**Team:** Leah Barbour, Andrew Cummings, Dinesh Nayak

For our project we were interested in the relationship between the weather and alcohol sales. Specifically, we wanted to look at a time period and compare the average temperature in the US and the sales of alcohol in the US. For the Extract portion, we found 2 CSVs online

* one that had monthly alcohol sales data in the USA for the last 30 years (<https://fred.stlouisfed.org/series/S4248SM144NCEN>)
* and one that had monthly average temperatures in the USA starting in 2000 (<https://www.ncdc.noaa.gov/cag/national/time-series/110/tavg/1/12/2000-2018?base_prd=true&begbaseyear=1901&endbaseyear=2000>).

To transform the data, we loaded the CSVs into python and used pandas to look at the data in dataframes. We determined that the first 4 rows in the weather data were irrelevant so we dropped those rows. Likewise the data in the last row was irrelevant so it was dropped as well. For both the tables we elected to change the column names to something more sensible that reflected the data (we changed the columns "DATE" to "MonthRecorded" and "S4248SM144NCEN" to "MonthlySales" in the alcohol table and changed "Contiguous U.S. Average Temperature" to "MonthTemp" and "Unnamed: 1" to "AvgTemp" in the weather table).

Since our data was related, we chose to store our data in a relational database. We created a weather\_db in Postgres and loaded our two dataframes into two separate tables:

* The temperature table temperature with the columns id SERIAL PRIMARY KEY, MonthTemp INT, and AvgTemp DOUBLE PRECISION
* And the drinking table with id SERIAL PRIMARY KEY, MonthRecorded DATE, and MonthlySales INT.

Because the date format in the two tables were different we had to transform the data using SQL. We used SQL to get the dates to match in a YYYY-MM-DD format. After this we joined the two tables together on the date and created a new third table. This new table further cleaned the data by dropping 97 rows that we present in the drinking table but not the temperature table.