Sulayman Birt (201342787) **Thamer Aljohani (201577595)**

Zihan Jin (201669791) **Wasif Shaukat (201589942)**

1. Introduction

In order to find out the effectiveness of a given procedure, rule, or policy the surrounding real-world events must be monitored, which requires the collection of data and its analyses. Data on traffic collisions are crucial for developing and monitoring local, national, and worldwide road safety policies. The Department of Transport collects such data, known as STATS19 records. The aim of this report is to analyse the patterns in this dataset for three key aspects:

1. Demographics of casualties
2. Different local authorities in which Accidents included killed or seriously injured (KSI) casualties.
3. Pedestrians who were KSI casualties

In analysing the dataset. regarding these avenues of interest, a picture will be constructed of car accident incidents across this period, such that road safety policies and infrastructure can be adjusted and developed in order to reduce the number of casualties. The publicly available data being analysed here is for the year 2019 and is taken from the UK government website, including files of attributes for individual accident events and the individual casualties involved in such events, along with a guide file that allows numbered keys to be transformed into categorical data.

2. Data Quality:

Before commencing analysis, the data must be profiled, which includes data characterisation and data quality checks. In order to wrangle the dataset and scout for potential new discoveries, the number of rows of the accident and casualty CSV files were compared. There were more rows in the casualties file which is explained by their being multiple casualties per accident, and when removing duplicates from the casualties file the number of rows were the same. The two files were merged based on accident index number, which lead to no missing values or empty cells and the correct number of rows, verifying an exact match between the two datasets.

To verify the data characteristics for instance cardinality, the total number of casualties was summated over the accident file was compared with the number of records in the casualty file which gave an exact match. For all the attributes where data was not either present or out of range “-1” was included in the cell, making it easier to identify null values and proceed with the analysis. Furthermore, there were null values in location data columns which were expected due to their mention in the guide file. These null or missing values were not removed from the data tables due to the ability of the Tableau plotting software to filter out such values.

There were some minor data quality issues in the guide files which lead to issues when converting the numbered keys to their nominal counterparts. Firstly, the second_road_class variable was missing the key for “-1” missing values, with these values having to be removed separately. Also, the local_authority_ons_district variable had 2 identical entries in the guide which led to errors when matching the keys, so this duplicate row had to be removed. These profiling steps ensured that data was complete, accurate and consistent.

3. Data Characterisation:

Variable	Description	Characterisation
accident_index	A long form unique reference number for each accident and/or casualty inclusive of the year	Ordinal

accident_year	Year in which the accident and/or casualty took place	Ordinal
accident_reference	A short form unique reference number for each accident and/or casualty	Ordinal
location_easting_osgr	Easting value giving accident location	Continuous
location_northing_osgr	Northing value giving accident location	Continuous
longitude	Longitude value giving accident location	Continuous
latitude	Latitude value giving accident location	Continuous
police_force	Police force of the accident	Nominal
accident_severity	Severity of the accident event	Nominal
number_of_vehicles	Number of vehicles involved in the accident	Discrete
number_of_casualties	Number of casualties involved in the accident	Discrete
date	Date of the accident	Ordinal
day_of_week	Day of the week of the accident	Ordinal
time	Time of the accicent	Ordinal
local_authority_district	Local authority district of the accident	Nominal
local_authority_ons_district	A different division of local authority district of the accident	Nominal
local_authority_highway	Another division of local authority district of the accident based on highways	Nominal
first_road_class	Class of the first road of the accident	Nominal
first_road_number	Number of the first road of the accident	Ordinal
road_type	Type of the road of the accident	Nominal

speed_limit	Speed limit of the road of the accident	Nominal
junction_detail	Type of the junction of the accident	Nominal
junction_control	Junction control at the accident	Nominal
second_road_class	Class of the second road of the accident	Nominal
second_road_number	Number of the second road of the accident	Nominal
pedestrian_crossing_human_control	Type of human control for pedestrian crossing at the accident	Nominal
pedestrian_crossing_physical_facilities	Type of pedestrian crossing at the accident	Nominal
light_conditions	Light conditions at the accident	Nominal
weather_conditions	Weather conditions at the accident	Nominal
road_surface_conditions	Road surface conditions at the accident	Nominal
special_conditions_at_site	Special conditions at the accident	Nominal
carriageway_hazards	Road hazards at the accident	Nominal
urban_or_rural_area	Type of area at the accident	Nominal
did_police_officer_attend_scene_of_accident	If police officer was at the accident	Nominal
trunk_road_flag	Road management at the accident	Nominal
lsoa_of_accident_location	Census data subdivision of area	Nominal
vehicle_reference	A unique reference number for each vehicle in an accident	Ordinal
casualty_reference	A unique reference number for each casualty in an accident	Ordinal
casualty_class	Class of casualty	Nominal

sex_of_casualty	Sex of casualty	Nominal
age_of_casualty	Age of casualty	Discrete
age_band_of_casualty	Age band of casualty	Nominal
casualty_severity	Severity of the casualty event	Nominal
pedestrian_location	Location of pedestrian casualty	Nominal
pedestrian_movement	Movement of pedestrian casualty	Nominal
car_passenger	Location of car passenger casualty	Nominal
bus_or_coach_passenger	Location of bus/coach passenger casualty	Nominal
pedestrian_road_maintenance_worker	If the pedestrian casualty was a road maintenance worker	Nominal
casualty_type	Type of the casualty	Nominal
casualty_home_area_type	Deprivation index of casualty	Nominal
casualty_imd_decile	Home area type of casualty	Nominal

4. Detailed Analysis

a. Demographics of Casualties

To study how casualties are differing demographically, the data for sex was first analysed, it was not surprising to see that most casualties are men, possibly due to a higher propensity for unsafe driving that leads to lower insurance premiums for females than for males, although this does include non-driver casualties, *(Figure 1).* The proportion of male casualties increased from slight to serious to fatal casualties, which highlights the greater risk to their life than to females. The sex of the driver or rider was also analysed, to make a comparison. *Table 1* shows that approximately two-thirds of casualties are when the driver or rider is male, supporting our previous notions that male drivers are more unsafe.

Looking at the deprivation values in *Table 2*, we clearly see that deprivation has a direct positive relationship with the number of casualties, but with the limitations of this dataset, we cannot comment on why there is such a relationship due to a lack of data on other social factors. When the age band of casualties was investigated, it was seen that number of casualties increases with age, especially increasing past the age of being able to drive, i.e. after the 16-20 age band, and continuing to increase until the 26-35 age band, before decreasing through middle and old age, with this clearly depicted in *Figure 2*. This could be due to maturity and experience leading to an increased sense of safety and road awareness, with sex also seeming to not have any bearing on the number of casualties with increasing age.

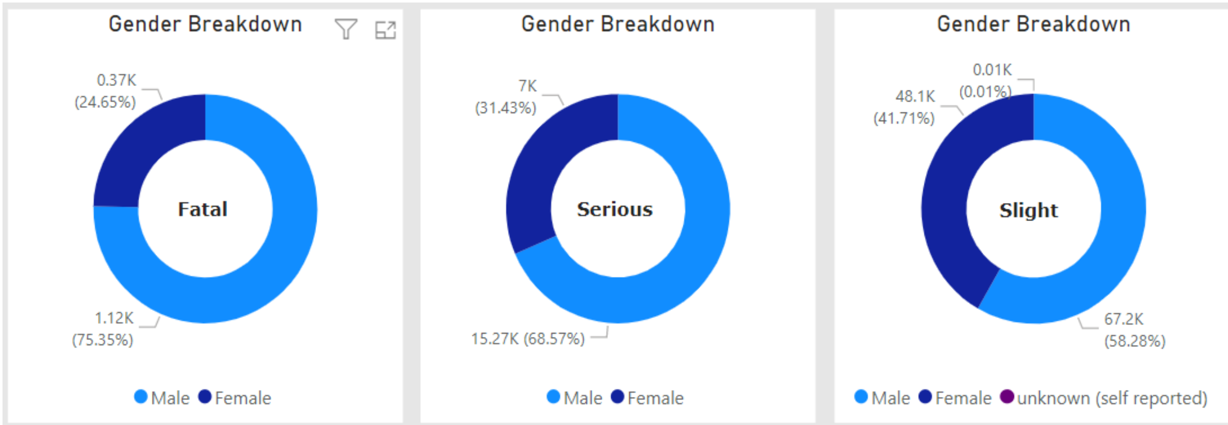

Figure 1

Table 1

Percentage of Casualties When The Driver Was Male or Female

Casualty Class	Sex Of Casualty	% of Total Count of Number Of Casualties along Casualty Class, Sex Of Casualty	Number Of Casualties
Driver or rider	Female	32.70%	29,752
	Male	67.30%	61,227

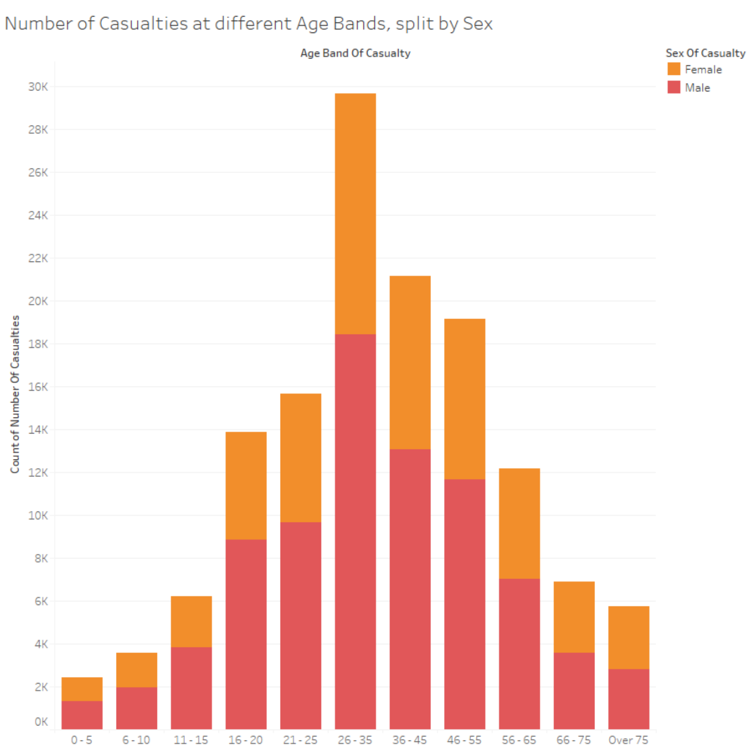

Figure 2

Table 2

Deprivation Level of Casualties

Casualty Deprivation Band	
Least deprived 10%	6.811%
Less deprived 10-20%	7.873%
Less deprived 20-30%	8.368%
Less deprived 30-40%	9.129%
Less deprived 40-50%	9.848%
More deprived 40-50%	10.275%
More deprived 30-40%	11.331%
More deprived 20-30%	11.929%
More deprived 10-20%	12.636%
Most deprived 10%	11.800%

b. KSI casualty accidents in different local authorities

There are more than 300 local authorities in this dataset, therefore meaning analysing individual districts would result in a level of detail that could not be contained within the scope of this report. Since these districts come under a certain police force, it is logical to analyse at a higher level by the police force, although there are 39 police forces in England which is still a very large number for visualisation. *Figure 3* shows the number of casualties for each police force area that controls the local authority regions, where it can be clearly seen that the metropolitan police have by far the highest number of casualties, which makes sense as London is the most populated city in England by far. The remaining local authorities have a similar number of casualties, and it appears that there is no relationship between urban and rural areas within each police force region.

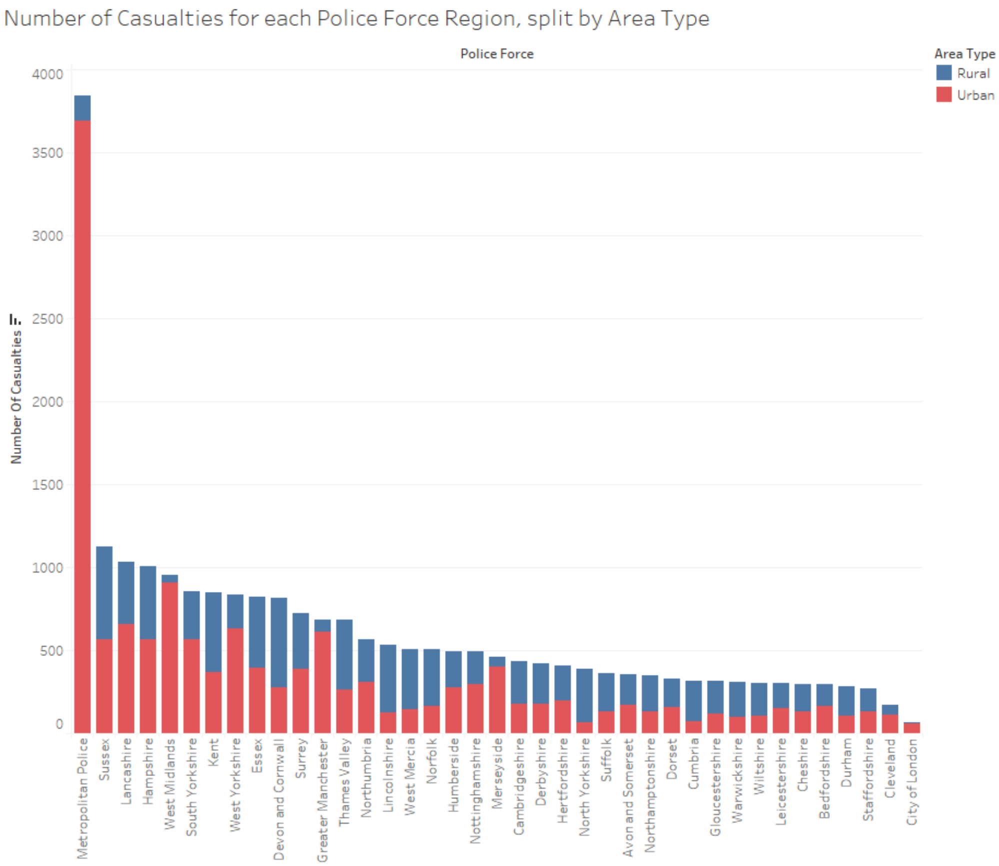

Figure 3

The five police forces having the most casualties were focussed on for the analysis shown in *figure 4* as otherwise there would be too many data points. The primary logical attribute to analyse was "*Did the police officer attend the scene of the accident"* because other attributes like weather and road conditions are not in the control of the local authority. This figure shows the top 5 police force regions having the most casualties. Looking at these results, the Hampshire police force has the best response to casualties as they made it to the scene 87% of the time, followed by Metropolitan, Sussex, Kent and West Midlands forces respectively.

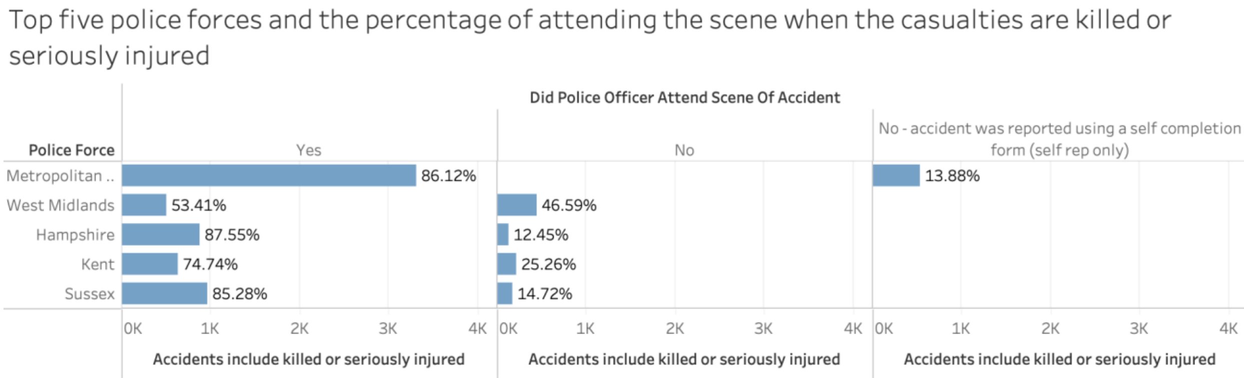

Figure 4

c. KSI casualty pedestrians

The group of KSI pedestrians has been split by age group, specifically into 2 pedestrian groups that are vulnerable to car accidents, school children from 6-20 years of age and elderly people of 66+ years. The number of casualties per hour of the day for these 5 age ranges is plotted in figure 5. For all groups it appears that the most casualties take place during daylight hours, increasing through the daytime and peaking in the afternoon for every group except 75+ years. For the 6-10 and 16-20 groups, there are similar patterns of an initial spike in casualties in the mid-morning and a peak spike in the afternoon. These can possibly be explained by accidents occurring when travelling to and from school or work in the morning or afternoon, which coincides with a rush hour

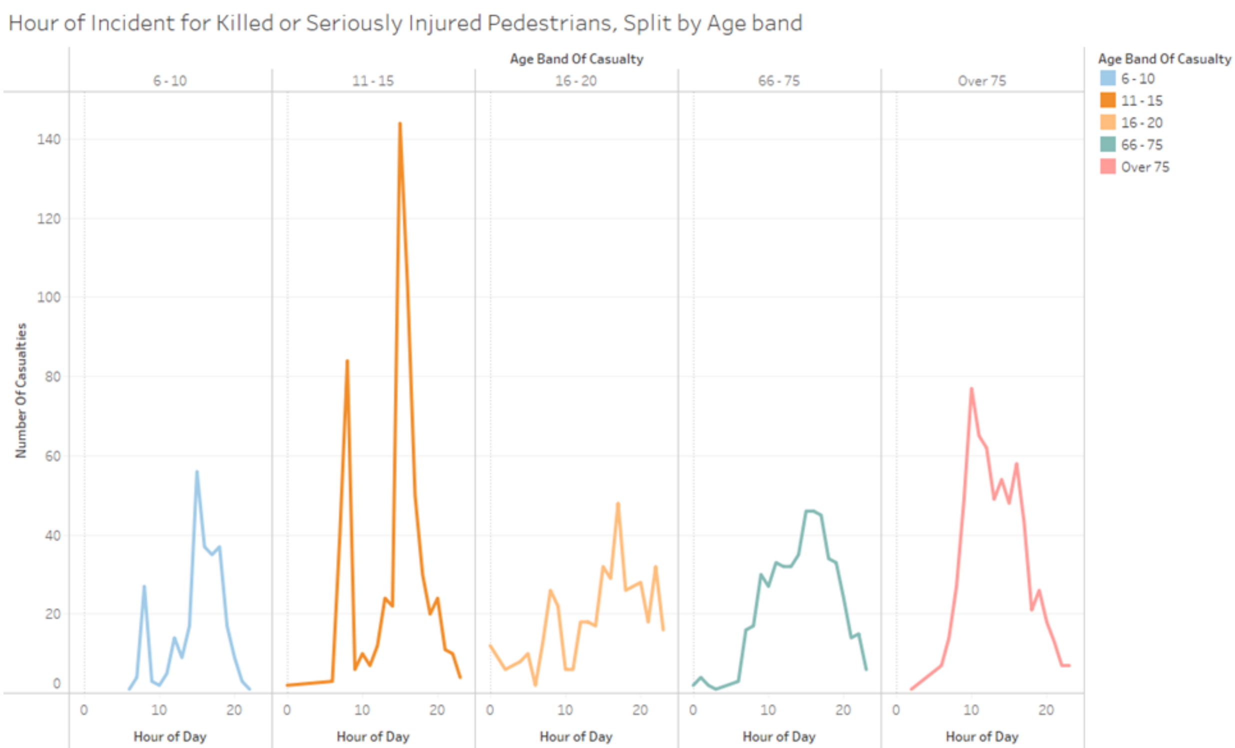

Figure 5

when there are more vehicles on the road. In terms of the 11–15-year group, we see the same trend but much more amplified with very high spikes in casualties in the mid-morning and afternoon that are much higher than for any other groups. This may be because this is the age at which children may start travelling to school without a parent looking after them (after 6-10), whilst them also not being of the age where they have road safety awareness (before 16-20). The 75+ sees the second highest peak, with a higher spike in the morning and a smaller spike in the afternoon, with the higher number of casualties than the 66-75 group being explained by the loss of senses and mobility experienced at this age.

Killed or Seriously Injured Pedestrian Locations

Pedestrian Location	
Crossing elsewhere within 50m. of pedestrian crossing	413
Crossing in zig-zag approach lines	28
Crossing in zig-zag exit lines	9
Crossing on pedestrian crossing facility	924
In carriageway, crossing elsewhere	2,291
In carriageway, not crossing	474
In centre of carriageway - not on refuge, island or central reservation	295
On footway or verge	472
On refuge, central island or central reservation	48
Unknown or other	491

Pedestrian Crossing Physical Facilities for Killed or Seriously Injured Pedestrians

Pedestrian Crossing Physical Facilities	
Central refuge	261
Data missing or out of range	4
Footbridge or subway	14
No physical crossing facilities within 50 metres	3,406
Pedestrian phase at traffic signal junction	650
Pelican, puffin, toucan or similar non-junction pedestrian light crossing	650
unknown (self reported)	66
Zebra	394

Figure 6

Also, *Figure 6* shows the proportion of different locations in which a KSI pedestrian was involved in an accident. For 2,291 of these incidents, the pedestrian was crossing away from any infrastructure such as at a crossing or within zig-zag lines, with this being the highest proportion by far. This is explained by the fact that the pedestrian in these cases is crossing unsafely and in a way such that they may not be immediately apparent, or their presence expected by drivers, leading to more accidents occurring. Following this is the group of 924 pedestrians who were crossing on the pedestrian crossing facility, which is expected to be a large group as most pedestrians crossing the road will be using these facilities. Furthermore, the proportion of different pedestrian crossing facilities at the accident location, supports these findings, as 3,406 pedestrians were killed or seriously injured where there was no crossing facility within 50 metres, further bolstering the idea that crossing in unsafe areas without using the provided safety infrastructure leads to more casualties.

5. Conclusions

To conclude the analysis, studying the demographic of casualties portrayed that most casualties are of men for all severity levels, most casualties are of age band 26-35 and deprivation is directly related to casualties, where increasing deprivation is accompanied by a greater number of casualties. While analysing the patterns between the top 5 highest-casualty police areas, the Hampshire police force led in terms of response to reaching the accident site. Finally, while analysing patterns for KSI casualties, Pedestrians of ages 6 years old to 20 years old ended up having by far the most casualties during school start and finish timings. More than 50% of the pedestrian casualties occurred due to there being no physical crossing facility within 50 meters, with these casualties mostly taking place directly on carriageways.