**Project 3 Proposal**

Group:

* Andrew Anastasiades – Repo Management, Flask, API
* Corey Anderson – Data Cleaning, Model
* Valerie Pippenger – Data Visualizations
* Alex Gainer – Data Cleaning, Heroku, Database
* Diana Kennen – Front End
* Rodney Daverman – html/css

Topic: Online Airline Arrivals

Data Sources:

* Bureau of Transportation Statistics
* Airline Service Quality Performance 234 (On-Time performance data)
  + Pre-Pandemic: 2015 -2019
  + Post-Pandemic: 2020-2021
  + <https://www.bts.gov/browse-statistical-products-and-data/bts-publications/airline-service-quality-performance-234-time>
  + <https://esubmit.rita.dot.gov/ViewReports.aspx>
* 23 million rows

Research Question:

* Can delays and cancellations be predicted based upon? Example features:
  + Departure city
  + Time of day
  + Day of the week
  + Airline
  + Tail number

Database:

* postgreSQL

Data Visualization:

* Dashboard
  + Probability of a delayed departure based upon flight characteristics such as:
    - Departure city
    - Time of day
    - Weekday
    - Airline
* Visualizations (D3.js, plotly, leaflet.js)
  + Summary statistics by
    - Airline
    - Time of day
    - Weekday
    - Delay cause
  + Bubble map of flight delays??
  + Map of airports or routes?

Modeling/Predictions:

* Neural Network or Deep Neural Network
* Logistic Regression with LASSO
* GridSearchCV to tune hyperparameters
* Model diagnostics: AUC & accuracy

Possible Cool Things

* H2o autoML
* Spark – big data