# Analyzing of Seattle road collisions and predicting their severity

# Analyzing and predicting road collisions is valuable for the City of Seattle

- Collisions impact many people: up to 480k people in 15 years in Seattle (610k inhabitants)

- It can lead to major properties damage and serious injuries

# Data acquisition and cleaning

- Data were acquired from the City of Seattle website: https://data.seattle.gov/Land-Base/Collisions/9kas-rb8d

- Data were formatted, continuous features were binned, new features were created, missing values were whether dropped or filled with the most common value, dummies variable were then created

- Finally, 31 features were used, with 134,843 instances for the train set and 33,863 for the test set

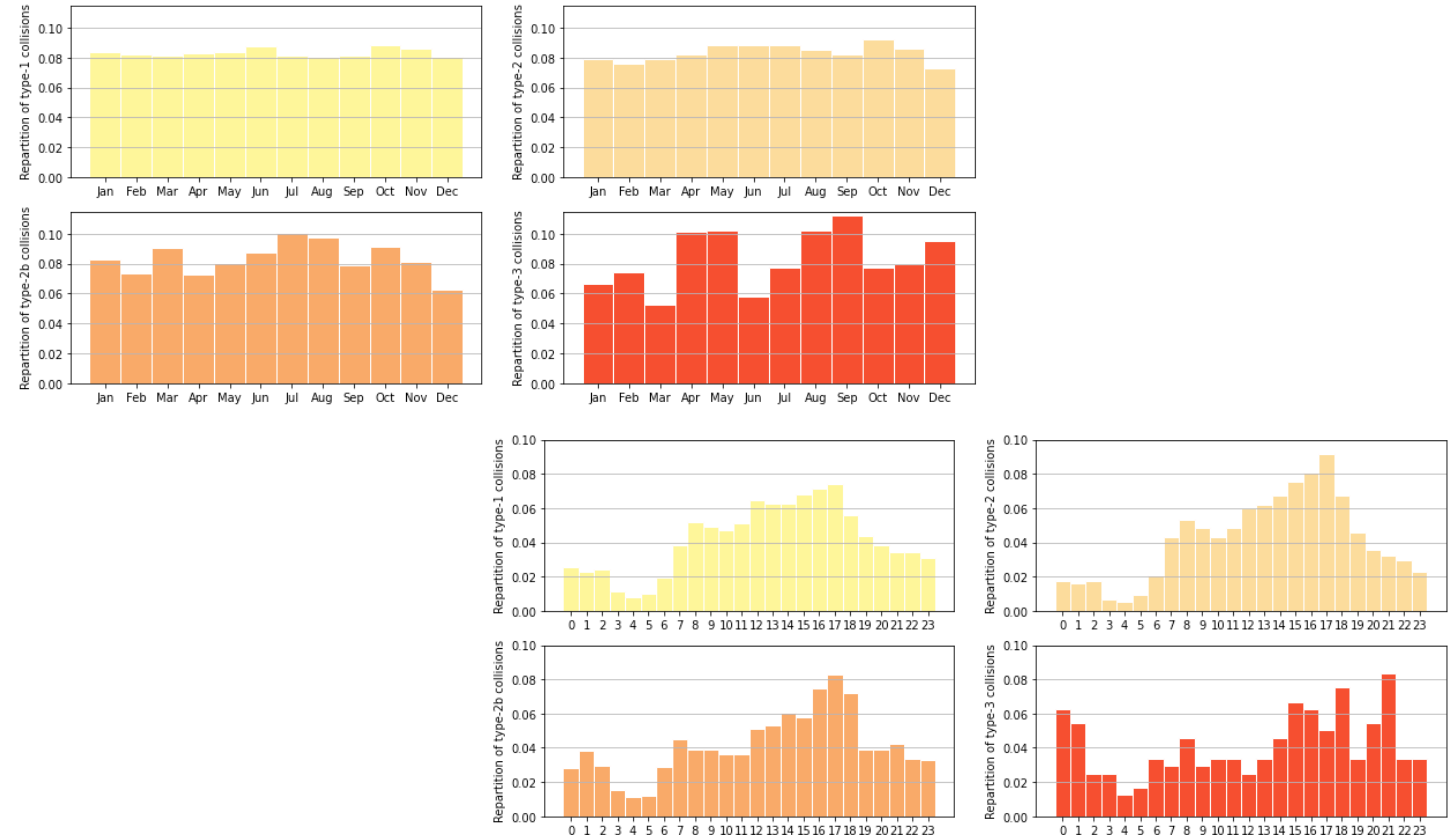# As we might think, collisions happen where there is traffic

## *Because of the place*

- Collisions mostly happen downtown and on big streets or avenues
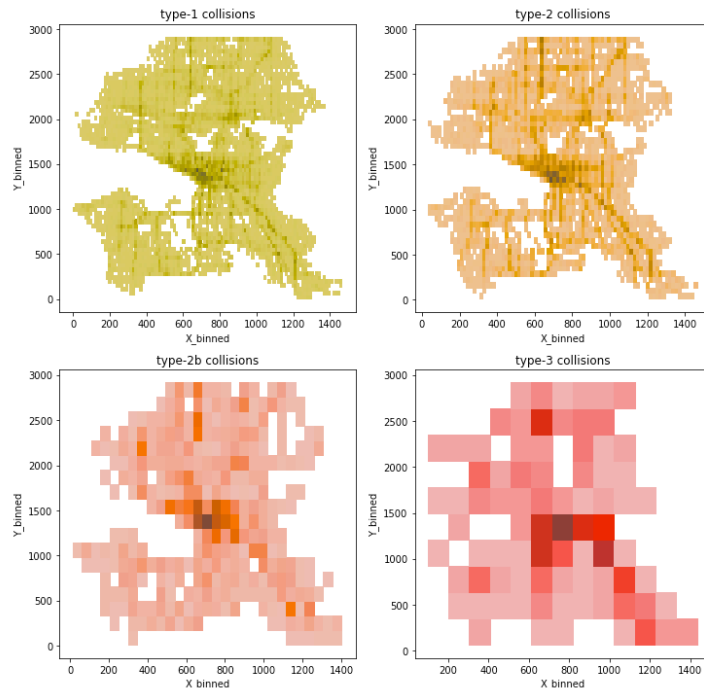


## *Because of the moment*

- Collisions happen less on weekend or during the night

# Collisions also happen when people pay less attention

## *Because they are tired*

- Collisions happen more on Fridays, at the end of the work-week

- Collisions happen more at 17h, at the end of the work-day

## *Because the driving conditions are good*

- More than 80% of the collisions happen when the weather is good

- More than 70% of the collisions happen when the road conditions are good

- More than 65% of the collisions happen in daylight

⇒ when driving conditions are good, people pay less attention than when they are bad, leading to more collisions
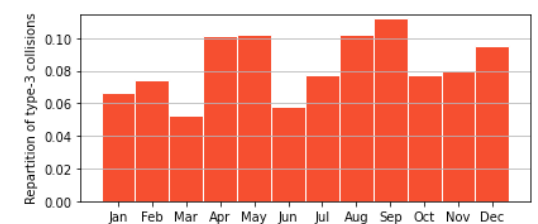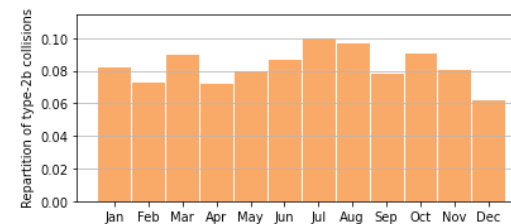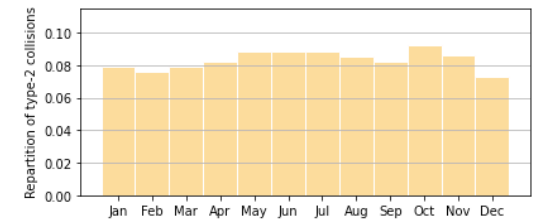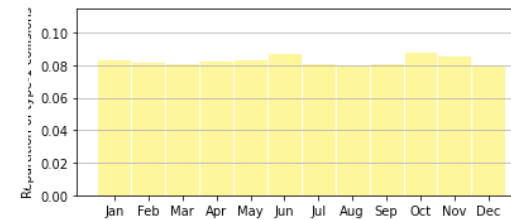
# Collisions also happen more on June, October and November

Collisions are not uniformly distributed across month (chi-squared test)
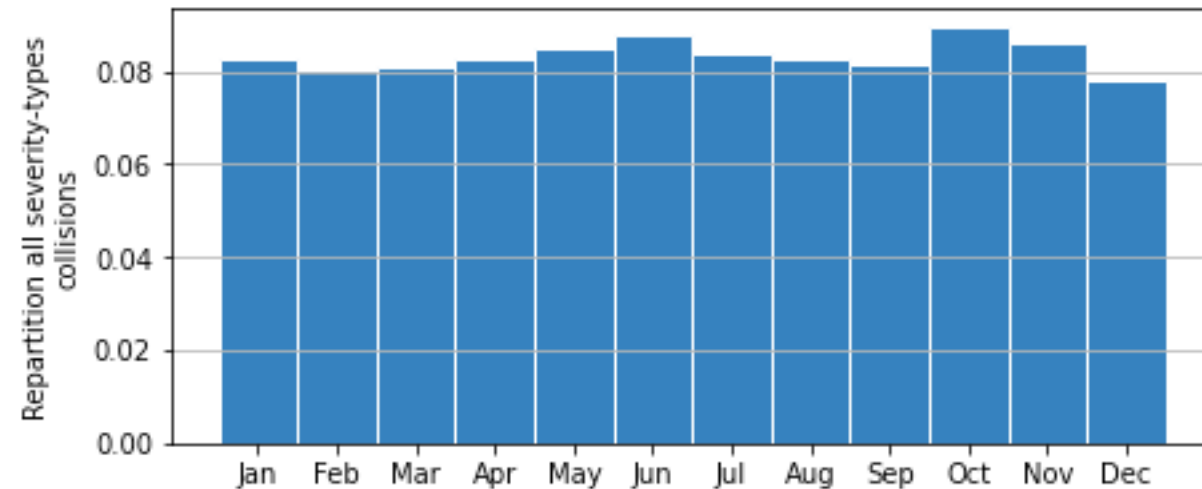


o They happen more on June, October and November
o And less on December

This statement stands for each severity-type collisions, except for the deadly ones (due to small amount of data)

# Predicting collision's severity to propose policies to prevent them

- 2 models used: a SGDClassifier (the best model here) and a DecisionTree (the easiest recommendation-giving model)

- Both gives the same results

- The most severe collisions can be characterized by:
  - involving pedestrians or cyclists, or
  - involving only 1 vehicle, or
  - happening at intersection, or
  - being head-on
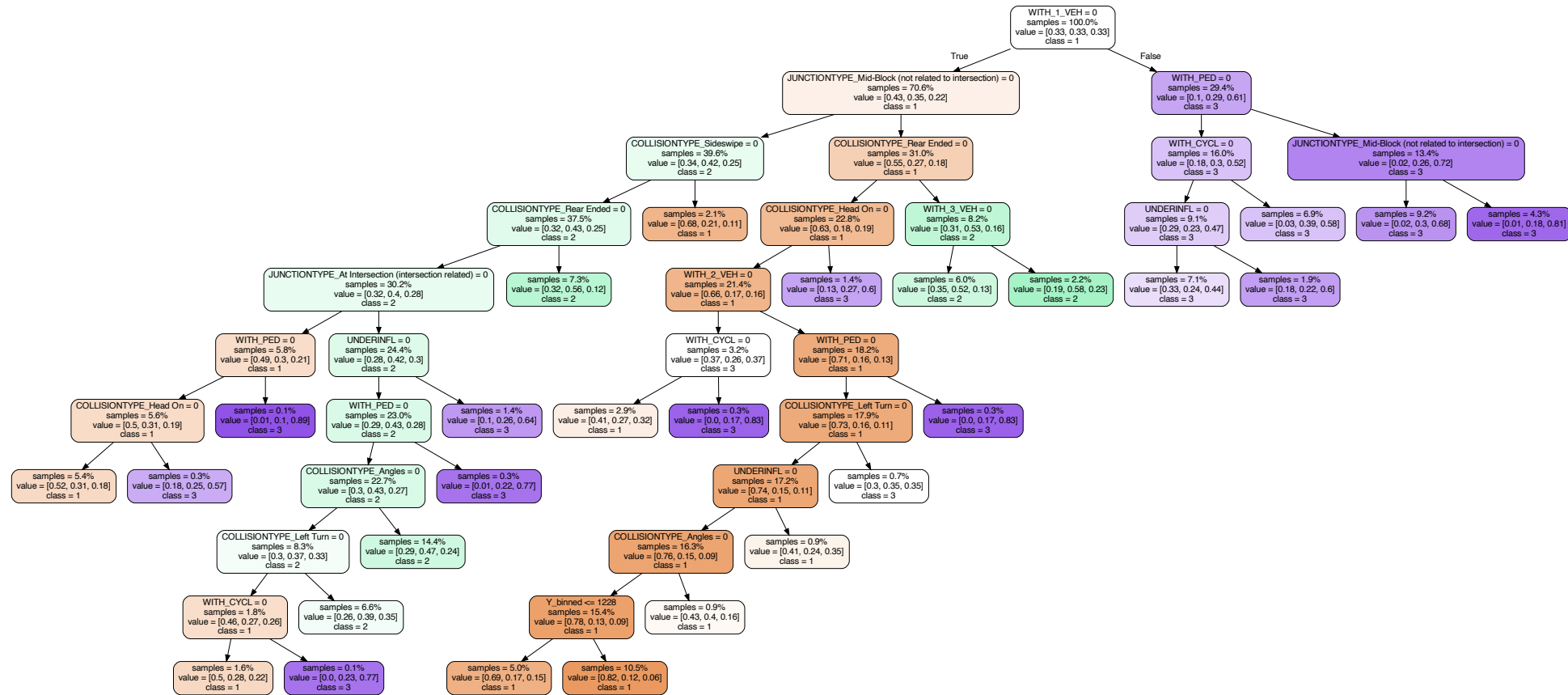
# Features of importance for the two models

**Impact and magnitude of features to predict most severe collisions with SGDClassifier**
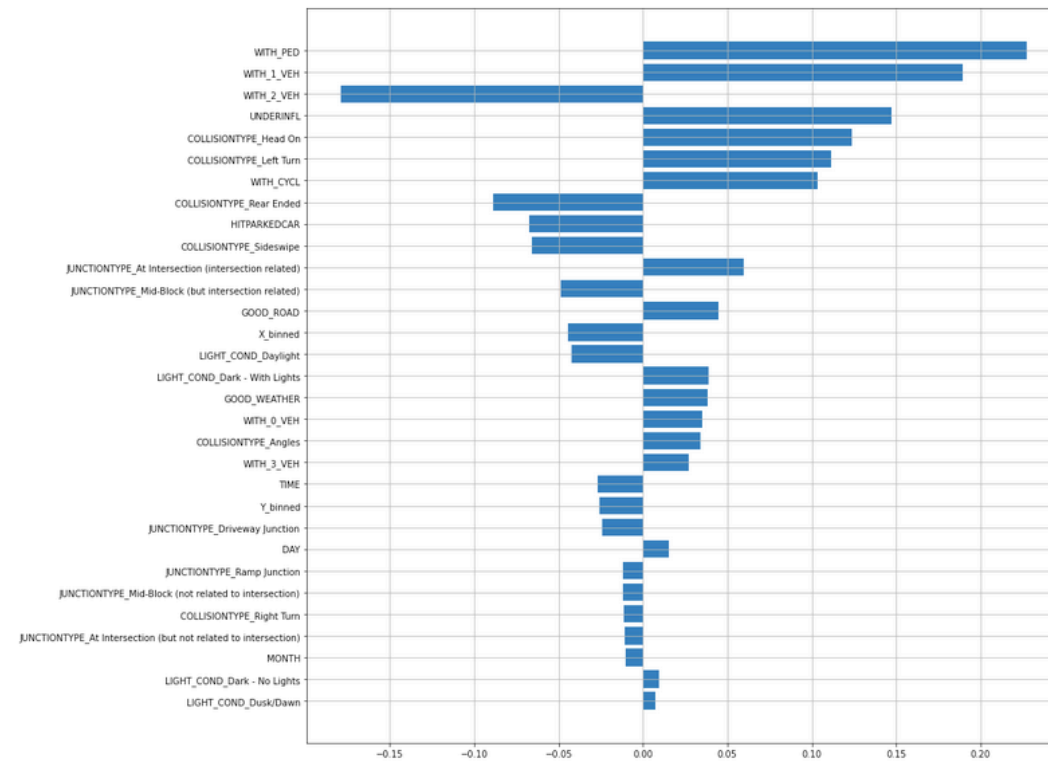
**Importance of features to predict severity of collisions with DecisionTree**
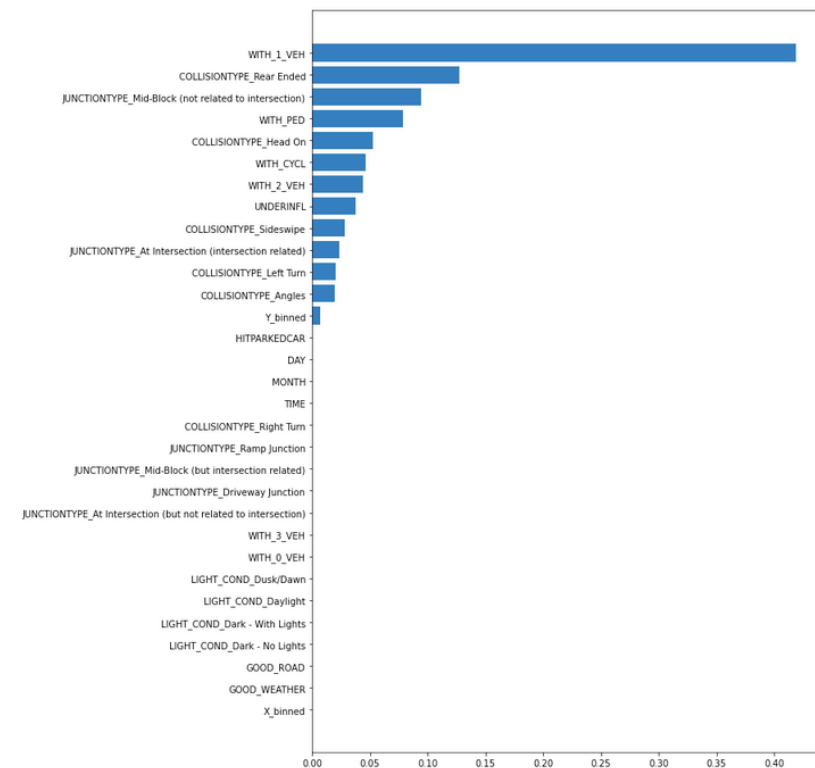
# Tree of decisions

# Conclusion

**Policy proposals to reduce collisions**

- Deploy police force more downtown or on big avenues, rather than on small streets

- Deploy police force more on Friday, rather than the weekend

- Deploy police force more around 17h, rather than around 4h

- Deploy police force more in June, October and November, rather than in December

**Policy proposals to prevent most severe collisions**

Deploy police force more, or/and create/develop urban amenities to:

- Improve sidewalk safety

- Create specific cycle paths

- Develop speed radars on head-on collision-prone blocks

- Improve safety of intersections (e.g. with radars)