Capstone Project Ideas:

1. **Project Idea 1:**

   (**Src. Kaggle**) - House Prices: Advanced Regression Techniques (+ Interest Rates)

   With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.
   Main skills to showcase:
   - Creative feature engineering
   - Advanced regression techniques like random forest and gradient boosting

   One more variable of Interest Rates (from Gov. website) can be added in the mix to spin it and make it little different.

   https://www.kaggle.com/c/house-prices-advanced-regression-techniques
   Data Size - 200kb

2. **Project Idea 2:**

   (**Src. Kaggle**) - Microsoft Malware predition (focus on Gaming machines?):

   The goal of this competition is to predict a Windows machine's probability of getting infected by various families of malware, based on different properties of that machine. Malware detection is inherently a time-series problem, but it is made complicated by the introduction of new machines, machines that come online and offline, machines that receive patches, machines that receive new operating systems, etc. While the dataset provided here has been roughly split by time, the complications and sampling requirements mentioned above may mean you may see imperfect agreement between your cross validation, public, and private scores.

   I can be specific and restrict the prediction to Gaming machines ONLY. This is different from what is asked in the competition.

   https://www.kaggle.com/c/microsoft-malware-prediction/data
   Data Size - 1 GB

3. **Project Idea 3:**
   (**Src. Data is Plural**) - Predict the evictions based on inflation in usa.

   This data is provided by The Eviction Lab at Princeton University.

There is state by state breakdown of various data-points like income, zip code, race etc.
A GeoJSON is also available.
Data got from Data is Plural - 2018.04.18 - Evictions.
Inflation data is available from IMF (International Monetary Fund) website.
Data needs to be cleaned and combined.

https://data-downloads.evictionlab.org
Data size: ~ 800 Mb

4. **Project Idea 4:** Kaggle Black Friday Dataset
   The dataset here is a sample of the transactions made in a retail store. The store wants to know better the customer purchase behavior against different products. Specifically, here the problem is a regression problem where we are trying to predict the dependent variable (the amount of purchase) with the help of the information contained in the other variables.

   https://www.kaggle.com/mehdidag/black-friday.