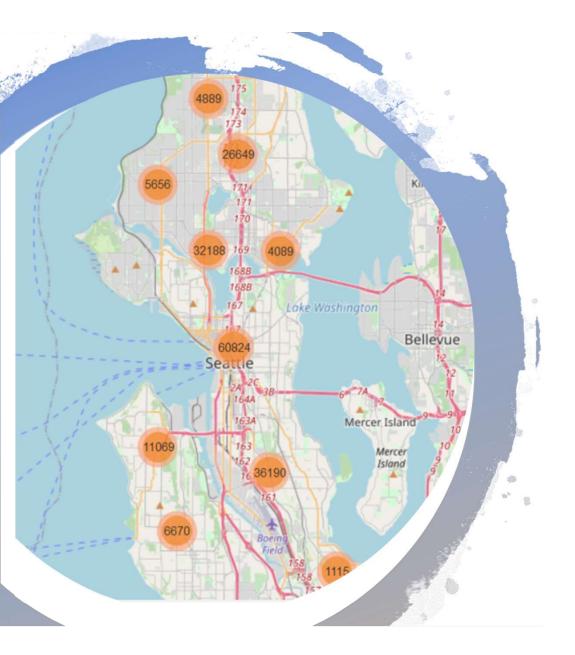


#### Problem statement

- Many car collisions occur every month causing damages to people's lives and economy.
- 2. Reasons of the incidents may include:
  - Bad organization of the traffic flows in certain places;
  - Poor road conditions in some locations;
  - Bad weather conditions;
  - Inattention of drivers;
  - And others
- 3. In order to take preventive measures, need to study statistical data and find the reasons.

# Data acquisition and cleaning

- The dataset includes information about 194,673 collisions of all types in Seattle from beginning of 2014 to May of 2020.
- 38 attributes in the initial dataset, 17 will be used for the analysis.

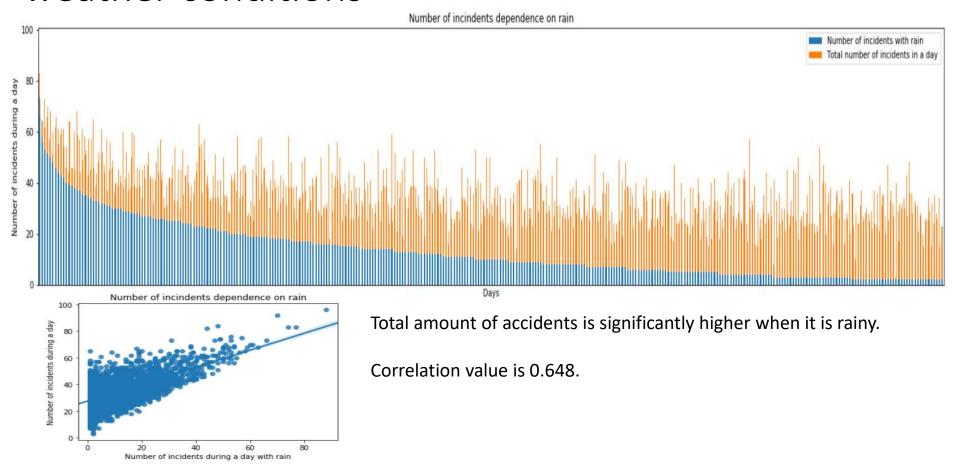


Large amount of accidents in certain locations

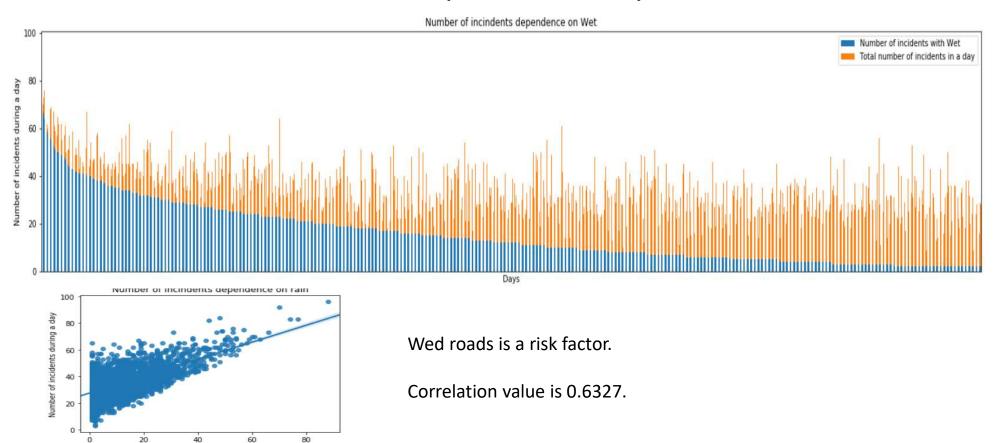
Many incidents occur in same places year over year.

Every place has to be studied individually to find its specific reasons.

# Rain is the most contributing factor among other weather conditions

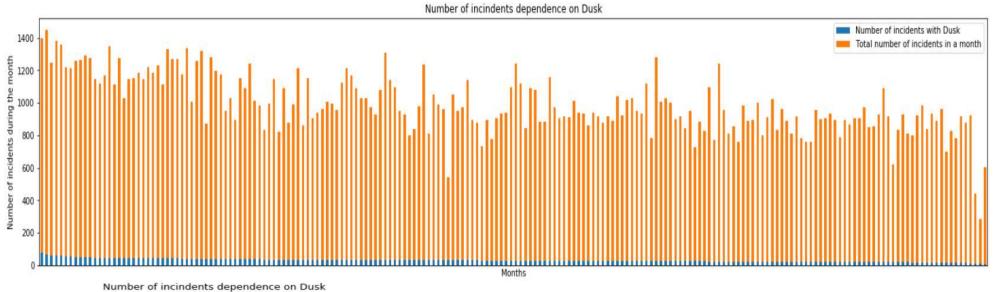


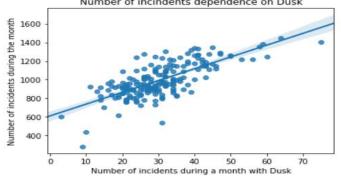
## Wet roads increase probability of an accident



Number of incidents during a day with rain

#### High correlation of accidents with dusk time



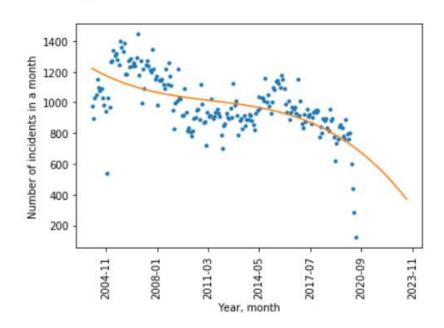


But low contribution to the total amount.

Correlation value is 0.726853.

## Decreasing trend needs to continue

3 2 -8.833e-24 x + 3.466e-14 x - 4.575e-05 x + 2.132e+04 The R-square value is: 0.39604732169598056



Regression analysis is not suitable for predicting values on extended time scale.

#### Conclusions

- 1. Certain places in the city need special attention in terms of organizing the traffic.
- 2. Rain, snow and wet and icy roads are massive contributing factors to the amount of collisions.
- 3. Special care needs to be taken by drivers in such conditions.
- 4. The model for predicting the frequency of accidents will not work on extended time scale.
- 5. Regression analysis of categorical variables can be considered as a further study.