Capstone 2 Ideas

* **Sentiment analysis of twitter data:** Our company has been collecting tweets that mention “BMW” through the use of a web-scraper for the past 3 years. It is possible that these tweets could be useful in determining trends with safety defects if users mention vehicle failures in the tweets. However, there is a great deal of noise in the dataset currently because tweets are about a wide variety of topics. If the tweets can be processed through sentiment analysis and then determined if they have positive, negative or neutral sentiment, then only the negative tweets can be used as an input to determine safety defects. This would greatly reduce the information that must be sifted through.
  + <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>
* **Predicting housing market prices:** Using regression, determine what parameters influence pricing of houses in a specific market. A close friend of mine has a property management company in Dallas, TX. They buy foreclosed properties and sell them to buyers as investment properties for rent. If a solid model was built based on Seattle’s data, they would be interested in seeing something similar built for the Dallas market so that they could determine how much capital is needed on a monthly basis for them to acquire houses in a certain area.
  + <https://www.kaggle.com/harlfoxem/housesalesprediction/data>
* **Predicting a rise in cryptocurrency:** Determine which crypto currency has the highest potential for value one year from today based on historical trends. Numerous investors have flocked to crypto currency for a chance at making large returns, but with the growing number of currencies to invest in, many people turn to the opinions of a few individuals (John McAfee) or chance for investment knowledge. Using historical data for the 15 or so cryptocurrencies found below, determine which currency has the most potential.
  + https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory/data