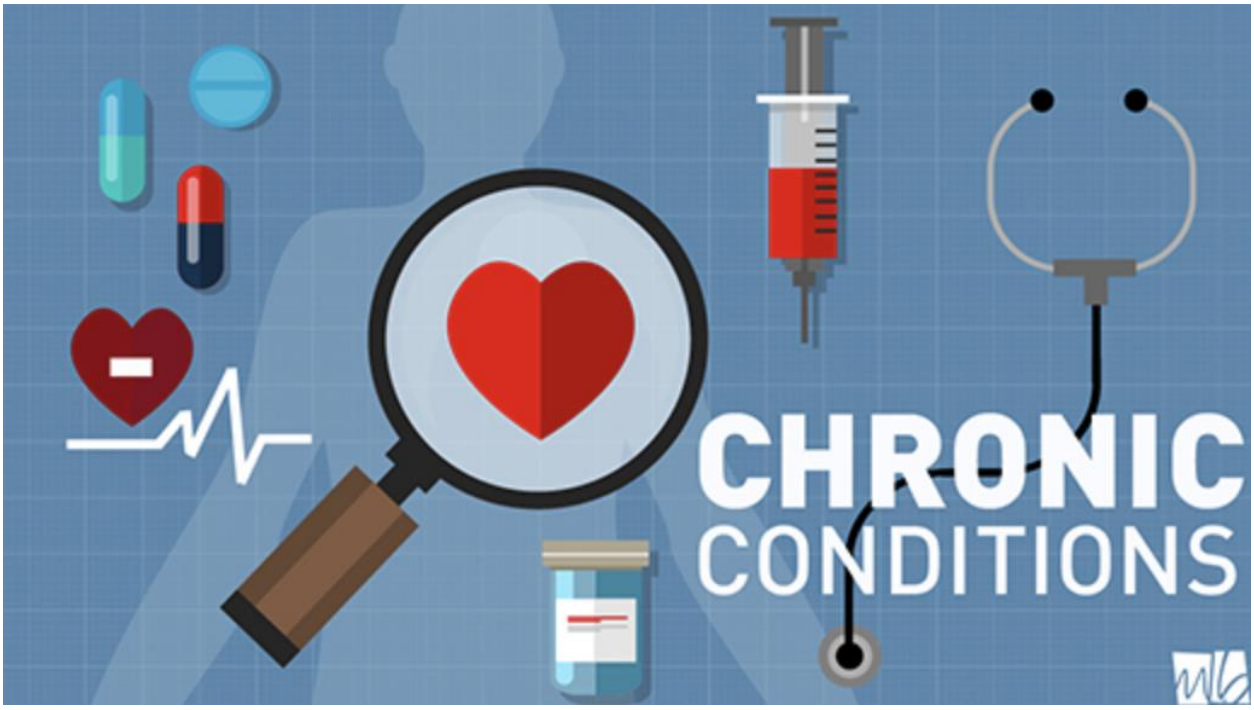
Chronic Conditions in United States

BY: Group 7 - Team STARKS

Final Project Report



Team Members

Abhinav Kaushik

Farooque Akhtar

Mayanka Verma

Acknowledgement

We would like to express appreciation to the Professor Vincent Lattuada for stimulating all the suggestions and the encouragement towards the project. He helped us to develop creative ideas towards the project and also encouraged us to complete the project successfully.

We would like to express the thanks of gratitude to our Teaching Assistant Ayush Jain for guiding us through the project who helped us solving the hindrances faced during the project. We were able to work comfortably and learn in depth under him.

Lastly, we would like to extend our thanks of appreciation to all members of our project who really worked effectively among ourselves and made this project successful with all the collaboration and the support towards each other. It was a great learning of both team work and the technical things.

Contents

INTRODUCTION

In just two words, the term “chronic condition” captures a huge swath of what ails America.

Chronic diseases are defined broadly as conditions that last 1 year or more and require ongoing medical attention or limit activities of daily living or both. Chronic diseases are the most prevalent and costly health conditions in United States. Approximately 45% of all Americans suffer from at least one chronic disease. Diseases such as cancer, obesity, diabetes, heart disease etc. are the leading causes of death, disability and reduced quality of life in the United States. They severely affect health and quality of life in US. It is a major driver of health care costs.

The chronic diseases focused in our study are:

Alcohol Abuse
Alzheimer’s Disease and Related Dementia
Arthritis (Osteoarthritis and Rheumatoid)
Asthma
Atrial Fibrillation
Autism Spectrum Disorders
Cancer (Breast, Colorectal, Lung, and Prostate)
Chronic Kidney Disease
Chronic Obstructive Pulmonary Disease
Depression
Diabetes
Drug Abuse/ Substance Abuse
Heart Failure
Hepatitis (Chronic Viral B & C)
HIV/AIDS
Hyperlipidemia (High cholesterol)
Hypertension (High blood pressure)
Ischemic Heart Disease
Osteoporosis
Schizophrenia and Other Psychotic Disorders
Stroke

The 21 chronic conditions are identified through Medicare administrative claims. A Medicare beneficiary is considered to have a chronic condition if the CMS administrative data have a claim indicating that the beneficiary received a service or treatment for the specific condition. Beneficiaries may have more than one of the chronic conditions listed above.

PROJECT OBJECTIVE

We have created a data-driven interactive dashboard that will help the decision makers to analyze large amounts of data and gain insights for making informed decisions regarding chronic diseases and identify the sectors that needs to be taken care of.

We have analyzed the spending (actual and standardized), readmissions into the hospital, emergency department visits. In addition to that we have considered the co-morbidity of chronic conditions i.e. The co-existence of 2 or more chronic conditions in a patient. We have analyzed the dataset across different states, age group, gender, medicare enrollment etc. This analysis helps us to identify the most prevalent and most expensive chronic conditions.

Also, we have identified the chronic conditions that are growing at a very high rate across years and the states that contributes the most for the chronic conditions prevalence. We also have identified the states where the prevalence is higher than the nation's average prevalence of a chronic condition.

In addition to this, we have analyzed the count of Medicare beneficiaries who have both Part A (Hospital Insurance) and Part B (Medical Insurance) coverage. Beneficiaries in Medicare Advantage represent the count of beneficiaries who are enrolled in a Medicare Advantage program. Fee-for-service (FFS) beneficiaries represent the count of beneficiaries who are enrolled in the Medicare FFS program (also known as original Medicare). Also, we have used the average age of beneficiaries to represent the average age of Medicare fee-for-service beneficiaries.

DATASETS

Datasets have been taken from https://www.cms.gov/.

We have taken multiple datasets under the heading Chronic Conditions.

1. **Chronic Conditions Prevalence Dataset for the year 2017 across Age, Gender and Medicare & Medicaid Enrollment.**

   We have 3 datasets under this title.

   i. **Prevalence -** This dataset includes all the information of prevalence of chronic diseases across various demographics and also the presence of multiple chronic diseases at the State Level. Prevalence estimates are calculated by taking the beneficiaries with a particular condition divided by the total number of beneficiaries in our fee-for-service population, expressed as a percentage. Different Datasets under this category. Information of prevalence of chronic diseases and prevalence of multiple chronic diseases.

   ii. **Mapping by Condition –** The prevalence expressed as total percentage of a chronic disease for being present as the only condition, when the patient has multiple conditions (0-6).

   iii. **Spending –** The total per capita spending on the chronic conditions on basis of state and other demographics and on basis of the presence of that condition when the patient suffers from multiple conditions.

2. **Comorbidity –** In this dataset we have the information of multiple combinations of chronic diseases when the patient suffers from more than one and the total prevalence and spending of each of this combination. The comorbidity has dataset in 2 forms.

   i. Dyads – When the patient suffers from 2 chronic conditions simultaneously. There are 210 dyads.

   ii. Triads - When the patient suffers from 3 chronic conditions simultaneously

3. **Utilization/Spending at State level for all beneficiaries. (2013 – 2017) -** This dataset includes the data at the state level for the chronic diseases mentioned within the scope of the project for the below:

   We have these 4 datasets mentioned below

a. **Per Capita Actual Spending** – It gives the information of per capita actual spending across states for different chronic diseases.
b. **Per Capita Standardized Spending –** It gives the information of per capita standardized spending across states for different chronic diseases.
c. **Emergency Department Visits –** Emergency Department visits include visits per 1000 beneficiaries where the beneficiary was released from the outpatient setting and where the beneficiary was admitted to an inpatient setting.
d. **Readmissions –** Hospital readmissions are expressed as a percentage of all admissions. A 30-day readmission is defined as an admission to an acute care hospital for any cause within 30 days of discharge from an acute care hospital. Except when the patient died during the stay, each inpatient stay is classified as an index admission, a readmission, or both.

4. **Chronic Conditions Prevalence across various Ethnicities -** We have datasets for prevalence of chronic diseases for each of the following ethnicities:
    a. Non-Hispanic White
    b. Black or African American
    c. Asian/Pacific Islander
    d. Hispanic
    e. American Indian/Alaska Native.

We are using these datasets to compare the prevalence of the chronic diseases across ethnicities.

5. **Medicare Beneficiaries Population and Demographics (2013-2016) –** It contains the information of Medicare Beneficiaries that fall under Part A and Part B type i.e. it represents the count of Medicare beneficiaries who have both Part A (Hospital Insurance) and Part B (Medical Insurance) coverage. Also, we have information about beneficiaries in Medicare Advantage. The Measures include the Medicare Advantage Participation Rate which represents the percent of Part A and Part B beneficiaries who are in enrolled in a Medicare Advantage program and is calculated by taking the count of beneficiaries in Medicare Advantage divided by the count of beneficiaries enrolled in Part A and B, expressed as a percentage. Other measure is Fee-for-service (FFS) beneficiaries represented by the count of beneficiaries who are enrolled in the Medicare FFS program.

PRE - PROCESSING OF DATASET

Most of the data preprocessing was done in Excel and MS SQL Server and some of it was done in R

Excel:

We have done the following for cleansing the data and for putting the data in a more readable format:

1. Removed the merged cells from our data files in Excel
2. Verified whether the column names for the same category of datasets was homogenous across various files
3. If they were not homogenous, due to minute differences, these were handled in Excel
4. We updated the column names to keep it homogenous for files capturing the same information for different years or for different demographics.
5. Handled the national rows table in the data file and stored them in a different file to a different file for validation of inputted rows in our fact tables

MS SQL Server:

1. Removed the special characters in the dataset.
2. Removed the records that only had null values

R:

1. Pivoted the data so that measures capturing the same kind of information were tracked in a single field instead of multiple. For example, we had data that had separate columns for each disease and separate columns for age groups. They were done in R, using the following code.

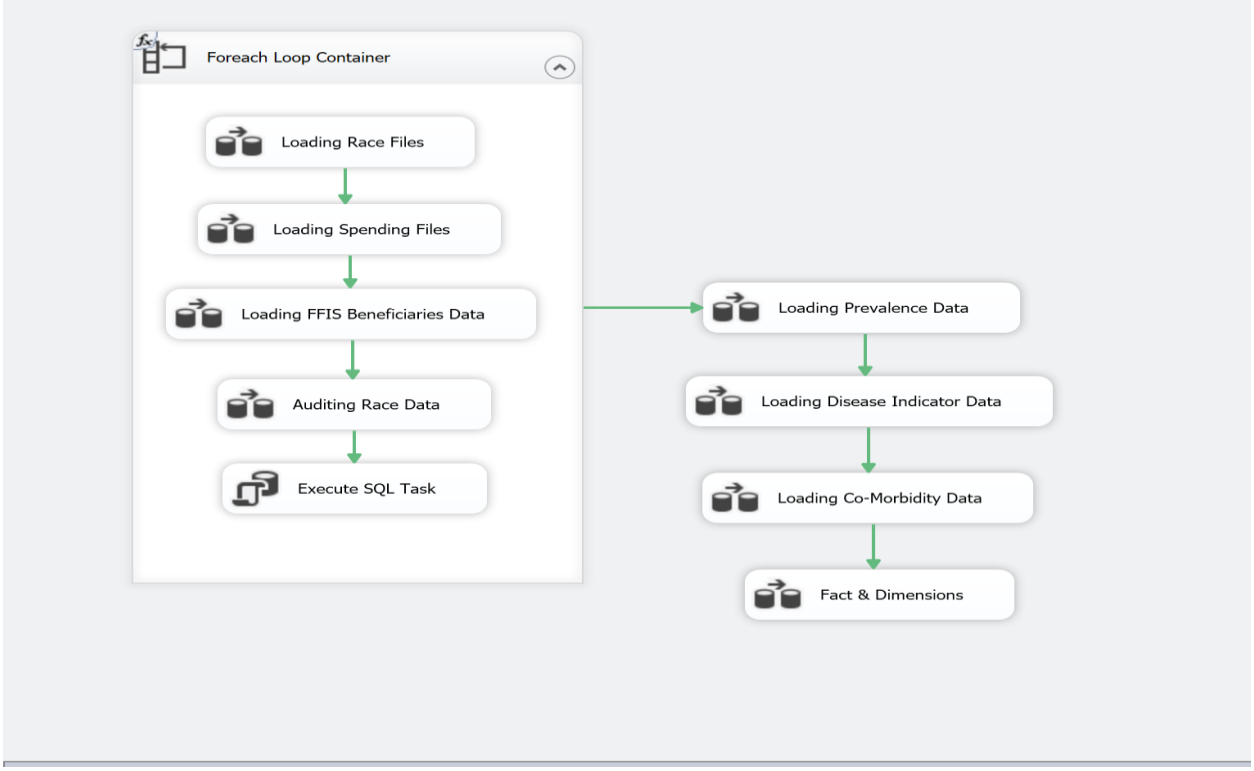
```
library (reshape2)
library(readxl)
setwd("/Users/mayankaverma/desktop/DWBI/")

race.df<-as.data.frame(read_excel("race.xlsx"))
View(race.df)
data.race<-as.data.frame(melt(race.df,id.vars = c("State","State FIPS Code","Year","Race")))
View(data.race)

data.df<-as.data.frame(read_excel("ActualSpendings.xlsx"))
data.act.sp<-as.data.frame(melt(data.df,id.vars = c("State","State FIPS Code","Year")))
colnames(data.act.sp)[5]<-"ActualSpending"
View(data.act.sp)

ED.df<-as.data.frame(read_excel("EDVisit.xlsx"))
data.ed<-as.data.frame(melt(ED.df,id.vars = c("State","State FIPS Code","Year")))
colnames(data.ed)[5]<-"EDVisit"
View(data.ed)

re.df<-as.data.frame(read_excel("Readmission.xlsx"))
readm<-as.data.frame(melt(re.df,id.vars = c("State","State FIPS Code","Year")))
colnames(readm)[5]<-"Readmission"
View(readm)

std.sp.df<-as.data.frame(read_excel("StandardizedSpending.xlsx"))
std.sp<-as.data.frame(melt(std.sp.df,id.vars = c("State","State FIPS Code","Year")))
colnames(std.sp)[5]<-"StandardizedSpending"
View(std.sp)

x<-merge(data.act.sp,data.ed, by=c("State","State FIPS Code","Year","variable"))
View(x)
```

DATA MODELING and WAREHOUSING

Data Integration and Data Warehousing:

1. We have loaded the data files using SSIS Packages into staging tables.
2. For the different data files that we have in different tabs in the same excel file, we have loaded them as different tables in SSMS using SSIS.
3. As we are using data files from 2013 – 2017, we have loaded the data files of one category for all years into one table.
4. For implementation of the above step, we have added a derived column for "year" information in SSIS.
5. For executing the above three steps, we have used foreach loop in SSIS to retrieve multiple data files from our local system and loading them in desired location in desired format.

Control Flow Task

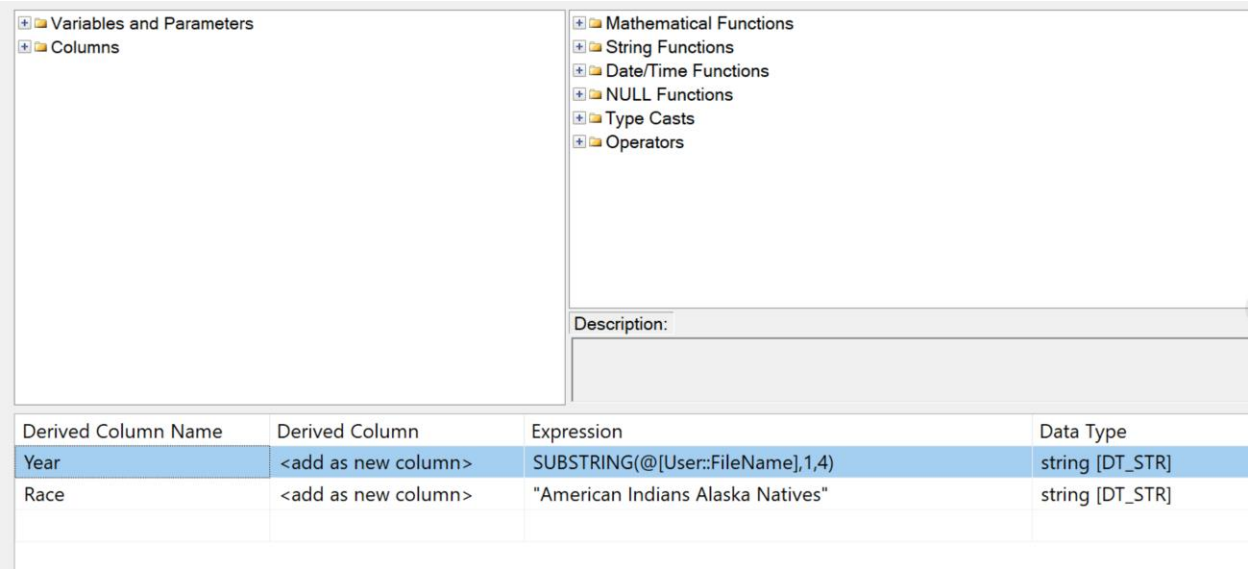


Loading the data files into staging table and adding the derived columns for demographics and Year:

Data Flow Task:

Adding a derived column for adding a year and race

Derived Column:



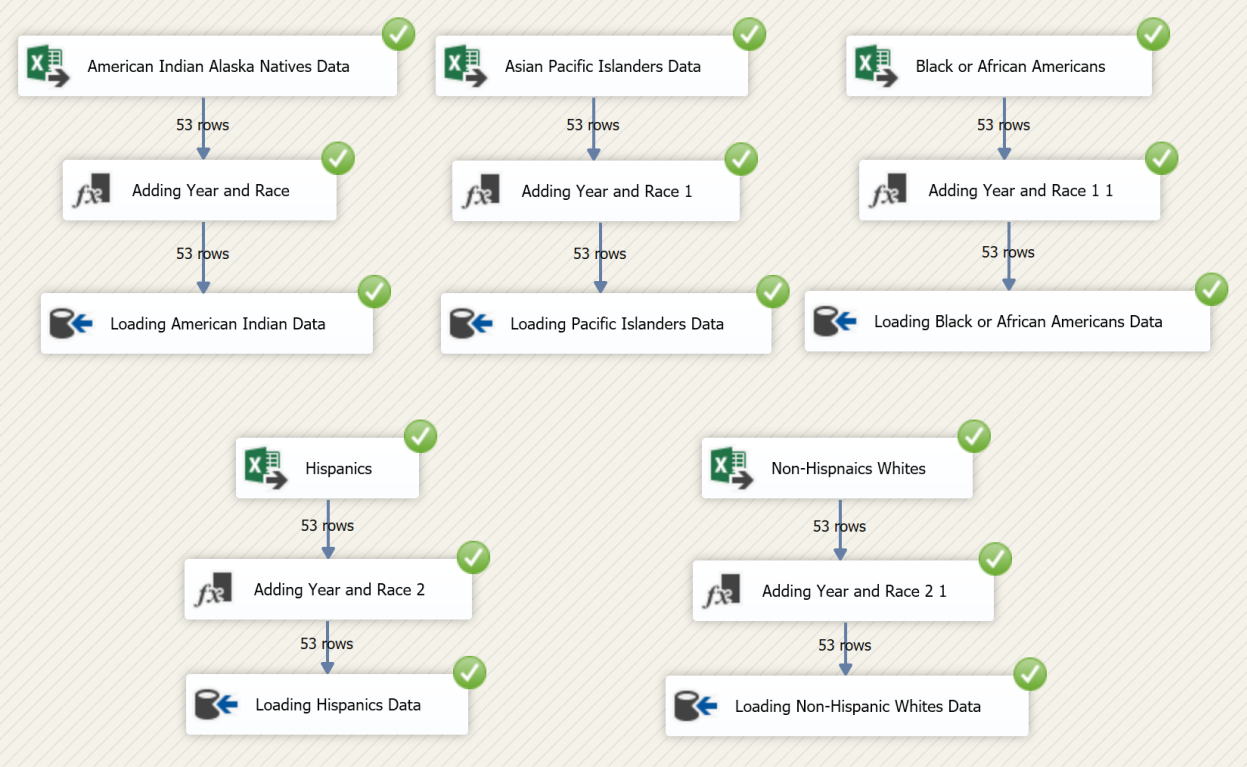
Derived Column Name	Derived Column	Expression	Data Type
Year	<add as new column>	SUBSTRING(@[User::FileName],1,4)	string [DT_STR]
Race	<add as new column>	"American Indians Alaska Natives"	string [DT_STR]

One Staging Table with the prevalence information across Races for 2013-2016:

100 % — Results | Messages

	Ischemic Heart Disease Prevalence (%)	Osteoporosis Prevalence (%)	Schizophrenia/Other Psychotic Disorders Prevalence (%)	Stroke Prevalence (%)	Year	Race
981	19.27	9.58	2.08	2.86	2015	Asian Pacific Islanders
982	18.76	2.39	8.25	3.8	2015	Black or African Americans
983	14.11	6.73	2.54	2.79	2015	Asian Pacific Islanders
984	22.73	5.98	2.51	2.51	2015	Asian Pacific Islanders
985	17.14	NULL	5.65	2.22	2015	Black or African Americans
986	15.9	5.61	3.05	2.65	2015	Asian Pacific Islanders
987	13.06	4.45	3.56	NULL	2015	Asian Pacific Islanders
988	22.6	2.35	5.13	4.33	2016	Black or African Americans
989	19.06	7.94	2.17	3.41	2016	Asian Pacific Islanders
990	31.11	6.73	2.74	4.05	2016	Non-Hispanic Whites
991	17.16	1.75	5.4	3.55	2016	Black or African Americans
992	16.3	5.42	1.7	2.59	2016	Asian Pacific Islanders
993	18.17	3.72	1.89	2.27	2016	Non-Hispanic Whites
994	18.25	9.07	1.98	3.18	2016	Asian Pacific Islanders
995	21.01	3.29	4.06	3.88	2016	Black or African Americans
996	25.52	3.71	6.5	4.87	2015	American Indians Alaska Natives
997	30.17	6.04	6.11	4.83	2015	American Indians Alaska Natives
998	17.67	3.57	2.53	2.4	2015	American Indians Alaska Natives
999	16.67	0	NULL	NULL	2015	American Indians Alaska Natives
1...	0	0	0	0	2015	American Indians Alaska Natives
1...	28.96	5.21	7.71	3.75	2015	American Indians Alaska Natives
1...	21.75	4.05	4.4	3.27	2015	American Indians Alaska Natives
1...	28.26	NULL	NULL	NULL	2015	American Indians Alaska Natives
1...	28.94	3.88	6.48	2.78	2015	American Indians Alaska Natives
1...	22.09	2.45	4.36	2.09	2015	American Indians Alaska Natives
1...	35.53	3.45	3.62	3.78	2016	American Indians Alaska Natives
1...	16.68	3.21	4.78	2.6	2016	American Indians Alaska Natives
1...	19.46	5.57	1.92	2.86	2016	American Indians Alaska Natives

Query executed successfully. | DESKTOP-FAOKAC9\SQLDEV16 (1... | DESKTOP-FAOKAC9\Abhina... | StagingTable | 00:00:00 | 1272 rows



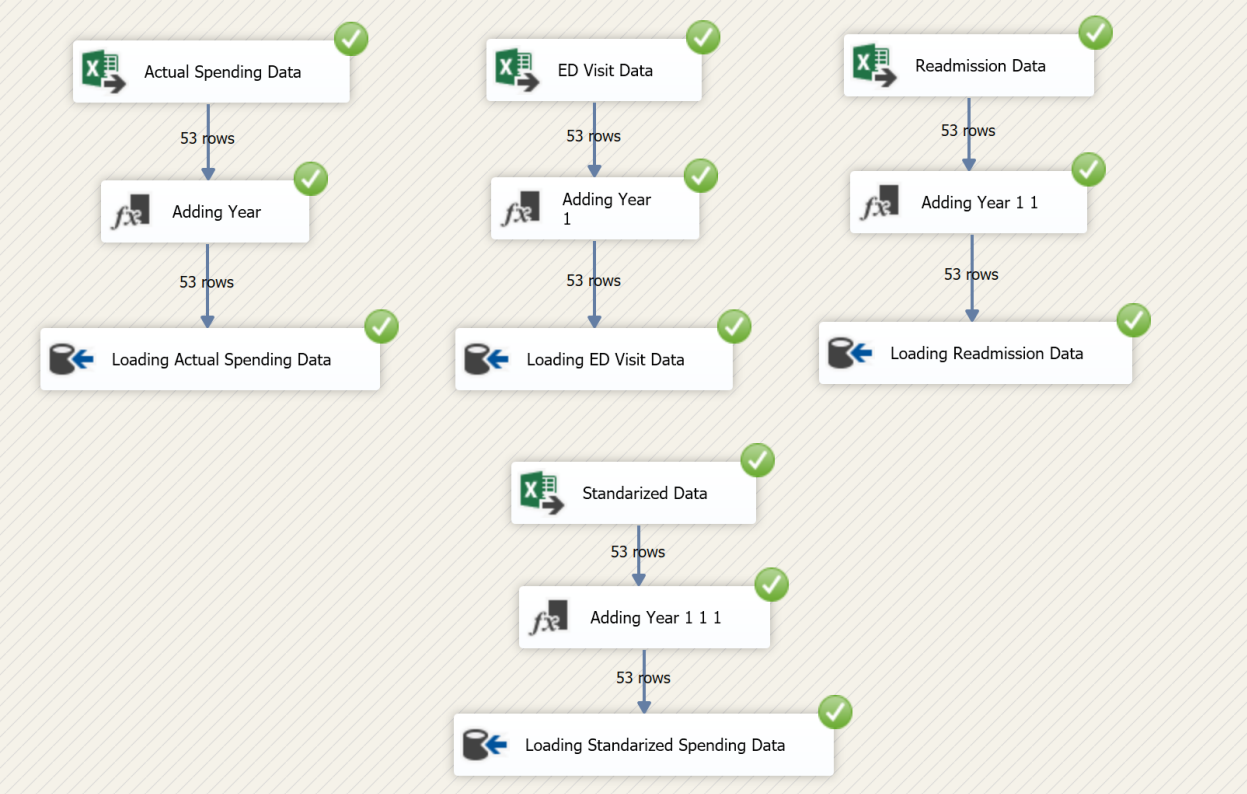

```
select * from Race
```

100 %

Results | Messages

	lence (%)	Ischemic Heart Disease Prevalence (%)	Osteoporosis Prevalence (%)	Schizophrenia/Other Psychotic Disorders Prevalence (%)	Stroke Prevalence (%)	Year	Race
1		31.45	6.14	3.11	4.1	2013	Non-Hispanic Whites
2		18.92	3.75	2.39	2.06	2013	Non-Hispanic Whites
3		26.01	6.76	1.71	3.58	2013	Non-Hispanic Whites
4		31.56	6.14	3.72	4.16	2013	Non-Hispanic Whites
5		26.66	7.14	3.62	3.63	2013	Non-Hispanic Whites
6		20.62	5.99	2.9	2.69	2013	Non-Hispanic Whites
7		28.18	7.59	4.25	3.8	2013	Non-Hispanic Whites
8		31.19	6.11	2.47	4.68	2013	Non-Hispanic Whites
9		21.3	6.48	2.94	3.02	2013	Non-Hispanic Whites
10		37.09	8.42	3.22	4.43	2013	Non-Hispanic Whites
11		28.8	5.98	3.02	3.92	2013	Non-Hispanic Whites
12		20.16	5.71	2.35	3.26	2013	Non-Hispanic Whites
13		20.7	4.78	3.15	2.35	2013	Non-Hispanic Whites
14		29.24	6.55	3.68	3.83	2013	Non-Hispanic Whites
15		29.92	6.33	4.61	3.77	2013	Non-Hispanic Whites
16		25.41	5.37	3.88	2.69	2013	Non-Hispanic Whites
17		27.17	6.17	4.01	3.24	2013	Non-Hispanic Whites
18		31.49	5.61	4.28	3.66	2013	Non-Hispanic Whites
19		34.58	6.49	3.81	4.45	2013	Non-Hispanic Whites
20		23.71	5.49	3.84	2.98	2013	Non-Hispanic Whites
21		29.35	6.96	3	4.21	2013	Non-Hispanic Whites

- Firstly, loaded the data for spending data from year 2013 to 2016 i.e. ActualSpending, EDVisit, Readmission, StandarizedSpending for different chronic conditions and per different states and using derived column for year



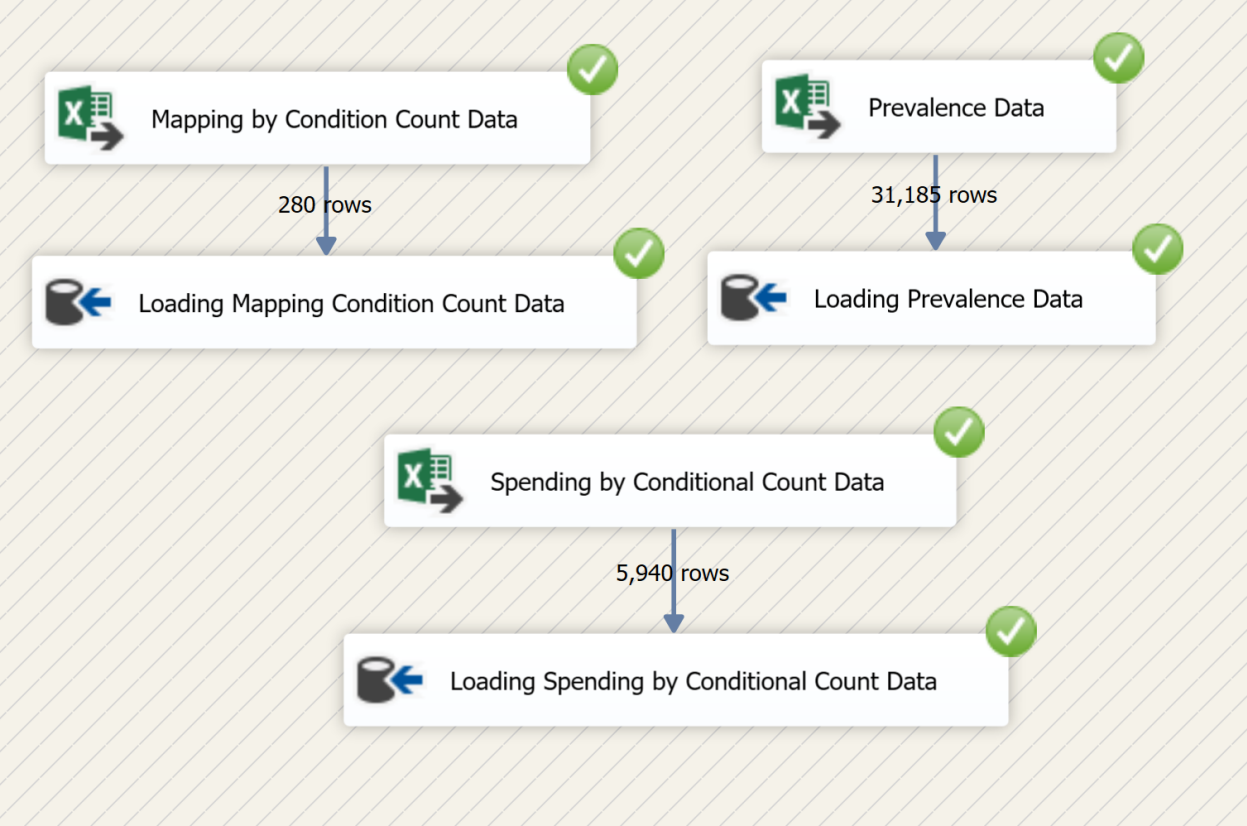

```
select * from ActualSpending
```

100 %

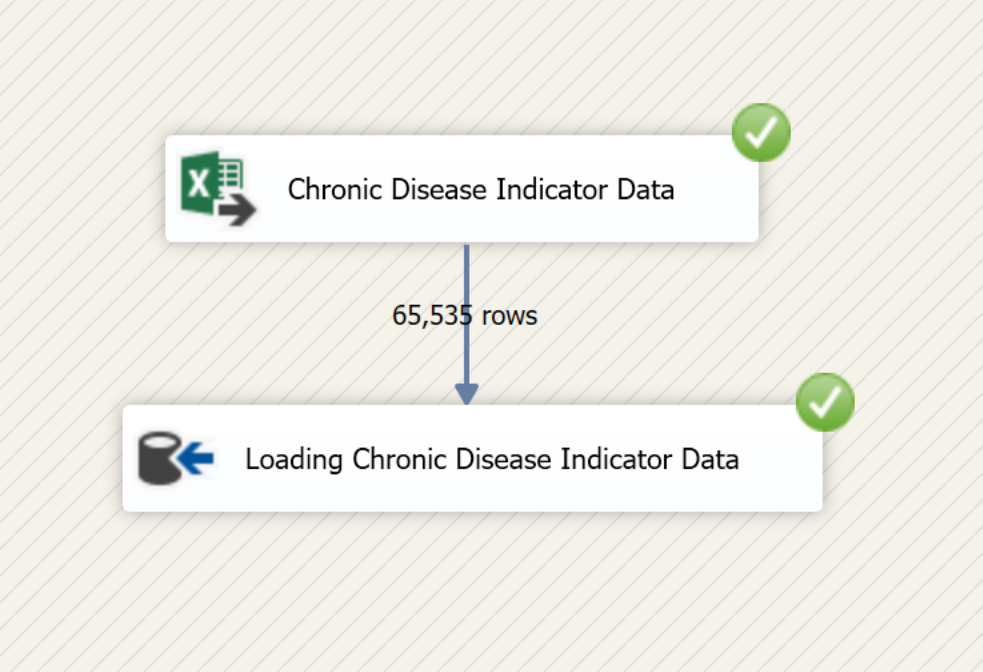
Results | Messages

	State	State FIPS Code	Alcohol Abuse	Alzheimer's Disease/Dementia	Arthritis	Asthma	Atrial Fibrillation	Autism Spectrum Disorders	Cancer
1	Alabama	01	21964.2285	19905.9487	13187.1586	18972.5819	20310.0642	5910.5606	17729.5701
2	Alaska	02	24900.2358	17489.2731	15450.1877	24812.8245	24157.3176	8483.3399	19970.1752
3	Arizona	04	25821.068	21368.5814	14656.0599	20382.5271	22872.876	8898.3676	18637.369
4	Arkansas	05	20850.4319	18861.3525	12998.0061	19882.4831	20065.066	7578.3507	17749.2127
5	California	06	31697.2057	27438.6482	17598.2984	26173.1853	28818.9543	10326.7473	21988.4095
6	Colorado	08	21887.5131	18984.4965	14419.4948	19554.7054	21252.1364	7497.3879	17599.2753
7	Connecticut	09	26698.327	23629.9767	17432.3525	23136.0443	26271.2294	8421.5311	22162.9454
8	Delaware	10	27523.9322	25112.1235	15204.9196	22307.0687	25152.3666	7224.5232	18580.9765
9	District of Columbia	11	32896.0697	26458.4134	17956.5967	27896.0464	34118.2449	10126.8693	22598.9844
10	Florida	12	27319.2068	25503.727	16906.1533	23790.6003	24920.7385	9729.1174	20105.9454
11	Georgia	13	23573.4449	21099.4979	13959.193	20084.5109	22074.6219	5983.0403	18119.169
12	Hawaii	15	19693.6412	15874.2373	12411.2117	17224.4922	20758.9204	5857.0464	14376.714
13	Idaho	16	20533.3899	16501.5743	13358.3571	19777.9672	18459.9627	5935.0951	17045.6908
14	Illinois	17	28204.5978	24492.317	16892.142	25056.051	26027.2875	9585.3987	20377.2945
15	Indiana	18	22787.6824	21645.1522	15150.88	22558.2617	23759.4408	6614.6321	19587.8255
16	Iowa	19	20118.1094	14934.8315	13134.5551	19992.2877	18899.1116	6947.3911	17135.7097
17	Kansas	20	20345.1756	18382.1533	14152.3956	20961.1143	21262.117	6138.6227	17866.2314
18	Kentucky	21	22117.1865	21122.0633	14079.5838	20860.4618	23945.414	6803.1711	19649.2417
19	Louisiana	22	26443.1044	26583.9035	16799.8098	24379.8675	25714.3878	7527.6962	19751.566
20	Maine	23	17874.359	17000.65	13695.1594	18666.1355	19590.4843	6030.2706	18697.7018
21	Maryland	24	32479.8408	25103.4317	17856.5	27461.8912	30162.7014	9414.4203	23675.6335

Stored data for prevalence for year 2017



Similarly, stored data for Chronic Disease Indicator, Co-Morbidity and FFIS Beneficiaries



Audit Table

- Audited and maintained records for the data of all the races for different years

```
select * from log_valid
```

100 %

Results | Messages

	PkgId	PackageName	year	Record_after_each_year	Load_Time
1	1	LoadingFiles	2013	265	2019-04-25 18:50:59.997
2	2	LoadingFiles	2014	530	2019-04-25 18:51:01.423
3	3	LoadingFiles	2015	795	2019-04-25 18:51:02.763
4	4	LoadingFiles	2016	1060	2019-04-25 18:51:04.053

- After loading the data, pivoted the required data for display in more understandable format using R
- After merging and pivoting the created facts and dimension tables using SQL

Dimension Tables:

```
--Gender Dimension Table--
create table Dim_Gender(Gender_id int identity(1,1) primary key, Gender varchar(100))
insert into Dim_Gender
Select distinct(Gender)
from Prevalence

select * from Dim_Gender

--Age Dimension Table--
create table Dim_Age(Age_id int identity(1,1) primary key, AgeGroup varchar(100))
insert into Dim_Age
Select distinct([Age Group])
from Prevalence

select * from Dim_Age

--Enrollment Dimension Table--
create table Dim_Enrollment(Enrollment_id int identity(1,1) primary key, Enrollment varchar(100))
insert into Dim_Enrollment
Select distinct(Enrollment)
from Prevalence

select * from Dim_Enrollment
```

Results | Messages

	Enrollment_id	Enrollment
1	1	All
2	2	Medicare & Medicaid
3	3	Medicare Only

Fact Tables:

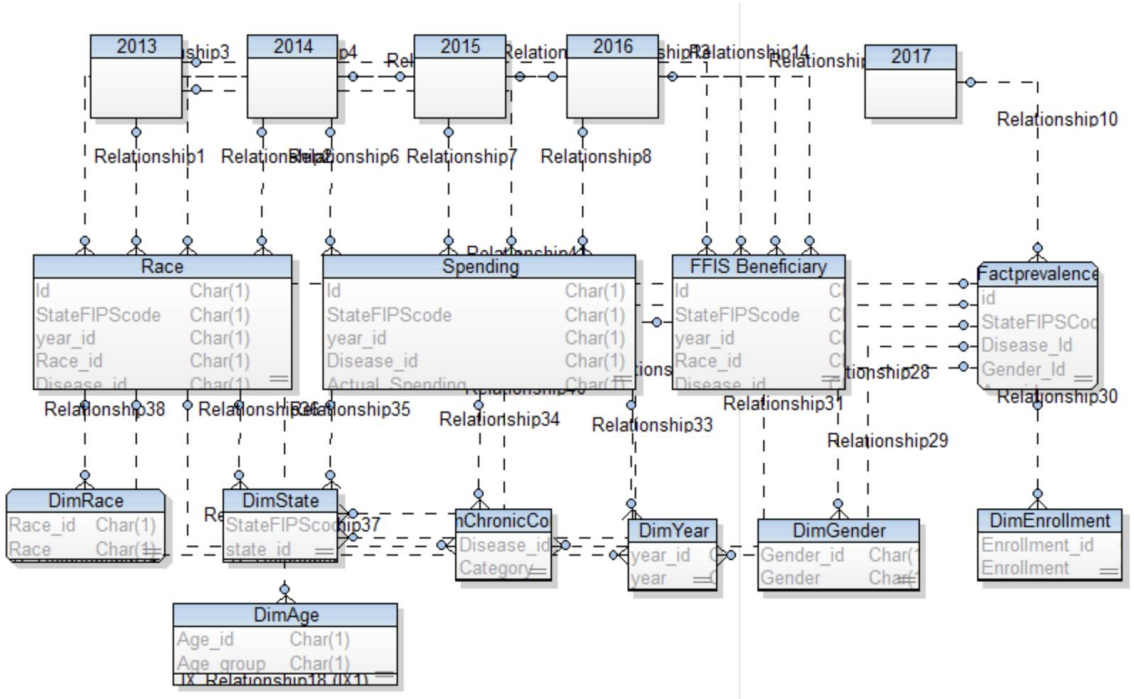
```
--Populating the Fact Table--

--Race Fact Table--
create table Fact_race_prevalence(id int primary key, [StateFIPSCode] int, year_id int, Race_id int, Disease_id int, Prevalence float,
                            foreign key([StateFIPSCode]) references Dim_state([StateFIPSCode]),
                            foreign key(Disease_id) references Dim_ChronicCondition(Disease_id),
                            foreign key(year_id) references Dim_Year(year_id),
                            foreign key(Race_id) references Dim_Race(Race_id))

insert into Fact_race_prevalence
--(StateFIPSCode,year_id,race_id,Disease_id,Prevalence)
select id,StateFIPSCode,year_id,race_id,Disease_id,[prevelance(%)]
from raceprev r
join Dim_ChronicCondition d on d.category=r.Disease
join Dim_State s on s.State=r.State
join Dim_year y on y.year=r.year
join Dim_Race rd on rd.Race=r.Race

select * from Fact_race_prevalence
```

0 %

Results | Messages

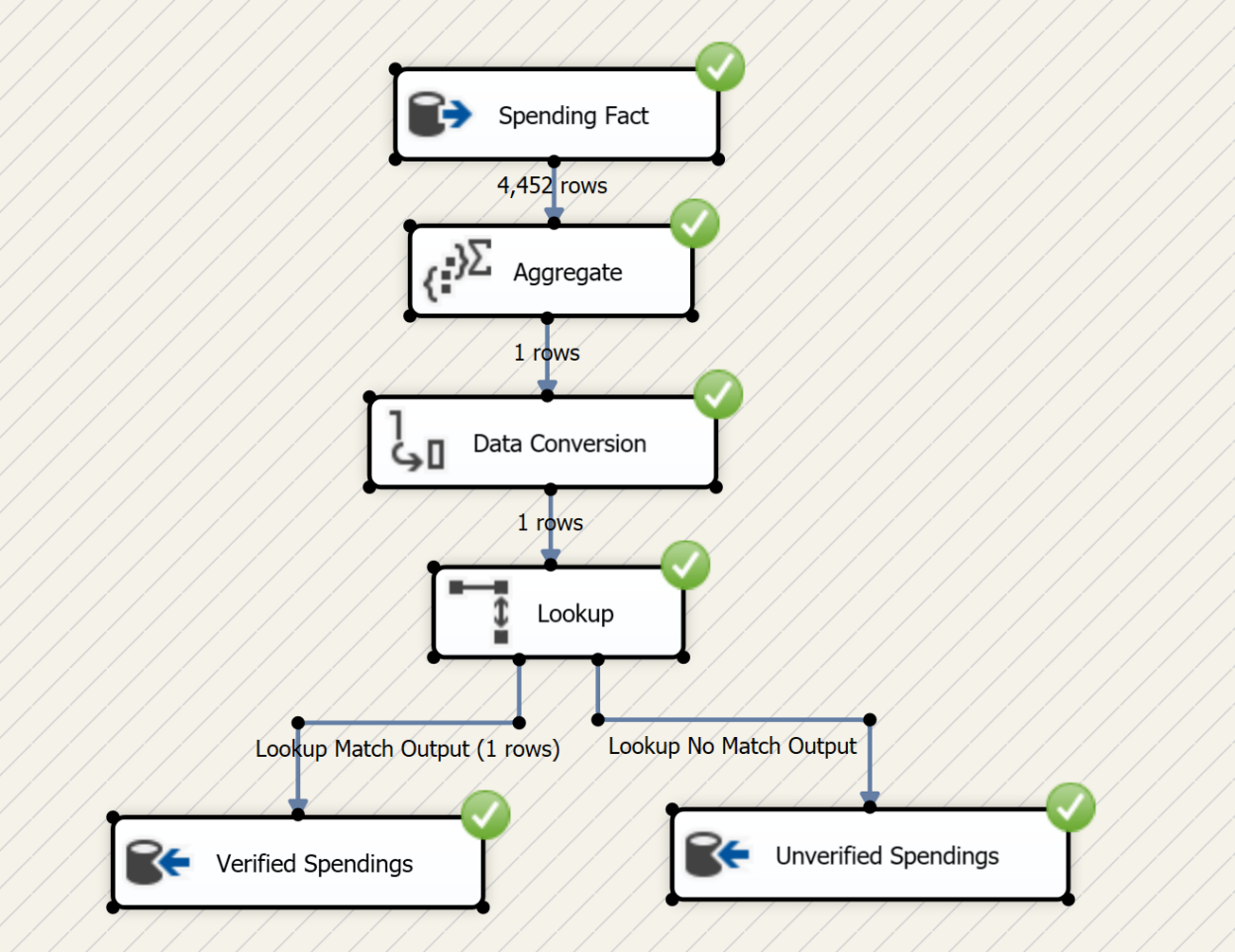
	id	StateFIPSCode	year_id	Race_id	Disease_id	Prevalence
1	1	1	3	4	7	1.56
2	2	2	3	4	7	2.08
3	3	4	3	4	7	2.08
4	4	5	3	4	7	1.89
5	5	6	3	4	7	2.24
6	6	8	3	4	7	3.22
7	7	9	3	4	7	3.82
8	8	10	3	4	7	2.12
9	9	11	3	4	7	2.95

Toad for our Dimensional Model



Validation of our Data Model

- Also validated the data for Fact_spending i.e. actual spending and standardized spending by taking average for both the spending across all the states



```
select * from [dbo].[Verified_Spendings]

select * from [dbo].[UnverifiedSpendings]
```

100 %

Results | Messages

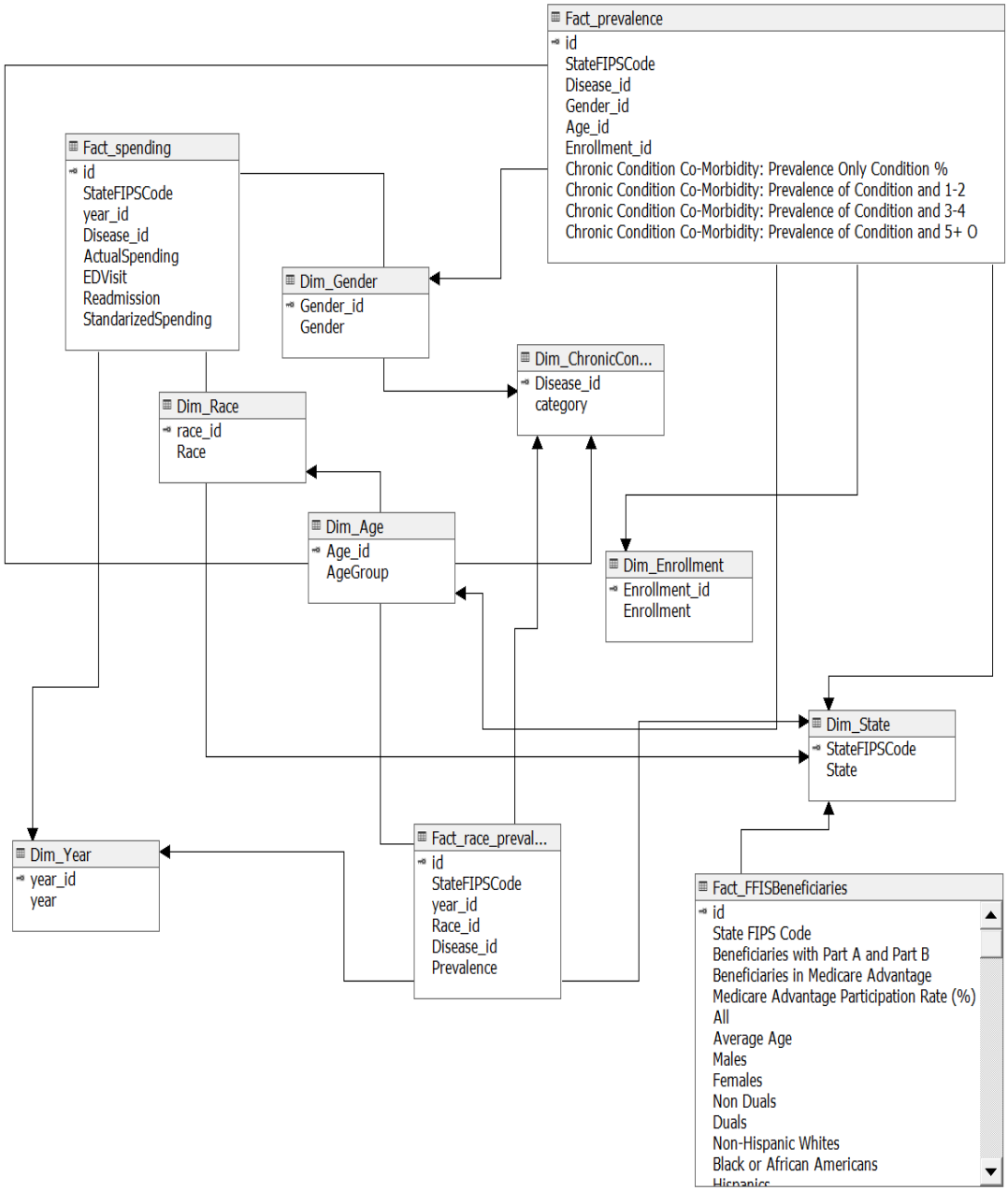
	ActualSpending	StandarizedSpending
1	20354.223078147428	19016.17111283723

	ActualSpending	StandarizedSpending

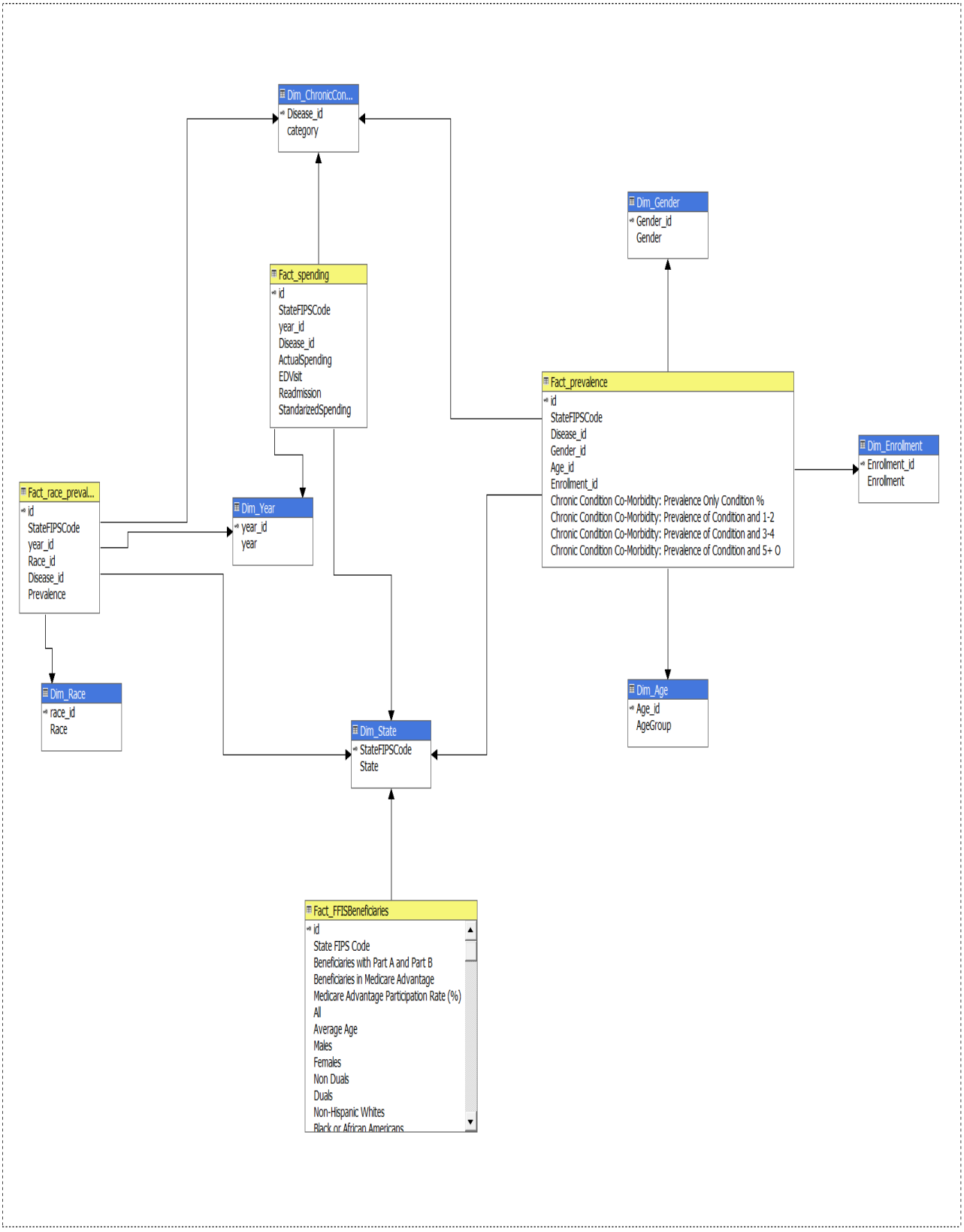
SSAS Cubes

Then Created SSAS Cubes using the fact and dimension tables

Data Source Views:

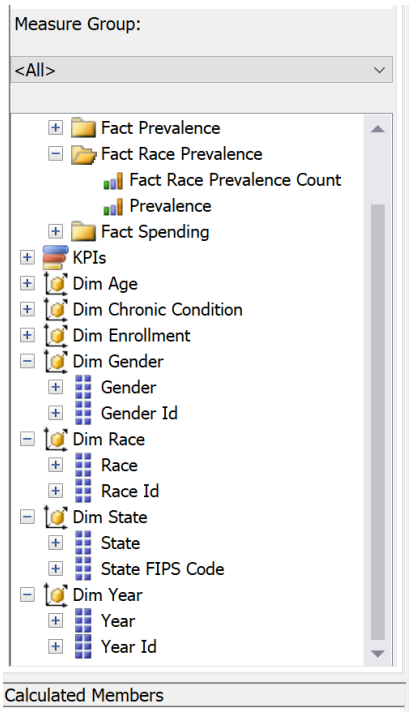


Cubes:


Dim_ChronicCon...
Disease_id
category
Dim_Gender
Gender_id
Gender
Fact_spending
id
StateFIPSCode
year_id
Disease_id
ActualSpending
EDVisit
Readmission
StandarizedSpending
Fact_prevalence
id
StateFIPSCode
Disease_id
Gender_id
Age_id
Enrollment_id
Chronic Condition Co-Morbidity: Prevalence Only Condition %
Chronic Condition Co-Morbidity: Prevalence of Condition and 1-2
Chronic Condition Co-Morbidity: Prevalence of Condition and 3-4
Chronic Condition Co-Morbidity: Prevalence of Condition and 5+ O
Dim_Enrolment
Enrollment_id
Enrollment
Fact_race_preval...
id
StateFIPSCode
year_id
Race_id
Disease_id
Prevalence
Dim_Year
year_id
year
Dim_Race
race_id
Race
Dim_State
StateFIPSCode
State
Dim_Age
Age_id
AgeGroup
Fact_FFISBeneficiaries
id
State FIPS Code
Beneficiaries with Part A and Part B
Beneficiaries in Medicare Advantage
Medicare Advantage Participation Rate (%)
All
Average Age
Males
Females
Non Duals
Duals
Non-Hispanic Whites
Black or African Americans

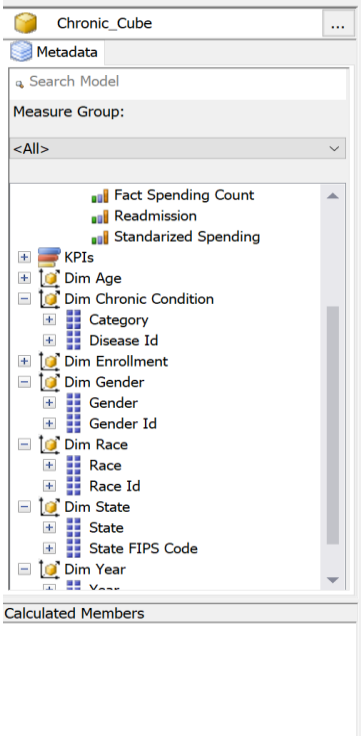
By browsing the cubes, we can see the outputs as per our requirements as following:

1.



State	Year	Prevalence
Alabama	2013	1451.99
Alabama	2014	1443.59
Alabama	2015	1476.9
Alabama	2016	1547.09
Alaska	2013	1109.1
Alaska	2014	1110.8
Alaska	2015	1116.52
Alaska	2016	1145.8
Arizona	2013	1290.03
Arizona	2014	1292.91
Arizona	2015	1312.06
Arizona	2016	1358.77
Arkansas	2013	1312.91
Arkansas	2014	1303.81
Arkansas	2015	1325.62
Arkansas	2016	1376.87
California	2013	1452.23
California	2014	1438.31
California	2015	1460.65
California	2016	1506.7
Colorado	2013	1166.78
Colorado	2014	1172.6
Colorado	2015	1175.07
Colorado	2016	1204.79
Connecticut	2013	1512.59
Connecticut	2014	1531.28
Connecticut	2015	1554.08

2.



Dimension | Hierarchy | Operator | Filter Expression

<Select dimension>

Category	Actual Spending	Standarized Spending
Alcohol Abuse	5005576.0975	4588374.7238
Alzheimer's Disease/Dementia	4541575.3096	4302343.8533
Arthritis	3169960.3311	3008909.0643
Asthma	4583200.8846	4279944.537
Atrial Fibrillation	4988990.6847	4654003.1396
Autism Spectrum Disorders	1579002.5677	1480146.1703
Cancer	4062049.9711	3830925.2984
Chronic Kidney Disease	5185166.5886	4851492.592
COPD	5213743.9346	4855432.095
Depression	4135697.0437	3896184.2723
Diabetes	3131049.7424	2962174.0848
Drug Abuse/Substance Abuse	5614273.3508	5148841.2297
Heart Failure	5752712.2054	5358890.25049999
Hepatitis (Chronic Viral B & C)	5351000.1574	4926595.6817
HIV/AIDS	3620055.544	3319086.1434
Hyperlipidemia	2803361.2828	2645728.2481
Hypertension	2911525.972	2748136.4524
Ischemic Heart Disease	3867773.6312	3625435.2776
Osteoporosis	3883994.1882	3675797.7565
Schizophrenia/Other Psychotic Disorders	4837672.4392	4557920.1652
Stroke	6297202.325	5867568.074

DIMENSIONAL MODEL – List of Tables

Fact Tables

1. Fact Spending – Measures Related to Spending, ED Visits, Readmissions etc.
2. Fact Race Prevalence – Captures prevalence across different ethnicities.
3. Fact FFIS Beneficiary – Captures information related to medicare beneficiaries
4. Fact Prevalence - Captures prevalence of different chronic conditions across different demographics, state etc.

Dimension Tables

1. Dim_Race
2. Dim_State
3. Dim_Age
4. Dim_Year
5. Dim_Enrollment
6. Dim_Gender
7. Dim_ChronicCondition

Junk Dimensions

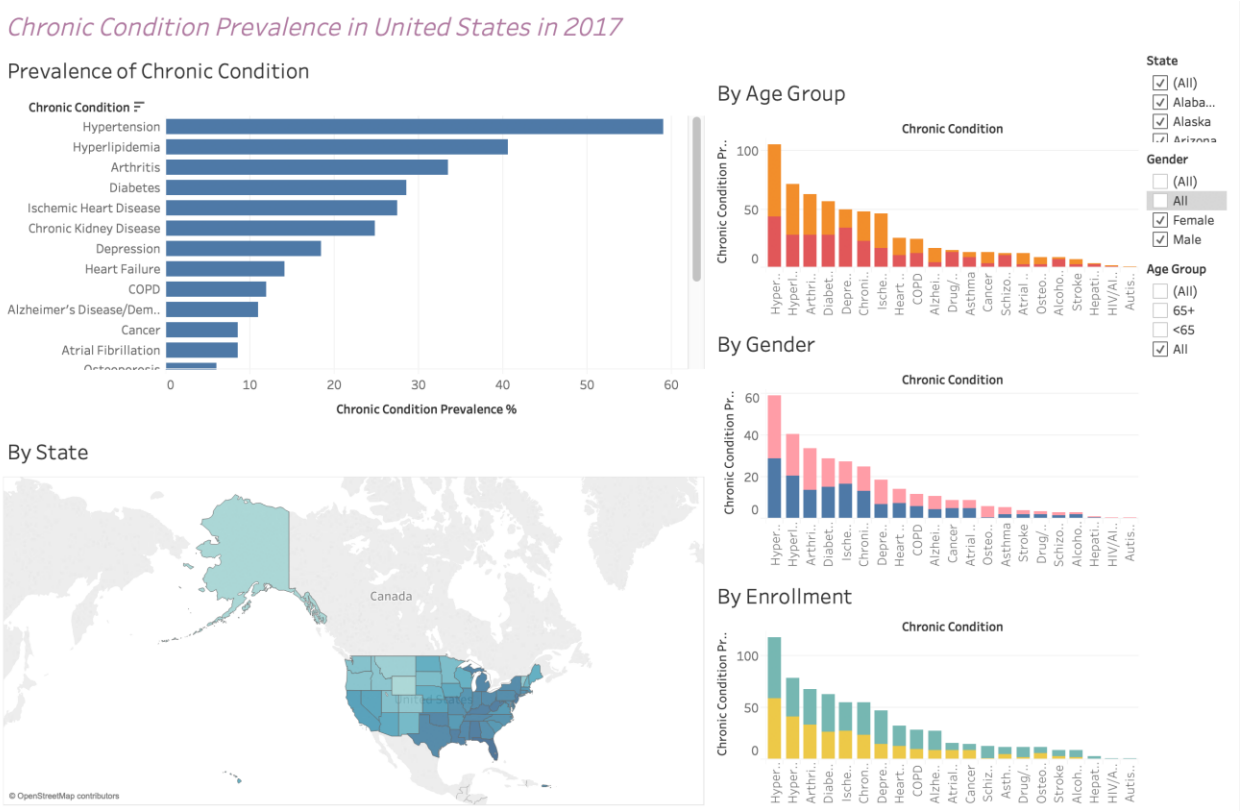
We have some junk dimensions as they do not have any information that can be used to join it to some dimension table.

1. **Co-morbidity –** This table captures the prevalence of various combinations of chronic diseases that are present in a person suffering from 2 or 3 chronic conditions simultaneously. They had each row with some combination of chronic diseases and did not capture any demographic data therefore it is used as junk dimension.

BUSINESS INTELLIGENCE DASHBOARD:

Our main objective in this stage was to create a user-friendly interactive dashboard that will be used by the decision makers to draw insights. We have visually analyzed various charts for the user to understand the basic gist of our analysis.

Visualizations by Mayanka Verma

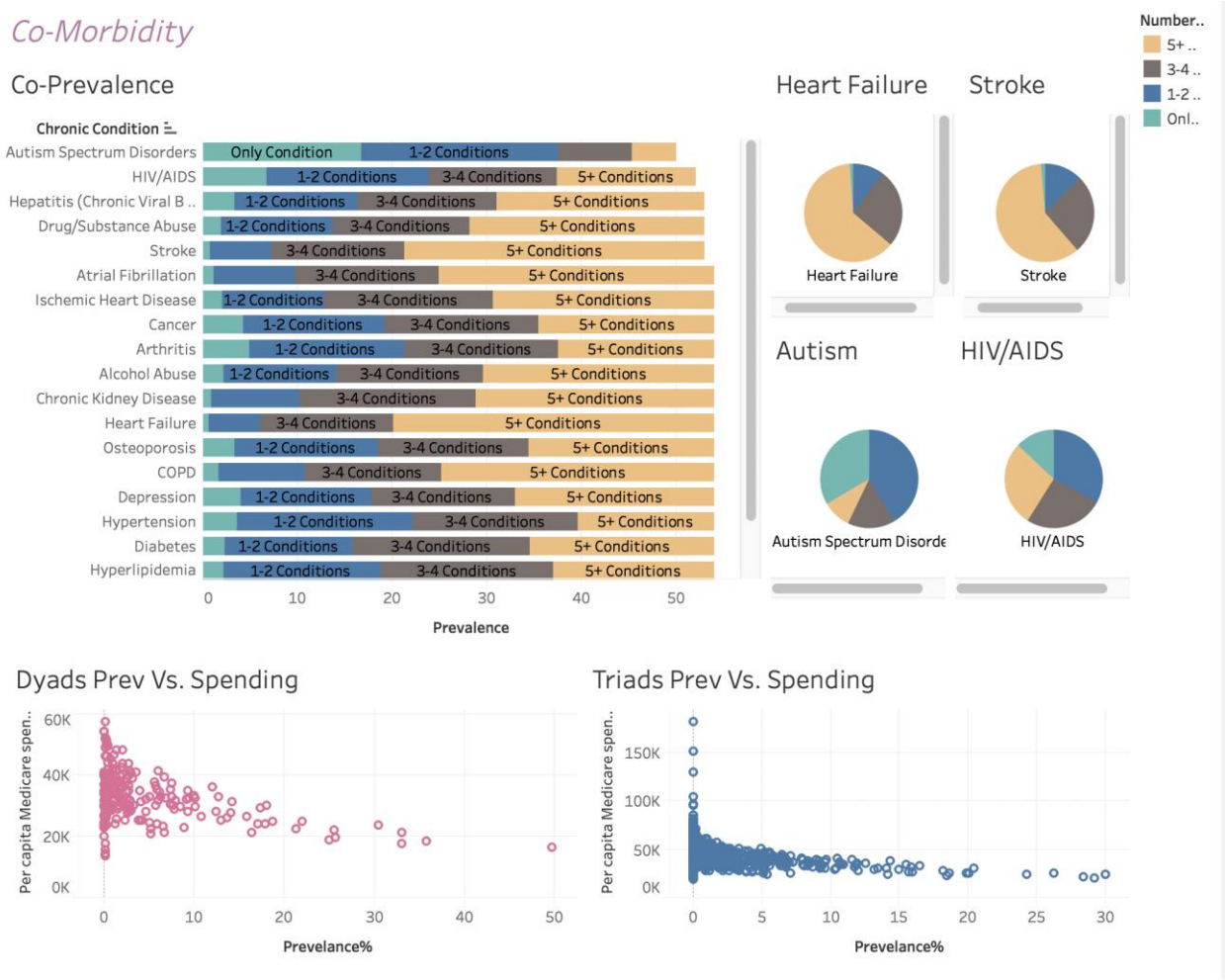


The first dashboard displays the prevalence of chronic conditions in the year 2017. We saw that hypertension and hyperlipidemia were the two most prevalent chronic conditions across United States in 2017. Hypertension is most prevalent with a significant margin, therefore we can assume that if a patient is suffering from multiple chronic conditions, then hypertension will most probably be one of them.

We saw the overall prevalence of the chronic conditions in United States across states. We found that the chronic conditions were more prevalent in the Eastern Coast in comparison to West. Florida and New Jersey are the states with the highest chronic conditions followed by West Virginia and Kentucky. Alaska

has a low prevalence of chronic conditions but it has a high prevalence of drugs and alcohol abuse. After that we saw that Alzheimer is a condition more prevalent for people over 65 years and Schizophrenia is more prevalent for people below the age 65.

Osteoporosis and asthma are significantly more prevalent for female in comparison to male. Patients suffering from drug substance abuse and Schizophrenia have both Medicare and Medicaid enrollment.

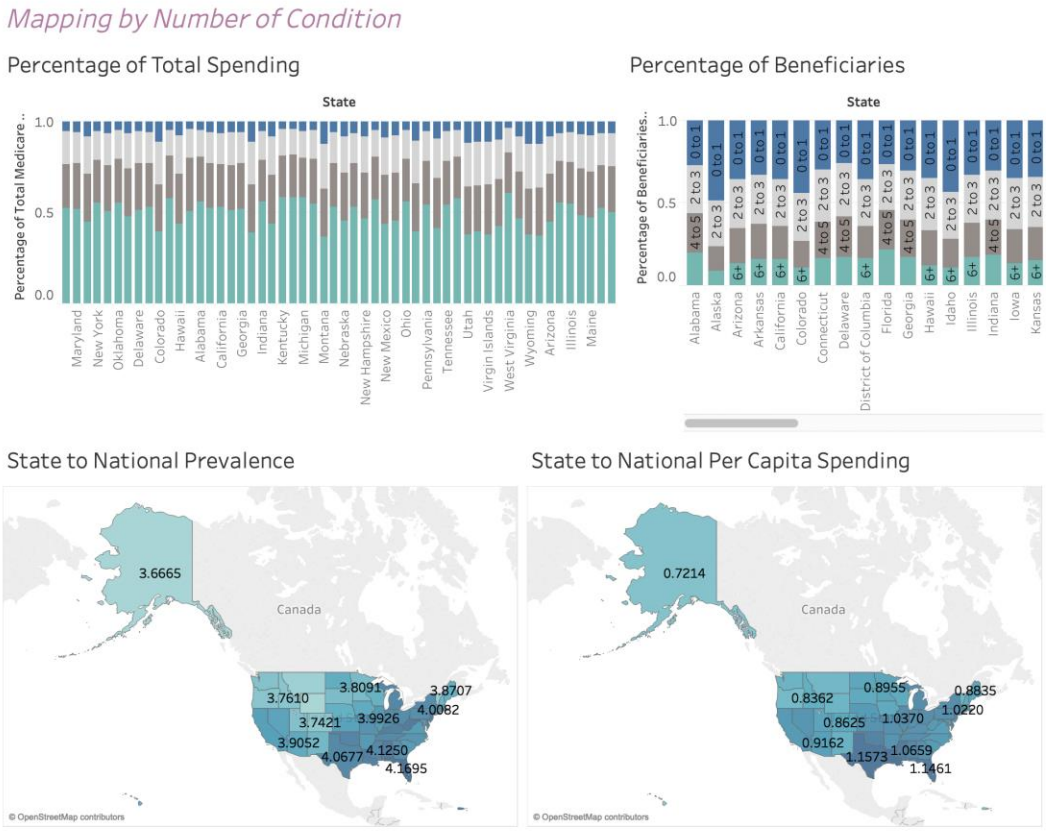


After realizing the overall prevalence of the chronic conditions, we wanted to see the comorbidity of the conditions that is when the patient suffers from multiple conditions simultaneously. In the first visualization, we are seeing that stroke, atrial fibrillation, heart failure, chronic kidney diseases are the ones that have a very low prevalence of occurring in only one condition i.e. the patient suffering from these conditions might be suffering from others as well. They are more clearly represented in the pie chart where we can see that Heart failure and stroke has the highest prevalence for 5 or more conditions. On the other hand, Autism and HIV Aids are more prevalent in a single condition or 1-2 conditions.

When analyzing the Dyads condition, we found that Hyperlipidemia and Hypertension, Arthritis and Hypertension, Asthma and Hypertension are the most frequently occurred dyads in patients. Hypertension is common for all of this. As we have already seen this in previous dashboard that hypertension is the most prevalent chronic condition, this satisfies our hypothesis that hypertension will be prevalent for most of the patients suffering from multiple chronic conditions.

When seeing the triads, we saw that heart diseases, hyperlipidemia, hypertension, arthritis, asthma are the most commonly prevalent ones.

Interesting insights from this visualization were that spending on triads were almost thrice of that of spending on dyads conditions. Also, we saw that as the condition becomes more prevalent, we see a significant decrease in the per capita spending. This could be due to the fact that the condition is more prevalent and therefore the treatment could have become more accessible for patients.

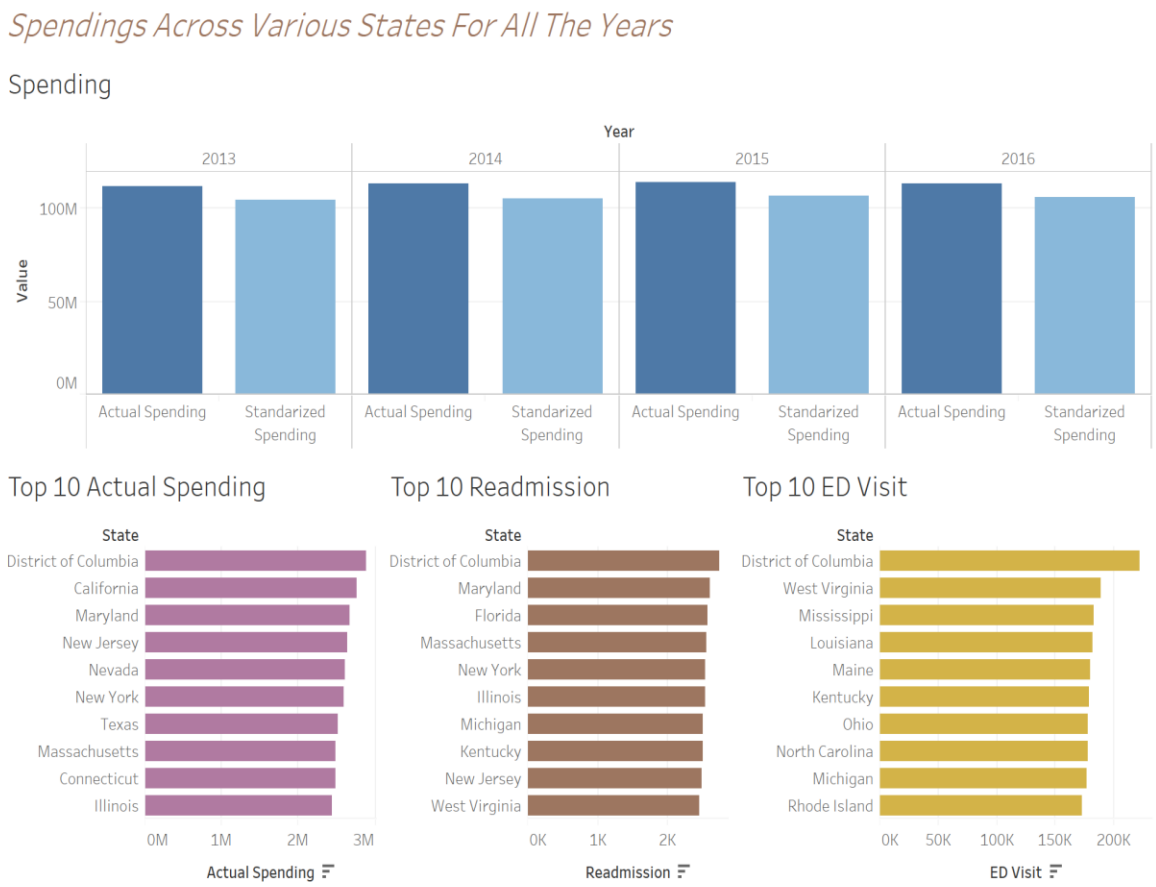


We can see that the percentage of total spending is highest when the person is suffering from 6+ chronic conditions. It is the least when the person has 0-1 chronic conditions. In contrary to that we saw that the

percentage of beneficiaries across the number of chronic conditions people were suffering from, people with 0-1 chronic conditions has the highest number of beneficiaries. This makes sense as the people having 0-1 chronic conditions are the major population who are beneficiaries and therefore, they have lesser per capita spending.

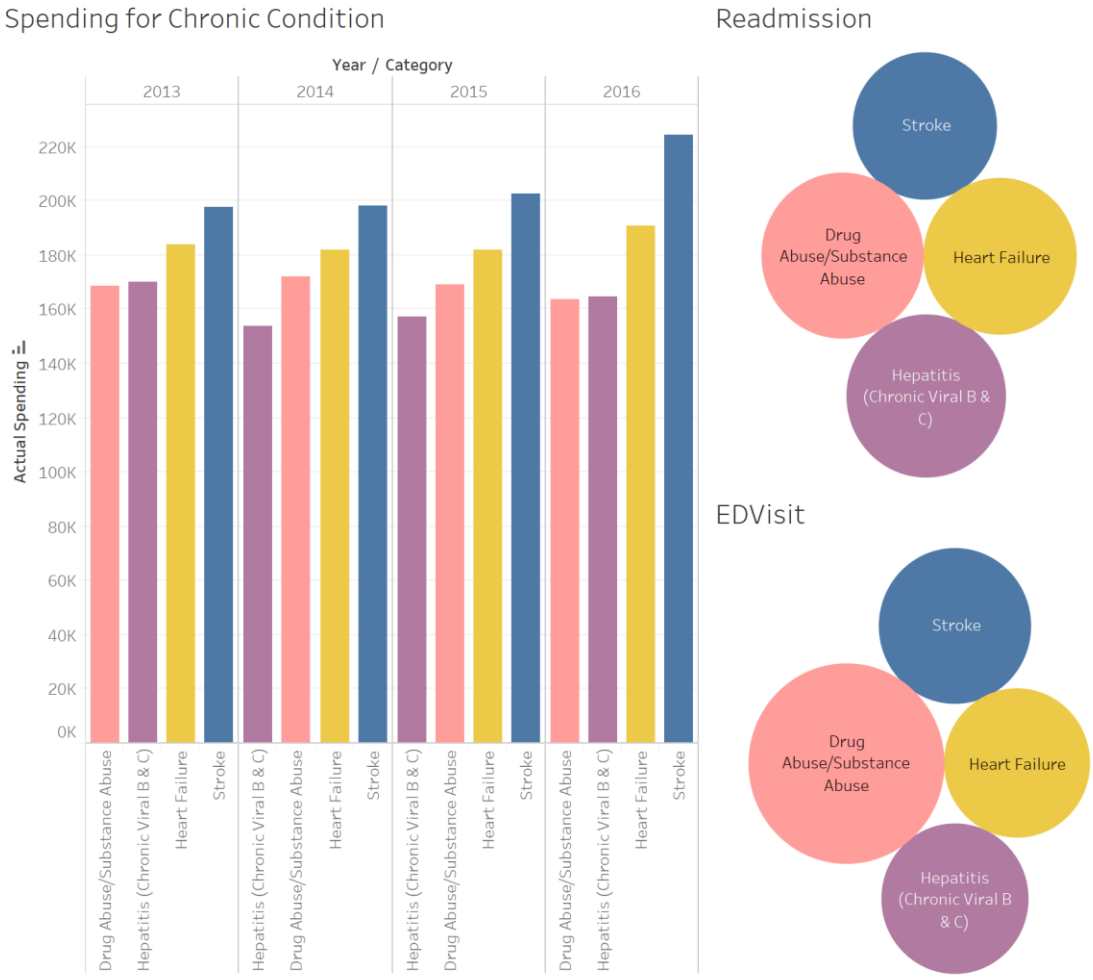
Next, we have seen the state to national ratio for prevalence to identify the states that have higher prevalence and per capita spending in comparison to the average prevalence across United States and the average per capita spending. States that have higher prevalence seems to have comparatively higher prevalence in comparison to overall United States.

Visualizations by Abhinav Kaushik



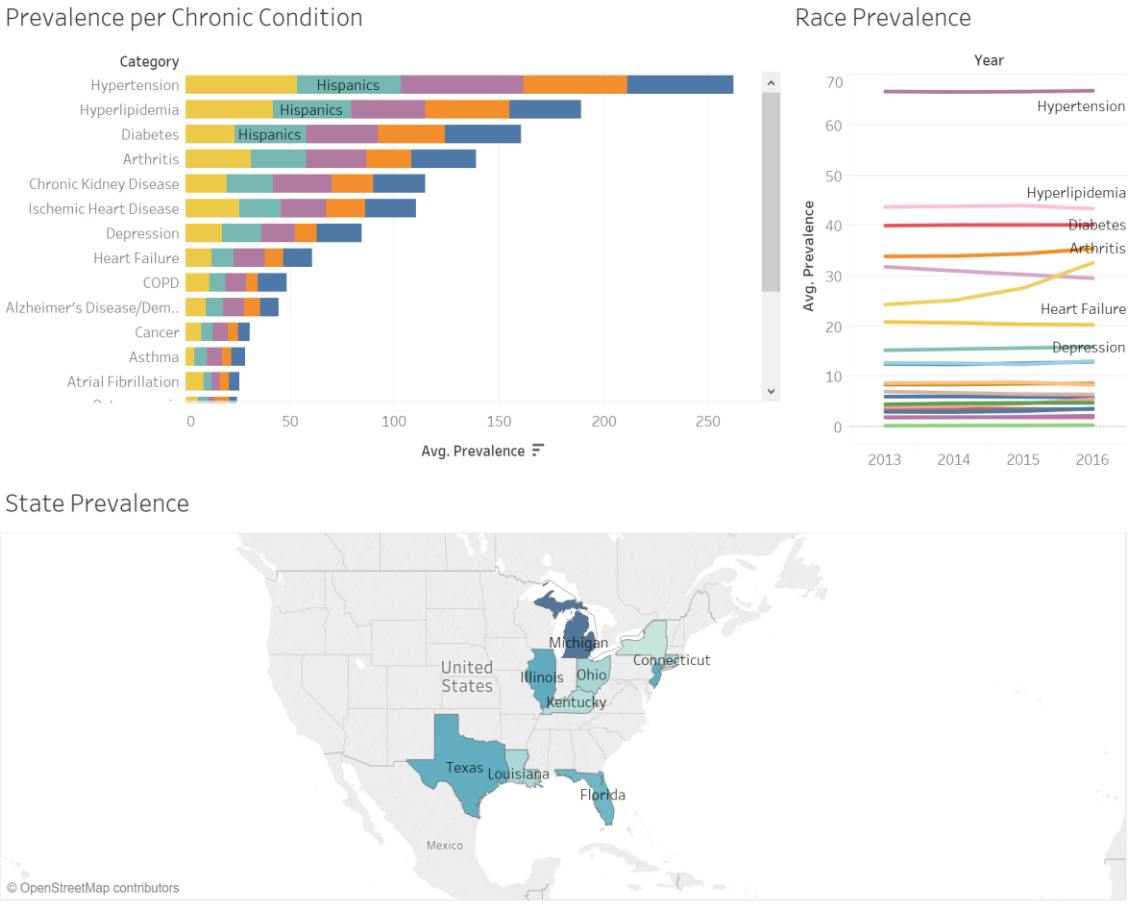
It can be seen that Actual Spending and Standardized Spending are almost same across all the 4 years i.e. from 2013 to 2016. There is no such difference in trends in these years. Actual spending is invested maximum in District of Columbia. It can be interpreted that since Readmission and Emergency Visit are both maximum in District of Columbia, therefore this could be the reason for maximum investment of Actual Spending in District of Columbia. It can also be interpreted that Top 10 Actual Spending and Top Readmission are very much related to each other because many states are common in both of them.

Chronic Condition Analysis in District Of Columbia



In District of Columbia, the maximum amount of Actual Spending is invested on Stroke Chronic Condition in each of the four years while least is invested on Drug Abuse/Substance Abuse in each year. Drug Abuse/Substance Abuse checking contributes most to the Readmission and Emergency Department visit for the year 2016, therefore it can be interpreted that treatment Drug Abuse is cheaper while Stroke is the most expensive chronic condition. It can also be interpreted that the Drug Abuse is a great problem in District of Columbia and thus, it should be taken into consideration and Government should take necessary measures to minimize it.

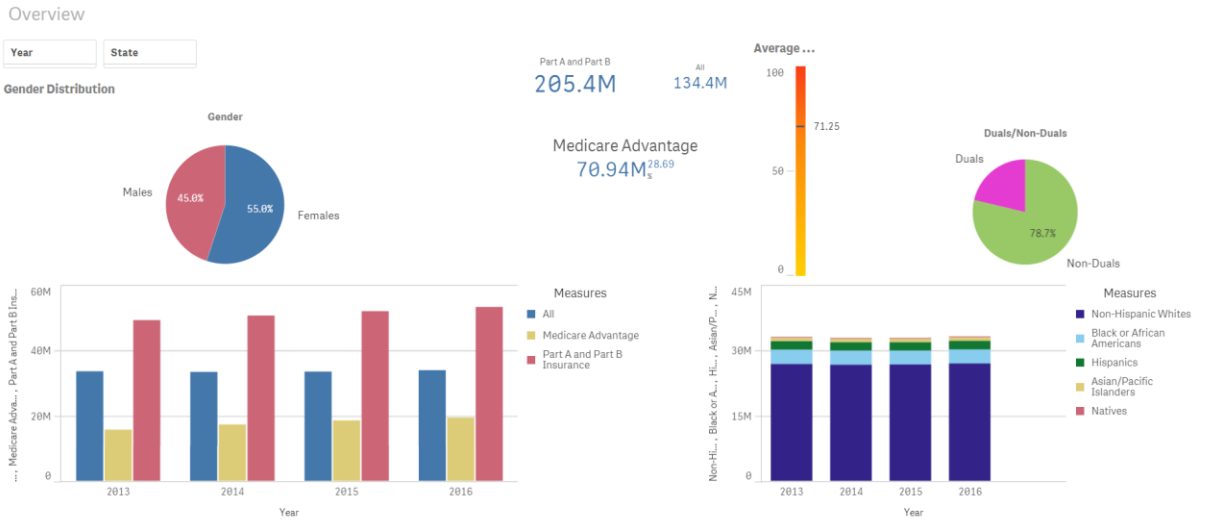
Prevalence Of Chronic Condition Across Races



It can be interpreted that prevalence is maximum for Hypertension chronic condition for all the races. And Black or African Americans are most prone to hypertension. They are also prone to other chronic condition in a significant quantity. By plotting the trend line for average prevalence of Black or African Americans for the years from 2013 to 2016, it can be seen that while other chronic condition remains almost constant but Chronic Kidney shows a significant increase in trend for the given years. Michigan, Illinois, Texas and New Jersey are the top 4 states where the average of prevalence is maximum for Black or African Americans.

Visualizations by Farooque Akhtar

FFIS Beneficiary Fact Table Visualisation

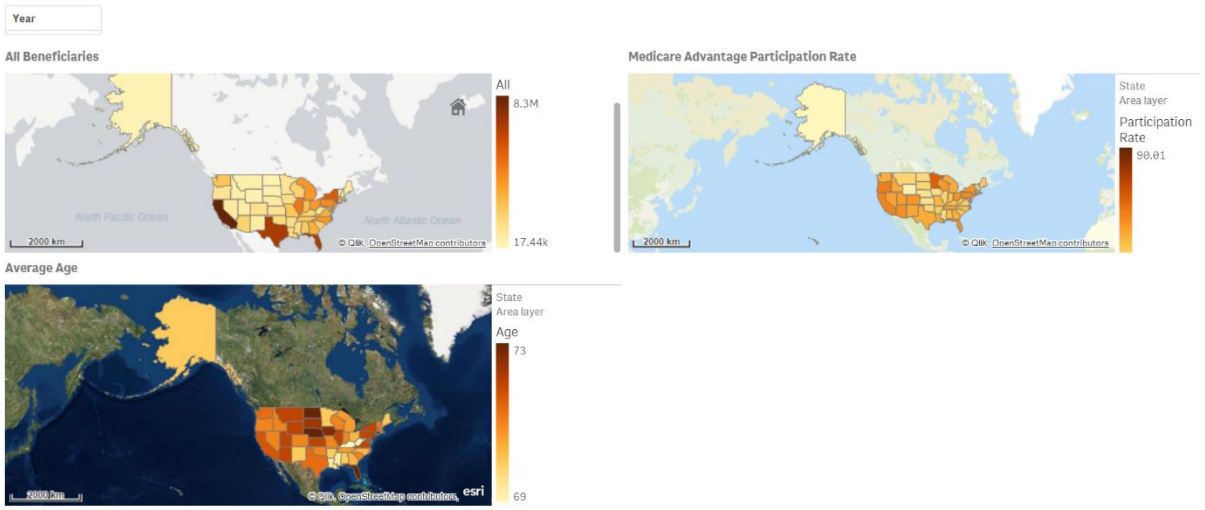


Total number of participants who have taken insurance are 205.5 M. Out of which 70.94M have taken Medicare Advantage.

The total number of Females are 55% whereas males are 45%.

The Average Age of participants is 71.25

A Trend Analysis of types of insurance have been made from 2013 to 2016. The next chart shows various Races across United States who have subscribed for Insurance. Non-Hispanic Whites have been maximum subscribers while Natives being the least subscribers.



The above Graphs represent the location of beneficiaries as well as average age across states.

Texas has the maximum All beneficiaries which means that is the total number of people who have taken various type of insurances. While Wyoming and North Dakota are the places with least number of subscribers.
Florida has maximum Average Age while Kentucky is the least. This data can be helpful for focusing on more subscribers.

REFERENCES

1. https://www.rand.org/blog/rand-review/2017/07/chronic-conditions-in-america-price-and-prevalence.html
2. Dataset - https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/index.html