

Solar Cells? Parking Tickets? Plane Flights? My partner, Sean Hall, and I (Julia Baratta) are considering working with one of the three datasets attached to this GitHub repository as .csv files.

Our first data set discusses domestic flight delays from different airports and carriers. Analysis of trends within this data set can help us determine the likelihood that a given flight will be delayed. This data is interesting because flight delay is a very common nuisance in the US. Although it is typically thought of as relatively random by most Americans, the data proves otherwise. We hypothesize that factors such as the type of airline carrier one uses, weather along the route, season, and time of the day all affect the probability of a flight being delayed. Analyzing this data set could prove useful for flyers when deciding on important flights. Additionally, this project could be taken a step further by predicting how likely it is that your checked luggage will be lost by the airline. This information could be used by passengers who are deciding between checking a bag and carrying on, or passengers who aren't sure if they should pack valuables in their checked bags.

The second table contains data on the net generation of energy from photovoltaic cells in the United States. Within this dataset, PV cell generation is broken down into the electrical power, commercial, industrial, and residential sectors. From this, we can determine which states are producing the most energy via photovoltaics, and how that has changed over the years. We can also isolate the breakdown of generation by sector within each state and see how this varies. This data can also be compared to a timeline of important breakthroughs in PVs, funding for clean energy, pricing of raw materials, political affiliation of the region, and average cloud cover in the state.

Our third dataset contains information about parking violations in major cities in North America. From this data, we can determine how variables like location, time of day, and car color and type affect the probability of getting a parking ticket. Analysis of this data can help isolate the intensity with which certain cities enforce parking violations, and even provide recommendations for where to park when all the open spots are taken. Additionally, learning about how make, model, year, and color of car affect frequency of parking tickets can help inform car buyers who often use street parking. This data could be applied further to the creation of an app that will tell a user where not to park based on a dataset of millions of parking violations. This app could also be updated by the user based on their successful and unsuccessful parking experiences, allowing the app to adjust its recommendations as newer, more accurate data is collected.

Collaboration Plan: We are going to set up a GitHub Repository to edit and share code. We will meet once a week on Thursdays from 12:30 pm - 2:00 pm. We are neighbors so we can meet more frequently if needed.