Capstone Project Ideas – David Kile

1) **US Election 2016 Review – Critical Analysis of US Polling Methodology and Proposal of New Prediction Methods**

   Client: News organization (e.g., fivethirtyeight.com), political consulting organizations, political parties.

   Datasets used: Kaggle election database (https://www.kaggle.com/benhamner/2016-us-election), publicly available polling data, alternative prediction data (e.g., google search results data).

   Research questions: Why were the 2016 US election polls so consistently inaccurate in comparison to the actual election outcomes? What methods differentiated the more accurate polls and models from those which were inaccurate? Are there patterns in the data that could be used to predict future inaccuracy? What biases contributed to these inaccuracies and how can these be avoided in future elections? How did this relate to the general inaccuracy of election coverage considering Trump's unexpected win? Furthermore, what methods other than traditional polling could be used in the future to corroborate or challenge polling results?

   Data product: Algorithm which provides an alternative method of predicting election results to corroborate or challenge the results of political polling (e.g., replication and refinement of Google search result analysis methods to predict election outcomes).

2) **US Election 2016 Review – Hidden Demographics and Campaign Strategy Effectiveness**

   Client: News organizations (e.g., fivethirtyeight.com), political consulting organizations, political parties.

   Datasets used: Kaggle election database (https://www.kaggle.com/benhamner/2016-us-election), publicly available demographic data (e.g., US census), polling data.

   Research questions: Which demographics (based on education level, geographic area, race, religion, etc.) were most strongly correlated with results for different candidates and political parties? What demographic groups have not been adequately studied or even identified, and can these explain unexpected demographic trends in the election outcome? Were these demographic groups effectively targeted by campaign resources (canvassing, etc.)? More generally, what story does this data tell about the current state of America's political landscape?

   Data product: Algorithm which analyzes demographic data and election results to identify statistically distinct voting demographics, with analysis focused on the relationship between these data and the success or failure of campaign activities.

3) **Analysis of 2015 U.S. Traffic Fatality Data**

   Client: U.S. Department of Transportation

   Datasets used: Fatality Analysis Reporting System (FARS) Database.

   Research questions: In August 2016, the U.S. Department of Transportation issued a call to action after finding that traffic fatalities increased by 7.2% from 2014 to 2015. The dept. posed several broad questions which they would like outside organizations and individuals to tackle. Two questions stood out to me as good areas to investigate further: What models can we develop to identify communities that might be at a higher risk for fatal crashes? How might we use studies of attitudes toward speeding, distracted driving, and seat belt use to better target marketing and behavioral change

campaigns? Details about the government call to action can be found at https://www.transportation.gov/fastlane/2015-traffic-fatalities-data-has-just-been-released-call-action-download-and-analyze.

Data product: Predictive model which identifies communities that are at a higher risk of fatal crashes, or which predicts the effectiveness of future behavioral change campaigns.