

Visualizations of traffic data

Amartya Das Sharma

May 2019

1 Introduction

The data used is from the public Kaggle dataset [1.6 million UK traffic accidents](#) retrieved on 27th April. The following section explains the data layout, with the remaining sections containing graphs comparing values of interest in the data. Explanations and observations are also provided.

2 The data

According to the Kaggle homepage, the dataset contains traffic data collected by the UK government over the period of 2005 - 2014 (excluding 2008). The datasets of interest are compiled summaries of police reports of traffic accidents over this time period. The information that the dataset contains for each report (i.e. the columns of the dataset) are the following:

Accident Index A unique ID given to the accident.

Easting and Northing Local British coordinates of the accident location.

Longitude and Latitude Global latitude and longitude of the accident location.

Police Force Denotes which police force attended the scene of the accident by unique ID. For example, 1 denotes metropolitan police, while 5 denotes the force at Merseyside. There are 51 unique codes for this field.

Accident Severity How severe the accident was. 1 stands for fatal accident, 2 for severe, 3 for slight.

Number of Vehicles Number of vehicles involved in the accident.

Number of Casualties Number of casualties in the accident.

Date Date the accident occurred in dd/mm/yyyy format.

Day of Week The day of the week. 1 stands for Sunday, 2 for Monday, until 7 for Saturday.

Time Time of the accident in 24h format.

Local Authority (District) Unique ID for the district. For example, 204 denotes Leeds, 350 denotes Boston. There are 416 unique codes for this field.

Local Authority (Highway) Unique ID for the highway authority. For example, E08000025 denotes Birmingham. There are 208 unique values in this field.

1st Road Class, 1st Road Number, 2nd Road Class, 2nd Road Number Values used only for junctions, denoting class and number of connecting roads.

Road Type 7 possible road types shown by ID. For example, 1 denotes roundabouts, 3 denotes dual carriageway, 2 denotes a one way street. -1 denotes data out of range or missing.

Speed limit Speed limit of the accident zone.

Junction Detail Type of junction. 9 total unique IDs denote 9 types of junctions. For example, 0 denotes that there is no junction within 20 meters, 1 denotes a roundabout and 2 denotes mini roundabouts.

Junction Control How the junction is controlled. 5 total unique IDs denote 5 different control mechanisms. For example, 2 denotes auto traffic signal, 3 denotes stop sign.

Pedestrian Crossing-Human Control Whether there is no pedestrian crossing within 50 meters (0), whether the crossing is controlled by school crossing patrol(1) or by an authorized person (2).

Pedestrian Crossing-Physical Facilities 6 unique IDs denoting the type of crossing. 0 denotes absence of physical crossing facilities within 50 meters. 1 denotes zebra crossing, 7 denotes footbridge or subway.

Light Conditions The condition of lighting. Values include daylight, darkness with street lights on, darkness with street lights available but unlit, darkness with no street lights etc.

Weather Conditions Values include fair with/without high winds, rainy with/without high winds, snowing with/without high winds, fog or mist, or other.

Road Surface Conditions Values include dry, wet or damp, snow, frost or ice etc.

Special Conditions at Site Whether there were any abnormal conditions at site, such as road works underway, road surface defective etc.

Carriageway Hazards Hazards relating to the road, such as whether there are other objects, whether there was a previous accident etc.

Urban or Rural Area In this field, 1 denotes an urban area, and 2 denotes rural.

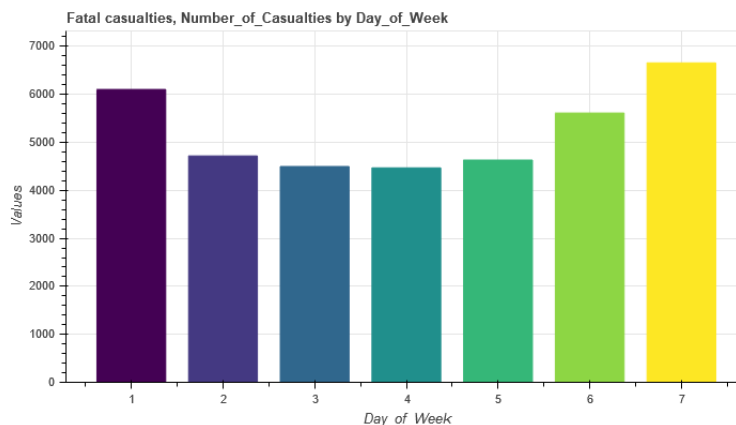
Did Police Officer Attend Scene of Accident Field contains yes, no and no (self reported accident via form).

LSOA of Accident Location LSOA areas are geographic areas uniquely used by England and Wales for reporting smaller area statistics. This field gives the area of the accident.

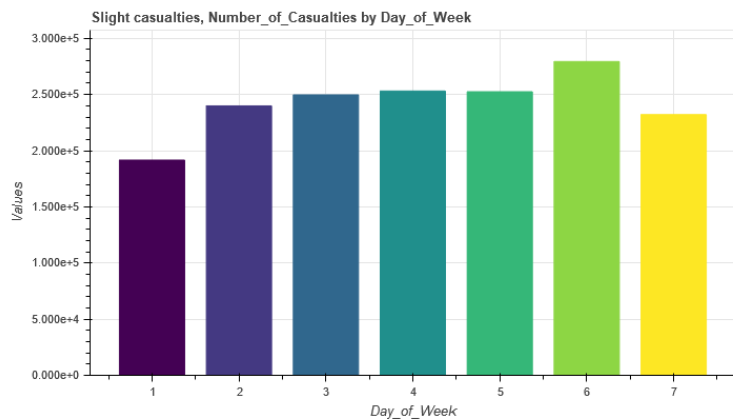
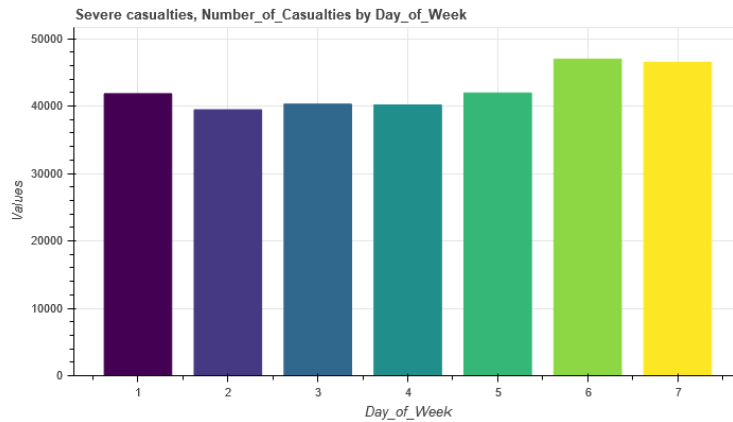
Year The year of the accident.

3 Day of week

Here we have graphs comparing number of casualties per day. As shown



there are more fatal casualties on the weekends, more slight casualties on

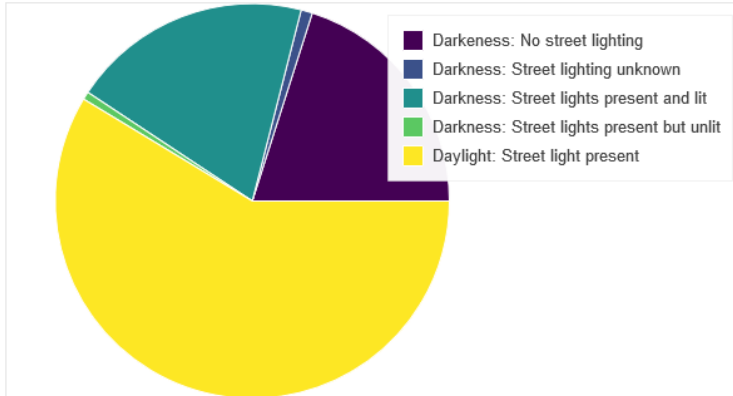


the weekdays, more severe casualties on Friday and Saturday. A possible explanation could be that people are more likely to be driving rashly and under influence on weekends.

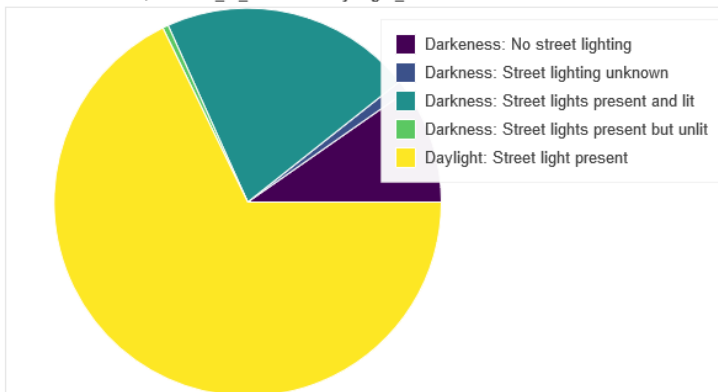
4 Lighting conditions

Here he have graphs comparing number of casualties based on the lighting conditions of the area. Most fatal casualties occurred in daylight on roads with streetlights (59%), followed by nighttime and no street lights (20%) and then nighttime and lit streetlights(19%). Most severe casualties again in daylight (67%), followed by streetlights being present and lit (21%) and then darkness without streetlights (10%). 73% of all slight accidents occur in daylight. Next biggest chunk is 19% for streetlights present and lit.

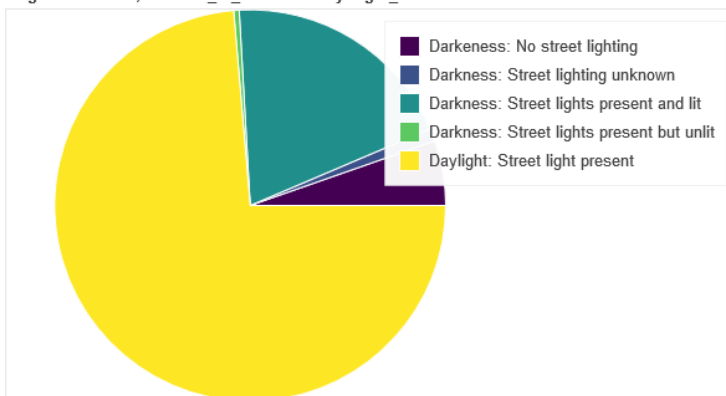
Fatal casualties, Number_of_Casualties by Light_Conditions



Severe casualties, Number_of_Casualties by Light_Conditions

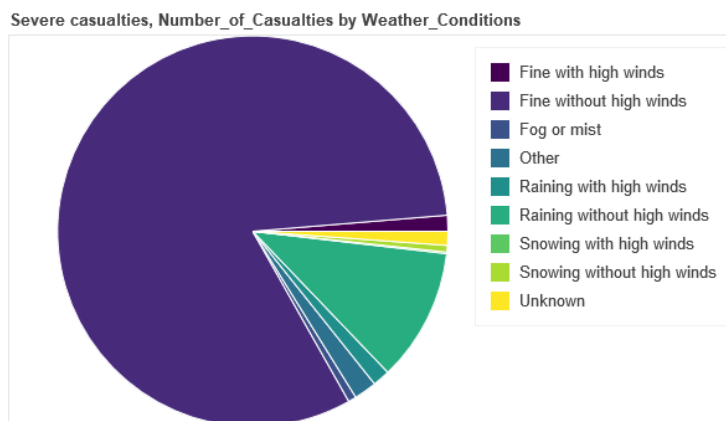
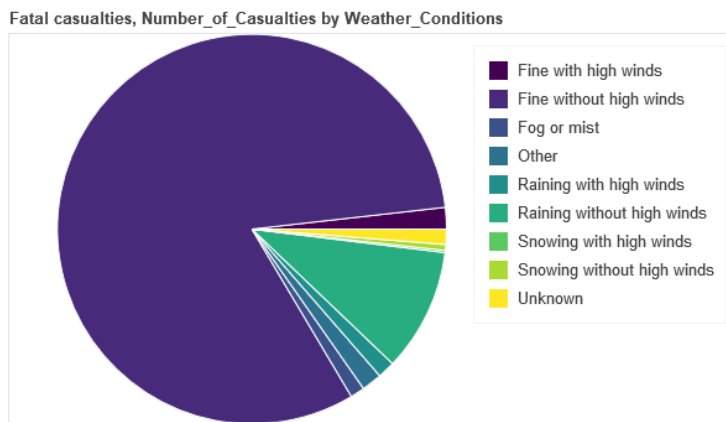


Slight casualties, Number_of_Casualties by Light_Conditions

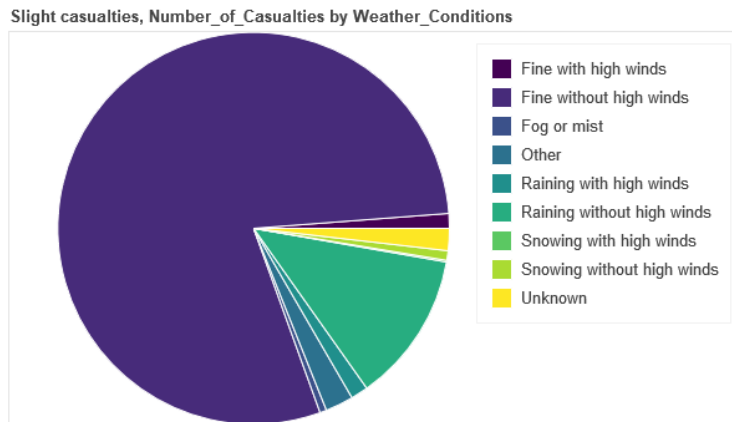


Most accidents consistently happen in the best lighting conditions, presumably because drivers are more careful when conditions are tougher to drive in. A lot of fatal accidents occur in darkness without streetlights, while severe and slight accidents occur in darkness with lights. It is presumable that accidents are more likely to be fatal when there are no streetlights.

5 Weather conditions

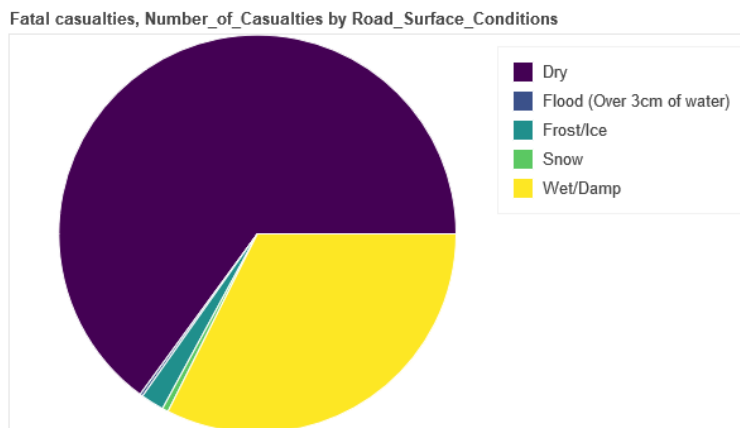


81.7% of all fatal casualties occur in fine weather without high winds, followed by 10% in raining without high winds. Data is almost identical for severe and slight casualties.



Again, this could be because drivers are more careful in adverse conditions, causing less accidents.

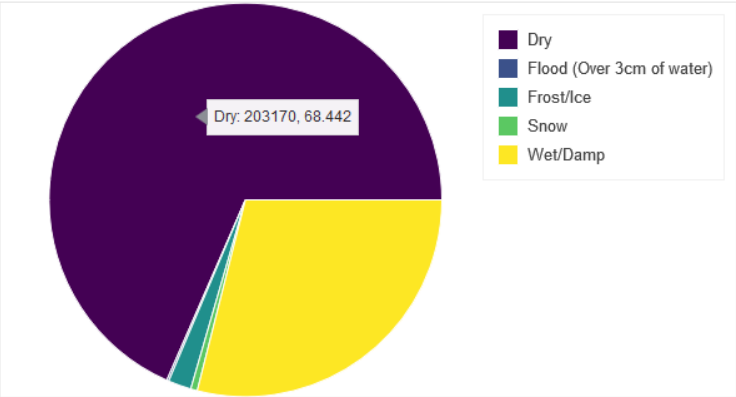
6 Road condition



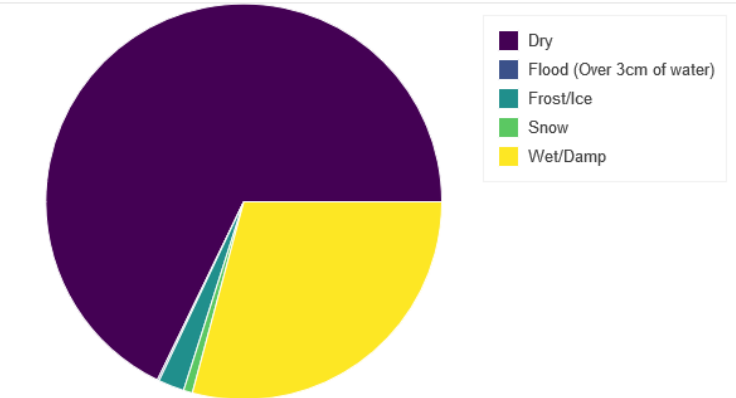
65% of all fatal casualties occur on dry roads, and 32.5% on wet/damp roads. Again, data is almost identical for severe (68.5%, 29%) and slight (67.8%, 29%) casualties.

Since road condition is dependent on weather, we can see that the data matches the weather condition data, leading us to the same belief that people are more careful in adverse conditions.

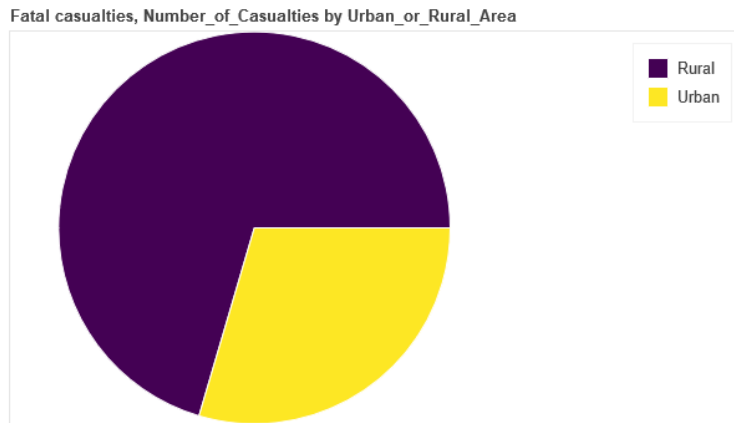
Severe casualties, Number_of_Casualties by Road_Surface_Conditions



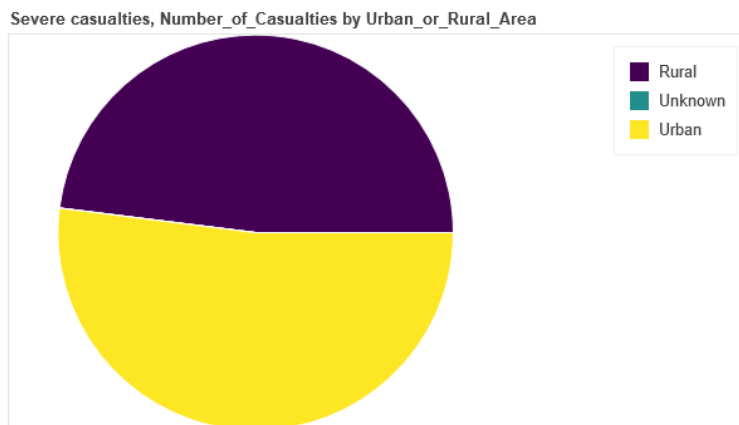
Slight casualties, Number_of_Casualties by Road_Surface_Conditions



7 Urban or Rural area

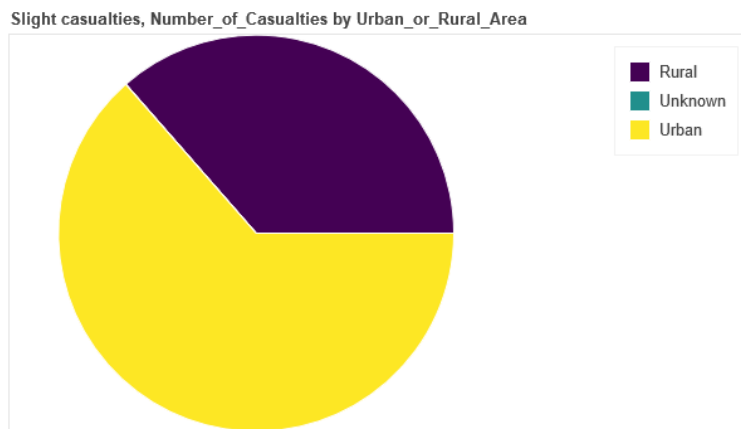


Of all fatal casualties, 29.5% were in urban areas, while 70.5% were in rural areas. The statistics flip for severe and slight accidents. Severe

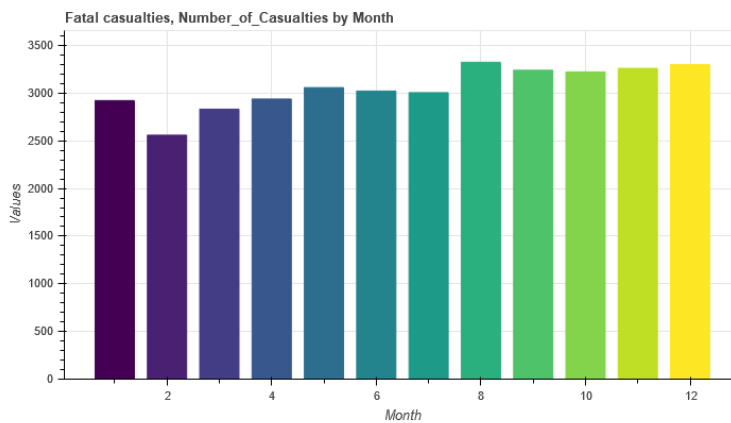


casualties were almost identical (52% urban, 48% rural) and for slight casualties, urban (63.5%) had more than rural (36.5%).

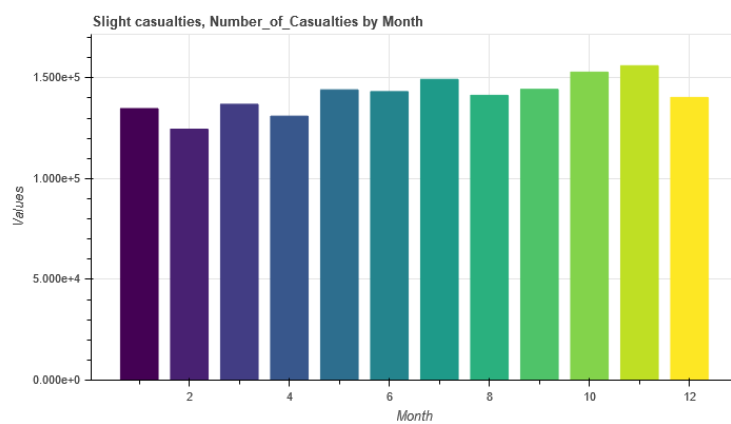
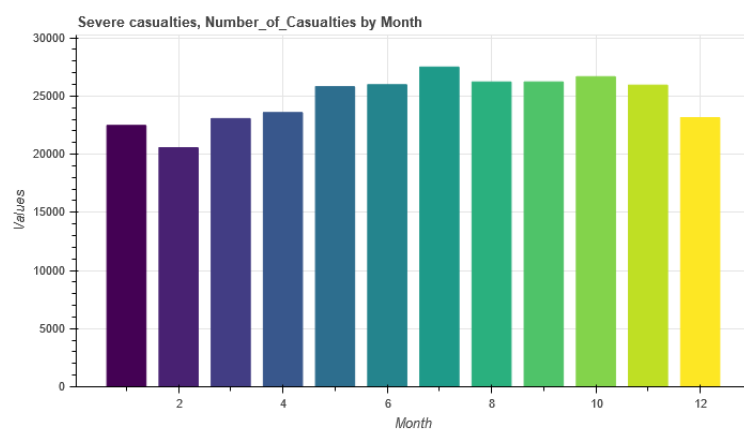
Since the data we are dealing with is made up of reported incidents, it is entirely possible that only fatal and mostly severe incidents from rural areas even get reported, which can explain the flip in the statistics. However, it is also possible that more casualties in rural areas are fatal because of a lack of resources for treatment, or lack of crossings and driver awareness.



8 Monthly analysis

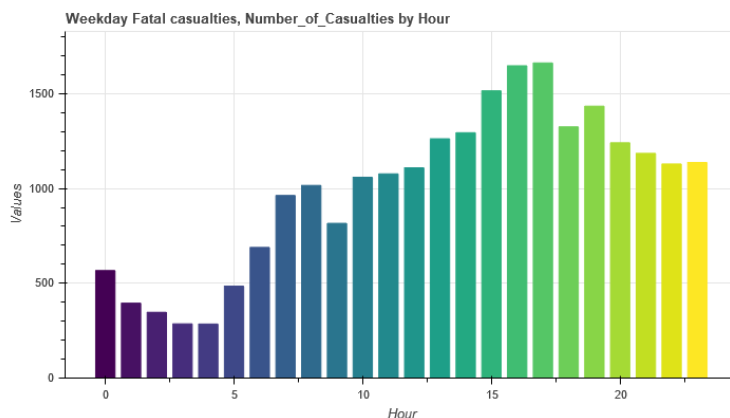


Provides no significant insight, numbers consistent amongst months except for February.

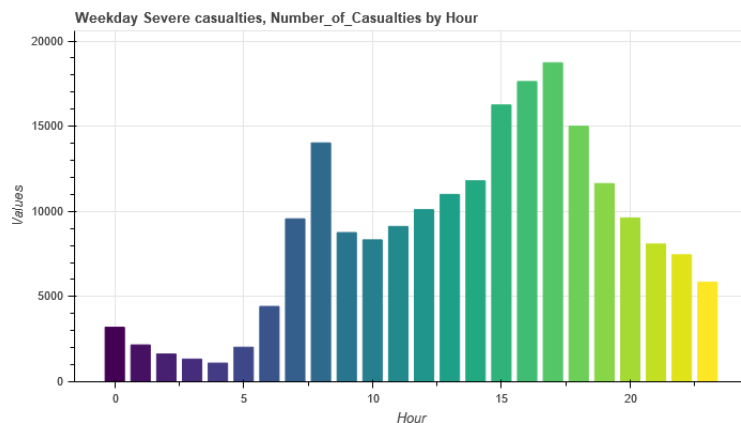


9 Hourly analysis

9.1 Weekdays

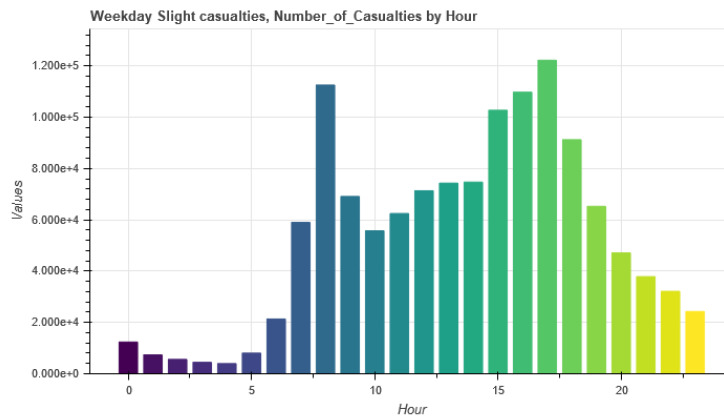


Fatal casualty numbers go up gradually from 04:00 (285), peaking at 17:00 to 17:59 (1662), then tapering off. Severe casualties follow a similar

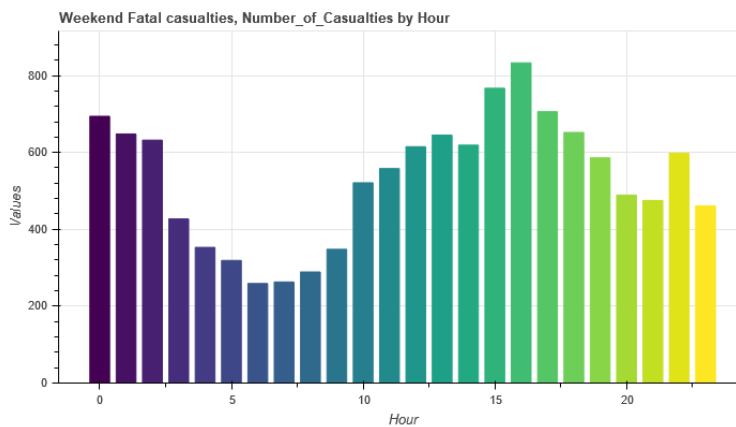


pattern but there is an abrupt spike at 08:00 to 08:59 (14021). Following this the number goes down to 8758 at 09:00, and increases gradually before rising abruptly from 15:00 to 17:59 (changing from 11798 to 16249 and above) before going back down. Slight casualties follow a similar pattern, with spikes at 08:00-08:59 and 15:00-17:59.

This obviously points to a rise in accidents around the start and end of office hours, when the traffic can be presumed to be at maximum.



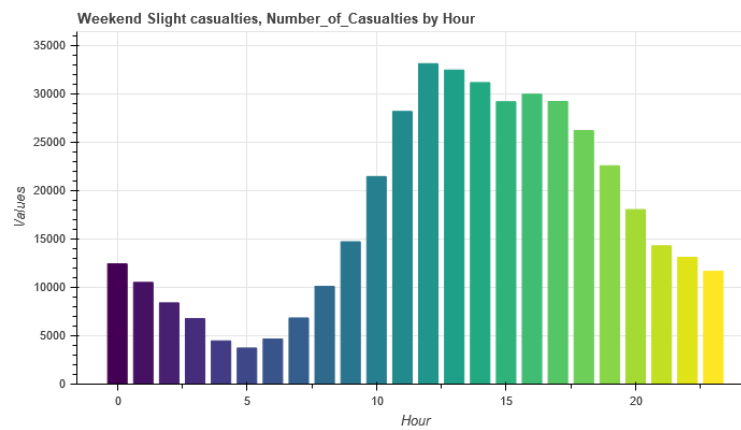
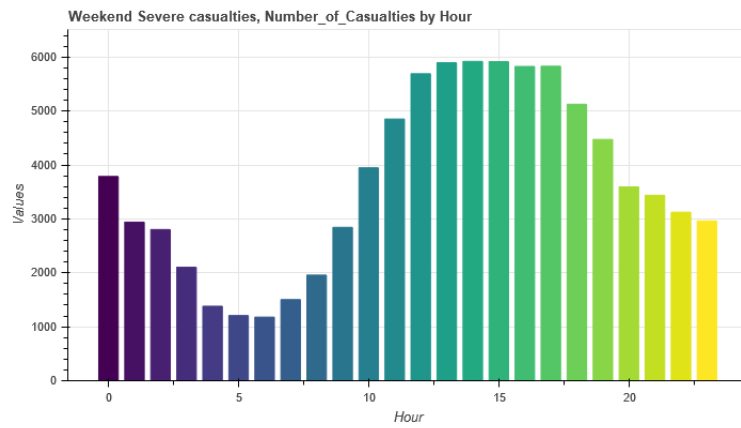
9.2 Weekends



Fatal casualties follow a smooth pattern, with spikes at 17:00-17:59 and 00:00-02:59. minimum values occur at 06:00 and 22:00.

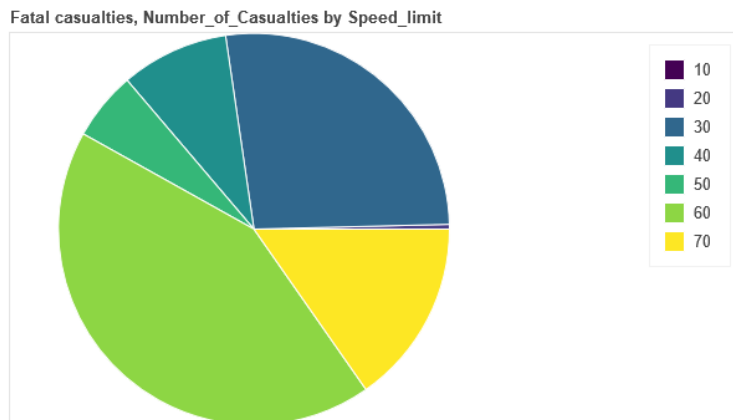
Severe and slight casualty numbers follow almost identical trends.

Comparing with weekdays, we can see that the peak accident prone hours have shifted, implying that the peak traffic hours shift similarly, and people go out late and return late on weekends.

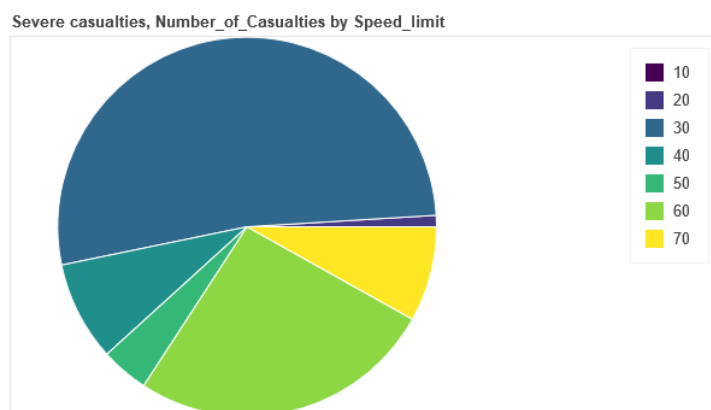


10 Speed limits

Speed limits in the data vary from 10 to 70 mph. Fatal casualties occur

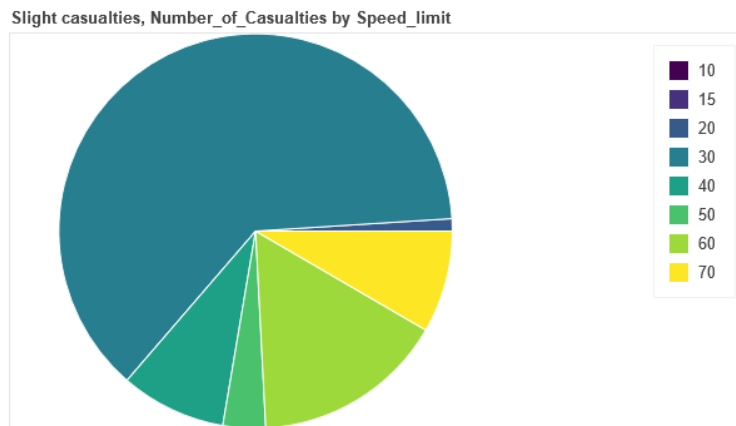


mostly at the higher speed limit areas (60 mph: 43%, 70 mph: 15.2%). 27% of fatal casualties occur at 30 mph speed limit areas. Most severe



casualties occur in areas with speed limit 30 mph (52.3%), followed by 60 mph(26%). Similarly, most slight accidents occur in areas with speed limit 30 mph (62.6%), followed by 60 mph(16%).

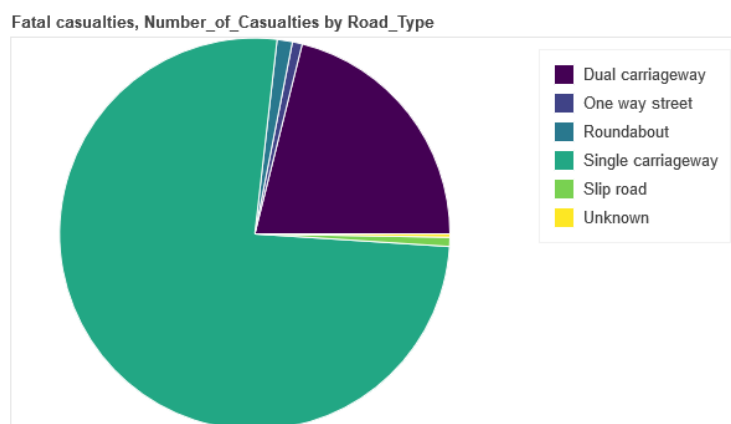
This indicates that as speed limits increase, more incidents turn fatal.



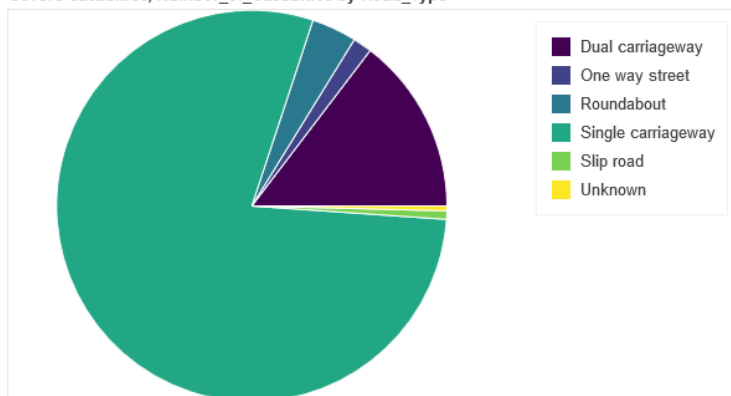
11 Road type

75.8% of fatal casualties occurred on single carriageways, followed by 21% on dual carriageways. Similar trends are seen for severe (79% single carriageway, 14.6% dual carriageway) and slight (73% single carriageway, 16.3% dual carriageway) casualties.

This could simply mean one-way roads are safer since there is no two way traffic, or it could mean that a proportionally large number of roads in the UK are single and dual carriageway roads.



Severe casualties, Number_of_Casualties by Road_Type



Slight casualties, Number_of_Casualties by Road_Type

