So now we will use the GDP to opting the former 50% state as our choosing pool.  Why is GDP?  Cause we want to ensure the economic impact can be directly observed.  Also to avoid any other factor that will cause a bias to the economy index.  We use the data from the state economic monitor, which includes US GDP information.  The first step is to read the file and set the Geography column, representing the state name, as the index column.Later on, we need to set up a benchmark of GDP because the pandemic begins in 2020 Q2, so we use column function to choose only the 2020 Q1 GDP as our benchmark.  As the picture shows.  Then we use nlargest function to arrange the state GDP and select the top 50% state as our candidate state. Then we build a new variable that stores the data only from the second row.  Because the first row is the data for the United States, not by the state, so we remove it.  Also, we transit the float into integer to better visualize our data.

The second step is to read the covid case.  We found the information from CDC and we set the State/Territory as the index column.  Then use left join to merge it with our previous statement, the 25 candidate states.  Also, we use fillna function to replace all the NA values with zero.

Then we need to choose five states to focus on. So we use nlargest to rank again. This time the ranking factor is the case ratio. Which means we pick up five most severe state among our candidate. Because we need to merge this file later, we also change the index name into “State”. Recalculate the confirmed case portion by state and insert as the first column to better visualize the percentage of confirmed cases.

The five states are Tennessee, Arizona, Wisconsin, New Jersey and Indiana.

We have target.

Then need two dataset to observe if the vaccination rate do affect the economic circumstances?

So we choose two index first is unemployment rate and next one obviously is vaccination rate.

We choose unemployment rate as our economic index, because we assume it can reflect the short term economic prosperity. Also, the disease will directly impact people and the unemployment rate is the most related to people than other economic index such as GDP or CPI.

So we start to import the unemployment rate by state, and set the index column by State also named after the column as State to satisfy the merged condition.

Then we start analyzing vaccination data. We set the Location as index column. Because the dataset is a daily accumulation so we use the pivot function to choose a specific date.. Because the Location from the original dataset is only abbreviation by state we insert the state name by merge function with abbreviation list to ensure we can merge this data with our other dataset. We also clean the NA data in the data set. Now we have vaccination data and the five most harsh states. We can now start to merge our dataset together to visualize.