**ETL Project Plan**

* Data
  + Population Data
    - Columns
      * Community Area \* (PK)
      * Population
  + PubHealth Indicators csv
    - Columns
      * Community Area \* (PK)
      * Community Area Name
      * Unemployment
      * Dependency
      * No High School Diploma
    - Transform
      * Drop NULLS
      * Community Area – change type to INT
  + Crime csv
    - Columns
      * ID
      * Community Area \* (PK)
      * Primary Type
    - Transform
      * Drop NULLS
      * Community Area – change type to INT
      * Crime rate (Calculation)
        + Drop rows where community area = 0
        + Group by cmty number
        + Calculate crime rate

# crimes / population \* 1,000

* + - * Drug related crime – Create variable
        + Primary Type – Narcotics & Other Narcotics Violation
        + Combine these two offenses into new variable
  + Merged Crime and Population
    - Merge on ‘Community Area’
* DB
  + Postgres
  + ERD Diagram
* Correlation
  + Crime rate and Unemployment
  + Unemployment and Education (No HS diploma)
  + Drug related crimes (created variable) and Dependency