**Road Accidents in the United Kingdom in 2004-2006**

Introduction

In the United Kingdom, people like to cycle a lot. This is true despite the often-gloomy weather. Growing environmental consciousness throughout the world and healthy lifestyle motivate people to cycle even more. It is therefore very important for the Government to provide safety on the roads for cyclists, pedestrians, and vehicle drivers. While permanent road signs for safety are installed throughout the country, I would like to see if it makes sense to install temporary road signs for cyclists in dangerous months and locations. In order to understand whether there is a need for such measures, I would like to identify such accident-prone locations and dangerous months. This should be possible by looking at the data that the Government has made available for the public to peruse.

Data

Department of Transportation in the UK made a report called “Reported fatal personal injury road accident and casualty data: GB 2006-2008” publicly available on data.gov.uk. The csv file provides details of each individual reported fatal personal injury road accidents in Great Britain for 2006 to 2008. The file also includes latitude and longitude of the locations where the accidents took place. I am particularly interested in the following columns:

1. Month\_of\_Accident
2. Longitude
3. Latitude
4. Pedal\_Cycles
5. Total\_Number\_of\_Accidents

The 1st Column will help me with seasonality of Pedal Cycle accidents (4), i.e. it will show if cyclists are more likely to get involved in accidents in particular months. The 2nd and 3rd Columns will help me identify hotspots for such accidents. I will make every attempt to visualize this data on the UK map and see if there is an opportunity for clustering.

Methodology

It was great to use correlation techniques and find relationship between different data. Even though I was looking at cyclists’ data, it is clear that such accidents never happen alone. There is always a counterparty, such as a car, pedestrian, or another bike. Since data is presented in such a way that the number of passengers is indicated only if they were hurt, we cannot know what the reason for the bike accident was. In the future, this database needs to be updated so that it becomes clear what caused the accident, because it is a useful information for policymakers.

Another methodology used was plotting map locations based on latitude and longitude data. Such clusters were identified, and it turned out that London is the capital of bike accidents. In addition, months of the year were plotted against number of accidents and we found the safest and most dangerous months for cyclists. This information would be more useful if we could also add weather conditions and daylight time in the UK in such months and see if there is correlation.

Results

The results have clearly demonstrated that the safest months are February and March for cyclists. It also became clear that the most dangerous month is October. It also became clear that the share of cyclist accidents among all accidents is 5%.

![Table

Description automatically generated]()

![Chart, bar chart

Description automatically generated]()

Discussion

While the results are interesting, they leave much room for interpretation. For example, why is October the most dangerous month for cyclists? It could be related with rain precipitation in this month, or it could be because of the shorter daylight, or some other time changes. To see if there is correlation, we would need to import weather data and run some correlations. Even if it turned out that there is a connection, however, it is important to understand that correlation does not mean causation. Still, the more factors we identify, the better. One other factor worth considering is overall traffic, i.e. any kind of increase in the number of cyclists or drivers can potentially lead to higher number of incidents. It would also be useful to see age breakdowns and occupations of those involved in accidents. For example, it could be that October is when classes start for a lot of students and many newcomers happen to be in new locations that they are not familiar with. This could lead to more accidents too.

Conclusion

Overall, we could see some hotspots and “hot” months for pedal cyclists in this project. What is also clear is that we need more data to make reasonable conclusions. Right now, we could only make some inferences. Part of the reason why this UK data is not so rich in information is because it comes from 2004-2006, when there was not much need for sophisticated data analysis in Python as now. Presumably. In addition, the difficulty with cyclist data is that a lot of slight accidents may not be reported and likely to be settled without police.