****

**Medicare Provider Utilization and Payment Data**

**Data Warehousing & Business Intelligence**

**GROUP 4**

**Authors:**

Rohit Appandaraju

Rohit Chandrasekhar

Mohamed Nadeem

Karthik Ranganathan

Table of Contents

[1.0 Summary 3](#_Toc5373081)

[1.1 Medicare 3](#_Toc5373082)

[1.2 About the data 3](#_Toc5373083)

[2.0 Techniques 4](#_Toc5373084)

[2.1 Cleaning the Data 4](#_Toc5373085)

[2.2 Additional Data 4](#_Toc5373086)

[2.3 Normalizing the Data for Storage 4](#_Toc5373087)

[2.4 Data Flow for Dim and Fact tables 4](#_Toc5373088)

[2.5 Visualization 5](#_Toc5373089)

[3.0 Findings 5](#_Toc5373090)

[4.0 Version 5](#_Toc5373091)

[5.0 References 5](#_Toc5373092)

[6.0 Workings 6](#_Toc5373093)

### 1.0 Summary

### 1.1 Medicare

Medicare is the United States federal government [health insurance](http://www.investinganswers.com/node/3562) program for Americans who are 65 years of age and older and some younger people with disabilities. When a U.S. citizen turns 65 years old, he may be eligible for Medicare benefits if he meets certain criteria. These benefits are intended to cover the costs of healthcare associated with advanced age. Eligible participants can choose from a number of plan options depending on their needs. All Medicare options require that [beneficiaries](http://www.investinganswers.com/node/5216) pay an affordable premium as well as co-payments for all covered healthcare services. Medicare comprises three separate sections that each deal exclusively with different parts of the healthcare system.1 Part A covers inpatient care and on-going treatments including hospital stays, in-home care, and nursing homes. Part B covers outpatient care and medical equipment needs like office visits, lab tests, x-rays, medications, and wheelchairs. Part C provides eligible participants with a variety of additional services to choose from depending on their needs. Part D provides prescription drug coverage.2

### 1.2 About the data

The dataset has various data points associated with inpatient cost, outpatient costs with respect to various procedures performed. Total expenditure related to inpatient and outpatient are explained these data points and gives insights related to government expenditure for the year 2012 to 2016. This gives critical insights to validate government actions to cut expenses for 2020 budget.

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. The PUF is organized by hospital and Medicare Severity Diagnosis Related Group (MS-DRG) and covers Fiscal Year (FY) 2013 through FY 2016 for this project.

Hospitals determine what they will charge for items and services provided to patients and these charges are the amount the hospital bills for an item or service. The Total Payment amount includes the MS-DRG amount, bill total per diem, beneficiary primary payer claims payment amount, beneficiary Part A coinsurance amount, beneficiary deductible amount, beneficiary blood deductible amount and DRG outlier amount.

Inpatient PUF users will be able to make comparisons between individual hospital-level charges and payments within local markets, and nationwide, for services that might be furnished in connection with an inpatient stay.

The Outpatient Utilization and Payment Public Use File (Outpatient PUF) presents information on common outpatient services provided to Medicare fee-for-service beneficiaries. The Outpatient PUF presents information on utilization, payment, and estimated hospital-specific charges for select Ambulatory Payment Classification (APC) Groups paid under the Medicare Outpatient Prospective Payment System (OPPS) for Calendar Years (CY) 2013 through 2016. The Medicare payment amount includes the APC payment amount, the beneficiary Part B coinsurance amount and the beneficiary deductible amount.

### 2.0 Techniques

### 2.1 Cleaning the Data

There are four excel files for inpatient and outpatient category across 4 years (2013-16). The files would be loaded through using SSIS Package. There have been several changes carried yearly with the data storage some new columns have been added. Some columns have been separated in two columns. Standardization of the excels sheets have been done to maintain uniform data columns and types across all the files. The initial excel sheets when converted in csv had a limitation of 60,000 rows. This created a bottleneck in the process of staging. This was overcome by finding alternate source of data with CSV files.

For Each Loop was used to upload both the inpatient and outpatient data into the staging table. The year of the data has been added with derived column.

### 2.2 Error handling, Auditing, Validating the Data and, SCD

Since the data has been uploaded in the CSV format. Any cells containing comma would create additional columns for the specific row thus throwing an error. The error output has been loaded in a separate table to analyze the problem and rectify it. The audit information gives the details of system, user and time of the load.

The validation of data was done by finding the unique hospitals and matching all the records of the outpatient and Inpatient database with it. Any records not matching where sent to the error table for further analysis. After determining the cause of error, the data were corrected for the issue and uploaded using SCD. In this specific case, there was one column which had a comma in it. Which caused 41 rows to be sent to the error table. This was corrected and uploaded through SCD pipeline.

### 2.3 Additional Data

Further there will be several derived columns added onto the original variables for better analysis of the data. Additional datasets would be required for analysis. Those datasets will be containing state wise data points such as per capita income4, population5, number of hospitals6, number of medical practitioners7, education levels8. The relation between expenditure in Medicare and the above data points will be determined by Pearson correlation. Thus, advising the different states on to reduce Medicare expense on long term future.

### 2.4 Normalizing the Data for Storage

We intend to normalize the dataset by utilizing the first 3 normal forms. We would be eliminating data redundancy, improving efficiency of queries, and fewer null values by creating multiple tables with primary keys and foreign keys.

### 2.5 Data Flow for Dim and Fact tables

After the analysis of the data in initial staging table, there would be certain measurements and requirements stated which would lead us to the final analysis reporting through visualization. The Dimension tables were created for various factors in the data. These included State, City, Hospital cubes. Further dimensions were created with additional data which were downloaded for analysis purposes.

### 2.6 Visualization

There will be two software’s used for visualization. Tableau and Power BI would be used primarily. Further we will be creating a Web Application in R Shiny for visualization which could be hosted in websites. The initial set of visualization reveals that three states in the US are heavy spenders in Medicare. California, Florida, New York and Texas contribute to 25% of the overall cost. The population in these states are also high.

### 3.0 Findings

We would like to find the different rates of inpatient and outpatient across states. We would like to see how Per Capita Income and Education levels are impacting the inpatient and outpatient visits. We would like to see how the average Medicare amount is varying according to the state. Also, what are the external factors involved in some states having more expensive treatment costs.

One major finding could be how helpful in terms of financial help does Medicare provide to the subscribers. Is its maximum coverage enough to cover majority of the cost associated with healthcare? Can recent growing US debt and controversy around Medicare cuts be justified. We could determine how it can impact the affordable care for millions of Americans depending on them.3

Having high percentage of tax in New York and California might also be the reason for this excessive expenditure. One solution to reduce the cost would be to promote medical tourism in states having low tax rates.

Our initial analysis of the Medicare seems to be some what puzzling. As we found the average life expectancy of US has been falling in the last couple of years. When the Medicare program was rolled out in 1966, the deductible was around 60$ and there were 3 million people who were eligible for it. As the government realized at that point that people under 65 years of age were mostly insured. There are currently 15 million people currently eligible for Medicare with average cost of deductible around 1300$. This often discourages people for seeking physician care. Medicare part B having a monthly payment of 185$ is also a grave concern9. One alternate solution to the Medicare is for government to subsidize private insurance and let people choose their insurance. Thus, creating competition in the market, which will help drive down the prices and offer better coverage.

### 4.0 Version

|  |  |  |
| --- | --- | --- |
| Version 1.0 | 03/28/2019 | Changes |
| Version 1.1 | 04/04/2019 | Enhanced Description, more briefing on data, Techniques updated on initial workings. |
| Version 1.2 | 04/12/2019 | Granular details of loading data, addition of Error handling, validation and SCD para, initial visualization results, update to findings. |
| Version 1.3 | 04/19/2019 | Updates to Workings with addition of dimension, fact table and Cubes creation, Update to finding with preliminary research on cost to patients in Medicare. |

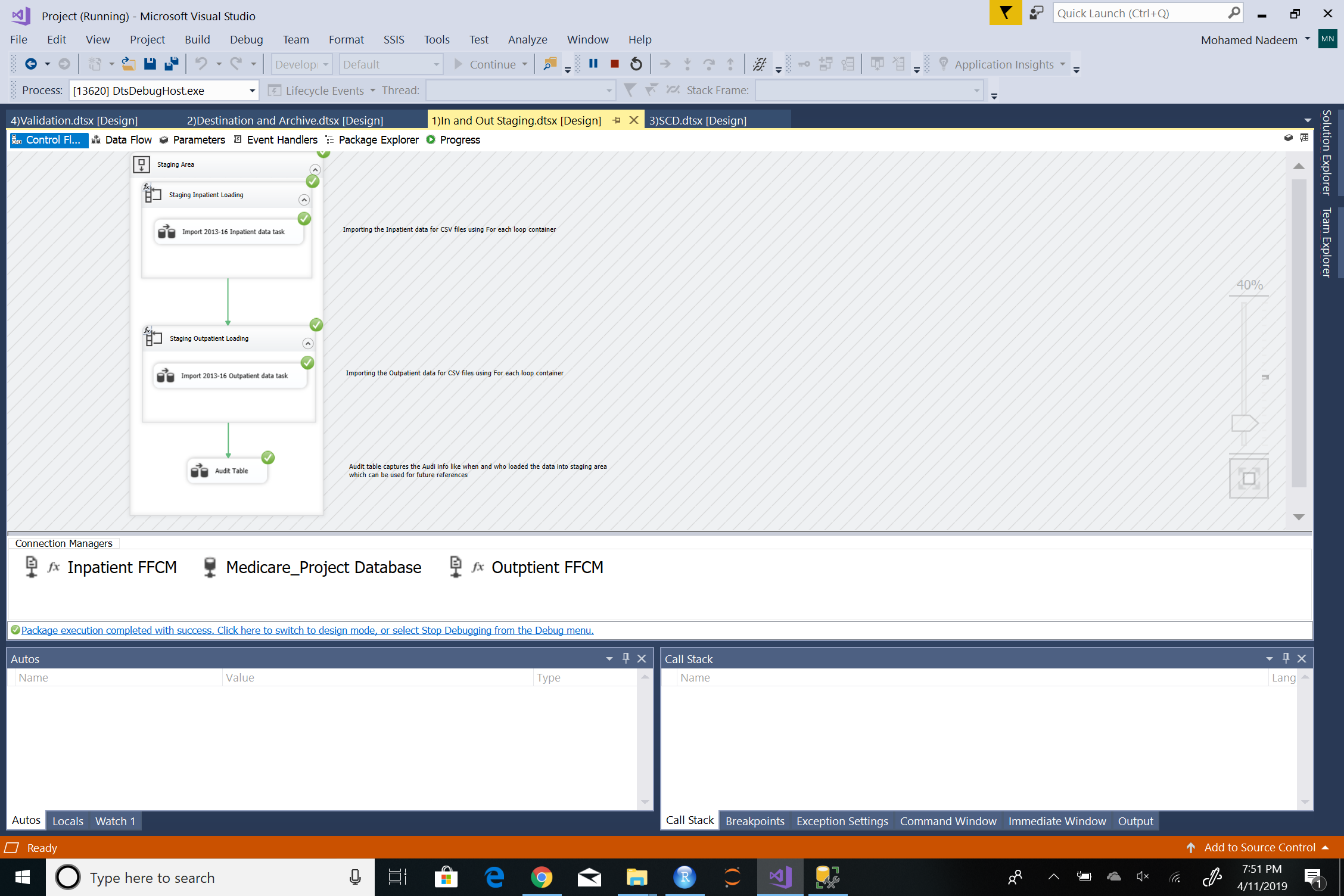
### 5.0 References

1. <https://www.merriam-webster.com/dictionary/Medicare>
2. <https://www.medicareinteractive.org/get-answers/medicare-basics/medicare-coverage-overview/original-medicare>
3. <https://www.vox.com/policy-and-politics/2019/3/12/18260271/trump-medicaid-social-security-medicare-budget-cuts>
4. <https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html>
5. <https://en.wikipedia.org/wiki/List_of_U.S._states_by_GDP_per_capita>
6. <https://data.medicare.gov/widgets/xubh-q36u>
7. <https://www.kff.org/other/state-indicator/total-active-physicians/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>
8. <https://en.wikipedia.org/wiki/List_of_U.S._states_by_educational_attainment>
9. <https://www.medicareadvantage.com/common-questions/does-medicare-cover-er-visits>

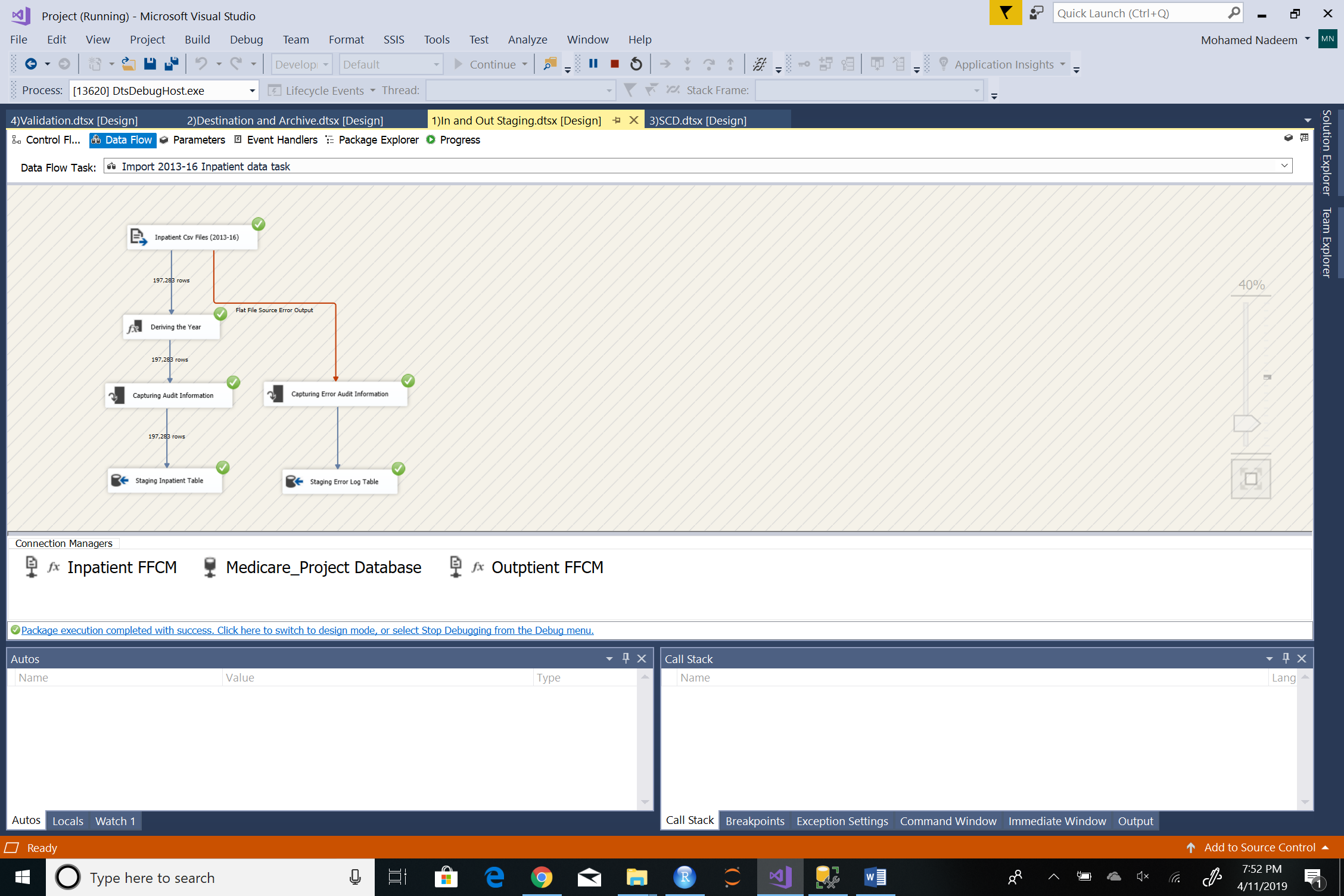
### 6.0 Workings

Staging:

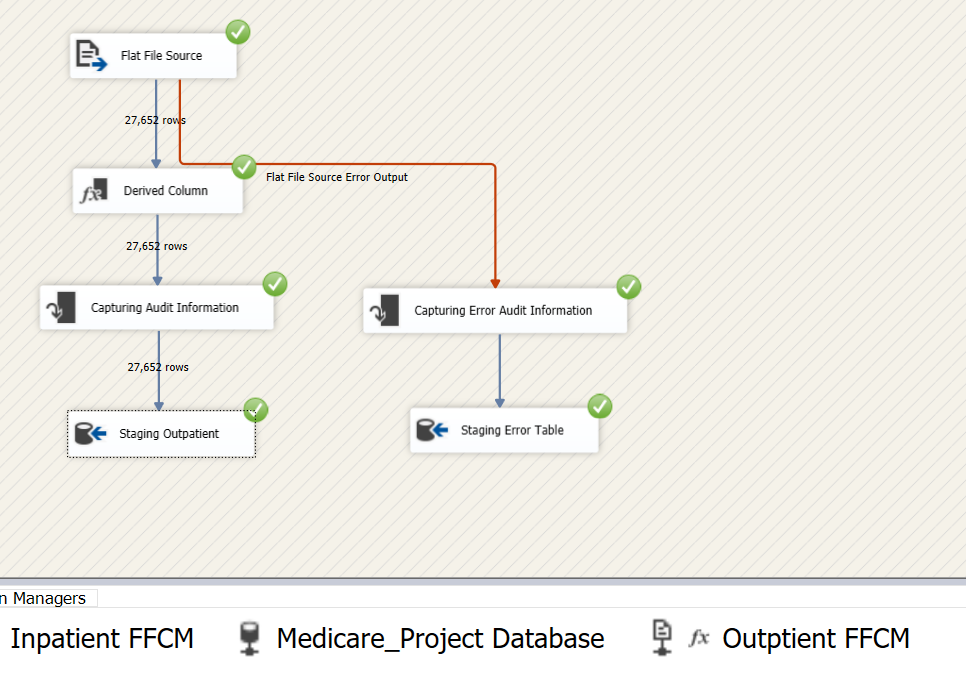
Step:1 The inpatient and outpatient files for medicare were loaded using For Each Loop



Step 2: Since the load was through CSV files, we realized there could be some errors. There for error handling was done. The file year was also added using derived columns

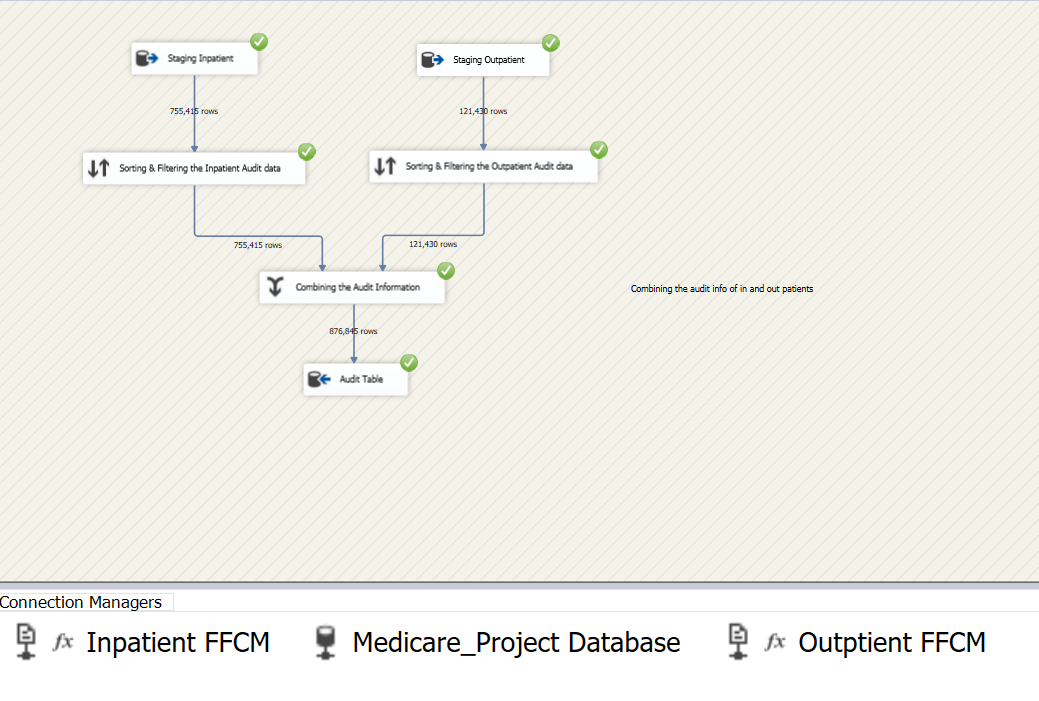


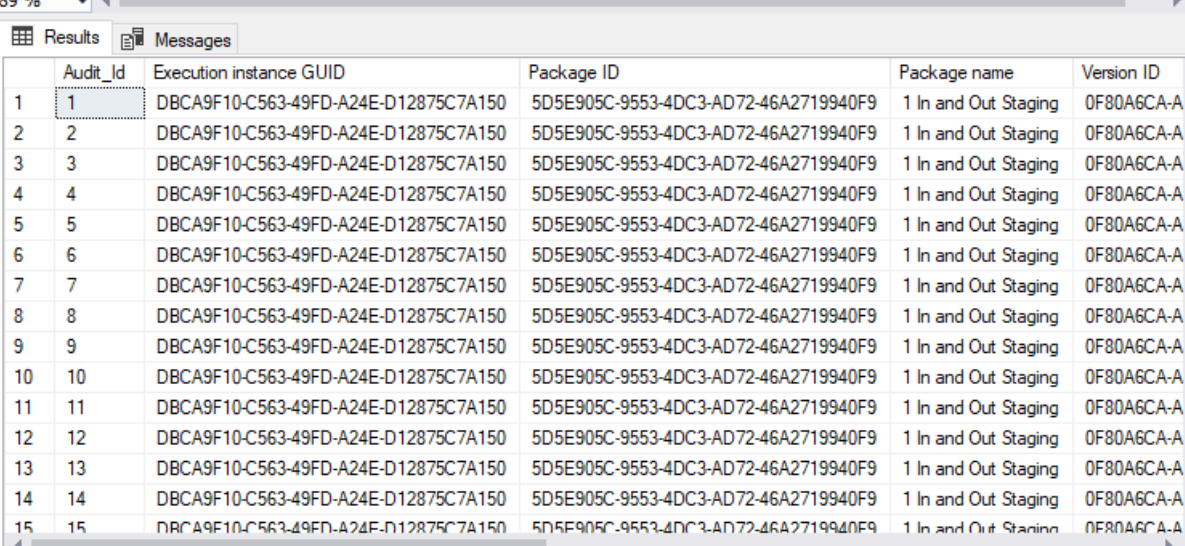
Step 3: Outpatient loading had similar data flow



Audit:

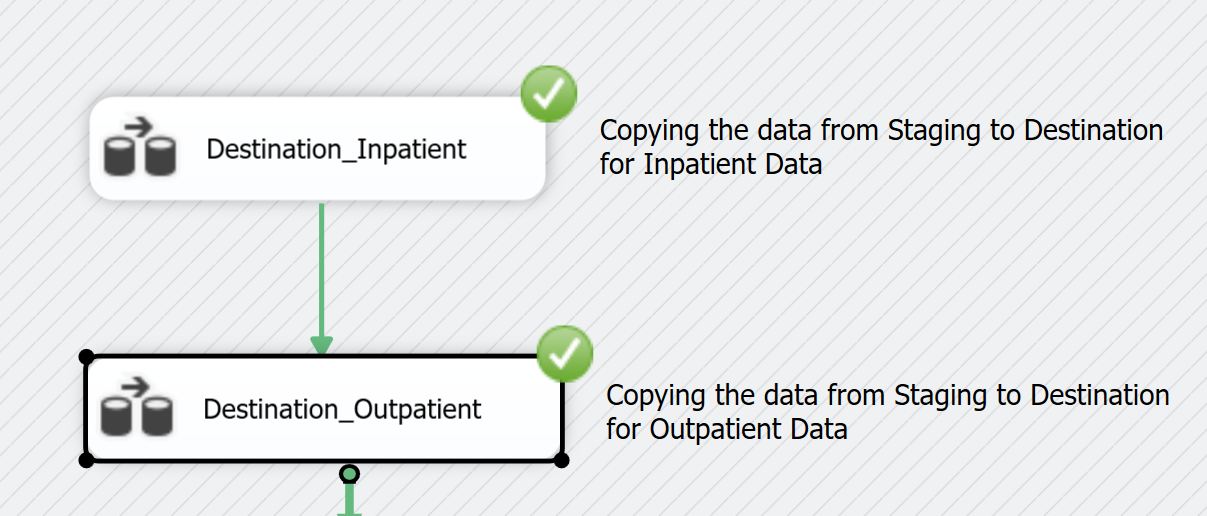
Step 4: We combined the audit of both tables for simplicity



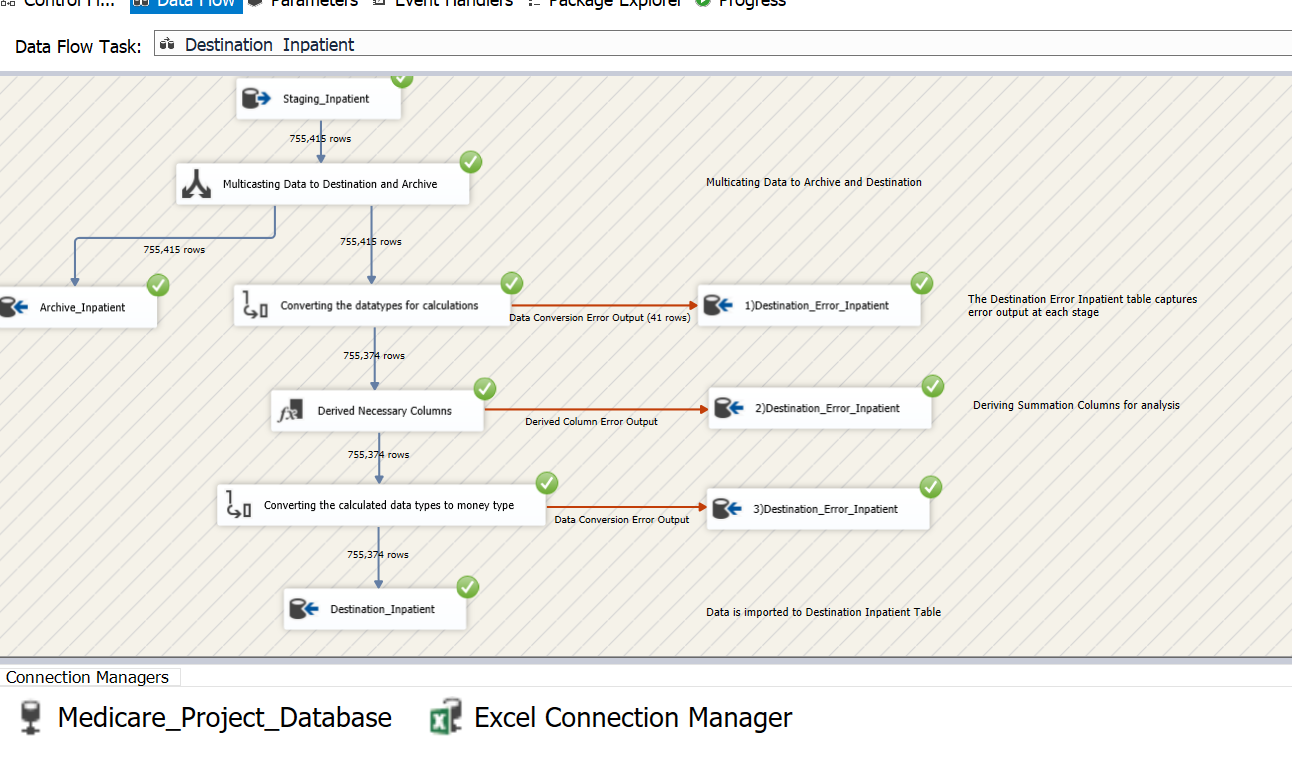


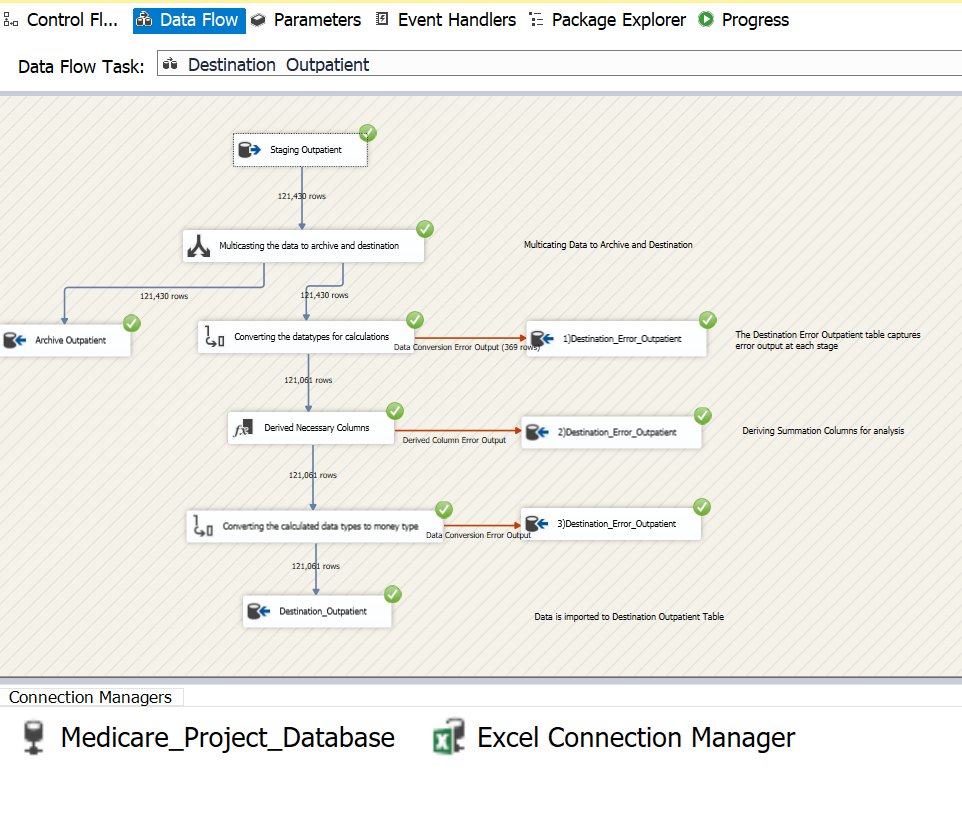
Destination and Archive:

Step 5: Destination table was used to store the final data from initial staging table after thorough review of the data.

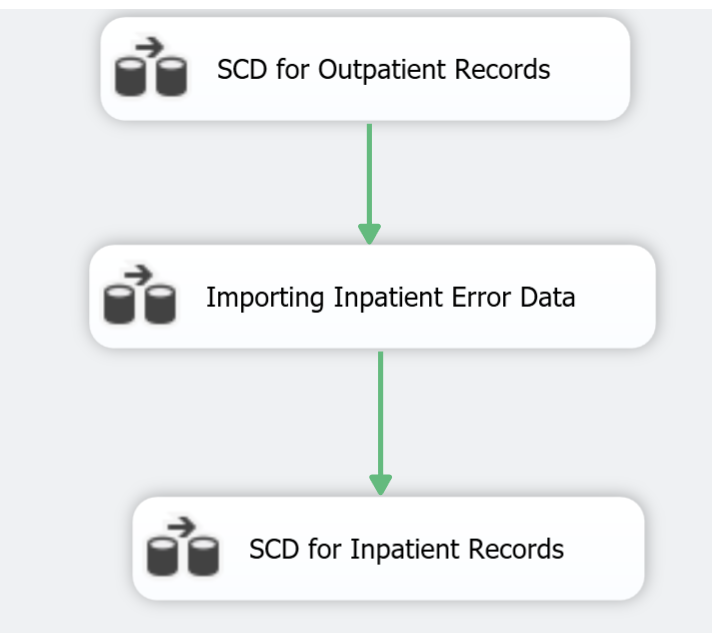


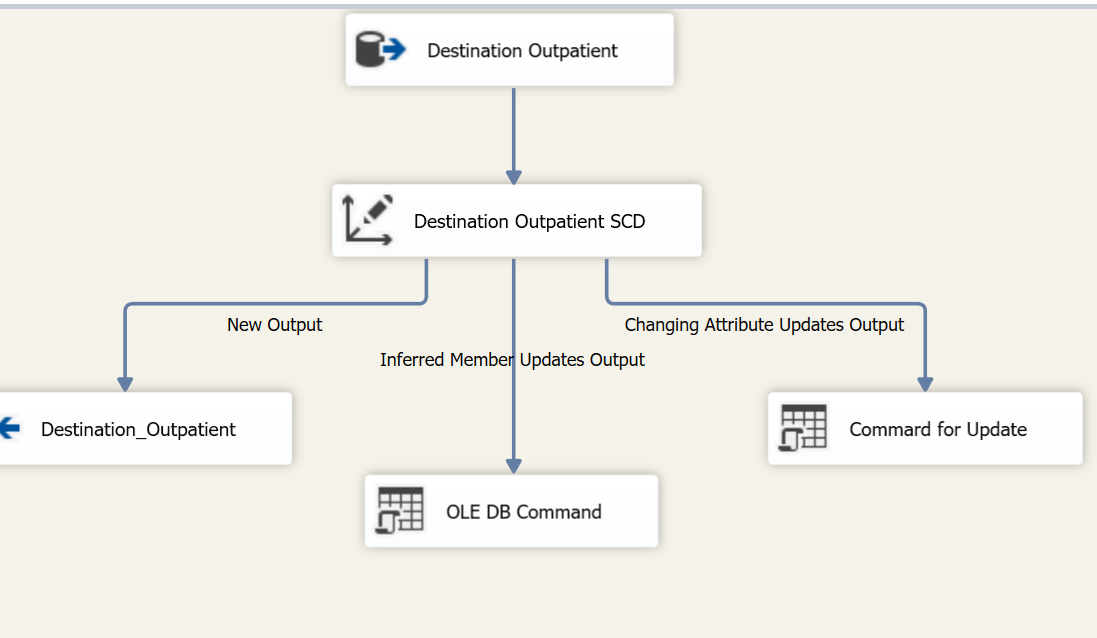
Step 6: In this data pipeline we carried out error handling and added necessary derived columns like the total cost for medicare and patients. As the initial data only contained averages



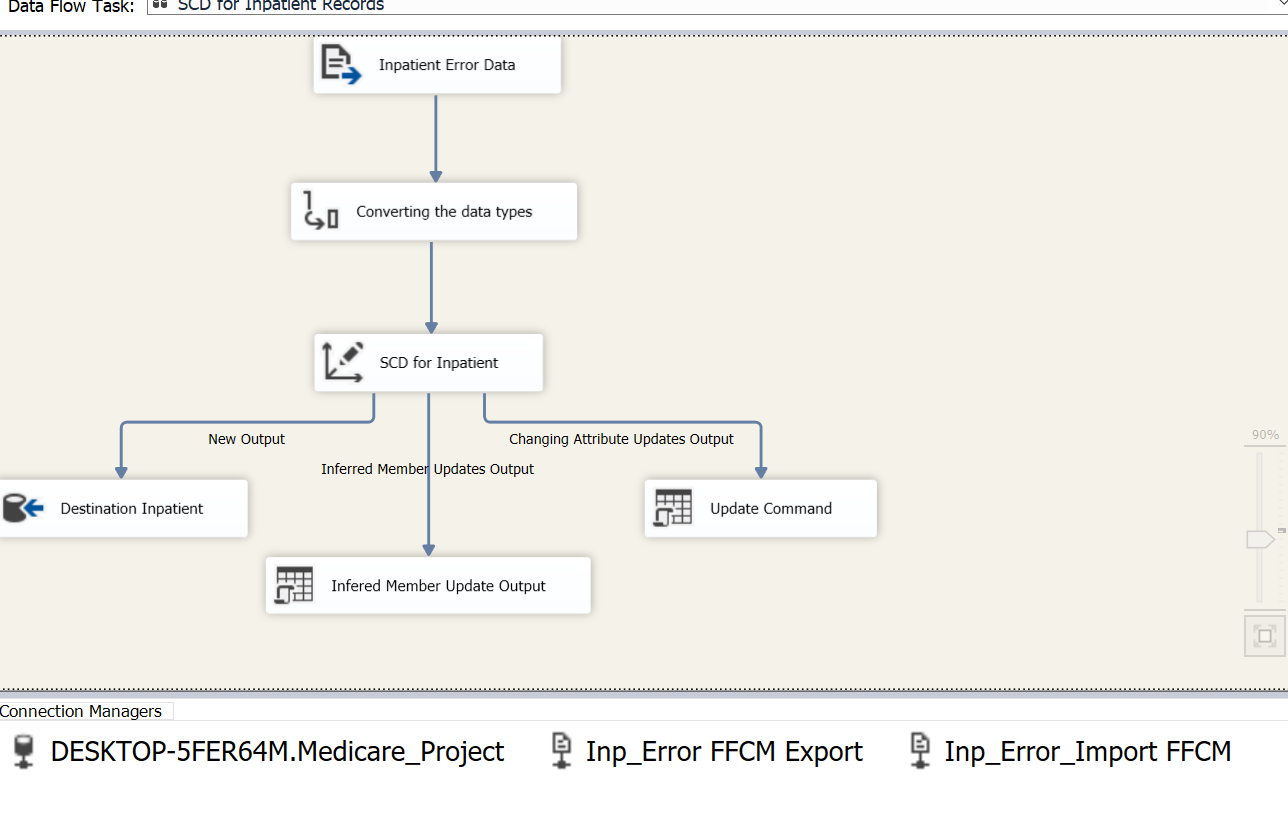


SCD:



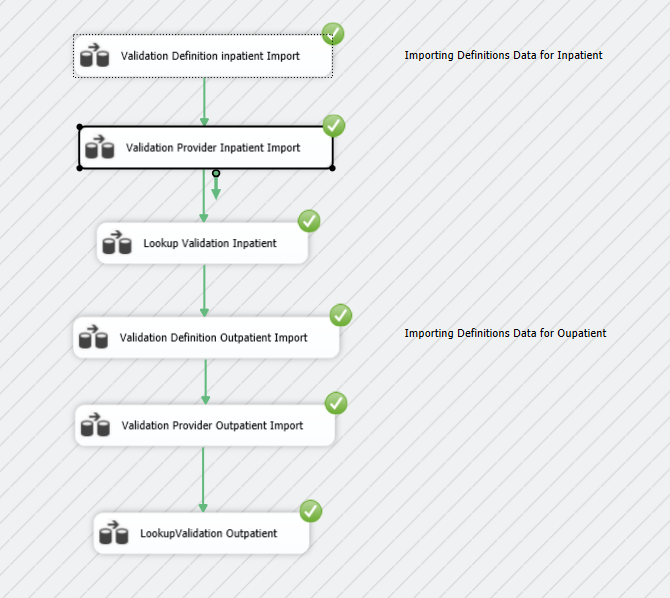


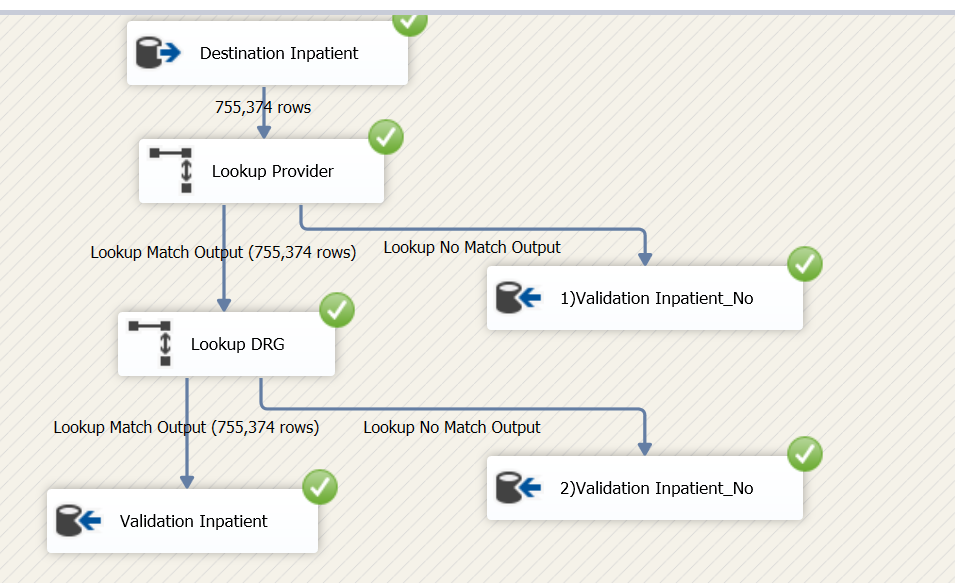
Step 7: A similar data flow was carried out for inpatient. Further we wanted to correct the 43 rows which had errors while loading into destination table. We realized this error was caused by a comma in one of the columns. We corrected the data in excel and loaded those rows through SCD.

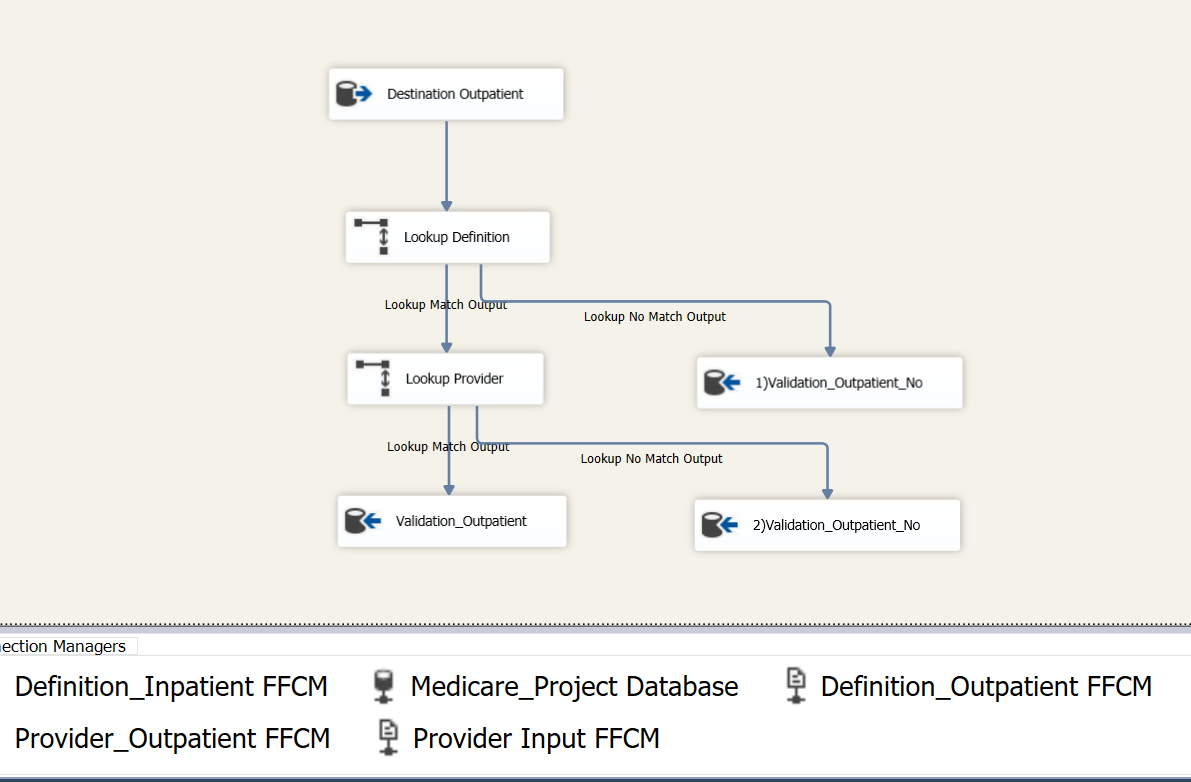


Validation:

Step 8: We wanted to make sure we had loaded all the data from initial datasets. Hence, we carried out data validation.

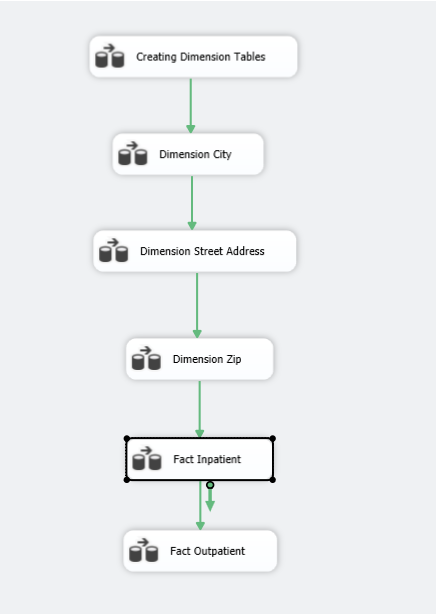




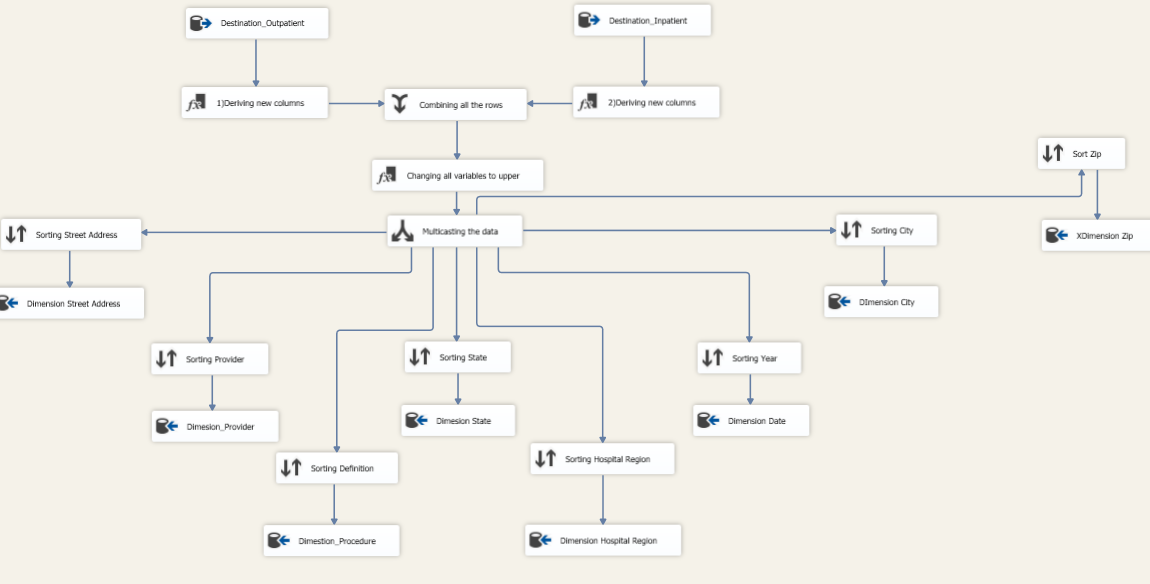


Dimension and Facts Data Flow Task:

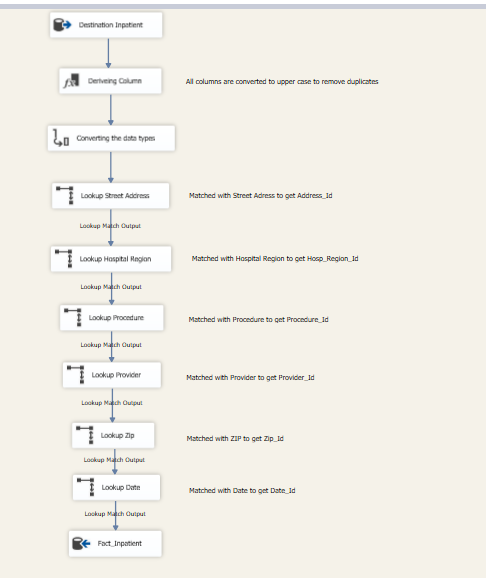
Step 9: Creating of facts and dimensions



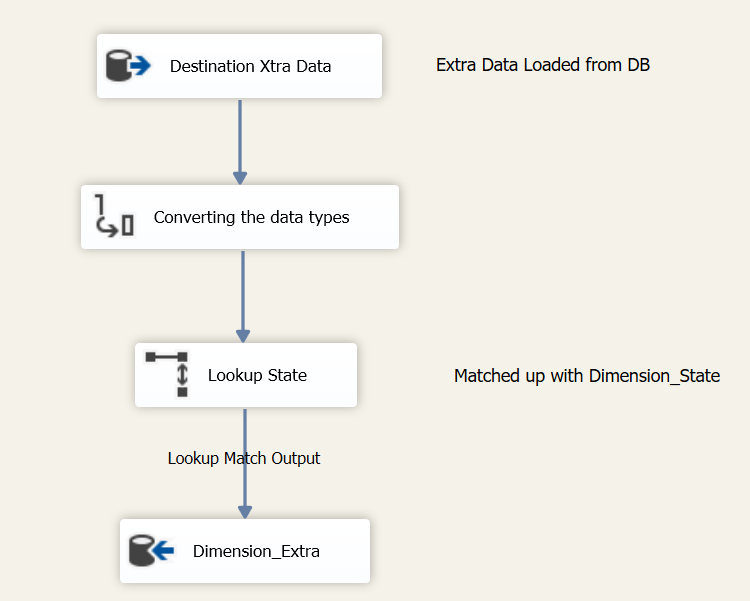
Dimension creation: We have created different dimensions here.



Fact Table Creation:

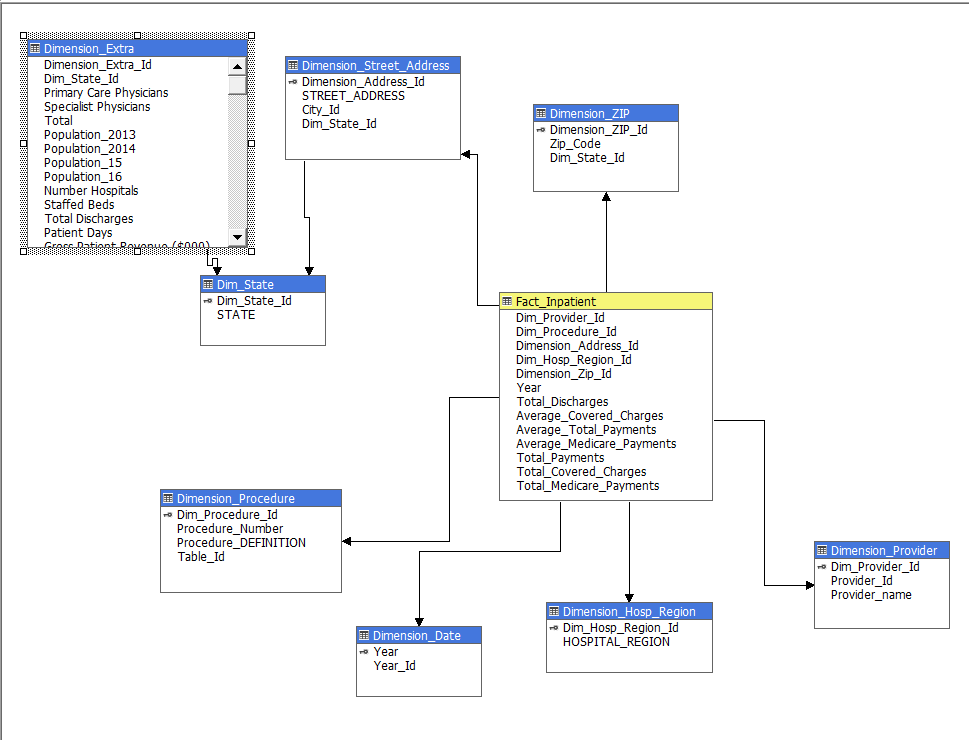


Supplementary Data Loading:



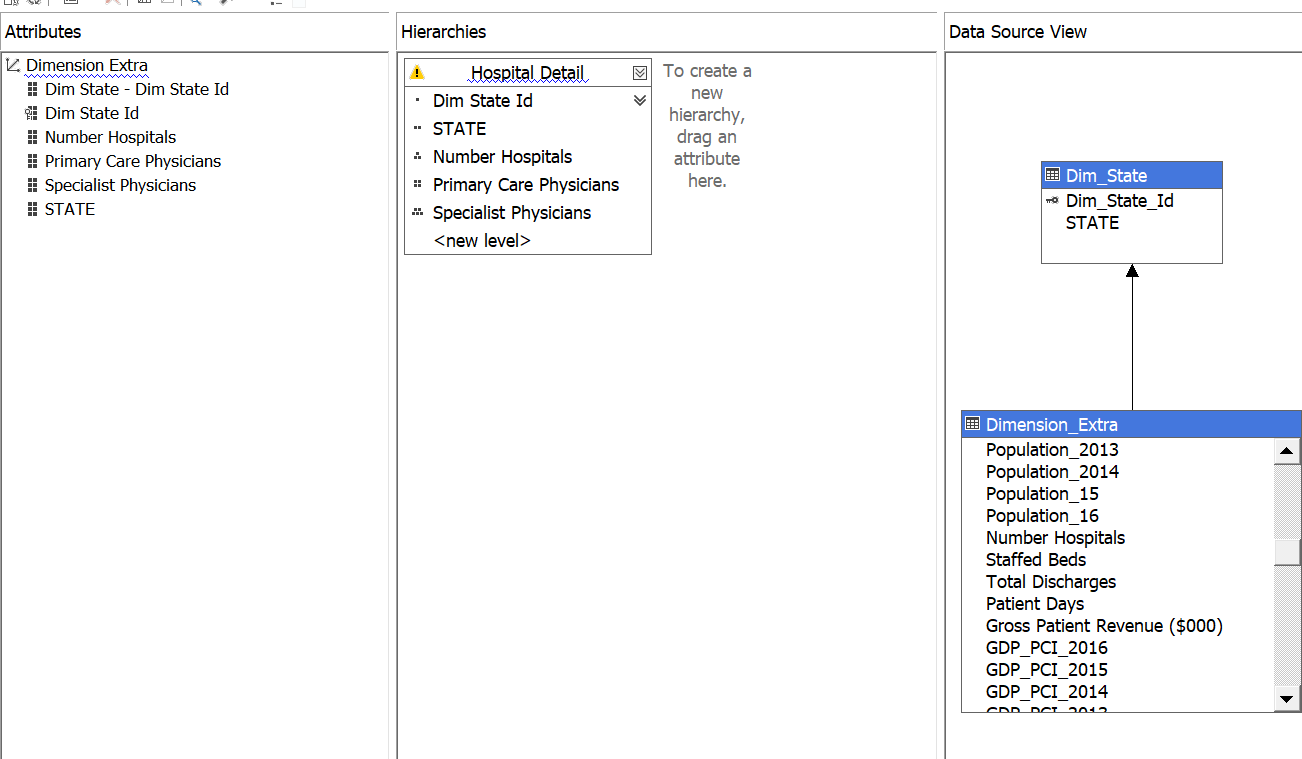
Fact Cubes Model (Snow flake Schema):

Step 10: Loading of Cubes in SSAS



Sample Hierarchies:

Step 11: Creating of hierarchies.



Cubes Analysis:

The below analysis displays the procedures covered by Medicare and money spent on each of them.

