Capstone Project Ideas:

1. Extracting location information from unstructured data

The goal of this project is to perform text mining on unstructured data such as news articles regarding terrorism events and extract location information such as the country, city, street address etc. of where the event took place. The accuracy should be at least 80%. I am open to using any open source tool along with Python to solve this problem.

Data: Data obtained using a web scraping tool such as webhose.io

2. Predicting crimes in Chicago

This project consists of crime data for Chicago from 2001 to 2017. The crime data include reported incidents with the exception of murders. There are a total of 22 attributes and more than 6 million records. My plan is to build a model to predict crime rates for some recent years such as 2015 or 2016. I also plan to create some visualization based on the results obtained.

Datasource: Kaggle (<https://www.kaggle.com/currie32/crimes-in-chicago>)

3. Performing Big Data Analysis on a NOSQL database

This project deals with performing some kind of big data analysis using tools such as Hadoop, Spark etc. on a NOSQL database such as MongoDB, Cassandra etc. I am not sure exactly what I can do. I am open to suggestions. Here are some potential data sources:

<https://www.lendingclub.com/info/download-data.action>

<http://www.econ.yale.edu/~shiller/data.htm>