Beau Tate

DTSC 3602

Data Collection, Joining, and Cleaning Notes

**GTD Data Set:**

**OECD Leading Composite Indicator:**

* Pulled data from OECD data base for G7 countries (minus the European Union) for dates ranging from February 05 through September 2021.
* Created a delta column that tracks the difference for each observation versus the third month prior. For example, to calculate the delta for April 2006 the delta column would hold the difference between the value for May 2006 and February 2006
* Left joined to the GTD by country code so that every GTD observation would have a relevant delta value

**OECD Immigration Data:**

GTD:

* Imported the whole GTD database
* Selected for relevant features including predictor and target
* Filtered for the G7 countries (minus the EU)
* Filtered for the years 2006 to 2020
* Created country codes to use to join other data frames