Questions:

1. Given prerace information, and race information perceived by the officer during a stop (race), to what extent can we predict stop likelihood?
   1. Do we have date frames for this question?:
      1. Yes, mostly; we have data frames containing:
         1. count of stops by prerace (**race\_predictions** [filtering NAs and Unknowns])
         2. count of stops by race (**race\_count)**
         3. **Do we need more for this question, or does it need to be reformulated?**
   2. What kind of visualization(s) should we use for this question?
      1. Bar chart for count of stops by race
   3. Do we have matplotlib visualization(s) for this question?
      1. One so far
   4. Do we have PNG visualizations?
   5. Have we analyzed the data for this question, for inclusion in the write-up?
2. Does time of day, binned into four categories (morning, afternoon, evening, and night), influence likelihood of a police stop?
   1. Do we have date frames for this question?:
      1. Yes; listed as Question 4 in the main branch, there are data frames containing:
         1. stops grouped into 4 time bins (**time\_of\_day**)
         2. time bins and precinct (**timeDF**)
         3. time bin, problem type, and problem count (**timeDF2**)
         4. time bin, race, and race count (**timeDF3**)
         5. time bin, problem type and by race (**timeDF4**)
   2. What kind of visualization(s) should we use for this question?
      1. ???
   3. Do we have matplotlib visualization(s) for this question?
      1. ???
   4. Do we have PNG visualizations?
   5. Have we analyzed the data for this question, for inclusion in the write-up?
3. Do specific precincts influence the likelihood of a police stop for individuals of a particular race & gender? (Given all stops within a precinct's boundaries and the subject's gender and race, can we predict what precincts are more likely to pull over a particular gender/race?) [correlation heatmap?] a. If time allows, normalize results by precinct race and gender from census data
   1. Do we have date frames for this question?:
      1. Yes; in main branch, we have data frames listing:
         1. stops per precinct by race (**stops\_per\_precinct\_by\_race**)
         2. precinct stops by gender (**precincts\_sex**).
         3. Likelihood of race matching by precinct (**precint\_match\_df**)
   2. What kind of visualization(s) should we use for this question?
      1. ???
   3. Do we have matplotlib visualization(s) for this question?
      1. ???
   4. Do we have PNG visualizations?
   5. Have we analyzed the data for this question, for inclusion in the write-up?
4. What is the likelihood of prerace matching actual race, for each race? Note: prerace can be reported by parties other than the officer, such as a citizen calling them in, or a police dispatcher relaying information
   1. Do we have date frames for this question?:
      1. yes; we have dataframes in Emanshu’s branch with:
         1. total match rate for prerace and race (**race\_predictions**)
         2. count of prerace for each race (**pre\_race\_count**)
         3. **we need a dataframe with a comparison between prerace and race, by race**
   2. What kind of visualization(s) should we use for this question?
      1. ???
   3. Do we have matplotlib visualization(s) for this question?
      1. ???
   4. Do we have PNG visualizations?
   5. Have we analyzed the data for this question, for inclusion in the write-up?
5. Does the type of police stop impact whether a person is searched? H1 If suspicious vehicle stops are related to a person being searched, then the percentage of searches will be higher than the regular police stops for the year; H0 If suspicious vehicle stops are not related to a person being searched, then the percentage of searches will be approximately equal for the year.
   1. Do we have date frames for this question?:
      1. no; may not be necessary depending on progress of other questions
         1. ?
   2. What kind of visualization(s) should we use for this question?
      1. ???
   3. Do we have matplotlib visualization(s) for this question?
      1. ???
   4. Do we have PNG visualizations?
   5. Have we analyzed the data for this question, for inclusion in the write-up?

Writeup Status: ?

Presentation Status: Outline posted on github

Presentation Requirements:

* Be at least 8-10 min. long
* Describe the core message or hypothesis for your project.
* Describe the questions you and your group found interesting, and what motivated you to answer them
* Summarize where and how you found the data you used to answer these questions
* Describe the data exploration and cleanup process (accompanied by your Jupyter Notebook)
* Describe the analysis process (accompanied by your Jupyter Notebook)
* Summarize your conclusions. This should include a numerical summary (i.e., what data did your analysis yield), as well as visualizations of that summary (plots of the final analysis data)
* Discuss the implications of your findings. This is where you get to have an open-ended discussion about what your findings "mean".
* Tell a good story! Storytelling through data analysis is no different than in literature. Find your narrative and use your analysis and visualization skills to highlight conflict and resolution in your data.

Presentation Guidance:

You are free to structure your presentations to your liking, but students tend to have success with the following format.

* Title Slide
  + Include the name of the Project and Group Members
* Motivation & Summary Slide
  + Define the core message or hypothesis of your project.
  + Describe the questions you asked, and why you asked them
  + Describe whether you were able to answer these questions to your satisfaction, and briefly summarize your findings
* Questions & Data
  + Elaborate on the questions you asked, describing what kinds of data you needed to answer them, and where you found it
* Data Cleanup & Exploration
  + Describe the exploration and cleanup process
  + Discuss insights you had while exploring the data that you didn't anticipate
  + Discuss any problems that arose after exploring the data, and how you resolved them
  + Present and discuss interesting figures developed during exploration, ideally with the help of Jupyter Notebook
* Data Analysis
  + Discuss the steps you took to analyze the data and answer each question you asked in your proposal
  + Present and discuss interesting figures developed during analysis, ideally with the help of Jupyter Notebook
* Discussion
  + Discuss your findings. Did you find what you expected to find? If not, why not? What inferences or general conclusions can you draw from your analysis?
* Post Mortem
  + Discuss any difficulties that arose, and how you dealt with them
  + Discuss any additional questions that came up, but which you didn't have time to answer: What would you research next, if you had two more weeks?
* Questions
  + Open-floor Q&A with the audience