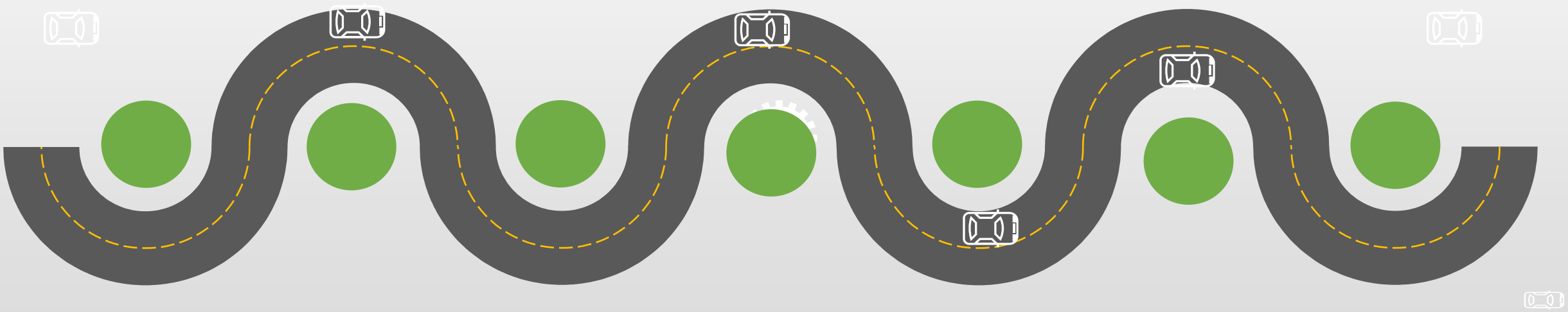
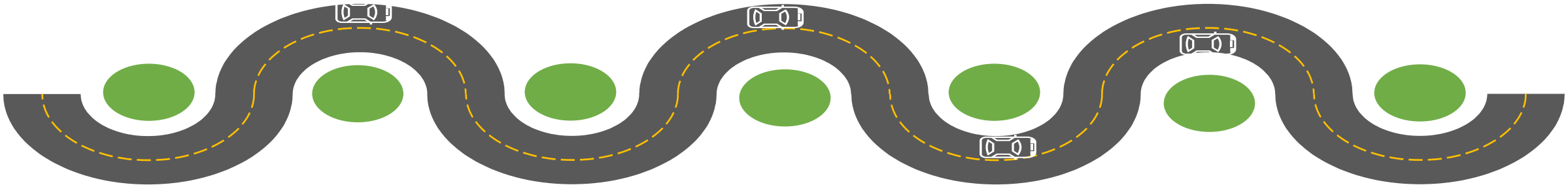
Seattle Car Accident Severity Analysis

Machine Learning Predictive Modeling



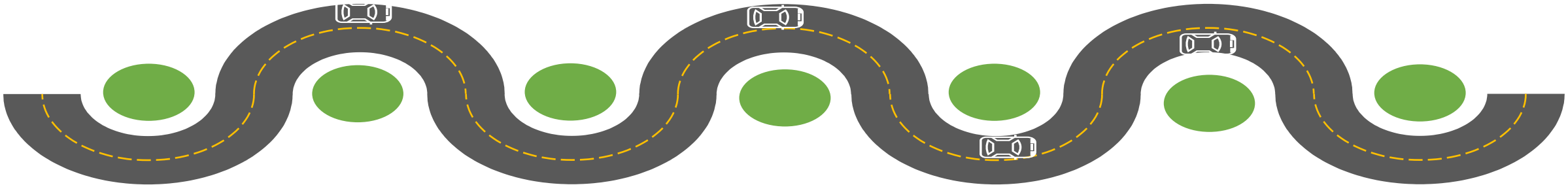
Crash Statistics

- Road traffic crashes kill 1.35 million people worldwide each year
  - first cause of death among children aged 5-14 and young adults aged 15-29
  - additional 20-50 million suffer non-fatal injuries, often resulting in long-term disabilities
- Several factors contribute to road crashes
  - unsafe road user behaviors
    - speeding
    - under influence of drugs or alcohol
    - inattention (e.g., reading texts, watching a squirrel)
  - environmental factors
    - weather
    - road conditions
    - light conditions



Seattle Accident Traffic Record

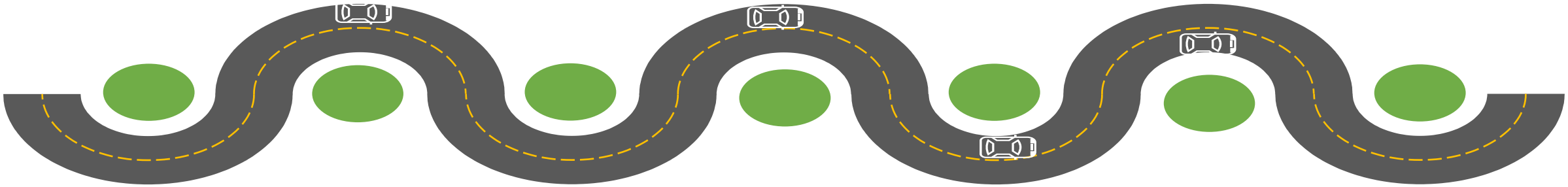
- A data subset to analyze the severity variables
  - 38 independent variables
  - 194,673 rows
- Dependent variable reporting severity levels
- Independent variables determining where severity relationships exist



Dependent variable

The dependent variable, “SEVERITYCODE”, contains data that correspond to different levels of severity. This data is a broader categorization of the data found in the Metadata file that accompanies the dataset.

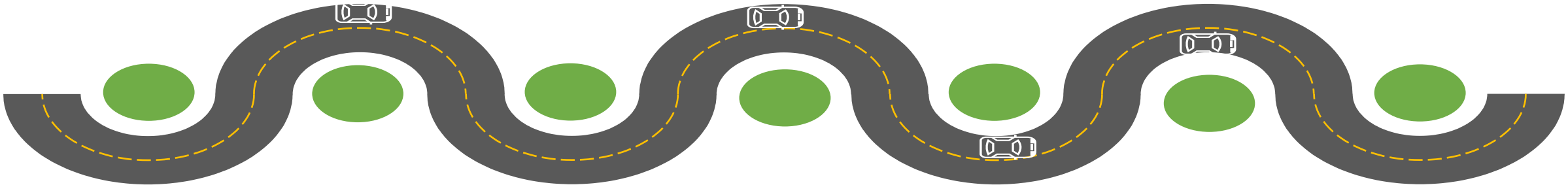
Attribute	Description
SEVERITYCODE	A code that corresponds to the severity of the collision: • 1—property damage • 2—injury



Dependent variable

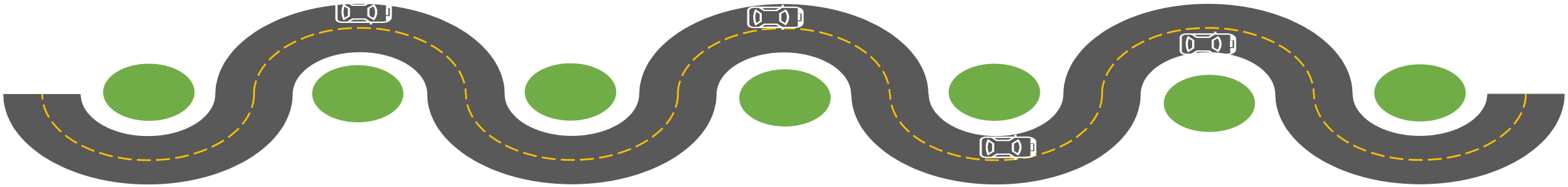
These variables are used to analyze WEATHER, ROADCOND and LIGHCOND to determine whether they have any relationship on the severity of the accident.

Attribute	Description
WEATHER	A description of the weather conditions during the time of the collision
ROADCOND	The condition of the road during the collision
LIGHTCOND	The light conditions during the collision



Methodology

- Extracted the dataset
- Determined what attributes were needed to train my machine learning models
- Dropped superfluous data that were not relevant to the project’s objective
- Converted the remaining environmental attributes to categories
- Assigned variables to the categories
- Balanced the severity data set
- Transformed the data for ingestion into the ML predictive models
- Ran the models
- Created a chart to illustrate the accuracy of the models using the data



Results Based on ML Predictive Modeling

Based on the accuracy percentages, we can see that by having median values, the environmental attributes (weather, road conditions and light conditions) have some, but not an overwhelming influence, on the severity of the accidents.

Algorithm	Jaccard	F1 - Score	Log Loss
KNN	0.55	0.53	NA
Decision Tree	0.55	0.47	NA
Logistic Regression	0.52	0.51	0.68