VISION ZERO CHICAGO

KEEPING THE STREETS SAFE

VISION ZERO CHICAGO

- 2012 – eliminate fatalities / severe injuries in 10 years
- **ALL** traffic accidents **ARE** avoidable
- 2016 – Work commenced
- Many major U.S. cities have joined the initiative.
  — S.F., L.A., N.Y., Austin, D.C.

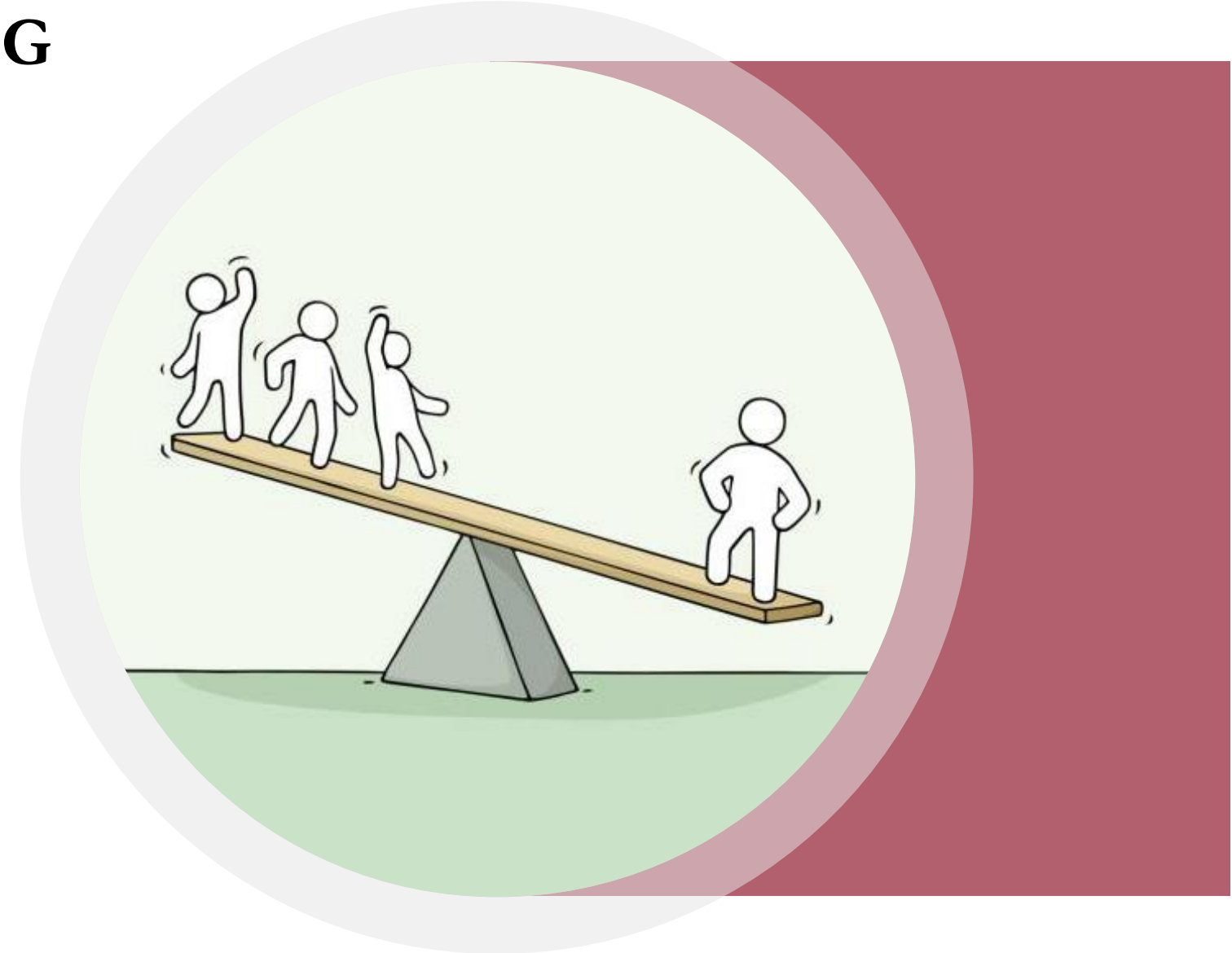
U.S. Facts:

- 3,700 deaths / day
- 20 to 50M suffer non-fatal injuries, often resulting in long-term disability
- Leading cause of death among people aged 5 to 29.

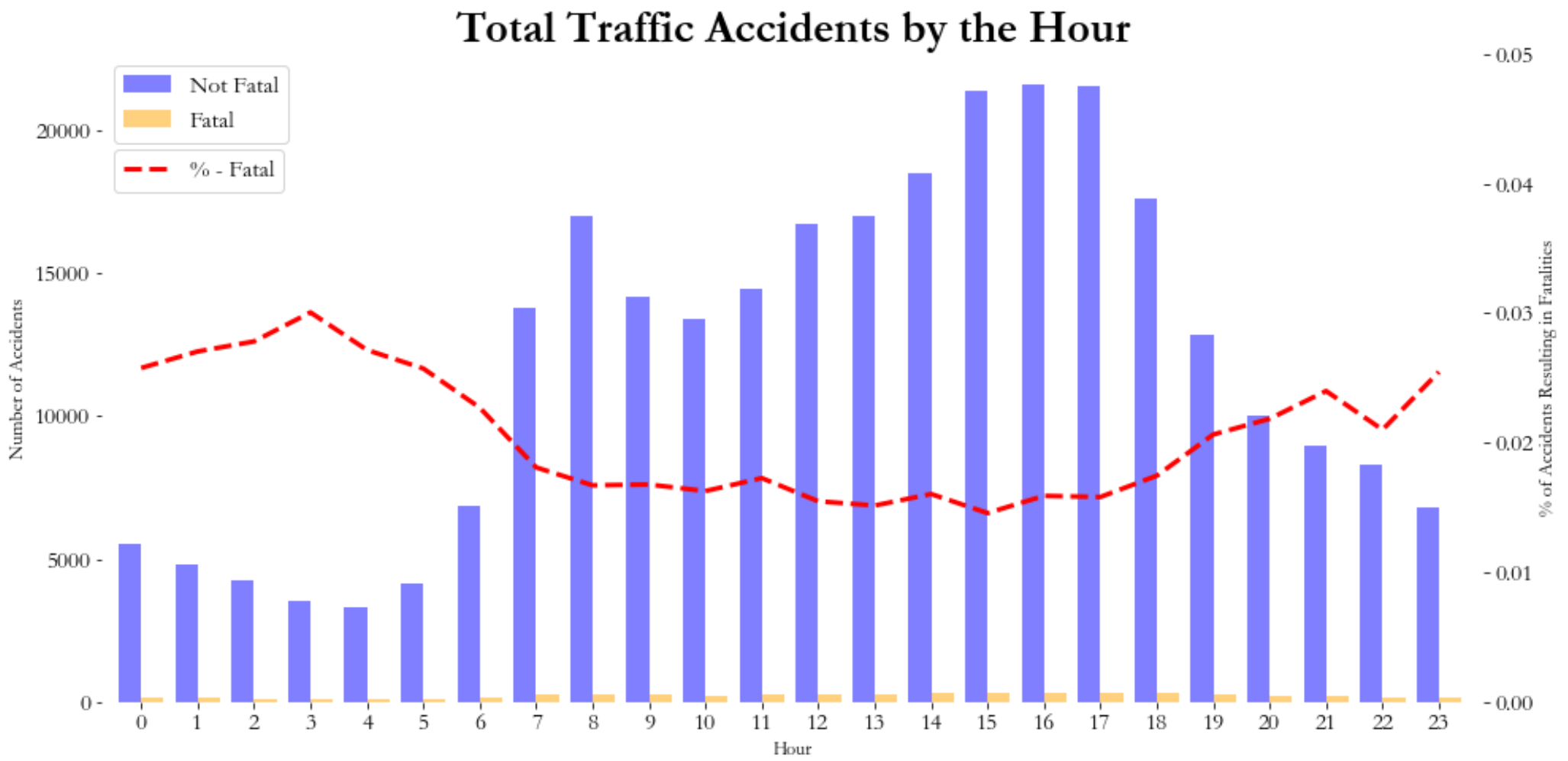
PREPARATION AND EXPORTATION

DATA AND PROCESSING

- Source - City of Chicago
- Sept ‘17 to Feb ’20
- >290k accidents
- ~50 features
- Created two classifications
  - “Non-fatal” and “Fatal”
  - Initially five
- Disregarded post-crash attributes
  - Prevented data leakage
- **Imbalanced dataset**
  - < 2.0% positive (i.e. fatalities)
  - Applied undersampling



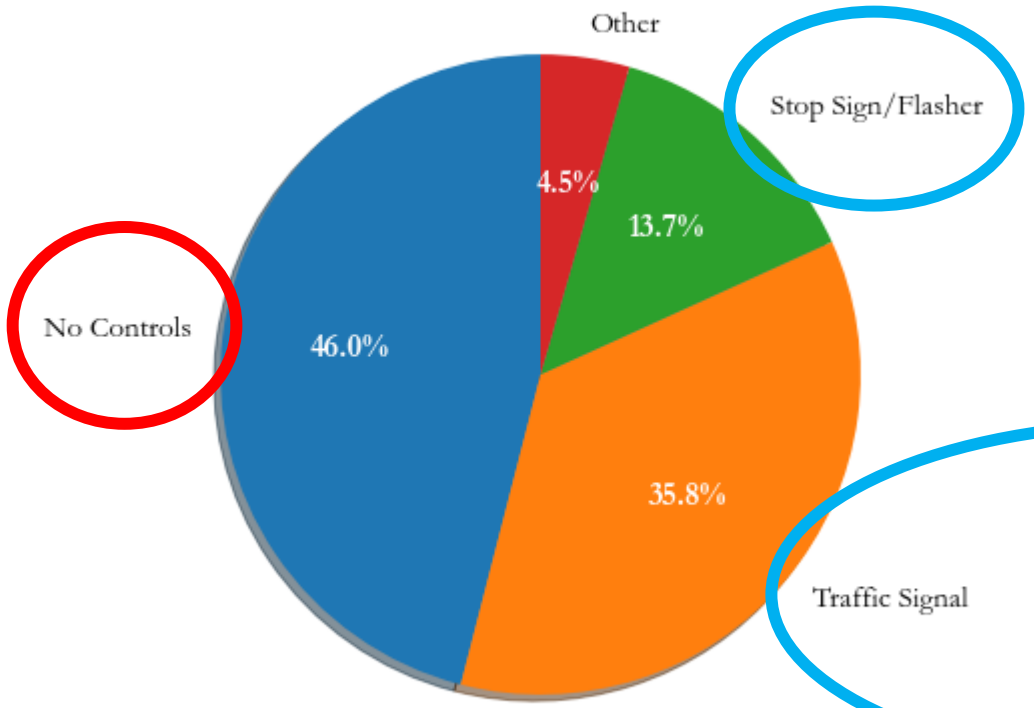
DOES THE SUN PROTECT DRIVERS? (1)(2)



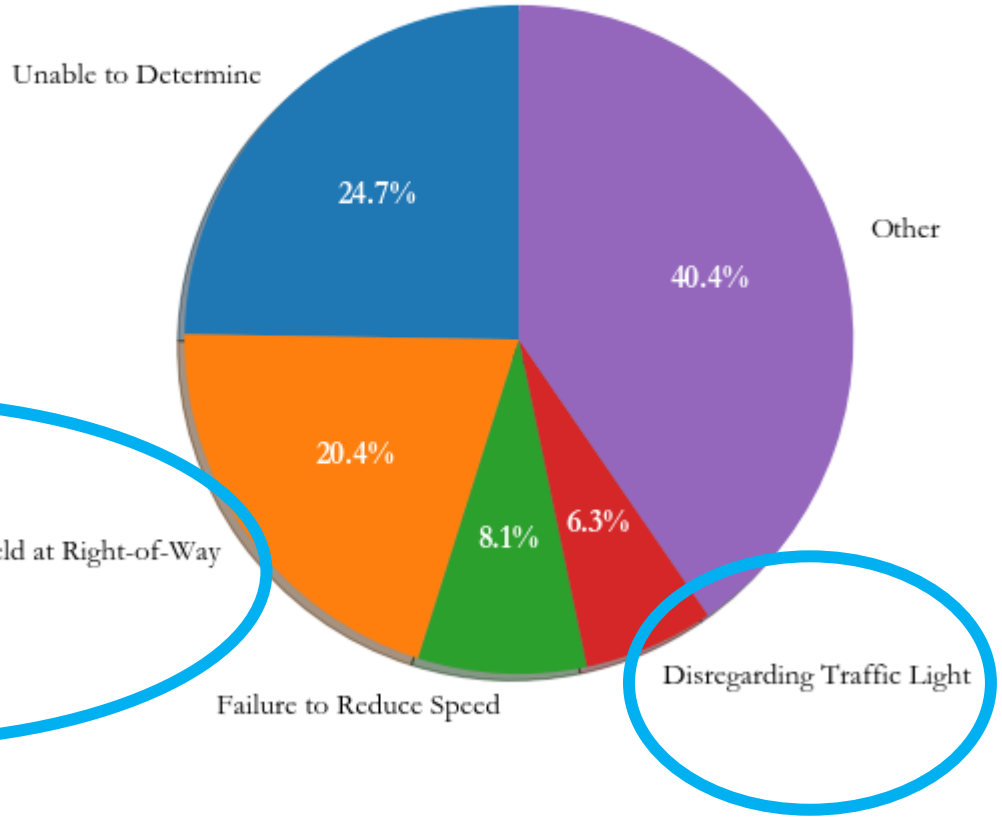
(1) Night and early AM is considered from 8:00 pm to 7:00 am.
(2) The percentage is derived from the hourly likelihood of a fatality.

DO CERTAIN STREET DESIGNS INCREASE RISK? (1)

Traffic Control at Scene



Primary Cause of Accident



(1) 5,358 accidents

MODEL EVALUATION AND SELECTION

GRADIENT BOOST – OPTIMAL MODEL

RESULTS / HYPERPARAMETES

Recall = 0.76

– Baseline models: 0.58 to 0.62

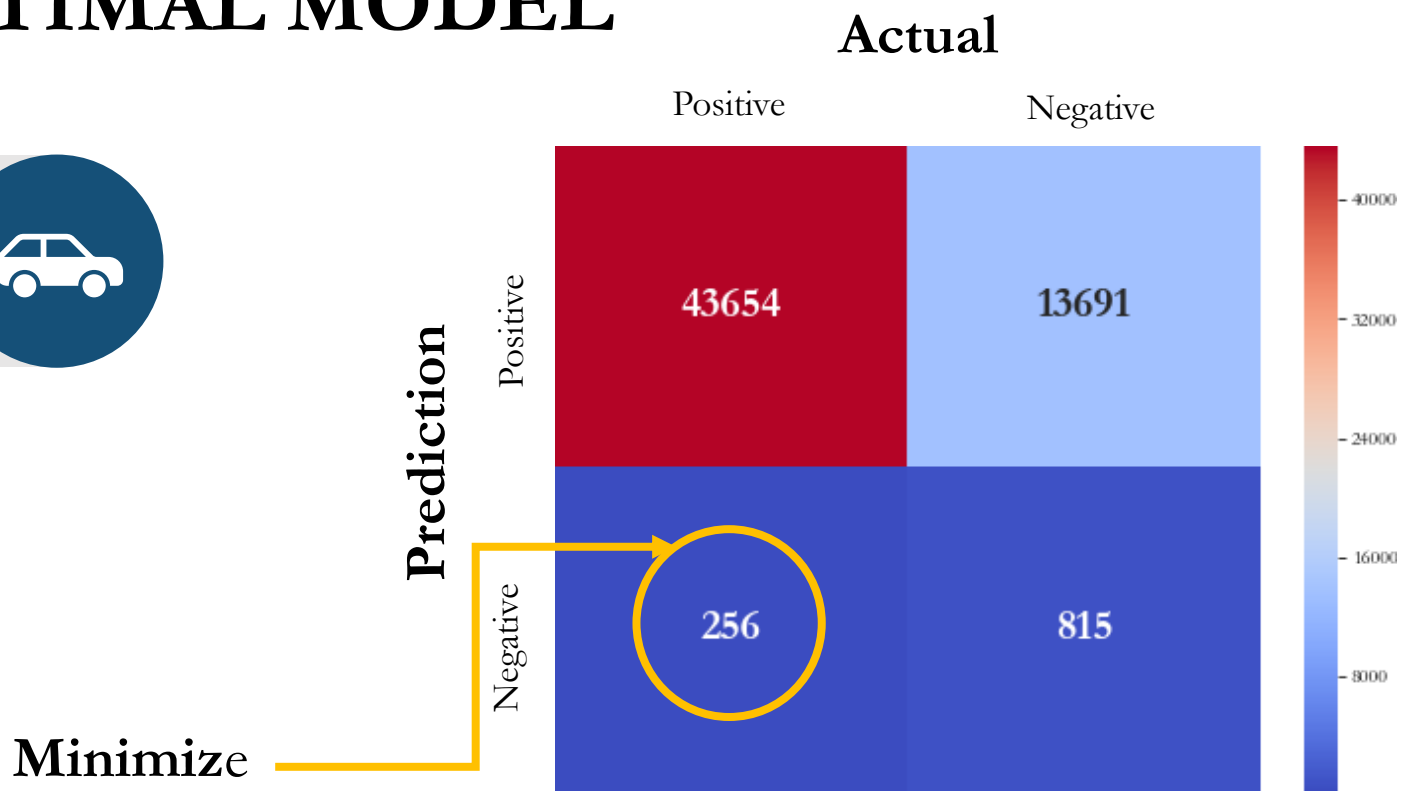
AUC Score = 0.813

Optimal Hyperparameters:

– Learning rate: 0.1

– Max. depth: 7

– Criterion: MSE

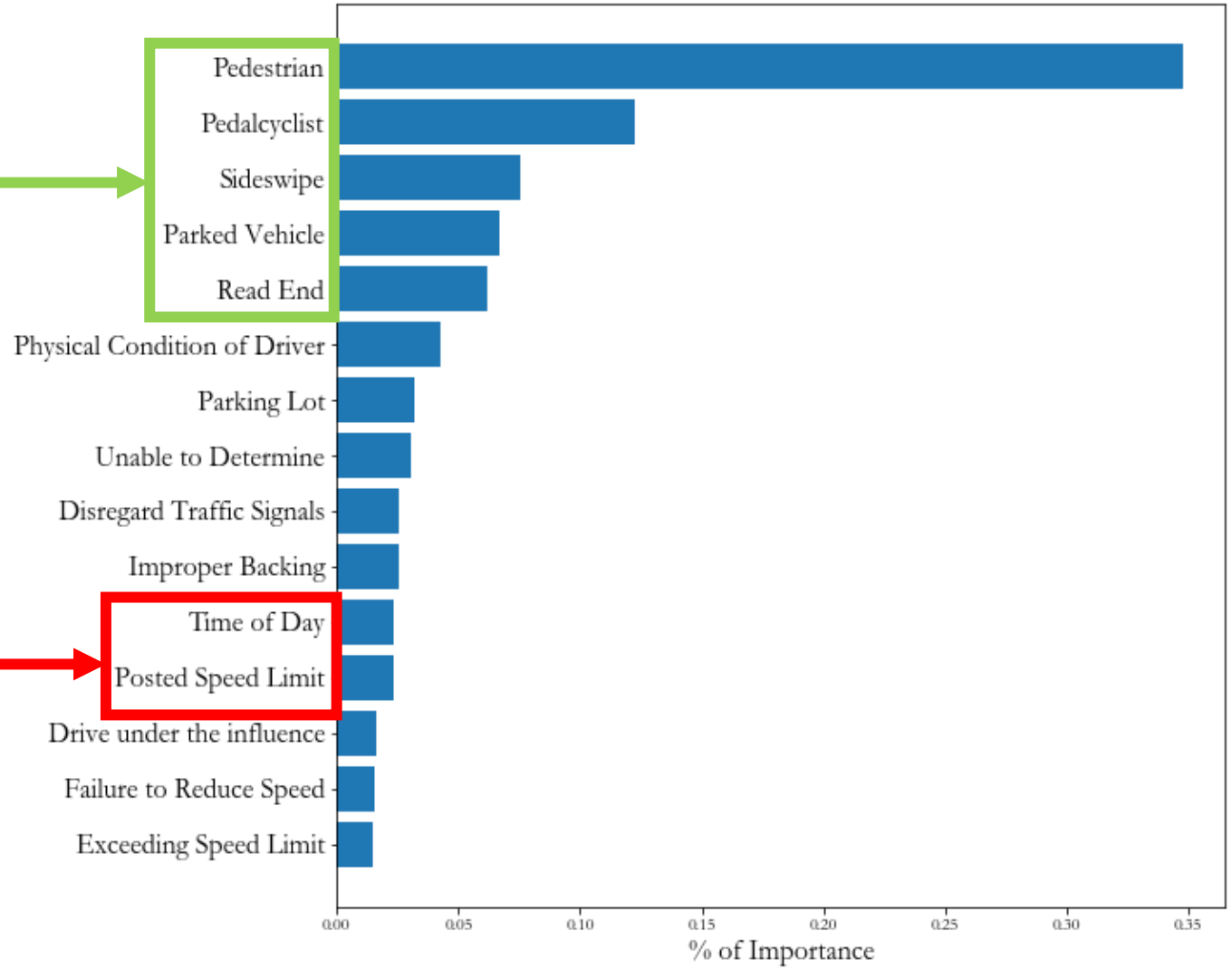


CONFUSION MATRIX

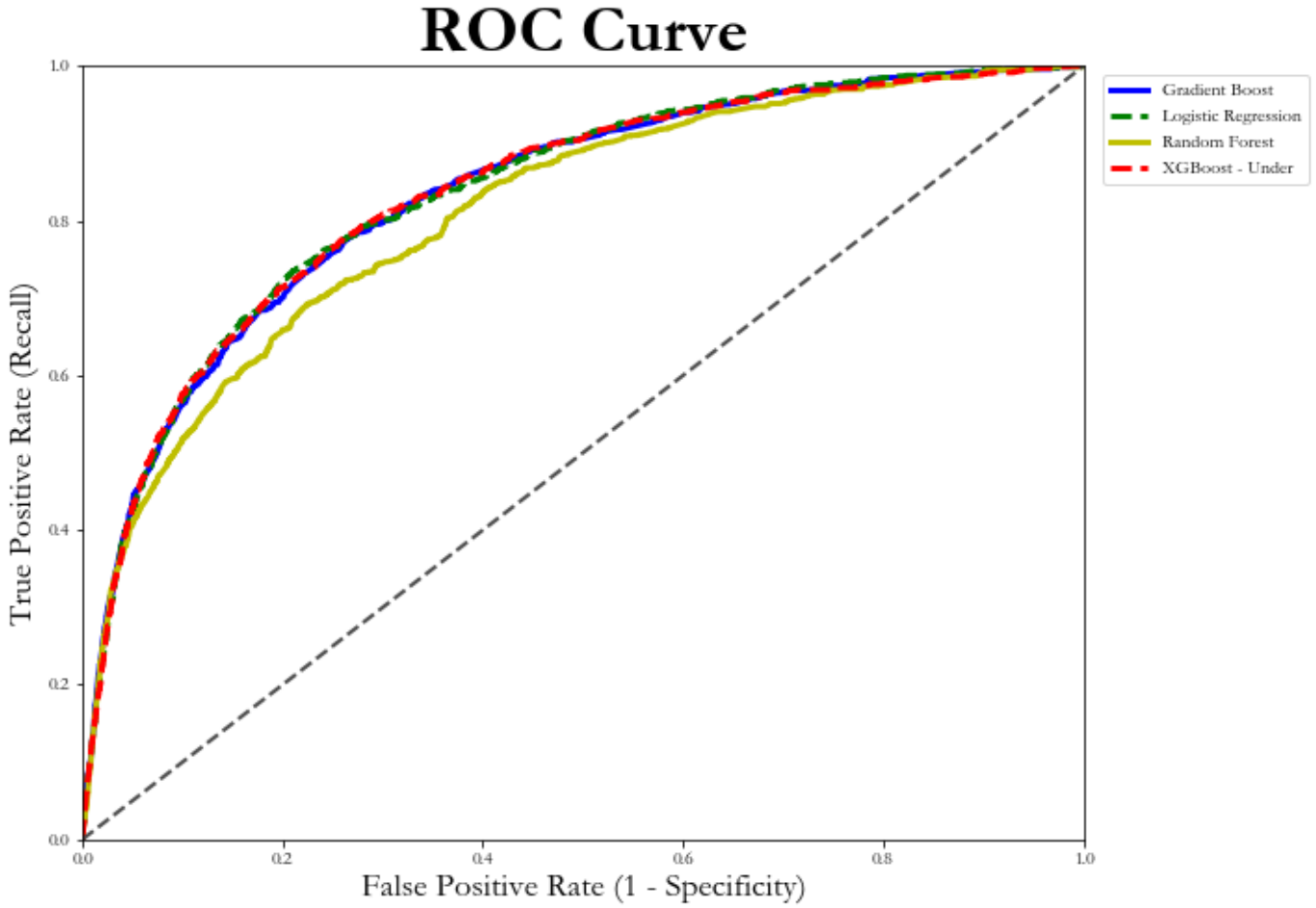
FEATURE IMPORTANCE – TOP 15 – GRADIENT BOOST

Initial points of contact.

Only numerical, i.e. not one-hot encoded, features.



MODEL COMPARISON



Model	AUC Score (1)
Gradient Boost	**0.813**
Logistic Regression	0.809
Random Forest	0.796
XGBoost	**0.813**

(1) Optimized model bases on AUC score using Randomized Search CV.

FURTHER EXPLORATION / SUGGESTIONS

EXPLORATORY / HYPOTHESIS TESTING

- Evaluate pedestrian and cyclist safety
  - Would the city benefit from pedestrian-only streets, e.g. S.F. Market Street?
  - Should more protection be provided for cyclists? Some European cities have bike lanes that are protected by parked cars.
- Perform hypothesis testing to see if accidents are more likely to result in a fatality accounting for the following features:
  - Intersections
  - Traffic controls
  - Time of day

MACHINE LEARNING

- Add geospatial mapping to include traffic volume.
- Include socio-economic data.
- Add traffic volume and traffic light / speed camera violations to the dataset.
- Incorporate traffic policing protocols.

THANK YOU

MORGANPSELL@GMAIL.COM