Initial Capstone Proposals

1. Traffic accidents--Can we target the roads on which fatal accidents are most likely to occur? At which times of day or year on those roads are such accidents more likely to occur? How often is alcohol impairment a factor? How often is weather a factor (cloudy, foggy, raining, snowing, icy roads, etc.)? What types of vehicle are most commonly involved (small cars, trucks, tractor trailers, motorcycles, etc.)?

One possible source is the 2015 Traffic Fatalities provided by NHTSA available here: https://www.kaggle.com/nhtsa/2015-traffic-fatalities.

This could be used to predict fatalities by time of day (binary early or late, for example) and time of year (Winter, Spring, Summer or Fall—factors could be created based on month of year given), and other correlative factors. This topic would be of interest to law enforcement agencies, as they can make the best choices in deploying resources in those areas and at those times preemptively (i.e., putting speed traps or DUI checkpoints in those areas and shortly before those times when traffic fatalities are likely to occur in an attempt to encourage safe driving behavior) and to ensure that resources are available to quickly respond at those times. It would also be of interest to auto insurance companies of individuals that live in those areas in setting their rates.

2.   Tech Stock Prediction—Can direction of technology stocks (i.e. hardware-producing companies such as IBM, Intel, etc.) be predicted based on performance of raw materials commodities (gold, platinum, etc.)? Additionally, how sensitive are they to changes in interest rates? What types of movements on the part of the metals would indicate a good buy or sell signal? Similarly, how might interest rate movements produce good buy or sell signals here? Are these relationships apparent over an annual period, a five year period and a ten year period or are they only apparent on certain time frames? This topic would be of interest to both individual and institutional investors as well as to the tech firms themselves.

Data can be obtained through Quandl and Yahoo Finance. For example, one could go to https://finance.yahoo.com/u/yahoo-finance/watchlists/tech-stocks-that-move-the-market to obtain a list of the top tech stocks. Then choose 4 or 5 (those that are hardware, rather than purely software oriented) such as Apple, Microsoft, Intel, NVidia and AMD, obtain and download their individual financial histories, focusing on the Adjusted Close column for each and combining each of these columns into a single dataset.

Metals data could be obtained through Quandl and the CME through the following link: https://www.quandl.com/data/CME-Chicago-Mercantile-Exchange-Futures-Data   where additional querying may be done to bring up data for different metals. (Each metal dataset apparently covers only one month of each year, so several would have to be obtained for each metal and joined appropriately, likely using the "Last" column there, analogous to Close in Yahoo Finance.) Proposed metals to use for comparison would be gold, platinum, palladium and copper.

Finally, interest rate (LIBOR) data could be obtained from FRED at the following link: https://fred.stlouisfed.org/categories/33003?t=1-year%3Busa&ob=pv&od=desc using the USD12MD156N column.

3. Credit Card Default prediction. Predict whether an individual will default on credit card payments. What is the typical age of those who default? Are men or women more likely to default? Are they more likely to be college educated or do they typically have a high school education or less? How much credit had they been initially granted? How frequently had they been late with payments in the past? This topic would be of interest to any lenders, whether banks issuing credit cards or banks engaged in mortgage lending, for example. (An individual that defaults on a credit card payment is likely to default on other financial obligations as well.)

At least one dataset is obtainable from UCI (although using data from Taiwan) available at https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients.