Background: We are at the beginning of a very deep, and likely very long, recession because of the Covid-19 pandemic. The history of recession in the 21st century speaks to a long road to recovery, with much volatility between areas and industries.

Question: Can we predict which areas of the US are most likely to recover from the recession?

Dataset: I will be using the Bureau of Labor Statistics Quarterly Census of Employment and Wages data from 2000 – 2019. I have classified the time periods into two recession/recovery periods: 2001 and 2007/8.

For each county in the United States, I will define "recovery" as whether or not the county has re-gained the same number of jobs as in their pre-recession "peak"- that is, the highest number of jobs before the recession catalyst. Recession catalysts are defined as 9/11/2001 (Q3 2001) and the stock market crash of 2008 (Q3 2008). While these are not the sole catalysts of the recessions in question, they are the events that had the most dramatic effects on the economy.

Features: I will use a combination of industry makeup of each area (defined by the BLS), as well as state political data (partisan control of legislative branch, governorship, state budget totals, etc.) as my feature set. County-level political data is too diffuse and granular to collect and operationalize for this project, and will not be used. Both industry makeup and political scenario features are calculated at the time of the recession catalyst.

Excluded data: A few counties are missing some key data from the QCEW. Additionally, some counties population data is not obtainable through census records. These records are excluded.

Puerto Rico and District of Columbia are also excluded, as their budget information is not available.

See data dictionary for complete column breakdown.

MVP: Create a model to evaluate feature importance, to identify the most important factors in whether an area will recover from the recession.

MVP+ : Use the model to predict which areas will recover from the 2020 recession.

MVP++: Include more granular industry makeup (going further down in the BLS definitions) to get at more specific factors that determine recession recovery.