* The Data
  + Main Dataset – (After Cleaning)
    - 193 Countries, 2005-2015
    - 2023 data points with 22 features (minus target feature)
      * Examples: GDP, Alcohol consummation, Population, Developing/Developed
  + 7 Separate dataset
    - 5 added water and sanitation features
    - 2 were to clean/replace features
* ETL
  + Issues:
    - Nulls, zeros, string to ints, and inconsistencies.
    - Population changed by factor of 10x year to year
    - Getting new clean data for `population` and `GDP` required conversions from wide to long.
    - Primary KEYS: Year and Country
      * Country names had lots of inconsistency.
  + Was an iterative process with the EDA and ML
* Database
  + AWS’ RDS to Postgres
  + Used psycopg2 to bridge python and PostgresSQL

A picture containing calendar

Description automatically generated