Capstone Project 1: Project Proposal

What is the problem you want to solve?

In 2015 there are lots of flight delays in the United States. I am going to predict what causes flight delays and Cancellations?

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

My client in this situation could be a range of individuals includes Airlines Passengers, The Department of Transportation/National Aviation System.
For Airlines Passenger, no one enjoys spending extra time at the airport because flight delays have kept their plane grounded. Based on the analysis passengers will be more mindful as to what airlines have the largest and lowest delays and cancellations, what days of the week are the best to flight and even what airports they should avoid.

The Department of Transportation(DOT)/National Aviation System implemented regulations to reduce delays and

provide more protection for consumers. As a result, most U.S. airlines are now prohibited from allowing a domestic flight to remain on the tarmac for more than three hours. Carriers are also required to notify passengers of known delays and provide accommodations.

What data are you using? How will you acquire the data?

I am using this free dataset from Kaggle: 2015 Flight Delays and Cancellations: Which airline should you fly on to avoid significant delays? It can be found @ https://www.kaggle.com/usdot/flight-delays
This data has 3 sources
airlines.csv-2 columns
airports.csv-7 columns
flights.csv-31 columns

Briefly outline how you'll solve this problem.

I am going to solve the problem by following these steps

- Explore and Clean Data: Using Data-wrangling With Pandas
- Finding Missing Values
- Filtering Data

- Sorting
- Merge & concatenation
- Build Visualizations(using matplotlib and seaborn)
- Explore linear regression with different regularization method
- Use Machine learning algorithms

What are your deliverables?

Code & Paper