ULB UNIVERSITÉ LIBRE DE BRUXELLES

Car accident analysis in the UK

INFO-H423 - Data Mining

Students:

Ait Oujkal Abdellatif 000430127
Outmane Mouad 000427221
Pappas Akilleas 000425456

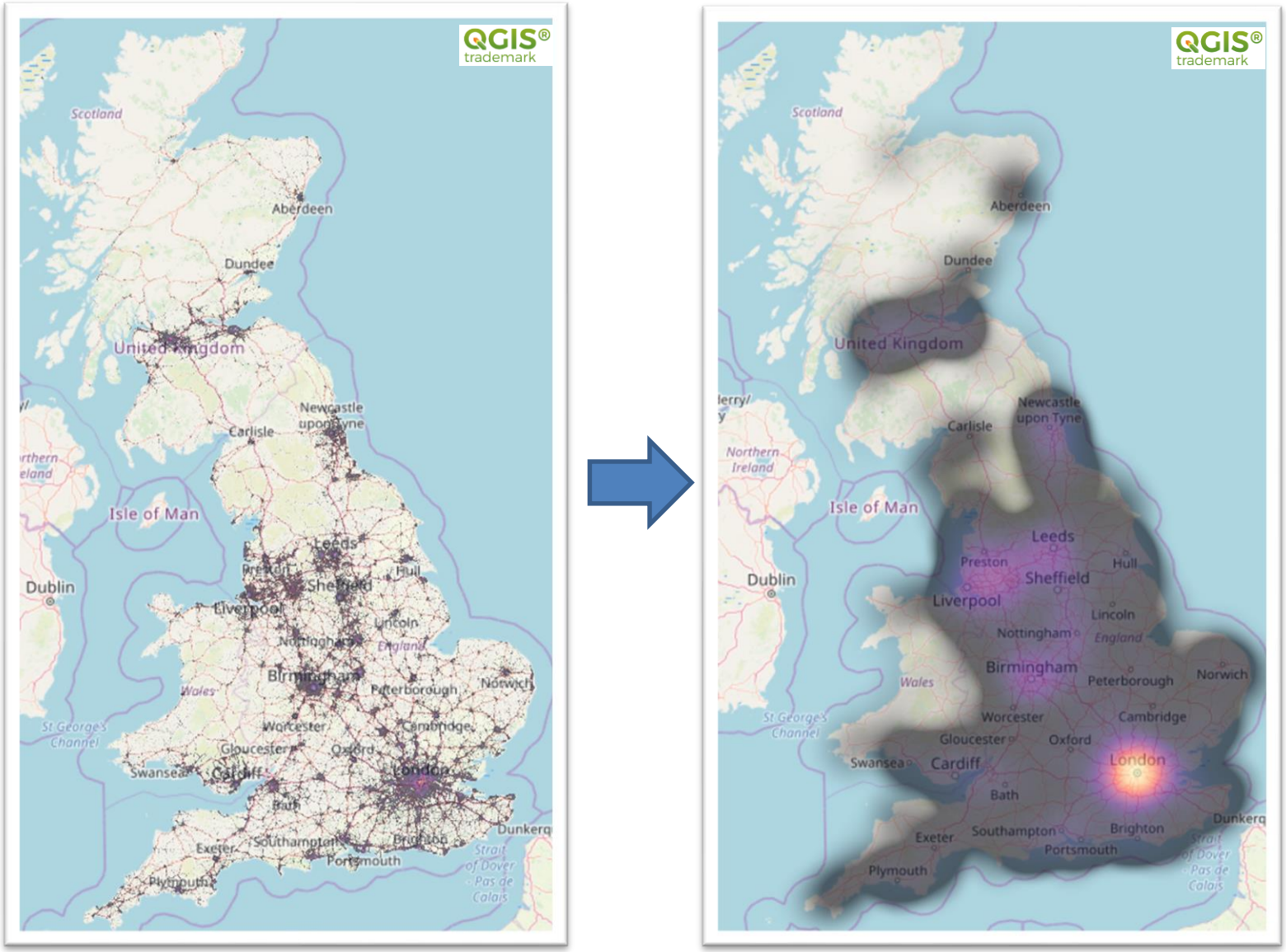
Supervisors:

Mahmoud Sakr

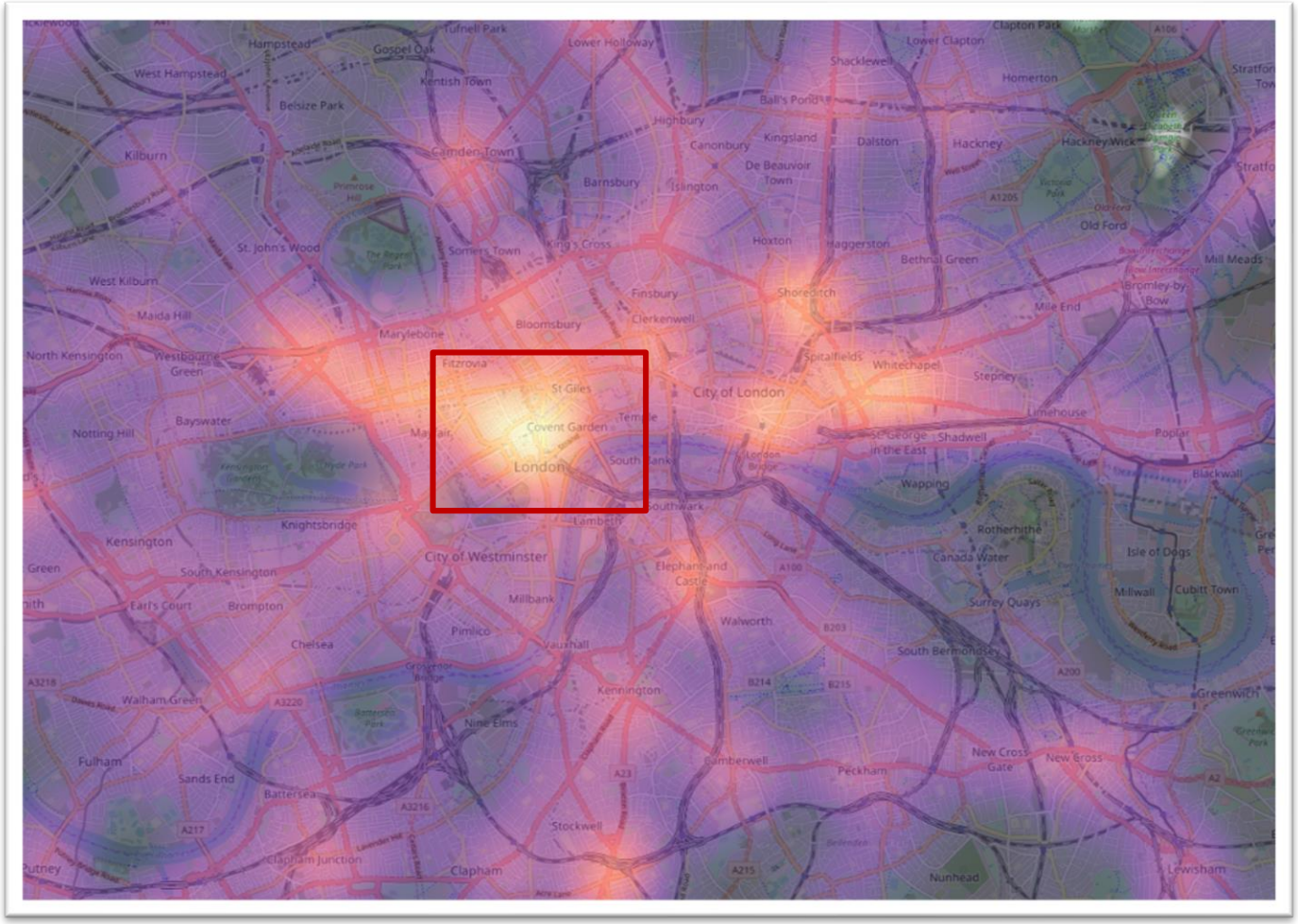
December 2019

Hotspots identification

Hotspots identification

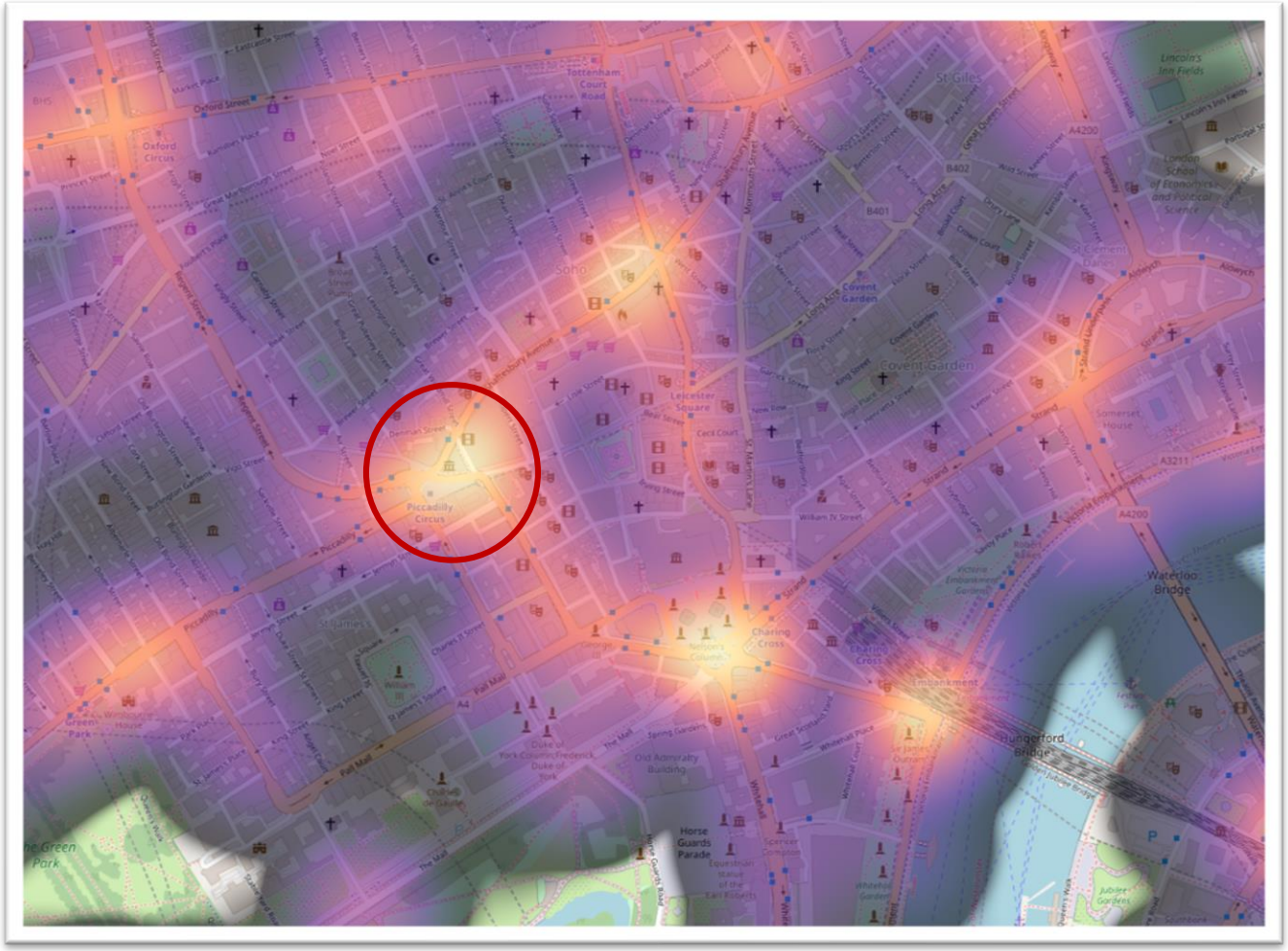


Hotspots identification



Zoom into London

Hotspots identification



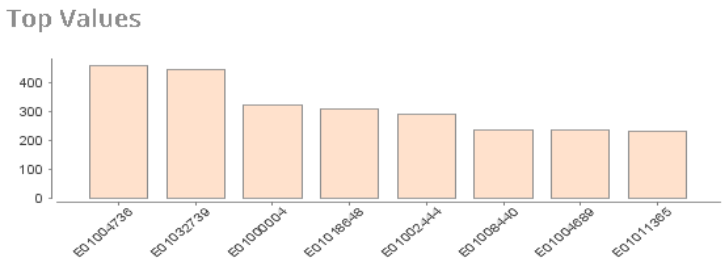
Zoom into London

Hotspots identification

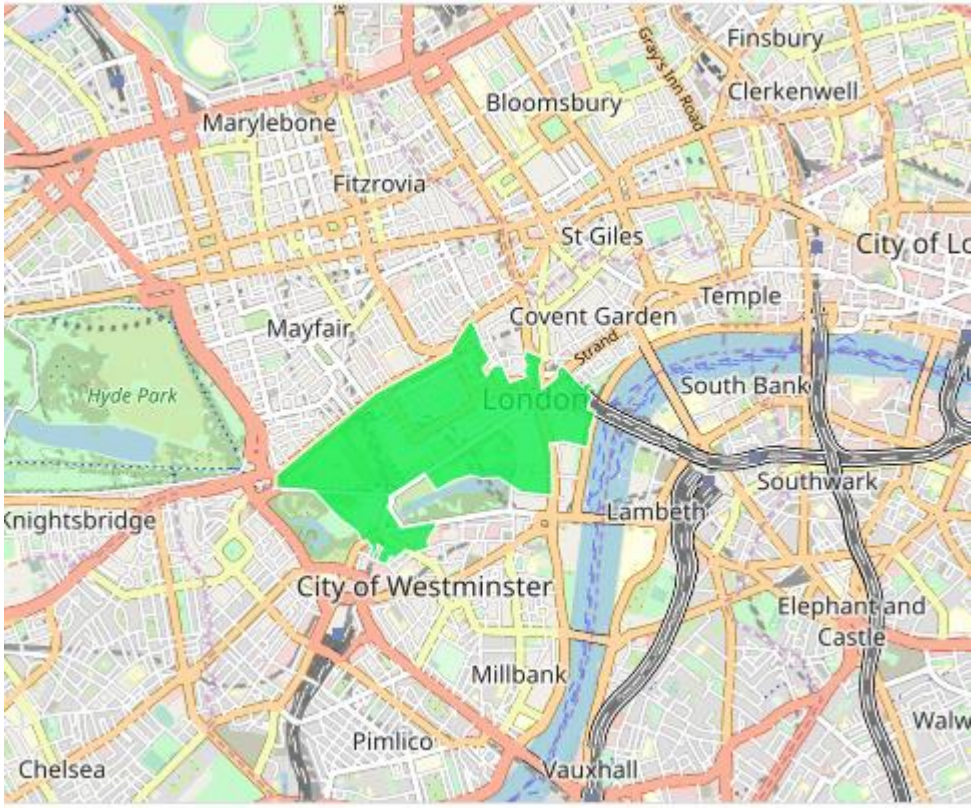
- known as a major traffic junction

Piccadilly Circus junction

Hotspots indentification



Histogram of LSOA attribute



E01004736 LSOA zone

Preprocessing

- Irrelevant features

Local_Authority_(Highway)	Pedestrian_Crossing-Human_Control	Did_Police_Officer_Attend_Scene_of_Accident
Date	2nd_Road_Number	Junction_Detail
Accident_Index	2nd_Road_Class	1st_Road_Number
1st_Road_Class	Special_Conditions_at_Site	Carriageway_Hazards
Junction_Control	LSOA_of_Accident_Location	Year

- Missing values:
  - Junction_Detail : 100%
  - Junction_Control : 40%

- None values:
  - Special_Conditions_at_Site : 97.78%
  - Carriageway_Hazards : 98.31%

Preprocessing

- Data Cast

Attribute name	Values	New values
Accident_Severity	**1**,**2** or **3**	**fatal** (1) or **non-fatal** (2 and 3)
Day_of_Week	from **1** to **7**	**true** if it is a day of week else **false**
Time	**all times**	**Night**, **Morning**, **Afternoon** and **Evening**

Patterns detection

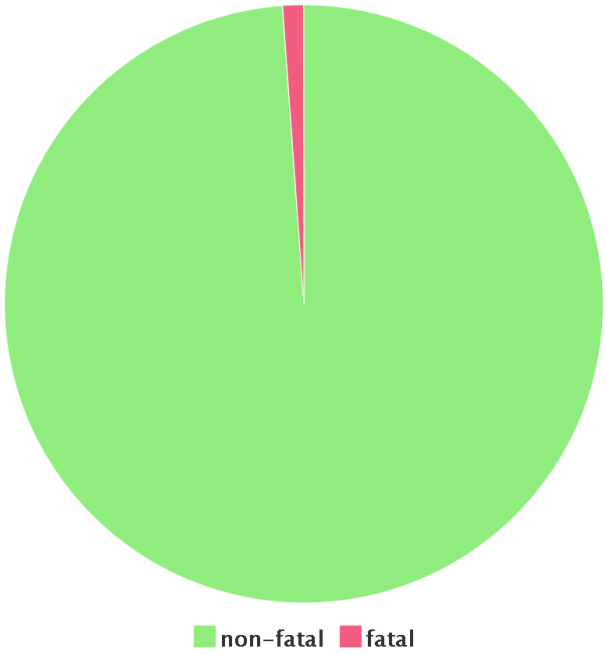
- FP-growth algorithm
- Frequent itemsets ⇨ patterns
- "Standard Pattern"

Patterns detection

Premise *Category*	Prem... *Number*	Conclusion *Category*	Lift *Number*	Confide... *Number*	Total Support *Number*
[Weather_Conditions = Fine without high winds, Light_Conditions = Darkeness: No street lighting]	2	[Accident_Severity]	4.002	0.046	0.002
[Weather_Conditions = Fine without high winds, Road_Type = Single carriageway, Light_Conditions = Darkeness: No street lighting]	3	[Accident_Severity]	3.855	0.044	0.001
[Speed_limit = 60, Light_Conditions = Darkeness: No street lighting]	2	[Accident_Severity]	3.769	0.043	0.001
[Road_Type = Single carriageway, Speed_limit = 60, Light_Conditions = Darkeness: No street lighting]	3	[Accident_Severity]	3.757	0.043	0.001
[Light_Conditions = Darkeness: No street lighting]	1	[Accident_Severity]	3.484	0.040	0.002
[Weather_Conditions = Fine without high winds, Road_Type = Single carriageway, Day_Of_Week, Speed_limit = 60]	4	[Accident_Severity]	3.393	0.039	0.001
[Road_Type = Single carriageway, Time = afternoon, Speed_limit = 60]	3	[Accident_Severity]	3.391	0.039	0.001
[Road_Type = Single carriageway, Light_Conditions = Darkeness: No street lighting]	2	[Accident_Severity]	3.295	0.038	0.001
[Time = afternoon, Speed_limit = 60]	2	[Accident_Severity]	3.217	0.037	0.001
[Weather_Conditions = Fine without high winds, Day_Of_Week, Speed_limit = 60]	3	[Accident_Severity]	3.151	0.036	0.001
[Weather_Conditions = Fine without high winds, Road_Type = Single carriageway, Road_Surface_Conditions = Dry, Speed_limit = 60]	4	[Accident_Severity]	3.125	0.036	0.002
[Road_Type = Single carriageway, Road_Surface_Conditions = Dry, Speed_limit = 60]	3	[Accident_Severity]	3.104	0.035	0.002
[Weather_Conditions = Fine without high winds, Road_Type = Single carriageway, Speed_limit = 60]	3	[Accident_Severity]	3.029	0.035	0.003
[Road_Type = Single carriageway, Day_Of_Week, Speed_limit = 60]	3	[Accident_Severity]	2.969	0.034	0.001
[Weather_Conditions = Fine without high winds, Road_Surface_Conditions = Dry, Speed_limit = 60]	3	[Accident_Severity]	2.877	0.033	0.003
[Weather_Conditions = Fine without high winds, Road_Type = Single carriageway, Light_Conditions = Daylight: Street light present, Road_Surface_Conditions = Dry, Speed_limit = 60]	5	[Accident_Severity]	2.863	0.033	0.002
[Road_Surface_Conditions = Dry, Speed_limit = 60]	2	[Accident_Severity]	2.857	0.033	0.003
[Road_Type = Single carriageway, Light_Conditions = Daylight: Street light present, Road_Surface_Conditions = Dry, Speed_limit = 60]	4	[Accident_Severity]	2.840	0.032	0.002
[Weather_Conditions = Fine without high winds, Speed_limit = 60]	2	[Accident_Severity]	2.816	0.032	0.003

Imbalanced Data

- 98% of non-fatal accidents and 2% of fatal accidents
- Not enough data to represent the minority class (non-fatal)



- Solution : Undersampling

Predictive model

- Random Forest
- Performs good compared to the state-of-art- models
- Parameters :
  - Number of trees: 60
  - Criterion: accuracy
  - Maximal depth: 6
  - Voting strategy: majority vote
- K-Fold cross validation (K=10)

Results

accuracy: 66.73% +/- 1.79% (micro average: 66.73%)

	true non-fatal	true fatal	class precision
pred. non-fatal	3673	1900	65.91%
pred. fatal	1625	3398	67.65%
class recall	69.33%	64.14%	

- Accuracy ⇨ 66,73%
- Class recall for non-fatal ⇨ 69,33%
- Class recall for fatal ⇨ 64,14%
- Class precision for non fatal ⇨ 65,91%
- Class precision for fatal ⇨ 67,65%

Questions?