**COVID Killers**

06.06.2020 Team Meeting Proposed Agenda:

1. **Review/Revise Segment Roles and Assignments**
2. **Review Segment Rubric Deliverables and capture actions (20.2.3 as guide)**
   1. Presentation DRAFT
   2. GitHub Repository
      1. README add – Communications Protocols
      2. 4 commits per team member for segment, min 8 commits total
   3. Machine Learning
      1. README and ML adds
         1. Describe data preprocessing
         2. Describe preliminary features, and selection
         3. Describe how data was split
         4. Explain Choice of Model – Random Forest
   4. Database
      1. MUST include at least 1 join in database (not Pandas)
   5. Dashboard
      1. Storyboard in Google Slides
      2. Describe tools used
      3. Describe interactivity
3. **Review and assign open actions**
4. **Agree on next meeting time/date**

**Open Actions:**

**README Inputs:**

**Gray Text - Already included (shown ONLY for order of information)**

Selected Topic: COVIDE-19 Cases in California

Reason for selection: The spread of COVID-19 has changed society, and everyone globally is searching for answers. Most prominently people want to go what is going to happen in the near future, will the virus continue to spread or will it subside and be controlled.

Description of Source data: Multiple sources of data will be transformed and combined to help us create future predictions for COVID cases

* COVID-19 daily cases by State and County
* California county demographics data; ethnicity, income, gender, and age

**Question Hoped to Answer:** Christina will submit a new description such as: The COVID Killers team will predict likelihood to contract COVID-19 symptoms and whether those symptoms will be severe/fatal based on individual demographics.

GitHub:

GitHub is being used to allow the team to collaborate and share progress on the project. A Project Repository has been created with a master branch and at least 1 individual branch for each team member.

Machine Learning Model:

Pandas will be used to clean the data and perform exploratory analysis. SciKitLearn is the Machine Learning library we'll be using to create a classifier, encode data, split data in to test and train data sets, scale the data, and test varying machine learning models for making future data predictions. Further analysis will be completed using Python. A random forest machine learning model will be created to determine the probability of individuals contracting COVID-19, which draw data from a Postgres database.

Database:

A Postgres database will hold the data used to predict COVID-19 impact by for individuals. We will integrate data from the database into our machine learning model. Also, the data tables will be exported as csv files to use in Tableau to display the data in a dashboard.

Dashboard:

Tableau will be used to display the data in a fully functioning interactive dashboard. It will be hosted on <https://public.tableau.com>. Users may visit the Tableau dashboard to view their likelihood of contracting COVID-19 based on their individual data; 1. county of residence, 2. age, 3. gender, and 4. income.

Presentation:

A Google Slides Presentation providing a comprehensive review of the COVID Killers project including:

* Selected topic
* Reason why they selected their topic
* Description of their source of data
* Questions they hope to answer with the data
* Description of the data exploration phase of the project
* Description of the analysis phase of the project
* Technologies, languages, tools, and algorithms used throughout the project
* Result of analysis
* Recommendation for future analysis
* Opportunities for improvement and proposed changes to the project approach

**End of README additions**

**TEAM Assignments – Segment 1 and Proposed Segment 2 (Not Final)**

