Data and Questions

In this project, I used Microsoft SQL Server Management Studio to analyze data on opioid overdose deaths as well as prescriber and prescription information in the United States. The data ranged between 1999 and 2014, and both datasets can be found on Kaggle. It is to be noted that I cleaned the data in excel for the Prescription and Prescriber Information dataset. I disposed of any and all drug classes that do not qualify as an opiate, and created an additional column using an aggregate function in order to total all of the prescriptions for each prescriber. My focus was to gain insight on the following questions:

- What percentage of the United States population dies each year due to opioids?
- How many total opioid overdose deaths were there between 1999 and 2014 in the United States, and what years were opioid overdose deaths most prevalent in that period?
- Calculate a rolling death count by state.
- How many total opioid overdose deaths were there between 1999 and 2014 in each state?
- What states had the highest rate of opioid overdose deaths compared to the population?
- What states had the highest rate of opioid overdose deaths per 100,000 people, most frequently between 1999 and 2014?
- Show states with the highest and lowest rates of opioid overdose deaths per 100,000 people in 2014.
- How many opioid prescriptions were prescribed in the United States in 2014?
- Show prescriber information of those who issued the highest amount of prescriptions in 2014.
- How many prescriptions on average were prescribed per provider in each state in 2014?
- What is the rate of prescriptions in comparison to the population in each state in 2014?

The Analysis

- **What percentage of the United States population dies each year due to opioids?**

This first question not only determined the death to population ratio for each year, but also the correlation between that ratio and the amount of prescriptions dispensed by US retailers in that year (in millions). I used the SQL statement below to display that information.

```
Select Year, SUM(Deaths) AS TotalDeaths, SUM(Population) AS USPopulation,
(SUM(Deaths)/SUM(Population))*100 AS DeathPercentageUS,
[Prescriptions Dispensed by US Retailers in that year (millions)]
From PortfolioProjectA..OpioidDeaths
Group By Year, [Prescriptions Dispensed by US Retailers in that year (millions)]
Order By DeathPercentageUS DESC;
```

	Year	TotalDeaths	USPopulation	DeathPercentageUS	Prescriptions Dispensed by US Retailers in that year (millions)
1	2014	29650	318857056	0.00929883765846474	196
2	2013	26023	316128839	0.00823177033842205	207
3	2012	24098	313914040	0.00767662382988668	217
4	2011	23757	311591917	0.00762439546851275	219
5	2010	22057	308745538	0.00714407085617542	210
6	2009	21311	306771529	0.00694686370324803	202
7	2008	20343	304093966	0.00668970853568334	196
8	2007	19286	301231207	0.00640239110418596	184
9	2006	18457	298379912	0.00618573813373871	174
10	2005	15742	295516599	0.00532694273461099	163
11	2004	14584	292805298	0.00498078419332426	155
12	2003	13699	290107933	0.00472203564319629	149
13	2002	12656	287625193	0.00440017088489185	142
14	2001	10123	284968955	0.0035523167778048	138
15	2000	9036	281421906	0.00321083746764191	126
16	1999	8627	279040168	0.00309166958357049	116

The results clearly show that the total deaths as well as the death percentage increased each year, and that each actually tripled in that 15 year period (or increased by 300%). As can also be seen, the United States population only increased by 15%. The amount of prescriptions dispensed increased by 90%.

This introduction into the analysis distinctly presents the harsh reality of the opioid epidemic of the United States.

- **How many total opioid overdose deaths were there between 1999 and 2014 in the United States, and what years were opioid overdose deaths most prevalent in that period?**

This question involves multiple SQL statements and results. Firstly, to show the total number within the 15 year period, and then breaking the number down into 3 separate 5 year periods.

```
--The total amount of opioid overdose deaths in the United States between 1999 and 2014

Select SUM(Deaths) AS Total_Deaths_in_the_US_From_1999_to_2014
From PortfolioProjectA..OpioidDeaths

--BREAKING THINGS DOWN INTO 5 YEAR PERIODS
--The total amount of opioid overdose deaths in the United States between 1999 and 2003

Select SUM(Deaths) AS Total_Deaths_in_the_US_From_1999_to_2003
From PortfolioProjectA..OpioidDeaths
Where Year IN (1999, 2000, 2001, 2002, 2003)

--The total amount of opioid overdose deaths in the United States between 2004 and 2009

Select SUM(Deaths) AS Total_Deaths_in_the_US_From_2004_to_2009
From PortfolioProjectA..OpioidDeaths
Where Year IN (2004, 2005, 2006, 2007, 2008, 2009)

--The total amount of opioid overdose deaths in the United States between 2010 and 2014

Select SUM(Deaths) AS Total_Deaths_in_the_US_From_2010_to_2014
From PortfolioProjectA..OpioidDeaths
Where Year IN (2010, 2011, 2012, 2013, 2014)
```

	Total_Deaths_in_the_US_From_1999_to_2014
1	289449

	Total_Deaths_in_the_US_From_1999_to_2003
1	54141

	Total_Deaths_in_the_US_From_2004_to_2009
1	109723

	Total_Deaths_in_the_US_From_2010_to_2014
1	125585

The results show that 43% of the total 289,449 opioid overdose deaths between 1999 and 2014 occurred between 2010 and 2014. 38% of them were between 2004 and 2009, and only 19% between 1999 and 2003.

- **Calculate a rolling death count by state.**

This question asks for the amount of deaths in each state, by year, from 1999 to 2014. It's visually beneficial and would be especially so if utilized in a dashboard on a software such as Tableau. Below I've included the SQL statement, in which I used Partition to create the rolling death count per state. Due to the size of the results, I've included only the results of the state of West Virginia, since in some of the following questions it proves to be significantly more prevalent in terms of deaths and prescriptions than other states.

```
Select State, Year, Deaths, Population, [Crude Rate],
SUM(Deaths) OVER (Partition by State Order by State, Year) AS RollingDeathCount
FROM PortfolioProjectA..OpioidDeaths
```

	State	Year	Deaths	Population	Crude Rate	RollingDeathCount
1	West Virginia	1999	34	1811799	1.9	34
2	West Virginia	2000	52	1808344	2.9	86
3	West Virginia	2001	150	1801481	8.3	236
4	West Virginia	2002	183	1805414	10.1	419
5	West Virginia	2003	203	1812295	11.2	622
6	West Virginia	2004	273	1816438	15	895
7	West Virginia	2005	154	1820492	8.5	1049
8	West Virginia	2006	299	1827912	16.4	1348
9	West Virginia	2007	347	1834052	18.9	1695
10	West Virginia	2008	387	1840310	21	2082
11	West Virginia	2009	193	1847775	10.4	2275
12	West Virginia	2010	495	1852994	26.7	2770
13	West Virginia	2011	596	1855364	32.1	3366
14	West Virginia	2012	513	1855413	27.6	3879
15	West Virginia	2013	518	1854304	27.9	4397
16	West Virginia	2014	585	1850326	31.6	4982

As can be seen above, the death count in West Virginia increased by 14,653% in just a 15 year period. These results indicate which states have the greatest increase in deaths over the 15 year period given.

- **How many total opioid overdose deaths were there between 1999 and 2014 in each state?**

This question provides a simple portrayal of the total death count for each state.

```
Select State, SUM(Deaths) AS TotalDeaths
From PortfolioProjectA..OpioidDeaths
Group by State
Order by TotalDeaths DESC
```

	State	TotalDeaths
1	California	27044
2	Florida	19919
3	New York	16156
4	Texas	15050
5	Ohio	13623
6	Illinois	13072
7	North Carolina	10413
8	Massachusetts	9923
9	Washington	9528
10	Maryland	9403
11	Pennsylvania	9304
12	Michigan	8748
13	Tennessee	7467
14	Arizona	7298
15	Virginia	7210
16	New Jersey	6961
17	Missouri	6460
18	Kentucky	6283
19	Georgia	6052
20	Nevada	5954

I've provided the top 20 instances. The results are to be expected with the highest contenders being the greatest in population as well.

- **What states had the highest rate of opioid overdose deaths compared to the population?**

This question breaks down the death percentage from the first question by state, determining the death to population ratio in 2014. I include only the top 15 results here.

```
Select State, Year, Deaths, Population, (Deaths/Population)*100 AS DeathPercentage
From PortfolioProjectA..OpioidDeaths
Where Year=2014
Order By DeathPercentage DESC
```

	State	Year	Deaths	Population	DeathPercentage
1	West Virginia	2014	585	1850326	0.0316160503608553
2	New Hampshire	2014	302	1326813	0.0227613084888375
3	Rhode Island	2014	211	1055173	0.0199967209168544
4	New Mexico	2014	416	2085572	0.019946566217805
5	Ohio	2014	2175	11594163	0.0187594395559214
6	Kentucky	2014	762	4413457	0.017265377231499
7	Massachusetts	2014	1161	6745408	0.0172117090619278
8	Utah	2014	471	2942902	0.0160046104151616
9	Maryland	2014	932	5976407	0.0155946541124124
10	Connecticut	2014	531	3596677	0.0147636276485211
11	Nevada	2014	413	2839099	0.0145468685664008
12	Tennessee	2014	892	6549352	0.0136196680221188
13	Delaware	2014	125	935614	0.013360210514165
14	Maine	2014	174	1330089	0.0130818313661717
15	Oklahoma	2014	504	3878051	0.0129962189770067

In the results we are immediately presented with another example of West Virginia setting itself apart, quite extremely, its death percentage being 28% higher than the runner up, New Hampshire. The rest only vary quite closely to each other.

- **What states had the highest rate of opioid overdose deaths per 100,000 people, most frequently between 1999 and 2014?**

The SQL statement I designed for this question shows the states with repeat high deaths rates over the 15 year period.

```
Select Top 20
State, Year, MAX([Crude Rate]) AS HighestDeathRate
From PortfolioProjectA..OpioidDeaths
Group By State, Year
Order By HighestDeathRate DESC
```

	State	Year	HighestDeathRate
1	West Virginia	2011	32.1
2	West Virginia	2014	31.6
3	West Virginia	2013	27.9
4	West Virginia	2012	27.6
5	West Virginia	2010	26.7
6	New Hampshire	2014	22.8
7	West Virginia	2008	21
8	Rhode Island	2014	20
9	New Mexico	2014	19.9
10	Nevada	2011	19.5
11	Nevada	2010	19.3
12	West Virginia	2007	18.9
13	Ohio	2014	18.8
14	Nevada	2009	18.5
15	Rhode Island	2013	18.5
16	Nevada	2012	17.9
17	Kentucky	2014	17.3
18	Massachusetts	2014	17.2
19	New Mexico	2013	16.8
20	Nevada	2008	16.5

As can be seen in the results, West Virginia once again is set apart from the rest, occupying all of the top 5 instances as well as 2 others for a total of 7 instances. New Hampshire appears again, but only once, in the 6th instance. Nevada appears in 3 instances, New Mexico in 3, and Rhode Island in 2. All of these states appeared in the top 15 instances of the previous questions results. It's notable that Ohio, Massachusetts, and Kentucky, appeared in both as well.

- **Show states with the highest and lowest rates of opioid overdose deaths per 100,000 people in 2014.**

This question shows the same death rates, but for all states within 2014. For the sake of saving space, I've provided the highest rated 10 and the lowest rated 10.

```
Select Distinct State, Year, MAX([Crude Rate]) AS HighestDeathRate
From PortfolioProjectA..OpioidDeaths
WHERE Year = 2014
Group By State, Year
Order By HighestDeathRate DESC
```

State	Year	HighestDeathRate	State	Year	HighestDeathRate
West Virginia	2014	31.6	California	2014	5.6
New Hampshire	2014	22.8	Iowa	2014	5.5
Rhode Island	2014	20	Montana	2014	5.3
New Mexico	2014	19.9	Idaho	2014	4.9
Ohio	2014	18.8	North Dakota	2014	4.6
Kentucky	2014	17.3	Texas	2014	4.5
Massachusetts	2014	17.2	Hawaii	2014	4.4
Utah	2014	16	Mississippi	2014	3.9
Maryland	2014	15.6	South Dakota	2014	3.9
Connecticut	2014	14.8	Nebraska	2014	3.2

The results allow for identifying the states with most prevalent death rates, while also demonstrating states that have potential for increasing prevalence. With a clear list of states that are not an issue at all, a public health official may eliminate them from a watchlist, or perhaps inquire into what sets them apart. Evaluating the factors that allow them such a low rate of death may assist in implementing interventions in the states with high rates of death.

- **How many opioid prescriptions were prescribed in the United States in 2014?**

This question has clear limitations, as the dataset used does not account for all of the prescriptions in the United States. However, for the sake of the analysis, it's useful to know what the total number is in the dataset.

```
Select SUM(TotalPrescriptions) AS USTotalOpioidPrescriptions
FROM PortfolioProjectA..PrescriberInfo
```

	USTotalOpioidPrescriptions
1	2178995

- **Show prescriber information of those who issued the highest amount of prescriptions in 2014.**

The prescriber information that is most useful here is the state and speciality, as well as total amount of prescriptions. However, I also selected NPI, credentials, and gender.

```
Select NPI, Gender, State, Credentials, Specialty, TotalPrescriptions
FROM PortfolioProjectA..PrescriberInfo
WHERE TotalPrescriptions<>0
ORDER BY TotalPrescriptions DESC
```

	NPI	Gender	State	Credentials	Specialty	TotalPrescriptions
1	1831150630	M	Florida	M.D.	Interventional Pain Management	16189
2	1750364121	M	Louisiana	MD	Physical Medicine and Rehabilitation	14674
3	1215199609	M	Florida	DO	Anesthesiology	6205
4	1225069065	M	Louisiana	M.D.	Interventional Pain Management	6149
5	1053392571	M	Indiana	MD	Pain Management	5890
6	1043205305	M	Florida	MD	Anesthesiology	5079
7	1124098603	M	Mississippi	MD	Family Practice	5051
8	1962518019	F	Ohio	MD	Pain Management	4687
9	1427075506	M	North Carolina	M.D.	Pain Management	4653
10	1891776399	M	Florida	MD	Physical Medicine and Rehabilitation	4500
11	1982823225	M	Arkansas	MD	Pain Management	4313
12	1376552661	M	Florida	M.D.	Physical Medicine and Rehabilitation	4197
13	1518170927	F	Oklahoma	M.D.	Anesthesiology	3987
14	1821351438	M	North Carolina	FNP	Nurse Practitioner	3893
15	1174527972	M	Kansas	M.D.	Anesthesiology	3767
16	1831339902	M	New York	M.D.	Neurology	3731
17	1851368898	M	Oklahoma	MD	Pain Management	3722
18	1073500260	F	Washington	ARNP	Nurse Practitioner	3717
19	1558300558	M	New York	RPA-C	Physician Assistant	3645
20	1710177555	M	Texas	DO	Physical Medicine and Rehabilitation	3634

The top 6 instances that have prescribed the most amount of prescriptions are considerably more reasonable, as they are either in pain management, anesthesiology, or physical medicine and rehabilitation. It's still helpful, however, to see which states they reside in. The first top prescriber that would be of greater interest is a Family Practice MD in Mississippi, and the next would be an FNP in North Carolina. Other notable takeaways are the high recurrence of providers based in Florida within the top 20 instances.

- **How many prescriptions on average were prescribed per provider in each state in 2014?**

For this question I utilized a CTE in order to determine the states with the highest average amount of prescriptions per provider.

```
WITH ProvidersVSPrescriptions AS (
Select TOP 51
State, COUNT(NPI) AS ProvidersPerState, SUM(TotalPrescriptions) AS PrescriptionsPerState
FROM PortfolioProjectA..PrescriberInfo
GROUP BY State
ORDER BY PrescriptionsPerState DESC
)
Select State, ProvidersPerState, PrescriptionsPerState, (PrescriptionsPerState/ProvidersPerState) AS PrescriptionsPerProvider
From ProvidersVSPrescriptions
Order By PrescriptionsPerProvider DESC
```

	State	ProvidersPerState	PrescriptionsPerState	PrescriptionsPerProvider
1	Alabama	344	62126	180.598837209302
2	Louisiana	354	61434	173.542372881356
3	Mississippi	193	30361	157.310880829016
4	Arkansas	216	31623	146.402777777778
5	Tennessee	552	77072	139.623188405797
6	Indiana	533	71208	133.598499061914
7	Nevada	155	20673	133.374193548387
8	Oklahoma	281	35384	125.921708185053
9	Georgia	613	76990	125.595432300163
10	Wyoming	38	4653	122.447368421053
11	North Carolina	778	92136	118.426735218509
12	Kansas	203	23139	113.985221674877
13	Kentucky	367	41093	111.970027247956
14	Arizona	509	53886	105.866404715128
15	South Carolina	390	40834	104.702564102564
16	Missouri	483	49632	102.757763975155
17	Michigan	872	89257	102.358944954128
18	Florida	1570	156559	99.7191082802548
19	West Virginia	199	19801	99.5025125628141
20	Montana	77	7192	93.4025974025974

What instantly struck me as the most interesting about the results to this question were the geographic similarities shared by the top instances. The top 5 instances are all in the Southeast region of the United States, and they are all adjoining each other. Within the top 20 instances, 12 of the 13 states of the Southeast region are found, with the exception of Virginia. Additionally, 17 of the top 20 instances are all adjoining one another, 12 being in the Southeast region and 5 being in the Midwest.

States shown here that have proven to have a greater prevalence of prescriptions or deaths, or both, in previous questions include: Mississippi, Nevada, North Carolina, Kentucky, Florida, and West Virginia.

- **What is the rate of prescriptions in comparison to the population in each state in 2014?**

For this question I used a CTE and then joined both tables in order to determine the correlation between prescriptions and population.

```
WITH ProvidersVSPrescriptions AS (
Select TOP 51
State, COUNT(NPI) AS ProvidersPerState, SUM(TotalPrescriptions) AS PrescriptionsPerState
FROM PortfolioProjectA..PrescriberInfo
GROUP BY State
ORDER BY PrescriptionsPerState DESC
)
Select p.State, p.ProvidersPerState, p.PrescriptionsPerState, o.Year, o.Population,
(p.PrescriptionsPerState/o.Population)*100 AS PrescriptionToPopulationRatio
From ProvidersVSPrescriptions AS p JOIN PortfolioProjectA..OpioidDeaths AS o
ON p.State = o.State
Where o.Year=2014
Order By PrescriptionToPopulationRatio DESC
```

	State	ProvidersPerState	PrescriptionsPerState	Year	Population	PrescriptionToPopulationRatio
1	Louisiana	354	61434	2014	4649676	1.32125335184645
2	Alabama	344	62126	2014	4849377	1.28111301719788
3	Tennessee	552	77072	2014	6549352	1.17678817690666
4	Indiana	533	71208	2014	6596855	1.07942345253913
5	West Virginia	199	19801	2014	1850326	1.07013574905179
6	Arkansas	216	31623	2014	2966369	1.06605078464614
7	Mississippi	193	30361	2014	2994079	1.01403469981921
8	Kentucky	367	41093	2014	4413457	0.931084181855629
9	North Carolina	778	92136	2014	9943964	0.926552026938151
10	Oklahoma	281	35384	2014	3878051	0.912417087861918
11	Michigan	872	89257	2014	9909877	0.900687263827795
12	Maine	147	11621	2014	1330089	0.873700932794723
13	South Carolina	390	40834	2014	4832482	0.844990214138407
14	Missouri	483	49632	2014	6063589	0.81852513420682
15	Arizona	509	53886	2014	6731484	0.80050699073191
16	Kansas	203	23139	2014	2904021	0.796791758737282
17	Wyoming	38	4653	2014	584153	0.796537893325892
18	Pennsylvania	1211	101698	2014	12787209	0.795310376173565
19	Ohio	981	91597	2014	11594163	0.790026843679876
20	Florida	1570	156559	2014	19893297	0.786993729596457

The top 20 instances here are almost exactly the same as the previous question, with the only difference being Maine, Ohio, Pennsylvania entered the running, excluding Montana, Georgia, and Kansas. The unique concern here would be Maine, placing at 12 out of 50 states. This is noteworthy, given its lower population and number of providers in comparison to other states. Also noteworthy would be the infamous southern states appearing again in the top 10.