PREDICTION OF HIT-AND-RUN ACCIDENTS IN NASHVILLE, TN

Zimu Su, Metis DSML bootcamp classification project

Hit-and-run accidents

- occur if someone who causes the accident leaves the scene without providing personnel information.

- cause severe injuries and properties damage without any penalty or blame.

- are considered as supplemental crimes.

The project aims at…

➢ identifying the significant factors to hit-and-run accidents in Nashville area.

➢ predicting probability of hit-and-run accident based on accident feature (collision type, location, etc.)

➢ Providing suggestion for Nashville Metro Police to decrease rate of hit-and-run accidents.
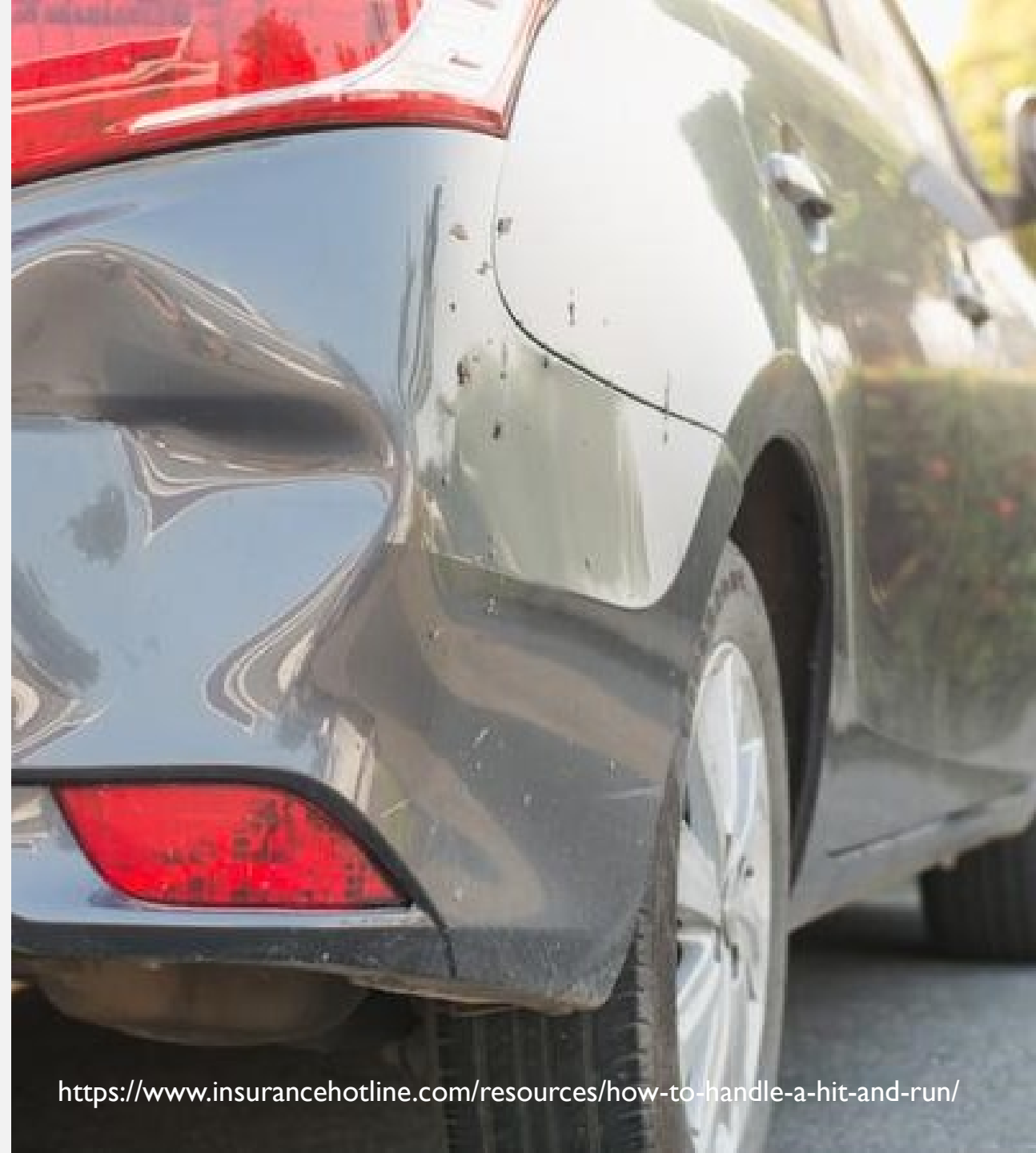
Nashville Downtown (photo by myself)

Data.Nashville.gov

Data Source:

- Nashville Open Data Portal:

  https://data.nashville.gov/Police/Traffic-Accidents/6v6w-hpcw
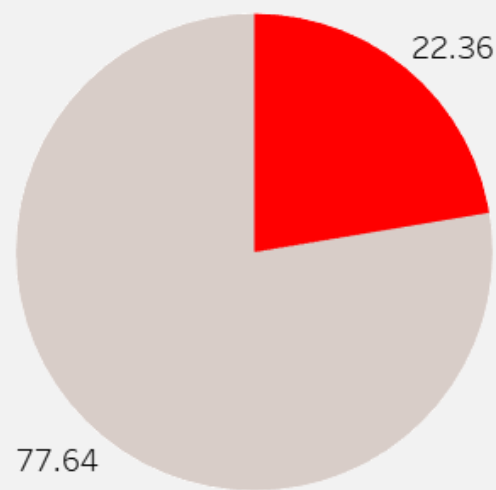
Data range:

- 01/01/2015 – 04/10/2022

- 213424 accidents information in total

Main Features:

- Date and Time, Number of Motor Vehicles, Number of Injuries, Number of Fatalities,

- Property Damage, Hit and Run, Collision Type, Harm type

- Weather, Illumination.

- Street Address, ZIP, Latitude, Longitude
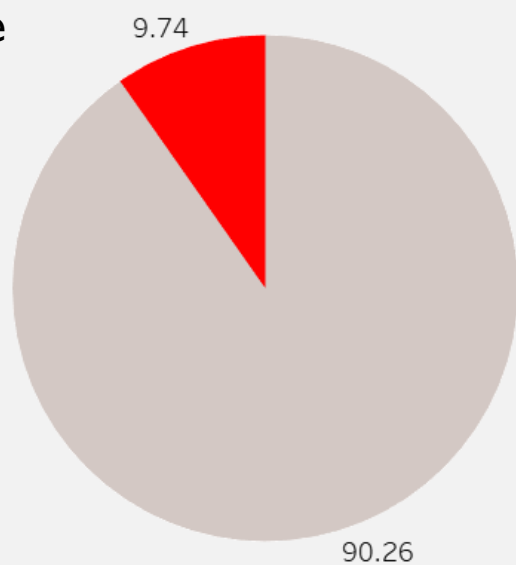
HIT-AND-RUN ACCIDENTS STATISTICS:
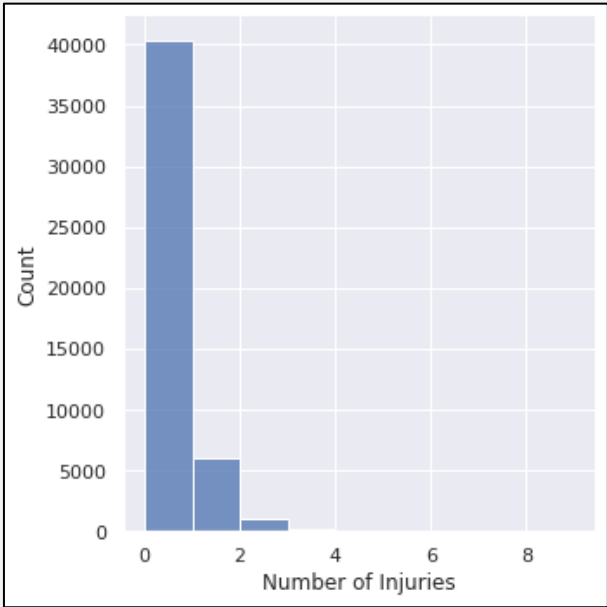
Ratio of hit-and-run
accidents:

22.36

77.64

Mostly 2 vehicles involved
in hit-and-run accidents:
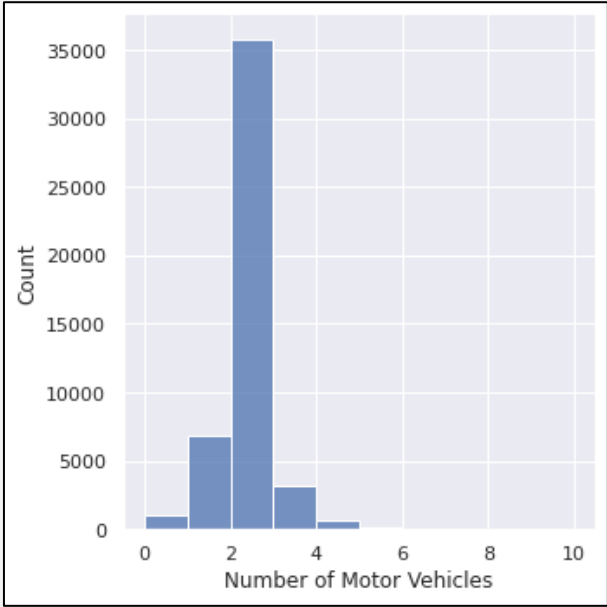
Ratio of property damage
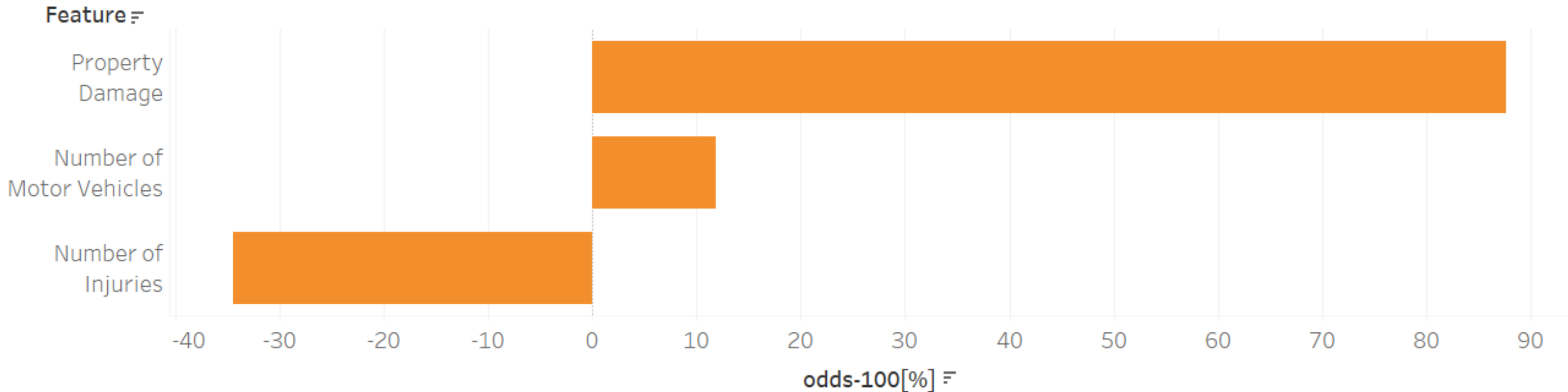in hit-and-run accidents:

9.74

90.26

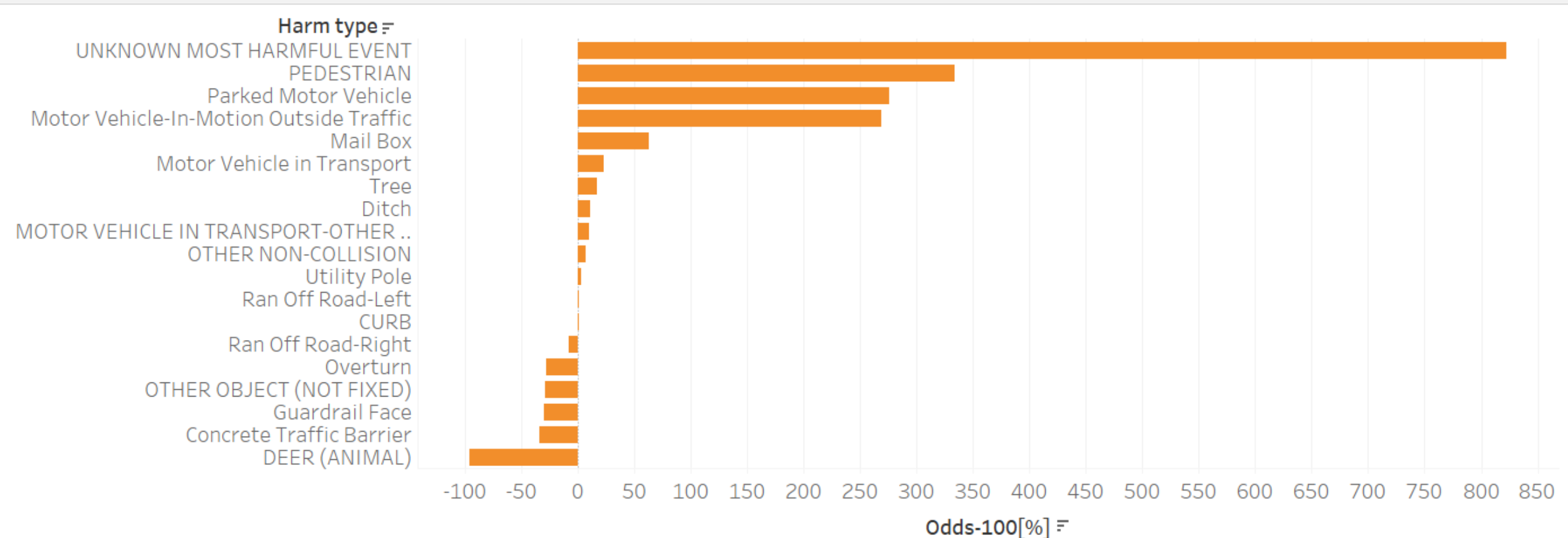~1/8 hit-and-run accidents
have injuries:

ALGORITHM AND TEST SCORES

- **Feature selection:**

  ➢ Number of Motor Vehicles, Number of Injuries;

  ➢ Property Damage, Collision Type, Harm type;

  ➢ Weather, Illumination condition.

- **Feature engineering:**

  ➢ One hot encoder for the features of collision type, weather, illumination, harm type.

  ➢ Standardize the continuous features: Number of Motor Vehicles, Number of Injuries.

  ➢ 83 feature in total.

- **Logistic regression algorithm with regularization (C=0.1) and balanced weight.**

  ➢ Optimal f1 scores: 0.456 at prob decision of threshold >=0.503, ROC AUC score: 0.73. Accuracy: 0.71.

- **Random forest algorithm.**

  ➢ Optimal f1 scores: 0.456 at prob decision of threshold >=0.249, ROC AUC score: 0.73. Accuracy: 0.71.
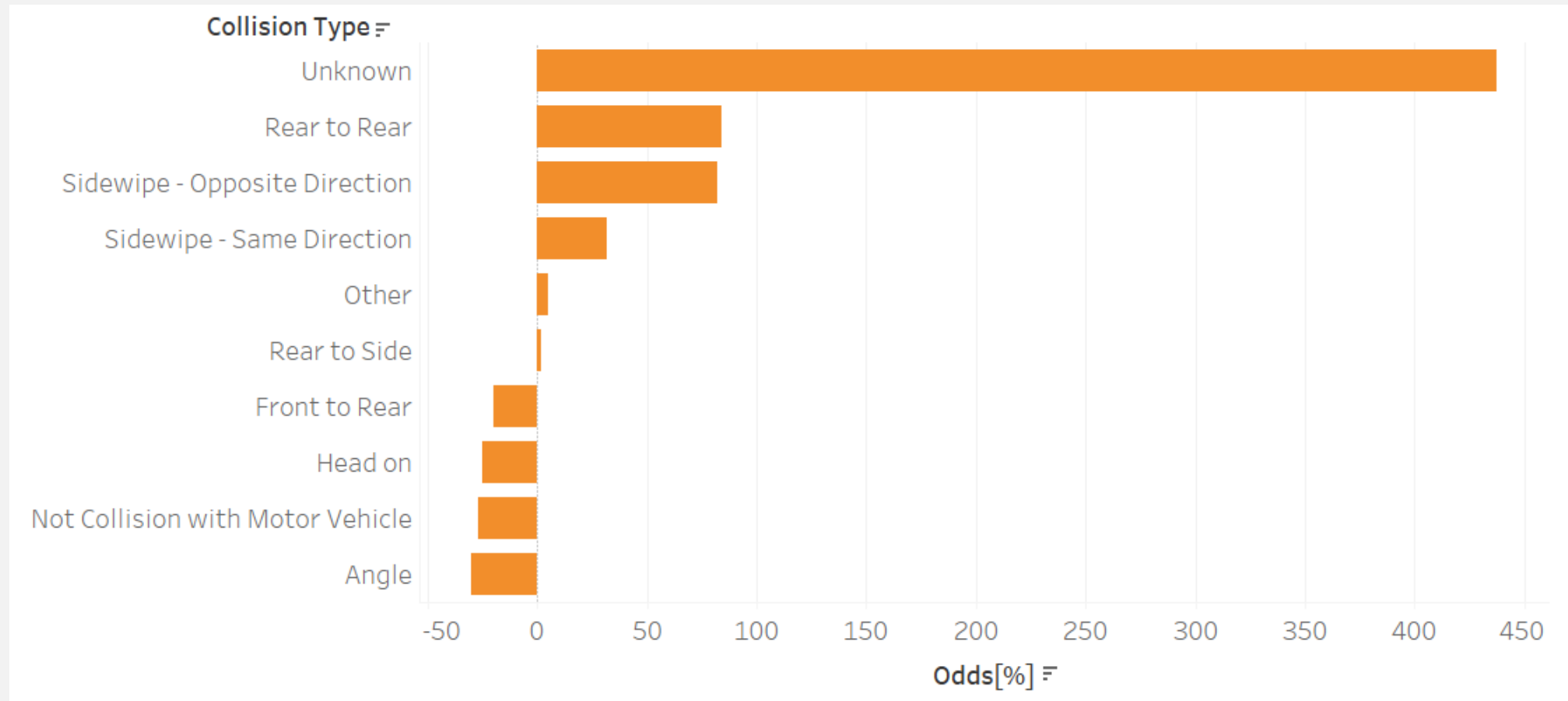
- Hit-and-run accidents are more likely to
  - ➢be caused with property damage (~1.9 odds)
 - ➢involve more vehicles (~1.1 odds)

- Hit-and-run accidents are more likely that
  - have unknown type of harm (~9 odds)
  - Pedestrian are hit (~4.5 odds)
  - occur for parked motor vehicles (~3.5 odds)
 - auto in outside traffic (~3.5 odds)

- The collision type of hit-and-run accidents are mostly unknown(5.5 odds), rear to rear (~2 odds), opposite sidewipe(~2 odds).
- The unknown type might be responsible for pedestrian or collision with parked motor.

- Besides unknown illumination, dark illumination is more likely to lead to hit-and-run accidents.
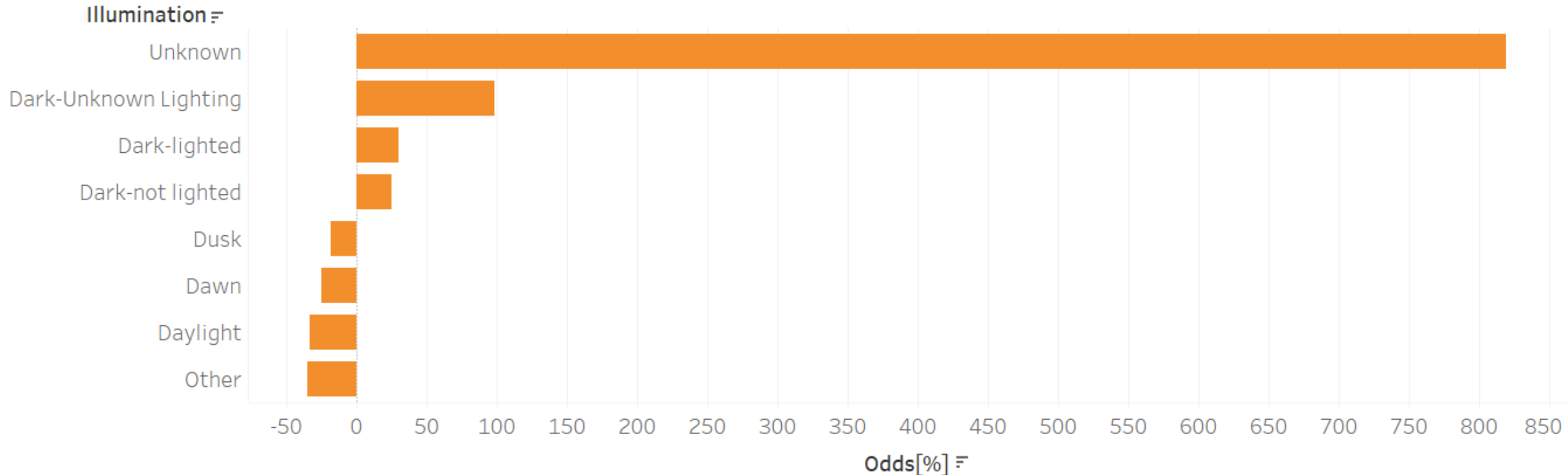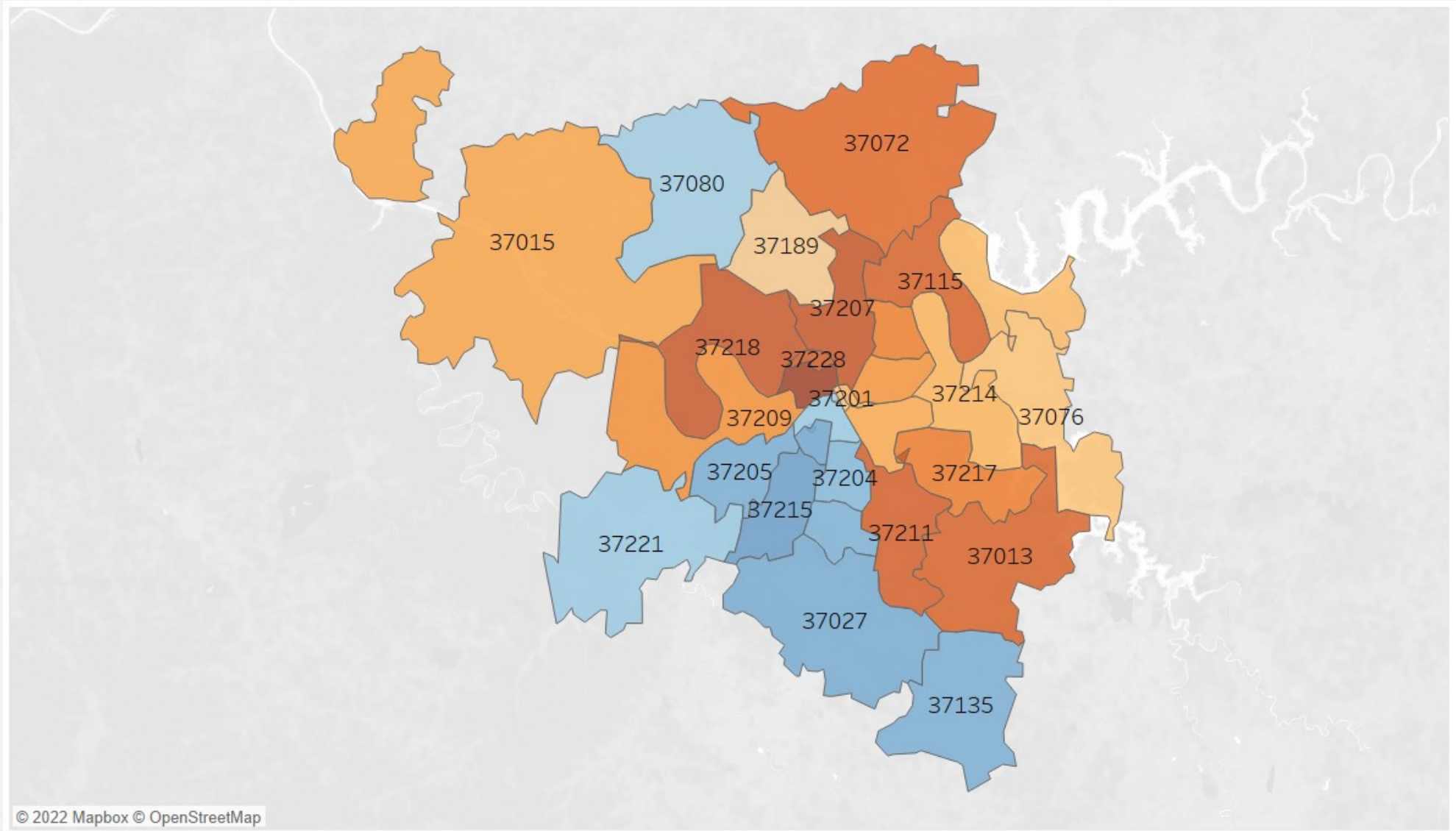
- Cloudy, fog, blowing snow are more likely to cause hit-and-run accidents.
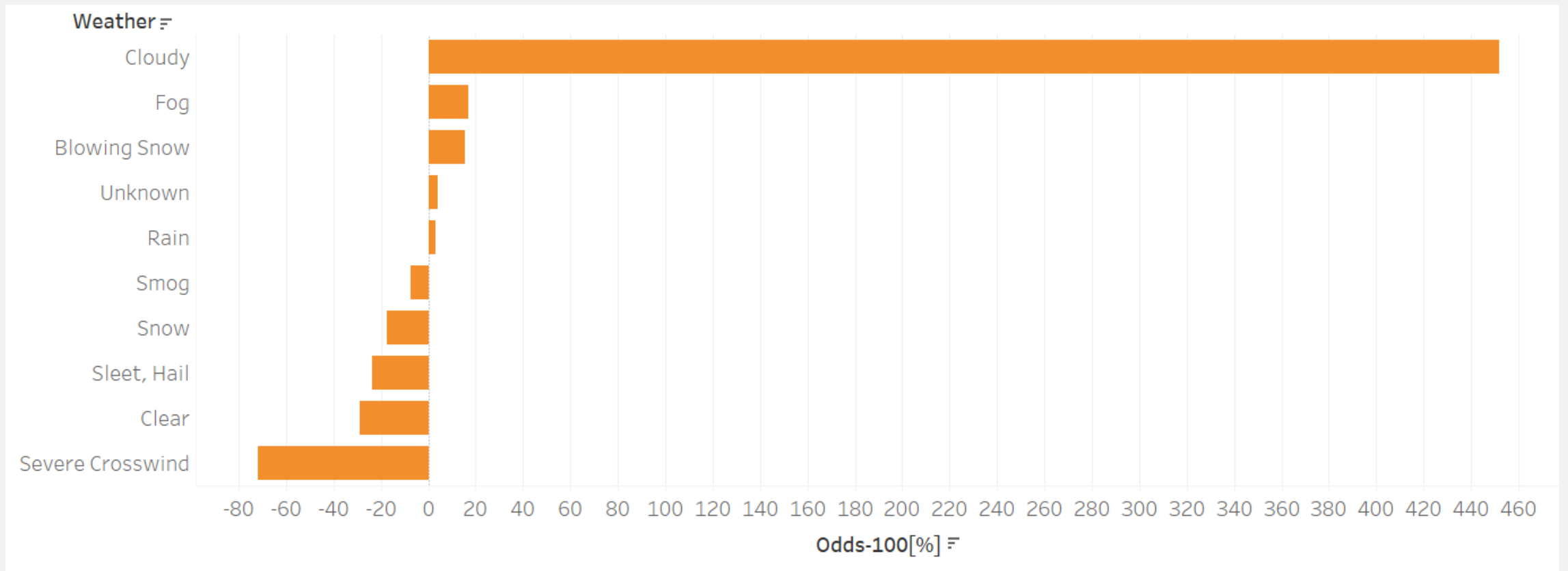
Odds of hit-and-run accidents

- Next step:
  - ➢ Check exact the meaning of each feature (such as unknown illumination condition);
  - ➢ Do more feature engineering to improve f1 score.
  - ➢ Check exact location of hit-and-run accidents (crime rate, road condition, etc.) and related to illumination condition (light or not at night)

Thank you!