

Week 5 Assignment

Week 5 assignment is due end of day on Sunday March 1st. You may work in a small group on the assignment. We'll look together at some of the most interesting student solutions in our Meet up on Tuesday March 3rd and/or our office hours on Wednesday March 4th.



It's quite common when analyzing data to pull data from SQL tables into an R dataframe (or a Python dataframe using pandas), and then perform downstream data cleansing and analysis.

Your task this week is to create a process that pulls information from the flights database into an R dataframe. There are a number of acceptable ways to do this, but you must provide the code in SQL and/or R so that I can reproduce your process in my environment.

Here is the information that you should load into your R data frame:

For flights from any of the three major NYC airports to Los Angeles Intl airport during the period between 2/23/2013 and 3/1/2013, please include the following information:

- Airport of origin
- Carrier
- Approximate temperature at scheduled time of departure
- Departure Delay
- Arrival Delay
- Air Time in Minutes
- Seating Capacity of Airplane

Probably the most straightforward approach (and 100% acceptable) is to create a .CSV file that contains all of the required information from SQL in a single result set, then loads this .CSV file into an R dataframe. There are other solutions, where all of the code is written in R.