Project

End to end data science project

3 homework assignments

Project proposal

Project presentation

Jupyter notebook final report

Can be data science for decision making, learning for science, etc

Project Ideas

1. 30 minutes or less!
   1. What airport to choose with the least chance of delays for airports within 30 minutes of each other?
      1. Kaggle Flight Delays https://www.kaggle.com/usdot/flight-delays
      2. Kaggle Airline delay data, <https://www.kaggle.com/giovamata/airlinedelaycauses>
      3. <http://stat-computing.org/dataexpo/2009/the-data.html>
      4. U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS)
   2. Scope
      1. US domestic arrivals only
      2. No more than 10 years of delay history
      3. Limit to select list of cities?
   3. Tasks
      1. Find historical datasets of airline delay data
      2. Clean and correlate the datasets
      3. Manually look at one set of airports
         1. Chicago
         2. New York
      4. Algorithm to find airport within a specified distance of somewhere
   4. Human centered concerns
      1. Help people save time waiting at airports
      2. Improve airport efficiency, aircraft system efficiency
      3. Definition of a delay may differ between people
      4. Final interface
         1. Pick city or airport?
         2. Maximum distance from airport to city?