Improve vehicle road safety by analysing accident data

submitted towards partial fulfilment of the criteria for award of PGPBA by GLIM

Students Name	Rohith Nayak, Sathyanarayana M N, Benny Mathew Kuruvila, George P Abraham, Tarun Kamal Khiani
Group No	5
Batch	PGPBA-BI JAN 2017
Project Mentor	Neelesh Singh
Program Director	Dr. P K Vishwanathan

Abstract

Throughout the world, roads are shared by cars, buses, trucks, motorcycles, mopeds, pedestrians, animals, taxis and other categories of travelers. Travel made possible by motor vehicles supports economic and social development in many countries. Yet each year, these vehicles are involved in crashes that are responsible for millions of deaths and injuries.

Although most countries have taken specific steps to road safety lots needs to be done to avoid fatal & serious accidents.

Identification of critical factors will enable countries to highlight and create awareness programs around these critical factors.

Techniques	Classification – Logistics Regression
Tools	R, Tableau
Domain	Public Services

Acknowledgements

We wish to place on record our deep appreciation for the guidance and help provided to us by our Mentor Mr. Neelesh Singh for guiding us in completing this project on time.

We would also like to place on record our appreciation for the guidance provided by Dr. P K Vishwanathan for giving us valuable feedback and being a source of inspiration in helping us to work on this project.

We certify that the work done by us for conceptualizing and completing this project is original and authentic.

Date: 24 November 2017
Place: Bangalore

Rohith Nayak
Sathyanarayana M N
Benny Mathew Kuruvila
George P Abraham
Tarun Kamal Khiani

GREAT LAKES INSTITUTE OF MANAGEMENT

Postgraduate Program Business Analytics and Business Intelligence

Certificate

This is to certify that the PGPBA Capstone Project titled "Improve vehicle road safety by analysing accident data" is a bonafide record of the project work carried out by "Rohith Nayak, Sathyanarayana M N, Benny Mathew Kuruvila, George P Abraham, Tarun Kamal Khiani" in fulfilment of requirements for the award of PGPBA-BI in Great Lakes Institute of Management.

December 2017

Neelesh Singh
Project Mentor

Dr. P K Vishwanathan
Program Director

Date: December 04, 2016
Place: Bangalore

Declaration

Improve vehicle road safety by analysing accident data

The Project report is submitted in partial fulfilment of academic requirements for PGPBA-BI in Great Lakes Institute of Management.

This report is a result of our analysis of UK Accident data from Kaggle and can also be found with the government of UK. All sections of the text, which has been obtained from other sources, are fully referenced. We understand that cheating and plagiarism constitute a breach of institute regulations and will be dealt with accordingly.

Signature	
Students Name	Rohith Nayak, Sathyanarayana M N, Benny Mathew Kuruvila, George P Abraham, Tarun Kamal Khiani
Group No	5
Batch	PGPBA-BI 2017
Date	24 November 2017

Table of Contents

Introduction

Throughout the world, roads are shared by cars, buses, trucks, motorcycles, mopeds, pedestrians, animals, taxis and other categories of travelers. Travel made possible by motor vehicles supports economic and social development in many countries. Yet each year, these vehicles are involved in crashes that are responsible for millions of deaths and injuries.

Consider the following:

- Each year, 1.25 million people are killed on roadways around the world.
- Each day, an estimated 3,400 people are killed globally in road traffic crashes involving cars, buses, motorcycles, bicycles, trucks, or pedestrians. Half of those people killed in crashes globally are pedestrians, motorcyclists, and cyclists.
- Road traffic injuries are estimated to be the eighth leading cause of death globally and the leading cause of death for young people aged 15–29.
- Current trends show that by 2030, road traffic injuries will become the seventh leading cause of death globally.
- Road traffic injuries place a huge economic burden on low- and middle-income countries and are estimated to cost US $518 billion globally and US $65 billion in low-income and middle-income countries, exceeding the total amount received in development assistance.

All countries have road safety policies, programs and awareness campaigns whether you are a driver, passenger, cyclist, or pedestrian, take the following steps to stay safe on the road:

- Use a seat belt in every seat, on every trip, no matter how short.
- Make sure children are always properly buckled in the back seat in a car seat, booster seat, or seat belt, whichever is appropriate for their age, height, and weight.
- Choose not to drive while impaired by alcohol or drugs, and help others do the same.
- Obey speed limits.
- Drive without distractions (such as using a cell phone or texting).
- Be alert when crossing streets, especially in countries where motorists drive on the left side of the road.
- Ride only in marked taxis and try to ride in those that have seat belts.
- Avoid riding in overcrowded, overweight, or top-heavy buses or minivans.

Although most countries have taken specific steps to road safety lots needs to be done to avoid fatal & serious accidents. Identification of critical factors will enable countries to highlight and create awareness programs around these critical factors.

Scope & Objective

The UK has one of the best road safety records in the world, but more needs to be done to prevent deaths (fatal) and serious injuries.

The UK police force classify accidents as

- **Fatal**: An accident in which at least one person is killed; other casualties (if any) may have serious or slightly injuries. Human casualties who sustained injuries which caused death less than 30 days after the accident.

- **Serious**: One in which at least one person is seriously injured but no is killed. An injury for which a person is detained in hospital as an "in-patient", or any of the following injuries whether or not they are detained in hospital: fractures, concussion, internal injuries, crushing, burns, severe cuts, severe general shock requiring medical treatment and injuries causing death 30 or more days after the accident.

- **Slight**: An injury of a minor character such as a sprain (including neck whiplash injury), bruise or cut which are not judged to be severe, or slight shock requiring roadside attention. This definition includes injuries not requiring medical treatment.

The UK police forces collect data on every vehicle collision in the UK on a form called Stats19. The data files provide detailed information on the accident, the vehicles involved and the causalities from 2004 -2015.

We intend to use this data to gain insights and understand factors influencing fatal & serious accidents so that we can propose specific campaigns and road safety measures that would reduce the number of serious accidents moving forward. This study though related to the UK has a relevance across the world.

Data Sources

Data Source	Details
Accidents	Accident records across 2004 to 2015 - Accident Severity - No. of vehicles - No. of causalities - Day & Time of accident - Location - Road Type - Weather Conditions - Light Conditions - Pedestrian Information
Vehicles	Details of all the vehicles involved in an accident - Vehicle Type - Vehicle Manoeuvre - Accident Location - Age of driver - Gender of driver - Engine Capacity - Age of vehicle - Driver in home area or not
Causalities	Details of all the casualties involved in an accident - Gender of casualty - Age of casualty - Casualty Type - Vehicle Reference - If pedestrian, location/movement/road maintenance worker - Car/Bus Occupant or not - Casualty in home area or not

For more details visit: https://www.kaggle.com/silicon99/dft-accident-data

Analytical Approach

- Aggregate the casualties, vehicles by accident index so that each row represents rich information about an accident

- Merge the casualties by accident and vehicles by accident with the accident data to arrive at data set which provides all information for an accident

- Explore the data and gain insights on the:
    - trend of accidents over time
    - relationship of accidents with the population of UK
    - accidents by time of day, day of the week
    - accidents by road conditions, junction, road type
    - accidents by light and weather conditions where accidents are
    - casualties with a focus on each category (pedestrian, occupant, driver, cyclist)
    - vehicles involved in the accident with a focus on each category
    - drivers and their characteristics involved in the accident
    - situations, factors that highly influence serious & fatal accidents

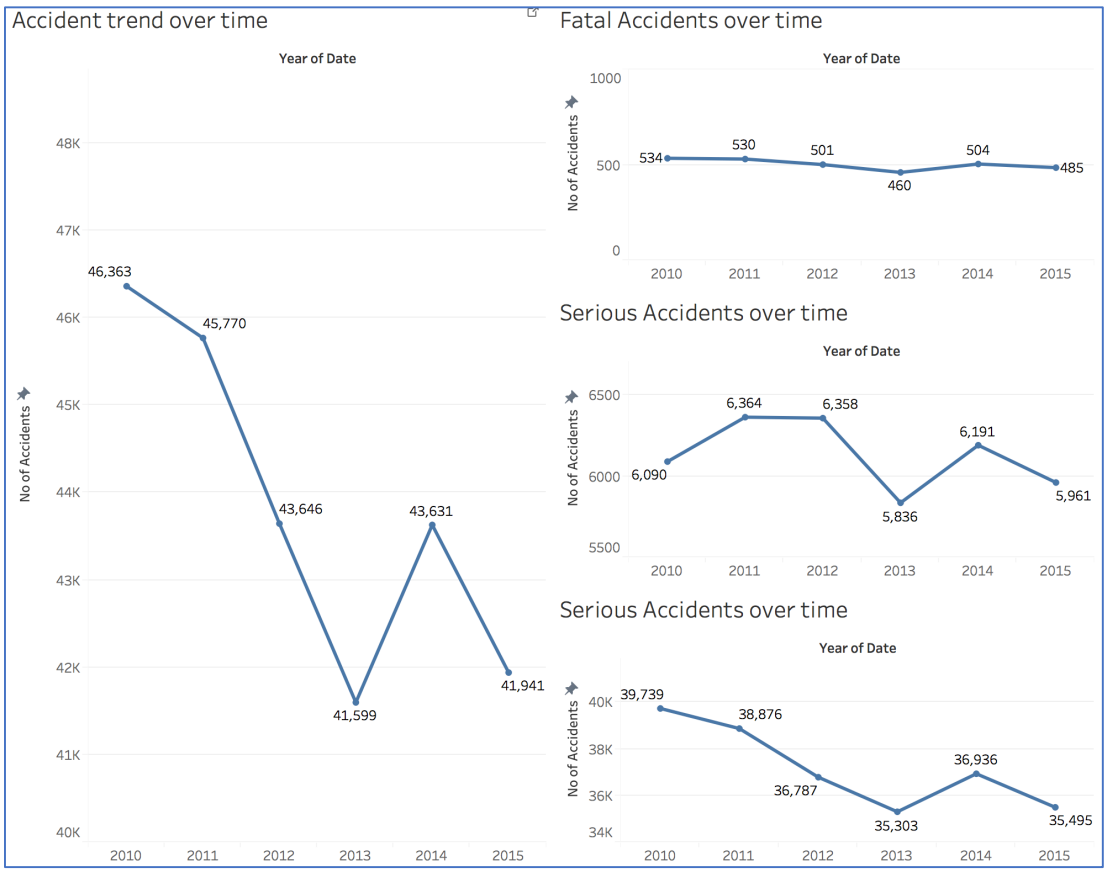
- Identify & Interpret factors that influence an accident to be fatal and serious
    - Accident Severity has 3 levels – 1 (Fatal), 2 (Serious), 3(Slight) hence this problem is a multi-class classification problem

    - **One vs Rest strategy**
      This strategy involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives.

    - Logistic regression to identify and quantify factors that influence an accident to be fatal & serious

Insights on Accidents

1. Declining trend of accidents across severity over time with a dip in 2013



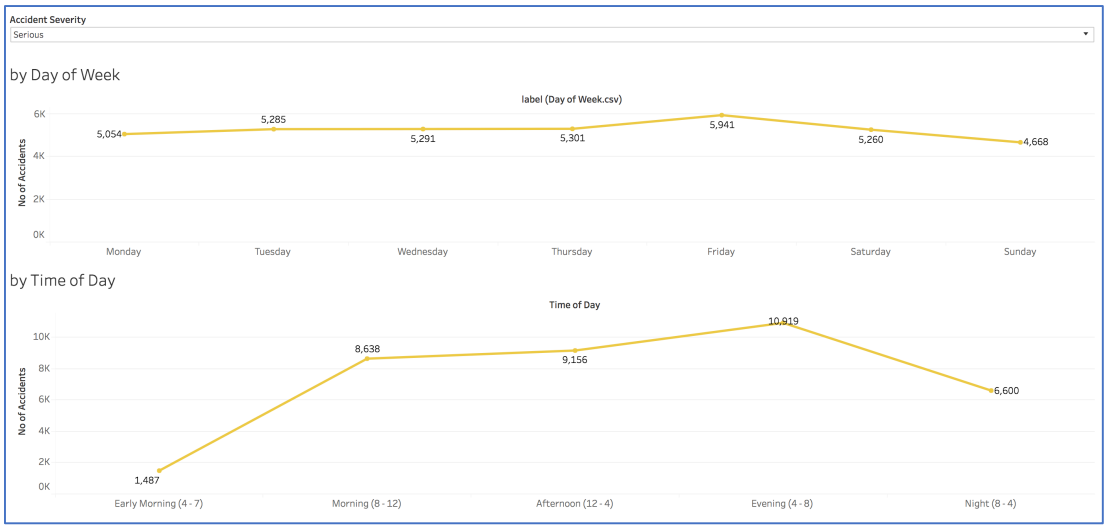
2. On an average, every year 14 out of every 100 accidents is serious, 1 out of every 100 accidents is fatal

Accident Severity as a % of total

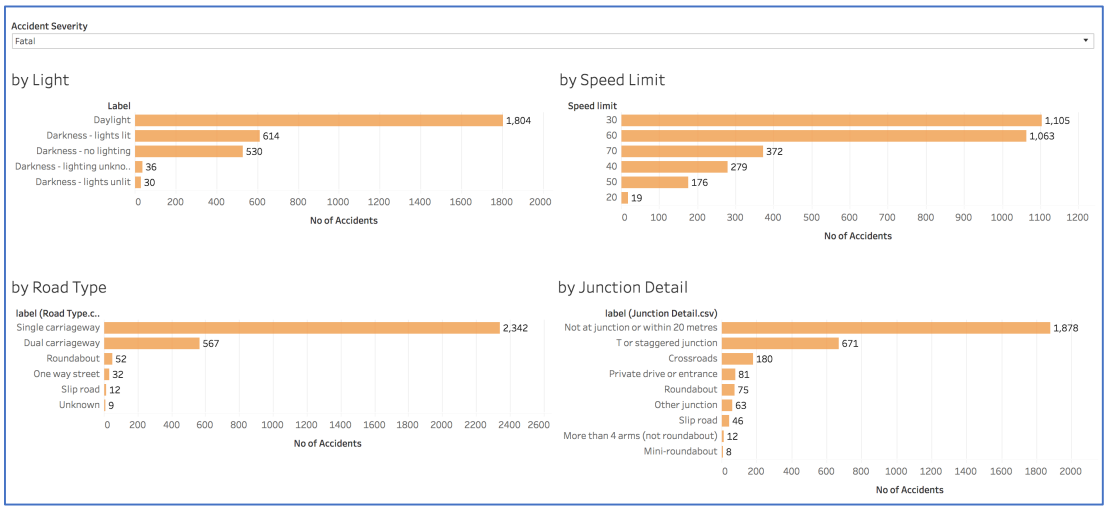
	Year of Date					
label (Accid..	2010	2011	2012	2013	2014	2015
Fatal	1.15%	1.16%	1.15%	1.11%	1.16%	1.16%
Serious	13.14%	13.90%	14.57%	14.03%	14.19%	14.21%
Slight	85.71%	84.94%	84.28%	84.87%	84.66%	84.63%

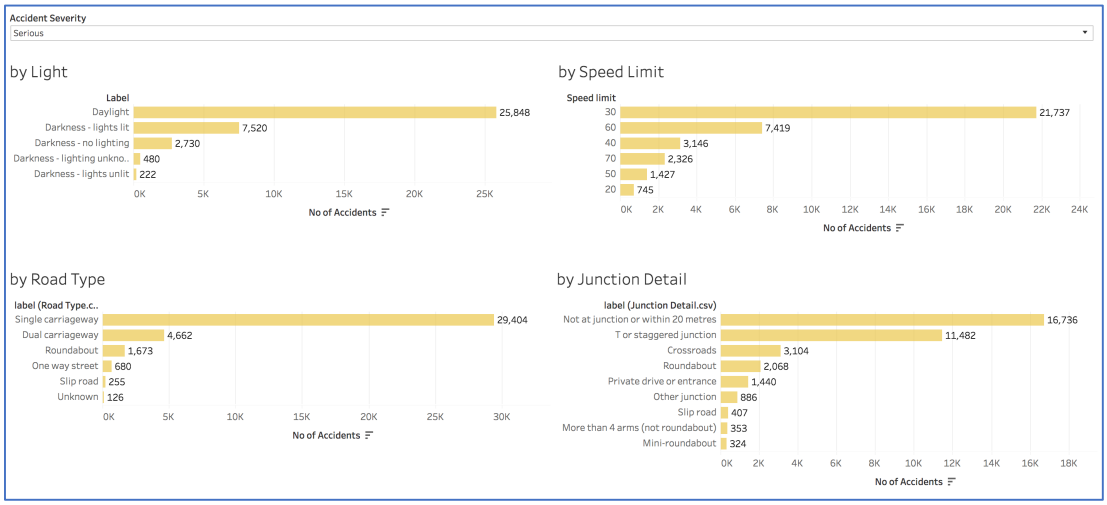
3. Fatal Accidents peak on Weekends and high numbers are observed from 8:00 PM to 4:00 AM

4. Serious Accidents peak on Fridays and most serious accidents occur from 4:00 PM to 8:00 PM

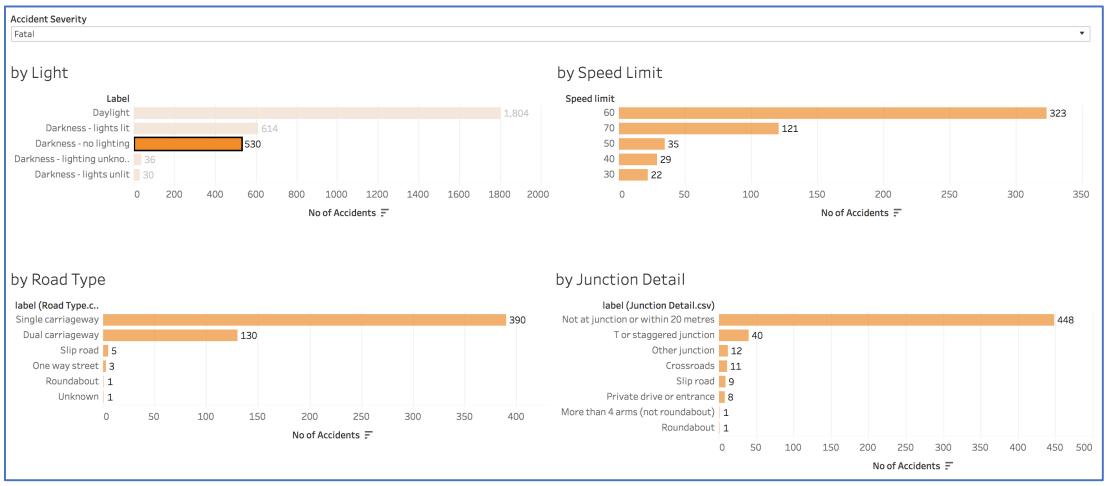


5. Most Fatal & Serious Accidents occur in daylight, on roads with 30 or 60 as speed limit, on single carriageways and at T/Staggered Junctions or not at a junction

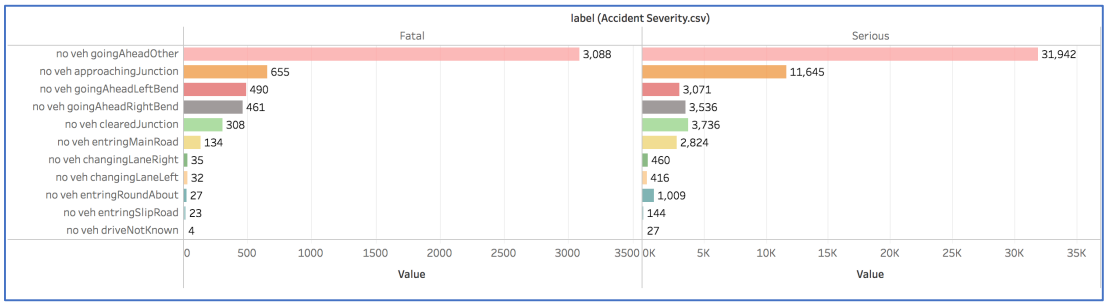




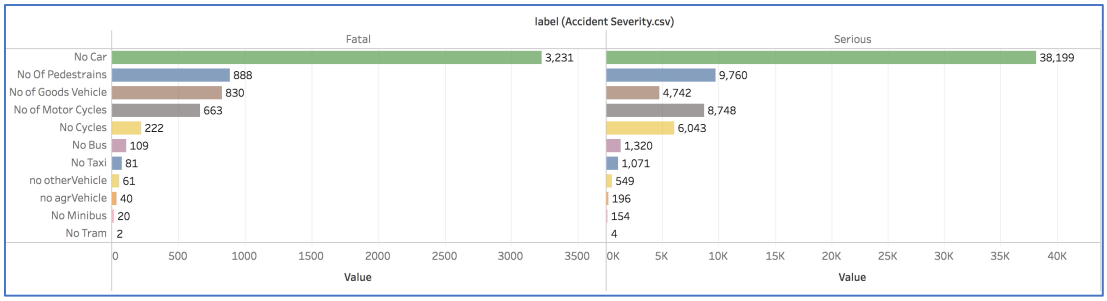
6. Fatal & Serious Accidents that in Darkness with no lighting occur at high speed limits of 60 & 70



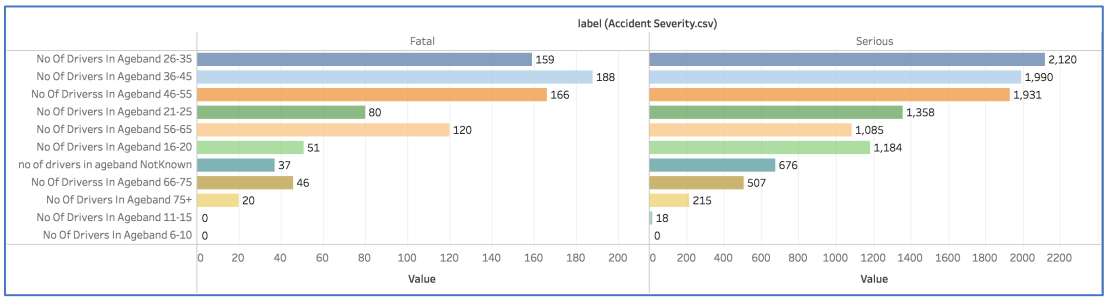
7. Overtaking and approaching junction are the prime situations when accidents are either fatal or serious and this is across vehicle categories



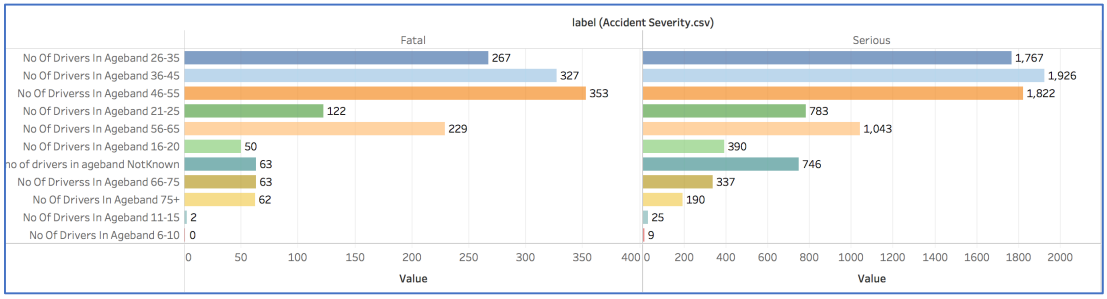
8. High number of cars, pedestrians, Goods Vehicles, Motor Cycles and Cycles are involved in Fatal or Serious Accidents



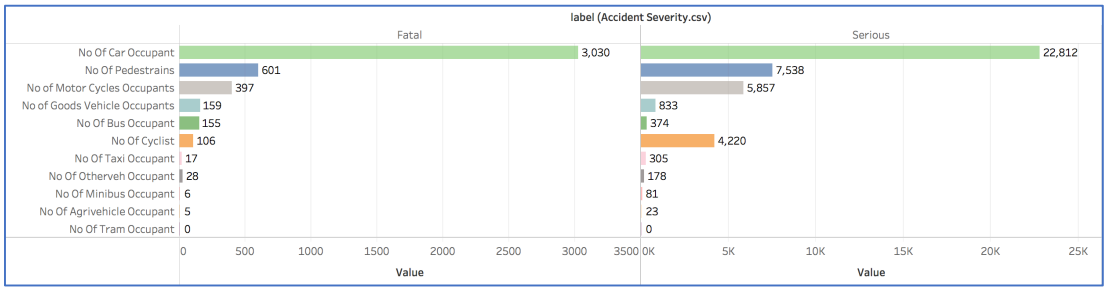
9. Car & Motor Cycle Drives in the age band 26 - 35, 36 - 45, 46 - 55 are primarily involved in Fatal & Serious accidents



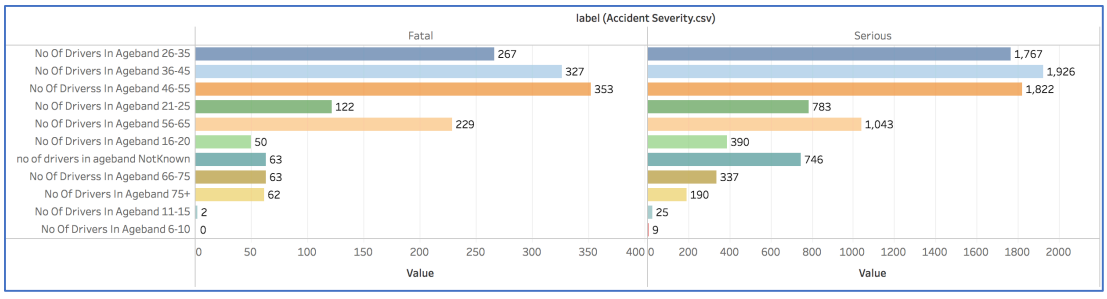
10. Goods Vehicle Drives in the age band 46 - 55, 36 - 45 are primarily involved in Fatal & Serious accidents



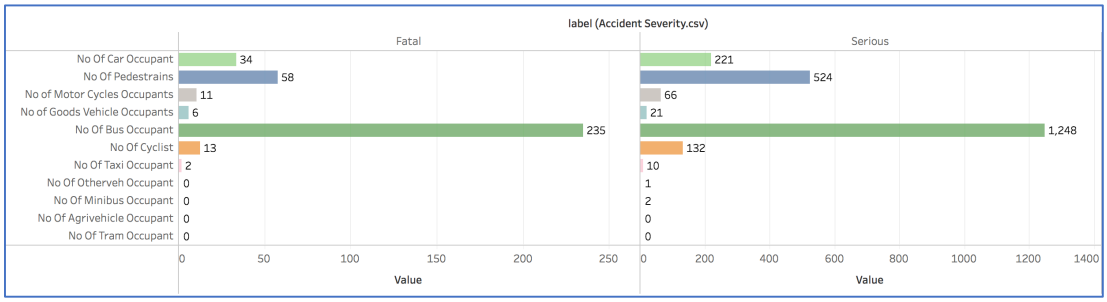
11. Cars cause the most no. of casualties. Car Occupants, Pedestrians (70%), Cyclist (50%) and Motor Cyclists (50%) are the ones affected



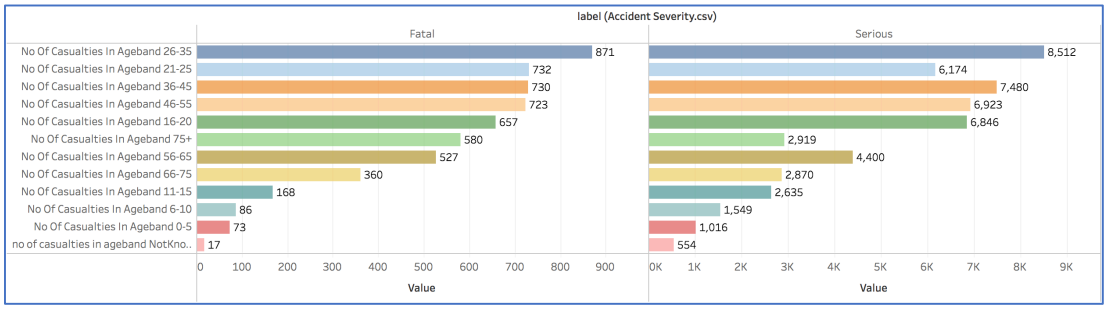
12. Goods Vehicles cause high no. of casualties. Pedestrians (25%), Cyclist (25%) and Motor Cyclists (33%) are the ones affected



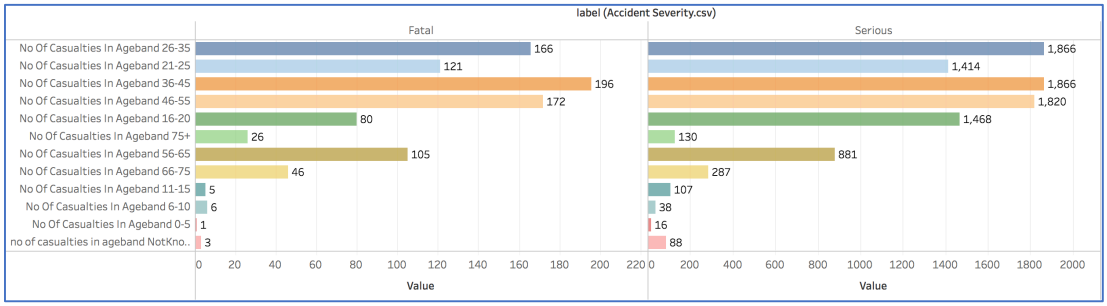
13. Trams, Agri Vehicles, Buses cause the least number of causalities



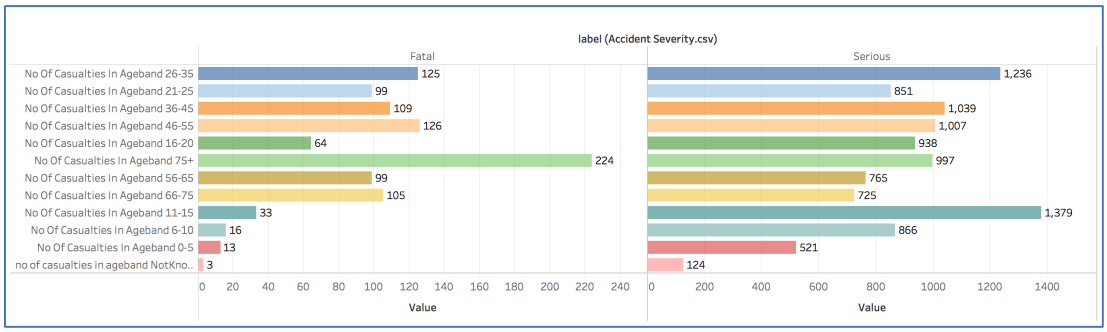
14. High no. of casualties is observed in the age groups of 26 – 35, 21 – 25, 36 – 45



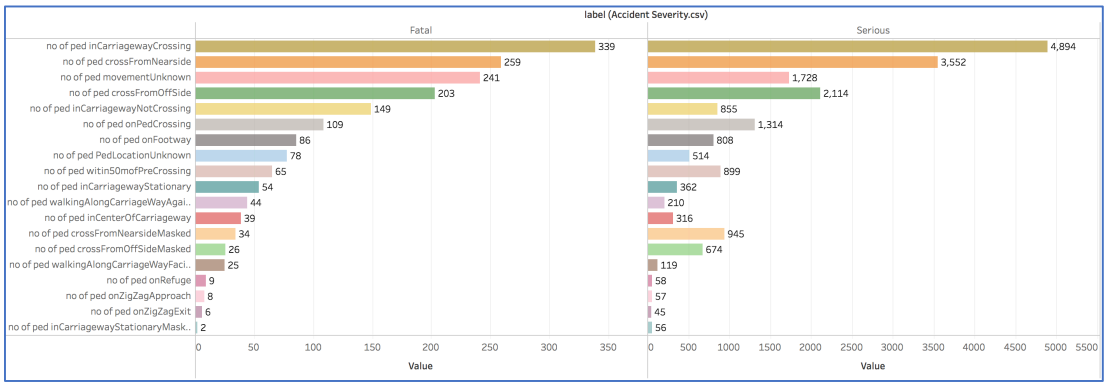
15. Fatality & Serious Injuries for Motor Cyclist (pillion & riders) are higher in the age group 36 – 45, 465 – 55 and 26 – 35



16. Fatality is high for Pedestrians over 75 years & Serious injuries are prominent for pedestrians in age group 6 – 15



17. Pedestrians are at high risk and crossing the carriageway, crossing from near side, offside are the primary reason for serious and fatal casualties



For more details visit
https://public.tableau.com/profile/tarun.khiani#!/vizhome/InsightsonAccidents/InsightsonAccidents

Data Preparation

Accident Severity has three levels – Slight, Serious and Fatal. We have converted this into two levels of Slight & Serious. Hence Accident Severity 'Serious' combines all accidents that are serious and fatal. The primary reason to do so is to convert it into a binary classification problem where we can use logistics regression to articulate factors that make an accident serious.

Accident Time was converted into a categorical variable with the following levels early morning, morning, afternoon, evening and night

Data from casualties and vehicles where aggregated by accident index and many features where created so that there is no information loss because of aggregation. This data was then merged with the accident data. The consolidated accident data from 2005 to 2015 now is 850 MB with 1,780,653 rows and 195 variables.

The number of accidents have decreased from 2005 to 2009 but the number of accidents from 2010 to 2015 are similar with very slight variation. Hence, we decided to pick 30% of the sample from 2010 to 2015 for our analysis.

Hence for building a model we now have a consolidated data file with 262,950 accidents and 195 variables.

Variable Selection

Variables like location, latitude, longitude, date, time in hours, no. of serious casualties, road numbers and if police visit the accidents spot or not were omitted while building the model as these variables can directly predict the result or variables that do not help in predicting or identification of factors leading to a serious accident.

All categorical variables where converted into dummies.

Variables with unknowns like `Junction_Control.-1`, `Weather_Conditions.-1`, `no_of_veh_in_ageNotKnown`, `no_of_ped_movementUnknown`,… were excluded. The reason for doing so is that such variables would not help in articulating the factors that make an accident serious.

This data is split into train (70%) and test (30%).

Fatal Accidents

- 3 Levels of Accident Severity were converted into 2 – 1 (fatal) & 0 (not fatal – serious & slight)
- Fatal Accidents account for 1.15% of the total accidents. For better prediction accuracy, the fatal accidents where oversampled in the training sample to arrive at 4.43%
- Logistics Regression Model was built on the training sample with 239 variables
- Eliminate variables that have a linear relationship with accident severity and have a high variance inflation factor. The reason to do so is that we wanted to eliminate the effect of multicollinearity and thus arrive at variables that can best explain serious accidents
- Eliminate variables that were not significant and had an odds ratio of less than 1

Serious Accidents

- 3 Levels of Accident Severity were converted into 2 – 1 (serious) & 0 (not serious – fatal & slight)
- serious Accidents account for 13.995% of the total accidents. For better prediction accuracy, the fatal accidents where oversampled in the training sample to arrive at 39.43%
- Logistics Regression Model was built on the training sample with 239 variables
- Eliminate variables that have a linear relationship with accident severity and have a high variance inflation factor. The reason to do so is that we wanted to eliminate the effect of multicollinearity and thus arrive at variables that can best explain serious accidents
- Eliminate variables that were not significant and had an odds ratio of less than 1

Factors influencing fatal accidents

The logistic model is **overall significant** as indicated by the likelihood ratio test.

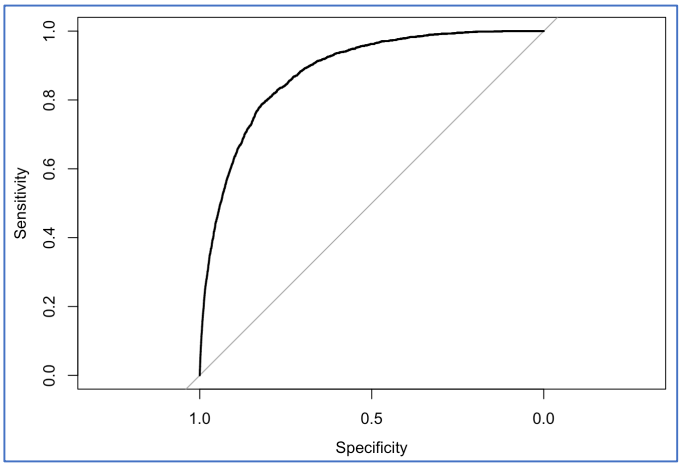
```
  #Df LogLik   Df Chisq Pr(>Chisq)
1 164 -25491
2   1 -34550 -163 18119  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The **McFadden R^2 is 0.2622068** which indicates that 26.22 % of the uncertainty of the intercept only model can be explained by the logistics regression model. A higher ratio (closer to 40 %) would be preferable but we can proceed at it indicates reasonable performance.

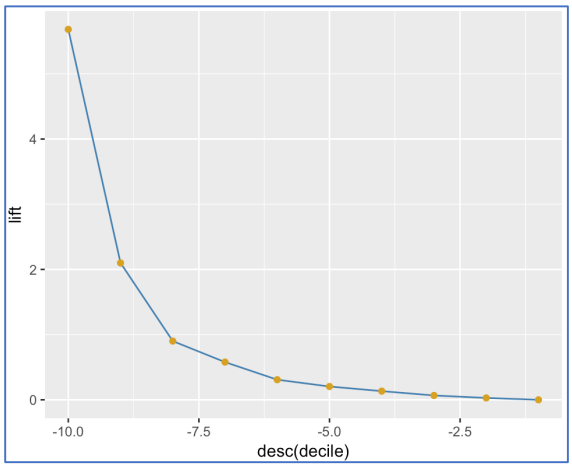
llh	llhNull	G2	McFadden	r2ML	r2CU
-2.549086e+04	-3.455015e+04	1.811857e+04	2.622068e-01	9.077532e-02	2.982442e-01

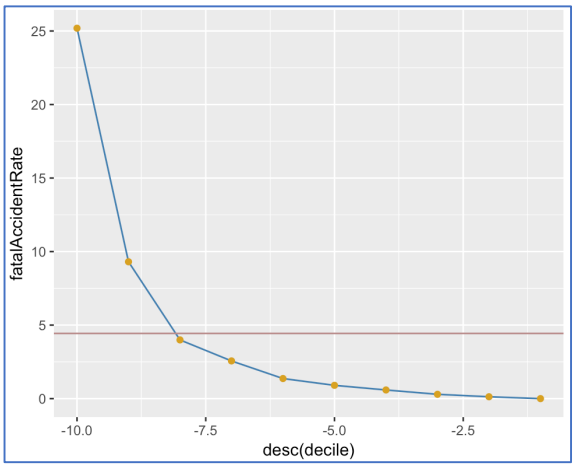
Classification Accuracy on the training sample

The Classification Accuracy is 94.21%, Sensitivity is 35.45%, Specificity is 94.21%, AUC is 88.14%, Lift is 5.68, 2.10 at 10th & 9th decile respectively, 77.82% of the fatal accidents are captured in first two (9th & 10th) deciles.



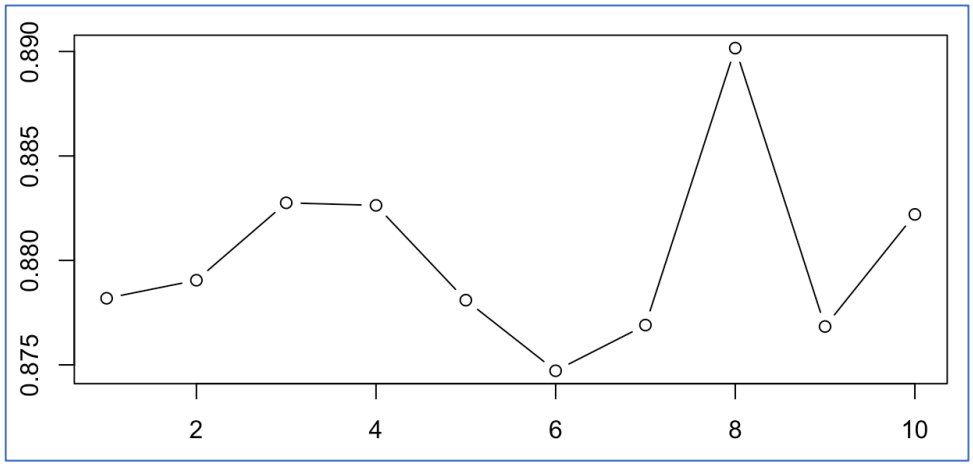

```
Confusion Matrix and Statistics

          Reference
Prediction      0      1
         0 176387   5448
         1   5568   2992

               Accuracy : 0.9421
                 95% CI : (0.9411, 0.9432)
    No Information Rate : 0.9557
    P-Value [Acc > NIR] : 1.0000

                  Kappa : 0.3217
 Mcnemar's Test P-Value : 0.2569

            Sensitivity : 0.35450
            Specificity : 0.96940
```





Decile	Lift	Fatal Accident Rate	Base Rate	Cumulative Fatal Rate
10	5.68	25.19	4.43	56.82
9	2.10	9.31	4.43	77.82
8	0.90	3.99	4.43	86.82
7	0.58	2.56	4.43	92.61
6	0.31	1.37	4.43	95.69
5	0.20	0.90	4.43	97.73
4	0.13	0.59	4.43	99.05
3	0.07	0.29	4.43	99.72
2	0.03	0.13	4.43	100.00
1	0.00	0.00	4.43	100.00

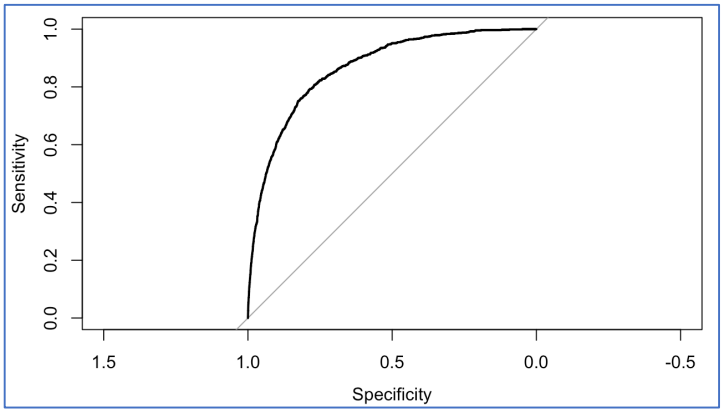
The AUC score for a 10-fold cross validation varies from 87.47% to 89.01% which indicates that the logistics model is robust.



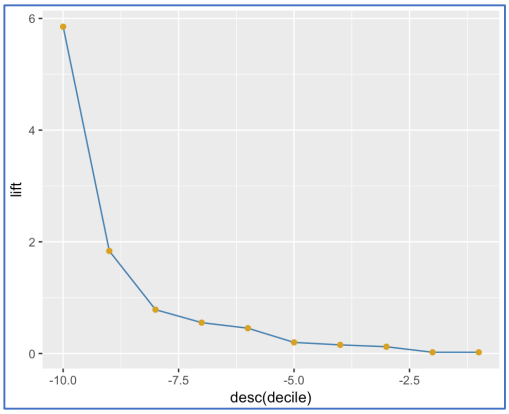
Classification Accuracy on the test sample

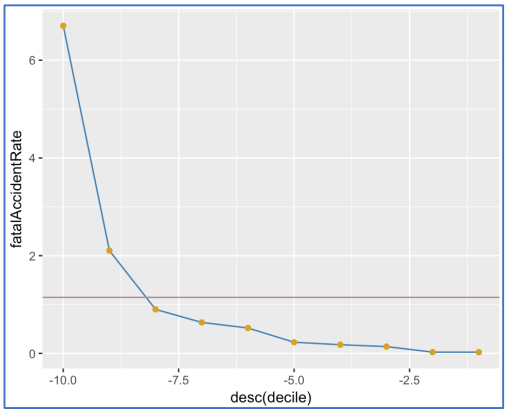
The Classification Accuracy is 96.23%, Sensitivity is 32.96%, Specificity is 96.96%, AUC is 86.57%, Lift is 5.85, 1.84 at 10th & 9th decile respectively, 76.88% of the fatal accidents are captured in first two (9th & 10th) deciles.

The overall model has reasonable accuracy and is robust.




```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 75610   606
         1  2371   298

               Accuracy : 0.9623
                 95% CI : (0.9609, 0.9636)
    No Information Rate : 0.9885
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1523
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.329646
            Specificity : 0.969595
```





Decile	Lift	Fatal Accident Rate	Base Rate	Cumulative Fatal Rate
10	5.85	6.71	1.15	58.52
9	1.84	2.10	1.15	76.88
8	0.79	0.90	1.15	84.73
7	0.55	0.63	1.15	90.27
6	0.45	0.52	1.15	94.80
5	0.20	0.23	1.15	96.79
4	0.15	0.18	1.15	98.34
3	0.12	0.14	1.15	99.56
2	0.02	0.03	1.15	99.78
1	0.02	0.03	1.15	100.00

Top 25 variables that influence fatal accidents

Variables	Estimate	Sig.	Odds	Interpretation
no_mobilityScooter	2.110291	***	8.2506	If the no. of mobility scooter increases by 1 then the odds to accident fatality increases by 8.25 times
no_of_casualties_in_ageband_11	1.694031	***	5.4414	If the no of occupants of age over 75 increases by 1 then the odds to accident fatality increases by 5.4
no_of_ped_inCarriagewayNotCrossing	1.526342	***	4.6013	If the pedestrian is in carriageway and are not crossing increases by 1 then accident fatality increases by 4.6
no_motocyle_over_500cc	1.435177	***	4.2004	If the no of motorcycles greater than 500cc increases by 1 then the odds to accident fatality increases by 4.2 times
no_of_ped_onZigZagApproach	1.28298	***	3.6074	If the pedestrian approaching Zigzag increases by 1 then the odds to accident fatal increases by 3.6 times
no_of_ped_crossFromOffSide	1.216794	***	3.3763	If the pedestrian crossing from Offside increases by 1 then the odds to accident fatal increases by 3.37 times
no_goodsVehicleOver7.5t	1.175373	***	3.2394	If the no. of goods vehicles of length over 7.5feet increases by 1, the odds of accident being fatal increases by 3.23 times
no_minibus	1.091722	***	2.9794	If the no. of minibuses increases by 1, the odds of accident being fatal increases by 2.98 times
no_motocyle_under_125cc	1.069278	***	2.9133	If the no. of motorcycles less than 125cc increases by 1 then the odds to accident fatality increases by 2.91 times
no_motocyle_under_500cc	1.064538	***	2.8995	If the no. of motorcycles less than 500cc increases by 1 then the odds to accident fatality increases by 2.89 times
no_agrVehicle	0.937071	***	2.5525	If the no. of agricultural vehicles increases by 1 then the odds to accident fatality increases by 2.55 times
no_of_ped_crossFromNearside	0.93512	***	2.5475	If the pedestrian crossing from nearside increases by 1 then the odds to accident fatal increases by 2.54 times
no_of_cyclist	0.837026	***	2.3095	If the no. of cyclists increases by 1 then the odds to accident fatality increases by 2.31 times
no_goodsVehicleUnder7.5t	0.78946	***	2.2022	If the no. of goods vehicles of length under 7.5feet increases by 1, the odds of accident being fatal increases by 2.2 times
no_of_ped_onZigZagExit	0.771697	**	2.1634	If the pedestrian exiting a zigzag increases by 1 then the odds to accident fatal increases by 2.16 times
no_of_ped_inCenterOfCarriageway	0.747917	***	2.1126	If the pedestrian in centre of carriageway increases by 1 then the odds to accident fatal increases by 2.11 times
no_of_ped_onRefuge	0.739848	***	2.0956	If the pedestrian is a refugee increases by 1 then accident fatality increases by 2.09

no_of_casualties_in_ageband_10	0.712718	***	2.0395	If the no. of people aged between 66 - 75 involved in casualties increases by 1, the odds of accident being fatal increases by 2.03 times
no_otherVehicle	0.690873	***	1.9955	If the no. of other vehicles increases by 1, then the odds of accident being fatal increases by 1.99 times

Factors influencing serious accidents

The logistic model is **overall significant** as indicated by the likelihood ratio test.

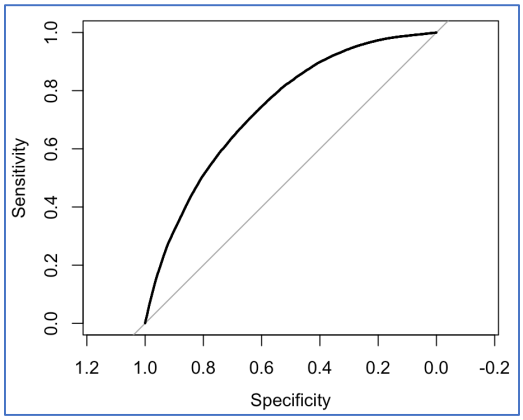
```
  #Df  LogLik   Df Chisq Pr(>Chisq)
1 176 -130577
2   1 -149073 -175 36992  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The **McFadden R^2 is 0.1240739** which indicates that 12.41% of the uncertainty of the intercept only model can be explained by the logistics regression model. A higher ratio (20% to 40 %) would be preferable but we can proceed at it indicates reasonable performance.

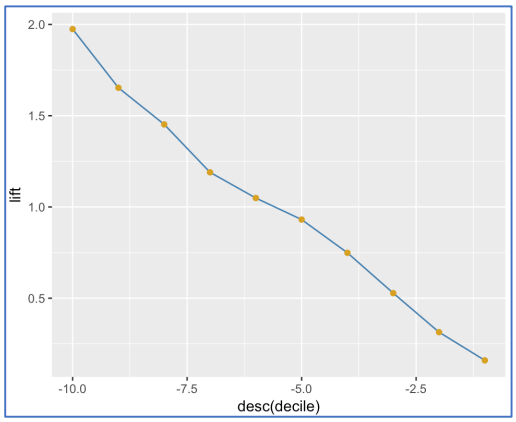
llh	llhNull	G2	McFadden	r2ML	r2CU
-1.305769e+05	-1.490730e+05	3.699214e+04	1.240739e-01	1.453151e-01	2.024120e-01

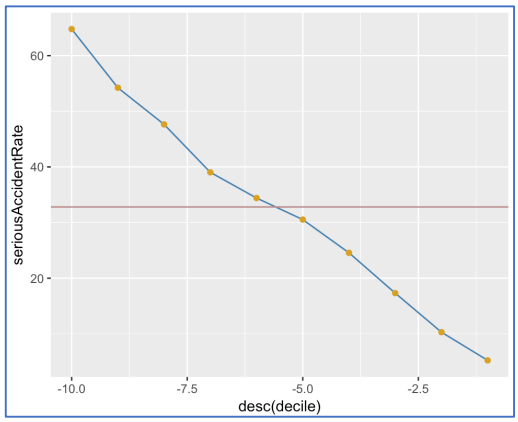
Classification Accuracy on the training sample

The Classification Accuracy is 70.08%, Sensitivity is 53.63%, Specificity is 78.11%, AUC is 73.88%, Lift is 1.97, 1.65 at 10th & 9th decile respectively, 73.19% of the serious accidents are captured in first five (10 to 6) deciles.



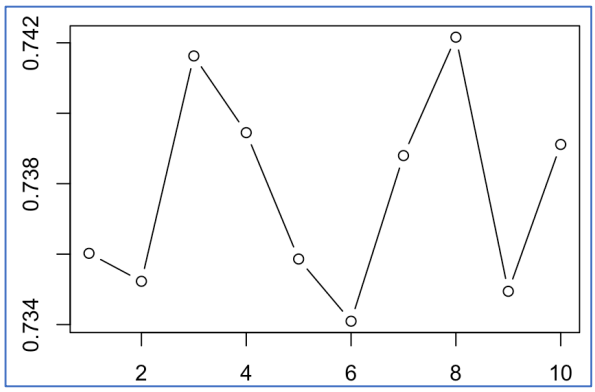

```
Confusion Matrix and Statistics

          Reference
Prediction      0      1
         0 123653  35832
         1  34652  41448

               Accuracy : 0.7008
                 95% CI : (0.699, 0.7027)
    No Information Rate : 0.672
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3187
 Mcnemar's Test P-Value : 8.959e-06

            Sensitivity : 0.5363
            Specificity : 0.7811
```





Decile	Lift	Serious Accident Rate	Base Rate	Cumulative Serious Rate
10	1.97	64.77	32.80	19.75
9	1.65	54.23	32.80	36.28
8	1.45	47.64	32.80	50.80
7	1.19	39.03	32.80	62.70
6	1.05	34.41	32.80	73.19
5	0.93	30.54	32.80	82.50
4	0.75	24.56	32.80	89.98
3	0.53	17.32	32.80	95.26
2	0.31	10.30	32.80	98.40
1	0.16	5.23	32.80	100.00

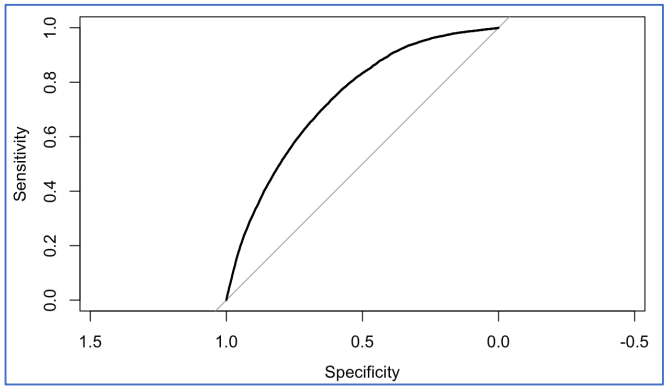
The AUC score for a 10-fold cross validation varies from 73.60% to 73.91% which indicates that the logistics model is robust.



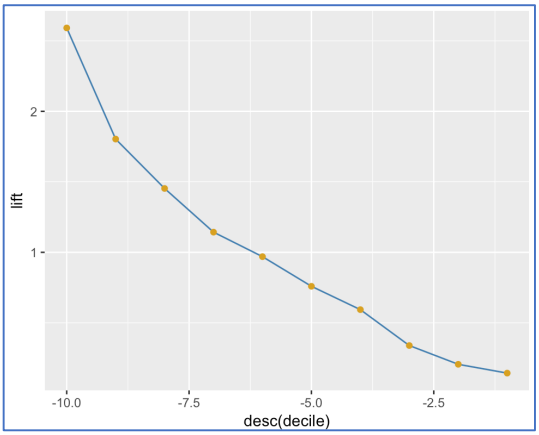
Classification Accuracy on the test sample

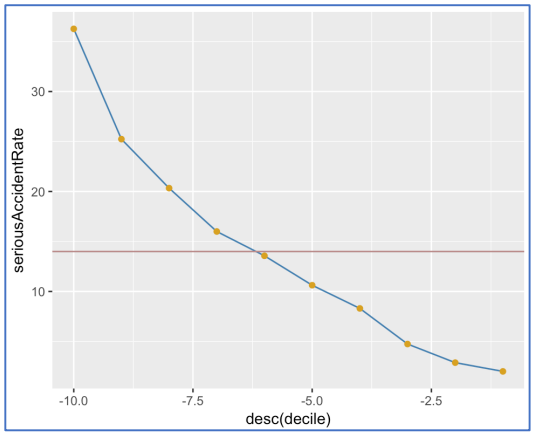
The Classification Accuracy is 74.75%, Sensitivity is 53.26%, Specificity is 78.25%, AUC is 73.91%, Lift is 2.59, 1.80 at 10th & 9th decile respectively, 79.60% of the serious accidents are captured in first five (10 to 6) deciles.

The overall model has reasonable accuracy and is robust.




```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 53089  5160
         1 14756  5880

               Accuracy : 0.7475
                 95% CI : (0.7445, 0.7506)
    No Information Rate : 0.86
    P-Value [Acc > NIR] : 1

                  Kappa : 0.231
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.53261
            Specificity : 0.78250
```





Decile	Lift	Serious Accident Rate	Base Rate	Cumulative Serious Rate
10	2.59	36.27	14.00	25.91
9	1.80	25.24	14.00	43.95
8	1.45	20.33	14.00	58.48
7	1.14	16.00	14.00	69.91
6	0.97	13.56	14.00	79.60
5	0.76	10.62	14.00	87.19
4	0.59	8.30	14.00	93.13
3	0.34	4.74	14.00	96.51
2	0.21	2.88	14.00	98.57
1	0.14	2.00	14.00	100.00

Top 25 variables that influence serious accidents

Variables	Estimate	Sig.	Odds	Interpretation
no_eMotorCycle	2.075103	***	7.9653689	If the no. of electric motorcycles increases by 1 then the odds to accident being serious increases by 7.96 times
no_motocyle_over_500cc	1.743162	***	5.7153885	If the no. of motorcycles over 500cc increases by 1 then the odds to accident being serious increases by 5.71 times
no_of_ped_crossFromOffSideMasked	1.431438	***	4.184714	If the pedestrian crosses from off side is masked increases by 1 then the odds to accident being severe increases by 4.18 times
no_motocyle_under_500cc	1.356186	***	3.8813619	If the no. of motorcycles below 500cc increases by 1 then the odds to accident being serious increases by 3.88 times
no_of_ped_crossFromNearsideMasked	1.342527	***	3.8287062	If the pedestrian crosses from nearside is masked increases by 1 then the odds to accident being severe increases by 3.82 times
no_of_ped_crossFromOffSide	1.161313	***	3.1941248	If the pedestrian crosses from off side increases by 1 then the odds to accident being severe increases by 3.19 times
no_motocyle_under_125cc	1.16022	***	3.190636	If the no. of motorcycles below 125cc increases by 1 then the odds to accident being serious increases by 3.19 times
no_of_cyclist	1.123112	***	3.074408	If the no. of cyclists increases by 1 then the odds to accident being serious increases by 1.12 times
no_of_ped_crossFromNearside	1.105508	***	3.0207574	If the pedestrian crosses from nearside increases by 1 then the odds to accident being severe increases by 3.02 times
no_of_casualties_in_ageband_11	0.966138	***	2.6277756	If the no of occupants of age over 75 increases by 1 then the odds to accident being serious increases by 2.6
no_of_ped_onZigZagExit	0.901909	***	2.4643038	If the pedestrian exiting a zigzag increases by 1 then the odds to accident being serious increases by 2.46 times
no_motocyle_under_50cc	0.884892	***	2.422723	If the no. of motorcycles below 50cc increases by 1 then the odds to accident being serious increases by 2.42 times
no_of_ped_inCarriagewayNotCrossing	0.801709	***	2.2293487	If the pedestrian in carriageway and are not crossing increases by 1 then the odds to accident being severe increases by 2.22 times
no_mobilityScooter	0.801103	***	2.227997	If the no. of mobilityscooter increases by 1 then the odds to accident being serious increases by 2.22 times
no_of_agrivehicle_occupant	0.720905	***	2.0562941	If the no. of agricultural vehicles increases by 1 then the odds to accident being serious increases by 2.05 times
no_of_ped_onFootway	0.661694	***	1.9380721	If the pedestrian on the footway increases by 1 then accident being serious increases by 1.93 times
no_of_casualties_in_ageband_10	0.656427	***	1.9278917	If the no. of people aged between 66 - 75 involved in casualties increases by 1, the odds of accident being severe increases by 1.92 times

no_of_ped_onZigZagApproach	0.603592	***	1.8286749	If the pedestrian approaching Zigzag increases by 1 then the odds to accident severity increases by 1.75 times
no_of_casualties_in_ageband_9	0.466107	***	1.5937778	If the no. of people aged between 56 - 65 involved in casualties increases by 1, the odds of accident being severe increases by 1.59 times
no_of_otherveh_occupant	0.359316	***	1.4323489	If other vehicle occupants increases by 1, the odds of accident being severe increases by 1.43 times

Summary & Recommendations

Fatal Accidents are more likely in the following situations

- Travelling in mobility scooters
- People over 65 years of age
- Pedestrians
    - on carriage way and not crossing
    - on carriageway and crossing
    - crossing before or after a zig zag junction
- Travelling on motor cycles
- Travelling in darkness with no street/carriageway with no lighting







Fatal Accidents can be reduced by the following measures

- Do's & Don't for those using mobility scooters and offer a 20-minute course to ensure safety
- Advice pedestrians, passengers on mobility scooters, cyclists to dress appropriately to be seen
    - Brightly colored clothing makes it easier for drivers to see you during the daytime
    - At night, wear reflective material on your shoes, cap, and jacket or on your arms or legs that bounce the car's lights off you and back to the driver. This light cautions the driver that there's something or someone moving and they should slow down and be careful
- Ensure that all streets & carriageways are lit-up when the sun sets
- Run an awareness drive for pedestrians once in a year – Walk for a cause and walk safely

Serious Accidents are more likely in the following situations

- Commuters in Electric Motor Cycles
- Motor Cycle Riders
    - Heavy motorcycle users (>500cc)
    - Middleweight motorcycle users (>300cc)
- Pedestrians
    - crossing from onside or offside irrespective of being masked or unmasked
    - exiting a zigzag
- Cyclists

Serious Accidents can be reduced by the following measures

- As per UK regulations, Electric motorbike speed is restricted to 15mph, has a motor of 250 watts output or less, and has pedals, and can be ridden by anyone aged 14 or over. They require no registration documents, number plates, tax disc or MOT. The rider does not need to be in possession of a licence. The current system should be changed.
    - Since now more powerful electric vehicles are in the market, proper registration and licensing mechanism should be in place for electric motorcycles as well
    - The e-motorcycle users are mostly of younger age (middle school to high school goers, college students) mostly less than 25 yrs. Proper rider training, road manners awareness campaigns should be mandatorily conducted at school and college levels.
    - Usage of proper ride gear, Practice of first aid treatment should be promoted.
- For usage of Heavy-weight Motorbikes (>500cc) followed by Middle-weight motorbikes (125cc – 500cc)
    - Issuance of Motorcycle Licenses should be stringent. Proper rider training should be given before issuing of such licenses.
    - Users riding without rider gear (not limited to helmets), should be considered as violating traffic rules and issued tickets.
    - Campaigns promoting best road practices, safety gear usages should be conducted frequently. These campaigns should be promoted through different rider communities thus ensuring to target the correct type of road users.
- Ensure that all streets & carriageways are lit-up when the sun sets
- Run an awareness drive for pedestrians once in a year – Walk for a cause and walk safely.
- Cyclists involve a large set of daily commuters in the UK.
    - Cyclists should be strictly advised to use only bike lanes. Violators should be penalized.
    - Usage of proper rider gear and first aid kit should be promoted among cyclists.

Future Enhancements

The study can be enhanced by addressing the topic in depth and considering the following aspects:

1. Different models for different casualty types (where pedestrians are involved, cyclists are involved, motor cyclist are involved). This would further help in identifying patters specific to the casualty type.

2. Model to consider interaction between variables. This would help in identifying hidden patterns on how a combination of variables make an accident fatal or serious.

References

- Government of UK: https://data.gov.uk/dataset/road-accidents-safety-data
- Kaggle: https://www.kaggle.com/silicon99/dft-accident-data
- Centre of Disease Control & Prevention: https://www.cdc.gov/features/globalroadsafety/index.html
- Everyone is a pedestrian: https://www.trafficsafetymarketing.gov/get-materials/pedestrian-safety/everyone-pedestrian
- Classification of electric motor cycles and rules in the UK: https://www.gov.uk/electric-bike-rules

GitHub Sources & Tableau Story

- UK Accident data set from Kaggle (raw data): https://drive.google.com/open?id=0B8J-8j4mM74ERTU1Zk51WTM1SjQ
- Data at different stages (processed data): https://drive.google.com/open?id=0B8J-8j4mM74EbUZDQlA2WVNQRlE
- GitHub: https://github.com/tkhiani/assignments/tree/master/captstone
- Tableau Dashboard: Insights on Accidents in the UK from 2010 – 2015: https://public.tableau.com/profile/tarun.khiani#!/vizhome/InsightsonAccidents/InsightsonAccidents