Machine Learning- Team 3

Members: Beth Emborsky, Brodie Armstrong, Ellen Rud

Topic: Covid-19 Hot Spot Prediction and Indicators at U.S. County Level

**Datasets:**

* Daily Cases and Deaths by County & hotspots:
  + https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html#hotspots
* Mobility Data:
  + <https://www.google.com/covid19/mobility/>
  + <https://docs.safegraph.com/docs/social-distancing-metrics>
  + <https://www.cuebiq.com/visitation-insights-covid19/>
* Population Dynamics: Gender, race, income,
  + U.S. Census or Kaggle or other dataset TBD
  + Area of Counties for Pop Density:
* TBD:
  + School Closures: ?
  + Events: ?
  + Mobility: Uber Movement
  + Testing Data
  + Healthcare Access
  + Tourism Revenue: Bureau of Labor Statistics

Rough Timeline:

Presentation is Tues 20th

1. Determine all sources – Tomorrow
2. Mock Up of End MVP- Ellen tomorrow/Sat
3. ERD/ Database Design- PostGreSQL- Beth- Sat Rough Draft
4. Manipulate and Clean Source Data- Google Colabs- Tuesday – Divide and Conquer TBD
5. Create Database-AWS or Heroku - Wed 15
6. Middle Layer: Data Pulls and Flask App – Brodie Wed 15
7. Machine learning:
   1. Random Forest to rank the parameters to hotspots
   2. Regression/ maybe Neural Network predict County Hotspots over time-Ellen/Beth
8. Front End: html, JavaScript, D3.js, bootstrap, CSS, - Ellen- rough draft: Tuesday 14