Project Ideas for Capstone Three

1. At what age should UFC fighters retire?
   1. My childhood hero, Anderson Silva, once reigned supreme in the middleweight division of the UFC. At one point, he was unstoppable, and he is still regarded by many as the greatest UFC fighter of all time. However, after his loss against Chris Weidman in 2013, many called on him to retire. Silva is now 45 and still fighting, in a sport where many fighters retire before the age of 38. Unfortunately, he has not been the same since his 2013 loss. After Weidman ended Silva’s UFC record for longest title reign and longest consecutive UFC wins, Silva has now won only 1 out of his last 8 fights. The issue of Silva’s lackluster return has always made me curious: How important is age when it comes to a UFC fighter’s performance? Is an age at which performance markedly decreases and fighters should safely retire?
   2. My dataset will come from a completed Kaggle dataset on UFC Fight Historical Data from 1993-2019: <https://www.kaggle.com/rajeevw/ufcdata>
2. When it comes to parking citations in Los Angeles, are certain areas more prone to receiving parking citations than others?
   1. Anyone who visits or lives in Los Angeles can tell you that parking there is a nightmare. Anytime I plan a trip there, I have to prepare to arrive 30 minutes in advance just to find some parking. Not only are spots crowded, many streets are filled with confusing signs that limit or restrict parking to specific times throughout the day. It is so easy to get confused by these signs that it is estimated that the city of Los Angeles brought in over $250 million in 2013 from parking tickets alone. I will use data to divide the city of Los Angeles into distinct sections and spot any patterns in parking citation behavior. Perhaps some areas are more prone to parking citations than others.
   2. My dataset will come from a compiled Kaggle dataset on Los Angeles parking citations: <https://www.kaggle.com/cityofLA/los-angeles-parking-citations?select=parking-citations.csv>
3. Can NLP predict whether a blog was written by a male or female?
   1. The stereotypes of differences between men and women are countless. Many, like the claim that women talk more than men, have been disproven by dozens of studies. However, the topic of word choice between the sexes has been a lot more nuanced. Studies have found that men speak more abstractly, while women were more focused on the details. However, these are qualitative conclusions drawn by human researchers. Is it possible, then, that a machine learning algorithm might be able to quantitatively detect these differences and use them to predict whether a new blog was written by a man or woman? Such a prediction may be difficult to accomplish, so there would be an equal, if not greater emphasis on the exploratory data analysis section of this project. I expect to derive some curious insights from my dataset, which contains a list of blogs, blogger IDs and other features, and the text of their blogs. I hope to design an aesthetically pleasing word cloud in the analysis section, as well as some histograms that will help gain a greater understanding of the data.
   2. Dataset used will come from The Blog Authorship Corpus: https://u.cs.biu.ac.il/~koppel/BlogCorpus.htm