Capstone project ideas

Here is my three ideas for capstone project.

1- Sberbank House price prediction

Housing costs demand a significant investment from both consumers and developers. And when it comes to planning a budget—whether personal or corporate—the last thing anyone needs is uncertainty about one of their biggest expenses. Therefore, many companies such as Zillow, a real-estate Marketplace, or, Sberbank, Russia's oldest and largest bank, helps their customers by making predictions about realty prices so renters, developers, and lenders are more confident when they sign a lease or purchase a building. Recently, Sberbank has challenged Kagglers to develop algorithms which use a broad spectrum of features to predict realty prices. They provided a rich dataset that includes housing data and macroeconomic patterns.

Although the housing market is relatively stable in Russia, the country's volatile economy makes forecasting prices as a function of apartment characteristics a unique challenge. Complex interactions between housing features such as number of bedrooms and location are enough to make pricing predictions complicated. Adding an unstable economy to the mix means Sberbank and their customers need more than simple regression models in their arsenal. An accurate forecasting model will allow Sberbank to provide more certainty to their customers in an uncertain economy.

The dataset has been released in the Kaggle (https://www.kaggle.com/c/sberbank-russian-housing-market)

2- Chicago crime prediction

The crime rate has increased in Chicago even more than New York and LA in combined. But now the police say they have got a solution. Police in Chicago are turning to big data and technology to help predict where and when violent crime might occur. They identify patterns and trends through Machine Learning and predictive model Analytics. One of the programs uses this data to predict possible crimes and locations. Then the police deploy officers to hotspots. This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. The dataset is released in Chicago Data Portal (https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2).

3- Expedia Hotel Recommendations

Planning your dream vacation, or even a weekend escape, can be an overwhelming affair. With hundreds, even thousands, of hotels to choose from at every destination, it's difficult to know which will suit your personal preferences. Should you go with an old standby with those pillow mints you like, or risk a new hotel with a trendy pool bar? Expedia wants to take the proverbial rabbit hole out of hotel search by providing personalized hotel recommendations to their users. This is no small task for a site with hundreds of millions of visitors every month! Currently, Expedia uses search parameters to adjust their hotel recommendations, but there aren't enough customer specific data to personalize them for each user. Therefore, Expedia has recently challenged Kagglers to contextualize customer data and predict the likelihood a user will stay at 100 different hotel groups. The dataset has been released in Kaggle. (https://www.kaggle.com/c/expedia-hotel-recommendations#timeline)