**ETL Report**

**Extract**

Data files related to Chicago Population, Crime, and Public Health Indicators were identified. CSV files were accessed and downloaded from the Chicago Data Portal and the Act for Children websites. Below is a list of datafiles used for this project:

* Chicago Public Health Data 2015
  + Data Source: <https://data.cityofchicago.org>
* Chicago Crime Data 2018
  + Data Source: <https://data.cityofchicago.org>
* Chicago Population Data 2017
  + Data Source: <https://www.actforchildren.org/>

**Transform**

Several transformations were performed to prepare the data for analysis. Each transformation per data source is described below:

* Chicago Public Health Data 2015
  + Select Columns
    - Community Area \* (PK)
    - Community Area Name
    - Unemployment
    - Dependency
    - No High School Diploma
  + Drop NULLS
  + Community Area – change variable type to STR
* Chicago Crime Data 2018
  + Select Columns
    - ID
    - Community Area \* (PK)
    - Primary Type
  + Drop rows where Community Area = 0
  + Drop NULLS
  + Community Area – change type to STR
  + Group by Community Area
* Chicago Population Data 2017
  + Select Columns
    - Community Area \* (PK)
    - Population
  + Community Area – change type to STR
  + Population – change type to INT
* Merged – Master Table
  + Merged on ‘Community Area’
  + Create variable – Crime rate (Calculation)
    - Calculate crime rate
      * Number of crimes / population \* 1,000

**Load**

The team decided to use PostgreSQL.

* PostgreSQL is an open source object-relational database system that uses SQL language
* PostgreSQL is a free and open-source relational database management system (RDBMS)

Three tables were created

* Public Health
* Crime
* Population

Tables merged in python using an SQL command

* Master\_table (merged all three tables)

**Analysis**

* Correlation
  + Crime Rate and Unemployment

A screenshot of a cell phone

Description automatically generated

* + Unemployment and Education (No HS diploma)

**A screenshot of a cell phone

Description automatically generated**