Northeastern University
DS 5230 – Unsupervised Machine Learning
December 8, 2017, Adwait Sahasrabhojanee, Sreejith Sreekumar, Xue Yu, Xuexian Li

Packet: Pattern Recognition in Accidents in London

1 Introduction

The UK government has put their efforts into building an aggregated collection of accident records in the England, Scotland and Wales regions over a span of ten years from 2005 to 2014. The data includes 1.6 million accident instances. Having access to this, one of the most comprehensive data sets on traffic, we work towards uncovering patterns and answer questions related to the potential causes and trends of accidents. We start our analysis on the whole of the UK area and eventually shift towards isolating and understanding urban cites where most accidents are concentrated, taking London as the candidate for our study.

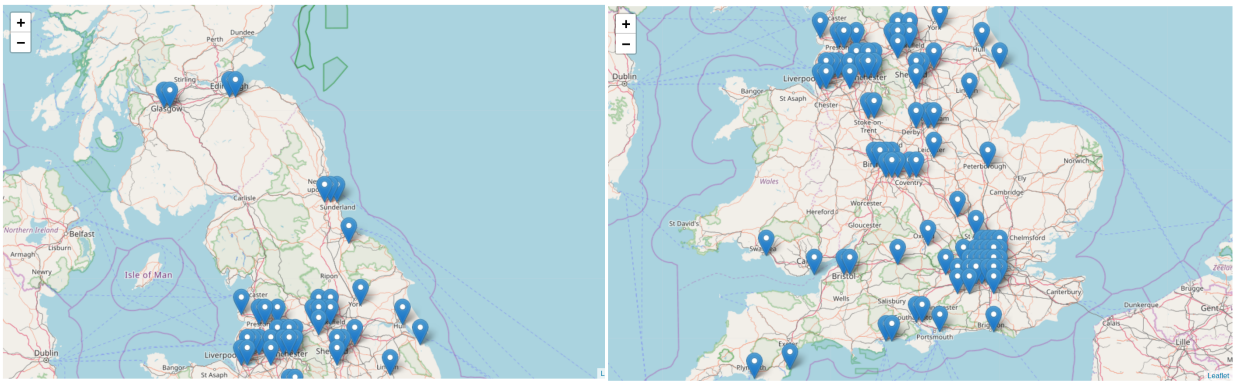
1.1 The Dataset

We consider the accident data from years 2009 to 2014 for our experiments. Every record of an accident includes the location (in geographical co-ordinates), the type of area, and the date and time of the accident.In addition, it includes the speed of the vehicle, the severity, the number of officers attended, the number of casualties, the weather conditions, the road and lighting conditions during the accident.

Furthermore, to understand the flow of traffic, we are given the Annual Average Daily Flow data, that tracks the amount of traffic which has been present on the roads at different points of time.

2 Exploring our data

We begin our analysis, identifying the areas where accident concentrations are the highest. The towns/cities of the accident were identified by decoding the geo-tags and the top 200 places where the concentration is the highest were isolated. The records from these 200 places were tagged on a map to see the spread of hotspots. As we can see from the map, these hotspots are centered around the main cities in the UK.



These regions, which we have identified to be the hotspots are also the highest populated cities[1], as well as the most traffic dense regions in the UK(Appendix, Fig 1).

3 Appendix

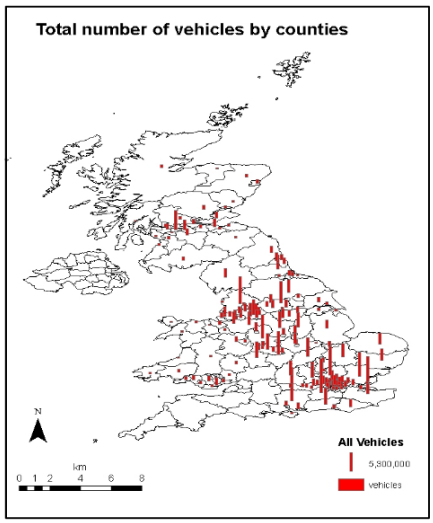


Figure 2: Frequency of vehicles per county

4 References