Bike accidents in the UK analysis

Data

<https://www.nature.com/articles/s41597-022-01333-2#Sec6>

CYCLANDS: CYCling geo-Located AccideNts, their Details and Severities

This dataset contains a collection of 30 datasets on geographically located cycling crashes across the world. This collection covers country, region, or city-wide areas in multiple easy-to-use formats, such as Comma Separated Values (CSV) files and GeoJSON files.

I am using the UK Gov data, 2005-2018

Step 1 – Clean the data

* Copy the data to a working sheet to maintain the original.
* Set date to UK, and order by date
* Check for duplicates, none found
* The accident index column has some weird values, some are just numbers and excel is formatting them as such. Change this so the column is formatted as text.
* Check speed limit column as it contains some speed limits of 15, 10 and 0. There are only.
* Used the 3d maps feature in excel to plot the locations of these roads on a map to find out where they are:
  + The 0 value is on a road so lets change that to 30
  + The 10 and 15 values seem to be in parks so that is probably the reason they are so low. So let’s keep them.

Start to analyse the data

* I want to see how speed limit affects the outcome of the accidents. E.g. are accidents less severe at lower speed limits.
* Things to consider:
  + The lower speed limits might reduce accidents. There might be near misses where an accident would have occurred at higher speeds. This won’t be captured in the data.
  + More cyclist may travel on roads with lower speed limits. So the relative accident rate won’t be captured by the data.
  + We can still analyse the severity of accidents that did occur at different speeds.
* Create pivot table to analyse data.