



# Car Accident Severity Project

# HELLO!

**I AM JETHIN CHANDRA**

I am here to give a presentation of this project.

You can find me at  
[@jethinpilli](#)





# Introduction

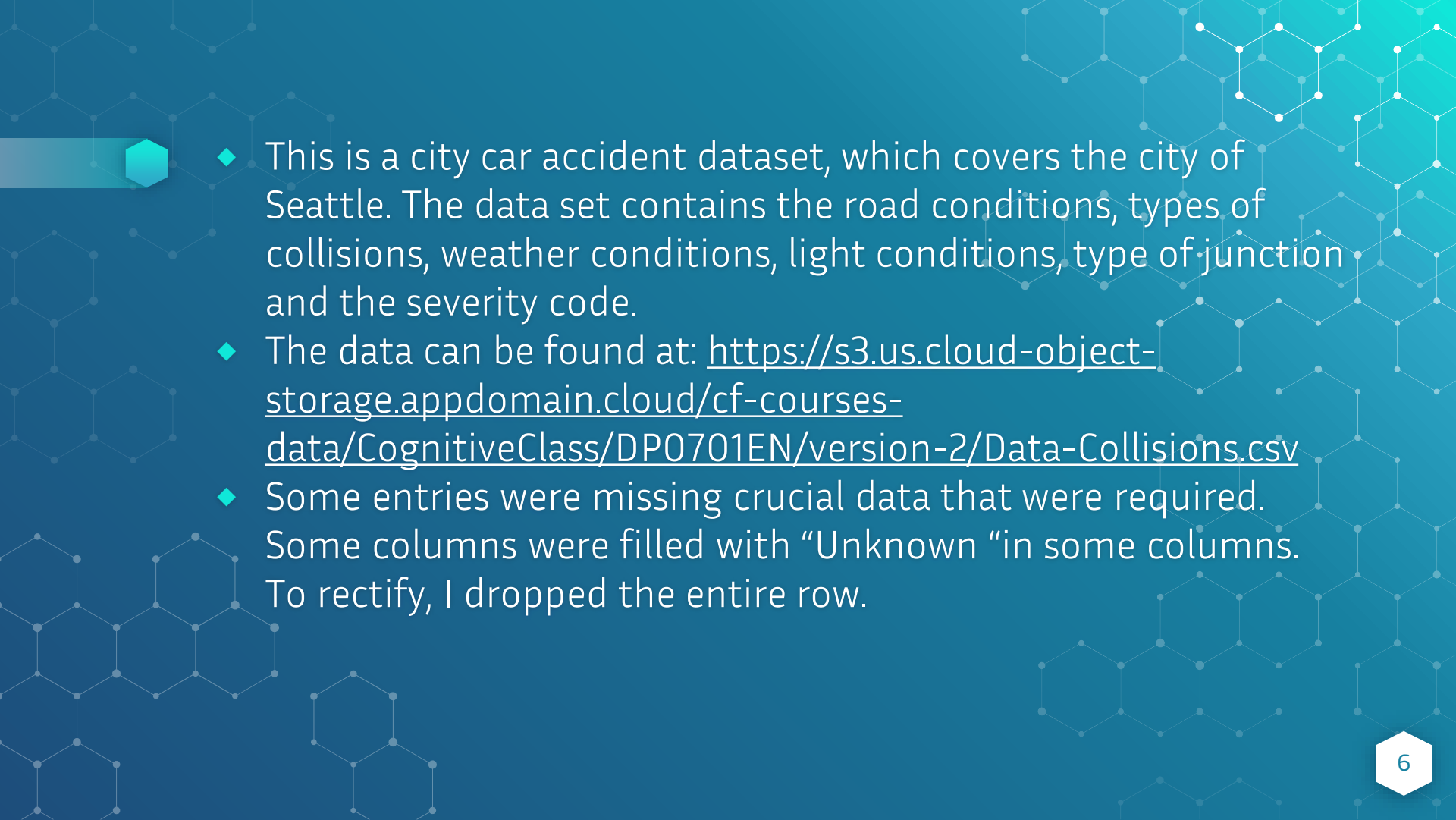
- ◆ The main business problem is about collisions in Seattle and conditions causing them. Intention is to build predictive model which will provide reasonable prediction if under certain condition is higher probability of collisions with certain severity (certain places on exact days) and which are the mostly involved on collision. Audience consists of people who lives in Seattle or travel through it and goal is to provide them information about current situation on a roads and possible dangers. Audience should be interested in this problem because knowing the relationships between conditions and likelihood of collision can save their money and life.





2

Data

- 
- ◆ This is a city car accident dataset, which covers the city of Seattle. The data set contains the road conditions, types of collisions, weather conditions, light conditions, type of junction and the severity code.
  - ◆ The data can be found at: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>
  - ◆ Some entries were missing crucial data that were required. Some columns were filled with “Unknown “in some columns. To rectify, I dropped the entire row.



3

# Methodology



# Exploratory Data Analysis

## ◆ Weather Condition

The proportion of L2 was higher when the weather condition is not good.



## ◆ Road Condition

The proportion of L2 was higher when the road condition is not good.



This shows that Weather and Road conditions certainly have a big effect on accident severity.





4

# Predictive Modelling

## Algorithms Considered for classification were KNN algorithm, Decision Tree classifier and Logistic Regression.

### KNN Algorithm

A k-nearest-neighbor algorithm, often abbreviated k-nn, is an approach to data classification that estimates how likely a data point is to be a member of one group or the other depending on what group the data points nearest to it are in.

### Decision Tree Classifier

Decision tree learning is one of the predictive modelling approaches used in statistics, data mining and machine learning. It uses a decision tree to go from observations about an item to conclusions about the item's target value.

### Logistic Regression

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression).



5

# Results



◆ The accuracy for the different predictive models are:

```
F1 score (KNN): 0.5333204189555325  
Accuracy socre(KNN): 0.6666565772746534
```

```
F1 score (Tree): 0.5333204189555325  
Accuracy socre(Tree): 0.6666565772746534
```

```
F1 score (log): 0.5333204189555325  
Accuracy socre(log): 0.6666565772746534
```

The expectation, going on this investigation, was that bad weather and poor road conditions was dangerous, and that this would be evidenced by a pattern of higher crash rates have been supported by evidence.



6

# Conclusion

- ◆ The objectives of the investigation are met.
- ◆ Road traffic injuries can be prevented. Effective interventions include incorporating road safety features.
- ◆ The model provides empirical evidence against speeding and road conditions.





# THANKS!

ANY QUESTIONS?

You can find me at:

- ◆ @jethinpilli
- ◆ jethinpilli@gmail.com

