write a brief report.

1 - where, what

2 - column description / data dictionary

3. Evaluate dataset
→ missing value, summary statistics

4. Integration
eval again

5. pre-processing data

Find the factors to develop a recommendation system based on the composite scores to determine which counties in PA are more at risk for "effects."

LABEL   A B C D E.

Score : 1 2 3 4 5

collect /
0. Understand the data set, and clean the data, integrate Data.

1. Determine Factors

2. Building the predictive models

3. Develop the score
→ how often updated?
→ get past scores?

1. Collect, understand, clean, and integrate the data.

2. variable selection (selecting factors) and building the predictive models (modeling the effects)

3. Develop the Score for each PA county

Internal Risk

IR_1, IR_2 - - - - - - - - IR_{10}

"factors"

Cause
- licensed treatment facilities
- center of excellent locations
- location / quantity of police stations
  - PacMAT locations
  - programs
- family member that is addicted
  - insurance
- number of pharmacies in area
  - single parents
  - education factors
  - mental health diagnoses
  → Ermp RATE
  → Age   - alcohol use
  → Income   - smoking
  → Dropoffs - prevalence of AA
              - self-medication for mental (issues)

Effects
- overdose #s
- addicted babies data
- Methadone rx data
- DOC - Felonies & misdemeanors drug-apred data
- future population increasing