Road Accident and Safety Data

Introduction

Between 2012 and 2019, the number of road deaths plateaued at around 1,850 per year, five per day (UK road death and casualty statistics, 2020). This report is based on 3 data sets comprising road safety data for 2019. In Figure 1, accident data has 32 columns including 117,536 entries. Figure 2 has vehicles data with 23 columns including 216,381 entries and figure 3 has casualties data with 16 columns including 153,158 entries.

Data cleaning, such as removing missing values or -1, changing numbers to words, and merging data connect by accident index, has been undertaken to improve data quality. Moreover, showing feature's importance which specific feature will have a larger effect on accident severity, and shows accident prediction by using machine learning. This report aims to recommend some actions that government could do based on an analysis of accident severity in various situations.

```
RangeIndex: 117536 entries, 0 to 117535
Data columns (total 32 columns):
 #   Column
---  ------
 0   Accident_Index
 1   Location_Easting_OSGR
 2   Location_Northing_OSGR
 3   Longitude
 4   Latitude
 5   Police_Force
 6   Accident_Severity
 7   Number_of_Vehicles
 8   Number_of_Casualties
 9   Date
 10  Day_of_Week
 11  Time
 12  Local_Authority_(District)
 13  Local_Authority_(Highway)
 14  1st_Road_Class
 15  1st_Road_Number
 16  Road_Type
 17  Speed_limit
 18  Junction_Detail
 19  Junction_Control
 20  2nd_Road_Class
 21  2nd_Road_Number
 22  Pedestrian_Crossing-Human_Control
 23  Pedestrian_Crossing-Physical_Facilities
 24  Light_Conditions
 25  Weather_Conditions
 26  Road_Surface_Conditions
 27  Special_Conditions_at_Site
 28  Carriageway_Hazards
 29  Urban_or_Rural_Area
 30  Did_Police_Officer_Attend_Scene_of_Accident
 31  LSOA_of_Accident_Location
dtypes: datetime64[ns](1), float64(4), int64(23),
```

Figure 1

```
RangeIndex: 216381 entries, 0 to 216380
Data columns (total 23 columns):
 #   Column
---  ------
 0   Accident_Index
 1   Vehicle_Reference
 2   Vehicle_Type
 3   Towing_and_Articulation
 4   Vehicle_Manoeuvre
 5   Vehicle_Location-Restricted_Lane
 6   Junction_Location
 7   Skidding_and_Overturning
 8   Hit_Object_in_Carriageway
 9   Vehicle_Leaving_Carriageway
 10  Hit_Object_off_Carriageway
 11  1st_Point_of_Impact
 12  Was_Vehicle_Left_Hand_Drive?
 13  Journey_Purpose_of_Driver
 14  Sex_of_Driver
 15  Age_of_Driver
 16  Age_Band_of_Driver
 17  Engine_Capacity_(CC)
 18  Propulsion_Code
 19  Age_of_Vehicle
 20  Driver_IMD_Decile
 21  Driver_Home_Area_Type
 22  Vehicle_IMD_Decile
dtypes: int64(22), object(1)
```

Figure 2

```
RangeIndex: 153158 entries, 0 to 153157
Data columns (total 16 columns):
 #   Column
---  ------
 0   Accident_Index
 1   Vehicle_Reference
 2   Casualty_Reference
 3   Casualty_Class
 4   Sex_of_Casualty
 5   Age_of_Casualty
 6   Age_Band_of_Casualty
 7   Casualty_Severity
 8   Pedestrian_Location
 9   Pedestrian_Movement
 10  Car_Passenger
 11  Bus_or_Coach_Passenger
 12  Pedestrian_Road_Maintenance_Worker
 13  Casualty_Type
 14  Casualty_Home_Area_Type
 15  Casualty_IMD_Decile
dtypes: int64(15), object(1)
```

Figure 3

Road accidents by hours of day and day of week with accident severity

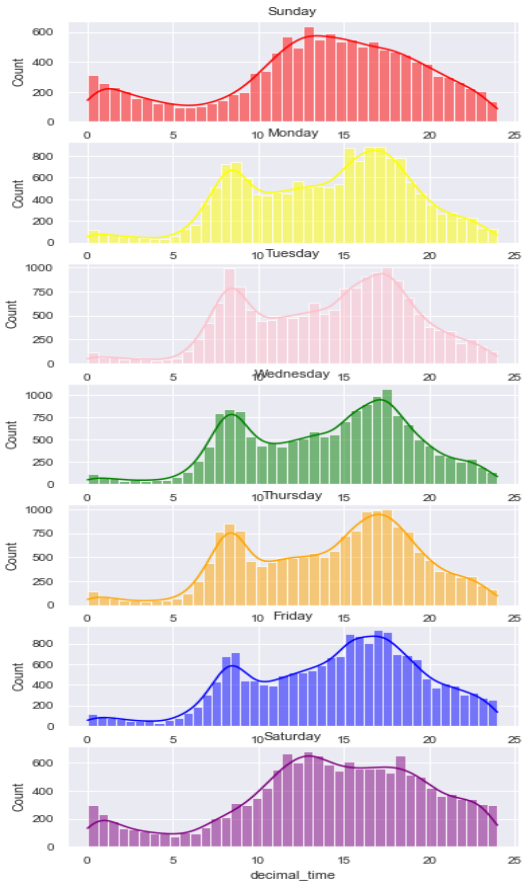

Figure 4

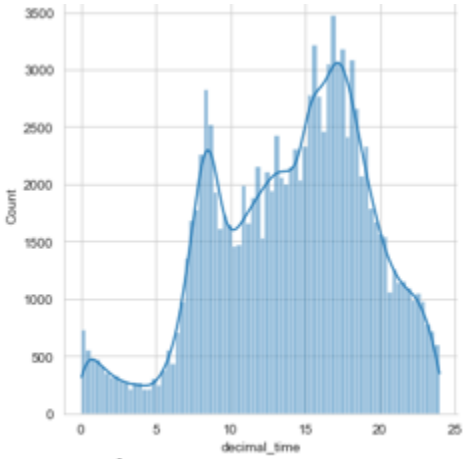

Figure 5

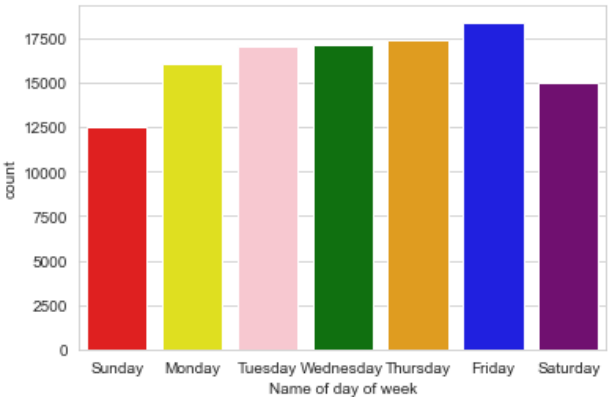

Figure 6

Figure 4 shows on weekdays daily work and working time generates the most accident. Especially, in the end, working time is the most accident followed by starting time respectively. Moreover, on Friday night the number of accidents is increasing, and affect to figure 6 shows Friday is the most accident. As a result, overall accident cases on weekends are less than on weekdays.

Name of Accident Severity	Fatal	Serious	Slight	sum
Name of day of week				
Friday	235	3670	14500	18405
Thursday	225	3489	13647	17361
Wednesday	238	3376	13560	17174
Tuesday	237	3294	13482	17013
Monday	224	3139	12738	16101
Saturday	260	3245	11492	14997
Sunday	238	2925	9356	12519

Table 1

Table 1 shows that most situation is slight, serious, and fatal respectively. Although Saturday hasn't accident too many, Saturday has the most fatal accident.

Motorcycle accidents by hours of day and day of week with accident severity

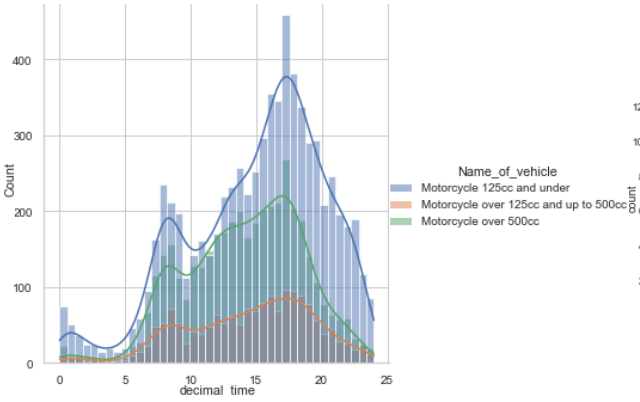

Figure 7

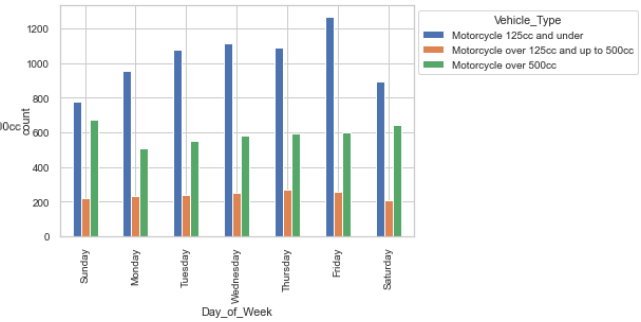

Figure 8

Figures 7 and 8 show motorcycles 125cc and under is the most accident at the end of working time and the most accident on weekdays followed by motorcycles over 500cc and motorcycles over 125cc and up to 500cc respectively. However, on weekends, the number of accidents decreased. On the one hand, the number of accidents on motorcycles over 500cc on weekends increased. Contrary to motorcycle 125cc.

Name of Accident Severity	Fatal	Serious	Slight	sum
Name_of_vehicle				
Motorcycle 125cc and under	57	1941	5174	7172
Motorcycle over 500cc	236	1842	2062	4140
Motorcycle over 125cc and up to 500cc	34	545	1097	1676

Table 2

Name of Accident Severity	Fatal	Serious	Slight	sum
Name of day of week				
Friday	42	689	1393	2124
Thursday	39	661	1252	1952
Wednesday	42	593	1310	1945
Tuesday	36	533	1286	1855
Saturday	63	639	1039	1741
Monday	34	568	1097	1699
Sunday	71	645	956	1672

Table 3

Tables 2 and 3 show the number of accidents on Sunday in the last place. However, accident severity on Sunday is the most fatal and a lot serious. Moreover, accident severity of motorcycles over 500cc is more violent because the number of serious fatal and serious is too much. Compared with the number of summations. In my opinion, motorcycles over 500cc reach high speed it more risks of serious and fatal in accidents severity.

Pedestrians involved in accident by hours of day and day of week with accident severity

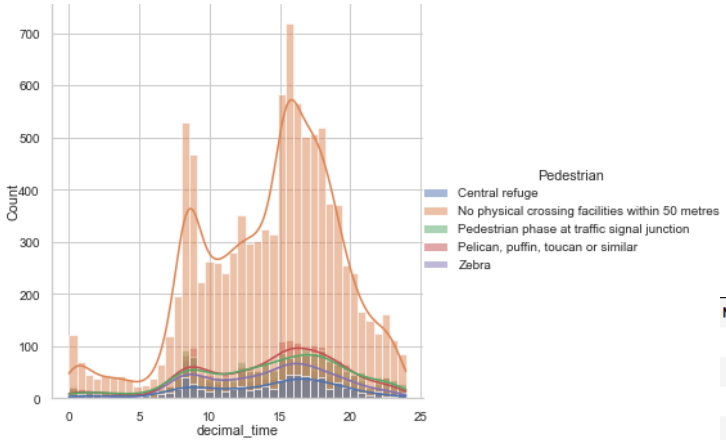

Figure 9

Name of Accident Severity	Fatal	Serious	Slight	sum
Pedestrian				
No physical crossing facilities within 50 metres	241	2873	7058	10172
Pelican, puffin, toucan or similar	44	604	1257	1905
Pedestrian phase at traffic signal junction	35	557	1235	1827
Zebra	16	290	974	1280
Central refuge	21	233	445	699

Table 4

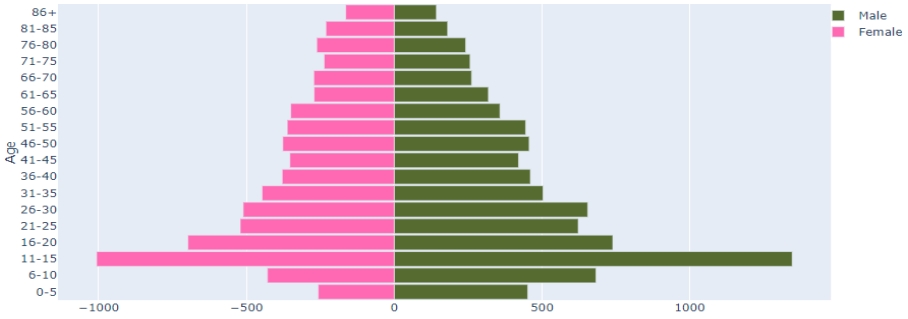

Figure 10

Table 4 shows that in most accidents in that pedestrians involve are no physical crossing facilities within 50 meters. Figure 9 shows pedestrian time is not the same as a motorcycle because in figure 10 most pedestrians who are involved in accidents are children and teenagers. Moreover, the number of accidents in the morning is the same as at start of office hours. However, students leave school is not at the same time as the end of working hours. It's faster than the end working hour.

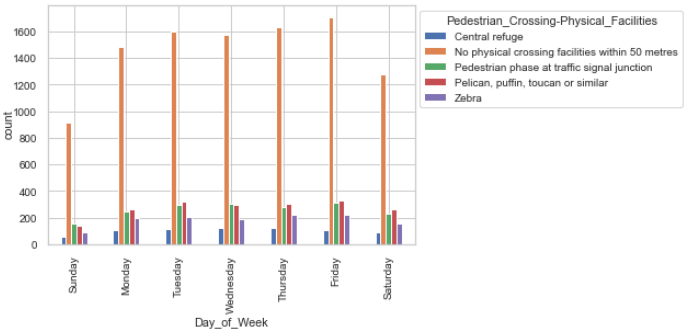

Figure 11

Name of Accident Severity	Fatal	Serious	Slight	sum
Name of day of week				
Friday	60	765	1854	2679
Thursday	62	733	1754	2549
Tuesday	42	695	1791	2528
Wednesday	50	687	1735	2472
Monday	42	634	1606	2282
Saturday	66	605	1341	2012
Sunday	35	438	888	1361

Table 5

Figure 11 and table 5 show Friday is the most accident. Although Saturday hasn't accident too many, Saturday has the most fatal accident.

Impact of daylight saving on road accidents in week start and stop

	Week_Saving_light	day_light_saving	Accident_Case
3	1 week before daylight saving start	No daylight savings	2250
1	1 week after daylight saving start	Daylight savings	2019
2	1 week before daylight saving end	Daylight savings	2131
0	1 week after daylight saving end	No daylight savings	2205

Table 6

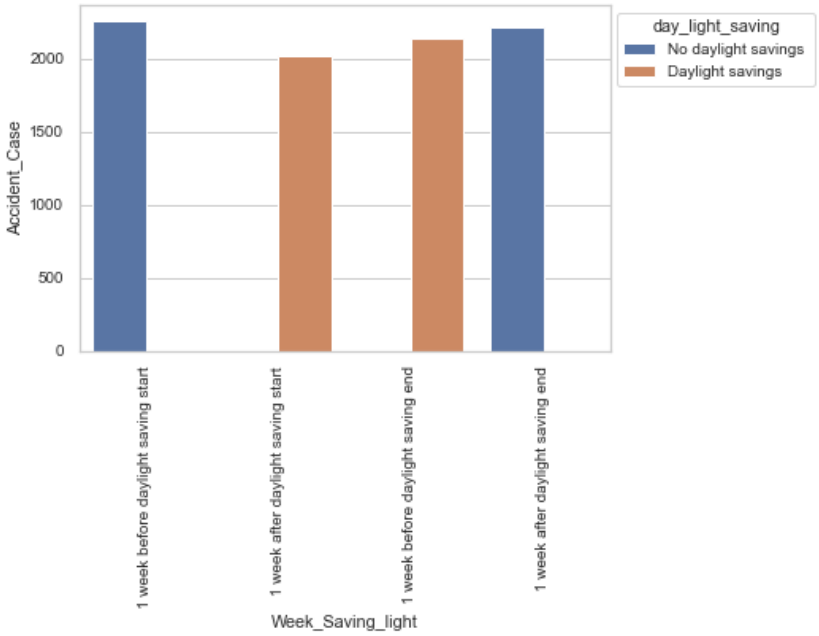

Figure 12

(Belam, 2019) implied that range of daylight saving is from 31 March 2019 until 27 October 2019. Figure 12 and table 6 show that 1 week before and 1 week after daylight saving (No daylight savings) are more accident cases than during daylight saving periods. In my opinion, no daylight saving in nighttime period is longer than daylight saving.

	Seasons	Accident_Case
1	Spring	4269
0	Autumn	4336

Table 7

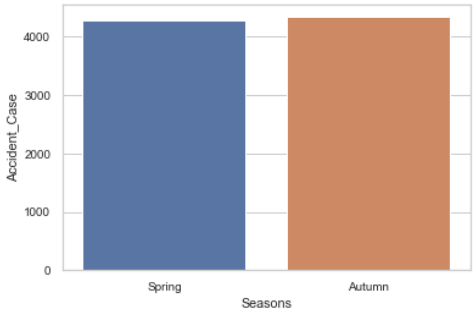

Figure 13

Table 7 and figure 13 show that there are more accidents in autumn when daylight savings end compared with spring, when daylight savings begin, probably because autumn has fewer hours of daylight than spring. 11:54 hours in March and 10:47 hours in October (Worlddata.info, no date).

Impact of sunrise and sunset on road accidents

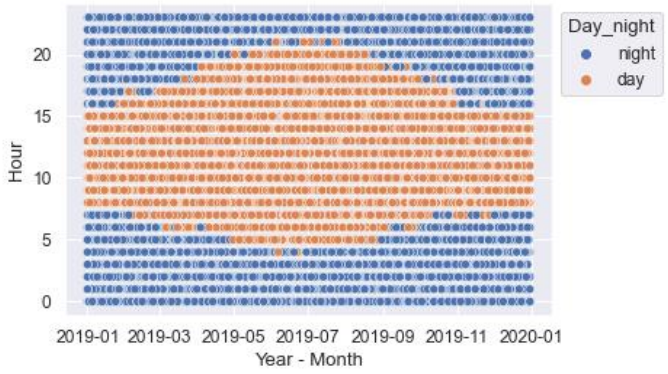

Figure 14

(Worlddata.info, no date) recorded the average sunrise and sunset times in the UK. Figure 14 shows sunrise and sunset time are different every month.

Sunrise

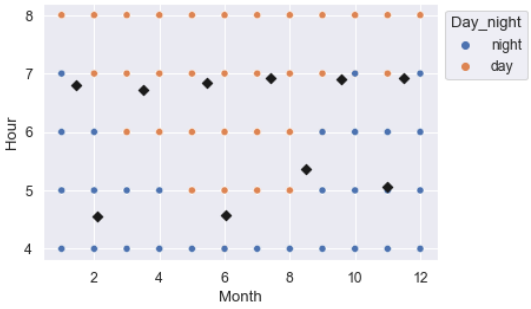

Figure 15

Sunset

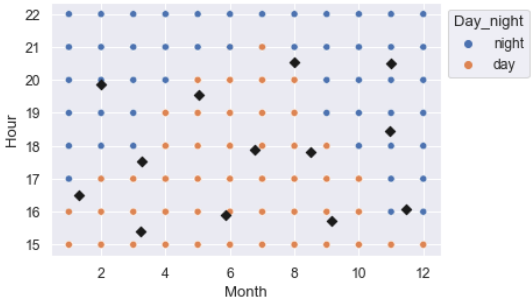

Figure 16

Sunrise

Figure 15 shows sunrise in 12 months from 04:00 to 08:00 depending on each month, most of the black points (cluster) around 7 am because many people go to office. Besides, 04:00 to 06:00 the cluster of accident cases happens during the month at nighttime.

Sunset

Figure 16 shows sunset in 12 months from 15:00 to 22:00 depending on each month, most of the black points (cluster) from 15:00 to 18:00 because of work timeout. Besides, the cluster of accident cases happens during the month during nighttime and twilight period.

As a result, first factor that affects accidents of sunrise and sunset is nighttime and second factor of accidents is twilight it clearly shows in sunset time cluster of accidents is between day and night.

Types of vehicles involved on road with accident severity

Name of Accident Severity Name_of_vehicle	Fatal	Serious	Slight	sum
Car	1519	19022	88035	108576
Pedal cycle	92	3080	10305	13477
Motorcycle	332	4433	8472	13237
Goods	389	2326	8990	11705
Bus	64	611	2563	3238
Other vehicle	57	373	832	1262

Table 8

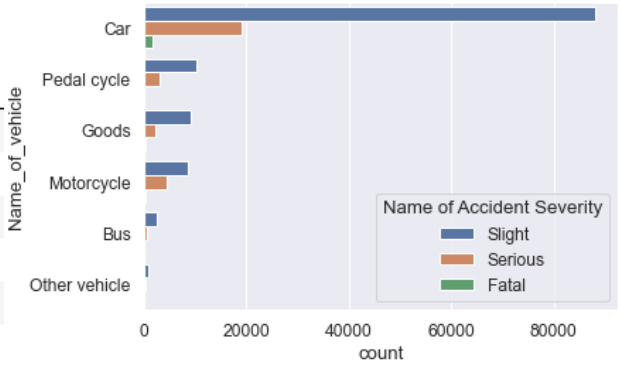

Figure 17

Table 8 and figure 17 show car is the most accident case in the UK followed by pedal cycle, motorcycle, and goods respectively. Most accident severity of vehicle is slightly followed by serious and fatal. Car is the most accident severity. However, most car accident severity trend to slight. On the other hand, the most risk of a serious accident is a motorcycle because the number of serious accidents is more than half of slight accidents. Therefore, they must be treated with caution.

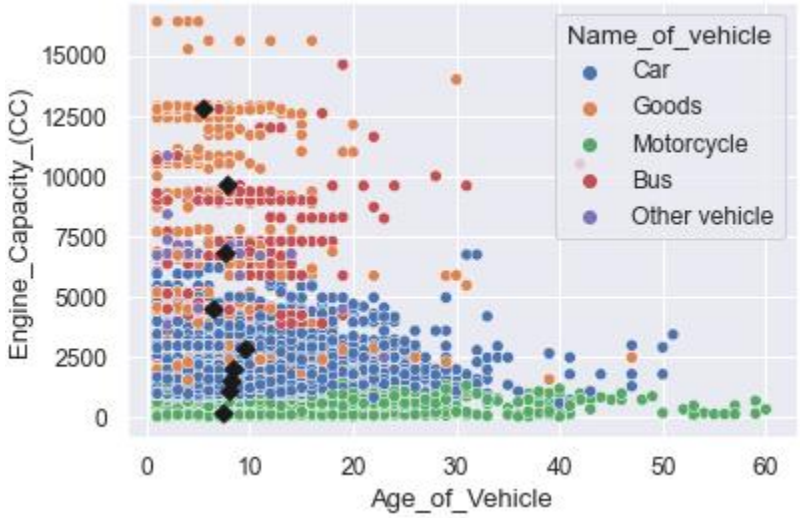

Figure 18

Figure 18 shows the number of clusters between engine capacity (CC) and age of vehicle. However, type of vehicle will decrease because some vehicle has no engine capacity. As a result, the most accident is vehicles 7-10 years an engine capacity lower than 3000cc.

Condition and situation of road accidents

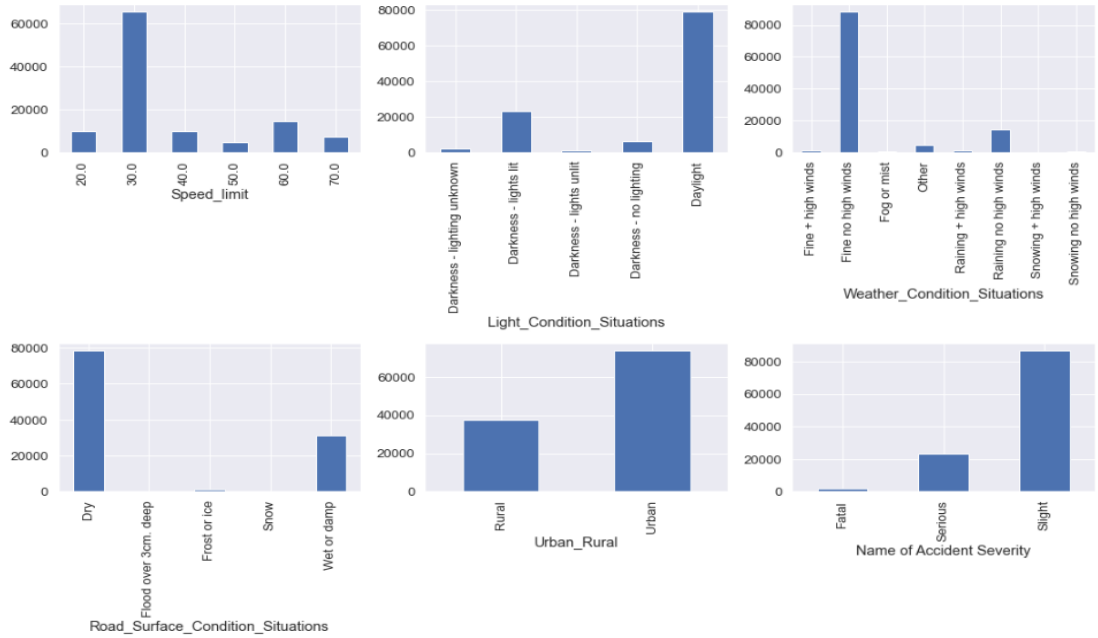

Figure 19

Figure 19 shows the conditions of accidents in the UK most accidents happen in urban area more than in rural area twice and accident severity in the UK are slight, serious, and fatal. The dangerous place is road where speed limit is 30.

	support	itemsets
0	0.793669	(Weather_Fine no high winds)
6	0.711570	(Light_Daylight)
2	0.706621	(Road_Dry)
7	0.669543	(Weather_Fine no high winds, Road_Dry)
10	0.593828	(Weather_Fine no high winds, Light_Daylight)

Table 9 Support rating on condition and situation.

Table 9 shows the number of weathers fine no high winds situation is the highest support followed by light condition is daylight and road surface condition situation dry.

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
(Road_Wet or damp, Light_Daylight)	(Weather_Raining no high winds)	0.153315	0.131417	0.070151	0.457558	3.481733	0.050002	1.601246
(Weather_Raining no high winds)	(Road_Wet or damp, Light_Daylight)	0.131417	0.153315	0.070151	0.533802	3.481733	0.050002	1.816150
(Weather_Raining no high winds)	(Road_Wet or damp)	0.131417	0.279317	0.127419	0.969581	3.471261	0.090712	23.692066
(Road_Wet or damp)	(Weather_Raining no high winds)	0.279317	0.131417	0.127419	0.456182	3.471261	0.090712	1.597194
(Road_Wet or damp)	(Weather_Raining no high winds, Light_Daylight)	0.279317	0.072713	0.070151	0.251151	3.454005	0.049841	1.238283
(Weather_Raining no high winds, Light_Daylight)	(Road_Wet or damp)	0.072713	0.279317	0.070151	0.964761	3.454005	0.049841	20.451555

Table 10 Lift rating on condition and situation.

Table 10 shows the relationship of many situations together by market basket analysis. Now let's focus on lift value. It showed light condition is daylight, road is wet or damp and weather raining with no high winds.

The most accident severity location on highways

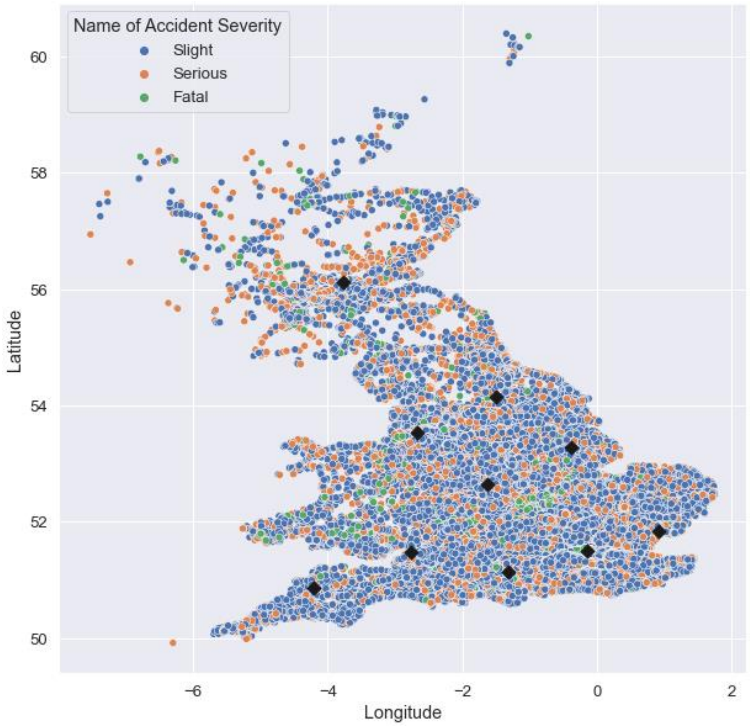

Figure 20

Name of Accident Severity Local_Authority_(Highway)	Fatal	Serious	Slight	sum
Kent	36.0	658.0	2911.0	3605.0
Surrey	27.0	649.0	2285.0	2961.0
Birmingham	14.0	410.0	2189.0	2613.0
Hampshire	33.0	586.0	1766.0	2385.0
Essex	38.0	655.0	1691.0	2384.0
Lancashire	38.0	683.0	1583.0	2304.0
West Sussex	26.0	499.0	1456.0	1981.0
Hertfordshire	24.0	354.0	1571.0	1949.0
Lincolnshire	48.0	405.0	1436.0	1889.0
Norfolk	34.0	423.0	1183.0	1640.0

Table 11

Figure 20 shows 10 clusters of accidents in the UK. Table 9 shows dangerous areas (highways) in the UK. It showed the most area is Kent, Surrey, and Birmingham respectively. Moreover, the most Fatal accident is in Lincolnshire.

Driver variables affect the outcome

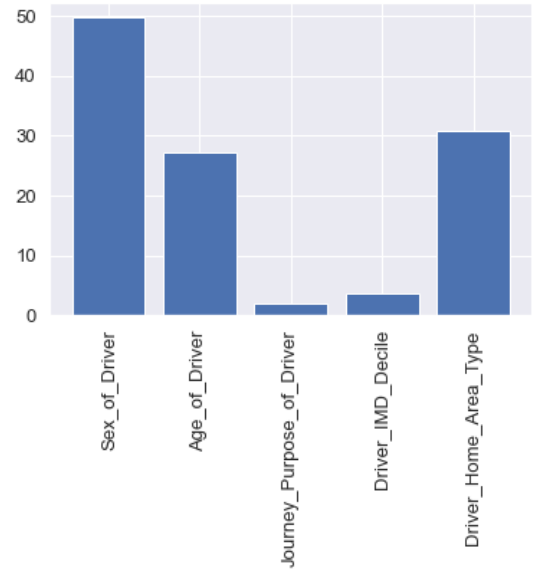

Figure 21

Figure 21 shows variables that driver affects to outcome are sex of driver, driver's home area type, and age of driver respectively. The graph below explains most variables related to outcomes such as the sex of driver and age of driver.

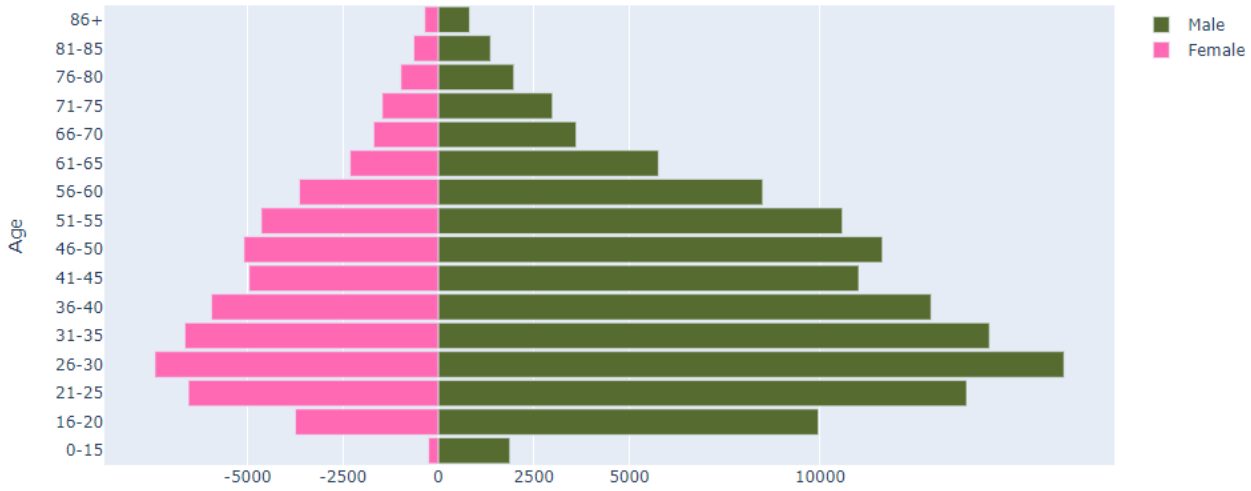

Figure 22

Figure 22 shows male tends to be more accident than female. (Gov.UK, 2015) implied that people can start driving a car when age more than 16 years old. Therefore, drivers lower than 15 years old drive a pedal cycle or smuggle drive car. Thus, it should focus on more than 16 years old. More and more people who are young adult (26-30 years old) is the most involved in the accident. Followed by 31 -35 and 21-25 respectively. As a result, the most age of the accident is 21 to 40 years old (working age).

Predictions about when and where accidents will occur accident severity and compared to government model

In the government model of accident severity, there is no fatal severity. Therefore, a comparison is made between serious and slight accidents.

My model

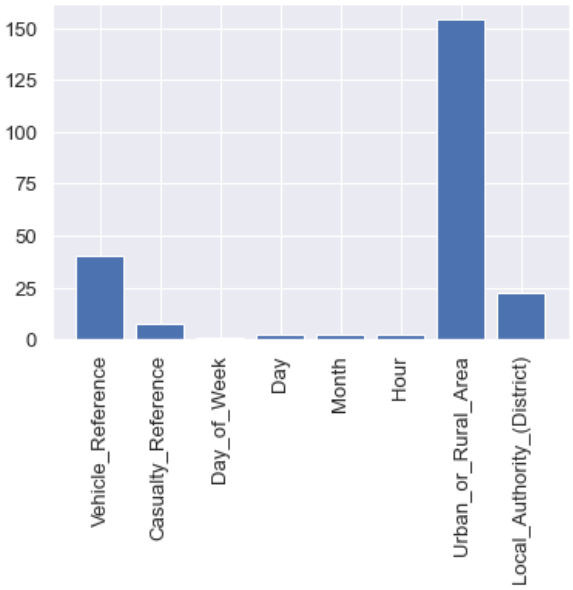

Figure 23

This uses the same features as the government models, such as Casualties reference and Vehicles reference, but includes additional features, such as when - Day of Week, Day, Month, Hour – and where - Urban or Rural Area, and Local Authority (District) in figure 23. Moreover, when and where variables that affect to outcome is urban or rural area.

```
DecisionTreeClassifier with Random Search
Train_accuracy 75.13898097618198
Test_accuracy 74.69400244798041
KNeighborsClassifier with Random Search
Train_accuracy 99.63788442903045
Test_accuracy 82.42809057527539
RandomForestClassifier with Random Search
Train_accuracy 99.68633651247004
Test_accuracy 86.3984088127295
NaiveBayes with Random Search
Train_accuracy 74.7564645279747
Test_accuracy 74.2579559363525
```

Figure 24

Figure 24 shows there are 4 classifiers to compare the model and the best result is Random Forest Classifier which uses a random search to improve the model.

Compared my model to the government model

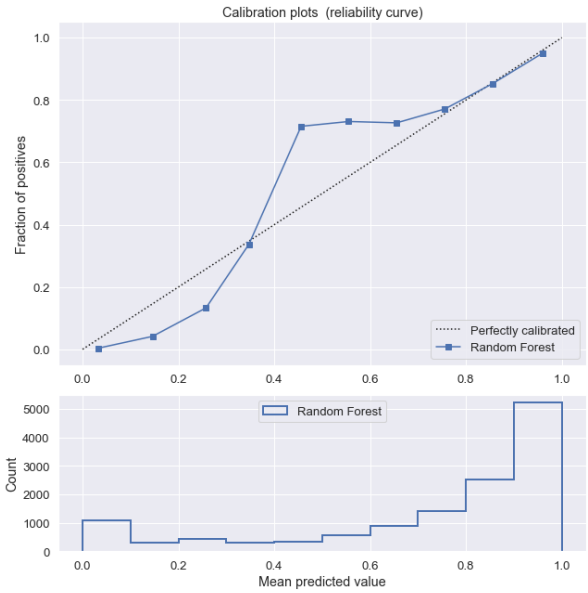

Figure 25 my model

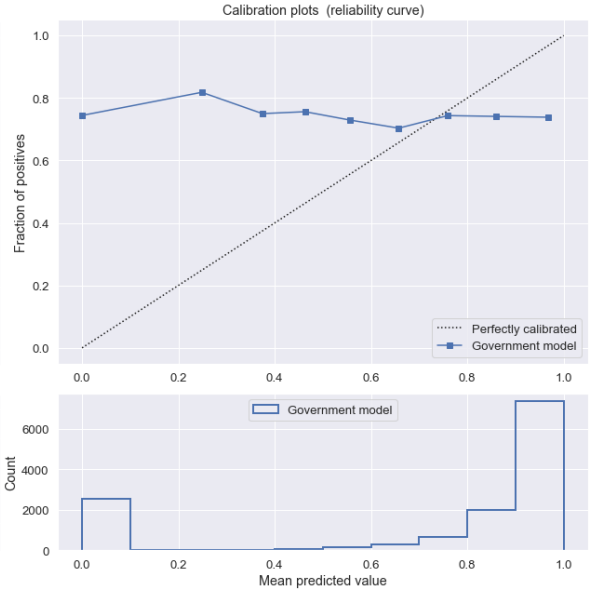

Figure 26 government model

Figures 25 and 26 show calibration plots and mean predicted value (probability value). The accident probability more than 0.5 is slight and probability lower than 0.5 is serious.

Critical thinking with my model and government model

The calibration plots show my model is better than government model because my model is closer to the diagonal line (perfectly calibrated).

However, calibration plots are not useful to compare models because inputs are not the same, my model contains when and where factors, whereas the government model does not.

The mean predicted value shows the government model is better than my model because my model has many distributions of probabilities and many values around 0.5, which is not good. On the other hand, the government model has lower distribution around 0.5.

Recommendation

There are four things that government should get involved in to decrease accident cases to be safety

Persuade more people to use public transport

• On weekdays most accidents happen during working time and studying time, government should decrease factors for people who use private transportation by increasing the number of buses. Especially, start working time and end. On the other hand, on the weekend, government should make more routes for people who go to travel long distances should use train.

Law enforcement and provide education

• For pedestrians, government law enforcement when they cross the road improperly and educate someone who misunderstands or acts carelessly because most accident casualties are students and teenagers. According to speed limit accidents, government law enforcement reduces speeds to 20 mph everywhere in urban areas.

Solved the environmental situation and Improve facilities

• Weather conditions, road surface, and light conditions have affected accidents. Government should have increased traffic signs for warnings and lamp posts because some places may lack them. It can cause accidents at night. Especially, no daylight saving that night longer than a day or sunrise, sunset at twilight time effect drivers is blur or cognitive decline.

• One area of controversy is smart motorways that have no hard shoulder and deaths have occurred. Government has been told to delay introduction of more smart motorways pending availability and analysis of evidence regarding incidence of accidents.

Competition and public relations

• A key issue is young drivers and always has been. There is a high proportion who drive without insurance. The issue is not properly addressed by government, in my view, and needs a PR campaign to address it, as well as stronger penalties.

Reference

Belam, M. (2019) “When do the clocks go forward in 2019 and could this be the last time?,” *The guardian*, 30 March.
Available at: https://www.theguardian.com/uk-news/2019/mar/26/when-do-the-clocks-go-forward-change-2019-daylight-savings-time
(Accessed: April 8, 2022).

Sunrise and sunset in the United Kingdom (no date) *Worlddata.info*.
Available at: https://www.worlddata.info/europe/united-kingdom/sunset.php
(Accessed: April 8, 2022).

Government Digital Service (2015) *Driving lessons and learning to drive*, *GOV.UK*.
Available at: https://www.gov.uk/driving-lessons-learning-to-drive
(Accessed: April 8, 2022).

UK road death and casualty statistics (2020) Brake.

Available at: https://www.brake.org.uk/get-involved/take-action/mybrake/knowledge-centre/uk-road-safety

(Accessed: April 15, 2022).