

**Healthcare Services Data Warehouse**[INFO 7290 DATA WAREHOUSE AND BUSINESS INTELLIGENCE]  
Prof: Vincent Lattuada

**TEAM 10: TECHUSKIES 2.0**  
YASH SHINDE (Leader)  
SWAPNIL DESHPANDE  
SHASHANK SAHU  
JAY SUVARNA

**Document Revision History:**

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Who | What | Version |
| 3/15/2020 | All Members | Decided on datasets and created initial project proposal. | 0.1 |
| 3/22/2020 | All Members | Altered proposal, decided data preprocessing protocol, Added high level view and ERD | 0.2 |
| 4/08/2020 | All Members | Added Data Overview Model, Designed Data Integration Diagram, Corrected ERD, and Added Data Source Definition and Staging Table Definition | 0.3 |
| 4/17/2020 | Jay Suvarna  Swapnil & Shashank  Yash & Shashank | ETL Process Design  ETL Process execution: Staging and destination table formation and data transfer.  Compilation and documentation | 0.4 |
| 4/21/2020 | Jay & Yash  Shashank  Swapnil  Yash | Visualizations on Tableau  Data Cubes formation, Slowly Changing Dimensions  ERD, Integration Diagram, Dimensional Model  PowerPoint presentation, Report. | 0.5 |

**Objective:**

* To develop data warehouse comprising of various healthcare related datasets taken from cms.org
* To create a Data Model for healthcare datasets used using ER Studio.
* To deliver insights of number of services, Medicare payment amounts, number of beneficiaries, the number of APC services, hospitals' average total estimated submitted charges.
* Insights about the payment services used and charges in healthcare domain over United States.
* Discover relatability between suppliers, providers and medical equipment, prosthetics, orthotics, and supplies.
* Visualize the relationships on various analytical tools such as Tableau and SSAS.

**Dataset Used:**

**1. The Physician and Other Supplier Public Use File**

**Link:** <https://data.cms.gov/Medicare-Physician-Supplier/Medicare-National-HCPCS-Aggregate-Summary-Table-CY/2zuc-y5mm/data>

**Description:** The Physician and Other Supplier Public Use File (Physician and Other Supplier PUF) provides information on services and procedures provided to Medicare beneficiaries by physicians and other healthcare professionals.  
**This dataset contains columns such as variety and number of services, Medicare payment amounts.**

**2. The Outpatient Hospital Utilization and Payment Public Use File**  
**Link:** <https://data.cms.gov/Medicare-Outpatient/State-Summary-of-Outpatient-Hospital-Charge-Data-b/iwge-h6r6>

**Descriptions:**The Outpatient Hospital Utilization and Payment Public Use File (Outpatient Hospital PUF) presents information on comprehensive APC(Ambulatory Payment Classification) (C-APC) services provided to Medicare fee-for-service beneficiaries. The Outpatient Hospital PUF presents information on the**number of beneficiaries, the number of APC services, hospitals' average total estimated submitted charges, the average Medicare allowed charges**(which includes Medicare regular provider payments and beneficiary cost-sharing payments),**the average Medicare regular provider payments, the number of APC services with outlier payments,** and the average Medicare outlier provider payments among those services.

**3. The Inpatient Utilization and Payment Public Use File**

**Link:**<https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/tcsp-6e99>  
**Description:** The Inpatient Utilization and Payment Public Use File (Inpatient PUF) provides information on inpatient discharges for Medicare fee-for-service beneficiaries. The Inpatient PUF **includes information on utilization, payment (total payment and Medicare payment), and hospital-specific charges for the more than 3,000 U.S. hospitals** that receive Medicare Inpatient Prospective Payment System (IPPS) payments.

4. **Durable Medical Equipment, Prosthetics, Orthotics and Supplies (DMEPOS) National Aggregate table  
Link:** <https://data.cms.gov/Medicare-Durable-Medical-Equipment-DME-/Medicare-Referring-DMEPOS-HCPCS-National-Aggregate/gduq-te83>

**Description:** The 2017 Referring Durable Medical Equipment, Prosthetics, Orthotics and Supplies (DMEPOS) National Aggregate table presents information on DMEPOS products and services provided to Medicare beneficiaries ordered by physicians and other healthcare professionals.

**This dataset contains columns related to durable medical equipment, prosthetics, orthotics, and supplies (DMEPOS).**

**Data Preprocessing:**

* Mentioned above are the flat file data sources each having historical data of last 3 years.
* First, we will create SSIS packages for multiple years for further process.
* We are intending to accumulate all SSIS packages of flat files into a staging table, also archive table to dump the row historic data.
* Post this, we will cleanse the data using LOOKUP tables and Monitoring SCDs as learned during course work to remove data anomalies.
* This clean data will then be moved to destination table (Enterprise DW).
* Later, using ER Studio we are planning to design a Star Schema from the data.
* Once executed, the database will be ready to use for visualizations using Tableau and Power BI, SSRS, and Talend (for profiling).

Analysis / Visualization

Staging Table

Physicians and Suppliers Flat files

Star

schema

Data Warehouse

Archive Table

Medical Equipment Flat file

Inpatient Flat File

Outpatient Flat File

|  |  |
| --- | --- |
| Project Name: | Healthcare Services Data Warehouse |
| Model: | Main Model |
| Team: | TechHuskies 2.0 |
| Version: | Version 0.5 |
| Review: |  |

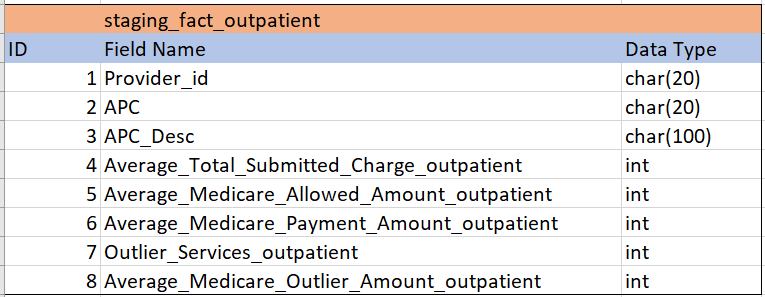
**Staging tables:**

All the four data sets will be loaded into four separate staging tables namely.

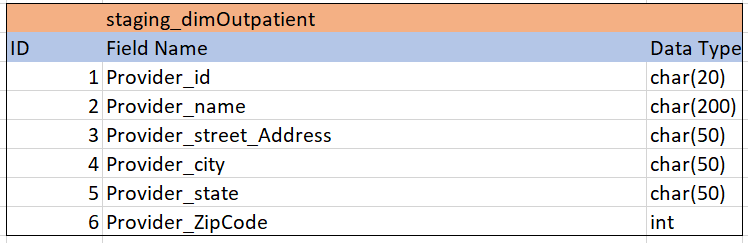
* staging\_fact\_outpatient
* staging\_dimOutpatient
* staging\_fact\_inpatient
* staging\_dimInpatient
* Medicare\_staging

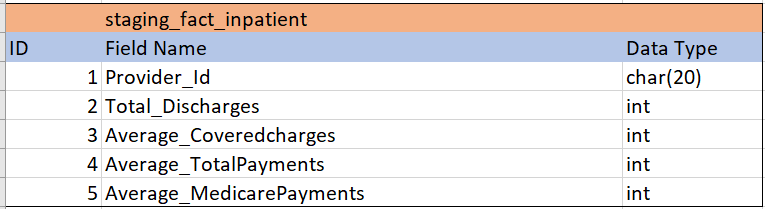
We have divided the inpatient and outpatient into fact and dimensions table separately.

**1. staging\_fact\_outpatient**

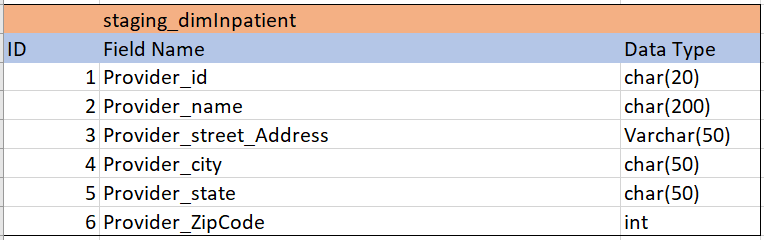


**2. Staging\_dimOutpatient**

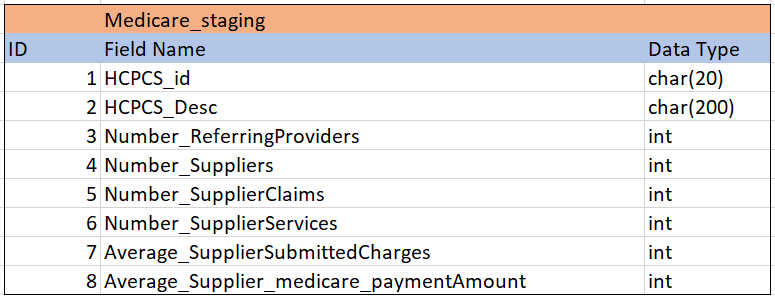


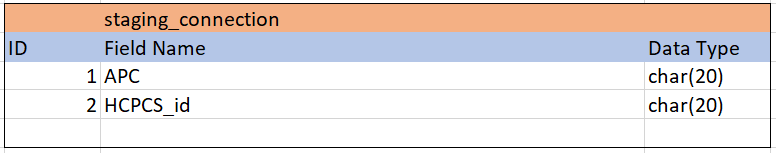
**3. staging\_fact\_inpatient**

**4. Staging\_dimInpatient**

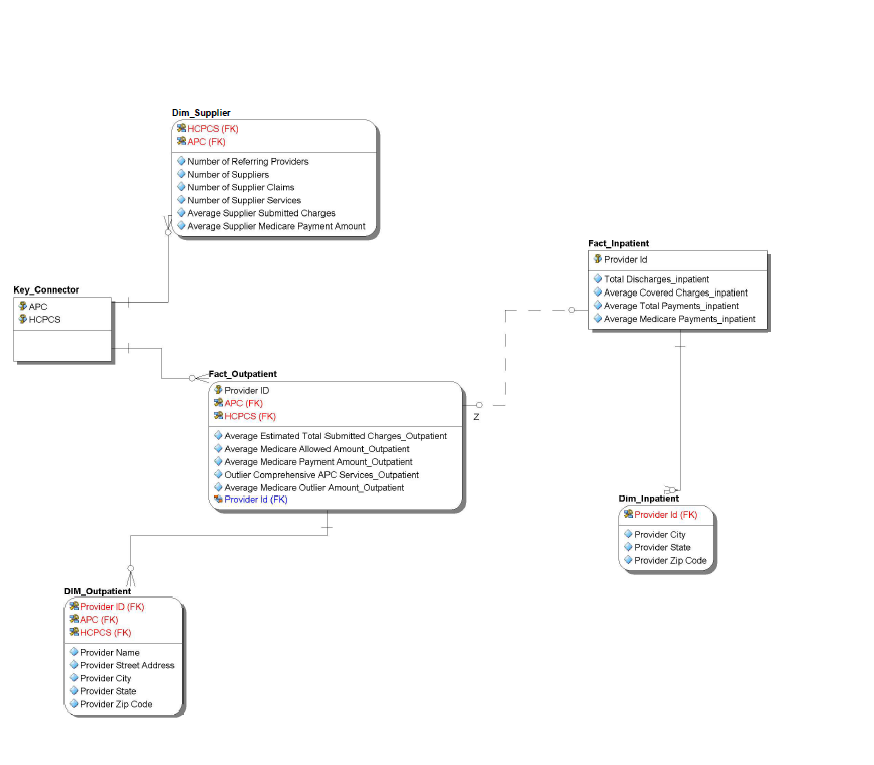


**5. Medicare\_staging**



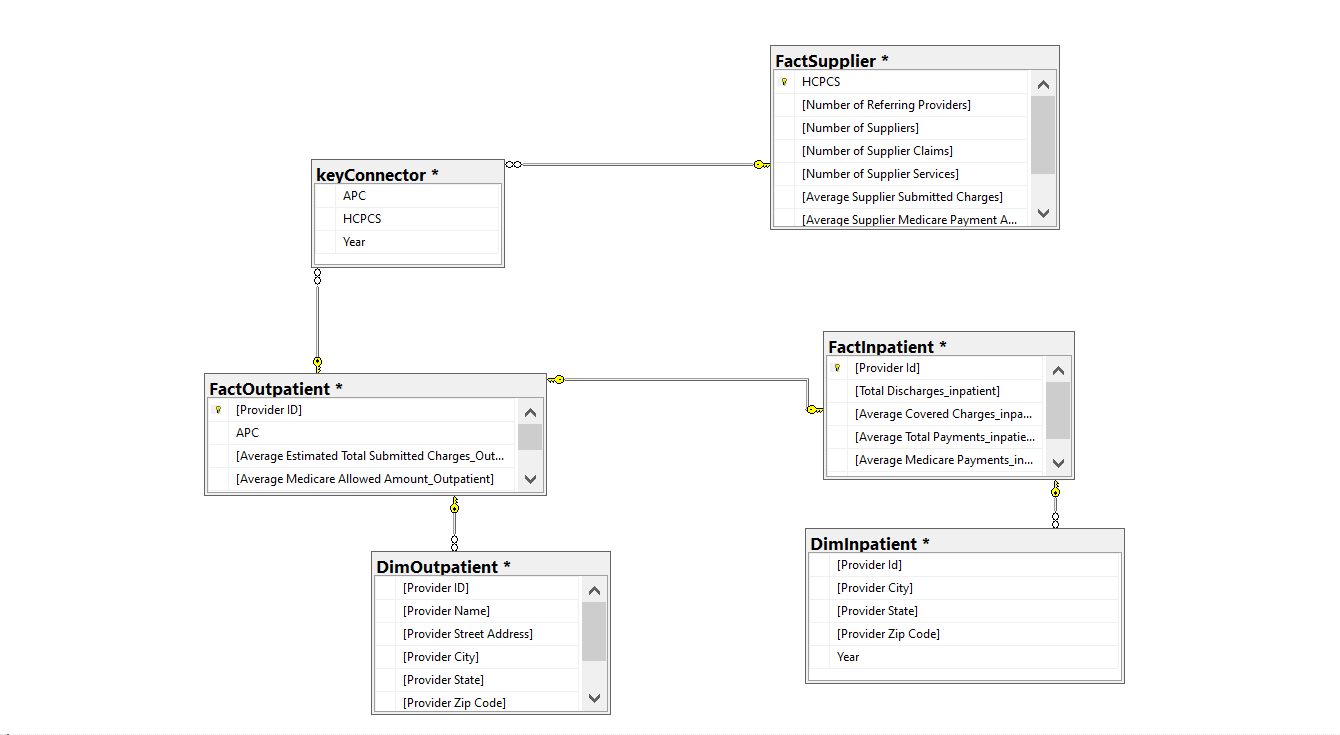
**6. Key\_Connector**

**ERD:**



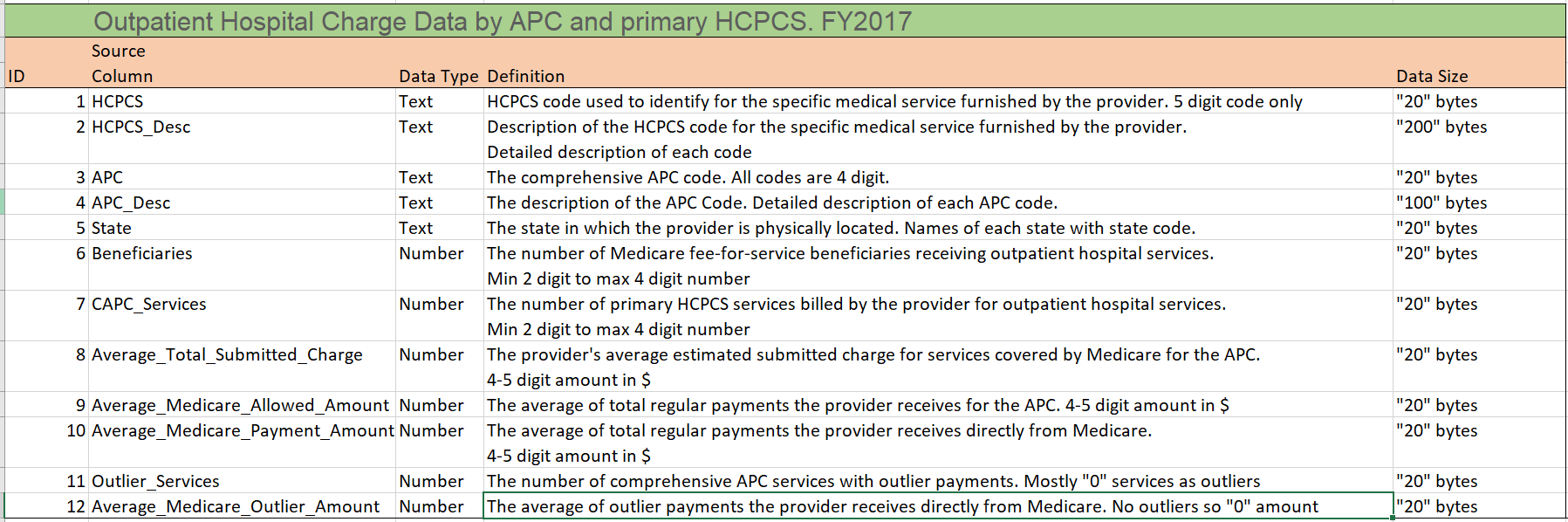
**Dimensional Model:**

The dimensional model below was created using the SQL Server Database Diagram and it shows the relationships between ever table in the Data Warehouse.

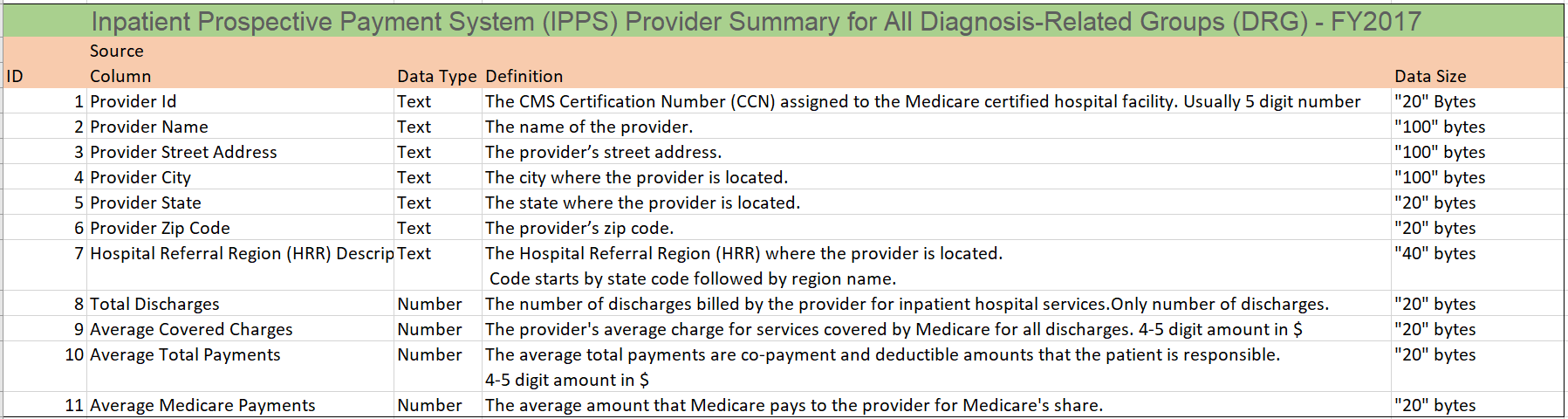


**Data Source Definitions:**

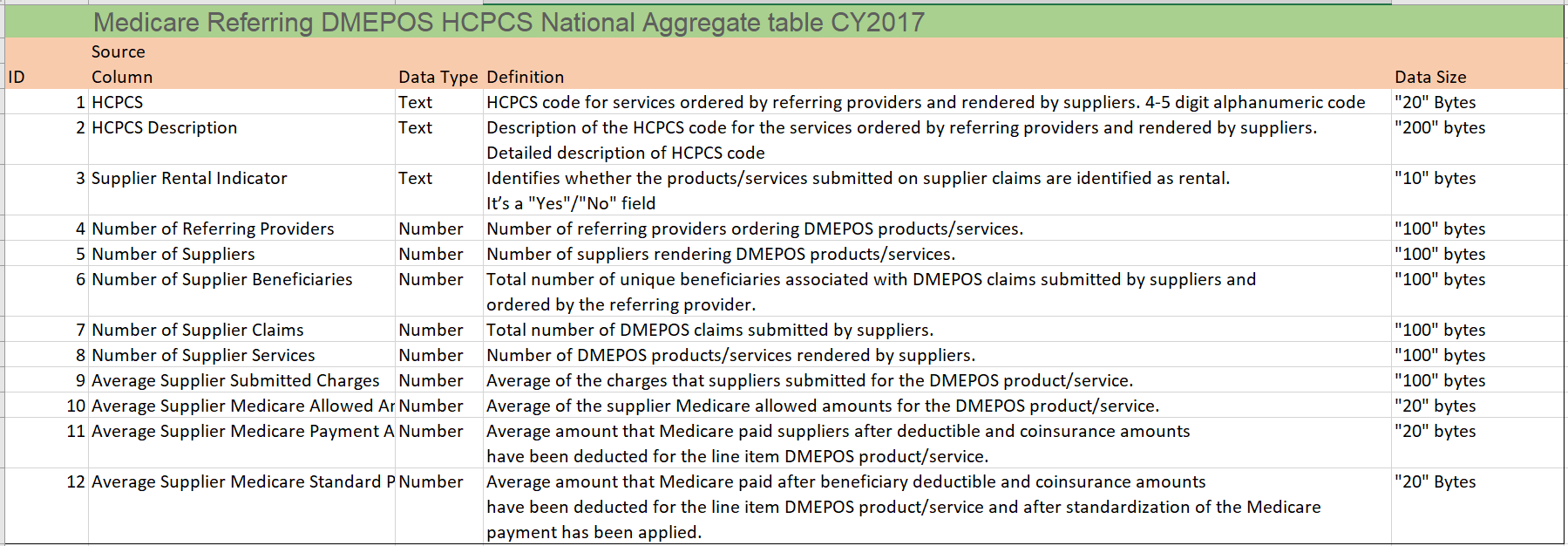
1. Outpatient Data Source

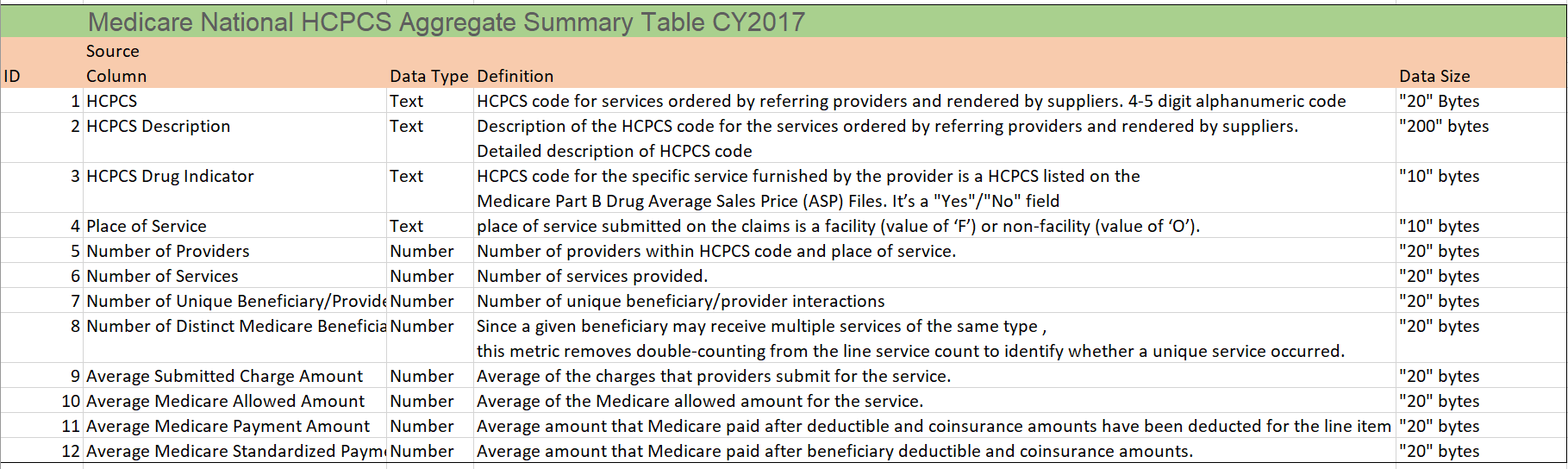


2. Inpatient Data Source



3. Medicare Supplier Data Source

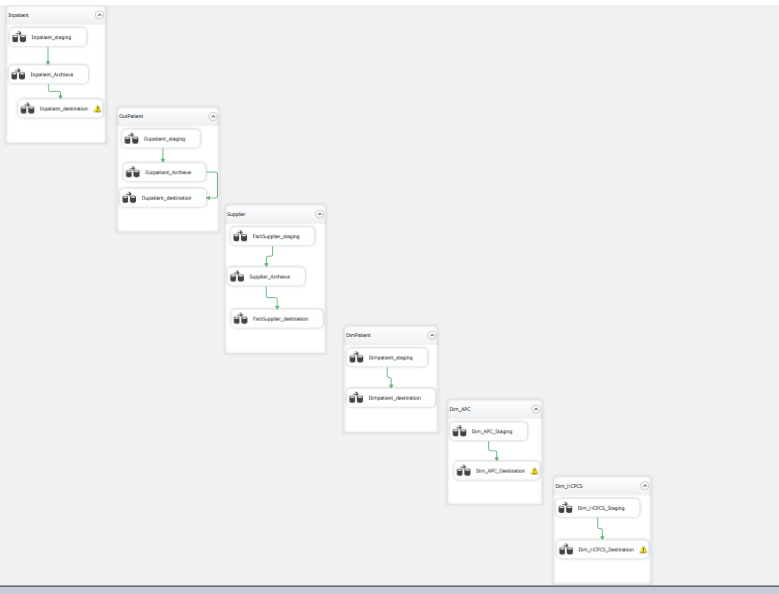




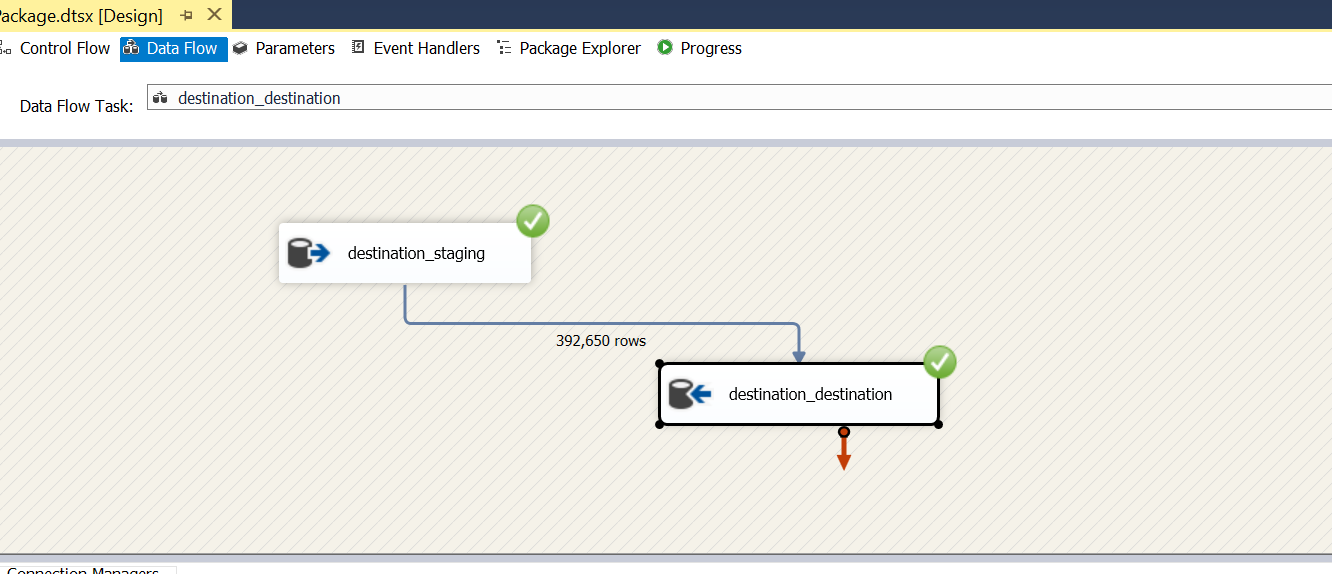
**ETL Mapping**

Here we will define the ETL process for our data flow, i.e. from source to staging and then staging to destination table.

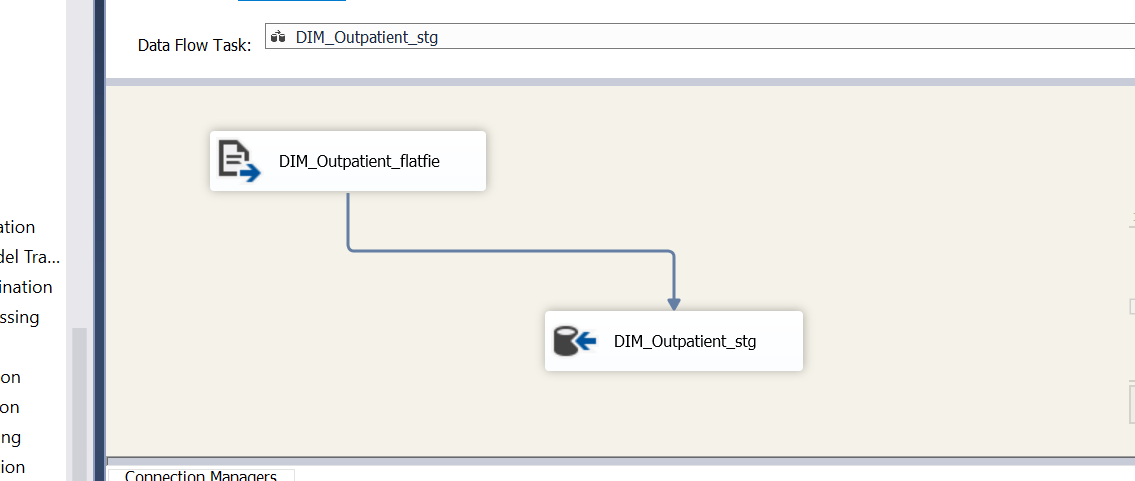
Below is the Overall Container for eacjh flow.

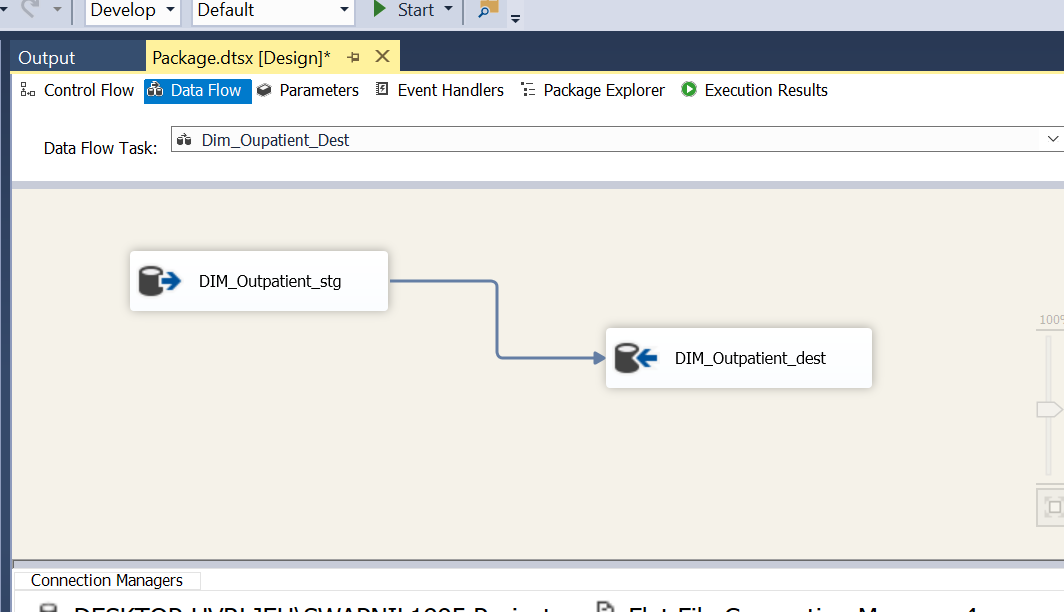


Example:



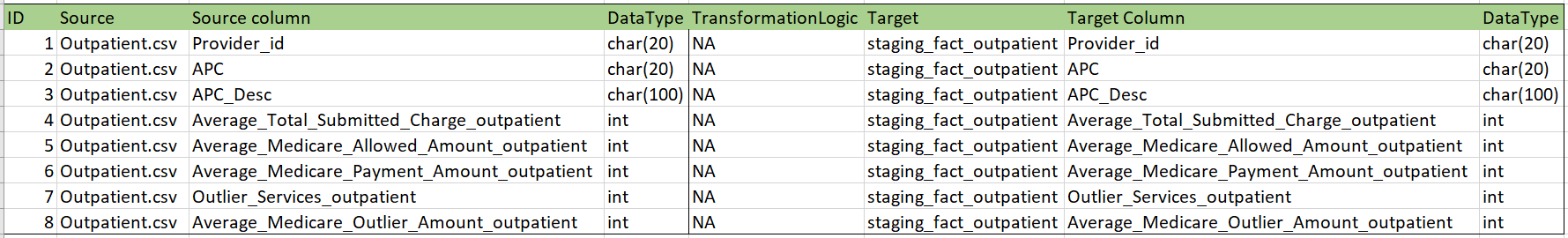
Shown below is the ETL flow of DIM\_Outpatient table. We have staged the flat-file and connected staging table to Destination table.

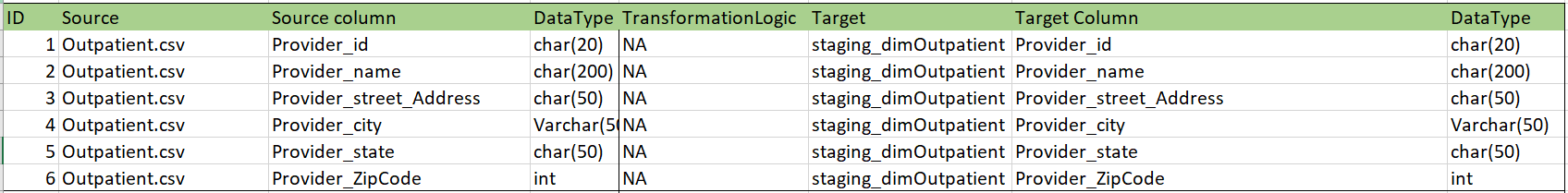


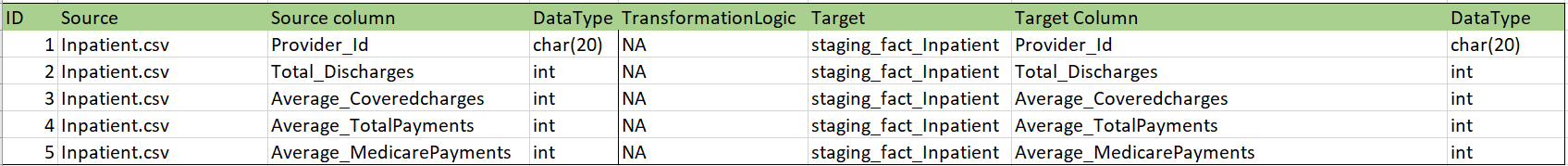


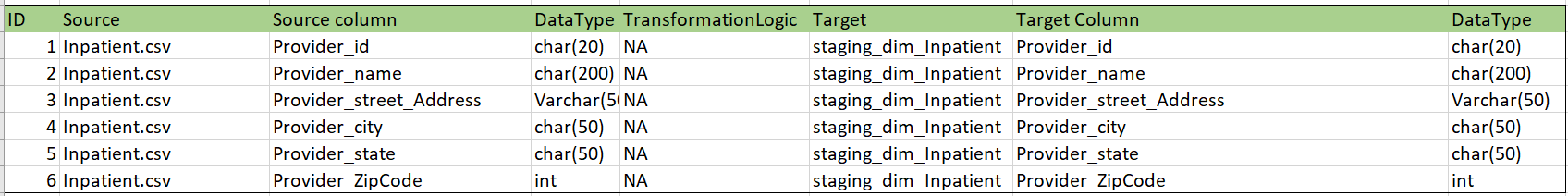
Similarly, all the tables in the ERD are staged and then connected to Destination table. We have mentioned all the attributes and corresponding datatypes.

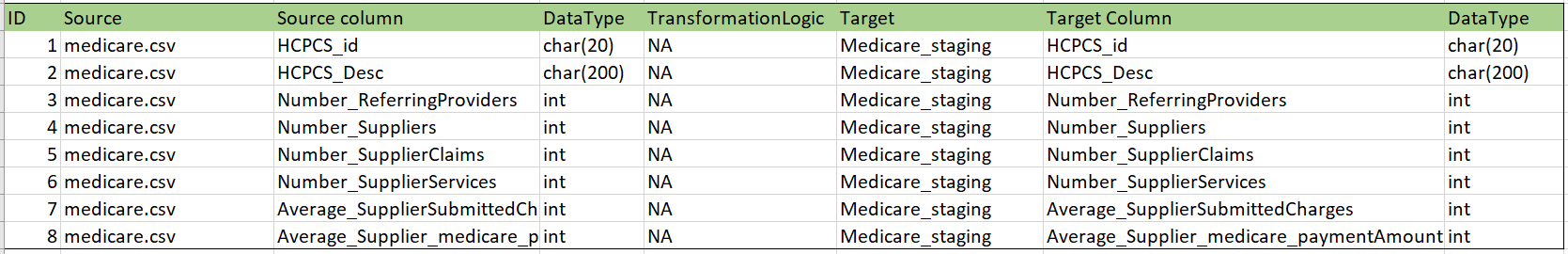
1. Flat File Source to Staging destination table



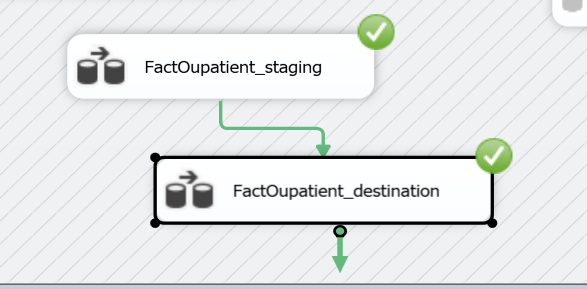


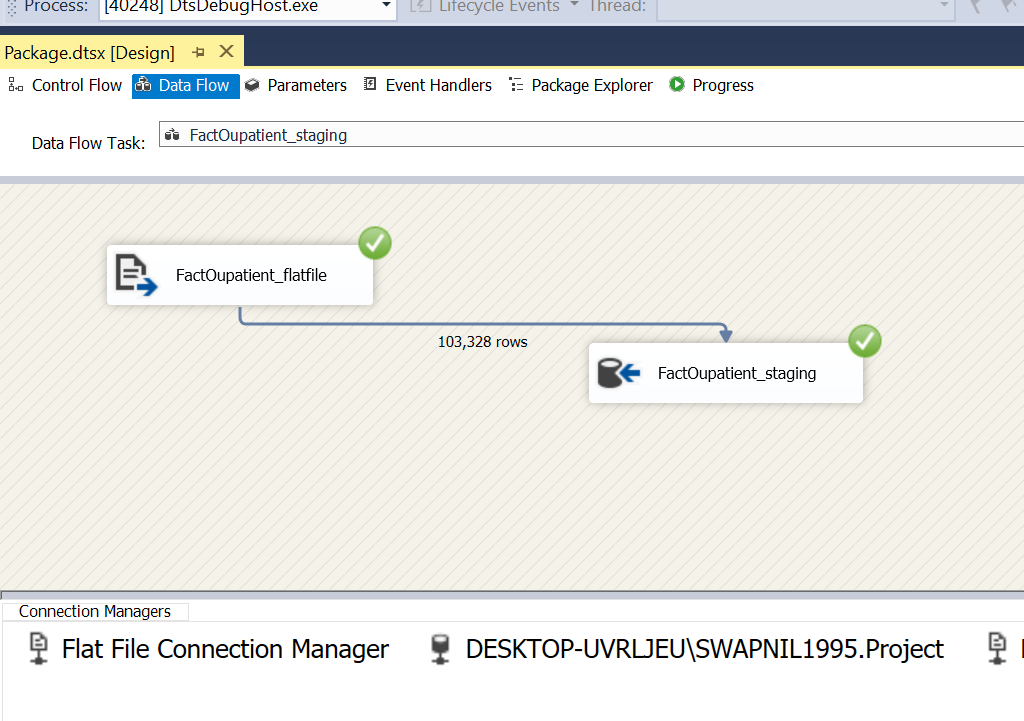


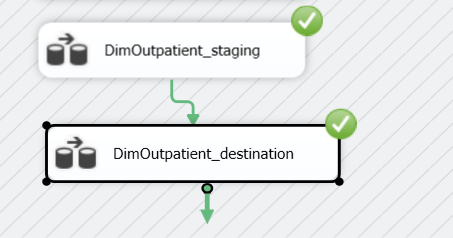


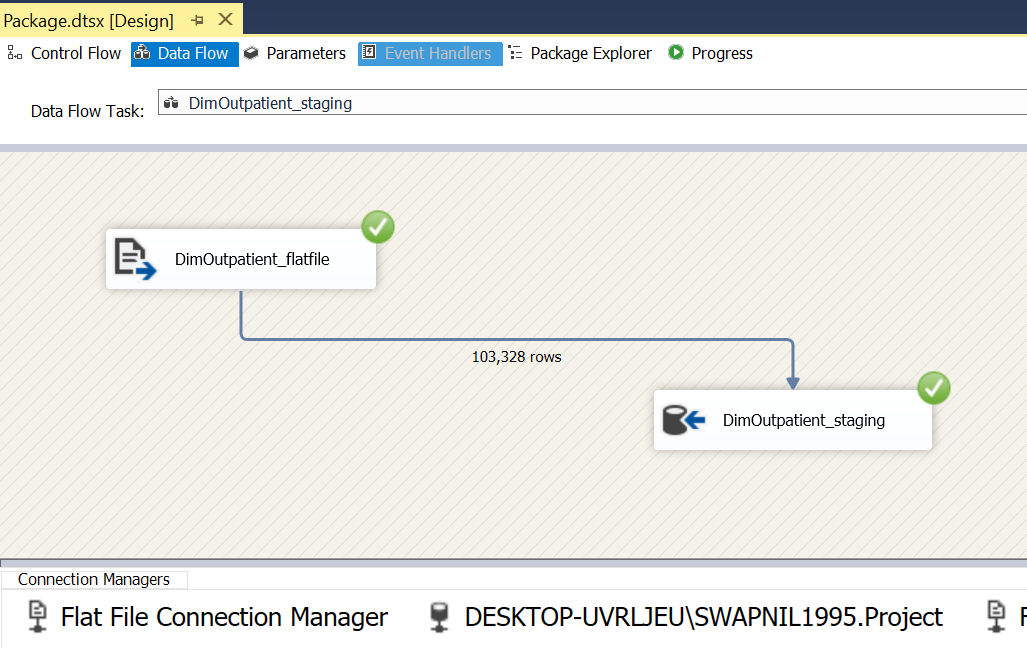


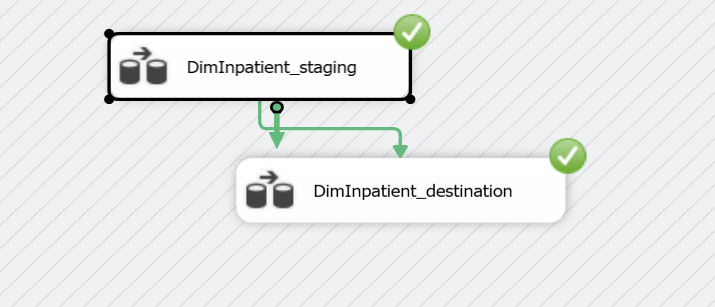
2.FlatFile Source to Staging Process Screenshots

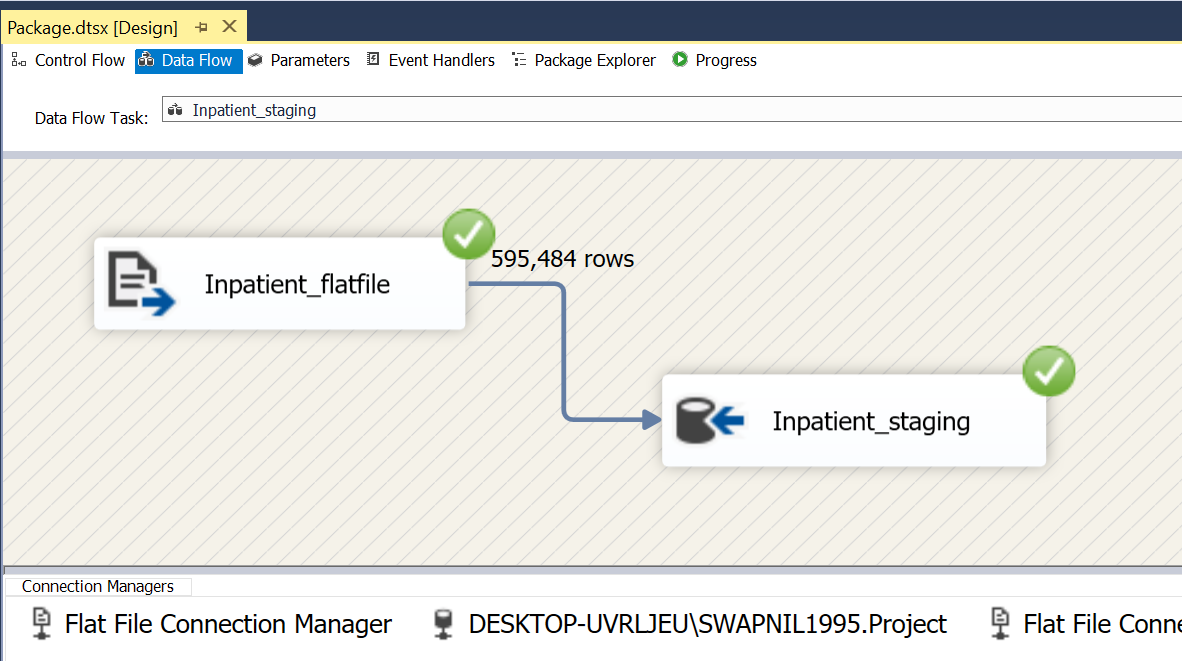


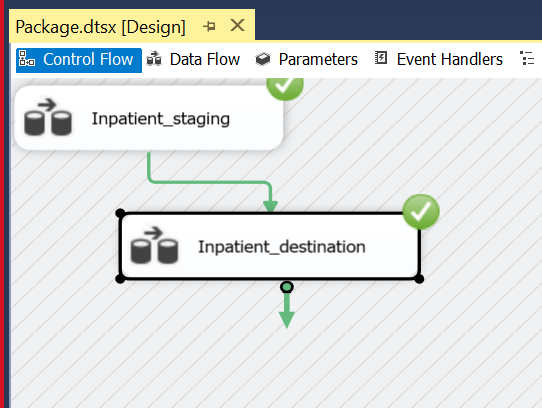


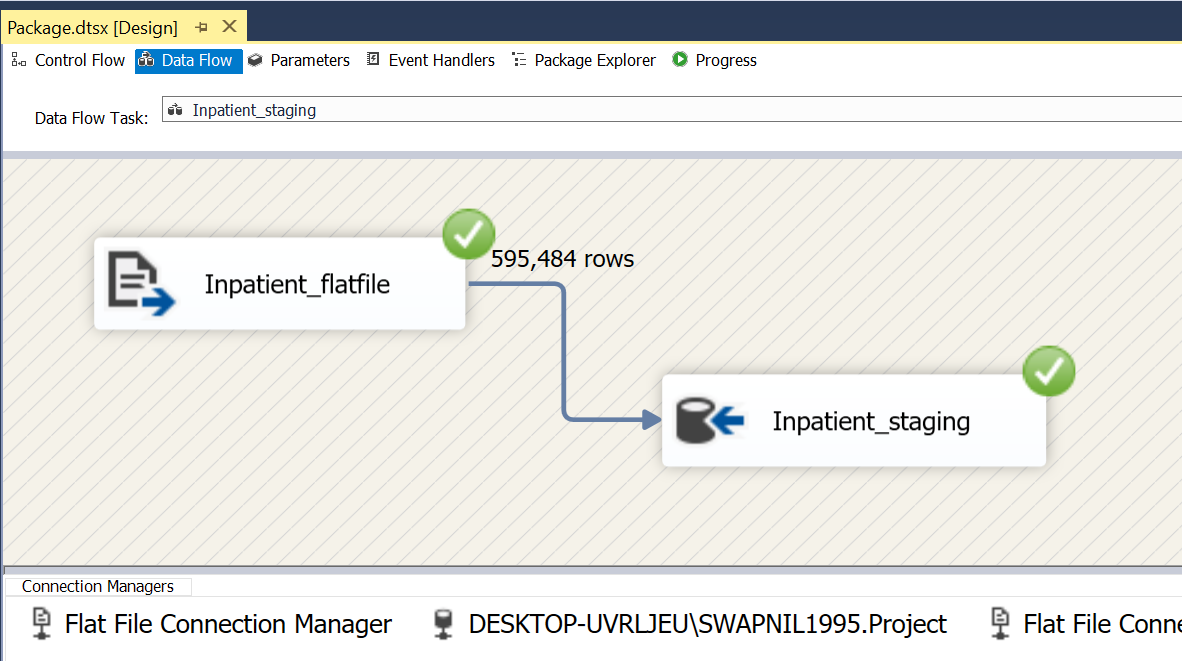


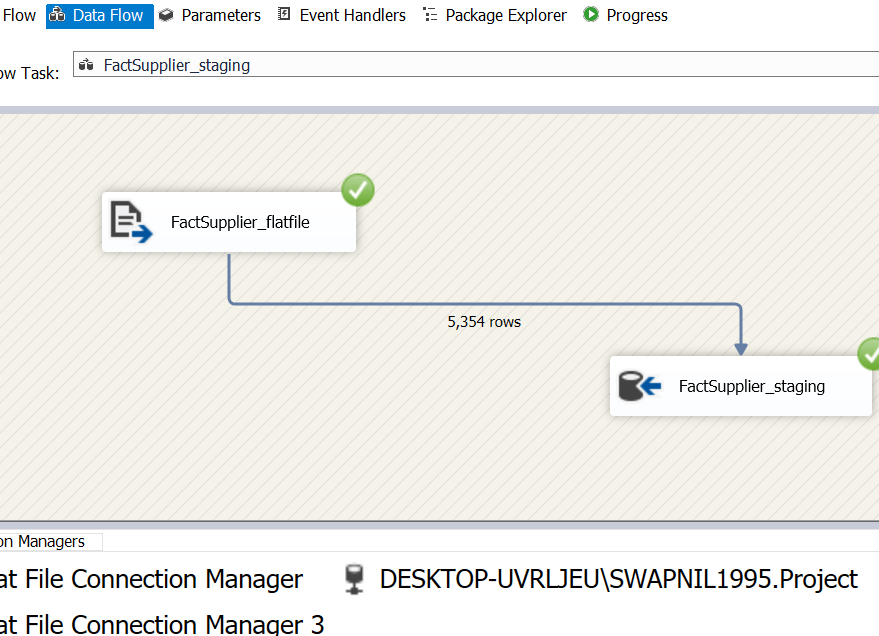


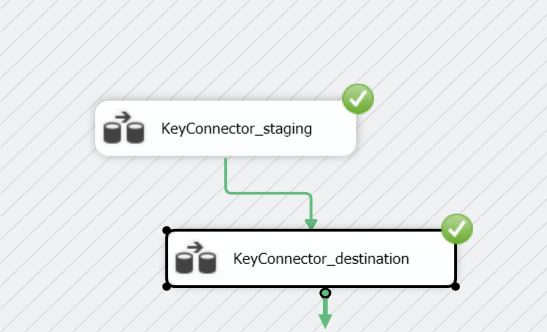


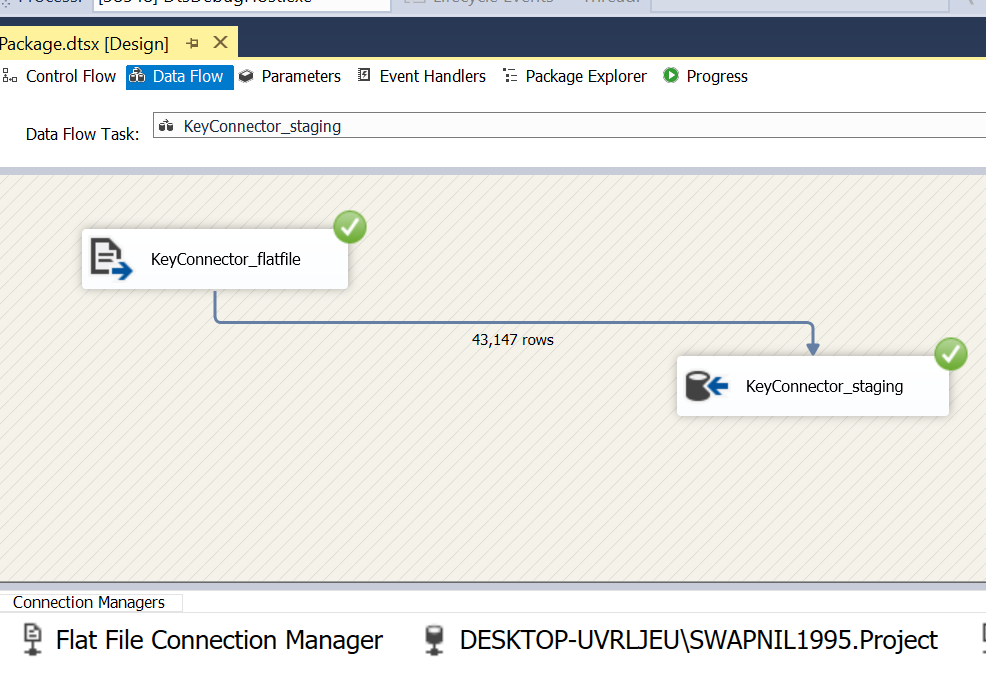




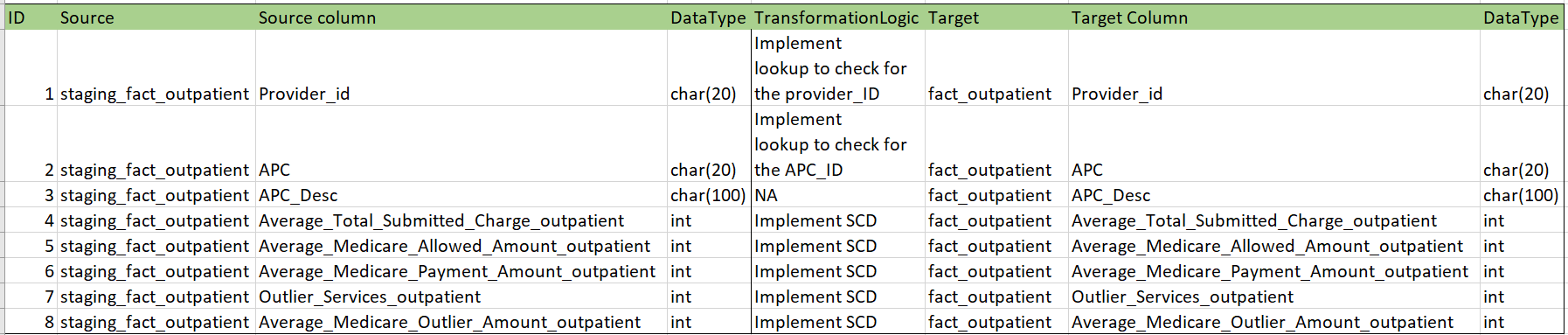


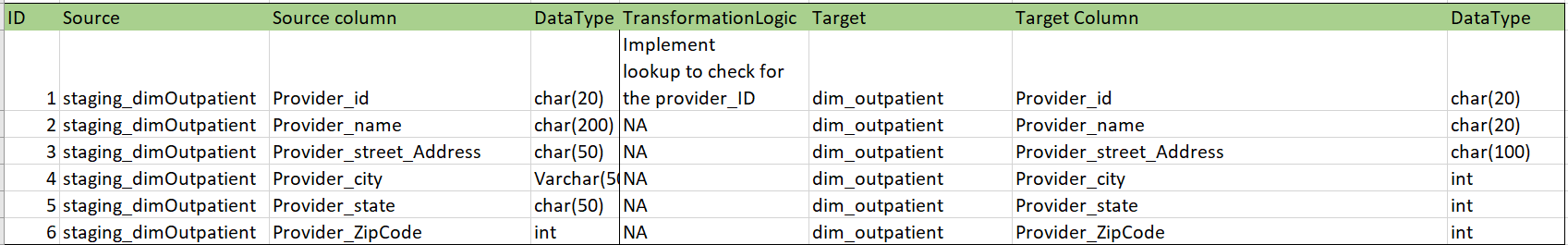


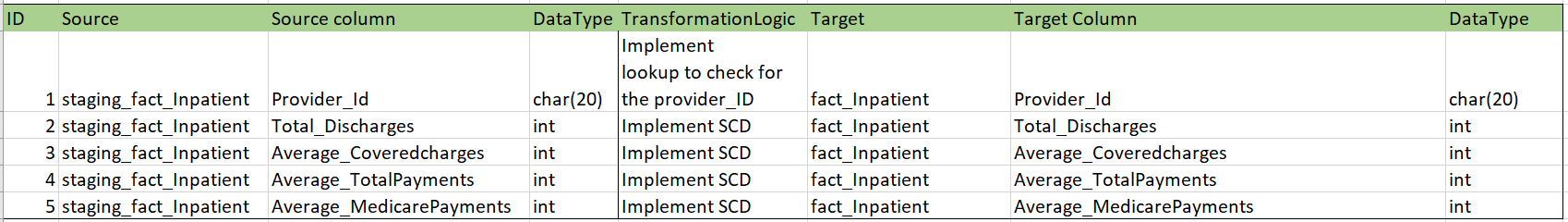


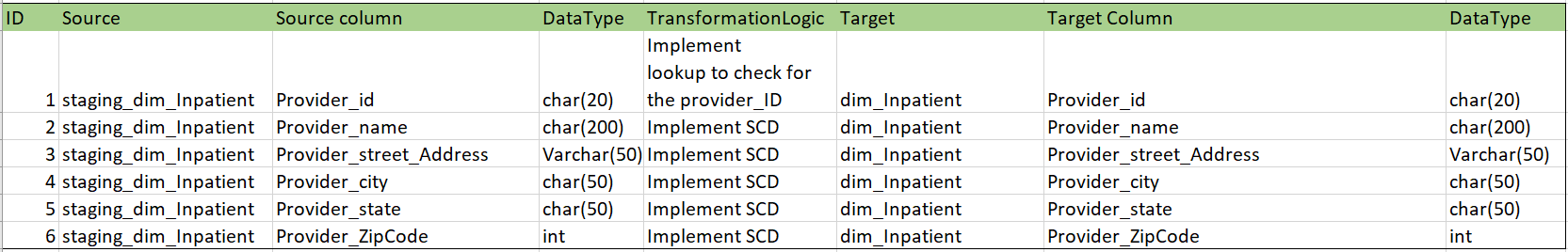


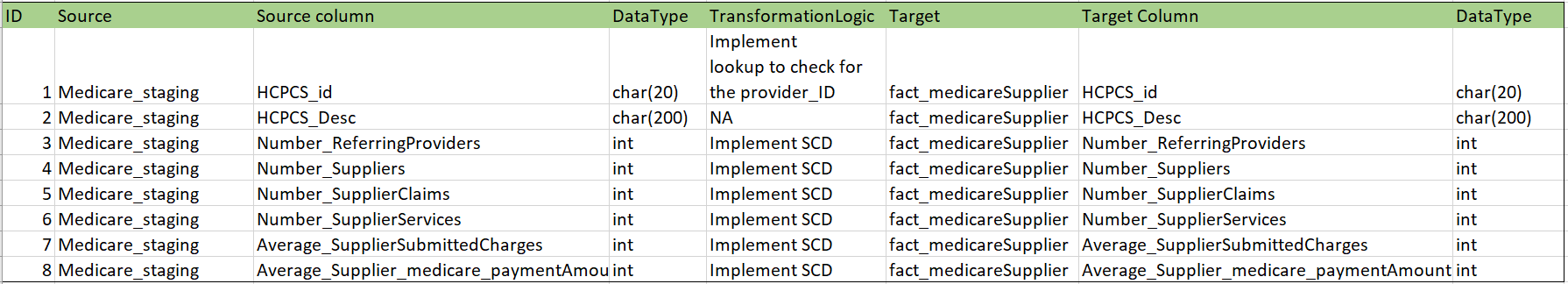
3. Staging Source to Final Destination Table



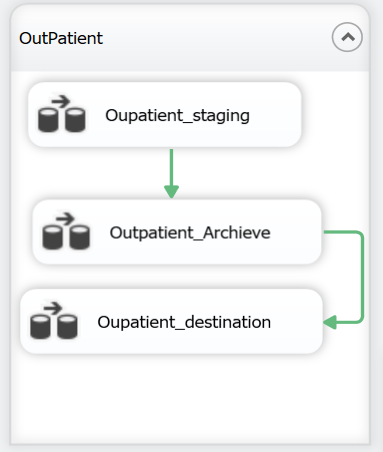


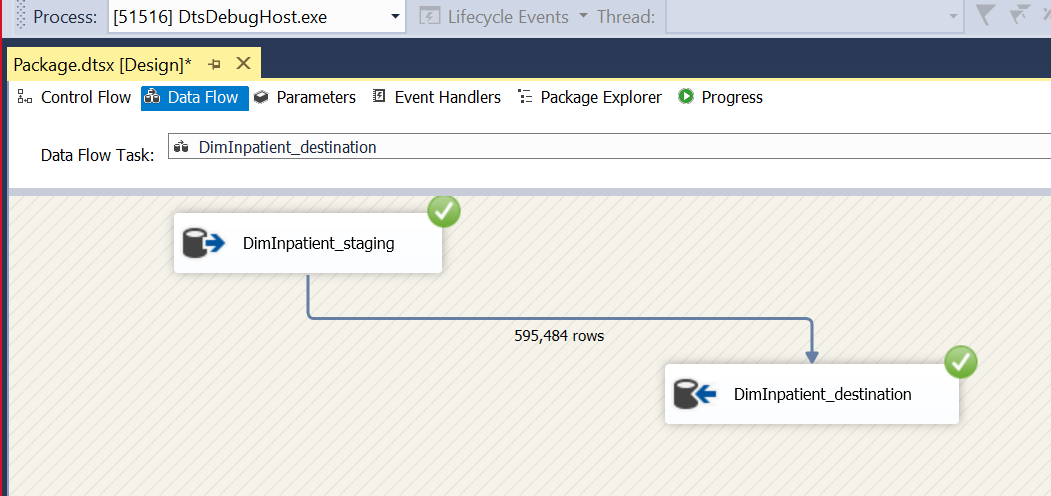


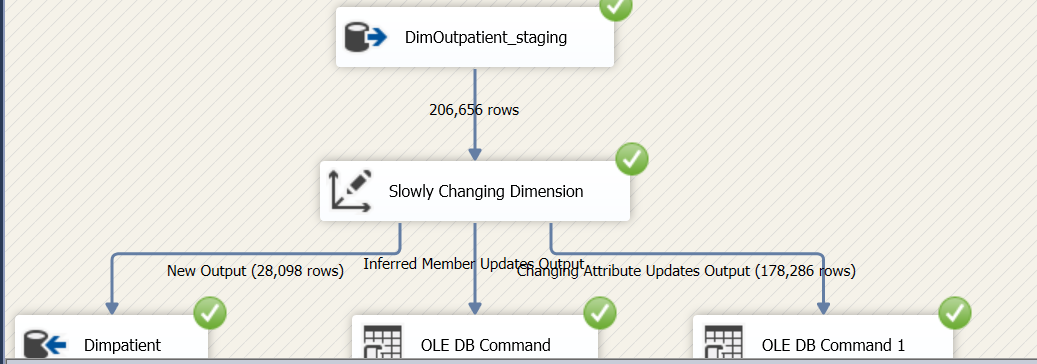


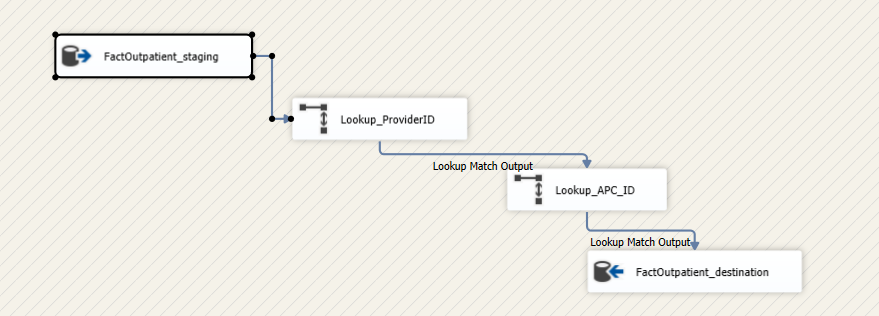


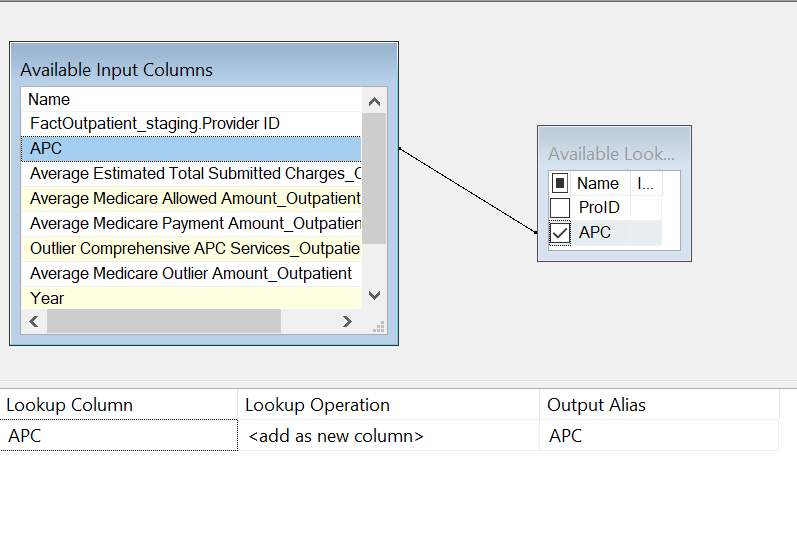
4. Staging Source to Archieve/Destination Process Screenshots

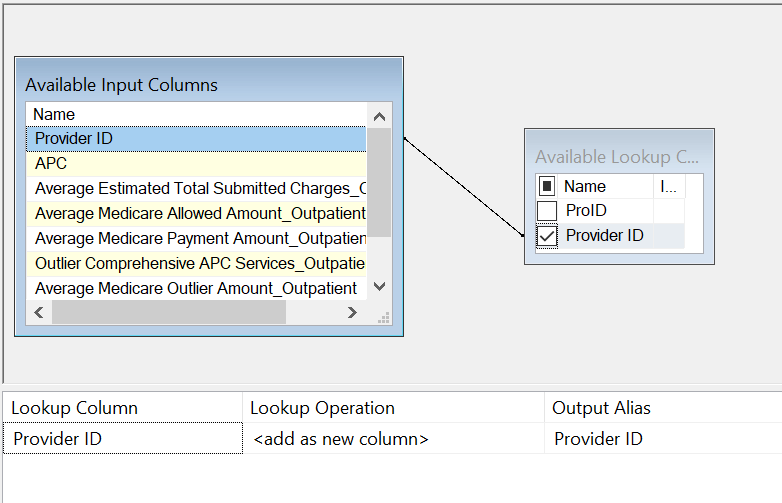








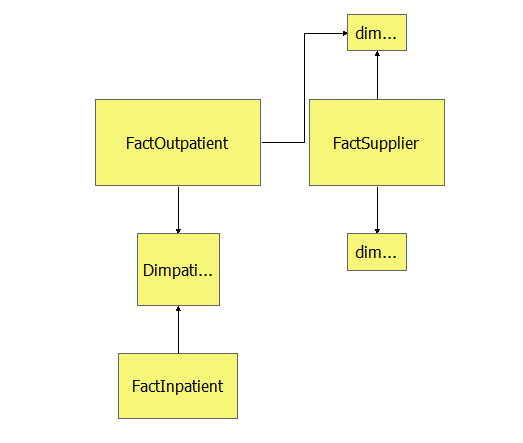


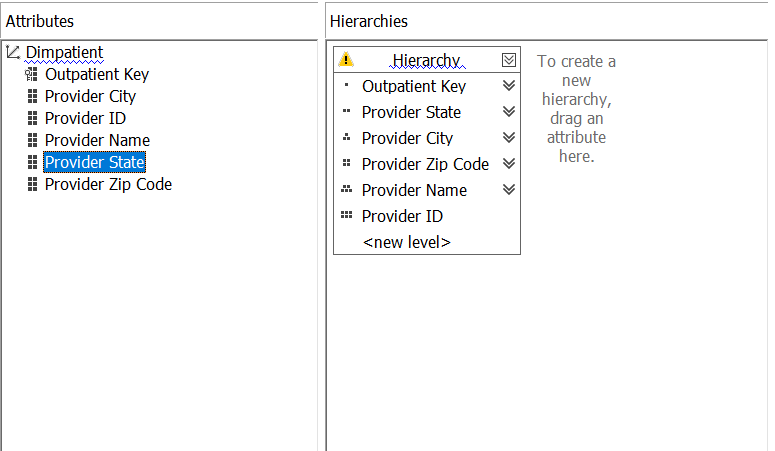


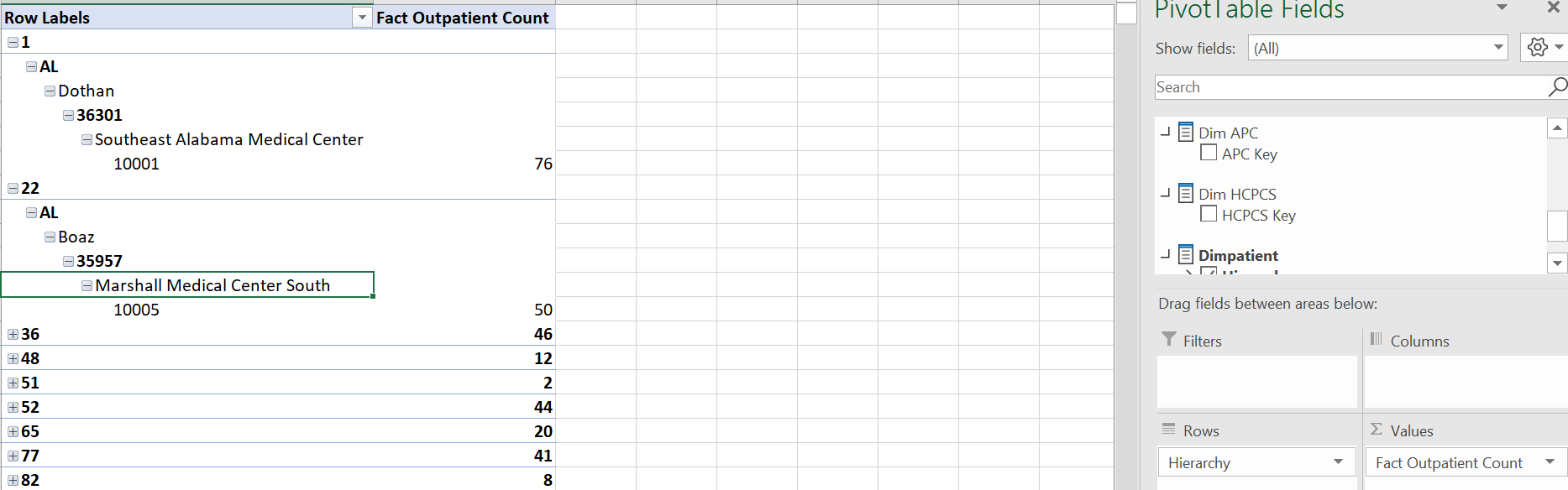
**Online Analytical Processing (OLAP Cubes):**

The Multidimensional OLAP Cube was creates using SQL Server Analysis Services (SSAS) in MS Visual Studio (SSDT).

Below are the screenshots of the dimensional hierarchies which are created for organization specific reports and also OLAP reports were further analyzed in MS Excel.

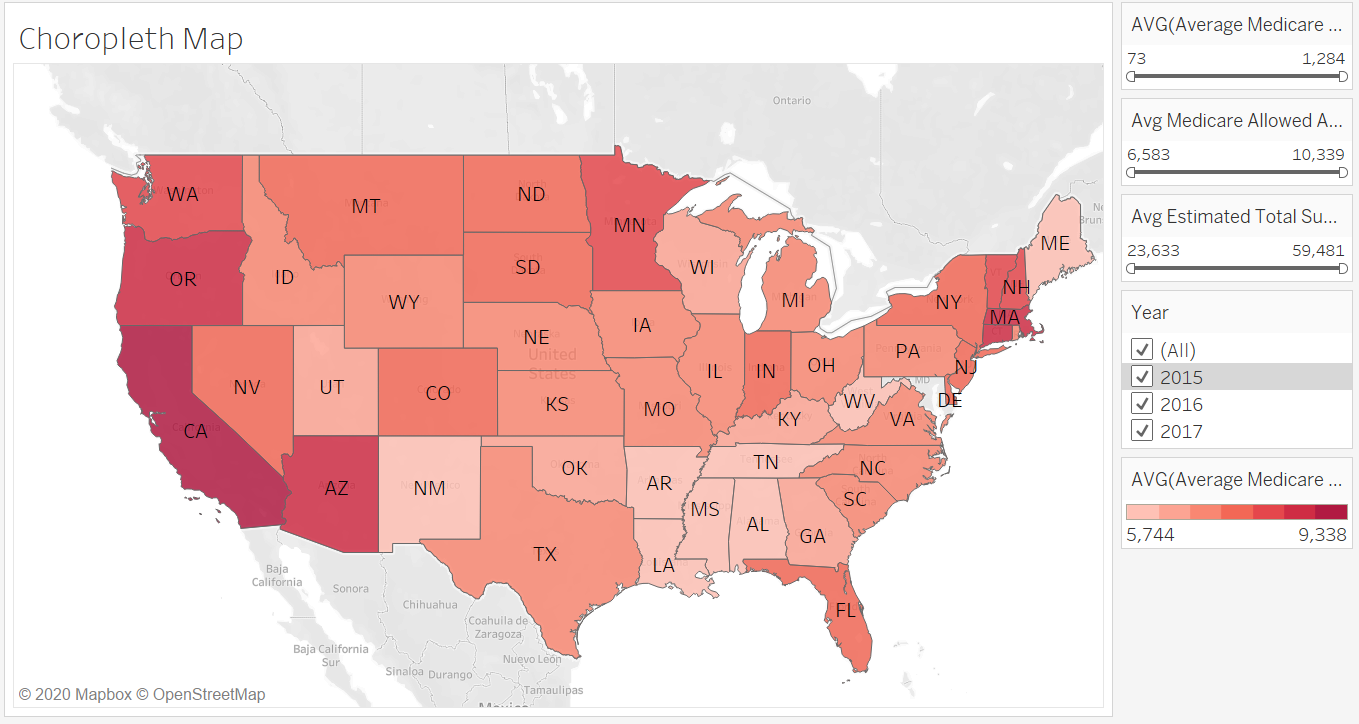




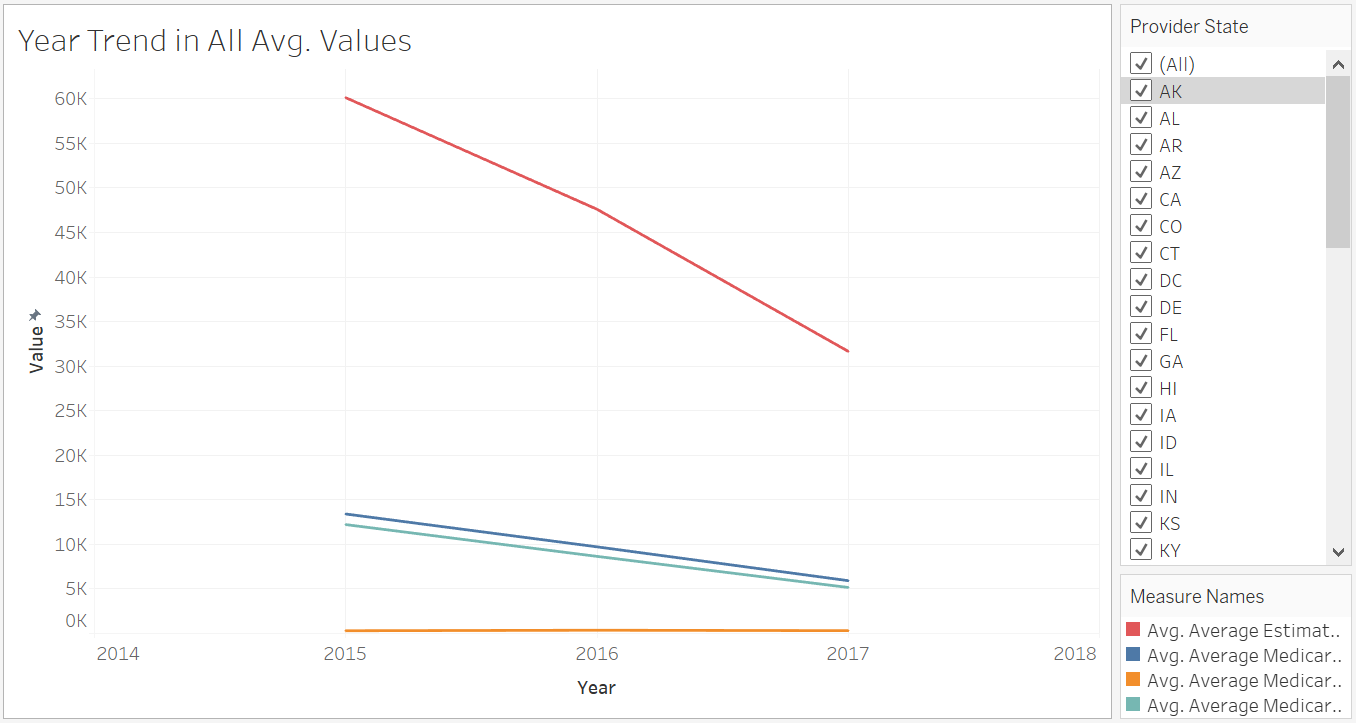


**Business Intelligence:**

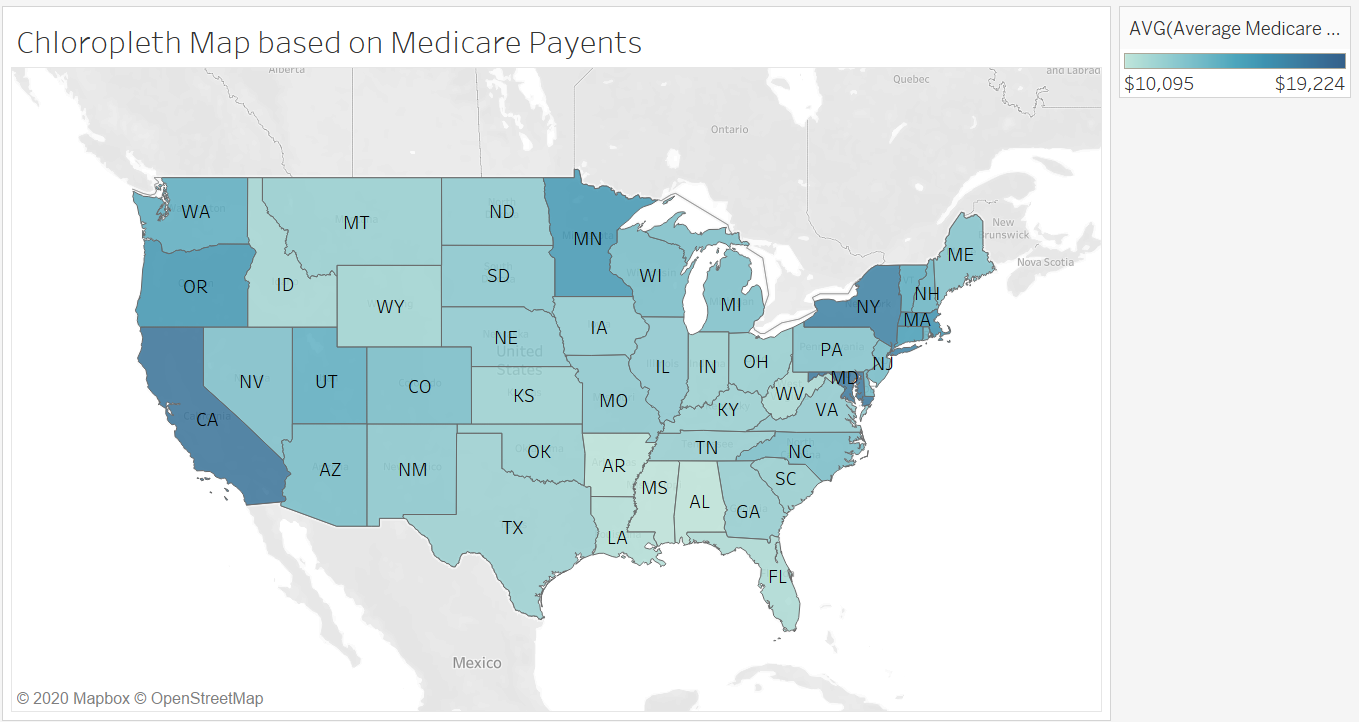
1. The following choropleth graph represents average values of **Medicare payment amt, Medicare allowed amt, Total submitted charges, Outlier amount. For Outpatient**

****

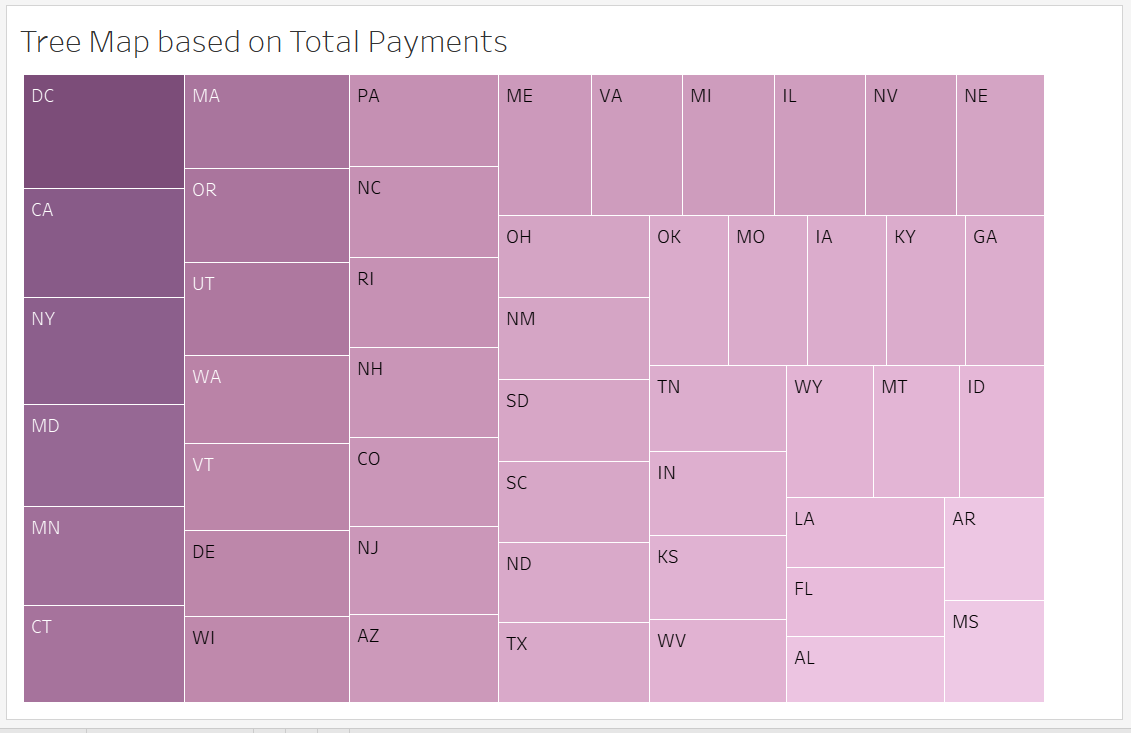
2. The following Line graph represents average values of **Medicare payment amt, Medicare allowed amt, Total submitted charges, Outlier amount. For Outpatient** trends over year **2015, 2016, and 2017 for Outpatient.**

****

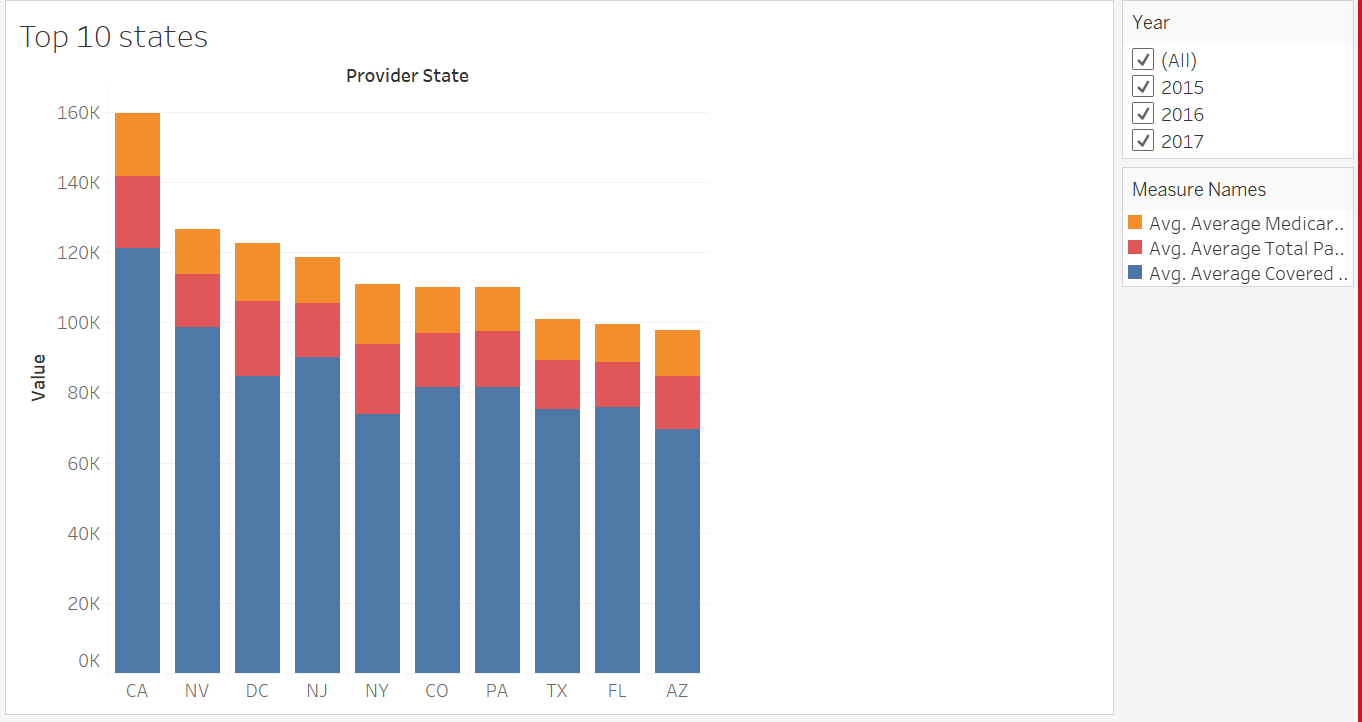
3. The following choropleth graph represents average values of **Medicare payment amt, Medicare payment, Total Payments, Total Discharges for inpatient.**

****

4. The following Tree-map represents average values of **Total Payment inpatient and sorted color gradient according to states**

****

5. The following Stacked Bar Chart represents average values of **Medicare payments, Total Payments, covered charges inpatient top 10 states in The US**

****

6. The following bar graph dashboard represents average values of **Medicare payment amt, Medicare allowed amt, Total submitted charges, Outlier amount. For Outpatient**

****

**Tools Use:**  
Create a dimensional data model for a Healthcare Data warehouse

* Load data from the data sources provided into your Health DWs

ETL tool:

* Talend Enterprise Data Integration
* SSIS
* ER Studio

DW database:

* Microsoft SQL Server - loaded using SSIS

Create visualizations of analyzed data using following BI tools:

* Tableau
* Power BI
* Talend
* SSRS

**Insights**

* Insights of state-wise average medical services expenditure such as ambulatory services.
* Analysis of usage of medical services over the years 2015-2017.
* Average cost, suppliers, beneficiaries based on specific medical services furnished by providers.

Thank You.