|  |
| --- |
|  |
| Opioid Crisis  Data Analysis  **ETL Group Project: Drew Seelig, Charles Biggar, Hayden Esman**  October 2019 |

Finding data, cleaning data, organizing data, establishing a data base with the proper structure, and being able to analyze data is extremely important in the modern world. Data is prevalent and in order to be able to drive meaningful results and solutions, data analysis can play a pivotal part.

This document will cover:

1. Executive Summary
2. Sources of Data
3. Data Dictionary
4. Data Cleaning
5. ERD diagram of our database
6. Our Table Schema
7. Five Queries Showing Analysis

# EXECUTIVE SUMMARY

**Opioids** are a major issue within the US. It is a leading cause of deaths. In order to understand this problem in more detail, we analyzed deaths in the US due to opioids from the CDC, where these deaths are occurring, who is prescribing opioids all in order to see if there are patterns that may lead to explanations of why this is occurring. The is a bold task, as there may be many environmental, social economic, and other factors that may be involved that were not taken into consideration as part of this analysis.

* **Largest State of Deaths:** Ohio, 161K over a 5-year period
* **Deaths by year:** Deaths are increasing year over year by about 10%, with 338K deaths in 2018 in the US
* **Region of country with most prescriptions:** CA, then FL which leads us to believe geography is not a leading indicator. Ohio is the 4th largest state for prescriptions with highest death count. CA and FL did not make the top 5 death count, also leading to a loose indicator of prescriptions to deaths.
* **Dr Type prescribing opioids that led to most deaths:** Male Drs are prescribing opioids 3x more than females. We don’t think this correlate to opioid deaths but rather just an interesting fact.
* **Gender of Dr that prescribed opioids that led to most deaths:** Family practice Drs are prescribing opioids at the largest rate with ~600K prescriptions. The CDC should conduct a study of why these prescriptions were made and if they were absolutely necessary.

The following describes the procedures we took through this exercise:

1. We looked into data related to opioid deaths
2. We found a data set that describes prescribed medicine, deaths by state, what type of DR. can prescribe opioids, and then found deaths by state by opioid
3. We then cleaned the data by ensuring that we can link by drug name, there were no null values, removed dupes, removed attributes that were not applicable to our data, cleaned like names, changed case of names.
4. We then looked at the data attributes and started to think about the appropriate tables and primary keys to link the data.

# SOURCES OF DATA

In order to conduct analysis of opioid deaths, we pulled data from the following three sources:

* **U.S. Opiate Prescriptions/Overdoses:** <https://www.kaggle.com/apryor6/us-opiate-prescriptions#overdoses.csv>
* **Drugs prescriptions with providers profile:** <https://www.kaggle.com/tajuddinkh/drugs-prescriptions-with-providers>
* **12 Month-ending Provisional Number of Drug Overdose Deaths from CDC:** <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm>

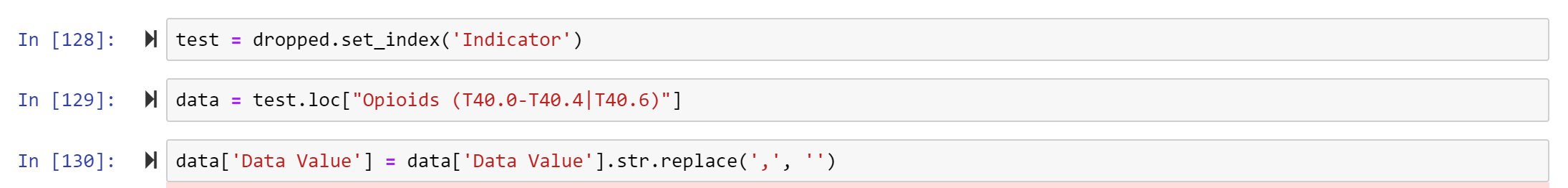
# DATA DICTIONARY

The data dictionary is important to understand the data we collected.



# DATA CLEANING

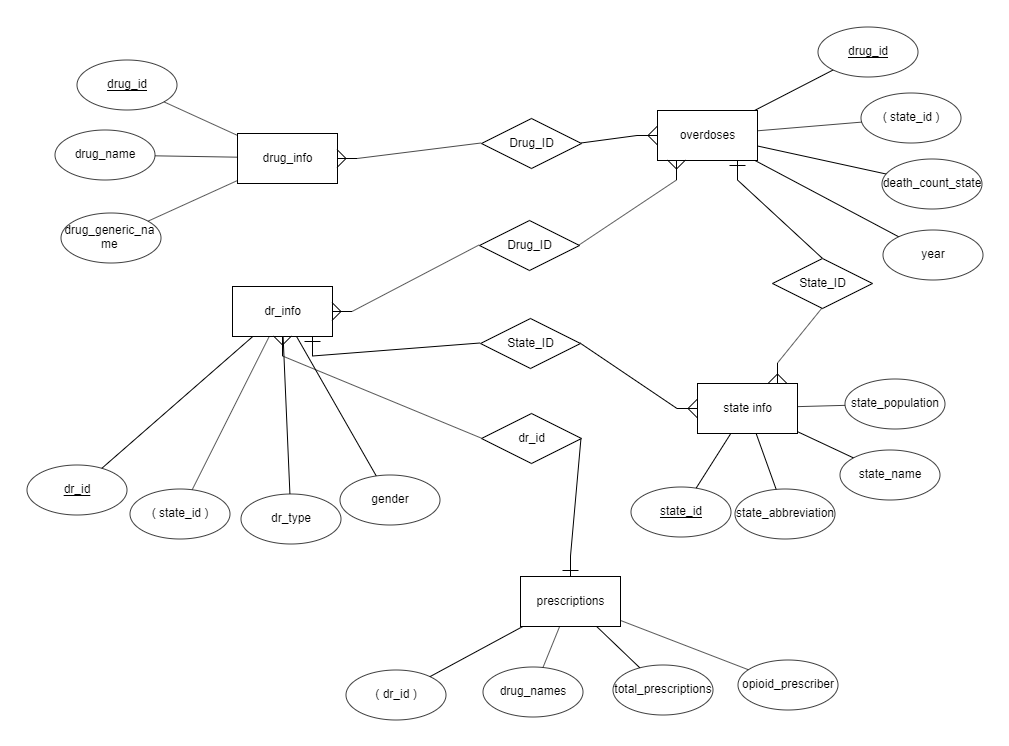
Naturally the data we collected was not clean. There was erroneous data, double counting, missing data, etc. We took many channels to clean the data, including dropping, normalizing, and combining to be able to produce results.



# 

# ERD DIAGRAM

The ERD diagram is a type of structural diagram for use in database design. This shows how we designed our database to study the opioid crisis.



# TABLE SCHEMA

-- ETL table schema

--Opioid Table

CREATE TABLE opioids (

drug\_id INT PRIMARY KEY,

drug\_name VARCHAR(30),

generic\_name VARCHAR (30)

);

-- State Table

CREATE TABLE states (

state\_id INT PRIMARY KEY,

state\_abbv VARCHAR(30),

state\_name VARCHAR(30),

state\_pop INT

);

-- Overdoses Table

CREATE TABLE overdoses (

state\_id INT,

year INT,

deaths INT,

FOREIGN KEY (state\_id) REFERENCES states(state\_id)

);

-- Doctor Table

CREATE TABLE prescribers (

doctor\_id INT PRIMARY KEY,

doctor\_type VARCHAR (100),

state\_id INT,

gender VARCHAR (1),

FOREIGN KEY (state\_id) REFERENCES states(state\_id)

);

-- Prescriptions Table

CREATE TABLE prescriptions (

doctor\_id INT,

ACETAMINOPHEN\_CODEINE INT,

DIPHENOXYLATE\_ATROPINE INT,

FENTANYL INT,

HYDROCODONE\_ACETAMINOPHEN INT,

HYDROMORPHONE\_HCL INT,

METHADONE\_HCL INT,

MORPHINE\_SULFATE INT,

MORPHINE\_SULFATE\_ER INT,

OXYCODONE\_ACETAMINOPHEN INT,

OXYCODONE\_HCL INT,

OXYCONTIN INT,

TRAMADOL\_HCL INT,

Total\_Opioid\_Prescriptions INT,

Opioid\_Prescriber VARCHAR (3),

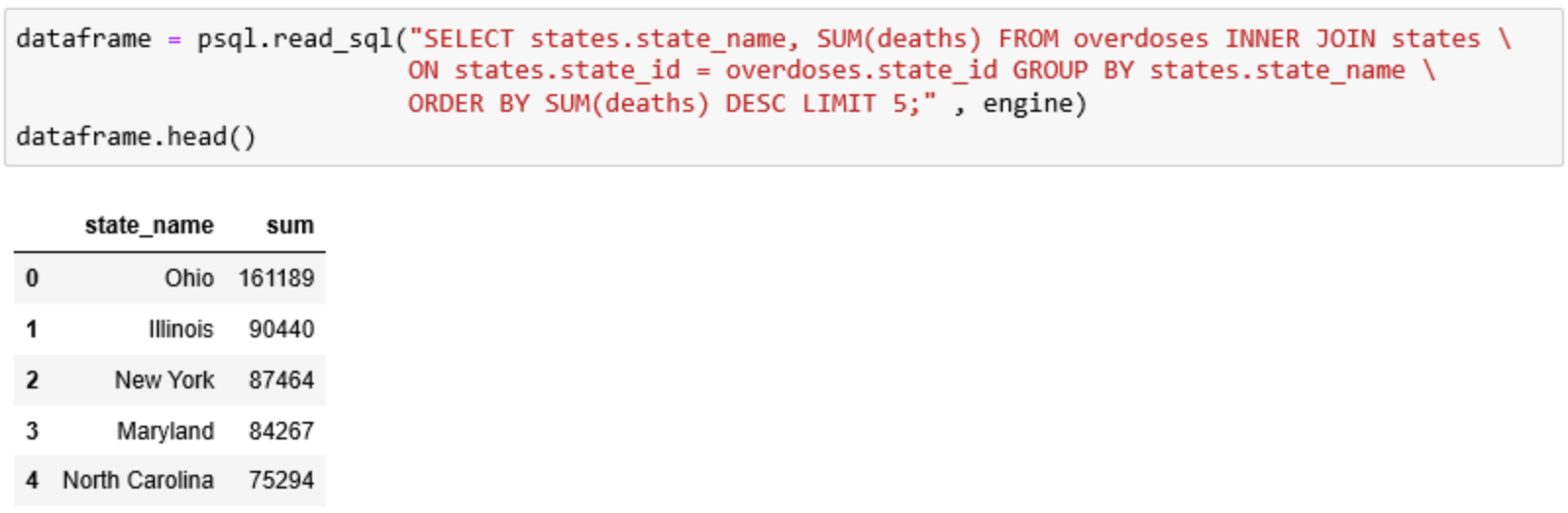
FOREIGN KEY (doctor\_id) REFERENCES prescribers(doctor\_id)

);

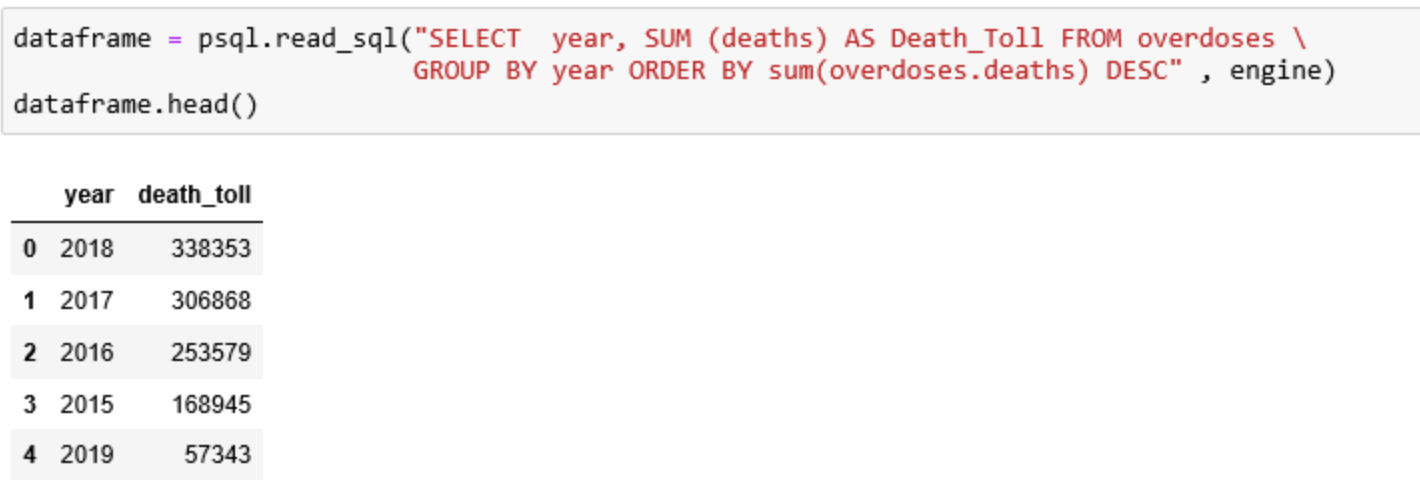
# QUERIES & RESULTS

Here are our queries and results:

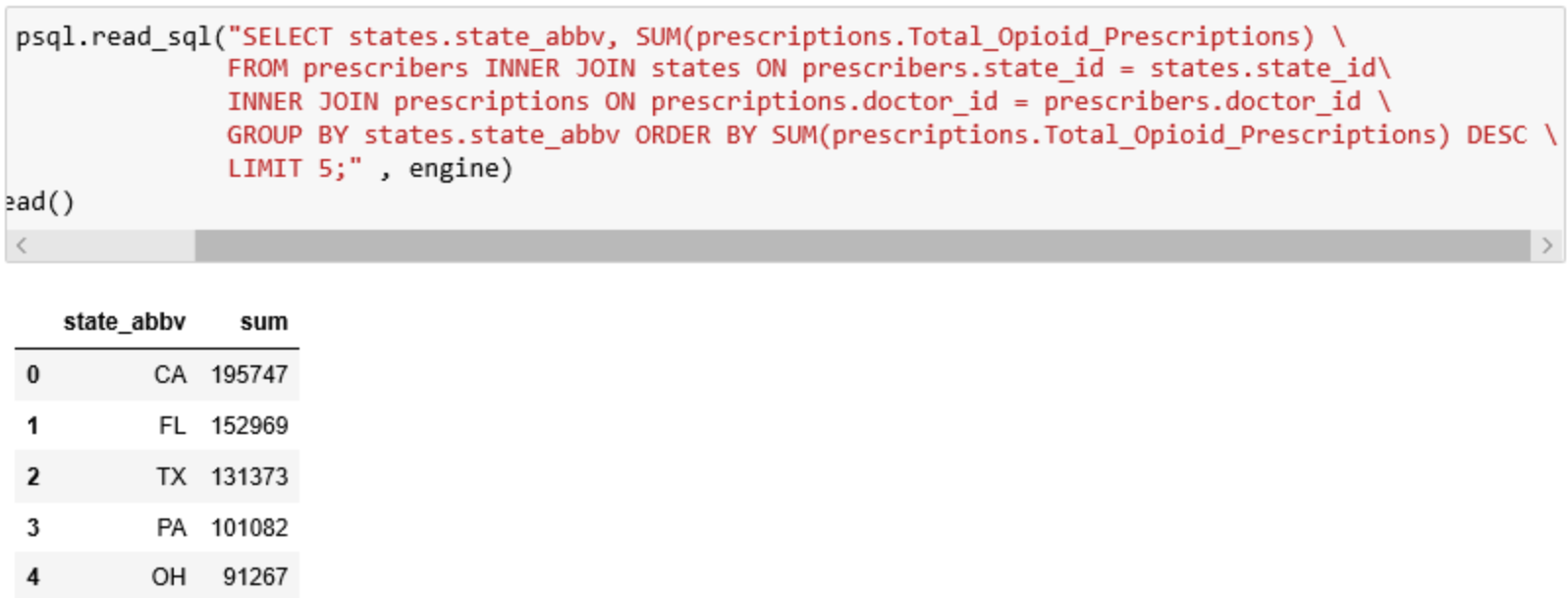
1. **Top 5 states of opioid deaths:**



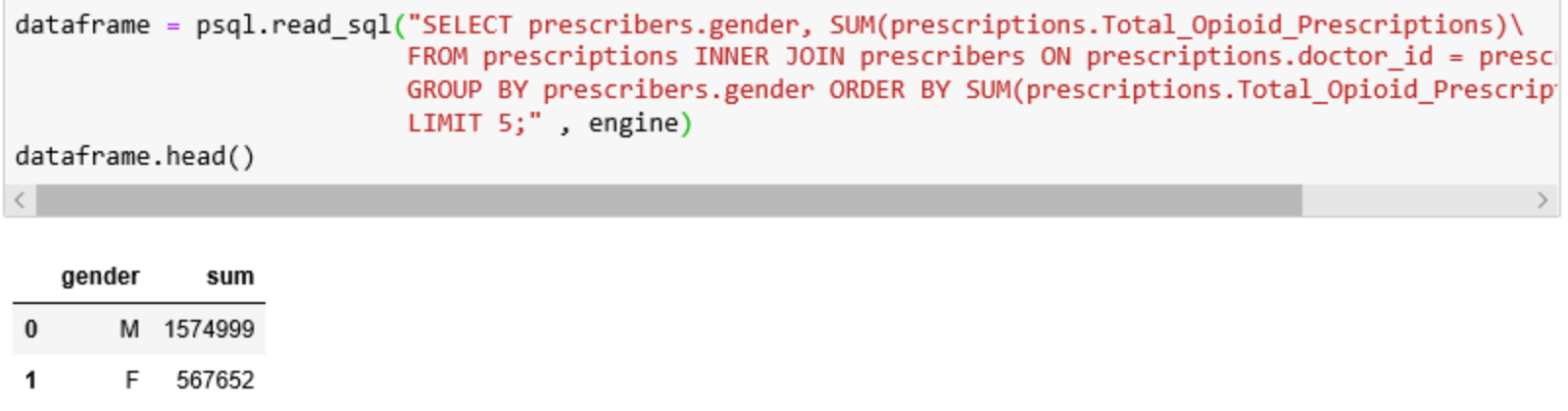
1. **Deaths by year:**



1. **States with most prescriptions:**



1. **Dr Gender of prescriptions:**



1. **Dr Type of prescriptions:**

