**ETL Project**

**Objective:**

Extract data from ‘multiple’ sources, transform and load the data so it can be easily accessed / queried by your business teams,

**Practical Question:**

What impact does income and alcohol consumption have on homicide rates?

Organization has multiple applications / systems:

Data will be pulled from the following three sources:

1. <https://www.kaggle.com/frankmollard/income-by-country>
2. <https://ourworldindata.org/alcohol-consumption>
3. https://data.worldbank.org/indicator/VC.IHR.PSRC.P5

**Approach:**

1. Locate source data / identify data sources – we pulled CSVs from the databases above
2. Extract data into files / DataFrame
3. Study the data – merge by year 2016
4. Clean data individually
5. Find a common value(s) / column(s)
6. Merge the data sets merge by country on year 2016
7. Upload the merged data into a database
   1. PostgreSQL – we are using
   2. Sqllite
   3. Mongo
8. Provide a toolset to be able to query the data loaded