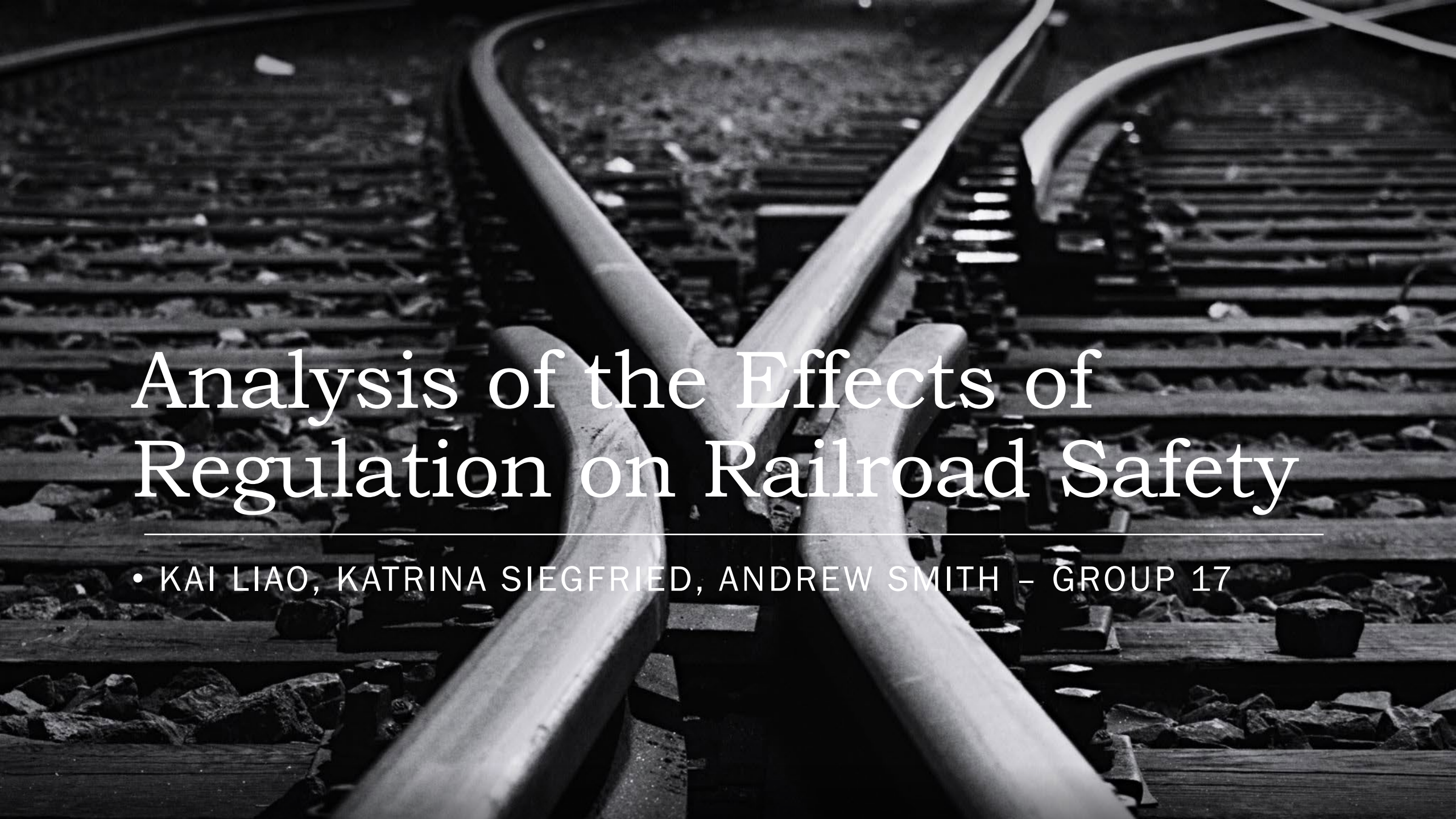


Analysis of the Effects of Regulation on Railroad Safety

• KAI LIAO, KATRINA SIEGFRIED, ANDREW SMITH – GROUP 17

Project Description

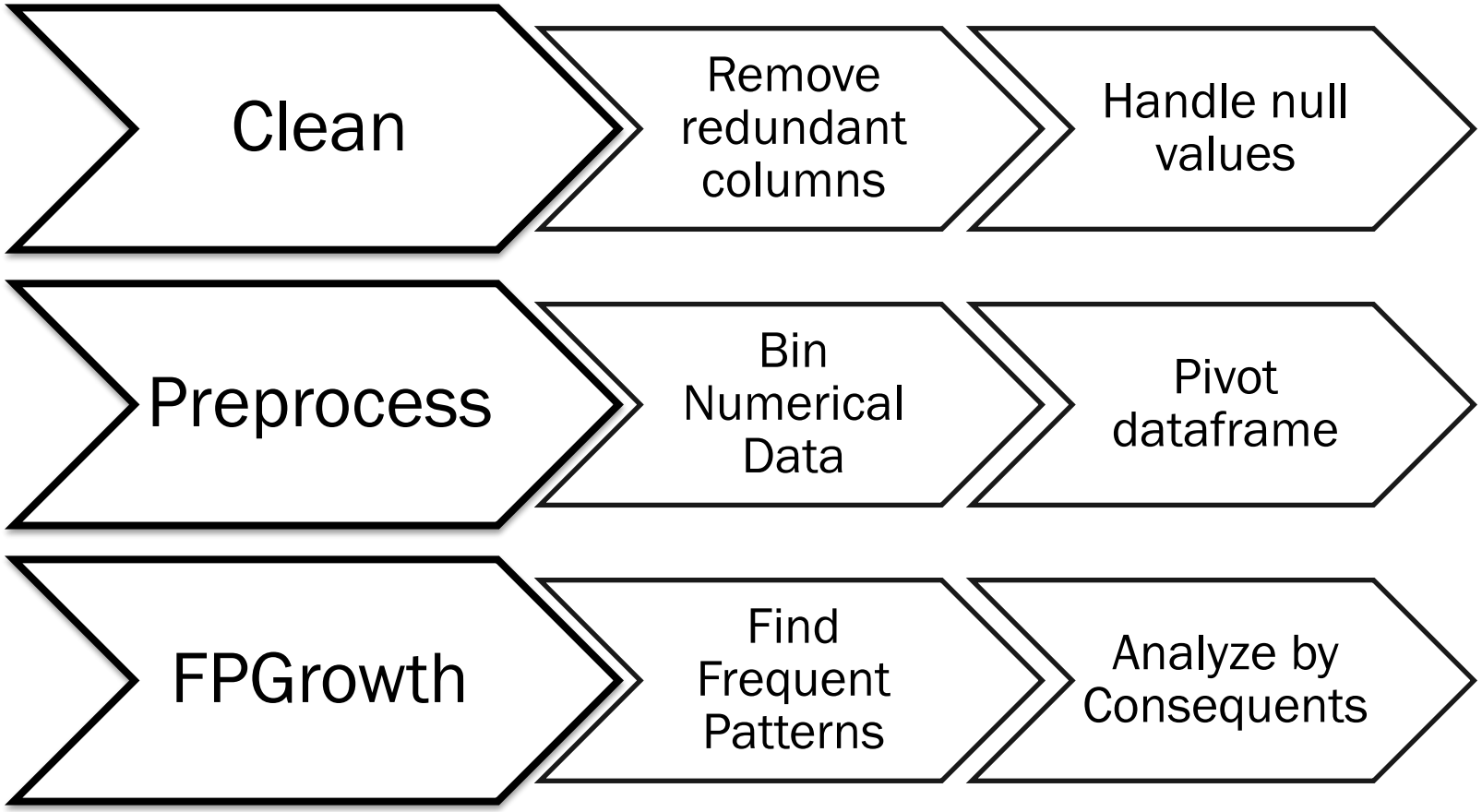
- Accidents at highway-railroad intersections cause tremendous losses of lives and resources. This project aims to consider the impacts of new regulations, locations of intersections, and the characteristics/topography of intersections to determine which features promote safety and which features do not.
  - Accident Analysis by Location
  - Effects of Intersection Characteristics

FP Growth Tools

Investigate frequent patterns and if they do or do not shed insight on attributes associated with incidents

Python	Pandas	Pyspark

FP Growth Method



FP Growth Results

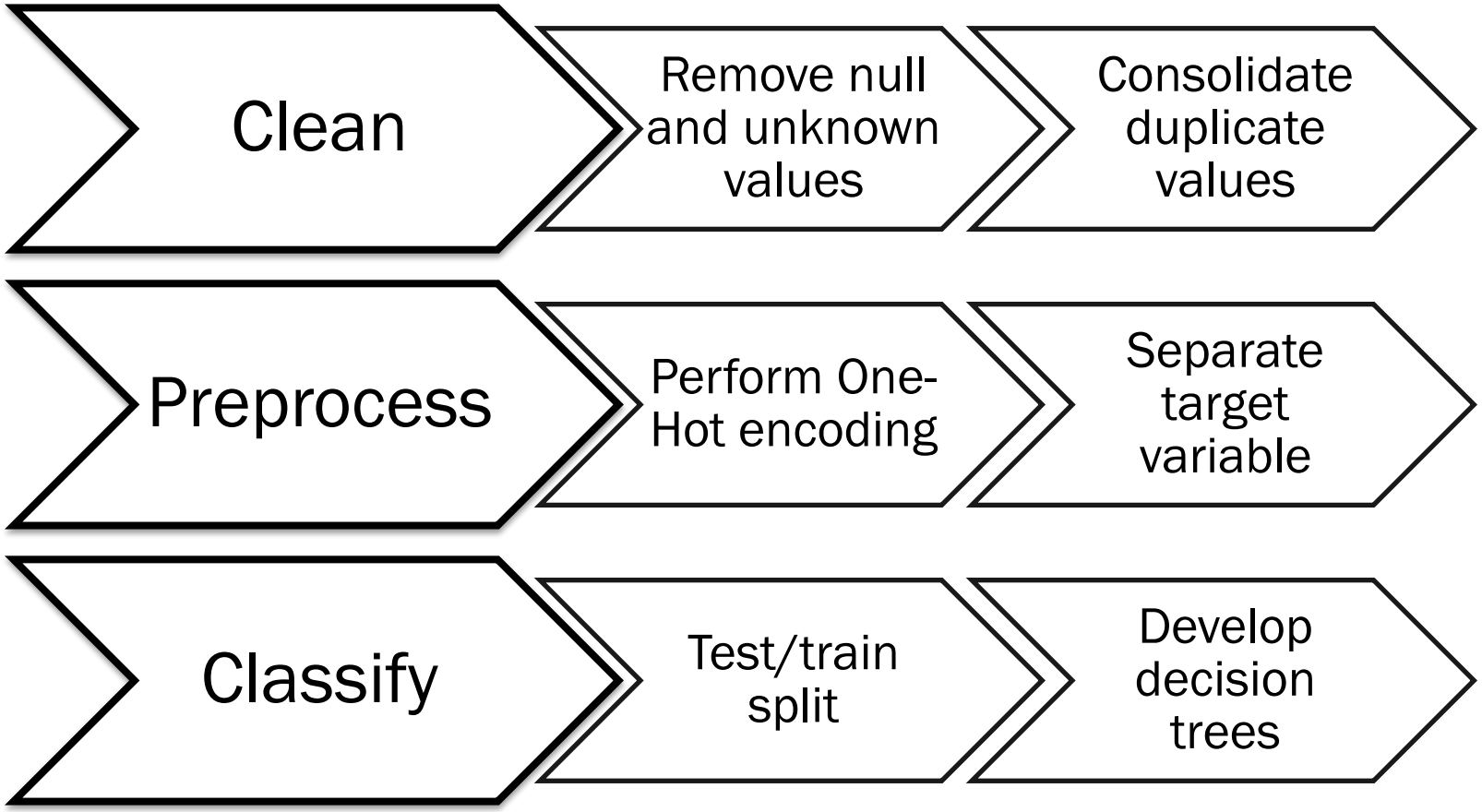
Pattern	Support	Lift	Confidence
Mortality High<-> Number of Train Cars Low	0.063	1.063	0.786
Casualty High <-> Number of Train Cars Low	0.063	1.062	0.786
State <-> City, Type of Track	0.542	1.128	0.826
Warning Connected to Signal, Crossing Illuminated <-> Mortality Low, Casualty Low	0.139	1.005	0.921

Intersection Characteristics

- How do the characteristics of an intersection effect the severity of incidents between railroads and highways?

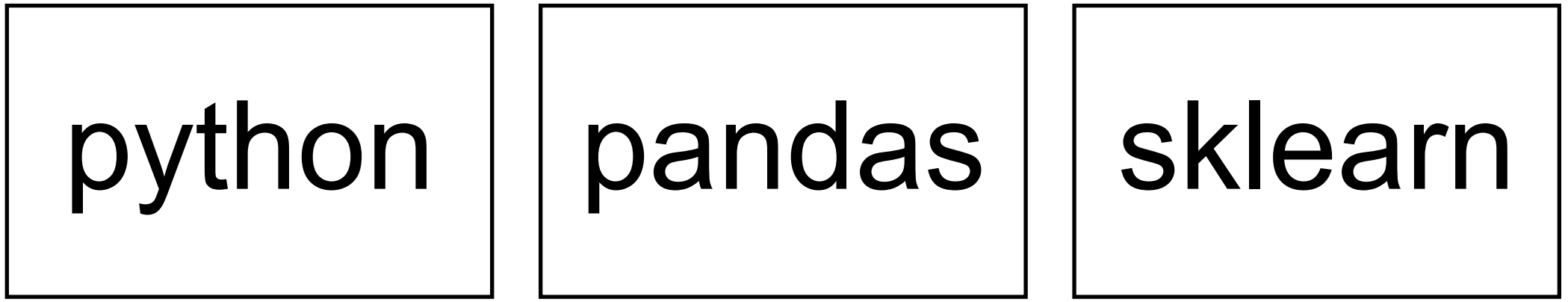
- **Method:**
- Decision Tree Classification

Decision Tree Methods

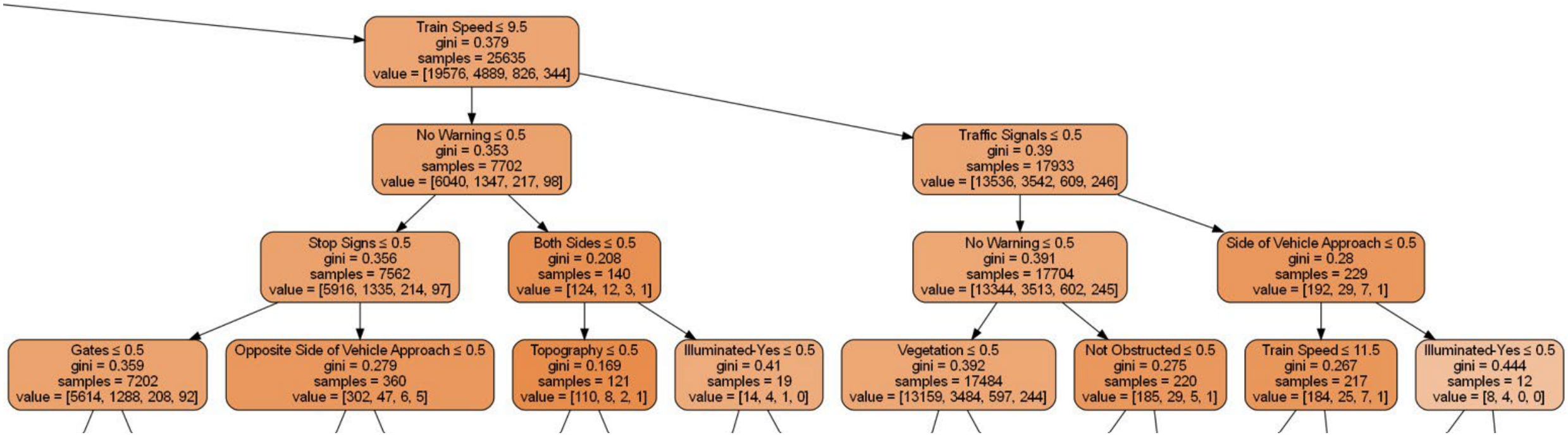


Intersection Classification Tools

- The Python Pandas library was used to manipulate data frames for cleaning and preprocessing. The sklearn library is used to perform one-hot encoding, decision tree classification, and classification scoring.



Intersection Classification Decision Tree



Intersection Classification Results

Injuries	Precision	Recall	F1-Score	Support
0	0.7278	0.9763	0.8339	34824
1	0.2374	0.0244	0.0443	10205
2	0.0598	0.0035	0.0066	1994
3+	0.0244	0.0011	0.0021	898

Accuracy: 0.7148

Intersection Classification Application

Prediction

- How many injuries can be expected at an accident for a given intersection?
- Model Developed

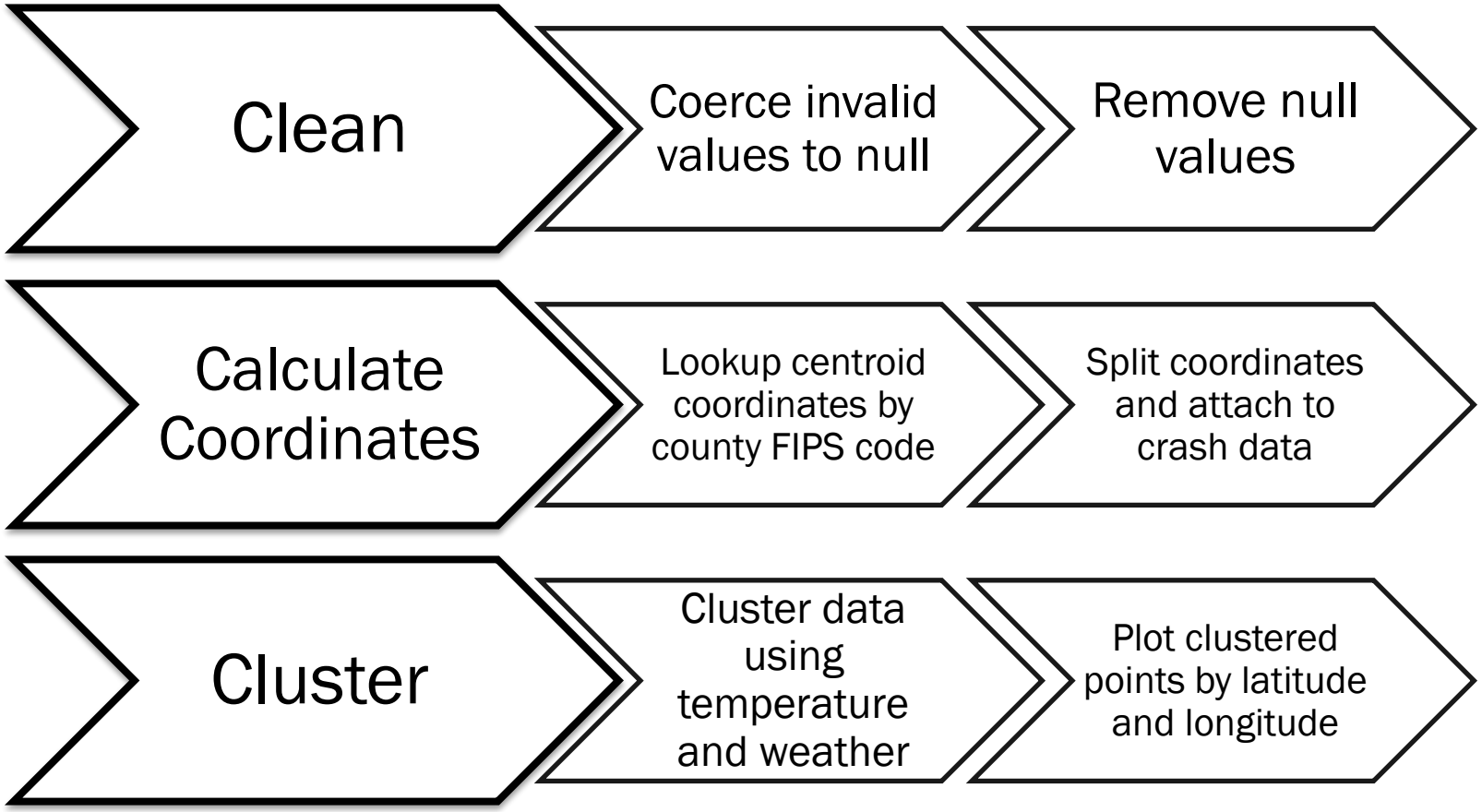
Regulation

- What characteristics appear to be driving increased number of injuries from accidents?
- Statistical analysis needed

Innovation

- What intersection characteristics can be improved upon to decrease the number of injuries experienced?
- Expert opinion needed

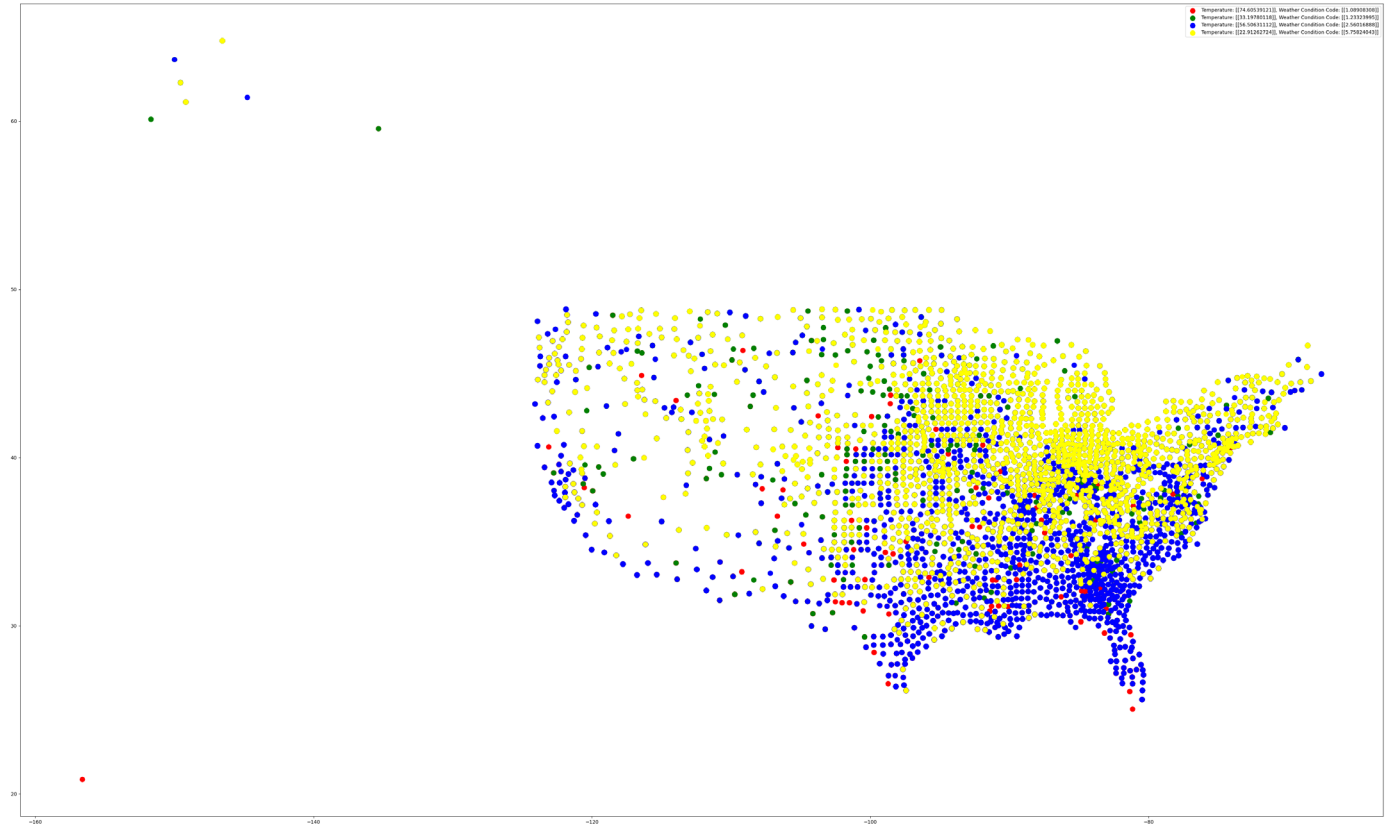
K-Means Locational Clustering Methods



K-Means Locational Clustering Tools

- Three python libraries were used. The pandas library was used to manipulate data frames for cleaning and preprocessing. The sklearn library was used to perform K-means clustering. The matplotlib library was used for visualization.

K-Means Locational Clusters



K-Means Locational Clusters Application

Freezing Snow

Northern states may need to focus on accident mitigation in blizzards

Cold Rain

Southern states may need to focus on accident mitigation in rainstorms

Clear Weather

- Inclement weather, rather than clear weather, represents the largest percentage of crashes and may be a large contributing factor