UK ROAD ACCIDENT SEVERITY

An Analysis, Visualization and Predictive Model

Introduction

- IBM Data Science Capstone Project

- Data Source:
  - Road Safety Data – Accidents 2018
  - UK Department for Transport
  - Data.gov.uk

Problem

- Thousands of road accidents each year

- How to improve road safety through a data science solution

- Machine Learning: A predictive model for crash likelihood based on location, weather conditions, road conditions
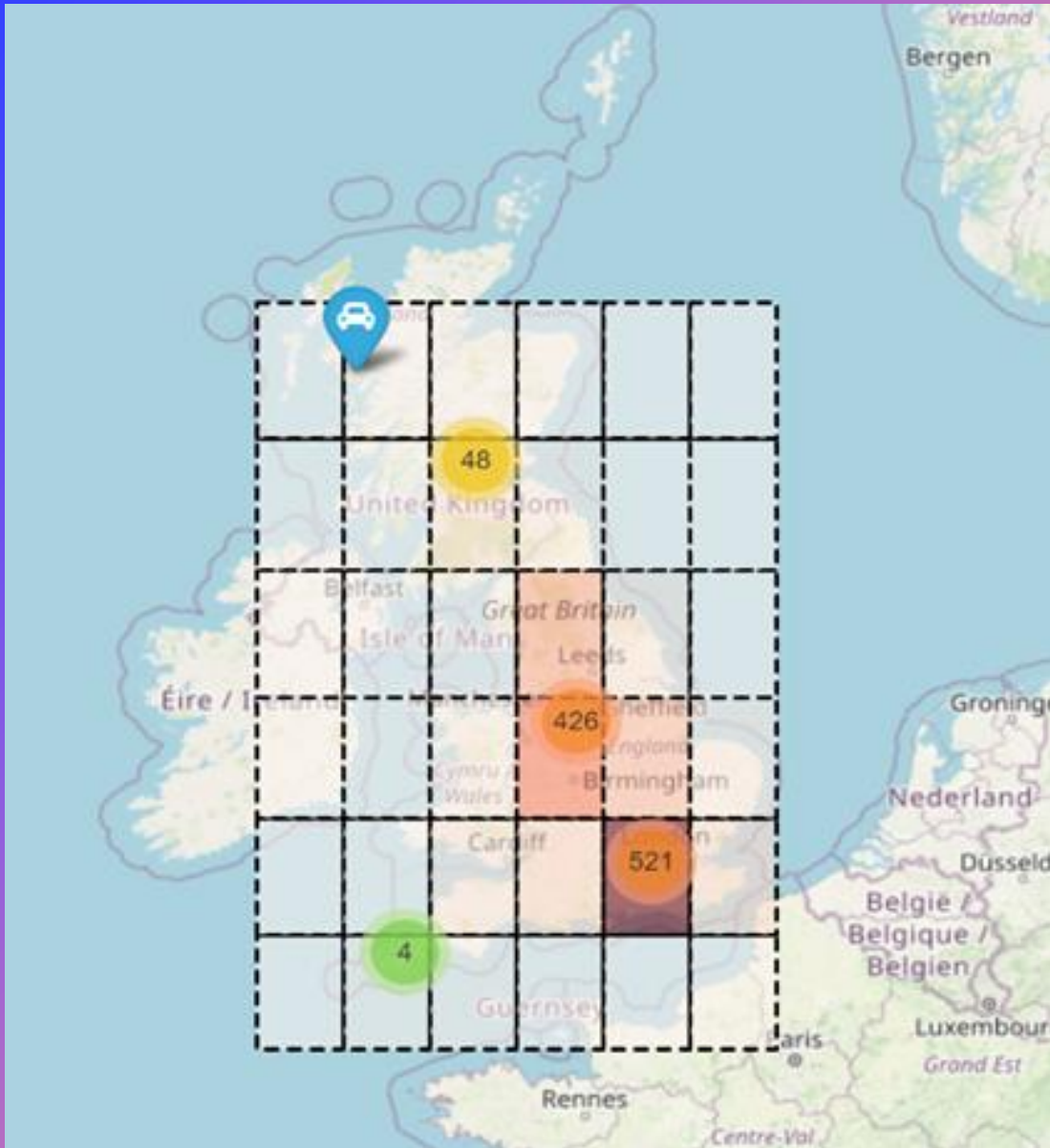
Data

- 122,635 samples
- 32 features

- Accident Severity
  - 1 – Fatal
  - 2 – Serious
  - 3 – Slight

INITIAL SOLUTION
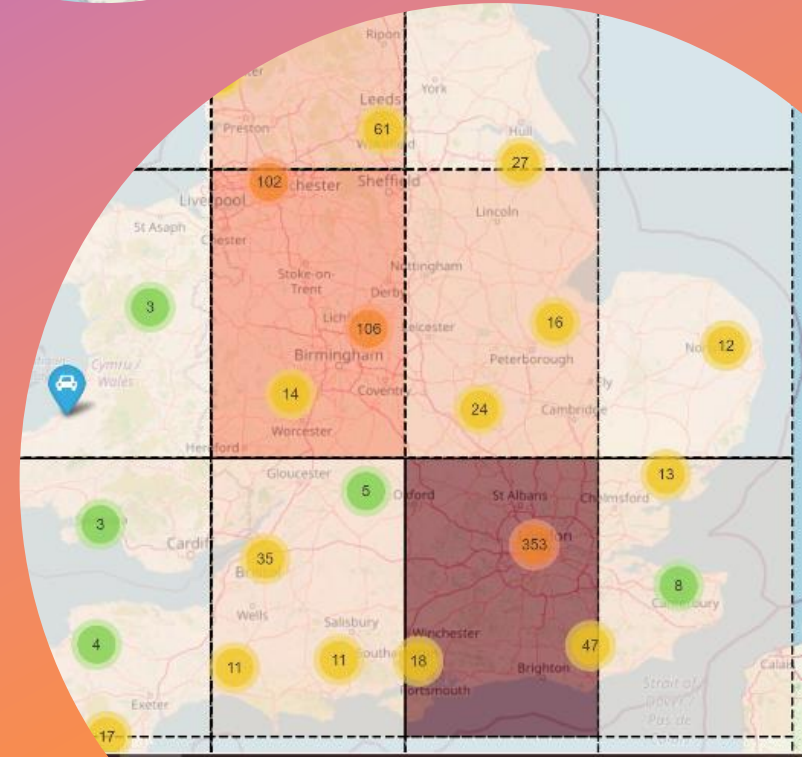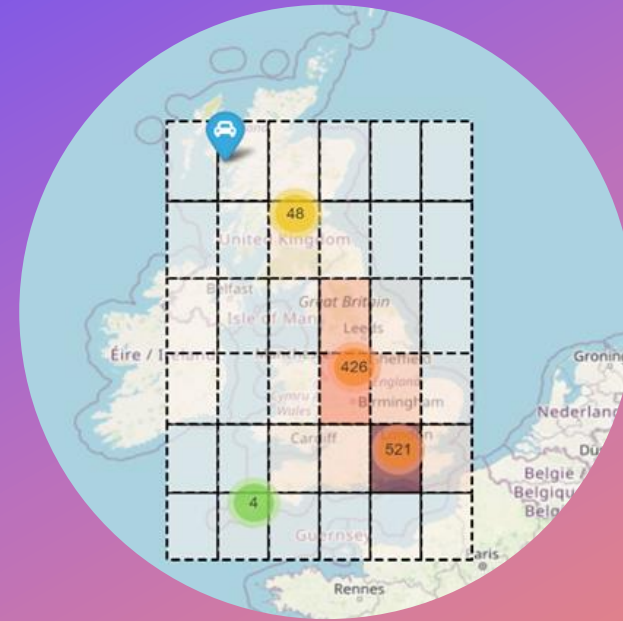
Map Visualization of Road Accidents

- Sample set of 1000 random accidents.

- Easy visualization of accident 'hot-spots'.

- Three clear 'high-risk' zones for chance of accident and severity in descending order:
  - London
  - Birmingham
  - Leeds

PREDICTIVE MODEL

Creating the Model

- Focus on 4 factors:
  - Day of week
  - Weather Conditions
  - Light Conditions
  - Road Conditions

- 3 Models:
  - Logistic Regression
  - Decision Tree
  - Random Forest

Decision Tree

```
Accuracy 79.71
              precision    recall   f1-score   support

         1    0.000000   0.000000   0.000000       334
         2    0.000000   0.000000   0.000000      4642
         3    0.797122   1.000000   0.887109     19551

 micro avg    0.797122   0.797122   0.797122     24527
 macro avg    0.265707   0.333333   0.295703     24527
weighted avg  0.635403   0.797122   0.707134     24527
```

Predicted	3	All
Actual		
1	334	334
2	4642	4642
3	19551	19551
All	24527	24527

- Scored best accuracy after adjustments.

- Successfully predicted 19551 serious accidents.

- Accuracy of 79.71%

Conclusion

- Great visualization system for localizing 'hot-spots'

- Predictive model needs more work and is not implementable in the real world
  - Expand with more data and features
  - Potentially include factors such as
    - Age of driver
    - Model of car
    - Speed

THANK YOU