Data Warehousing & Business Intelligence

CMS Medicare & Medicaid in United States of America

Professor: Dr. Le Thi Nhan

20C14001 - Le Duong Tuan Anh
20C14002 - Vo Tien Dat

01

INTRODUCTION

Analyze the problems and hypotheses

02

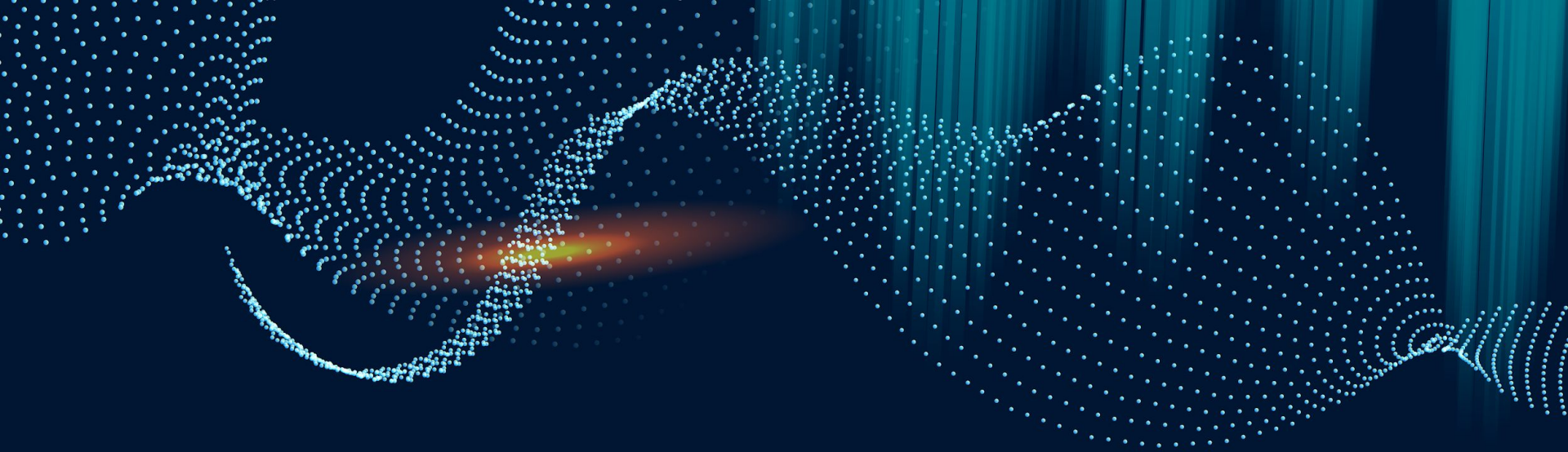
SOLUTION

ETL Process
Data warehouse architecture

03

CONCLUSION

Report visualization
Dashboard and Conclusion



01 INTRODUCTION

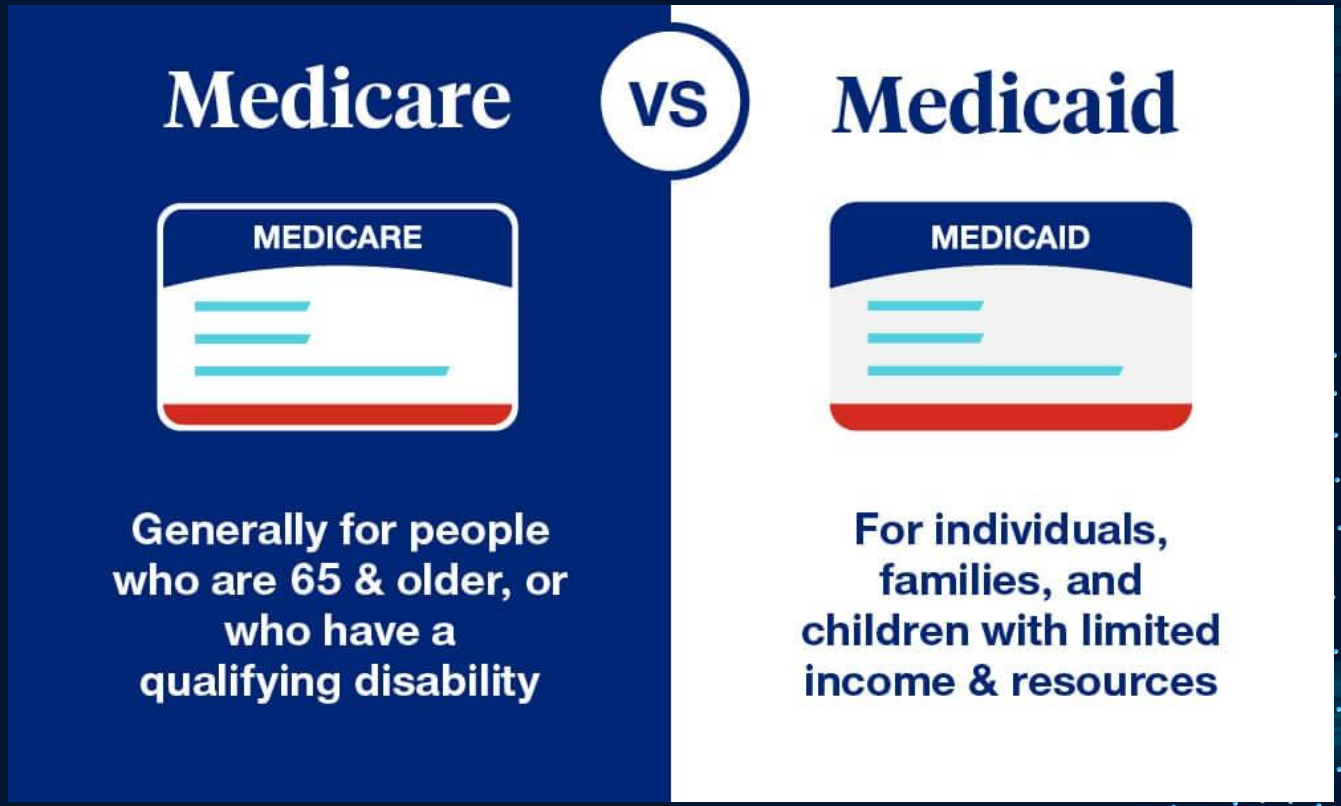
Analyze the problems

Introduction

The healthcare industry is seeing an increase in Data. We want to visually understand the US Government spending on **Medicare** and **Medicaid** across the 50 states and provide recommendations for resource allocation.

- Healthcare in the United States is expensive! An accident causes not just physical, emotional and psychological damage, but economic too.
- Nearly 80% of all data is unstructured, and medical records can be some of the hardest to work with.
- Approximately 15% of the population is over 65 years of age.
- Approximately 12% of the population falls below the poverty line.

What is Medicare / Medicaid?



Hypotheses

1. States where the number of people whose age is >65 is high.

— Medicare spending is also high

3. Government health spending is uniform between 2 genders.

— Male / Female

2. States which have high levels of poverty.

— Medicaid spending is high

4. States which have high income should have lower Medicaid spending

— People is able to afford Private Insurance.

Our Goals

Test the Efficiency

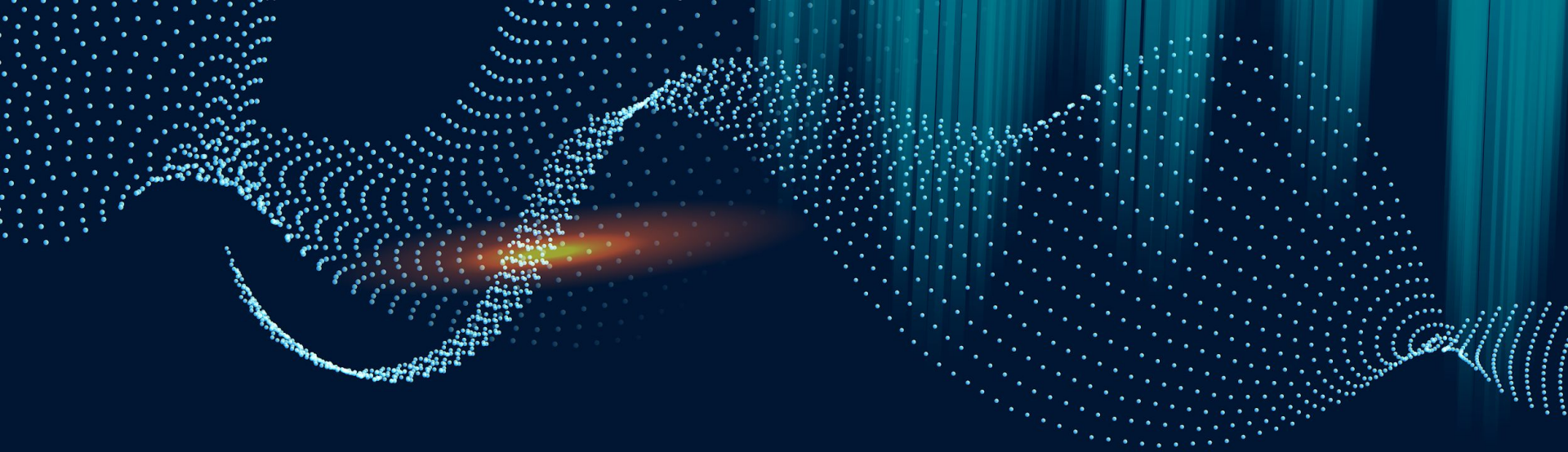
Test the dispersion of funds of 2 government health programs.

Visually understand

Understand government spending on Medicare & Medicaid.

Provide Recommendations

Understand and provide recommendations for resource allocation.



02 SOLUTION

ELT Process
Data Warehouse Architecture

The Flowchart

Understand the Data

- Collect health insurance and related data that could help solving business problems.
- Investigate data description for the selected datasets.
- Understand features of dataset their relations.

ETL

- Create tables (MSSQL) and define their relationships (PK-FK).
- Use Visual Studio to:
  - Extract the data from the cleaned CSV files, process (aggregate, lookup) and load it into tables via SSIS.
  - Create cubes, using **Galaxy schema**.

Business Intelligence

- Use MDX to query data in cubes.
- Use Power BI for visualizations.

Data Selection

Health Insurance Data

aug_2021__health_insurance_summary.csv

	uid	age	state	coverage	month	year
1	356a192b7913b04c5…	26	Alabama	Employment Based	8	2021
2	da4b9237bacccdf19c…	38	Alabama	Employment Based	8	2021
3	77de68daecd823bab…	30	Alabama	Employment Based	8	2021
4	1b6453892473a467d…	35	Alabama	Employment Based	8	2021
5	ac3478d69a3c81fa62…	22	Alabama	Employment Based	8	2021
6	c1dfd96eea8cc2b627…	19	Alabama	Employment Based	8	2021
7	902ba3cda18838015…	38	Alabama	Employment Based	8	2021
8	fe5dbbcea5ce7e2988…	26	Alabama	Employment Based	8	2021
9	0ade7c2cf97f75d009…	21	Alabama	Employment Based	8	2021

sep_2021__health_insurance_summary.csv

	uid	age	state	coverage	month	year
1	b0893feab4f95bfa705…	26	Alabama	Employment Based	9	2021
2	8bf40ad328a32b07f9…	29	Alabama	Employment Based	9	2021
3	fd9ecb15b957dbd844…	23	Alabama	Employment Based	9	2021
4	5c35bac30bfd01ff148…	55	Alabama	Employment Based	9	2021
5	2870f5069de61c181f…	58	Alabama	Employment Based	9	2021
6	f2237fd399760dba92…	26	Alabama	Employment Based	9	2021
7	b6710b8fbbefd671ab…	64	Alabama	Employment Based	9	2021
8	fe24fa600c156360c9…	50	Alabama	Employment Based	9	2021
9	07ebf9c73261d095b6…	40	Alabama	Employment Based	9	2021

UID: User ID (encryption)

Age: User age

State: Country state

Coverage: type of insurance

Month: Statistical month

Year: Statistical year

Data Selection

Income Data

aug_2021__income_summary.csv

	State	Personal Income (Mili…	Per Capital Income (…	Regional Price Parity …	Month	Year
1	Alabama	184785	38070	86.8	8	2021
2	Alaska	41461	56202	105.6	8	2021
3	Arizona	267361	39217	96.2	8	2021
4	Arkansas	113924	38257	87.4	8	2021
5	California	2103669	53949	113.4	8	2021
6	Colorado	277732	50971	103.2	8	2021
7	Connecticut	246709	68822	108.7	8	2021
8	Delaware	45058	47727	100.4	8	2021
9	District of Columbia	49276	73505	117.0	8	2021

sep_2021__income_summary.csv

	State	Personal Income (Mili…	Per Capital Income (…	Regional Price Parity …	Month	Year
1	Alabama	189162	38918	86.6	9	2021
2	Alaska	41283	55674	105.4	9	2021
3	Arizona	280120	40546	95.9	9	2021
4	Arkansas	118698	39722	86.9	9	2021
5	California	2212691	56308	114.4	9	2021
6	Colorado	288103	52097	103.0	9	2021
7	Connecticut	247887	69094	108.7	9	2021
8	Delaware	45574	47837	100.2	9	2021

State: Country state

Personal Income:

Per Capital Income

Regional Price Parity

Month: Statistical month

Year: Statistical year

Data Selection

Poverty Data

aug_2021__poverty_summary.csv

	State	Age Group	Gender	Race	Poverty Rate	Month	Year
1	Alabama	0-18	Male	NonHispanic_White	0.002431	8	2021
2	Alabama	0-18	Male	AfricanAmerican	0.006188	8	2021
3	Alabama	0-18	Male	Hispanic	0.006851	8	2021
4	Alabama	0-18	Male	Asian	0.001989	8	2021
5	Alabama	0-18	Male	AmericanIndian	0.0	8	2021
6	Alabama	0-18	Male	OtherPacific_Islander	0.00442	8	2021
7	Alabama	0-18	Male	TwoOrMore_Races	0.003757	8	2021
8	Alabama	0-18	Female	NonHispanic_White	0.003179	8	2021
9	Alabama	0-18	Female	AfricanAmerican	0.008092	8	2021

sep_2021__poverty_summary.csv

	State	Age Group	Gender	Race	Poverty Rate	Month	Year
1	Alabama	0-18	Male	NonHispanic_White	0.0018	9	2021
2	Alabama	0-18	Male	AfricanAmerican	0.00468	9	2021
3	Alabama	0-18	Male	Hispanic	0.00504	9	2021
4	Alabama	0-18	Male	Asian	0.0018	9	2021
5	Alabama	0-18	Male	AmericanIndian	0.00216	9	2021
6	Alabama	0-18	Male	OtherPacific_Islander	0.00396	9	2021
7	Alabama	0-18	Male	TwoOrMore_Races	0.0027	9	2021
8	Alabama	0-18	Female	NonHispanic_White	0.0024	9	2021

State: Country state

Age Group

Gender

Race

Poverty Rate

Month: Statistical month

Year: Statistical year

Data Selection

Medicare Data

aug_2021__medicare_summary.csv / sep_2021__medicare_summary.csv

	State	Age Group	Gender	Race	Total Cost	IP Actual Cost	OP Actual Cost	Prescribed Drugs	Hospice Benefits	Federally Qualified H…	Rehabilative Services	Home Health Services	Month
1	Alabama	65+	Male	NonHispanic_White	1921209071.0	564083878.5	257153959.7	97725516.12	91126131.15	6957896.836	47393058.43	106087037.3	8
2	Alabama	65+	Male	AfricanAmerican	767833286.0	225442605.2	102774535.5	39057125.7	36419605.65	2780803.439	18941180.5	42398903.74	8
3	Alabama	65+	Male	Hispanic	115343713.8	33865928.71	15438763.63	5867151.125	5470943.557	417731.5594	2845339.14	6369152.1389999995	8
4	Ala												
5	Ala												
6	Ala												
7	Ala												
8	Ala												
9	Ala												

	State	Age G…	Gen…	Race	Total Cost	IP Actual Cost	OP Actual Cost	Prescribed Drugs	Hospice Benefits	Federally Qualified H…	Rehabilative S…	Home Health …	Month
1	Alabama	65+	Male	NonHispanic_White	312926121.4	574505301.2	263701673.1	107766433.8	94094425.37	6939541.086	49737176.62	109159653.6	9
2	Alabama	65+	Male	AfricanAmerican	125549352.0	230497754.4	105799969.7	43237061.42	37751703.43	2784219.108	19955094.41	43796036.3	9
3	Alabama	65+	Male	Hispanic	19496486.2	35793862.86	16429616.06	6714258.241	5862440.174	432359.7736	3098815.06	6801061.126	9
4	Alabama	65+	Male	Asian	5773766.606000001	10600136.25	4865531.5430000005	1988387.0159999998	1736126.2419999999	128040.735	917695.3596	2014093.166	9
5	Alabama	65+	Male	AmericanIndian	2145114.6769999997	3938245.066	1807680.122	738741.0096	645019.1245	47570.689739999994	340949.3176	748291.5584	9

State
Age Group
Gender
Race
Total Cost
IP Actual Cost
OP Actual Cost
Prescribed Drugs
Hospice Benefits
Federally Qualified Health Center
Rehabilative Services
Home Health Services
Month
Year

Data Selection

Medicaid Data

aug_2021__medicaid_summary.csv / sep_2021__medicaid_summary.csv

	State	Age Group	Gender	Race	Total Cost	IP Actual Cost	OP Actual C…	Prescribed Dr…	Hospice Benefits	Federally Qualified H…	Rehabilative Services	Home Health Services	Month	Year
1	Alabama	0-18	Male	NonHispanic_White	65300528.4	60524390.99	10870665.36	47127579.28	4340293.614	2694632.9110000003	13455830.46	5266846.583000001	8	2021
2	Alabama	0-18	Male	AfricanAmerican	26098106.69	24189268.48	4344586.35	18835078.71	1734648.227	1076941.013	5377777.295	2104955.77	8	2021
3	Alal													
4	Alal													
5	Alal													
6	Alal													

	State	Age Group	Gender	Race	Total Cost	IP Actual Cost	OP Actual Cost	Prescribed Drugs	Hospice Benefits	Federally Qualified H…	Rehabilative Services	Home Health …	Month	Year
1	Alabama	0-18	Male	NonHispanic_White	9798964.617	67150619.95	9536148.11899…	54071767.49	4858336.95	2699271.506	13594293.45	5855789.788	9	2021
2	Alabama	0-18	Male	AfricanAmerican	3931450.823…	26941556.63	3826005.99600…	21694179.25	1949217.448	1082976.988	5454177.893999999	2349406.339	9	2021
3	Alabama	0-18	Male	Hispanic	610512.7228	4183738.73399…	594138.2565	3368876.538	302692.8493	168174.8696	846976.1281	364837.9504	9	2021
4	Alabama	0-18	Male	Asian	180799.6547	1238988.952	175950.4554	997672.435	89640.65871	49803.97166	250826.8636	108044.555	9	2021
5	Alabama	0-18	Male	AmericanIndian	67172.09392	460318.8119	65370.48169	370663.0227	33303.99473	18503.55891	93189.14725	40141.55342	9	2021
6	Alabama	0-18	Male	OtherPacific_Islan…	3766.659472	25812.26983	3665.634487	20784.8424	1867.513723	1037.582743	5225.5596590000005	2250.928229	9	2021
7	Alabama	0-18	Male	TwoOrMore_Races	254563.4027	1744479.236	247735.7974	1404708.9319999998	126212.8024	70123.30037000001	353160.7403	152125.2328	9	2021

State
Age Group
Gender
Race
Total Cost
IP Actual Cost
OP Actual Cost
Prescribed Drugs
Hospice Benefits
Federally Qualified Health Center
Rehabilative Services
Home Health Services
Month
Year

Data Selection

External Data

state.csv

	State	Abbreviation
1	Alabama	AL
2	Alaska	AK
3	Arizona	AZ
4	Arkansas	AR
5	California	CA
6	Colorado	CO
7	Connecticut	CT

race.csv

ID	Race
1	NonHispanic_White
2	AfricanAmerican
3	Hispanic
4	Asian
5	AmericanIndian
6	OtherPacific_Islander
7	TwoOrMore_Races

coverage.csv

ID	Coverage Type
1	Employment Based
2	Direct Purchase
3	Covered by TRICARE
4	Medicaid
5	Medicare
6	Uninsured

age_group.csv

ID	Age Group
1	0-18
2	19-64
3	65+

gender.csv

ID	Gender
1	Male
2	Female

datetime.csv

id	month	year
1	8	2021
2	9	2021

Table Overview

Health Insurance
- Age Group
- State
- Type of Coverage
- Year

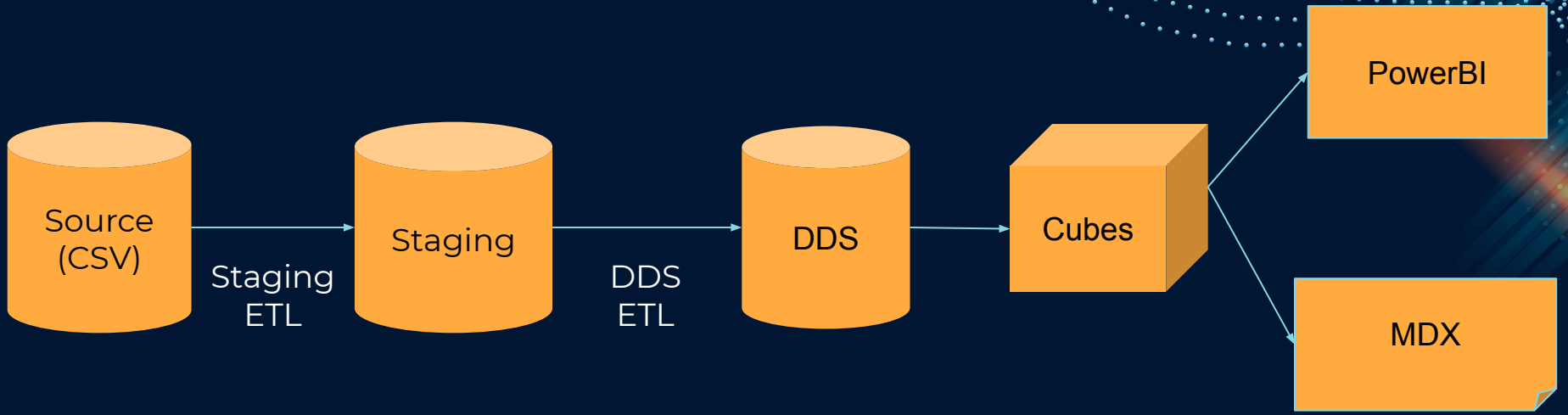
Poverty
- State
- Age Group
- Race
- Poverty Rate
- Year

Income
- State
- Personal Income (Millions of Dollars)
- Per Capital Income (Dollars)
- Regional Price Parity
- Year

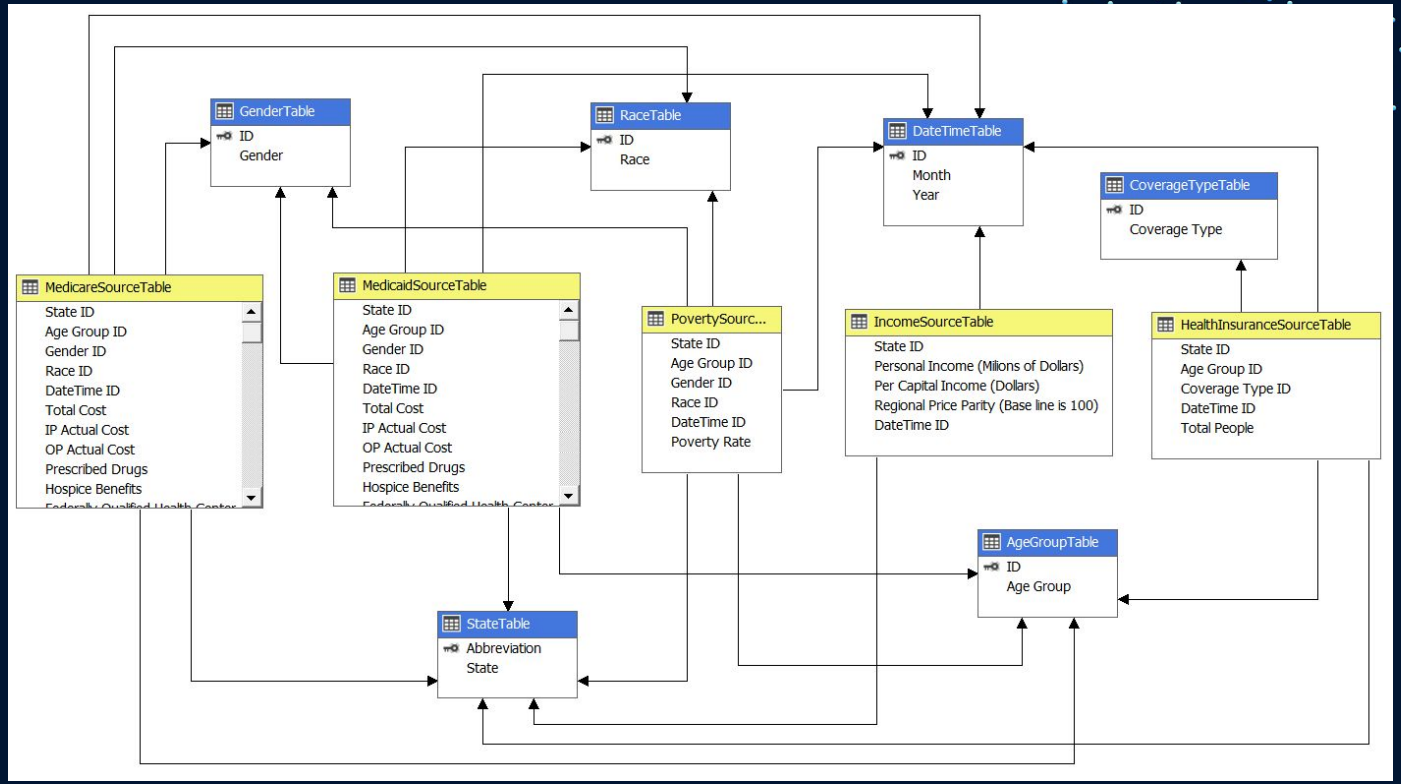
Medicare
- State
- Age Group
- Gender
- Race
- Total Cost
- IP ACtual Cost
- OP Actual Cost
- Prescribed Drugs
- Hospice Benefits
- Federally Qualified Health Centre
- Rehabilitative Services
- Home Health Services
- Yeara

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
- Yeara

Warehouse Overview

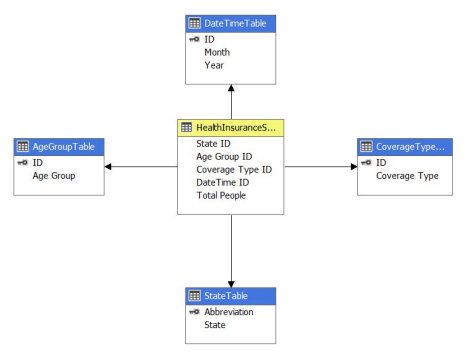


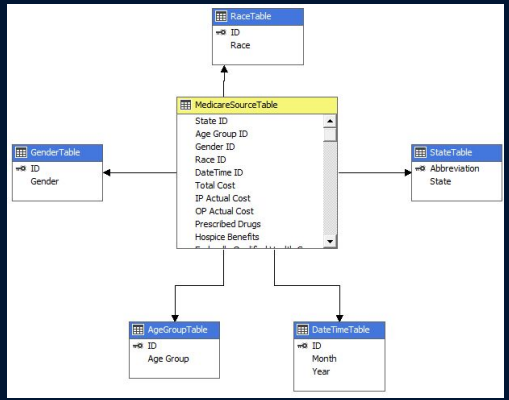
Warehouse Relationships

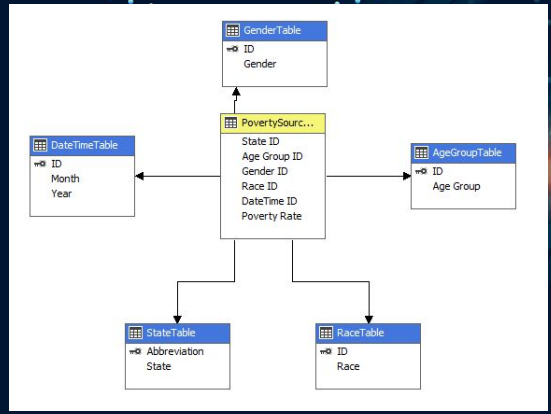


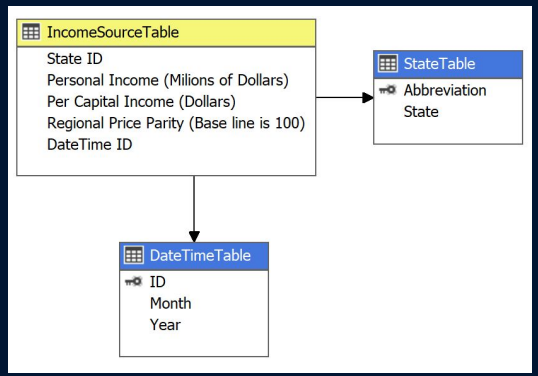
Warehouse Relationships

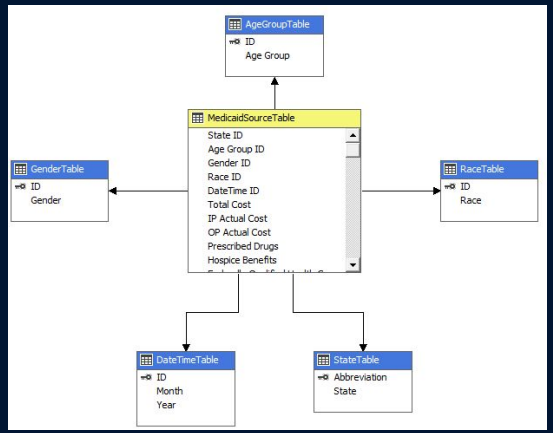
Split to Star Schema







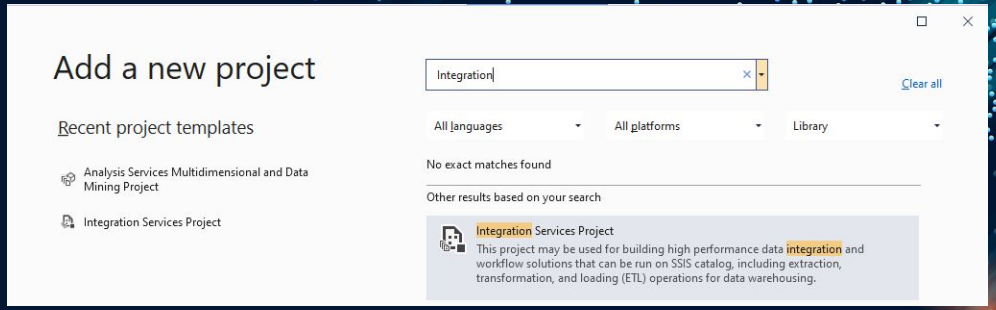




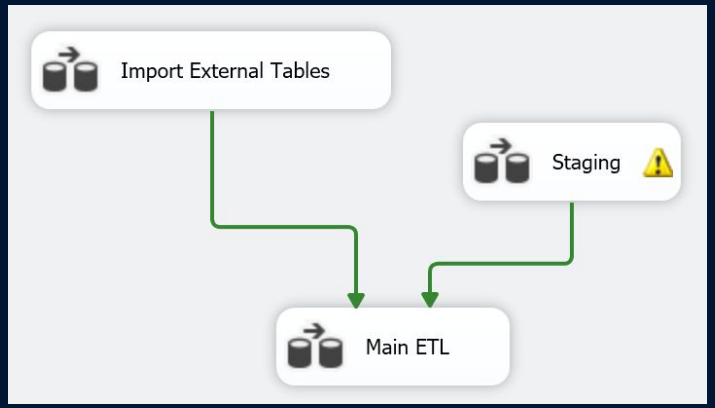
ETL

Tool: Integration Service Project

(VS2019 Community Version)

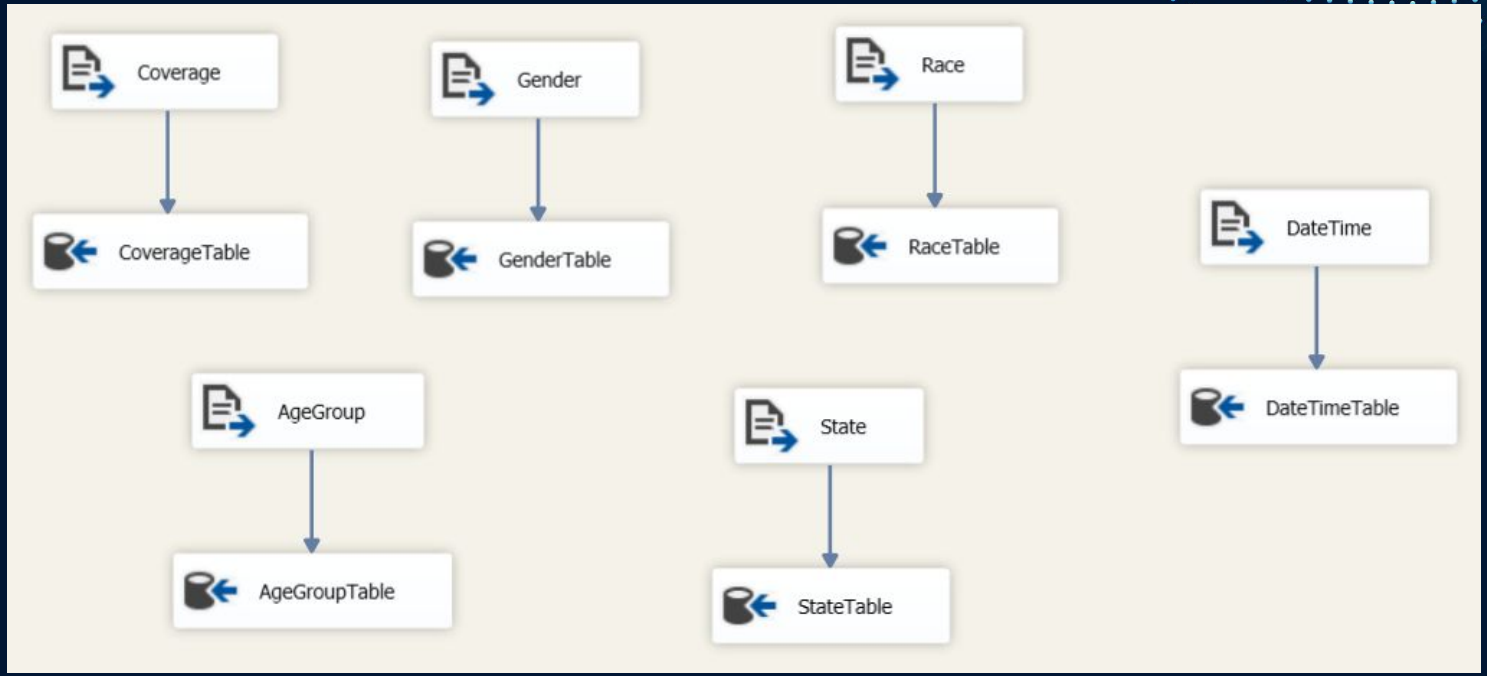


Main flow:



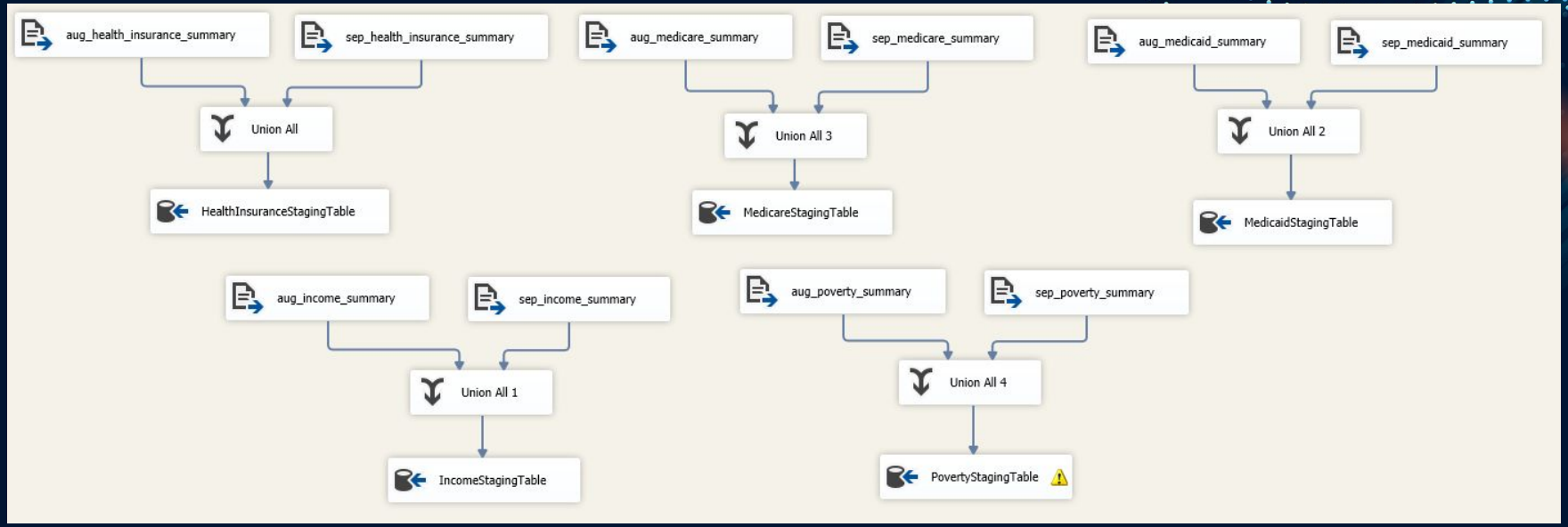
ETL

Import External Tables



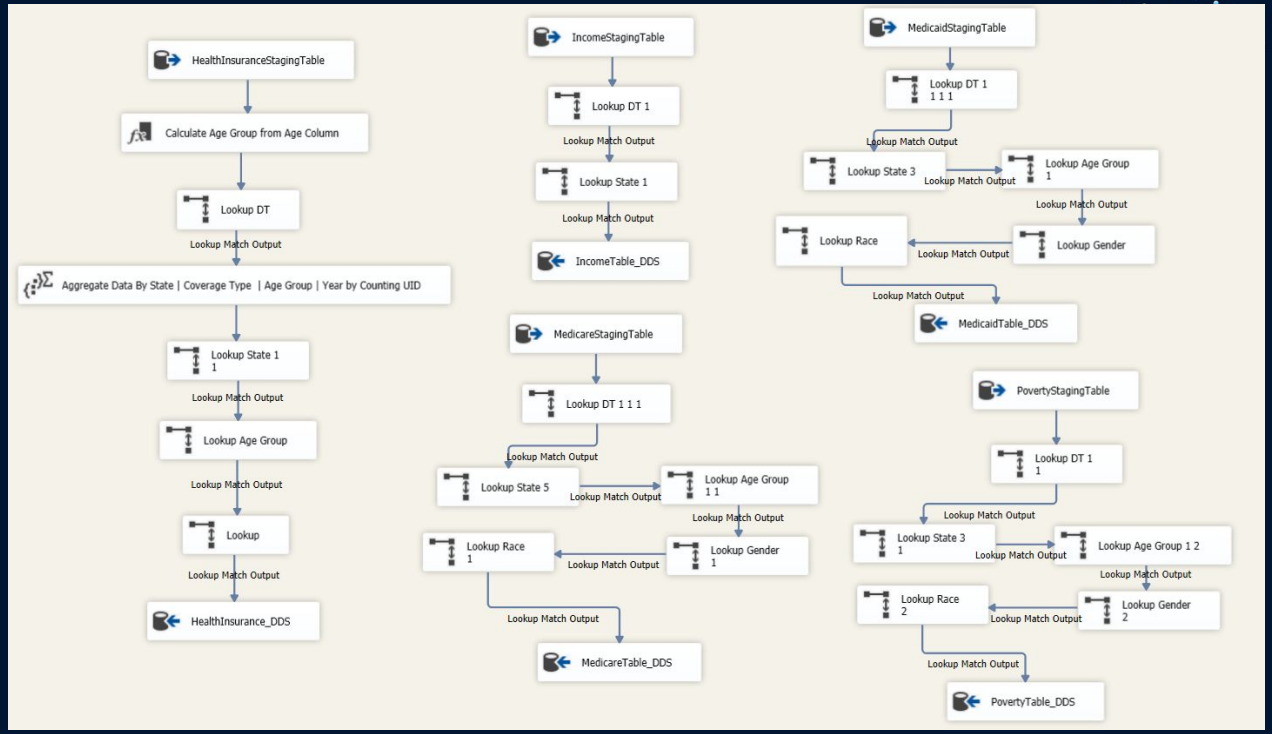
ETL

Staging


aug_health_insurance_summary
sep_health_insurance_summary
Union All
HealthInsuranceStagingTable
aug_medicare_summary
sep_medicare_summary
Union All 3
MedicareStagingTable
aug_medicaid_summary
sep_medicaid_summary
Union All 2
MedicaidStagingTable
aug_income_summary
sep_income_summary
Union All 1
IncomeStagingTable
aug_poverty_summary
sep_poverty_summary
Union All 4
PovertyStagingTable

ETL

DDS


HealthInsuranceStagingTable
Calculate Age Group from Age Column
Lookup DT
Lookup Match Output
Aggregate Data By State | Coverage Type | Age Group | Year by Counting UID
Lookup State 1 1
Lookup Match Output
Lookup Age Group
Lookup Match Output
Lookup
Lookup Match Output
HealthInsurance_DDS
IncomeStagingTable
Lookup DT 1
Lookup Match Output
Lookup State 1
Lookup Match Output
IncomeTable_DDS
MedicareStagingTable
Lookup DT 1 1 1
Lookup Match Output
Lookup State 5
Lookup Match Output
Lookup Age Group 1 1
Lookup Match Output
Lookup Race 1
Lookup Match Output
Lookup Gender 1
Lookup Match Output
MedicareTable_DDS
MedicaidStagingTable
Lookup DT 1 1 1 1
Lookup Match Output
Lookup State 3
Lookup Match Output
Lookup Age Group 1
Lookup Match Output
Lookup Race
Lookup Match Output
Lookup Gender
Lookup Match Output
MedicaidTable_DDS
PovertyStagingTable
Lookup DT 1 1
Lookup Match Output
Lookup State 3 1
Lookup Match Output
Lookup Age Group 1 2
Lookup Match Output
Lookup Race 2
Lookup Match Output
Lookup Gender 2
Lookup Match Output
PovertyTable_DDS

SSAS

Tool:

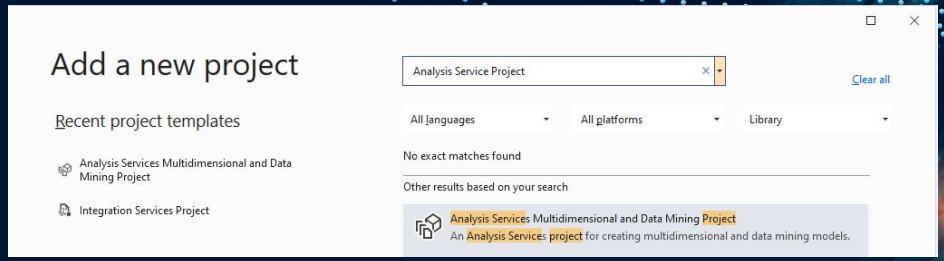
1. Analysis Services Multidimensional and Data Mining Project

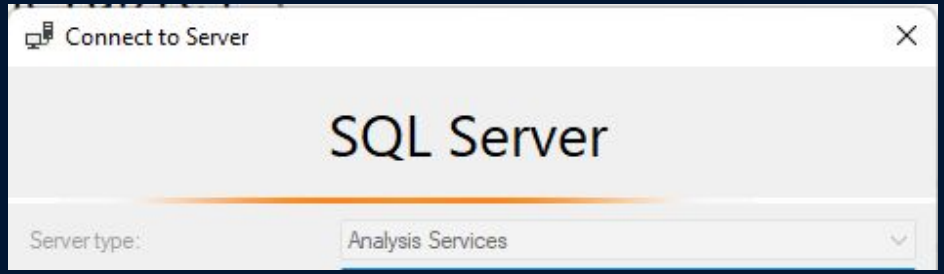
(VS2019 Community Version)

2. Microsoft Analysis Server

(to deploy cubes)

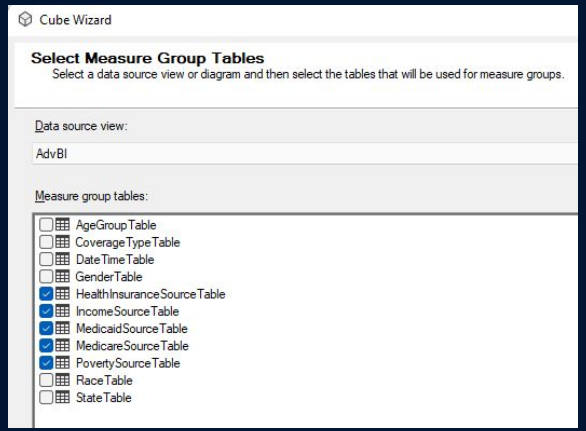
(MS SQL Server 2021)





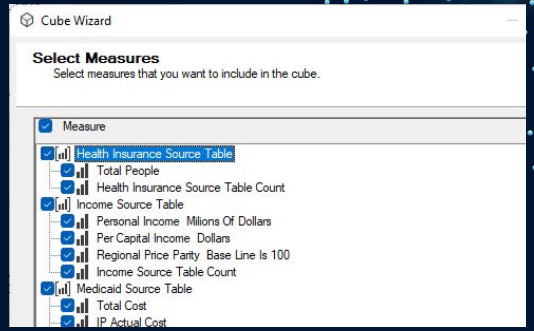
SSAS

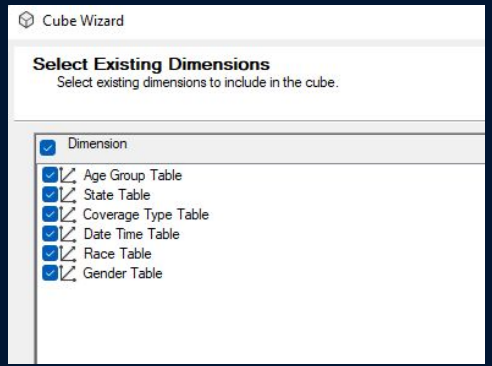
Build Cubes



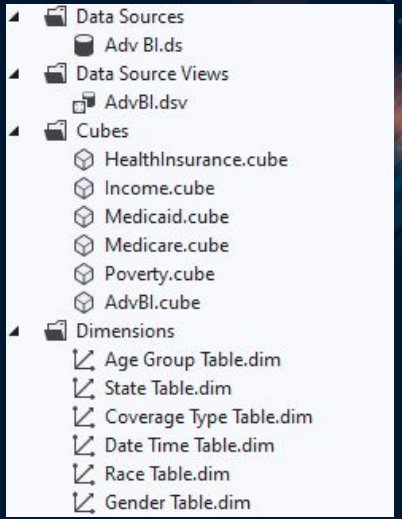
(1) Select fact tables

(2) Select measures





(3) Select dimensions

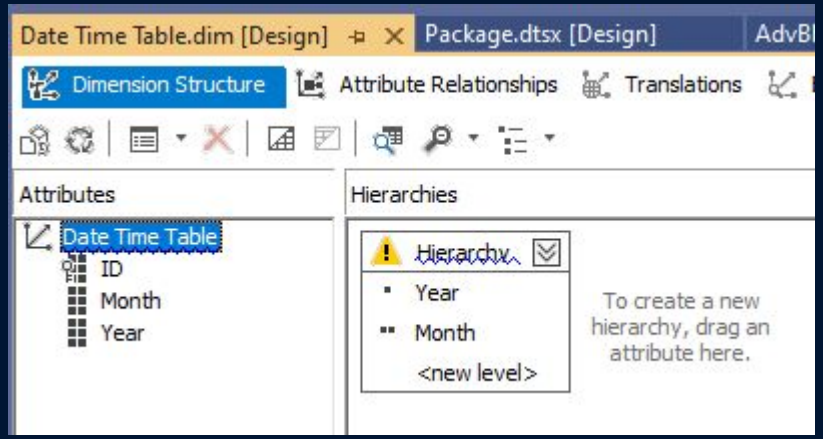


SSAS

Hierarchy on Dimension Structure

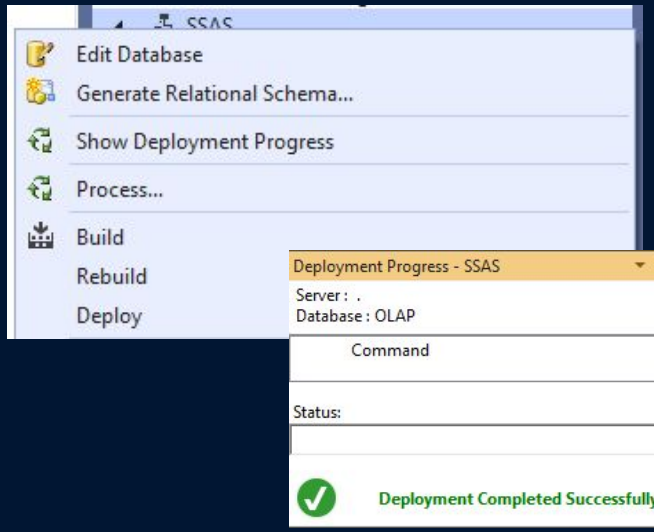
Insight: DateTime → [Year —> Month]

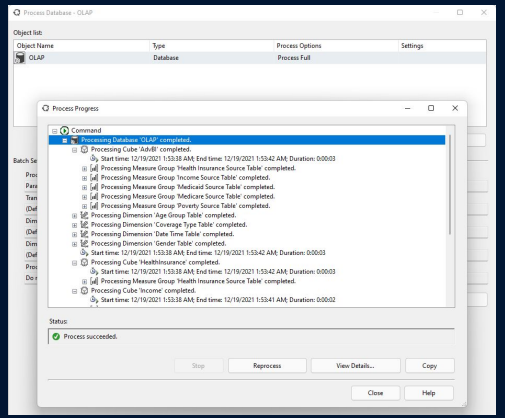
→ Edit hierarchies

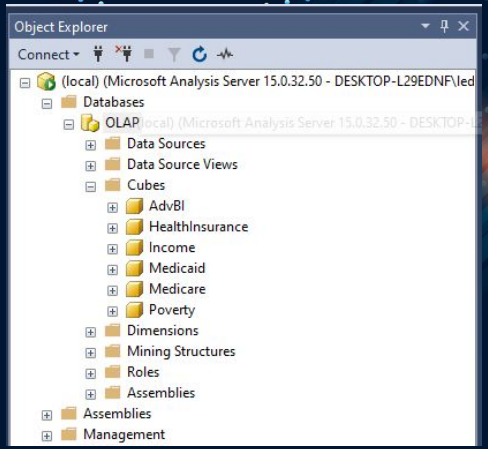


SSAS

Deploy Cubes to Microsoft Analysis Server







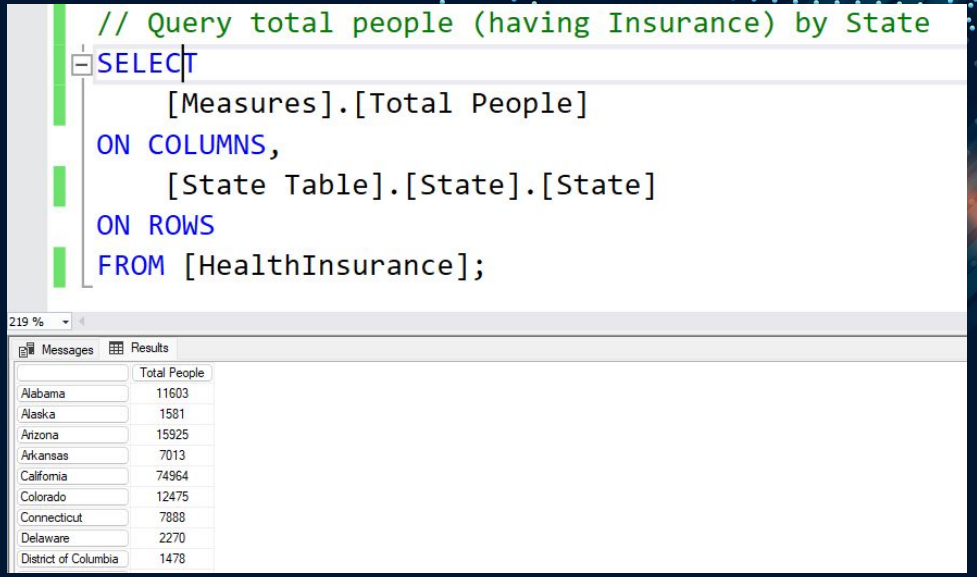
Build → Deploy → Process

SSAS

MDX Demo

Total people (having Insurance) by State

```
SELECT
    [Measures].[Total People]
ON COLUMNS,
    [State Table].[State].[State]
ON ROWS
FROM [HealthInsurance];
```

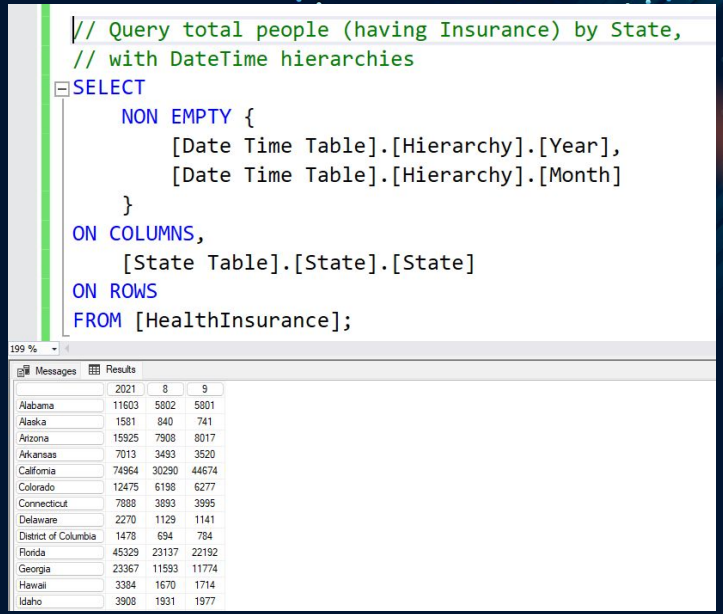


SSAS

MDX Demo - with hierarchies

Total people (having Insurance) by State, with hierarchies of DateTime Table

```
SELECT
    NON EMPTY {
        [Date Time Table].[Hierarchy].[Year],
        [Date Time Table].[Hierarchy].[Month]
    }
ON COLUMNS,
    [State Table].[State].[State]
ON ROWS
FROM [HealthInsurance];
```



	2021	8	9
Alabama	11603	5802	5801
Alaska	1581	840	741
Arizona	15925	7908	8017
Arkansas	7013	3493	3520
California	74964	30290	44674
Colorado	12475	6198	6277
Connecticut	7888	3893	3995
Delaware	2270	1129	1141
District of Columbia	1478	694	784
Florida	45329	23137	22192
Georgia	23367	11593	11774
Hawaii	3384	1670	1714
Idaho	3908	1931	1977

SSAS

MDX Demo - Complicated Query

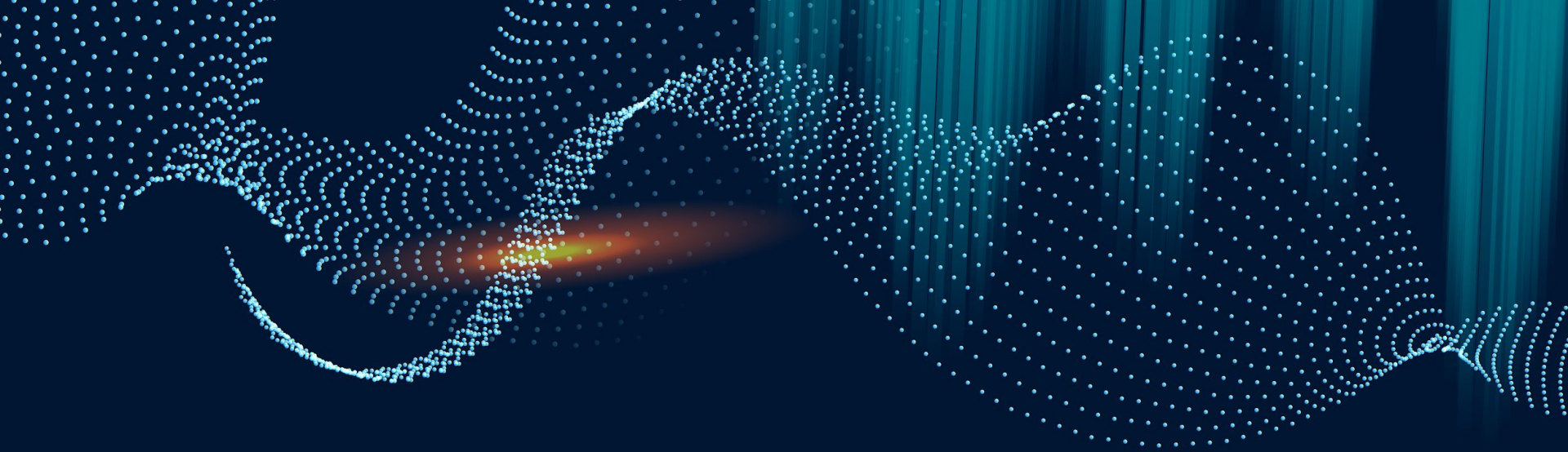
Get Avg Poverty Rate by Age Group on for every state

```
WITH
  MEMBER [Measures].[Avg Poverty Rate] AS
  AVG(EXISTING [Age Group Table].[Age Group].[Age
Group].Members,
      [Measures].[Poverty Rate])
SELECT NON EMPTY  [Measures].[Avg Poverty Rate] ON COLUMNS,

       NON EMPTY  [State Table].[State].[State] ON ROWS
FROM [Poverty];
```

Messages | Results

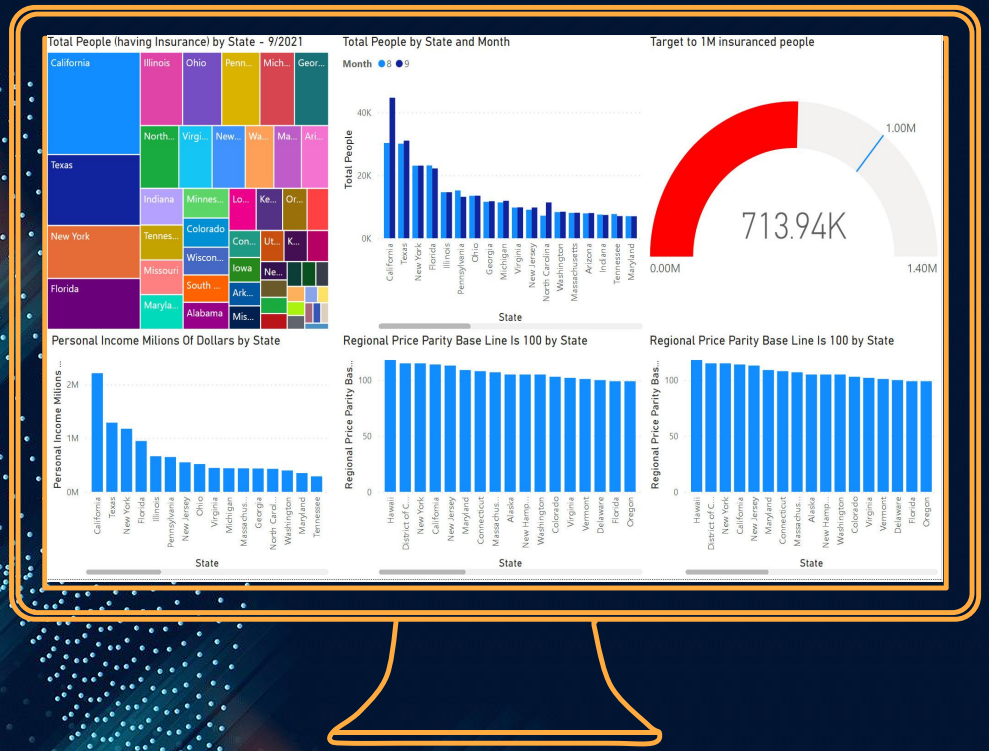
	Avg Poverty Rate
Alabama	0.110819999827072
Alaska	9.98399991658516E-03
Arizona	8.33559995517135E-02
Arkansas	0.128215999109671
California	4.27320001763292E-02
Colorado	2.17120000452269E-02
Connecticut	0.01400000002468
Delaware	0.027598000044236
District of Columbia	4.76879999623634E-02
Florida	5.33999996841885E-02
Georgia	7.90859992848709E-02
Hawaii	7.12000002386048E-03
Idaho	3.86080001480877E-02
Illinois	4.08999999053776E-02
Indiana	6.02520003449172E-02
Iowa	4.33019997435622E-02
Kansas	3.60000003129244E-02
Kentucky	0.115664000157267



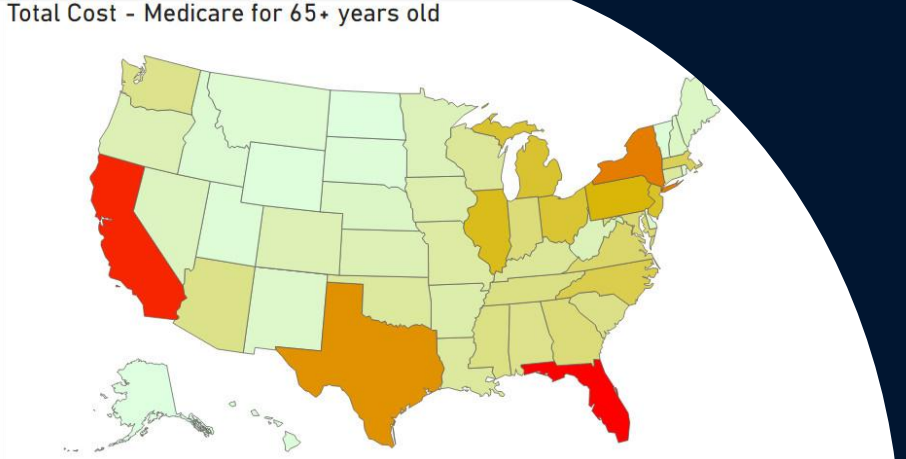
03 CONCLUSION

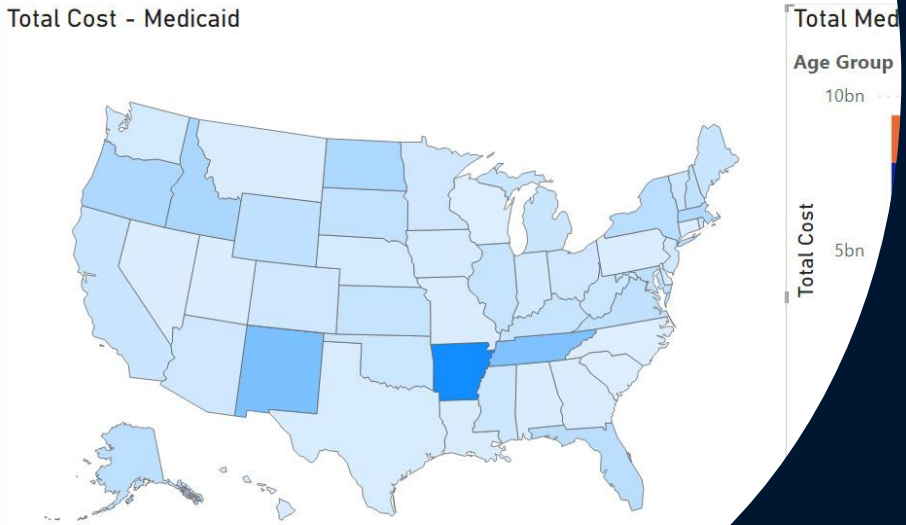
Report visualization
Dashboard

Visualization



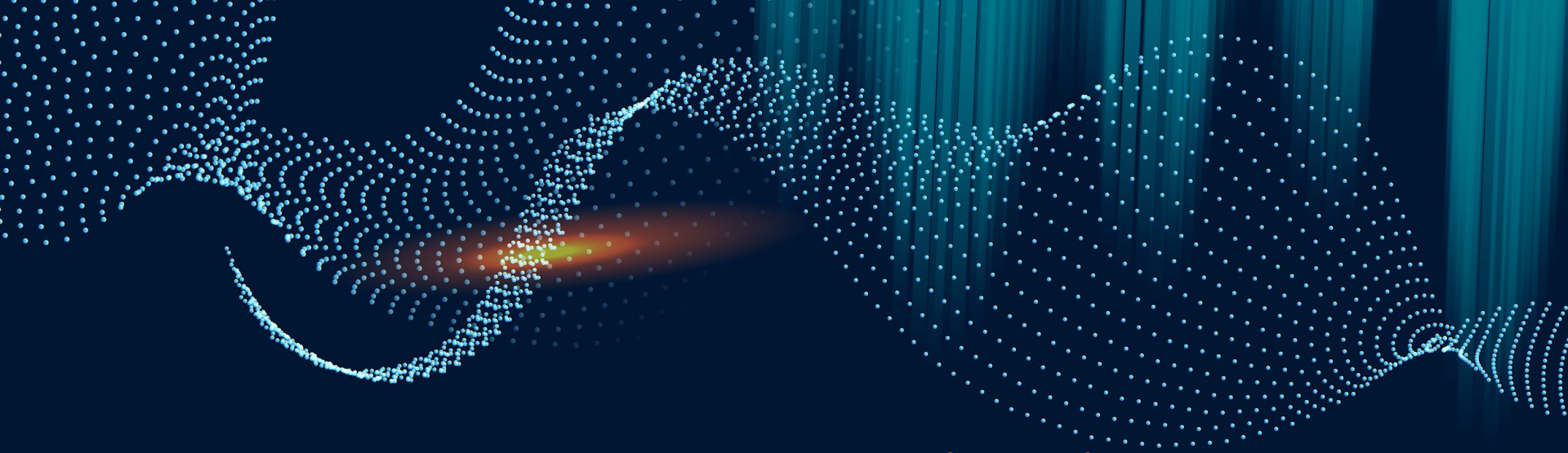
[Power BI Dashboards →]





Conclusion

- The Medicare and Medicaid spending is related to living standards and individual income across states.
- The state (like Hawaii) which have high Regional Price Parity but the number of people having insurance is low.
- Total health expenditure shows the rationality of distribution by race and state.
- The government should move the fund of spending in Medicaid to states with lower incomes and age group 19-64 to improve the health of the main labour force.



Thank you for your listening!