/\*

Submitted by: Murli Manohar Singh

Date: 07/19/2019 01:20AM

Description: To explain the steps followed for analysis

\*/

/\*Step 1: **Extract** the dataset using the flat file link:

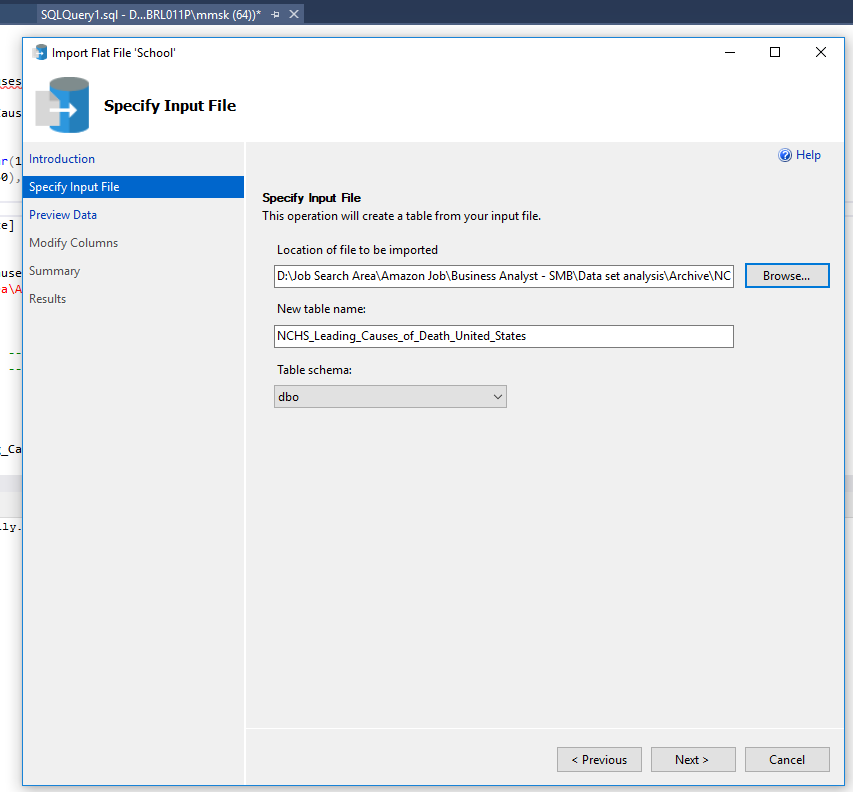
<https://catalog.data.gov/dataset/age-adjusted-death-rates-for-the-top-10-leading-causes-of-death-united-states-2013>

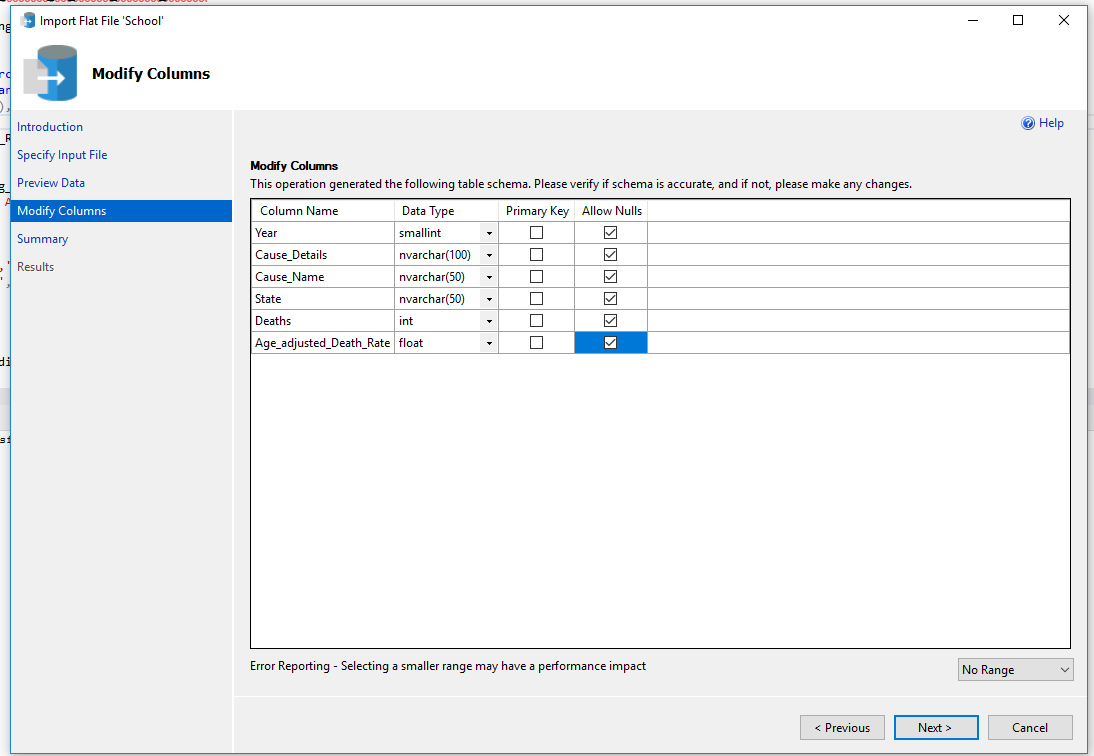
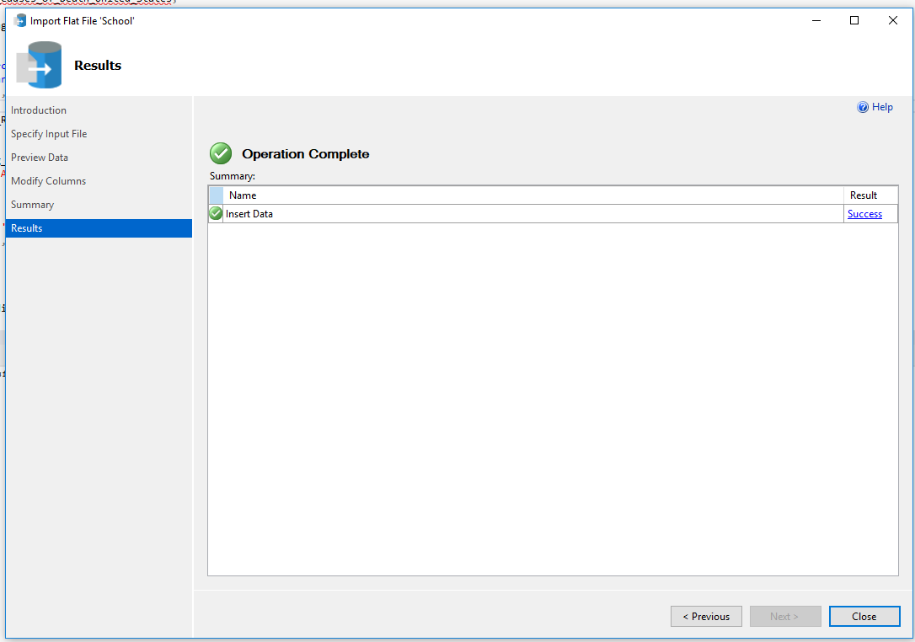
\*/

/\*

Step 2: **Transform & Load** Data set into SQL Server:

\*/





/\*

Step 3: Write SQL queries to fetch numbers for defined KPIs/Metrics

\*/

/\*

1. For insights point 1 (per analysis summary document):\*/

Select inn1.\*,CONVERT(DECIMAL(10,1),(inn1.Total\_deaths\*100.00)/inn2.Total\_Sum\_Of\_deaths) as [% contribution]

from (select Cause\_name,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name !='All causes'

group by Cause\_name

) as inn1

, (select Sum(deaths) as Total\_Sum\_Of\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016

) as inn2

order by 3 desc;

/\*

Note:

1. I don’t need All other causes of death. So, I have filtered it out using Cause\_name !='All causes'

2. Also analysis is for 10 years and dataset has data only until 2016. So, from 2007 to 2016.

3. Extract the result set in Excel or connect to tableau for graphs that I built for the Summary document

\*/

**/\* Below queries are for Four biggest states (California, Texas, Florida and New York) and same concepts mentioned in “Note” sections are applied here as well \*/**

-- For California

select inn1.\*,CONVERT(DECIMAL(10,1),(inn1.Total\_deaths\*100.00)/inn2.Total\_Sum\_Of\_deaths\_cali) as [% contribution]

from (select Cause\_name,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name !='All causes' and state='California'

group by Cause\_name

) as inn1

, (select Sum(deaths) as Total\_Sum\_Of\_deaths\_cali

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and state='California'

) as inn2

order by 3 desc;

-- For Texas

select inn1.\*,CONVERT(DECIMAL(10,1),(inn1.Total\_deaths\*100.00)/inn2.Total\_Sum\_Of\_deaths\_tx) as [% contribution]

from (select Cause\_name,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name !='All causes' and state='Texas'

group by Cause\_name

) as inn1

, (select Sum(deaths) as Total\_Sum\_Of\_deaths\_tx

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and state='Texas'

) as inn2

order by 3 desc;

-- For Florida

select inn1.\*,CONVERT(DECIMAL(10,1),(inn1.Total\_deaths\*100.00)/inn2.Total\_Sum\_Of\_deaths\_flo) as [% contribution]

from (select Cause\_name,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name !='All causes' and state='Florida'

group by Cause\_name

) as inn1

, (select Sum(deaths) as Total\_Sum\_Of\_deaths\_flo

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and state='Florida'

) as inn2

order by 3 desc;

-- For New York

select inn1.\*,CONVERT(DECIMAL(10,1),(inn1.Total\_deaths\*100.00)/inn2.Total\_Sum\_Of\_deaths\_nyc) as [% contribution]

from (select Cause\_name,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name !='All causes' and state='New York'

group by Cause\_name

) as inn1

, (select Sum(deaths) as Total\_Sum\_Of\_deaths\_nyc

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and state='New York'

) as inn2

order by 3 desc;

/\*

1. For insights point 2 (per analysis summary document):\*/

Select top 3 inn1.Cause\_Name,CONVERT(DECIMAL(10,1),(inn2.Total\_deaths\_2016-inn1.Total\_deaths\_2007)\*100.00/inn1.Total\_deaths\_2007) as [% increase in 10 years]

from (select Cause\_name,Sum(deaths) as Total\_deaths\_2007

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year=2007 and Cause\_name !='All causes'

group by Cause\_name

) as inn1

join

(select Cause\_name,Sum(deaths) as Total\_deaths\_2016

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year=2016 and Cause\_name !='All causes'

group by Cause\_name

) as inn2

on inn1.Cause\_Name=inn2.Cause\_Name

order by 2 desc;

/\*

Note:

1. Fetched top 3 death cause rate (% increase in 10 years)

2. I don’t need All other causes of death. So, I have filtered it out using Cause\_name !='All causes'

3. Analysis is for 10 years and dataset has data only until 2016. So, from 2007 to 2016.

4. Extract the result set in Excel or connect to tableau for graphs that I built for the Summary document

\*/

-- for Alzheimer''s disease

select Cause\_name,year,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name ='Alzheimer''s disease'

group by Cause\_name,Year

order by 2 asc;

-- for Unintentional injuries or accident

select Cause\_name,year,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name ='Unintentional injuries'

group by Cause\_name,Year

order by 2 asc;

/\*

Note:

1. Please note 'Unintentional injuries' is same as accident. So, in the BI tools I have updated the label to Accident to fit the graph labels better.

\*/

-- for Suicide

select Cause\_name,year,Sum(deaths) as Total\_deaths

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2007 and 2016 and Cause\_name ='Suicide'

group by Cause\_name,Year

order by 2 asc;

/\*

1. For insights point 3 (per analysis summary document):\*/

select inn1.Year,inn1.Total\_deaths\_next,CONVERT(DECIMAL(10,2),(inn1.Total\_deaths\_next-inn2.Total\_deaths\_previous)\*100.00/inn2.Total\_deaths\_previous) as [death rate over previous year]

from

(select year,Sum(deaths) as Total\_deaths\_next

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2006 and 2016

group by Year) as inn1

join

(select year,Sum(deaths) as Total\_deaths\_previous

from NCHS\_Leading\_Causes\_of\_Death\_United\_States

where year between 2006 and 2016

group by Year) as inn2

on inn1.Year=inn2.Year+1

order by inn1.Year;

/\*

Note:

1. Analysis is for 10 years and dataset has data only until 2016.

2. Extract the result set in Excel or connect to tableau for graphs that I built for the Summary document

\*/