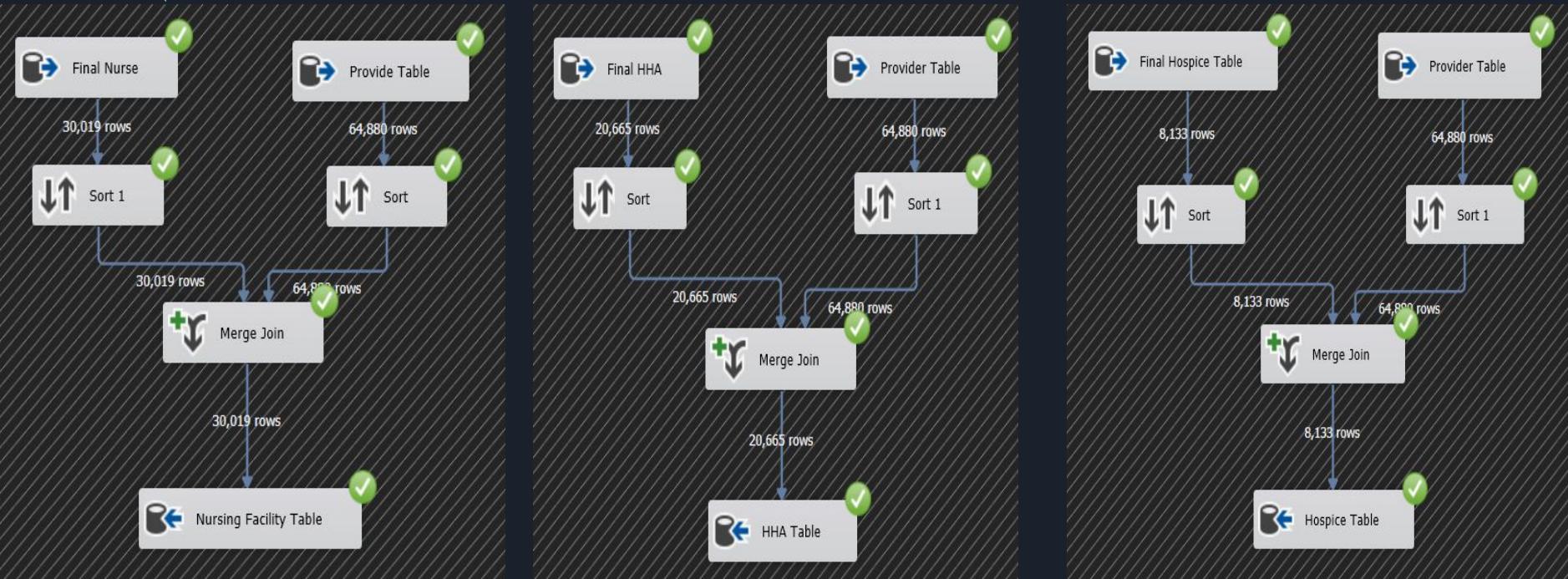
Data Flow for Provider Table



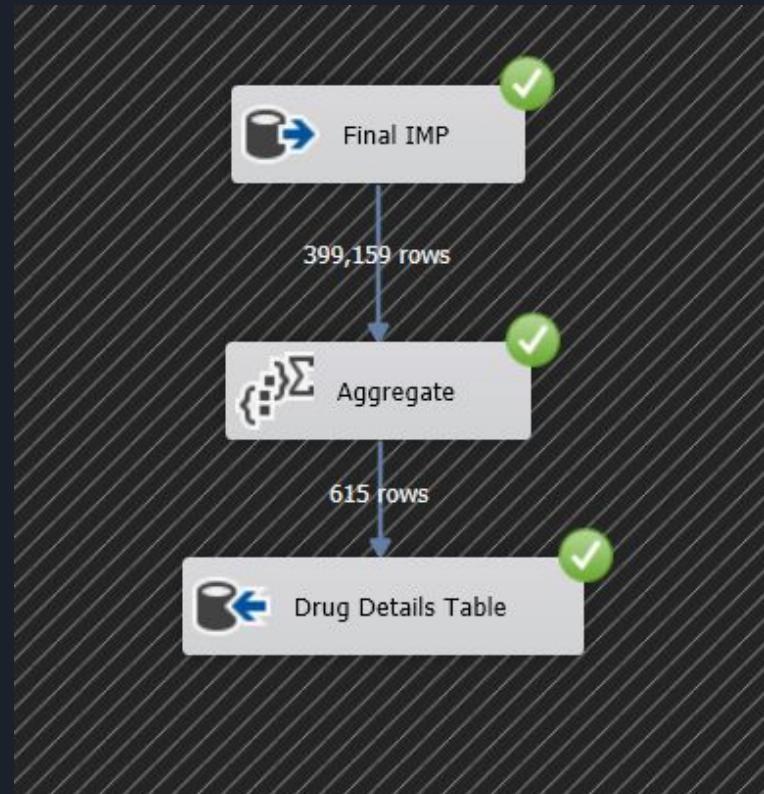
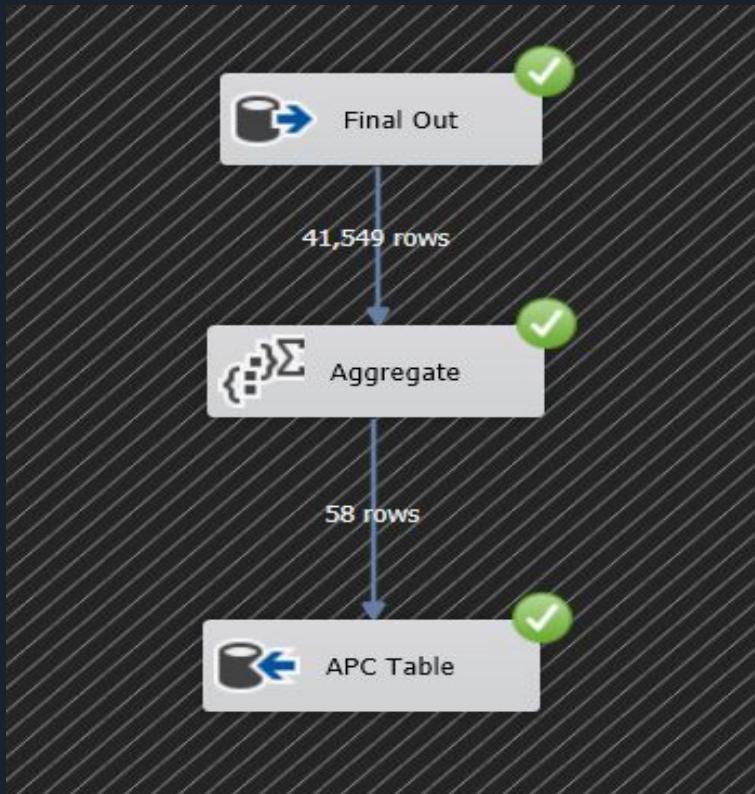
Data Flow for Inpatient and Outpatient Table



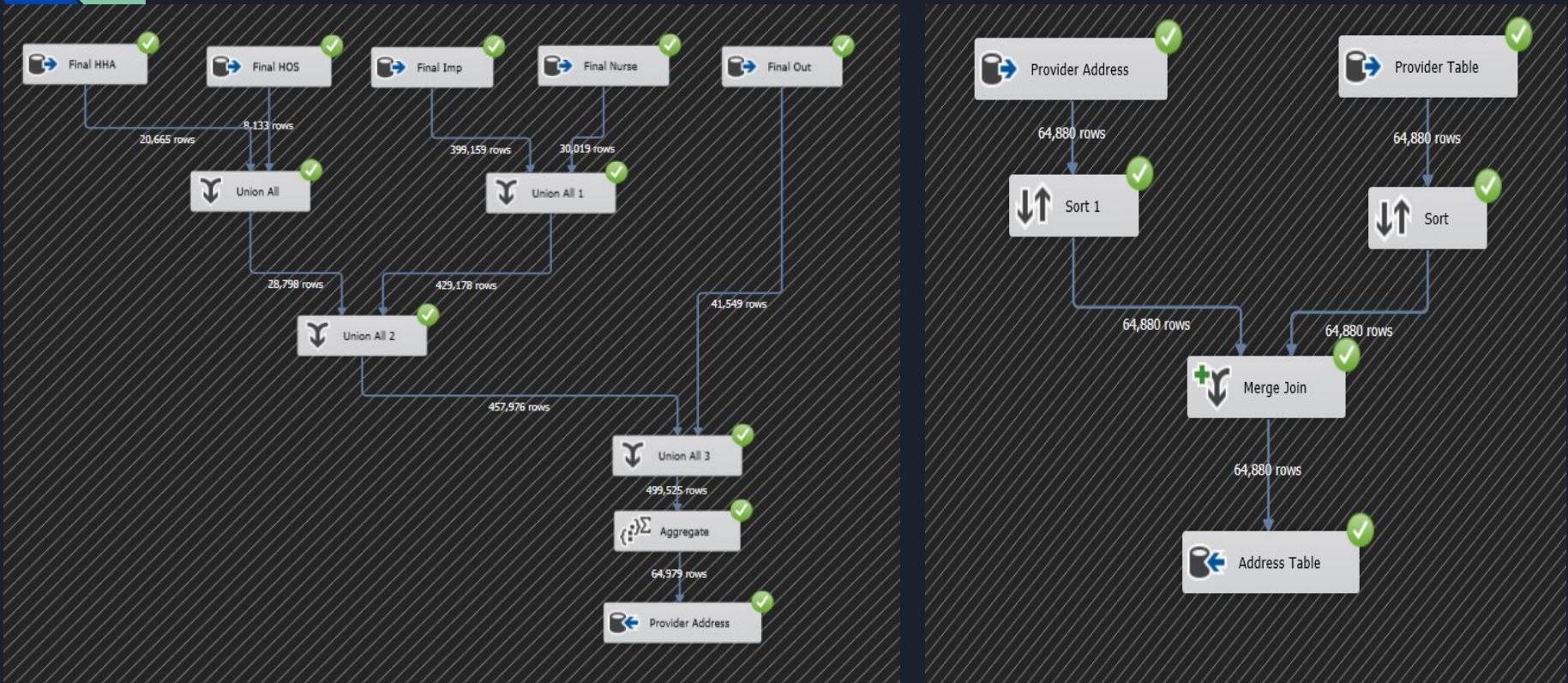
Data Flow for Nursing, Hospice and Home Health Agency Table



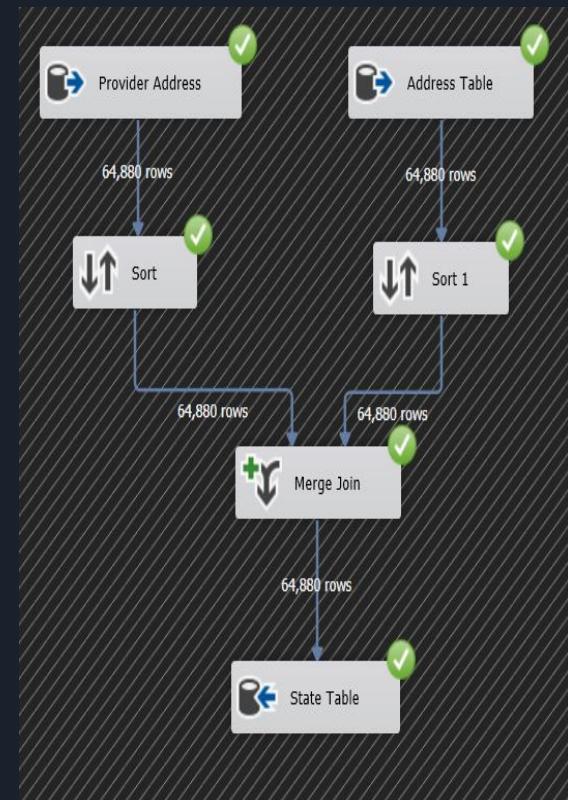
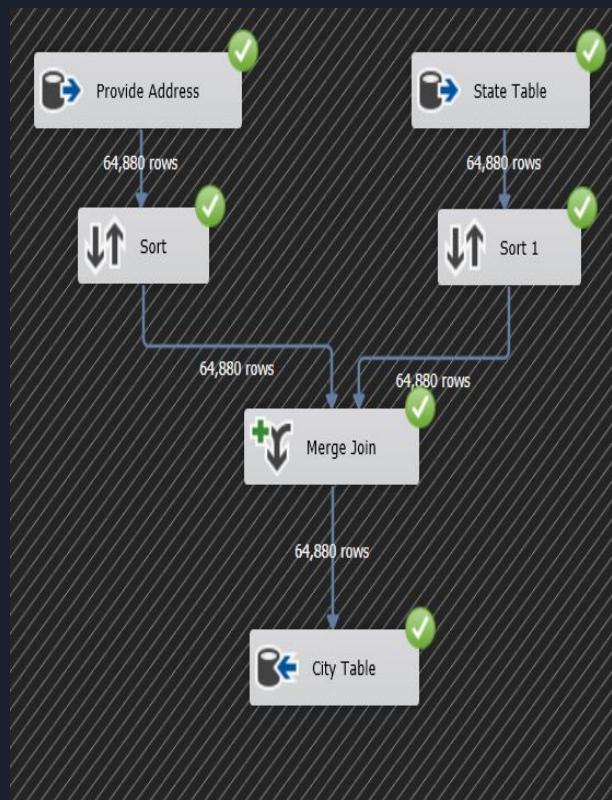
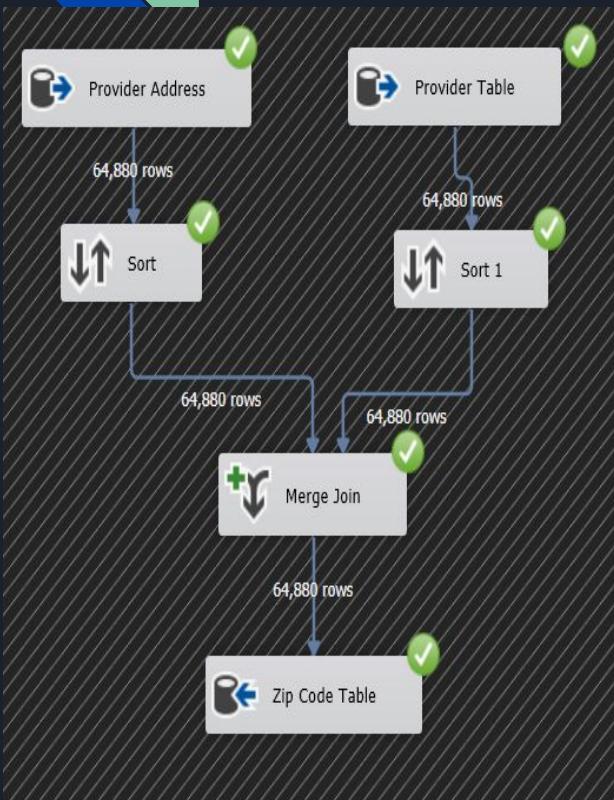
Data Flow for APC Details and Drug Details Table



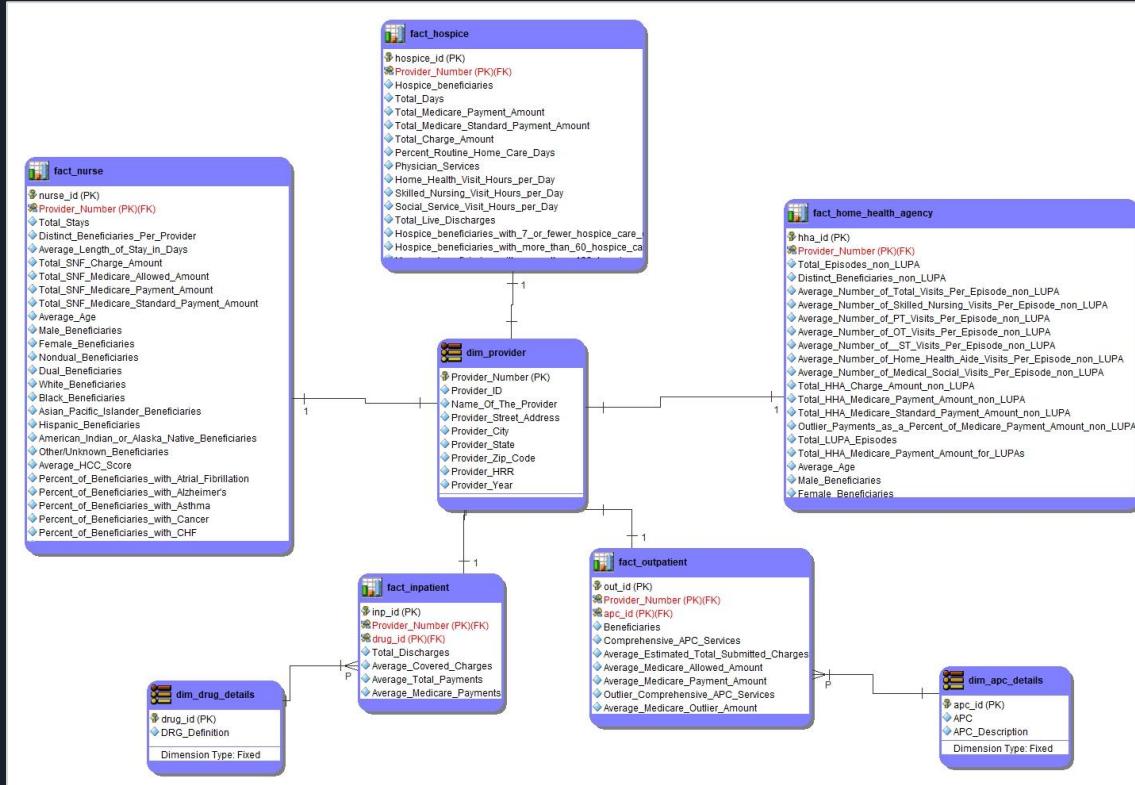
Data Flow for Provider Address and Address Table



Data Flow for Zip Code, City, and State Tables

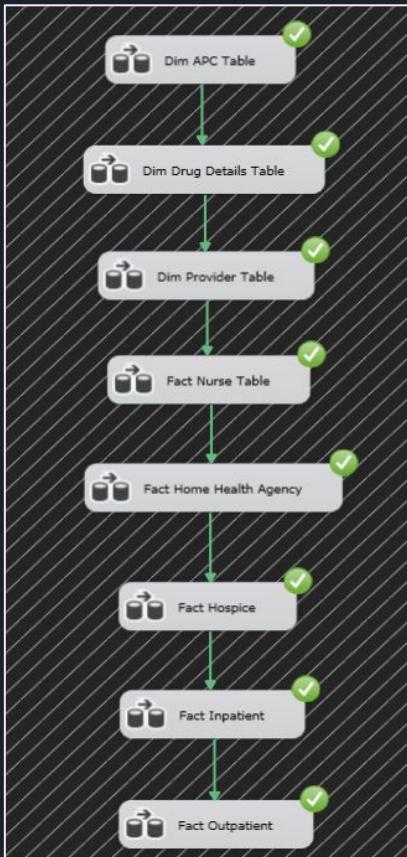


Data Mart : Dim Fact Model



- ERStudio
- 5 Fact Table
- 3 Dimensions
- Purpose:
Dim Fact Model
Multidimension
Data Mart
Visualization

ETL SSIS Package



- Software:

Visual Studio

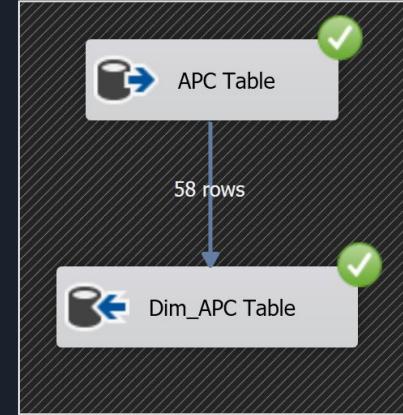
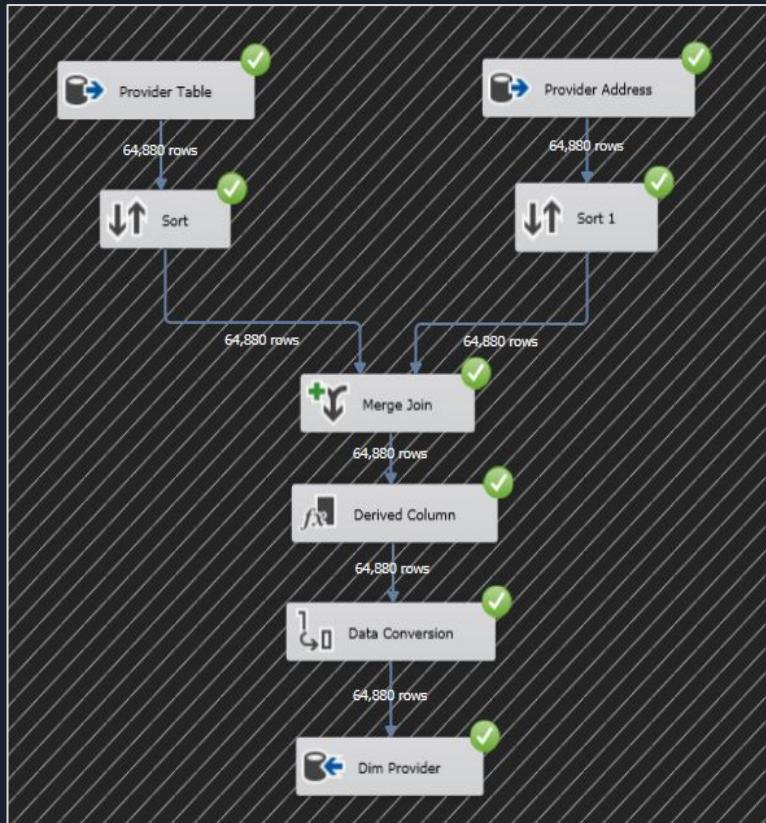
- Purpose:

Table Creation

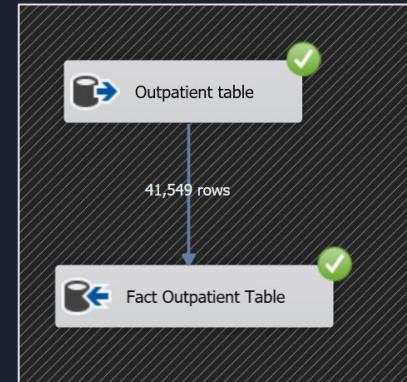
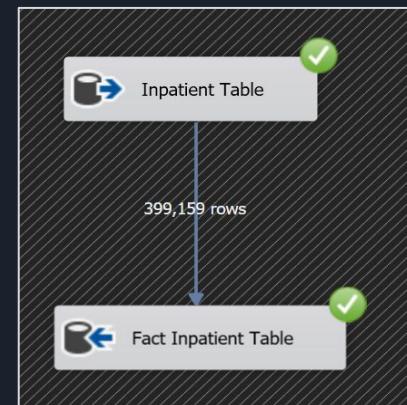
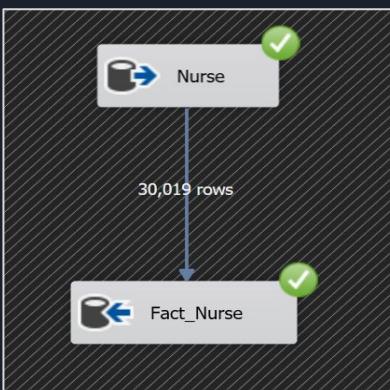
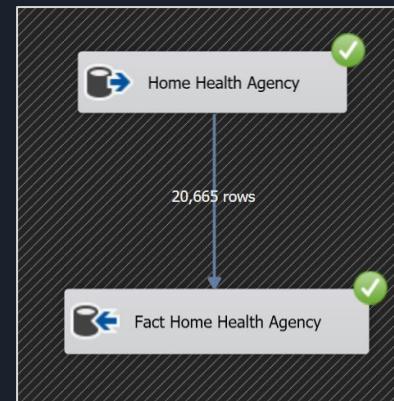
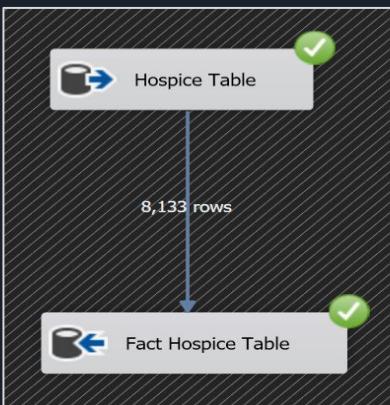
Relationship (Parent-Child)

Data Loading (From ERD Schema)

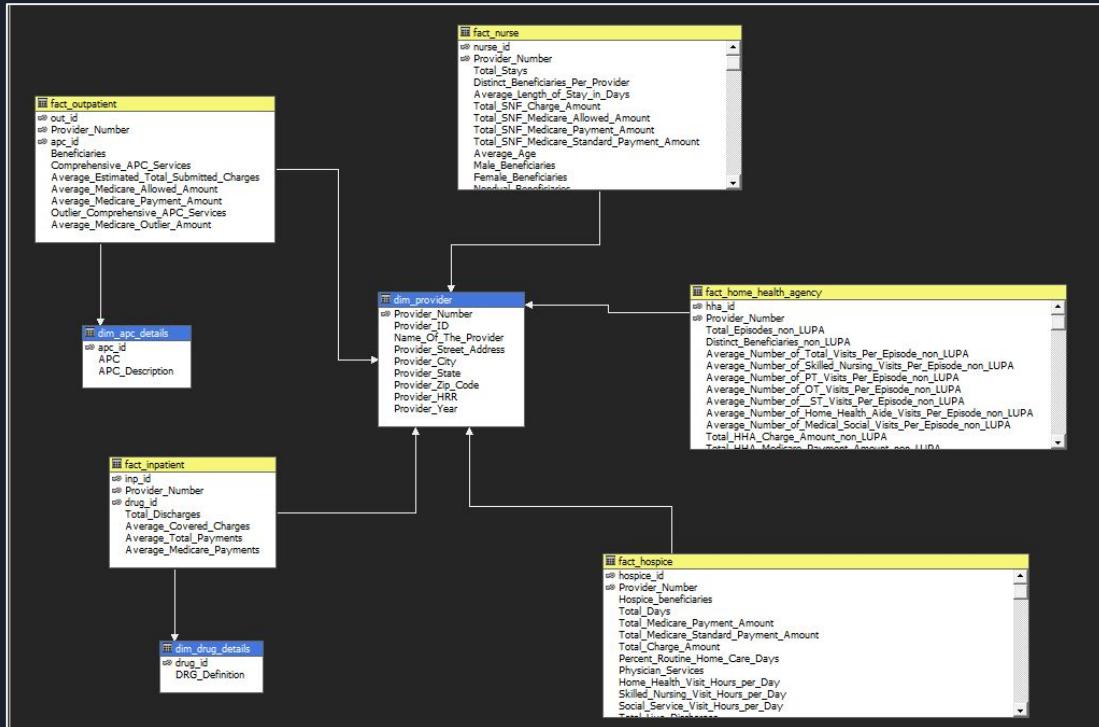
Dimension Table for Fact-Dim Model



Fact Table for Fact-Dim Model



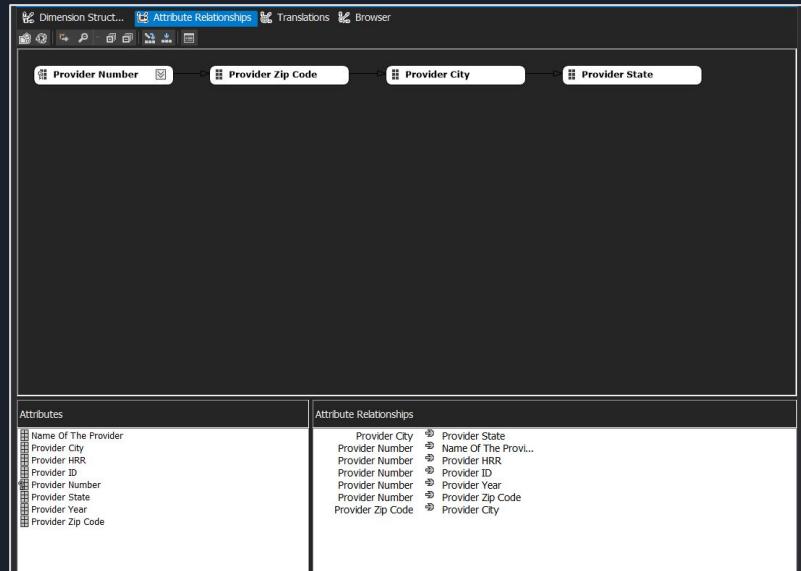
OLAP Multidimensional Cube



- Software: Visual Studio
- Mode: Multidimensional
- Fact Tables: 5
- Dimensions: 3
- Purpose: Cube Process Hierarchy Data Browse

Hierarchy

The screenshot shows the SSIS Dimension Designer interface. The main window displays the 'Hierarchy' tab under the 'Dim Provider.dim' dimension. A tooltip indicates: 'To create a new hierarchy, drag an attribute here.' The 'Attributes' pane on the left lists attributes: Dim Provider, Name_of_The_Provider, Provider_ID, Provider_HNR, Provider_ID, Provider_Number, Provider_State, Provider_Year, and Provider_Zip_Code. The 'Properties' pane at the bottom shows the 'Basic' section with 'Name' set to 'Dim Provider'. The 'Solution Explorer' pane on the right shows the project structure, including the OLAP cube and dimensions.



Cube Processing

The screenshot shows a Microsoft Visual Studio interface with several windows open, illustrating the process of cube deployment and its results.

Deployment Progress - OLAP_Project_dim_Fact: This window displays deployment details for the 'OLAP_Project_dim_Fact' project. It shows the server as 'LAPTOP-0AHSUBJSVS2019' and the database as 'OLAP_Project_dim_fact'. A message at the bottom states 'Deployment Completed Successfully'.

Process Cube - Project Dim Fact: This window shows the 'Object list' for the 'Project Dim Fact' cube. It includes columns for Object Name, Type, Process Options, and Settings. The entry for 'Project Dim Fact' is listed as a 'Cube' with 'Process Full' selected.

Process Progress: This window provides a detailed log of the processing activity. It lists the following completed tasks:

- Processing Cube 'Project Dim Fact' completed.
 - Start time: 4/23/2020 9:24:33 PM; End time: 4/23/2020 9:24:44 PM; Duration: 0:00:11
 - [m] Processing Measure Group 'Fact Home Health Agency' completed.
 - [m] Processing Measure Group 'Fact Hospice' completed.
 - [m] Processing Measure Group 'Fact Inpatient' completed.
 - [m] Processing Measure Group 'Fact Nurse' completed.
 - [m] Processing Measure Group 'Fact Outpatient' completed.

Solution Explorer: This window shows the solution structure for 'OLAP_Project.dim.Fact'. It includes a 'Data Sources' node with 'Project Dim Fact.ds', a 'Cubes' node with 'Project Dim Fact.cube' (which is selected), and other nodes like 'Dimensions', 'Mining Structures', 'Roles', 'Assemblies', and 'Miscellaneous'.

Properties: This window displays properties for 'Project Dim Fact.cube'. Key details include:

- Location**: File Name: 'Project Dim Fact.cube', Full Path: 'C:\Users\sshai\source\repos\OLA...'.
- Object Model**: Object ID: 'Project Dim Fact', Object Name: 'Project Dim Fact'.
- File Name**: Specifies the name of the file.

Taskbar: The taskbar at the bottom of the screen shows the Windows Start button, a search bar with 'Type here to search', and various pinned application icons. The system tray indicates the date as '4/23/2020' and the time as '9:25 PM'.

Cube Browser

OLAP_Project_dim_Fact - Microsoft Visual Studio

File Edit View Project Build Debug Team Database Cube Tools Test Analyze Window Help

Deployment Progress - OLAP_Project... Default Start

Server Explorer Toolbox

Project Dim Fact.cube [Design] Quick Launch (Ctrl+Q) Sign in

Language: Default

Command

Deployment Completed Successfully

Deployment Progress SSIS Toolbox

Deployment Progress - OLAP_Project... Default Start

Server : LAPTOP-0AHSUBJS5S2019
Database : OLAP_Project_dim_Fact

Project Dim Fact.cube [Design]

Dimension Hierarchy Operator Filter Expression Parameters

<Select dimension>

Provider ID Provider Zip Code Provider City Provider State Average Covered ... Average Medical...

Provider ID	Provider Zip Code	Provider City	Provider State	Average Covered ...	Average Medical...
010001	36301	dothan	AL	15412288.04	2908880.79
010005	35957	boaz	AL	2691396.62	918069.69
010006	35631	florence	AL	11798320.12	2093967.64
010007	36467	opp	AL	493263.44	184975.6
010008	36049	luvner	AL	78614.49	22941.05
010011	35235	birmingham	AL	9432917.8	1798313.08
010012	35968	fort payne	AL	2926631.23	346164.74
010016	35007	alabaster	AL	8442502.2	1258812.89
010018	35233	birmingham	AL	65221.46	21285.08
010019	35660	sheffield	AL	2824408.99	646337.32
010021	36360	ozark	AL	520530.09	213303.88
010022	35960	centre	AL	201829.28	55068.41
010023	36116	montgomery	AL	10208458.35	2379427.49
010024	36106	montgomery	AL	10396259.18	1942594.24
010029	36801	opelika	AL	5803144.8	2387214.64
010032	36278	wedowee	AL	87901.01	26703.46
010033	35233	birmingham	AL	46134267.85	8879934.3900
010034	36078	talasse	AL	212972.73	102975.48
010035	35058	culman	AL	3780934.37	1009010.83
010036	36420	andalusa	AL	1914373.68	307119.93
010038	36201	anniston	AL	3404558.55	323847.94
010039	35801	huntsville	AL	37628691.08	5626102.93
010040	35903	gadsden	AL	26938827.93	1984550.13
010044	35570	hamilton	AL	227680.21	66641.41

Solution Explorer

Search Solution Explorer (Ctrl+Shift+F)

Solution 'OLAP_Project.dim_Fact' (1 project)

- OLAP_Project.dim_Fact
 - Data Sources
 - Project Dim Fact.ds
 - Data Source Views
 - Project Dim Fact.dsv
 - Cubes
 - Project Dim Fact.cube
 - Dimensions
 - Dim Provider.dim
 - Dim Drug Details.dim
 - Dim Apc Details.dim
 - Mining Structures
 - Roles
 - Assemblies
 - Miscellaneous

Solution Explorer Team Explorer

Properties

Project Dim Fact Cube

ProcessingMode Regular

ProcessingPriority 0

Source Project Dim Fact (Data source)

Visible True

Basic

Description

ID Project Dim Fact

Name Project Dim Fact

Name Specifies the name of the object.

Properties Getting Started (SSIS)

Add to Source Control

Type here to search

Windows Taskbar

System tray: ENG 9:36 PM IN 4/23/2020

Data Visualisation

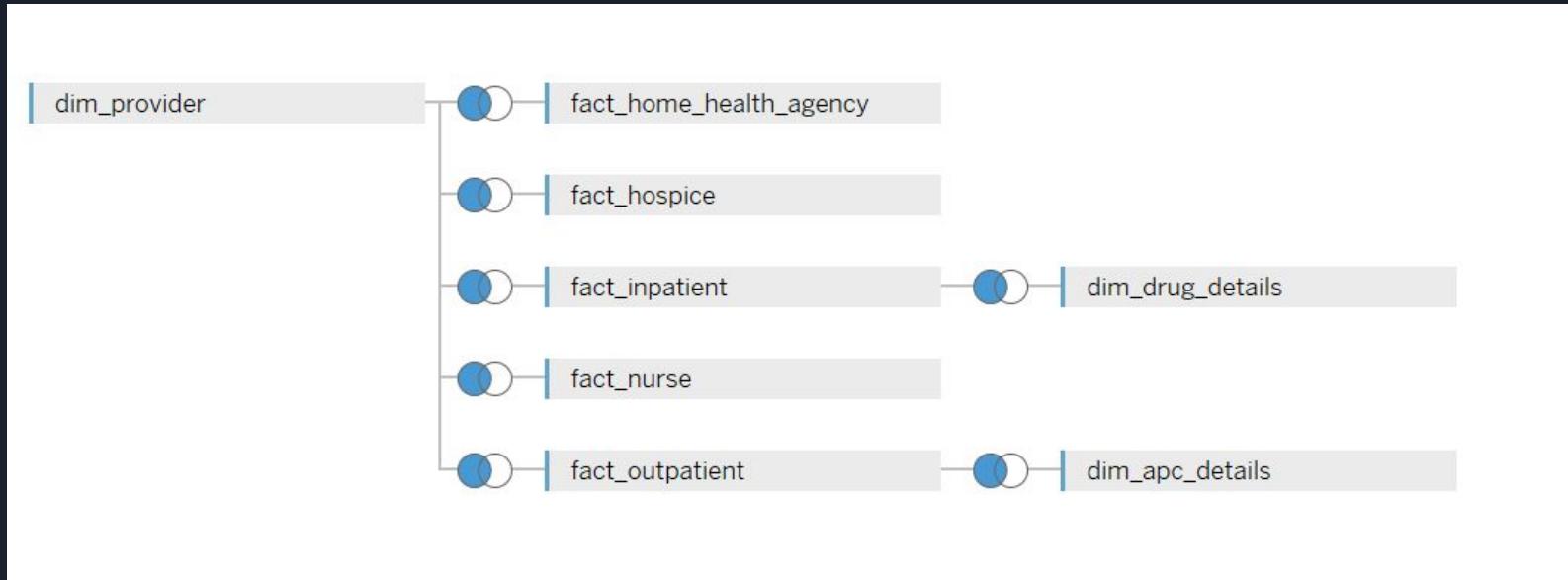


Tableau Dashboard: Health Services Comparison

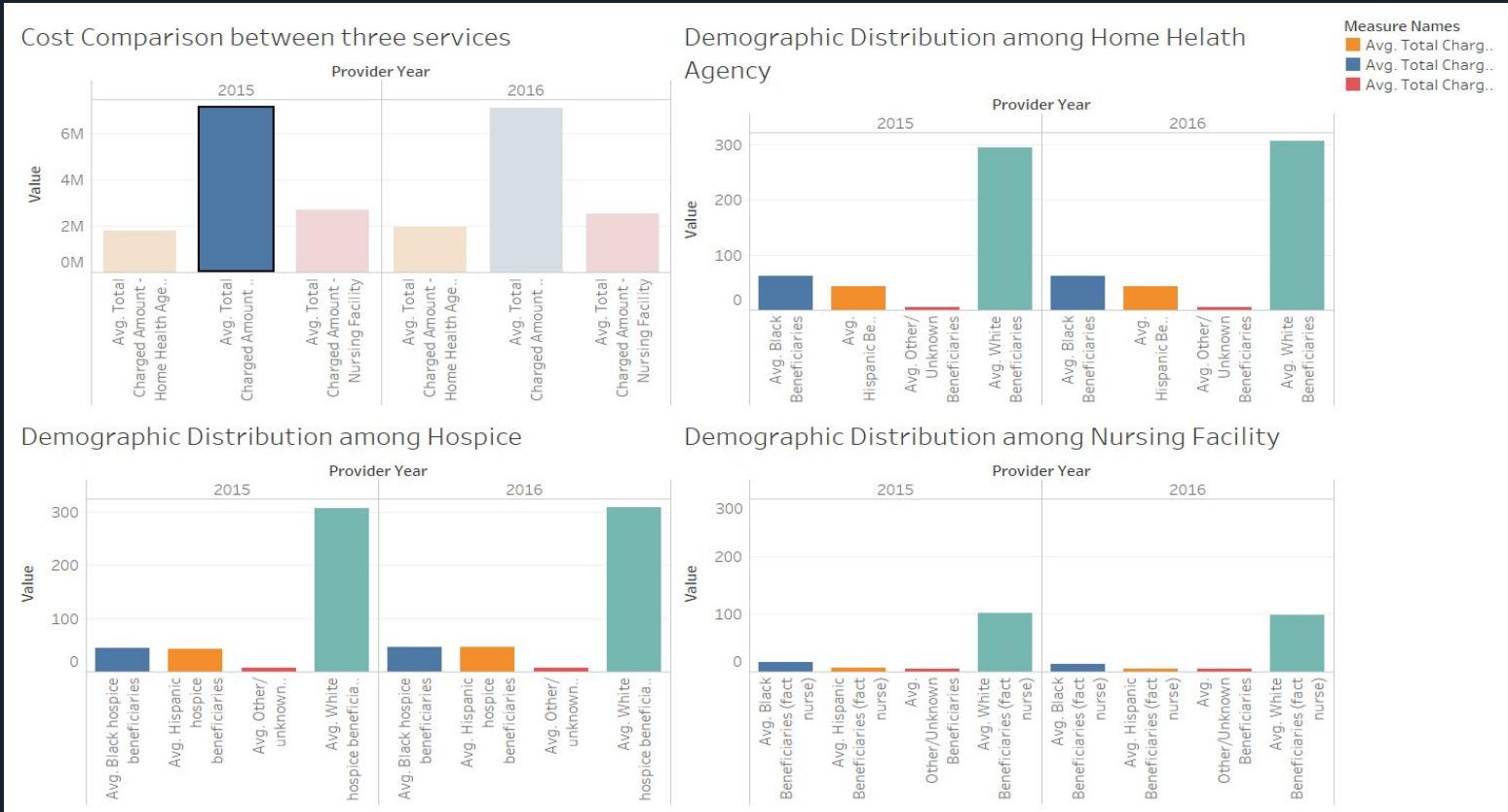


Tableau Dashboard: Disease Comparison in Nursing and Home Health Agency

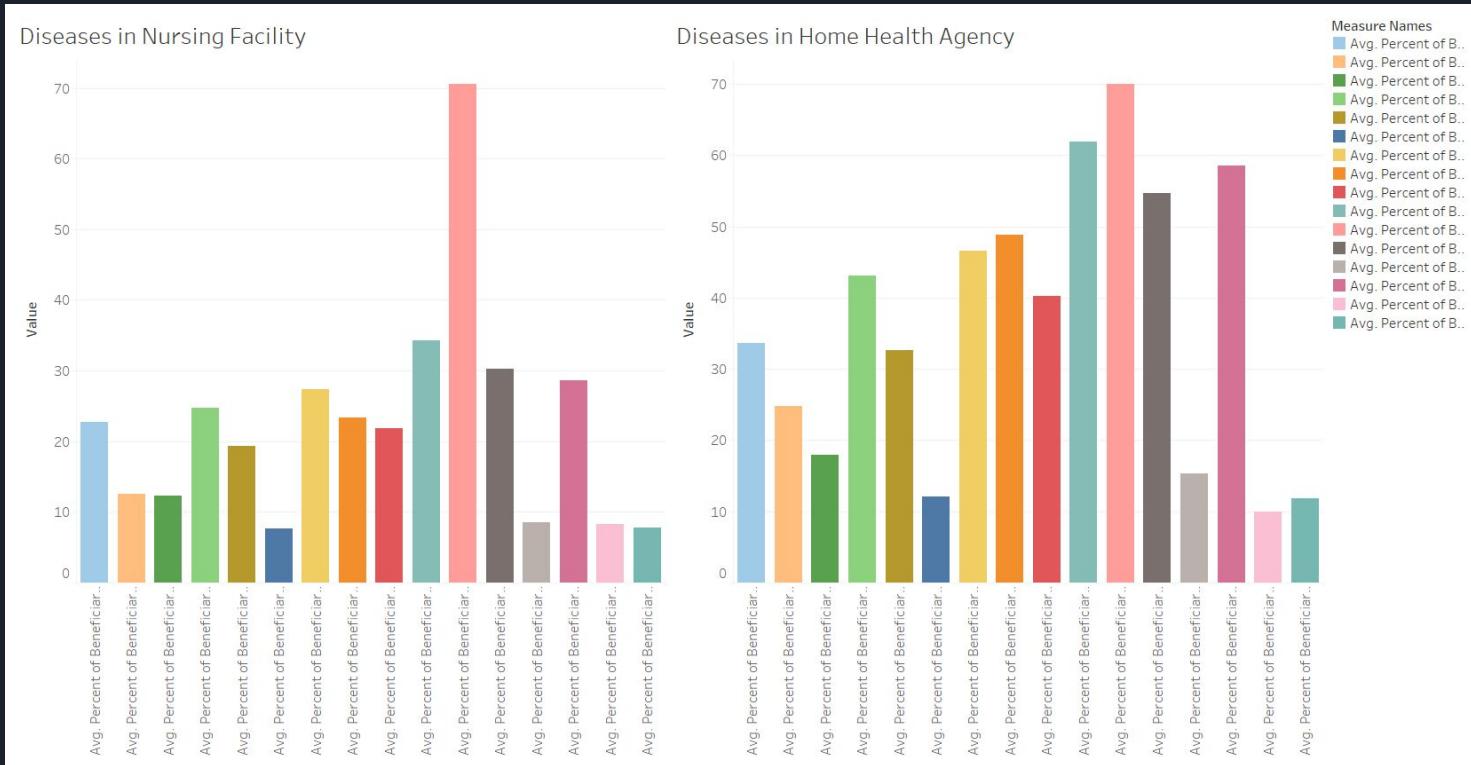


Tableau Dashboard: Beneficiary Distribution and their HCC Score

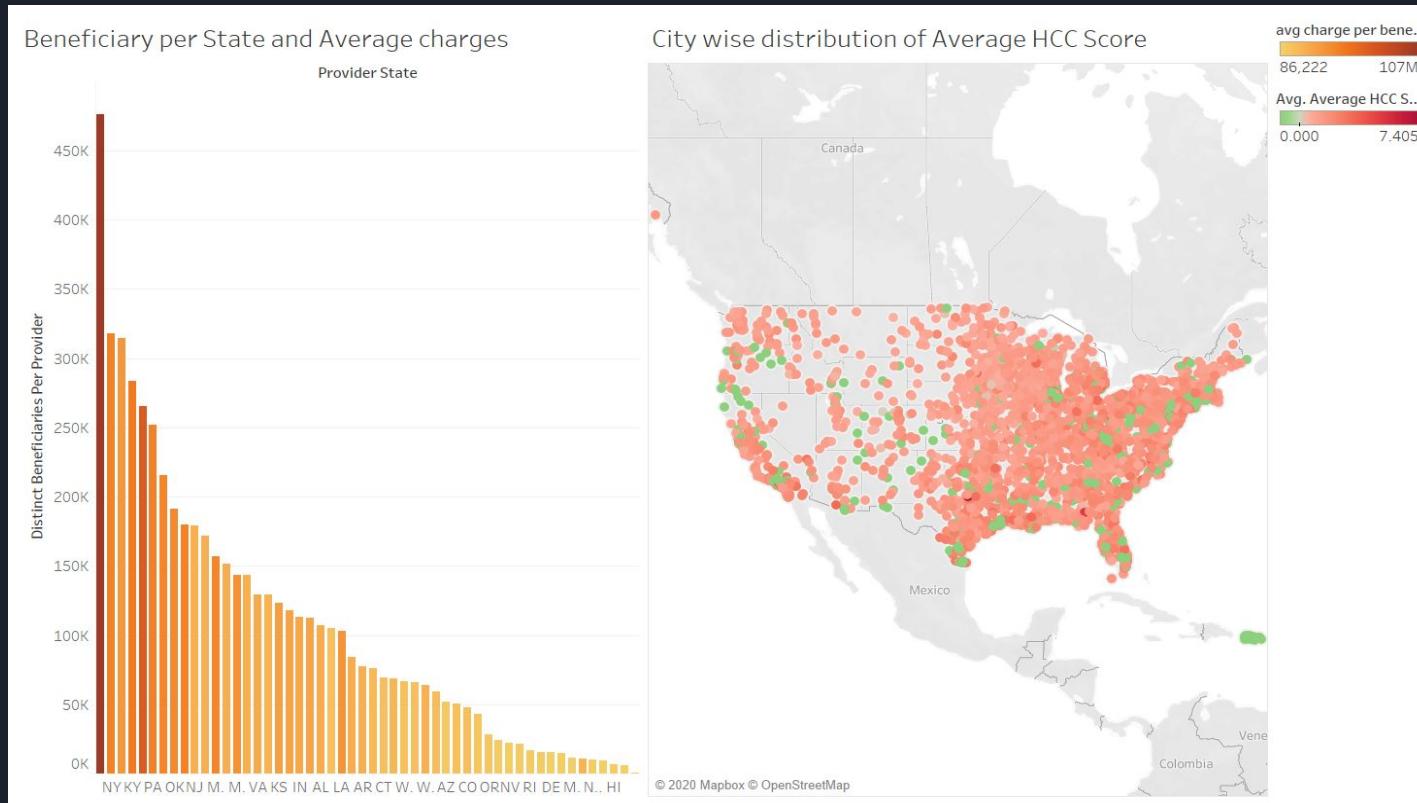
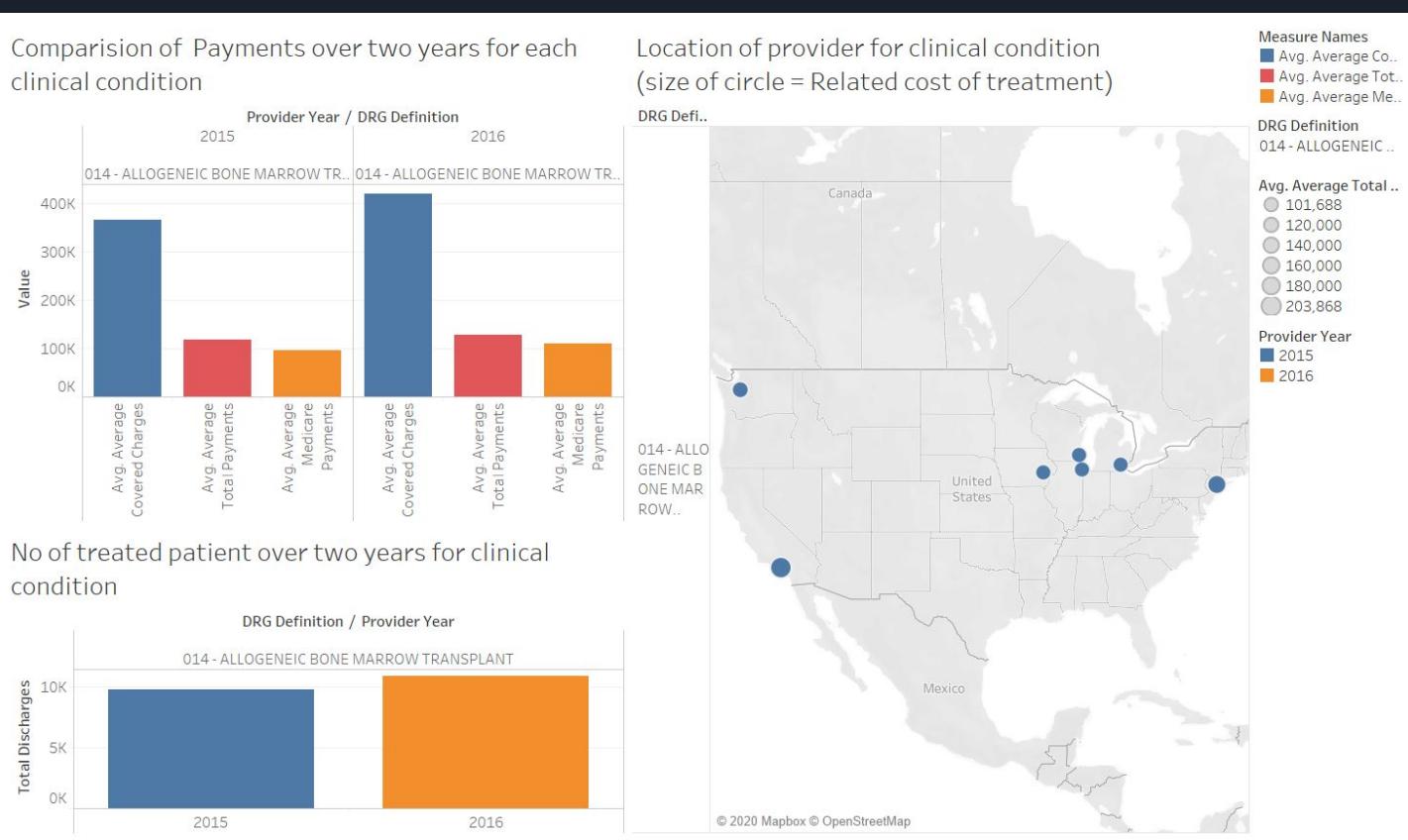


Tableau Dashboard: Drug Definition wise Payment, Discharge and Provider Distribution



Thank you!