PROJECT REPORT

Data Warehousing & Business Intelligence

Project by

Ad-Hawks

Group 2

Akshat Karambe

Maitri Brahmbhatt

Mayur Bhat

Nupur Dongare

Professor

Vincent Lattuada

Northeastern University
College of Engineering

Department of Mechanical & Industrial Engineering
334 Snell Engineering Center
360 Huntington Avenue
Boston, MA 02115

Version Control

Sl. No.	Version	Document Name	Changes	Date
1	V1	Project Report	First Document	03/28/2019
2	V2	Project Report	Added data description tables and Overview figures	04/03/2019
3	V3	Project Report	Added the toad model and Acronyms and Abbreviations	04/11/2019
4	V4	Project Report	Added the toad model and its DDL in the appendix section	04/26/2019

Acknowledgement

We would like to take this opportunity to acknowledge and express our sincere appreciation to the individuals who have helped us finish this project effectively.

Firstly, we would like to thank Professor Vincent Lattuada, College of Engineering, Northeastern University, for guiding us through the entire process of report preparation by providing valuable feedback at all stages of development and for teaching us the necessary tools required to efficiently tackle problems and overcome setbacks during the implementation of the project.

Secondly, we would like to thank our peers for providing an inducive atmosphere to understand, learn, contemplate and apply concepts and ideas to our individual projects.

Last but not least, we would like to thank Mr. Ayush Jain, Teaching Assistant for Datawarehouse and Business Inelligence for all the inputs, tips and tricks, and constructive criticism provided through the course of the semester ensuring that we are proficient in the concepts and fundamentals rather than only the implementation.

Table of Contents

List of Acronyms and Abbreviations

FFS	Fee-For-Services
IP	In-Patient
OP	Out-Patient
FQHC	Federally Qualified Health Center
HH	Home Health
INT	Integer
NN	Not Null

List of Figures

List of Tables

1.0 Introduction

1.1 Purpose

Healthcare in the United States is expensive! An accident causes not just physical, emotional and psychological damage, but economic too. Approximately 15% of the population is over 65 years of age. To aid people who are 65 and above, the government provides "Medicare" – A health program. Approximately 12% of the population falls below the poverty line. To aid people who fall in the poverty category, the government provides "Medicaid" – A health program with varying coverage based on the state. Our goal is to test out the efficiency of dispersion of funds for these two government health programs and test hypotheses with visual indicators which throws light on areas of improvement.
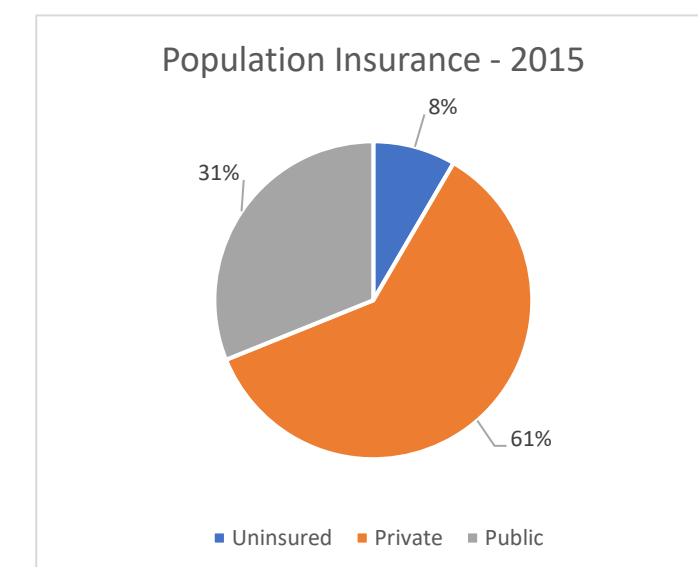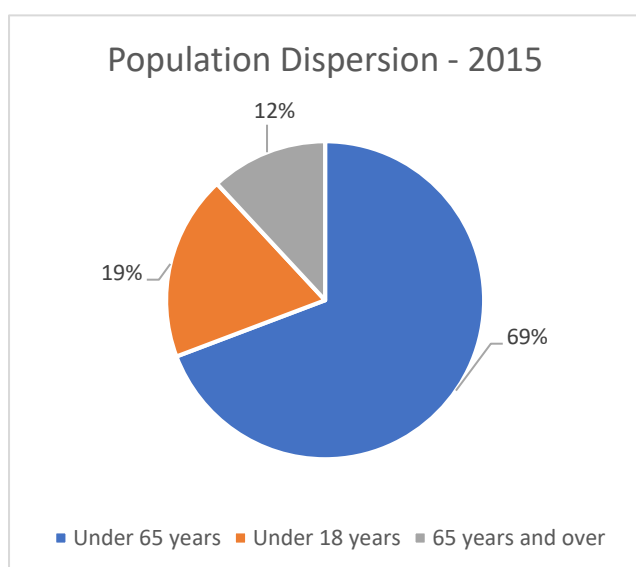

Figure 1: Population dispersion in the year 2015 by age group

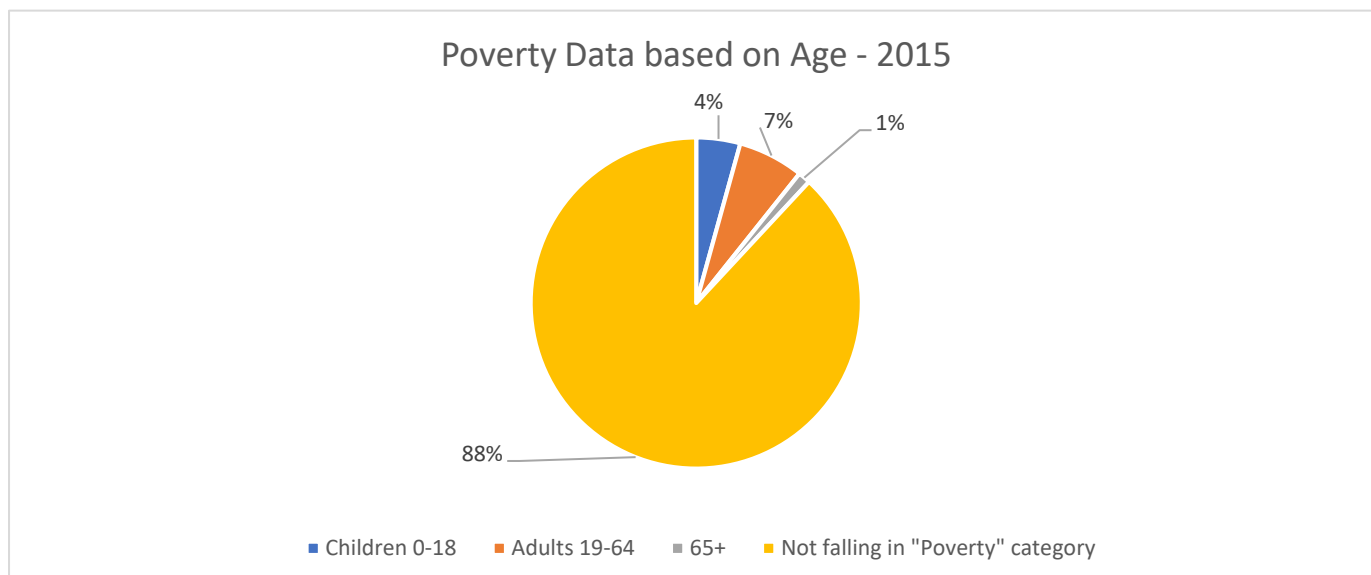Figure 2: Type of Population Insurance in the year 2015 by age group

Figure 3: Total Poverty in the year 2015 by age group

1.2 Scope

The scope of this project is to visually understand government spending on Medicare and Medicaid across the 50 states and provide recommendations for resource allocation.

Deliverables:
   a) Project report giving a high level view of the data, method of analysis and outcome.
   b) Visualization dashboard providing a holistic view of hypothesis testing in a visual manner.
   c) A presentation detailing the process of the project and to answer custom queries through visualization.

Requirements:
   a) Medical data for Medicare expenditure and enrollment for atleast 2 years.
   b) Medical data for Medicaid expenditure and enrolment for atleast 2 years.
   c) Demographics data for North America
   d) Population data for North America
   e) Poverty data for North America

Metodology:
   I.   Build 10-12 hypotheses like below:
         a. States where the number of people whose age is >65 is high, Medicare spending is also high.
         b. States which have high levels of poverty will have high Medicaid spending.
         c. Government health spending is uniform between 2 genders.
         d. States which have high income should have lower Medicaid spending owing to people being able to afford Private Insurance.

   II.   Test these hypotheses visually and report out the findings.

Out of scope:

   a) Other terretories are not included in our analysis as data on these are vastly limited.

Flowchart

This section gives us a general flow of the entire project.

Data Selection

- Medicare and Medicaid data from CMS.gov
- Income, Poverty data from Census.gov
- Health Insurance Dispersion data ACS-Census.gov

Understanding the data

- Went through the data description for the selected data sets
- Understood and selected the required and vital features in the data and their relations
- Took notes and marked the required deductions
- Based on the acquired knowledge decided the hypothesis to work on

Data Scrubbing

- Cleaned the data by removing the unnecessary attributes
- Added some calculated fields with the help of excel

ETL

- Designed a data base in toad and with the generated DDL scripts created a database in SSMS
- Extracted the data from the cleaned csv files and loaded it into tables via SSIS, first in staging and then in destination
- This cleaned and transformed data is loaded into dimension and fact tables of Data Warehouse
- Used Star schema to create cubes based on the data investigation

Business Intelligence

- Used tableau for visualizations
- These visuals will gives us insights about the data
- Interactive reports are combined into dashboards to visualize correlations between all data sets

2.0 Table Overview

Health Insurance Dispersion
- State
- Type of Coverage
- Number of People
- Age Group
- Year

Income
- State
- Personal Income (Millions of Dollars)
- Per Capital Income (Dollars)
- Regional Price Parity (Base line is 100)
- Year

Poverty
- State
- Age Group
- Race
- Poverty Rate
- Year

Medicaid
- State
- Age Group
- Gender
- Race
- Total Cost
- IP Actual Cost
- OP Actual Cost
- Prescribed Drugs
- Hospice Benefits
- Federally Qualified Health Centre
- Rehabilitative Services
- Home Health Services
- Year

Medicare
- State
- Age Group
- Gender
- Race
- Total Cost
- IP Actual Cost
- OP Actual Cost
- Prescribed Drugs
- Hospice Benefits
- Federally Qualified Health Centre
- Rehabilitative Services
- Home Health Services
- Year

2.1 Table Description

a. Health Insurance Dispersion

This entity contains data pertaining to the type of health care coverage that the people in the United States have categorized based on State, Age Group, Race and Year. The entity has been populated with data from multiple sources. It is an approximation based on weighted population growth as the census is collected once every 10 years. It has information about the number of people who are Uninsured and Insured through various sources like Medicaid, Medicare, Private Insurance and Public Insurance.

b. Income

This table has state wise income for the years 2015 and 2016. It will have attributes like Personal Income, Per Capita Income and Regional Price Parity apart from State and Year columns. This is used to generate a few calculated attributes like Metric C. This gives a wholistic view of Income data categorized by State, Age group, Gender and Race.

c. Poverty

This table will comprise of poverty levels in each state for years 2015 & 2016. We are considering the poverty levels in each state by age group, gender and race to get the required hypothesis.

Note: Poverty at a state level is consistent across age groups as the values are in percent and would make sense to utilize them as percentages itself rather than numbers.

a. Medicare

This table contains 100 percent of Medicare claims for beneficiaries who are enrolled in the fee-for-service (FFS) program as well as enrolment and eligibility data. It has calendar years 2015-2016 and has information on demographics, spending, and service utilization for Medicare beneficiaries in different parts of the country. State wise cost spent on Inpatient and Outpatients is considered too

a. Medicaid

This table has state-by-state total expenditures by program for the Medicaid Program, Medicaid Administration and CHIP programs. It has 21 attributes and 318 records. Attributes are based on total costs per state by age group, gender and race. State wise cost spent on Inpatient and Outpatients is considered too.

3.0 Warehouse Overview

We intend to make a warehouse which has 5 dimensions and 6 fact tables. The general overview is as shown below.
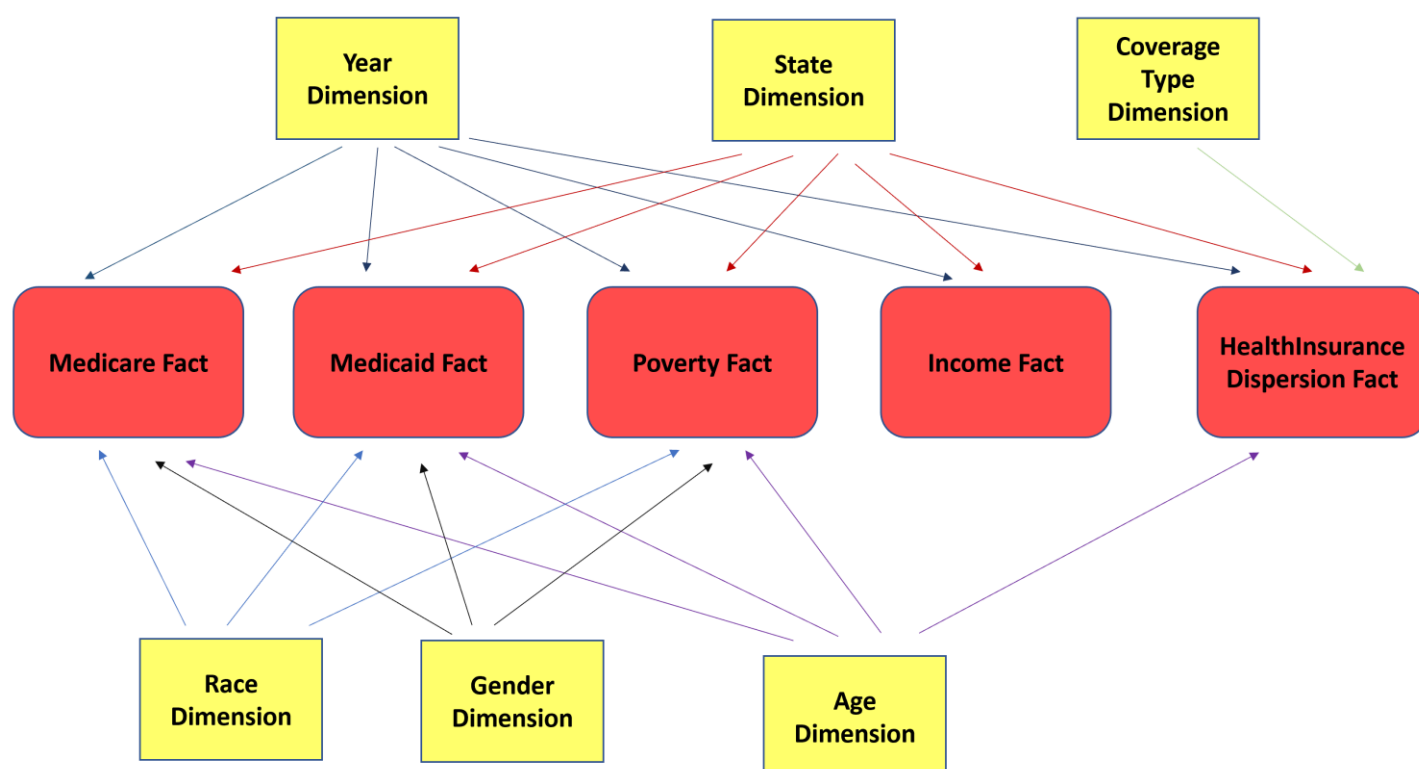

Figure 5: Overview of the Data warehouse with the Facts and Dimensions
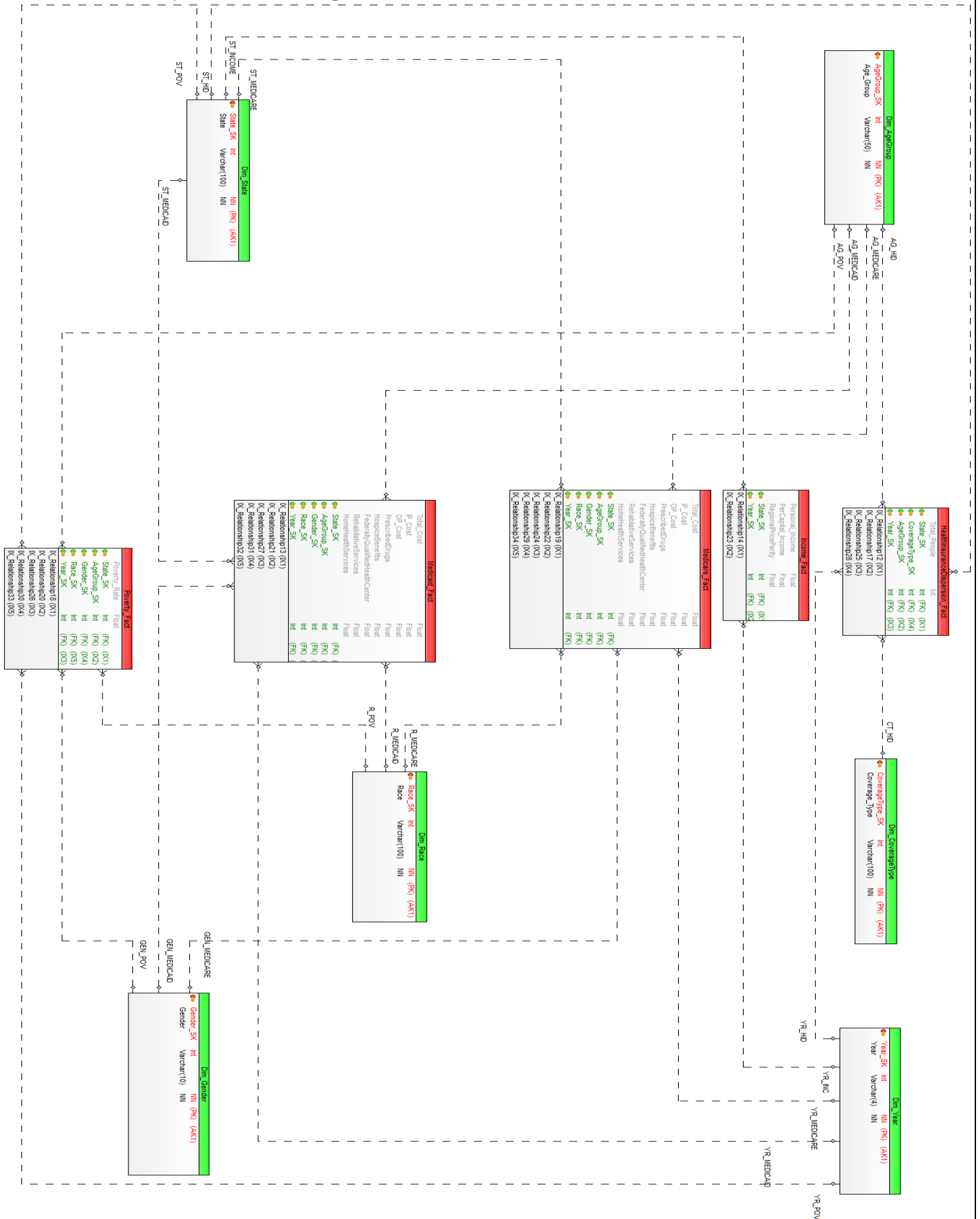
3.1 Warehouse Physical Model Design

Figure 6: Warehouse Physical Model Design with relationships between Dimensions and Facts

3.1 Entity Report

Full Name	Primary Key	#Attributes
Dim_AgeGroup	AgeGroup_SK	2
Dim_CoverageType	CoverageType_SK	2
Dim_Gender	Gender_SK	2
Dim_Race	Race_SK	2
Dim_State	State_SK	2
Dim_Year	Year_SK	2
HealthInsuranceDispersion_Fact		5
Income_Fact		5
Medicaid_Fact		13
Medicare_Fact		13
Poverty_Fact		6

3.2 Dimension Report

Dim_State

Business Name (Attribute Name)	Datatype	Constraints	Definition	Example Data
State_SK	Int	NN (PK)	This is the surrogate key which will be the primary key for the state dimension which is an autoincrmented value	1
State	VARCHAR (100)	NN	This will hold the name of the state	Alabama

Dim_Year

Business Name (Attribute Name)	Datatype	Constraints	Definition	Example Data
Year_SK	Int	NN (PK)	This is the surrogate key which will be the primary key for the year dimension which is an autoincrmented value	1
Year	VARCHAR (4)	NN	This will show the year	2015

Dim_CoverageType

Business Name (Attribute Name)	Datatype	Constraints	Definition	Example Data
CoverageType_SK	Int	NN (PK)	This is the surrogate key which will be the primary key for the coverage type dimension which is an autoincrmented value	1
Coverage_Type	VARCHAR (4)	NN	This column will have different health insurance coverage	Medicare

13

Dim_Race

Business Name (Attribute Name)	Datatype	Constraints	Definition	Example Data
Race_SK	Int	NN (PK)	This is the surrogate key which will be the primary key for the race dimension which is an autoincrmented value	1
Race	VARCHAR (100)	NN	This will have names of all the race type	Hispanic

Dim_AgeGroup

Business Name (Attribute Name)	Datatype	Constraints	Definition	Example Data
AgeGroup_SK	Int	NN (PK)	This is the surrogate key which will be the primary key for the age group dimension which is an autoincrmented value	1
Age_Group	VARCHAR (100)	NN	This will hold different age groups	0-18

Dim_Gender

Business Name (Attribute Name)	Datatype	Constraints	Definition	Example Data
Gender_SK	Int	NN (PK)	This is the surrogate key which will be the primary key for the age group dimension which is an autoincrmented value	1
Gender	VARCHAR (10)	NN	This will hold different genders	Male

3.3 Fact Report

HealthInsuranceDispersion_Fact

Business Name (Attribute Name)	Datatype	Constraints	Definition
Total_People	Int		This will hold a a number for health insurance subscribers for each state, coverage type, age group and year in thousands
State_SK	Int	(FK)	This is the surrogate key for Dim_State
CoverageType_SK	Int	(FK)	This is the surrogate key for Dim_CoverageType
AgeGroup_SK	Int	(FK)	This is the surrogate key for Dim_AgeGroup
Year_SK	Int	(FK)	This is the surrogate key for Dim_Year

Income_Fact

Business Name (Attribute Name)	Datatype	Constraints	Definition
Personal_Income	Float		Income received by, or on behalf of, all persons from all sources in that particular state in Millions of Dollars
PerCapita_Income	Float		Calculated as the total personal income of the residents of a given area divided by the population of the area in dollars
RegionalParity	Float		Regional price levels expressed as a percentage of the overall national price level for a given year having Base line is 100
State_SK	Int	(FK)	This is the surrogate key for Dim_AgeGroup
Year_SK	Int	(FK)	This is the surrogate key for Dim_Year

Medicare_Fact

Business Name (Attribute Name)	Datatype	Constraints	Definition
Total_Cost	Float		Actual Medicare total costs inclusive of all the benefits
IP_Cost	Float		
OP_Cost	Float		
PrescribedDrugs	Float		
HospiceBenefits	Float		
FederallyQualifiedHealth Center	Float		
RehabilativeServices	Float		
HomeHealthServices	Float		
State_SK	Int	(FK)	This is the surrogate key for Dim_State
AgeGroup_SK	Int	(FK)	This is the surrogate key for Dim_Gender
Gender_SK	Int	(FK)	This is the surrogate key for Dim_AgeGroup
Race_SK	Int	(FK)	This is the surrogate key for Dim_Year
Year_SK	Int	(FK)	This is the surrogate key for Dim_Year

Poverty_Fact

Business Name (Attribute Name)	Datatype	Constraints	Definition
Povert_Rate	Float		
State_SK	Int	(FK)	This is the surrogate key for Dim_State
AgeGroup_SK	Int	(FK)	This is the surrogate key for Dim_Gender
Gender_SK	Int	(FK)	This is the surrogate key for Dim_AgeGroup
Race_SK	Int	(FK)	This is the surrogate key for Dim_Year
Year_SK	Int	(FK)	This is the surrogate key for Dim_Year

Medicaid_Fact

Business Name (Attribute Name)	Datatype	Constraints	Definition
Total_Cost	Float		
IP_Cost	Float		
OP_Cost	Float		
PrescribedDrugs	Float		
HospiceBenefits	Float		
FederallyQualifiedHealthCenter	Float		
RehabilativeServices	Float		
HomeHealthServices	Float		
State_SK	Int	(FK)	This is the surrogate key for Dim_State
AgeGroup_SK	Int	(FK)	This is the surrogate key for Dim_Gender
Gender_SK	Int	(FK)	This is the surrogate key for Dim_AgeGroup
Race_SK	Int	(FK)	This is the surrogate key for Dim_Year
Year_SK	Int	(FK)	This is the surrogate key for Dim_Year

3.4 Index Report

Full Name	Unique	Entity
IX_Relationship12	NO	HealthInsuranceDispersion_Fact
IX_Relationship17	NO	HealthInsuranceDispersion_Fact
IX_Relationship25	NO	HealthInsuranceDispersion_Fact
IX_Relationship28	NO	HealthInsuranceDispersion_Fact
IX_Relationship14	NO	Income_Fact
IX_Relationship23	NO	Income_Fact
IX_Relationship13	NO	Medicaid_Fact
IX_Relationship21	NO	Medicaid_Fact
IX_Relationship27	NO	Medicaid_Fact
IX_Relationship31	NO	Medicaid_Fact
IX_Relationship32	NO	Medicaid_Fact
IX_Relationship19	NO	Medicare_Fact
IX_Relationship22	NO	Medicare_Fact
IX_Relationship24	NO	Medicare_Fact
IX_Relationship29	NO	Medicare_Fact
IX_Relationship34	NO	Medicare_Fact
IX_Relationship18	NO	Poverty_Fact
IX_Relationship20	NO	Poverty_Fact
IX_Relationship26	NO	Poverty_Fact
IX_Relationship30	NO	Poverty_Fact
IX_Relationship33	NO	Poverty_Fact

3.5 Relationship Report

Full Name	Type	Parent Entity	Child Entity	Card.
AG_HID	Non-identifying	Dim_AgeGroup	HealthInsuranceDispersion_Fact	1:N
AG_MEDICAID	Non-identifying	Dim_AgeGroup	Medicaid_Fact	1:N
AG_MEDICARE	Non-identifying	Dim_AgeGroup	Medicare_Fact	1:N
AG_POV	Non-identifying	Dim_AgeGroup	Poverty_Fact	1:N
CT_HID	Non-identifying	Dim_CoverageType	HealthInsuranceDispersion_Fact	1:N
GEN_MEDICAID	Non-identifying	Dim_Gender	Medicaid_Fact	1:N
GEN_MEDICARE	Non-identifying	Dim_Gender	Medicare_Fact	1:N
GEN_POV	Non-identifying	Dim_Gender	Poverty_Fact	1:N
R_MEDICAID	Non-identifying	Dim_Race	Medicaid_Fact	1:N
R_MEDICARE	Non-identifying	Dim_Race	Medicare_Fact	1:N
R_POV	Non-identifying	Dim_Race	Poverty_Fact	1:N
ST_HID	Non-identifying	Dim_State	HealthInsuranceDispersion_Fact	1:N
ST_INCOME	Non-identifying	Dim_State	Income_Fact	1:N
ST_MEDICAID	Non-identifying	Dim_State	Medicaid_Fact	1:N
ST_MEDICARE	Non-identifying	Dim_State	Medicare_Fact	1:N
ST_POV	Non-identifying	Dim_State	Poverty_Fact	1:N
YR_HID	Non-identifying	Dim_Year	HealthInsuranceDispersion_Fact	1:N
YR_INC	Non-identifying	Dim_Year	Income_Fact	1:N
YR_MEDICAID	Non-identifying	Dim_Year	Medicaid_Fact	1:N
YR_MEDICARE	Non-identifying	Dim_Year	Medicare_Fact	1: N
YR_POV	Non-identifying	Dim_Year	Poverty_Fact	1:N

4.0 ETL

4.1 Data cleaning and transformation

The data was acquired from CMS.gov, kff.org and other approved sources.

The primary files are detailed as follows:

Data Source Name	Number of attributes	Number of Rows	Data quality(0-5)	Year	Source
Poverty by Age(Percentages)	3	55	5	2015 and 2016	kff.org
Poverty by Age(Number)	3	55	5	2015 and 2016	kff.org
Poverty by Gender(Percentages)	2	55	5	2015 and 2016	kff.org
Poverty by Gender(Number)	2	55	5	2015 and 2016	kff.org
Poverty by Race(Number)	7	55	5	2015 and 2016	kff.org
Poverty by Race(Number)	7	55	5	2015 and 2016	kff.org
Medicare - State County All Table	245	3250	4	2015	cms.gov
Medicare - State County All Table	245	3250	4	2016	cms.gov
Medicaid Speding	12	7231	3	2016	cms.gov
Medicaid Spending	960	52	5	2015	cms.gov
Medicaid by Age(Percentages)	3	52	5	2015 and 2016	kff.org
Medicaid by Age(Number)	3	52	5	2015 and 2016	kff.org
Medicaid by Gender(Percentages)	2	52	5	2015 and 2016	kff.org
Medicaid by Gender(Number)	2	52	5	2015 and 2016	kff.org
Medicaid by Race(Number)	7	52	5	2015 and 2016	kff.org
Medicaid by Race(Number)	7	52	5	2015 and 2016	kff.org
RPP and Income data	4	56	5	2015	census.gov
RPP and Income data	4	56	5	2016	census.gov
Health Insurance Dispersion Data	208	28	4	2015	cms.gov
Health Insurance Dispersion Data	208	28	4	2016	cms.gov

The above data is from raw source and requires a significant amount of cleaning, transposing, calculations and transformations.

Some of the issues faced during the transformation process are listed below:

a) Data is inconsistent across the various years
b) Data which should be presented in our source staging tables are present in multiple tables.
c) Some data needs to be transposed to make sense.
d) Error rates in population data were throwing off the totals in our tables. Needed to be handled properly.
e) Income data wasn't available at a demographic level, this needs some calculated data.
f) The first pass at building the data cube threw errors due to certain measures not being present in all fact tables.

Methods used to clean and transform data:

a) Utilized Excel Macros to insert blank rows in the source file.
b) Used advanced lookup like VLOOKUP and HLOOKUP to get data at a state/age group/gender/race level.
c) Missing values have been changed from N/A to 0 as there was no method to accurately predict values.
d) Base population fact was created which included the percent of population based on gender, race, age group and state. This gives us an understanding of the scale of the population. Eg: 0.533% is the percent of the population of people in Alabama who are male, 0-18 and Non-hispanic White compared to the entire population.
e) Medicare and Medicaid data which is available only at the age group level can now be transformed to indicate costs and expenditures at a demographic level.
f) Loaded data from semi-formatted csv files into Rstudio to clean up and make transformation whereever necessary.

Final Data is detailed as follows:

Data Source Name	Number of attributes	Number of Rows	Data quality(0-5)	Year
Health Insurance Dispersion	5	1836	5	2015
Health Insurance Dispersion	5	1836	5	2016
Income	5	51	5	2015
Income	5	51	5	2016
Poverty	6	4452	5	2015
Poverty	6	4452	5	2016
Medicare Data	13	4368	5	2015
Medicare Data	13	4368	5	2016
Medicaid Data	13	4452	5	2015
Medicaid Data	13	4452	5	2016

4.2 Staging, Destination and Archive

First we loaded all the data into staging and then to the destination and archive. While doing so I inserted the information of the files which are being loaded in the there respective tables.
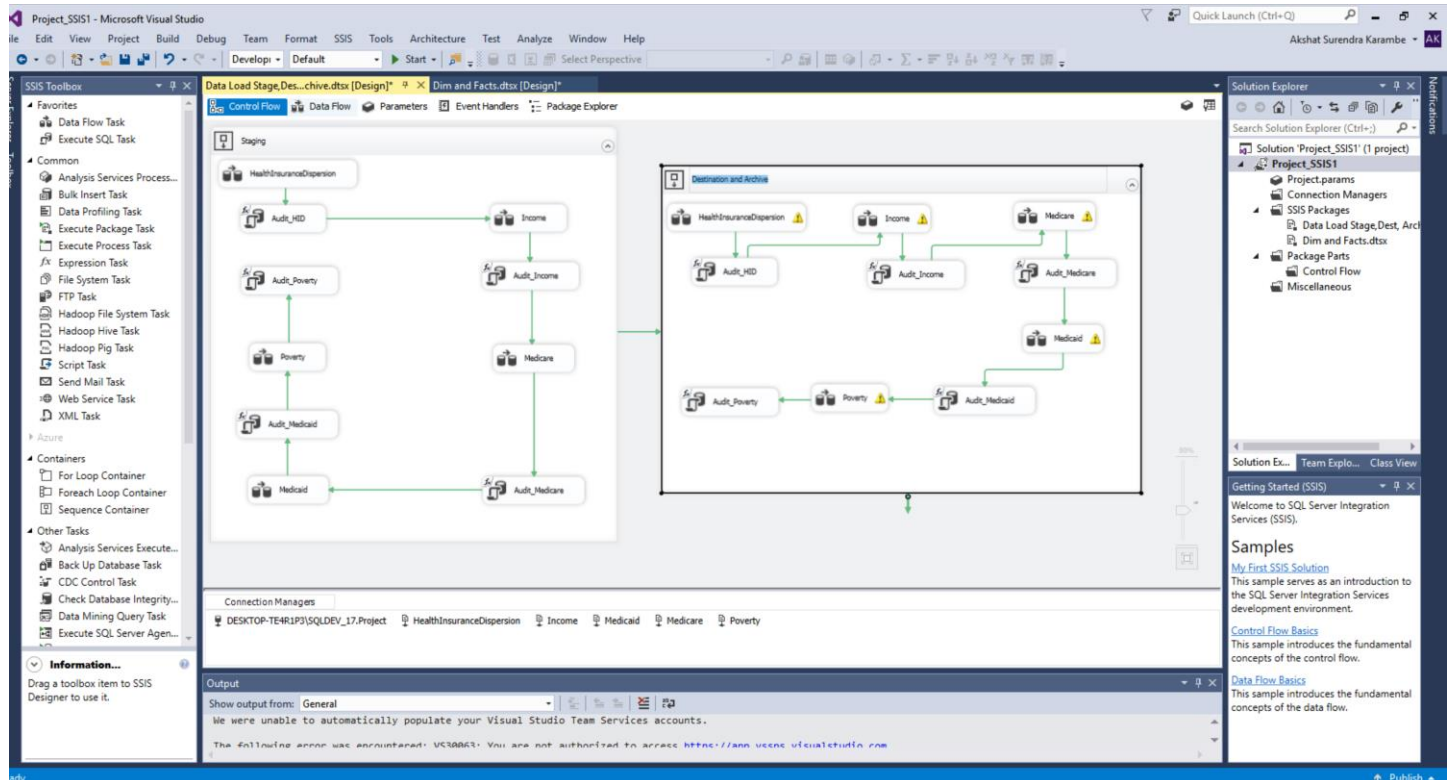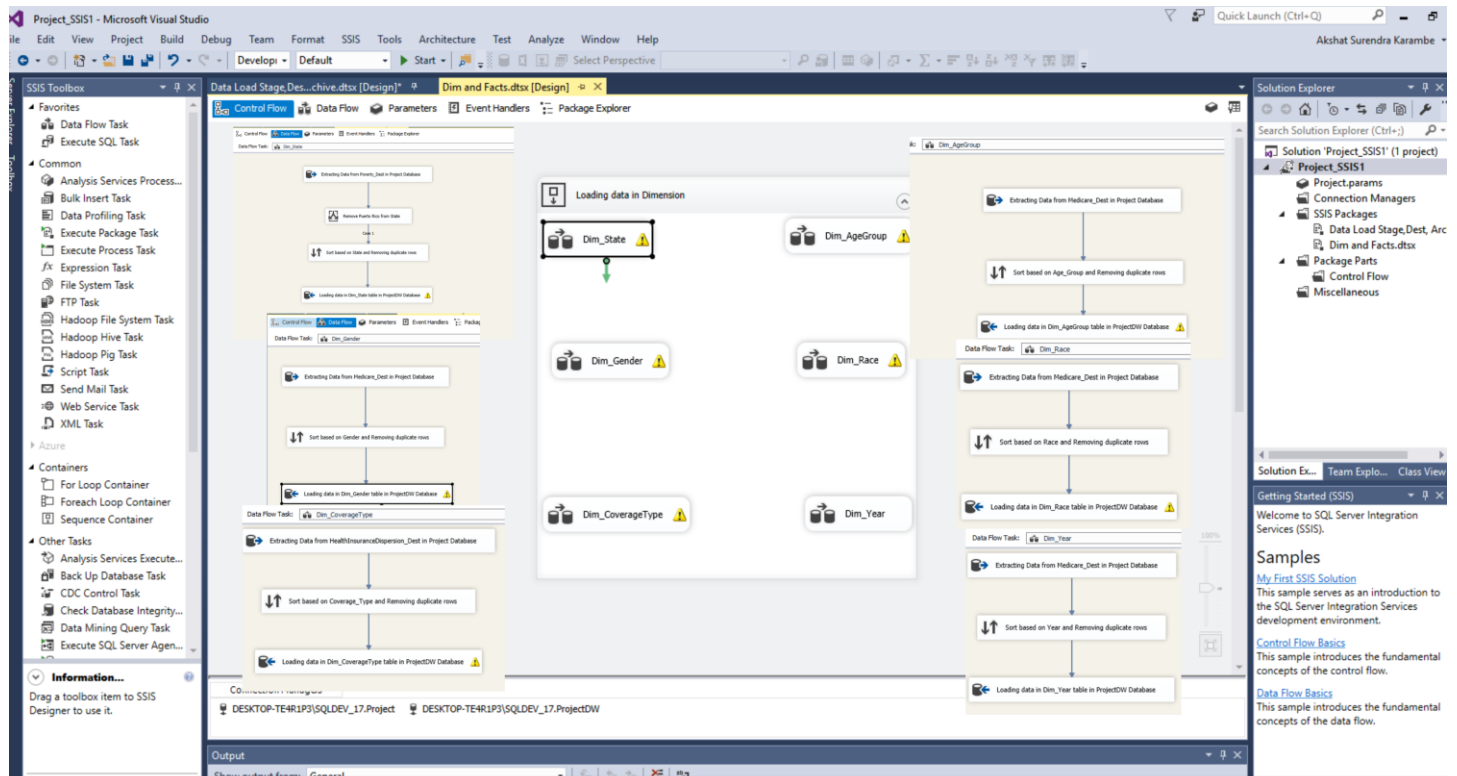

Here we can see the exact number of rows being loaded in each table along with the date and time of each load
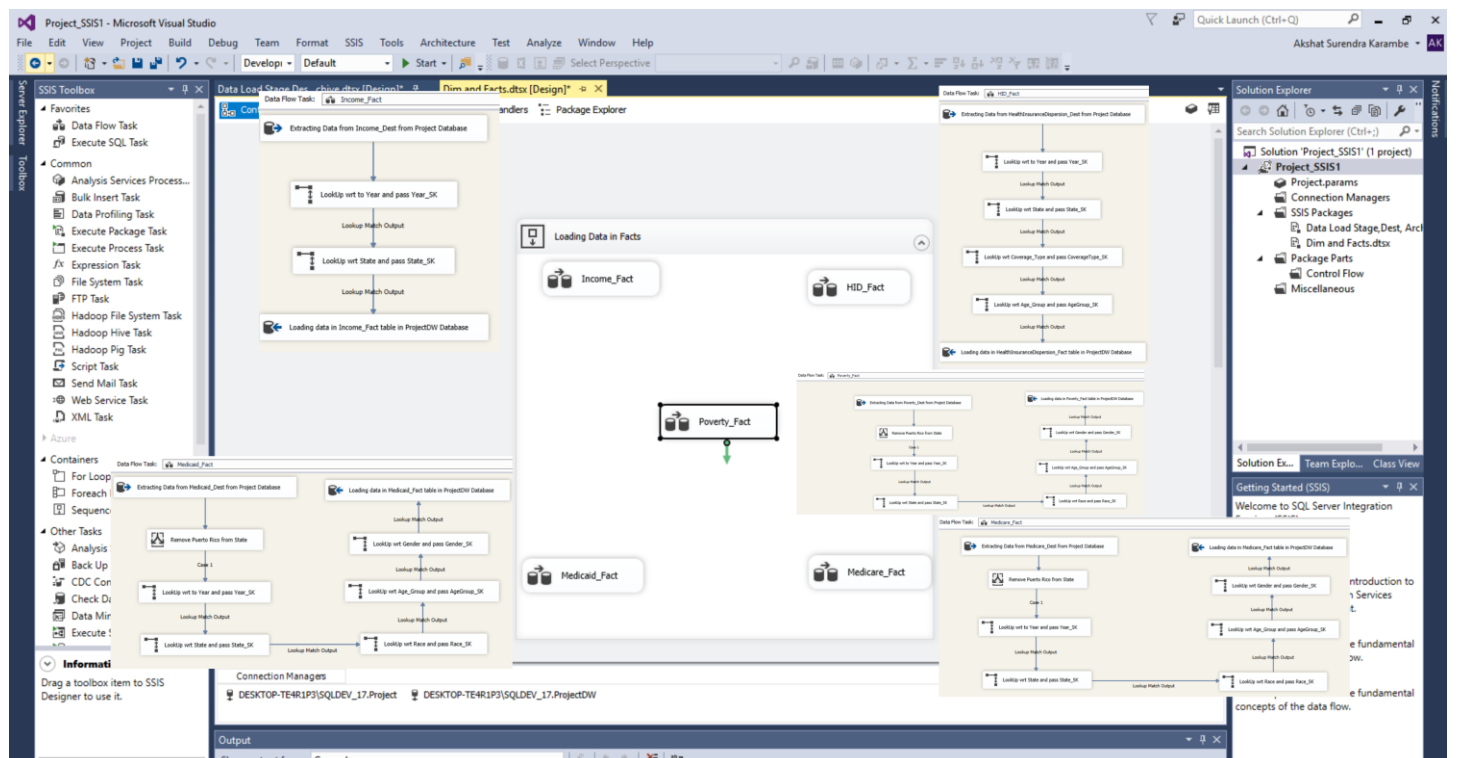
4.3 ETL: Dimensions and Facts

Extracted the data from destination table and loaded it in the required dimension tables in the warehouse
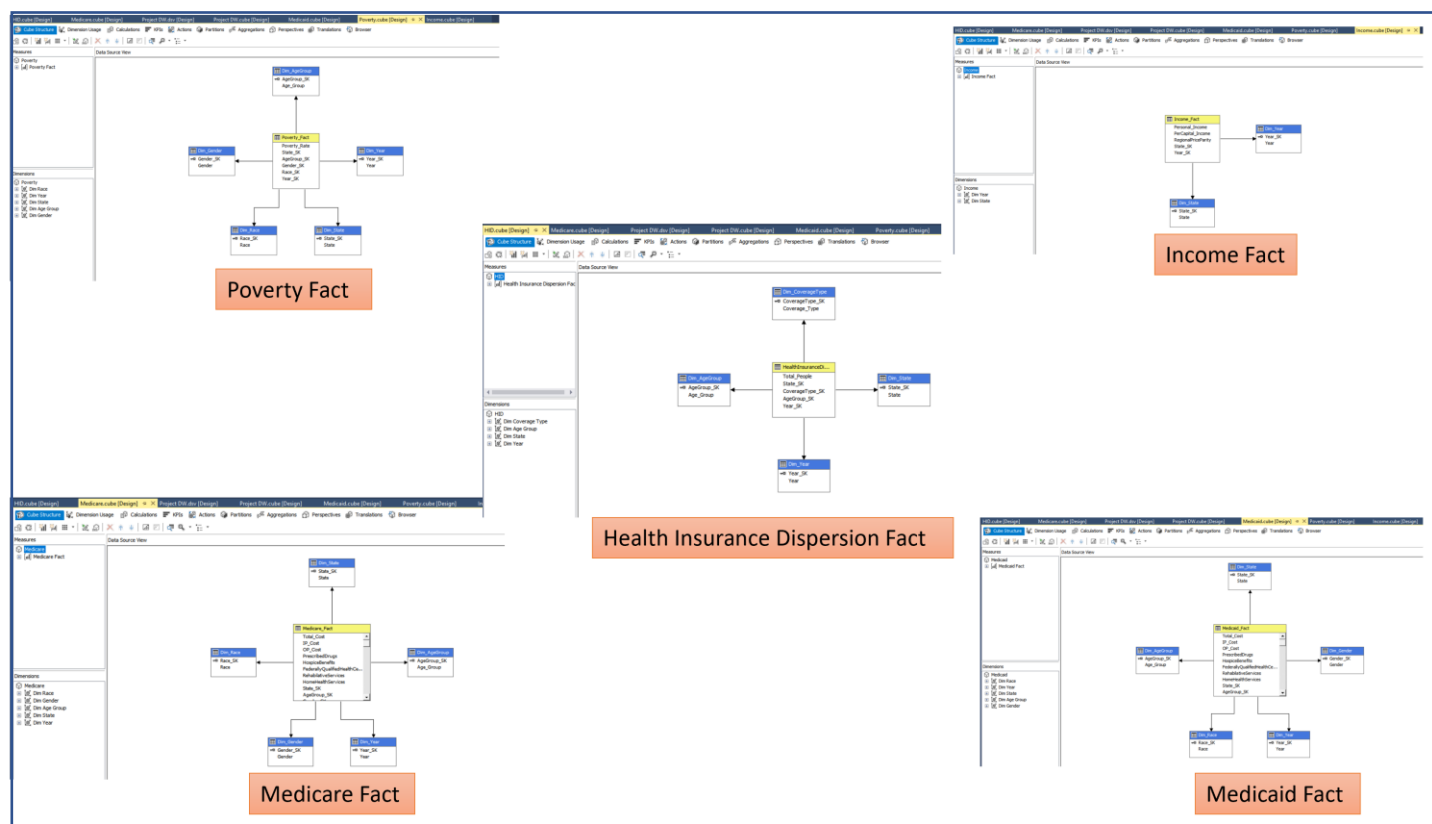

Extracted the data from destination table and loaded it in the required facts tables

4.4 SSAS: Cubes

After the etl we moved on to SSAS where we had to generate cubes. The figure below shows us all the cubes being generated.

Poverty Fact

Income Fact

Health Insurance Dispersion Fact

Medicare Fact

Medicaid Fact

5.0 Visualization

Group Visualizations:

MEDICAID DEMOGRAPHICS DASHBOARD:
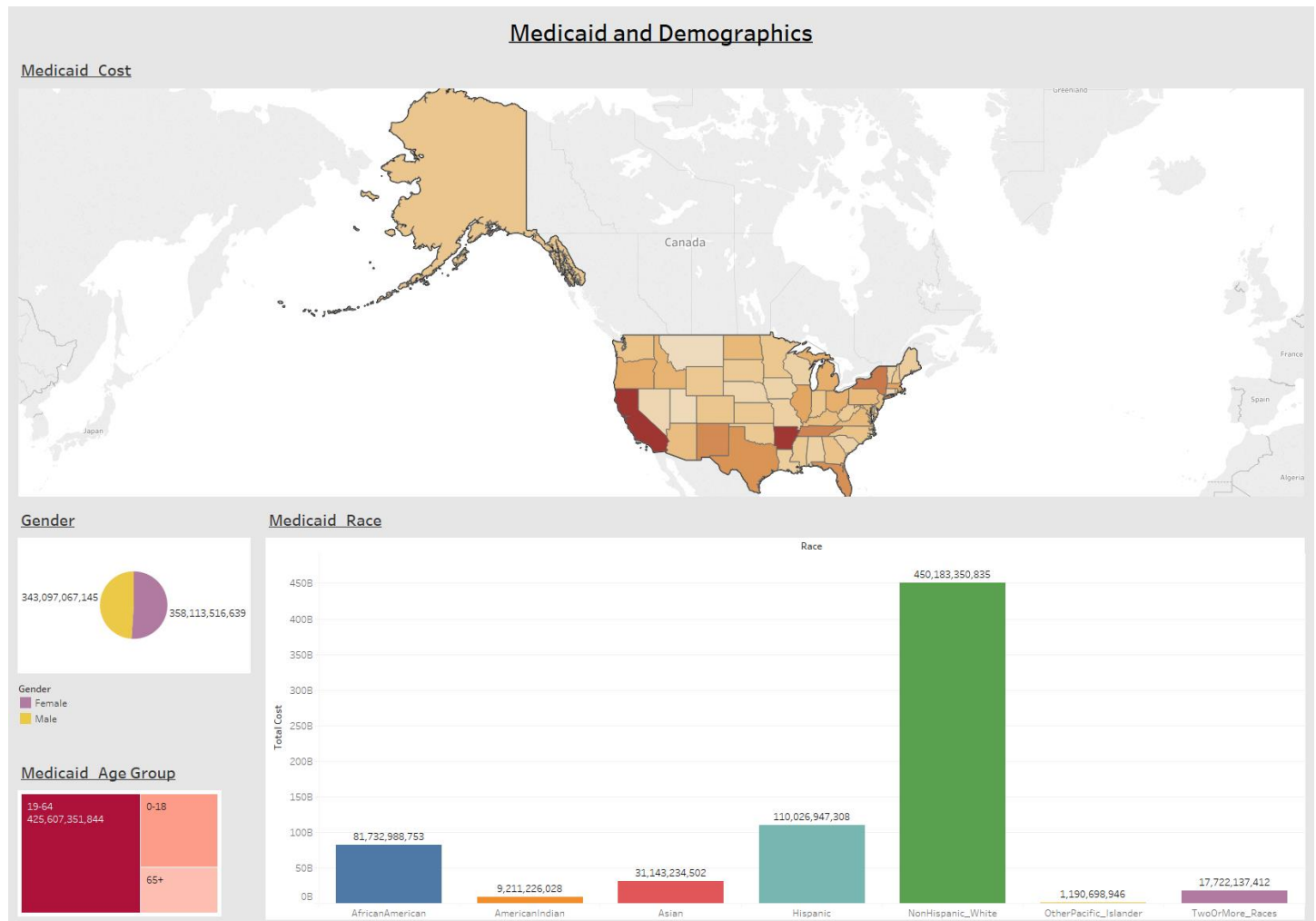
This Dashboard shows Demographic data For Medicaid such as Race, Gender and Age Group

MEDICARE DEMOGRAPHICS DASHBOARD:

This Dashboard shows Demographic data For Medicare such as Race, Gender and Age Group

Individual Visualizations:
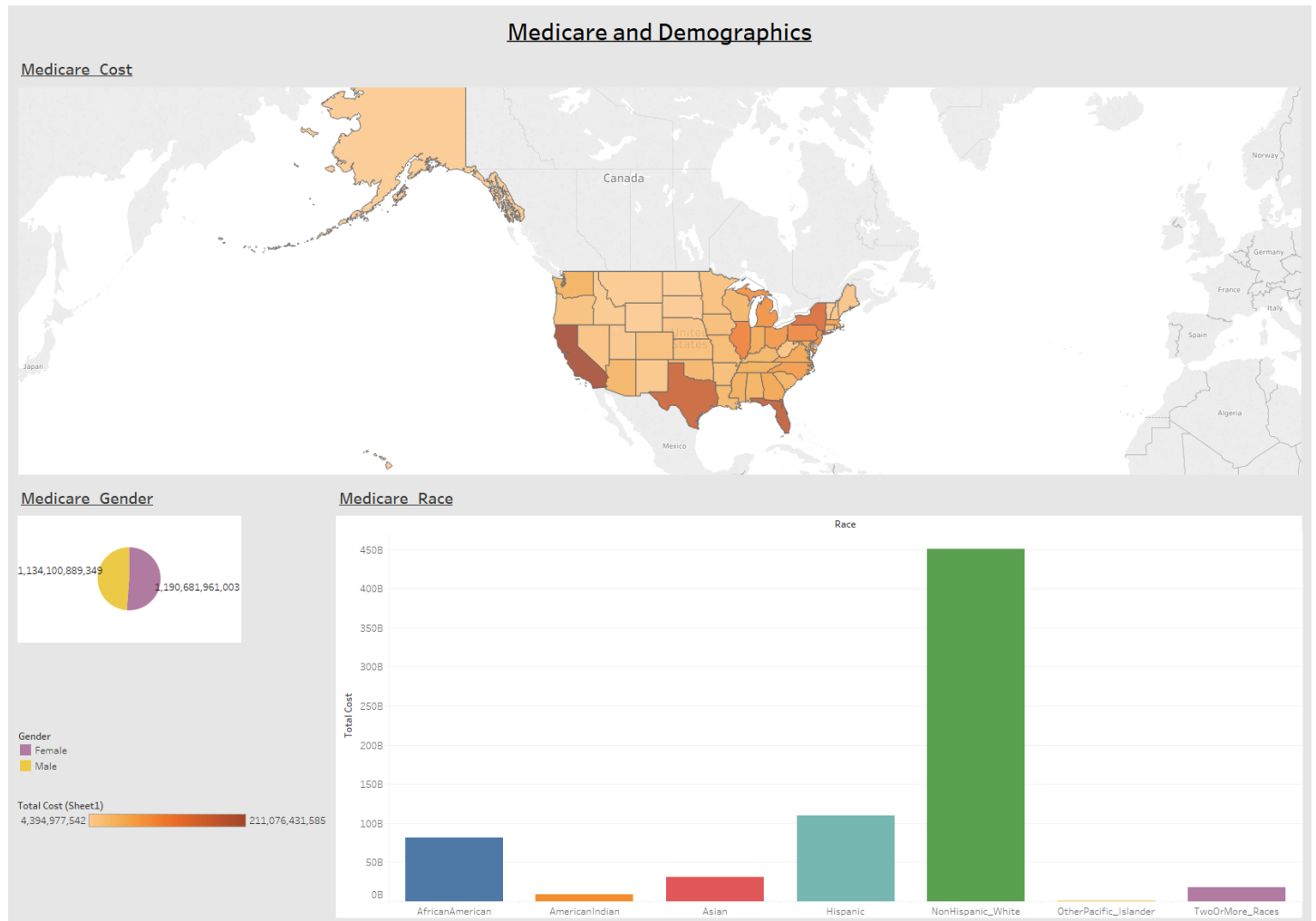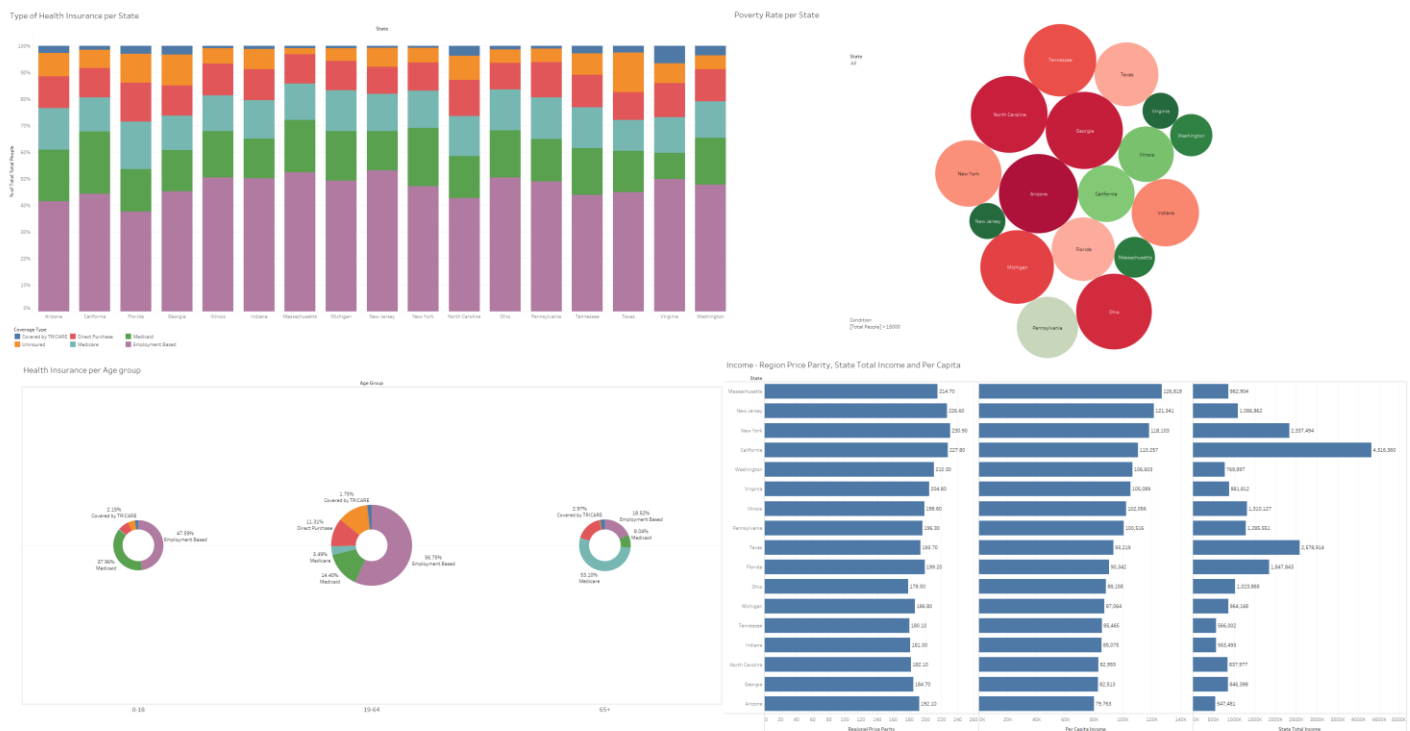

- Employment based Insurance is more in 19-64 Age group which makes more sense as this is the working Age group
- Medicare Insurance is more in the 65+ Age group than in others, this also makes sense as Medicare is for the older people and disabled.
- The Uninsured are more in 19-64 and less in older people.
- In bar graph for Income, we can see the top 10 states sorted based on Per Capita Income in descending order. We are going to test these with the hypothesis that the states having higher per capita income will have less number of uninsured people and the employment based insurances will be higher and vice versa.
- Using the bar graph of Income, we can see that Massachusetts has the highest per capita income and then if we check in Type of Health Insurance per State graph about the dispersions we can see that that employment-based insurance is High and the Uninsured are very less. So, by using the Income graph and the above graph we can say that per capita income is one factor contributing to the health insurance coverage you choose or not.
- Here the 2 states Mississippi and New Mexico have the most poverty rate and if compared with the earlier graph we can see that these states are amongst the least in the employment based insurance and also the government spending is more on Medicaid in these states.

Metric C variation across States
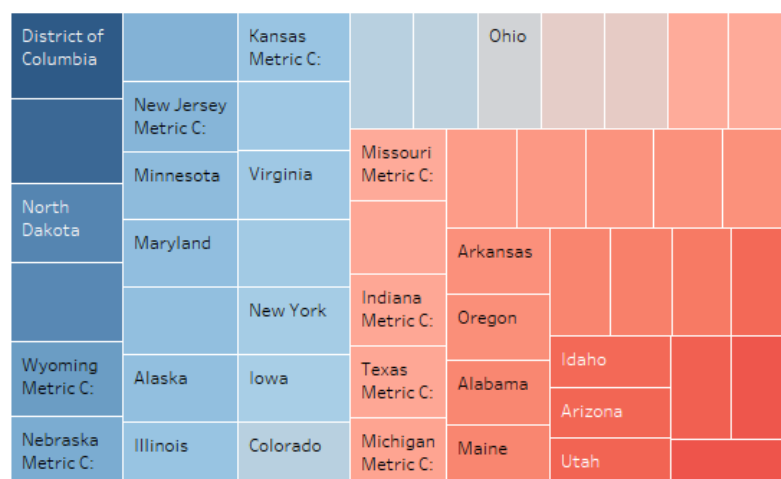
Metric C
0.8282 1.3198

Year
✔ 2015
✔ 2016
☐ Unknown
☐ Unknown

Age Group
✔ 0-18
✔ 19-64
✔ 65+
✔ Unknown
✔ Unknown

Coverage type by age in States where population > 15,000,000
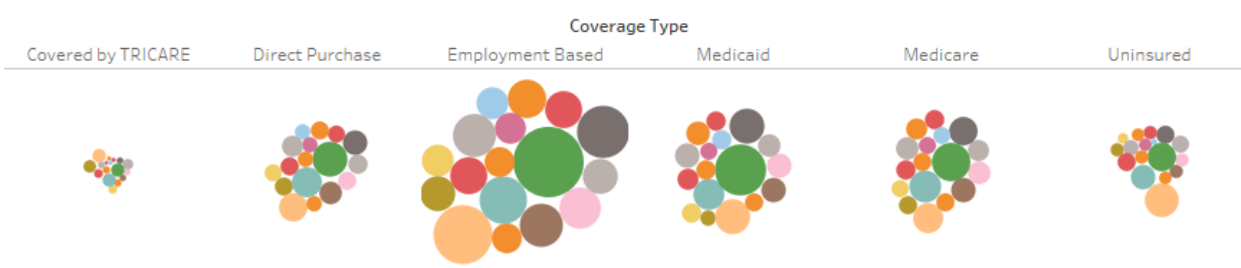
- In a particular state, if the cost of living is high, the poverty rates will be lower
- A calculated metric is used. Metric C = (Per capita income/RPP)/(National average per capita Income/ Baseline RPP)
- Metric C >1 => Higher than National Average.
- Higher the population of a state, higher will be employment.
- As a result, employer funded insurance will be higher. Especially in the 19-64 age group.

More Poverty > More Medicaid
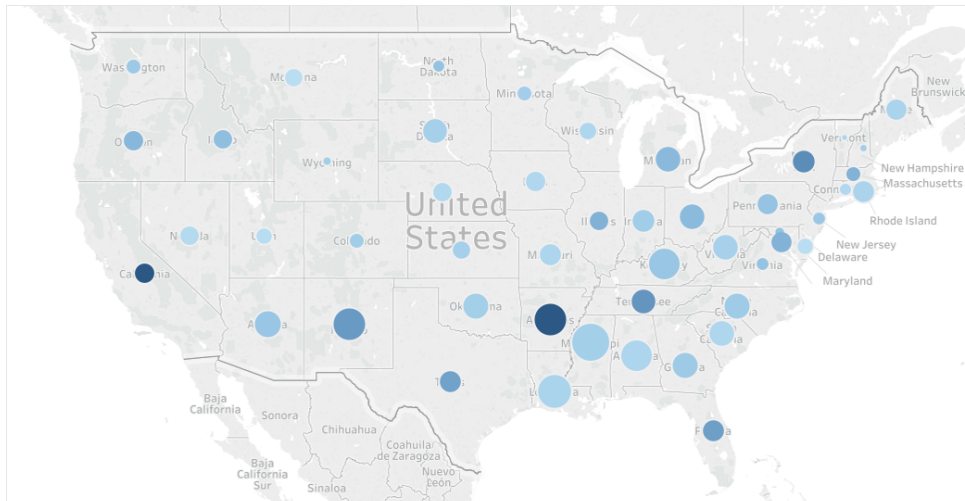
Less Income > More Medicaid

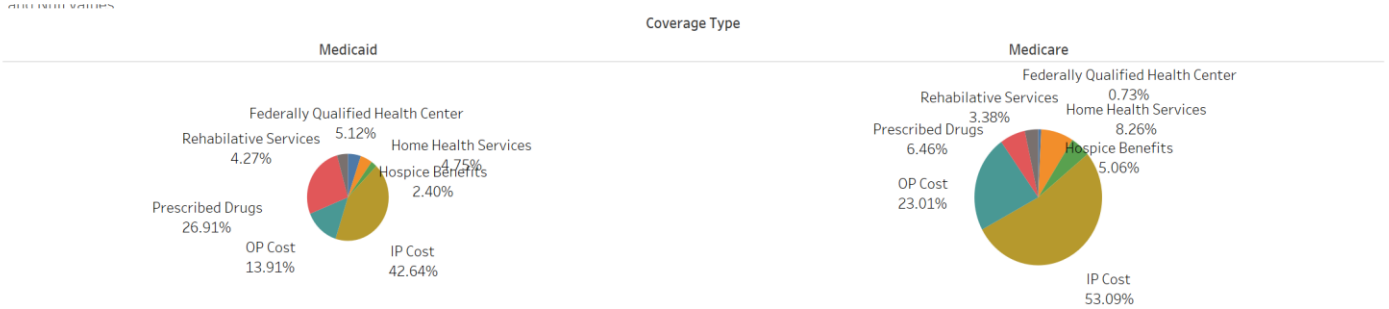| State	
California	2,158,180
Texas	1,289,457
New York	1,168,747
Florida	923,922
Illinois	655,064
Pennsylvania	642,776
New Jersey	543,431
Ohio	511,934
Virginia	440,906
Michigan	432,084
Massachusetts	431,452
Georgia	423,199
North Carolina	418,989
Washington	384,949
Maryland	342,727
Minnesota	283,257
Tennessee	283,001
Colorado	282,918
Indiana	281,747
Arizona	273,741
Wisconsin	267,607
Missouri	259,443
Connecticut	247,298
Louisiana	199,310
South Carolina	191,791
Alabama	186,974
Oregon	181,121
Oklahoma	172,877
Kentucky	171,735
Iowa	143,795
Kansas	137,311
Nevada	124,593
Utah	121,318
Arkansas	116,311
Mississippi	105,049
Nebraska	93,730
New Mexico	79,585
New Hampshire	74,538
Hawaii	70,538
West Virginia	67,425
Idaho	64,984
Maine	57,775
Rhode Island	53,053
District of Columbia	50,560
Delaware	45,316
Montana	43,980
North Dakota	41,878
Alaska	41,372
South Dakota	41,251
Wyoming	32,570
Vermont	30,819

More Income > Less Poverty (2)

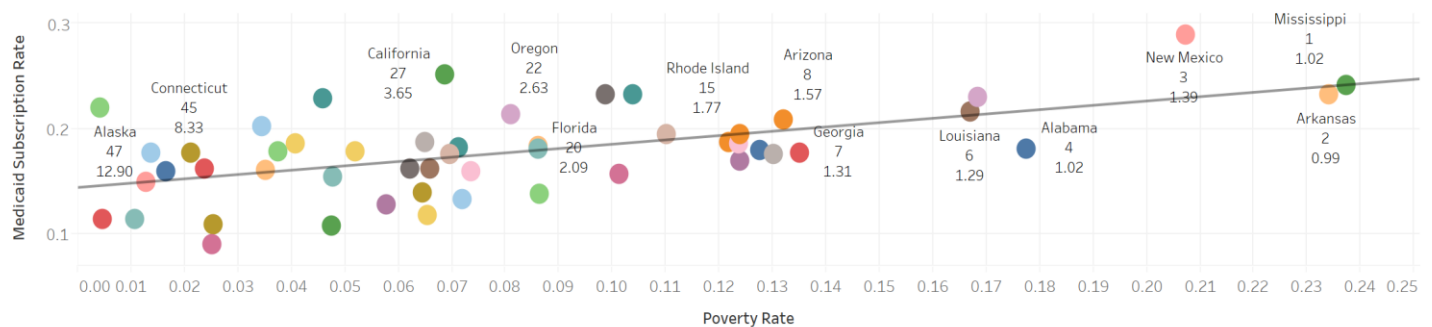| State	Avg. Per Capital Income
Mississippi	0.2550
West Virginia	0.1121
New Mexico	0.1855
Alabama	0.1662
Kentucky	0.1735
South Carolina	0.1083
Arkansas	0.1923
Idaho	0.0579
Arizona	0.1250
Utah	0.0429
Georgia	0.1186
North Carolina	0.1176
Montana	0.0550
Indiana	0.0904
Missouri	0.0816
Tennessee	0.1044
Nevada	0.0627
Maine	0.0755
Michigan	0.1078
Ohio	0.1142
Oklahoma	0.1163
Oregon	0.0693
Florida	0.0801

- In particular state, if the per capital income is high, then the poverty rate should be less.
- In particular state, if the poverty rate is high, then the total cost for Medicaid should be more.
- In particular state, if total personal income is Less, then the total cost for Medicaid should be more.

Medicaid-Medicare Comparison

and Null values

Coverage Type

Medicaid | Medicare

Medicaid pie chart:
- Federally Qualified Health Center 5.12%
- Rehabilitative Services 4.27%
- Home Health Services 4.75%
- Hospice Benefits 2.40%
- Prescribed Drugs 26.91%
- OP Cost 13.91%
- IP Cost 42.64%

Medicare pie chart:
- Federally Qualified Health Center 0.73%
- Rehabilitative Services 3.38%
- Home Health Services 8.26%
- Prescribed Drugs 6.46%
- Hospice Benefits 5.06%
- OP Cost 23.01%
- IP Cost 53.09%

Medicaid Coverage across the States

- Medicaid Subscription Rate is the ratio of Medicaid Subscribers to the total number of health insurance subscribers in the particular state.
- To check our hypothesis of Medicaid SUbscription being directly proportional to Poverty Rate, the ratio of MSR to PR was considered as Medicaid's coverage metric and should atleast be 1 to unfail our working hyothesis.
- However it fails for the state of Mississipi, where the ratio is below 1, showing there can be improvement in Medicaid coverage, however overall the Medicaid Coverage is satisfactory.
- More Services can be added under Home Health Services for Medicare as they seem to constitute a large portion of costs.
- Prescribed Drugs Services can be ramped up for Medicaid as they are a cover a large proportion of Medicaid Costs.

6.0 Conclusion

In conclusion, we can say that Medicare and Medicaid spending is consistent in states with lower income/higher percentage of poverty. But in states where the income levels are higher, we see some disparity in spending patterns. Our recommendation would be that the government prioritizes their spending on regions with lower income to maximize the effectiveness of resource allocation for improved health.

Appendix

Appendix-A: Dataset links

Links to all the data sets that have been used in the project

CMS Research, Statistics, Data & Systems
https://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems.html

Income and Poverty Data
https://www.census.gov/topics/income-poverty/income/data/tables.html

State Health Facts for other supplymentary tables
https://www.kff.org/statedata/

CMS government data
https://data.cms.gov/

Appendix-B: DDL Script for DataWarehouse

```sql
/*
Created: 15-04-2019
Modified: 22-04-2019
Model: Project Dim and Facts
Database: MS SQL Server 2017
*/


-- Create tables section -------------------------------------------------

-- Table Medicaid_Fact

CREATE TABLE [Medicaid_Fact]
(
 [Total_Cost] Float NULL,
 [IP_Cost] Float NULL,
 [OP_Cost] Float NULL,
 [PrescribedDrugs] Float NULL,
 [HospiceBenefits] Float NULL,
 [FederallyQualifiedHealthCenter] Float NULL,
 [RehabilativeServices] Float NULL,
 [HomeHealthServices] Float NULL,
 [State_SK] Int NULL,
 [AgeGroup_SK] Int NULL,
 [Gender_SK] Int NULL,
 [Race_SK] Int NULL,
 [Year_SK] Int NULL
)
go

-- Create indexes for table Medicaid_Fact

CREATE INDEX [IX_Relationship13] ON [Medicaid_Fact] ([State_SK])
go

CREATE INDEX [IX_Relationship21] ON [Medicaid_Fact] ([AgeGroup_SK])
go

CREATE INDEX [IX_Relationship27] ON [Medicaid_Fact] ([Year_SK])
go

CREATE INDEX [IX_Relationship31] ON [Medicaid_Fact] ([Gender_SK])
go

CREATE INDEX [IX_Relationship32] ON [Medicaid_Fact] ([Race_SK])
go

-- Table Medicare_Fact

CREATE TABLE [Medicare_Fact]
(
 [Total_Cost] Float NULL,
 [IP_Cost] Float NULL,
 [OP_Cost] Float NULL,
 [PrescribedDrugs] Float NULL,
 [HospiceBenefits] Float NULL,
 [FederallyQualifiedHealthCenter] Float NULL,
 [RehabilativeServices] Float NULL,
 [HomeHealthServices] Float NULL,
 [State_SK] Int NULL,
 [AgeGroup_SK] Int NULL,
 [Gender_SK] Int NULL,
 [Race_SK] Int NULL,
 [Year_SK] Int NULL
```

```
)
go

-- Create indexes for table Medicare_Fact

CREATE INDEX [IX_Relationship19] ON [Medicare_Fact] ([State_SK])
go

CREATE INDEX [IX_Relationship22] ON [Medicare_Fact] ([AgeGroup_SK])
go

CREATE INDEX [IX_Relationship24] ON [Medicare_Fact] ([Year_SK])
go

CREATE INDEX [IX_Relationship29] ON [Medicare_Fact] ([Gender_SK])
go

CREATE INDEX [IX_Relationship34] ON [Medicare_Fact] ([Race_SK])
go

-- Table HealthInsuranceDispersion_Fact

CREATE TABLE [HealthInsuranceDispersion_Fact]
(
 [Total_People] Int NULL,
 [State_SK] Int NULL,
 [CoverageType_SK] Int NULL,
 [AgeGroup_SK] Int NULL,
 [Year_SK] Int NULL
)
go

-- Create indexes for table HealthInsuranceDispersion_Fact

CREATE INDEX [IX_Relationship12] ON [HealthInsuranceDispersion_Fact] ([State_SK])
go

CREATE INDEX [IX_Relationship17] ON [HealthInsuranceDispersion_Fact] ([AgeGroup_SK])
go

CREATE INDEX [IX_Relationship25] ON [HealthInsuranceDispersion_Fact] ([Year_SK])
go

CREATE INDEX [IX_Relationship28] ON [HealthInsuranceDispersion_Fact] ([CoverageType_SK])
go

-- Table Income_Fact

CREATE TABLE [Income_Fact]
(
 [Personal_Income] Float NULL,
 [PerCapital_Income] Float NULL,
 [RegionalPriceParity] Float NULL,
 [State_SK] Int NULL,
 [Year_SK] Int NULL
)
go

-- Create indexes for table Income_Fact

CREATE INDEX [IX_Relationship14] ON [Income_Fact] ([State_SK])
go

CREATE INDEX [IX_Relationship23] ON [Income_Fact] ([Year_SK])
go
```

```sql
-- Table Poverty_Fact

CREATE TABLE [Poverty_Fact]
(
 [Poverty_Rate] Float NULL,
 [State_SK] Int NULL,
 [AgeGroup_SK] Int NULL,
 [Gender_SK] Int NULL,
 [Race_SK] Int NULL,
 [Year_SK] Int NULL
)
go

-- Create indexes for table Poverty_Fact

CREATE INDEX [IX_Relationship18] ON [Poverty_Fact] ([State_SK])
go

CREATE INDEX [IX_Relationship20] ON [Poverty_Fact] ([AgeGroup_SK])
go

CREATE INDEX [IX_Relationship26] ON [Poverty_Fact] ([Year_SK])
go

CREATE INDEX [IX_Relationship30] ON [Poverty_Fact] ([Gender_SK])
go

CREATE INDEX [IX_Relationship33] ON [Poverty_Fact] ([Race_SK])
go

-- Table Dim_State

CREATE TABLE [Dim_State]
(
 [State_SK] Int IDENTITY(1,1) NOT NULL,
 [State] Varchar(100) NOT NULL
)
go

-- Add keys for table Dim_State

ALTER TABLE [Dim_State] ADD CONSTRAINT [PK_Dim_State] PRIMARY KEY ([State_SK])
go

ALTER TABLE [Dim_State] ADD CONSTRAINT [State_SK] UNIQUE CLUSTERED ([State_SK])
go

-- Table Dim_AgeGroup

CREATE TABLE [Dim_AgeGroup]
(
 [AgeGroup_SK] Int IDENTITY(1,1) NOT NULL,
 [Age_Group] Varchar(50) NOT NULL
)
go

-- Add keys for table Dim_AgeGroup

ALTER TABLE [Dim_AgeGroup] ADD CONSTRAINT [PK_Dim_AgeGroup] PRIMARY KEY ([AgeGroup_SK])
go

ALTER TABLE [Dim_AgeGroup] ADD CONSTRAINT [State_SK] UNIQUE CLUSTERED ([AgeGroup_SK])
go

-- Table Dim_Race
```

32

```sql
CREATE TABLE [Dim_Race]
(
 [Race_SK] Int IDENTITY(1,1) NOT NULL,
 [Race] Varchar(100) NOT NULL
)
go

-- Add keys for table Dim_Race

ALTER TABLE [Dim_Race] ADD CONSTRAINT [PK_Dim_Race] PRIMARY KEY ([Race_SK])
go

ALTER TABLE [Dim_Race] ADD CONSTRAINT [State_SK] UNIQUE CLUSTERED ([Race_SK])
go

-- Table Dim_Gender

CREATE TABLE [Dim_Gender]
(
 [Gender_SK] Int IDENTITY(1,1) NOT NULL,
 [Gender] Varchar(10) NOT NULL
)
go

-- Add keys for table Dim_Gender

ALTER TABLE [Dim_Gender] ADD CONSTRAINT [PK_Dim_Gender] PRIMARY KEY ([Gender_SK])
go

ALTER TABLE [Dim_Gender] ADD CONSTRAINT [State_SK] UNIQUE CLUSTERED ([Gender_SK])
go

-- Table Dim_Year

CREATE TABLE [Dim_Year]
(
 [Year_SK] Int IDENTITY(1,1) NOT NULL,
 [Year] Varchar(4) NOT NULL
)
go

-- Add keys for table Dim_Year

ALTER TABLE [Dim_Year] ADD CONSTRAINT [PK_Dim_Year] PRIMARY KEY ([Year_SK])
go

ALTER TABLE [Dim_Year] ADD CONSTRAINT [State_SK] UNIQUE CLUSTERED ([Year_SK])
go

-- Table Dim_CoverageType

CREATE TABLE [Dim_CoverageType]
(
 [CoverageType_SK] Int IDENTITY(1,1) NOT NULL,
 [Coverage_Type] Varchar(100) NOT NULL
)
go

-- Add keys for table Dim_CoverageType

ALTER TABLE [Dim_CoverageType] ADD CONSTRAINT [PK_Dim_CoverageType] PRIMARY KEY ([CoverageType_SK])
go

ALTER TABLE [Dim_CoverageType] ADD CONSTRAINT [State_SK] UNIQUE CLUSTERED ([CoverageType_SK])
go
```

```
-- Create foreign keys (relationships) section -----------------------------------------------

ALTER TABLE [HealthInsuranceDispersion_Fact] ADD CONSTRAINT [ST_HID] FOREIGN KEY ([State_SK])
REFERENCES [Dim_State] ([State_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicaid_Fact] ADD CONSTRAINT [ST_MEDICAID] FOREIGN KEY ([State_SK]) REFERENCES
[Dim_State] ([State_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Income_Fact] ADD CONSTRAINT [ST_INCOME] FOREIGN KEY ([State_SK]) REFERENCES [Dim_State]
([State_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [HealthInsuranceDispersion_Fact] ADD CONSTRAINT [AG_HID] FOREIGN KEY ([AgeGroup_SK])
REFERENCES [Dim_AgeGroup] ([AgeGroup_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Poverty_Fact] ADD CONSTRAINT [ST_POV] FOREIGN KEY ([State_SK]) REFERENCES [Dim_State]
([State_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicare_Fact] ADD CONSTRAINT [ST_MEDICARE] FOREIGN KEY ([State_SK]) REFERENCES
[Dim_State] ([State_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Poverty_Fact] ADD CONSTRAINT [AG_POV] FOREIGN KEY ([AgeGroup_SK]) REFERENCES
[Dim_AgeGroup] ([AgeGroup_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicaid_Fact] ADD CONSTRAINT [AG_MEDICAID] FOREIGN KEY ([AgeGroup_SK]) REFERENCES
[Dim_AgeGroup] ([AgeGroup_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicare_Fact] ADD CONSTRAINT [AG_MEDICARE] FOREIGN KEY ([AgeGroup_SK]) REFERENCES
[Dim_AgeGroup] ([AgeGroup_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Income_Fact] ADD CONSTRAINT [YR_INC] FOREIGN KEY ([Year_SK]) REFERENCES [Dim_Year]
([Year_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicare_Fact] ADD CONSTRAINT [YR_MEDICARE] FOREIGN KEY ([Year_SK]) REFERENCES [Dim_Year]
([Year_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [HealthInsuranceDispersion_Fact] ADD CONSTRAINT [YR_HID] FOREIGN KEY ([Year_SK]) REFERENCES
[Dim_Year] ([Year_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Poverty_Fact] ADD CONSTRAINT [YR_POV] FOREIGN KEY ([Year_SK]) REFERENCES [Dim_Year]
([Year_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go
```

```
ALTER TABLE [Medicaid_Fact] ADD CONSTRAINT [YR_MEDICAID] FOREIGN KEY ([Year_SK]) REFERENCES [Dim_Year]
([Year_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [HealthInsuranceDispersion_Fact] ADD CONSTRAINT [CT_HID] FOREIGN KEY ([CoverageType_SK])
REFERENCES [Dim_CoverageType] ([CoverageType_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicare_Fact] ADD CONSTRAINT [GEN_MEDICARE] FOREIGN KEY ([Gender_SK]) REFERENCES
[Dim_Gender] ([Gender_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Poverty_Fact] ADD CONSTRAINT [GEN_POV] FOREIGN KEY ([Gender_SK]) REFERENCES [Dim_Gender]
([Gender_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicaid_Fact] ADD CONSTRAINT [GEN_MEDICAID] FOREIGN KEY ([Gender_SK]) REFERENCES
[Dim_Gender] ([Gender_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicaid_Fact] ADD CONSTRAINT [R_MEDICAID] FOREIGN KEY ([Race_SK]) REFERENCES [Dim_Race]
([Race_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Poverty_Fact] ADD CONSTRAINT [R_POV] FOREIGN KEY ([Race_SK]) REFERENCES [Dim_Race]
([Race_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go


ALTER TABLE [Medicare_Fact] ADD CONSTRAINT [R_MEDICARE] FOREIGN KEY ([Race_SK]) REFERENCES [Dim_Race]
([Race_SK]) ON UPDATE NO ACTION ON DELETE NO ACTION
go
```