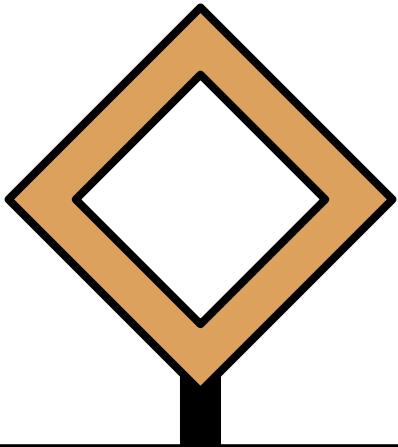


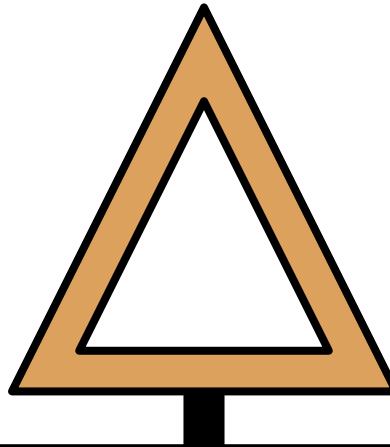
# **ROAD ACCIDENTS**

Data Analysis by Maëlys, Natalia, and Miguel

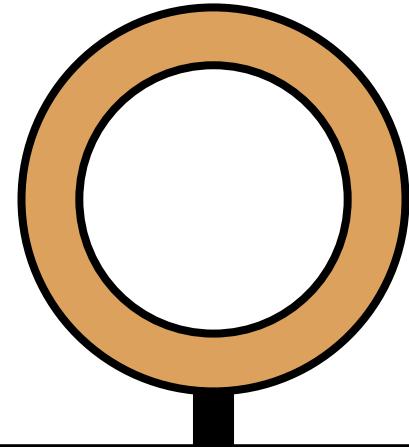
# Context - Auto Accidents in the US



**6M car accidents  
in the / year**



**38K deaths due to  
car accidents / year**



**3-4 car collisions in  
a person's lifetime**



# Data Overview

**Data Collection:** through APIs which broadcast traffic data captured by the **US state department** of transportation, **law enforcement**, and traffic **cameras / sensors**

kaggle

**Objectives:** extracting cause and effect rules to **predict car accidents** and studying the **impact of precipitation or other environmental stimuli** on accident occurrence

**Target Variable:** Severity of Accident

**Explanatory Variables:** environment, weather, location...

**Model:** Multinomial Classifier (4 classes numbered 1-4)

49

US States  
in Data

7

Years of  
Data

7.7

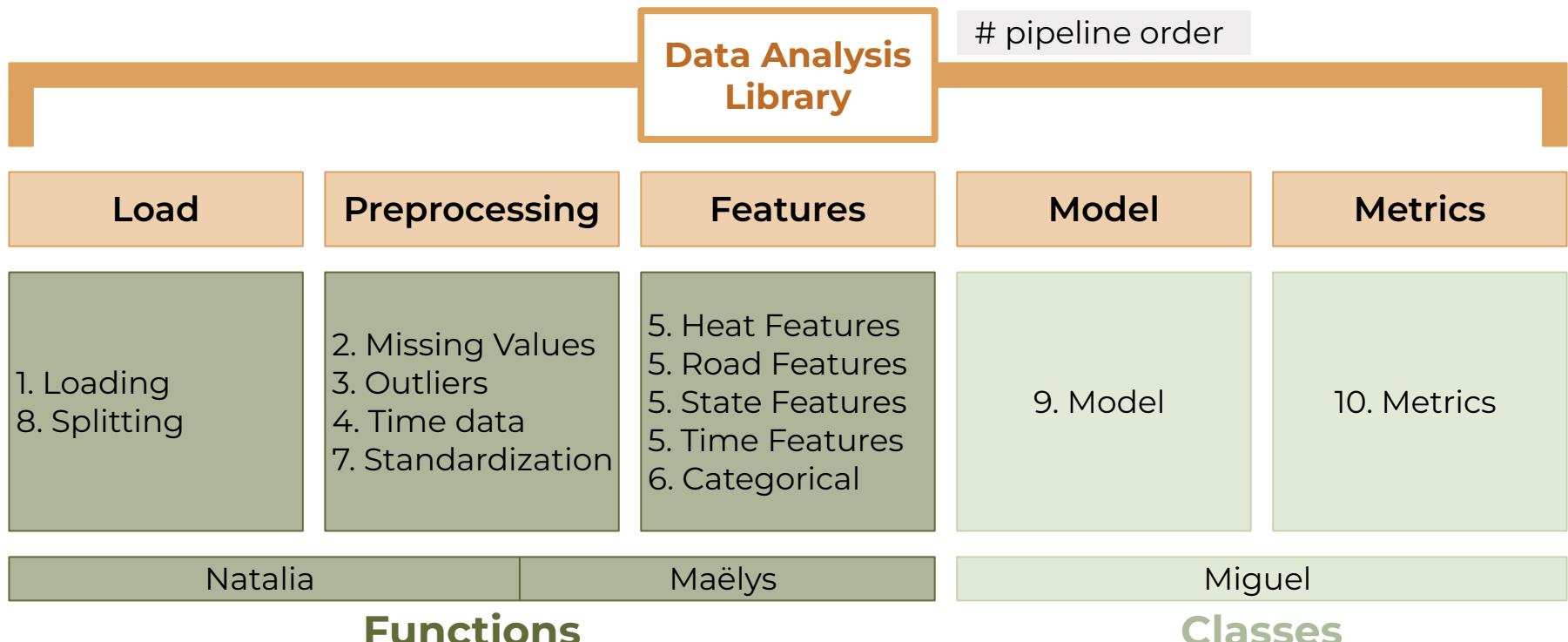
Million  
Records

0.1

Million in  
Sample



# Library Architecture



# Preprocessing

Missing Values	Outlier Treatment	Standardization	Time Data
Drop_row_na			
Fill_na_mean	Winsorize	Standardization	Time_type
Fill_na_0			



# Features

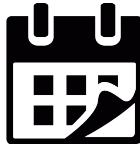
Functions that **transform or create** new features based on available data will fall under **feature engineering** and make up the modules in our features subpackage.

# Features - Time

## Month: DateTime Value Extraction

Using the Following Code:

```
data[column].dt.month
```



**Output:** [1, 2, 3, 4, 5,  
6, 7, 8, 9, 10, 11, 12]



## Duration: DateTime Difference

Using the Following Formula:

```
End Time - Start Time
```



**Output:** hours

# Features - Road Features (sum)



Amenity



Bump



Crossing



Give Way



Junction



No Exit



Railway



Round-about



Station



Stop



Traffic Signal



Turning Loop



# Features - Heat

**FAHRENHEIT** → **CELSIUS**

Using the Following Conversion Formula:

$$C = \frac{5}{9} (F - 32)$$

Where:

**C** = degrees Celsius

**F** = degrees Fahrenheit



## HEAT INDEX

*What the temperature feels like to the human body when relative humidity is combined with air temperature.*

Classification	Heat Index
Caution	80 - 90 °F
Extreme Caution	90 - 103 °F
Danger	103 - 124 °F
Extreme Danger	> 125 °F

# Features - States

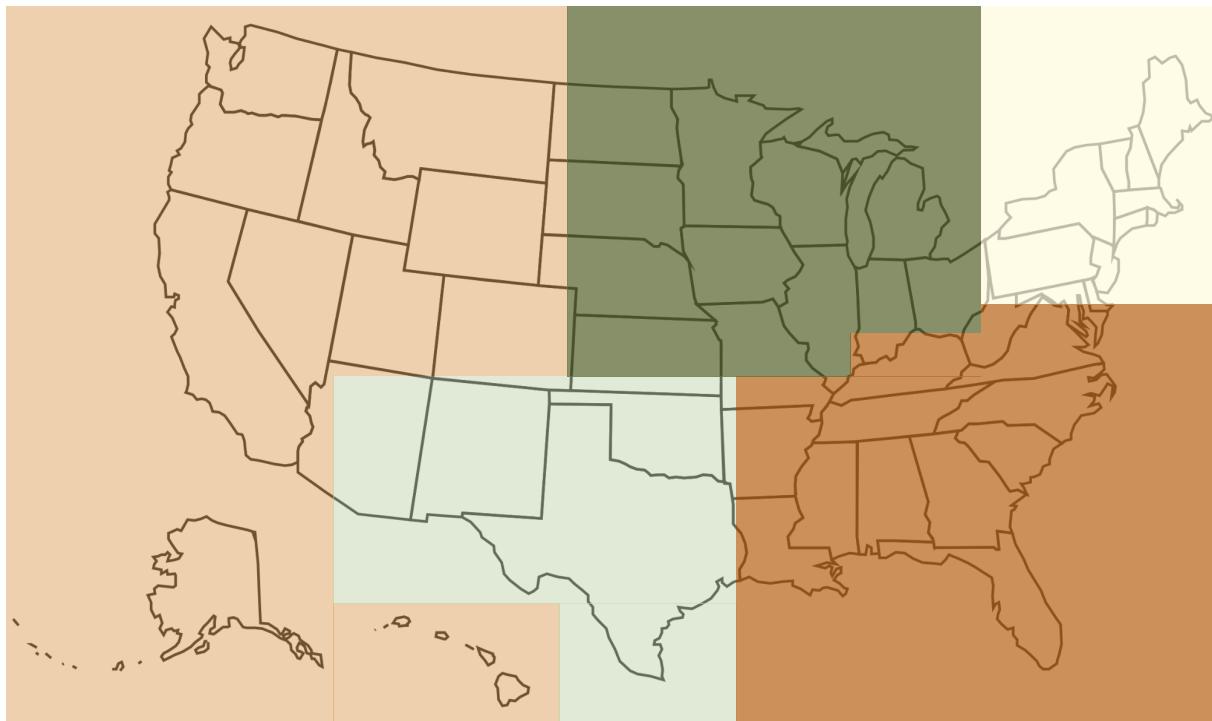
Midwest

West

Northeast

Southeast

Southwest 



# Features - Categorical

## One Hot Encoding

### Multi-Category Variables

Variables:

- Region
- Wind Speed bins

Original	N E S W
North	1 0 0 0
East	0 1 0 0
South	0 0 1 0
West	0 0 0 1

## Binary Encoding

### Binary Variables

Variables:

- Sunrise Sunset
- Civil Twilight
- Nautical Twilight
- Astronomical Twilight

Original	Binary
Sunrise	1
Sunset	0

## Binning

### Numerical to Categorical

Variables:

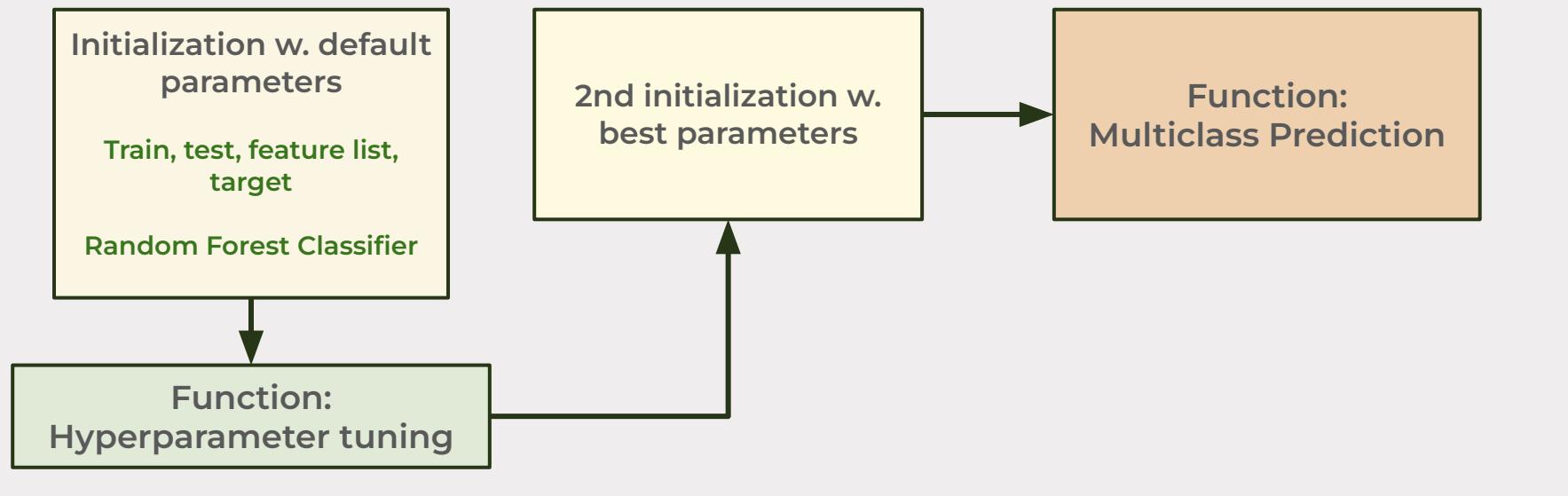
- Wind Speed (mph)

Original	Bins
0	Low
4	Medium
10	High



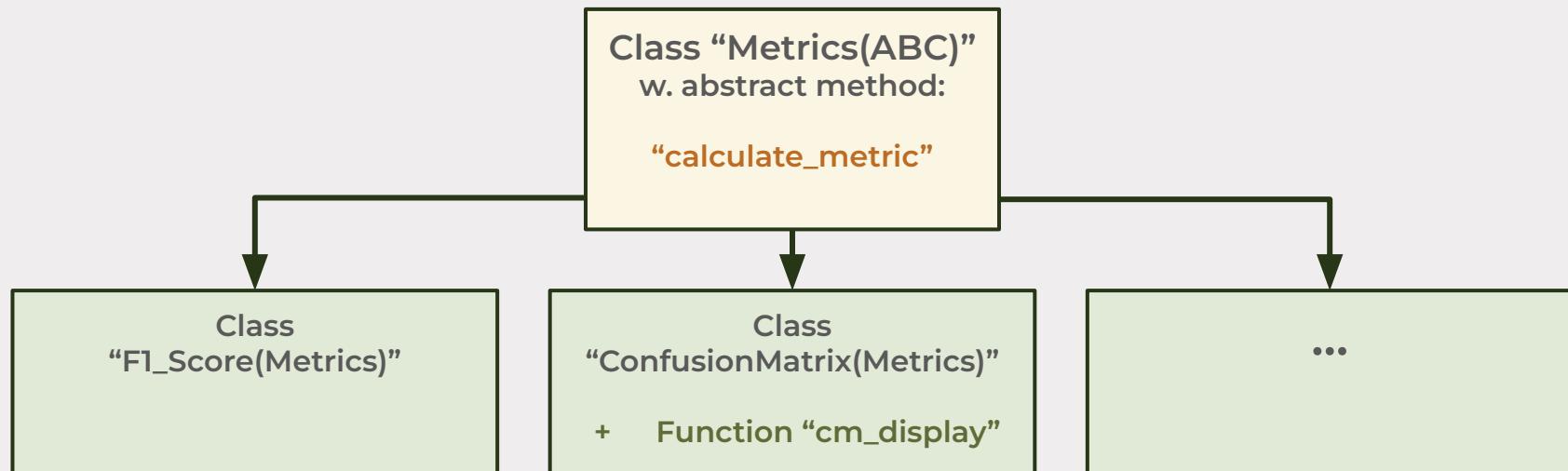
# Model

## Class Structure



# Metrics

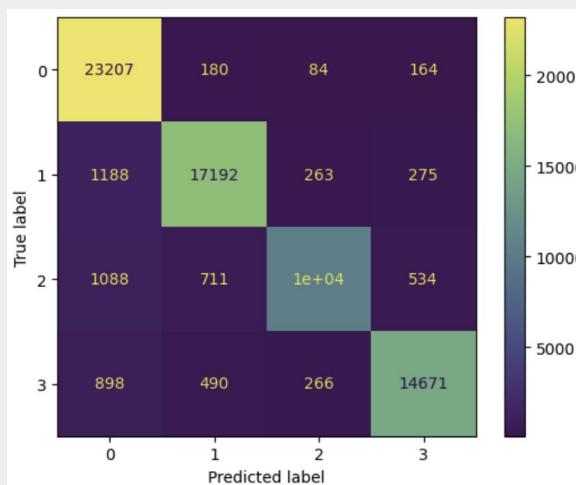
## Abstract Class Structure



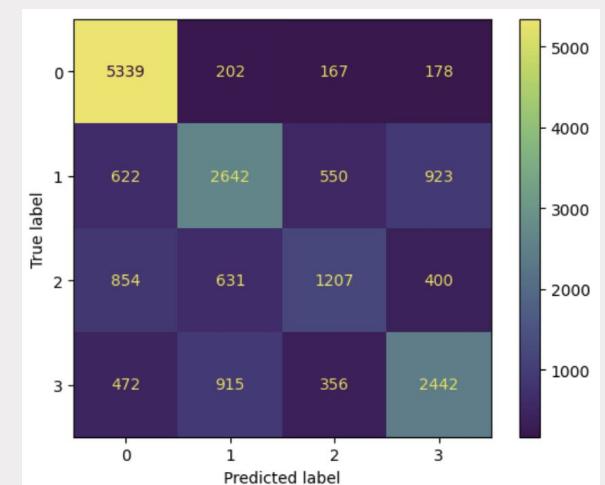
# Metrics

## Confusion Matrix Display

Class  
“ConfusionMatrix(Metrics)”  
+ Function “cm\_display”



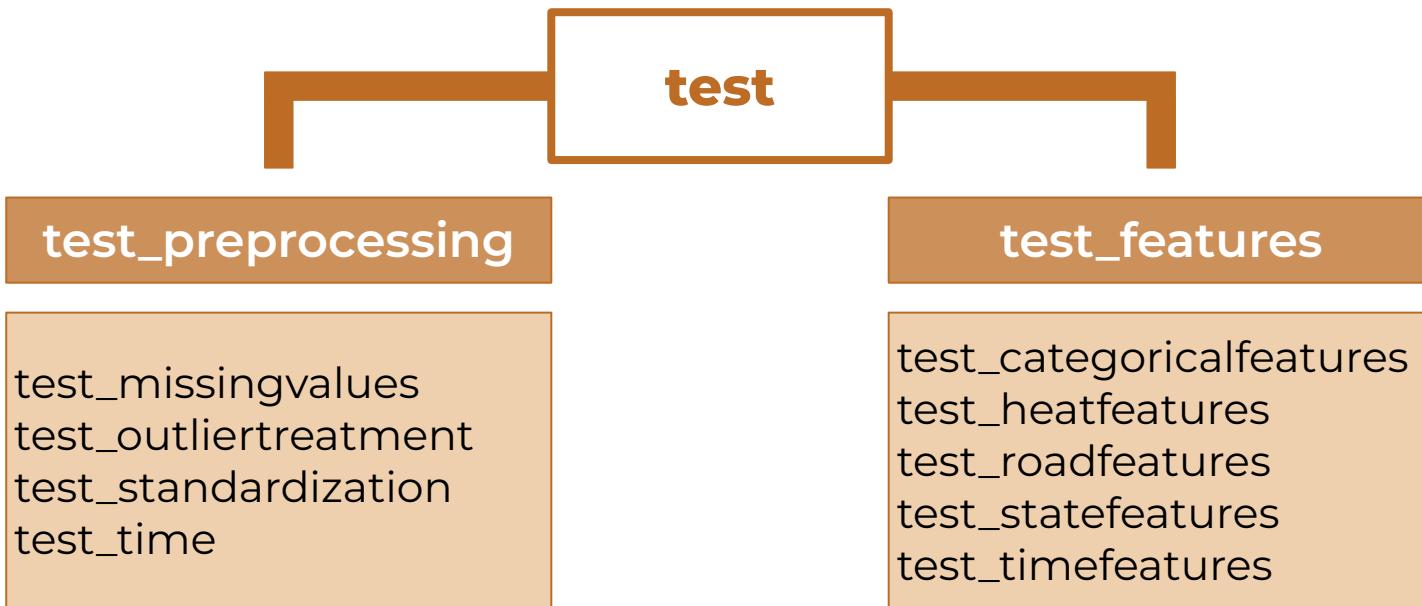
For Train



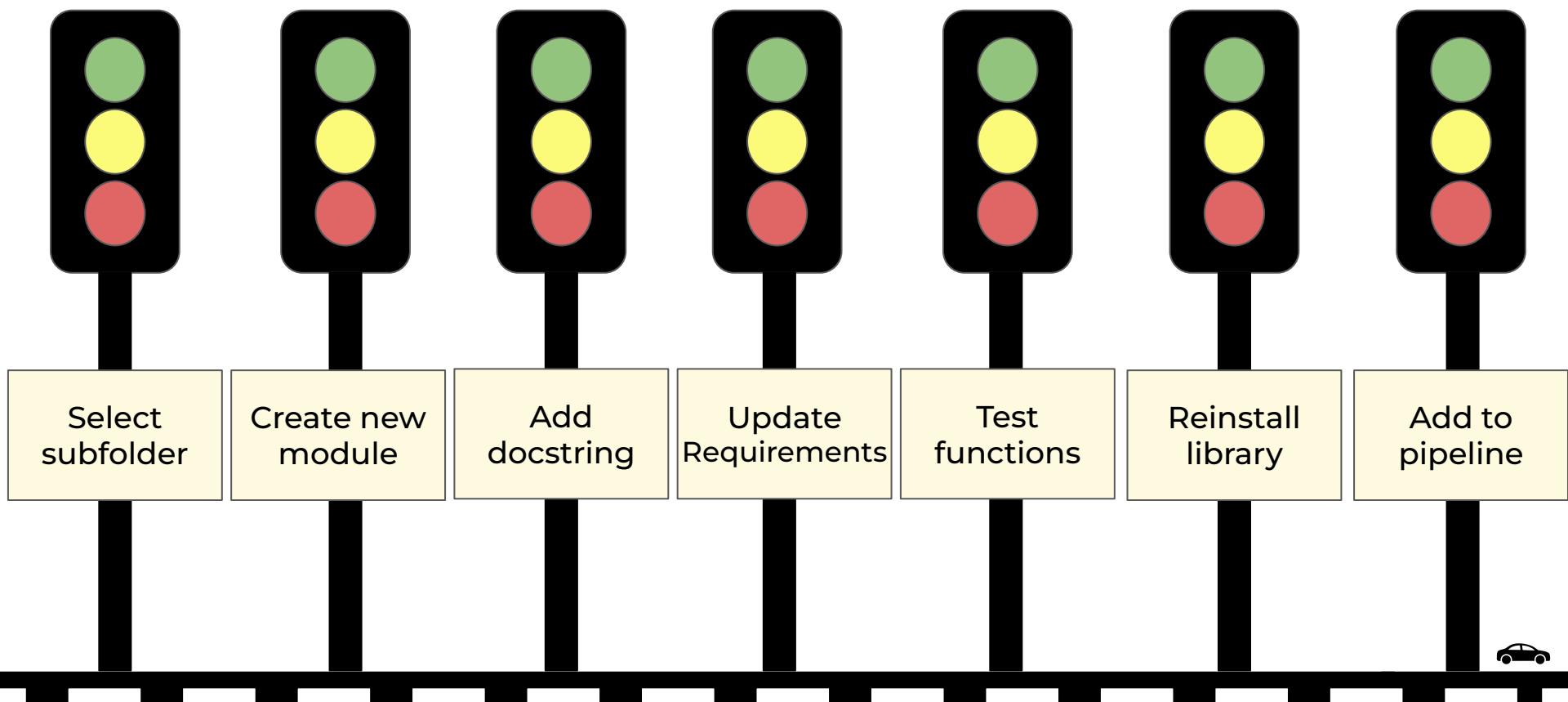
For Test

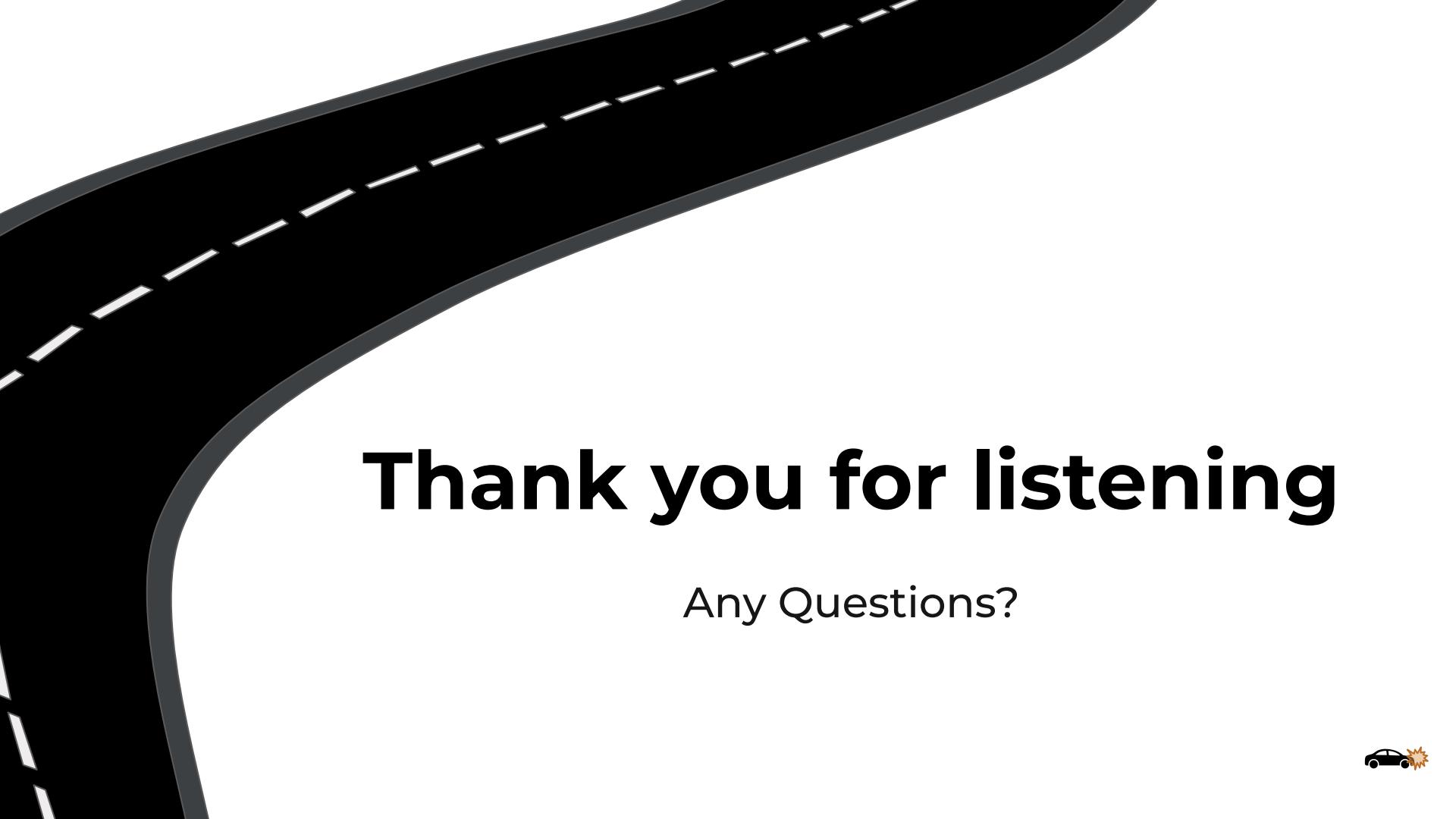


# Tests



# Scaling the Library





# **Thank you for listening**

Any Questions?



# Citations

## Context:

- average person experiences three to four auto collisions in their lifetime:  
<https://pl-law.com/what-is-my-probability-of-getting-into-a-car-accident/>
- Car Statistics:  
<https://www.ddlawtampa.com/resources/car-accident-statistics-you-need-to-know-in-2021/>

## US Road Accidents Data:

- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", 2019.
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.

Source for Classification of Heat Index: <https://www.weather.gov/ama/heatindex>