Accident Risk Index

Post Graduate Program in Data Science Engineering

Location: **Mumbai** Batch: **JAN'22**

Submitted by

Mardav Patel

Mrunal Lanjewar

Neha Gond

Shantanu Suryavanshi

Vishal Zanje

Mentored by

Mrs. Srikar Muppidi

Table of Contents

Abstract:

The data set is a records of accidents happened in United Kingdom. Its contains every details of accidents that happened in England, Scotland and Wales. The dataset contains 6,00,000 records and 27 variables. Following our the steps we did for building model.

1. Checking Outliers
2. Checking Correlation matrix and corresponding P values of T statistics from OLS Model .
3. Design a Model using Random Forest Regressor to visualize important features and remove insignificant variables.
4. Applying various Machine Learning Regression Models and evaluation of the model.
5. Choose the model with best accuracy score.

Introduction:

According to Semi Government Research Department (IBEF) "Domestic automobiles production increased at 2.36% Compound annual growth rate (CAGR) between FY16-20 with 26.36 million vehicles being manufactured in the country in FY20.Overall, domestic automobiles sales increased at 1.29% CAGR between FY16-FY20 with 21.55 million vehicles being sold in FY20". Rise in the number of vehicles on the road making road more prone to accident which lead to higher pay out to insurance firms. Government can also take advantage of this analysis report to control accident rate.

Domain and Feature Review:

Spike in sales of vehicles which lead to increase in number of accidents which eventually turn out as more payouts to insurance firms. Also Suggesting pre-emptive plan for insurance firms to control losses by finding risk of road accident to reduce insurance claim.

Data Set information:

The data set consists of 6 lakh rows divided into train and test data. It has 27 columns in total.

Problem Statement:

On the basis of target variable get the risk of accident to reduce insurance firm payouts.

Variable Categorization with description:

Numerical:

Police Force, Number of Vehicles, Number of Causalities, Day of week, Local Authority_(District), 1st_Road_Class, 1st_Road_Number, Speed Limit, 2nd_Road_Class, 2nd_Road_Number, Urban or Rural Area

Categorical:

Date, Time, Local Authority_Highway, Road_Type, Pedestrian_Crossing-Human_control, Pedestrian_Crossing-Physical Facilities, Light Conditions, Weather conditions, Road Surface Conditions, Special Conditions at site, Carriage Hazards, did Police officer attend scene of accident, state, postcode, country

Description of Features:

For Numerical:

- **Accident_ID**: Id of the accident
- **Police_Force**: Number of police force deployed at the accident scene
- **Number_of_Vehicles**: Number of vehicles involved in accident
- **Number_of_Casualties**: Deaths in the accidents
- **Day_of_Week**: Sunday,Monday,...etc
- **Local_Authority_(District)**: Cannot decide based on state. It should be category and not numerical
    - Can be differentiated based on England (east west..etc),Wales(east, west, etc..), Scotland(east, west, etc..)
- **Speed_limit**: Speed limit on the road

- **Urban_or_Rural_Area**: Type of area at which accident happened
- **1st_Road_Class**: Different types of roads
    - **1**: **Motorways -**
    - **3**: **A roads** – major roads intended to provide large-scale transport links within or between areas
    - **4**: **B roads** – roads intended to connect different areas, and to feed traffic between A roads and smaller roads on the network
    - **5**: **classified unnumbered** – smaller roads intended to connect together unclassified roads with A and B roads, and often linking a housing estate or a village to the rest of the network. Similar to 'minor roads' on an Ordnance Survey map and sometimes known unofficially as **C roads**
    - **6**: **unclassified** – local roads intended for local traffic. The vast majority (60%) of roads in the UK fall within this category
- **1st_Road_Number**: Number of road
- **2nd_Road_Class**: Types are same as 1st road class.
- **2nd_Road_Number**: Number of road

For categorical:

- **DATE**: date when accident took place
- **TIME**: There is 1368 unique time so it can be saperated into categories but will take a lot of time.
- **Local_Authority_(Highway)**: Local Authority that can enact traffic rules. (Different state wise)
    - **'EHEATHROW'**: Hearthrow is an airport in England.
    - **'W06000020'**: This is Torfaen is a county (Guessing similar to a state but not a state since there are three states given in the data) in Wales.
    - E in the data is for England and S is for Scotland
- **Road_Type**: Carriageway stands for highway. single carriageway is two lane highway. Roundabout is circle. Dual Carriageway is 4 lane highway. Slip Road is service road.
- **Pedestrian_Crossing-Human_Control**: A form of pedestrian crossing that gives priority to pedestrians or cycles crossing a road. Controlled crossings should be contrasted with a uncontrolled crossings, which do not give priority to pedestrians, and which typically take the form of subtle road markings, sometimes combined with a central refuge.
    - **'None within 50 metres'**: I am guessing none within 50 meters of accident took place.
    - **'Control by other authorised person'**: 'Control by school crossing patrol' Controlled by someone
- **Pedestrian_Crossing-Physical_Facilities**: Physical facilities available for pedestrain
    - **No physical crossing within 50 meters**: None within 50 meters of accident (guess)
    - **Zebra crossing**: zebra crossing
    - **Pedestrian phase at traffic signal junction**: Traffic signal for pedestrian
    - **non-junction pedestrian crossing**: A type of pedestrain crossing

    - **Central refuge**: A stop where pedestrain can stop before finishing to cross the road
    - **Footbridge or subway**: a bridge or underground tunnel for pedestrain
- **Light_Conditions**: Light condition during accident
    - **Daylight: Street light present**: daylight and street light also present
    - **Darkness: Street lighting unknown**: Unknown condition
    - **Darkness: Street lights present and lit**: Present and lit during accident
    - **Darkeness: No street lighting**: Absence of street light
    - **Darkness: Street lights present but unlit**: Not lit street light
- **Weather_Conditions**: Wheather conditions during accident
    - **Fine without high winds**:
    - **Raining without high winds**:
    - **Fine with high winds**:
    - **Raining with high winds**:
    - **Unknown**: Don't know
    - **Snowing without high winds**:
    - **Snowing with high winds**:
    - **Other**: Don't know
    - **Fog or mist**:
- **Carriageway_Hazards**: Different Hazards on the highways
    - **Any animal (except a ridden horse)**: Animal on road expect a horse
    - **Pedestrian in carriageway (not injured)**: A pedestrian on the road
    - **Dislodged vehicle load in carriageway**: Vehicle load on highway (what kind of load is debatable)
    - **Involvement with previous accident**: Not sure what it says
- **state**:
    - Alba / Scotland: Alba is another name for Scotland
    - Cymru / Wales: Cymru is another name for Wales

Target Variable:

Accident_Risk_Index (mean casualties at a postcode) = sum(Number_of_casualities)/count(Accident_ID)

With the help of the formula mentioned above we will derive the Target variable with the help of feature engineering.

Data Preprocessing:

Check for missing value treatment:

Time, Road Surface Conditions, special conditions at site has null values.

A] Time:

We have first converted the time on the basis of day and night and then took the mode and replace the value by it.

B] Road_Surface_condition:

- Due to 74% data contains Dry in the Road_Surface_Conditions naturally above dataframe shows dry more than other calss labels for weather_conditions variable
- This trend will be seen in other features as well due to 74% data has Dry class in road surface conditions. Hence, we will replace the null value by mode.

C] Special condition at site:

- It isn't necessary that there is going to be some condition on the road hence null value will be replaced with None.

Changing Data Types of wrongly classified features:

All the below variables were wrongly classified, so according to data we changed the data type of these variables to **object data types**

- Day_of_Week
- Local_Authority_(District)
- Urban_or_Rural_Area
- 1st_Road_Class
- 2nd_Road_Class
- 1st_Road_Number
- 2nd_Road_Number

Dropping of columns:

These variables do not explain the accident risk index much so we have not considered it, for building are machine learning models. The variables that are drop are stated below-

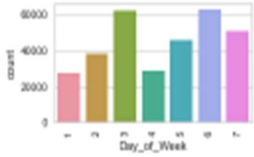
- Date
- Time
- Local_Authority_(Highway)
- Postcode
- 2nd Road_Number

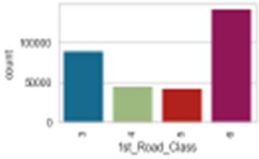
Removing Outliers:

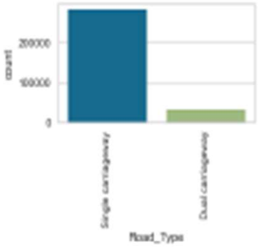
Since the dataset is of insurance company, we cannot remove the outliers because the extreme values also play important role in predictions. Hence removing the extreme values will not make our model more effective.

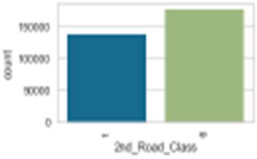
EDA:

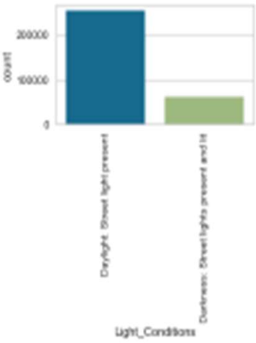
Categorical univariate analysis:

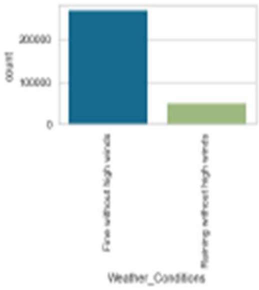


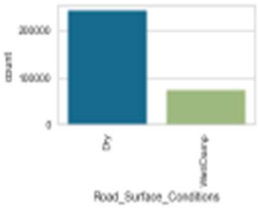


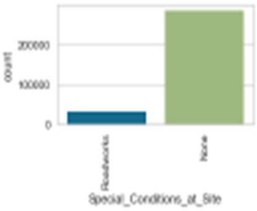


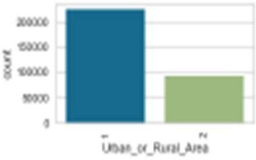


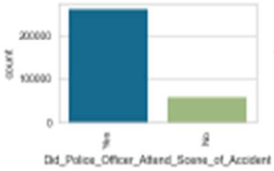


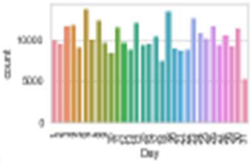


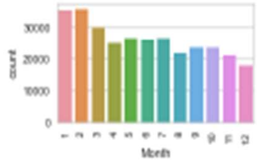


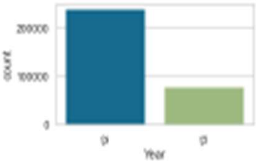


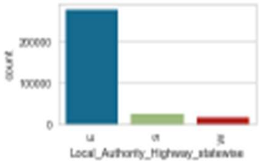








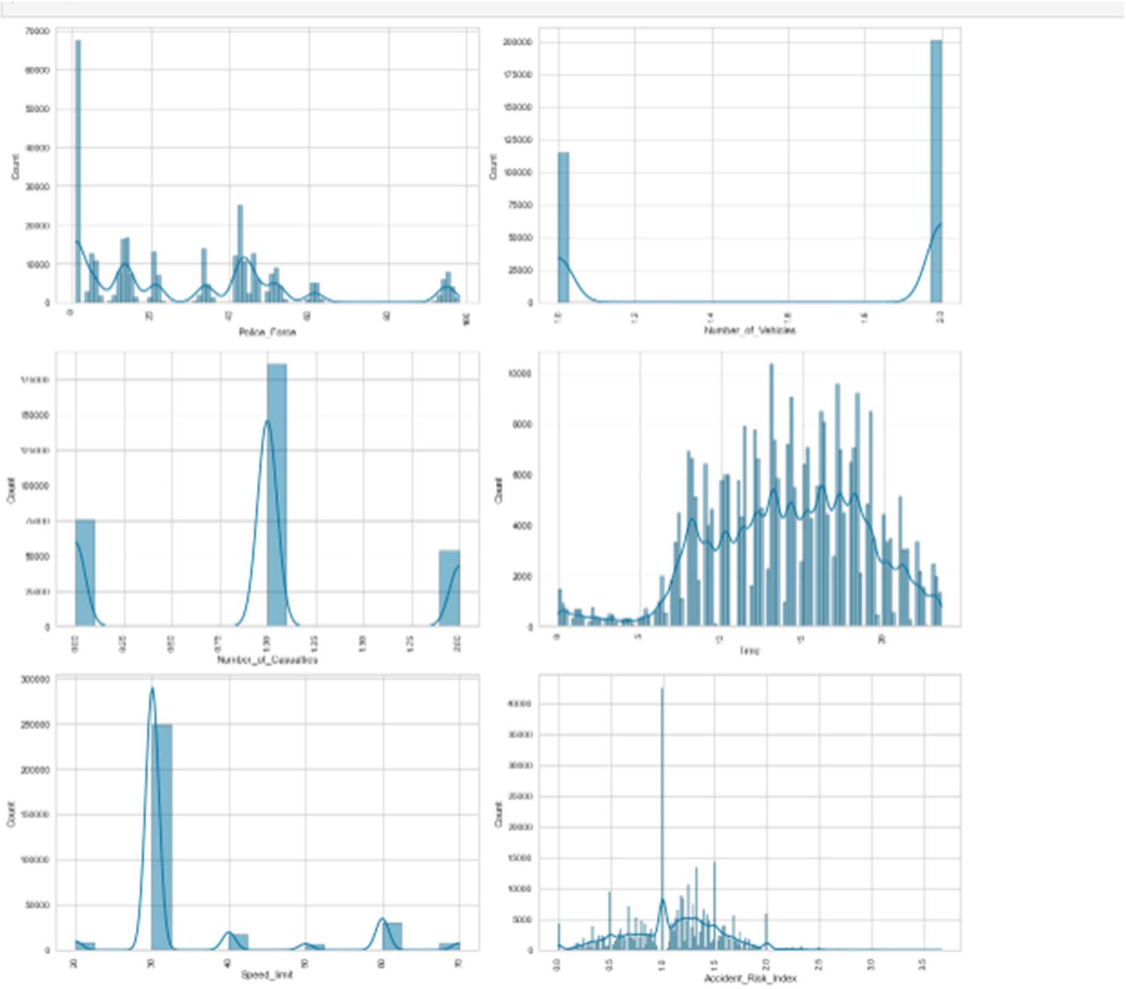




Categorical univariate analysis:

- Day_of_Week: 6th and 3rd day of week seem more frequently occuring than other classes. 4th day is the least occuring
- 1st_Road_Class: 6 road class is seen more frequently than other classes followed by 3 road class. 4 is slightly more than 5 road number.
- Road Type:Single Carriageway is seen most in the data. Presence of Dual carriageway is very less compared to the former type of carriageway
- 2nd_Road_Class: 6 road calss is seen more than 1.
- Light_Conditions: Daylight:Street light present is seen more than Darkness: Street lights present and lit.
- Weather_Conditions: Fine without high winds is more frequent than Raining wighout high winds.
- Road_Surface_Conditions: Dry condition is seem to be occuring more than other condition
- Special_Conditions_at_Site: None is seen more frequent.
- Urban_or_Rural_Area: Urban area is seen more frequently than Rural area.
- Did_Police_Officer_Attend_Scene_of_Accident: During most accidents police officer attended the scene of accident.
- Day: 6TH date and 19th dates are seen more than other dates. 6th date seen slightly more than 19th date. 31st is the least seen date here since not all months have 31st.
- Month:1st and 2nd months are seen more frequent than other months. 12th month seen least in the data
- Year: Year 2012's data is provided more than the year 2013
- Local_AUTHORITY_Highway_statewise: There are more authority data is available form England. Scotland and Wales are seen quite less in comparison to England

Numerical univariate Analysis:



- Police_Force: There are 0 Police Force seen in lot of districts. Between 45 and 55 follows the former one. There are up to 86 police force availble for few districts.

- Number_of_Vehicles: The data has 2 number of vehicles involved in the accident more than 1.

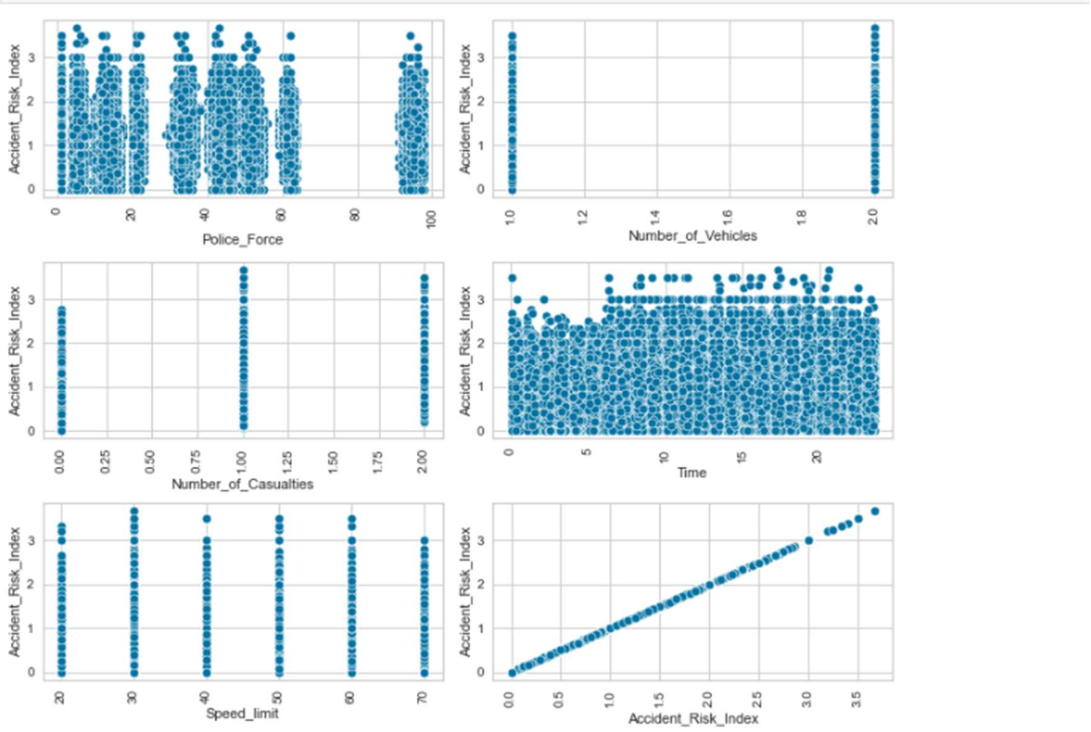
- Number of casualties: There are 1 number of casualty occuring more than other followed by 0. And 2 number of casulaty occuring least.

- Speed_limit: 30 speed limit is seen more frequent than other speed limit. Followed by 60 than 40. Least seen speed_limit is 70.

- Accident_Risk_Index: The target is normal and no need to transform it. The target has seen 1 Risk index the most. The Range of target is from 0 to 2.5.

Bivariate Analysis:

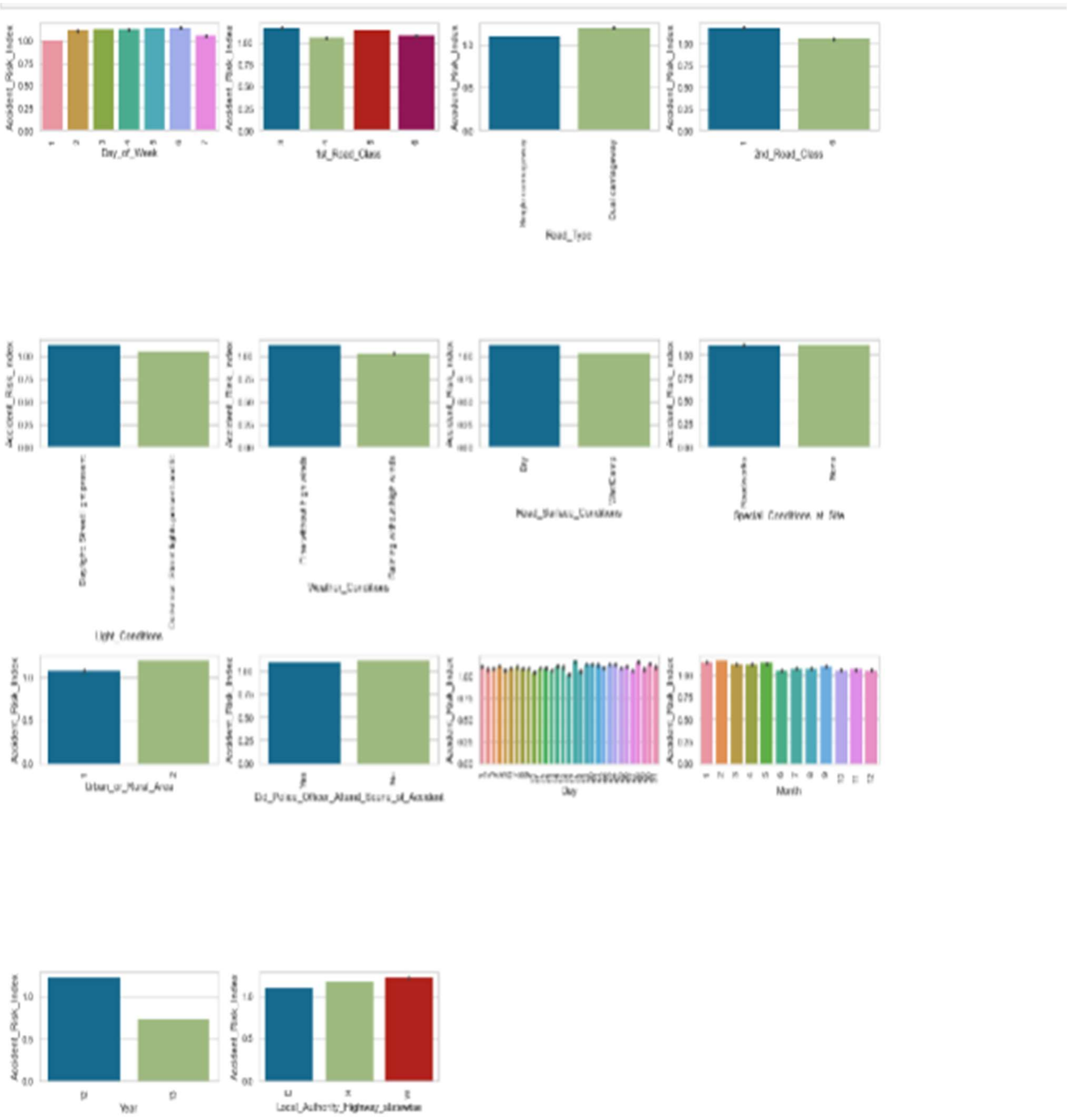
Numerical Bivariate Analysis: Numerical columns vs target(Numerical)



- Police force with Accident_Risk_Index tells clearly that there are clusters that are formed. We can see approx. 8 clusters from the plot
- Number of Casualities tells us that the casualities in the data set are of 3 different types i.e 0,1,2 respectively.
- Number of vehicles are of 2 types i.e 1 and 2
- Speed limit are of 20,30,40,50,60,70 respectively.
- In the time plot we cannot see the pattern bcoz the data is very noisy.

Categorical Bivariate Analysis:Categorical columns vs target(Numerical)

-



- Day_of_Week: All day of week seem to have same accident risk indedx except for 1st and 7th where 1st being the lowest.

- 1st_Road_Class: The class 3rd has more risk index than othe classes followed by 5th class which is slightly lower than calss 3. 6th class is slightly less than 4th class

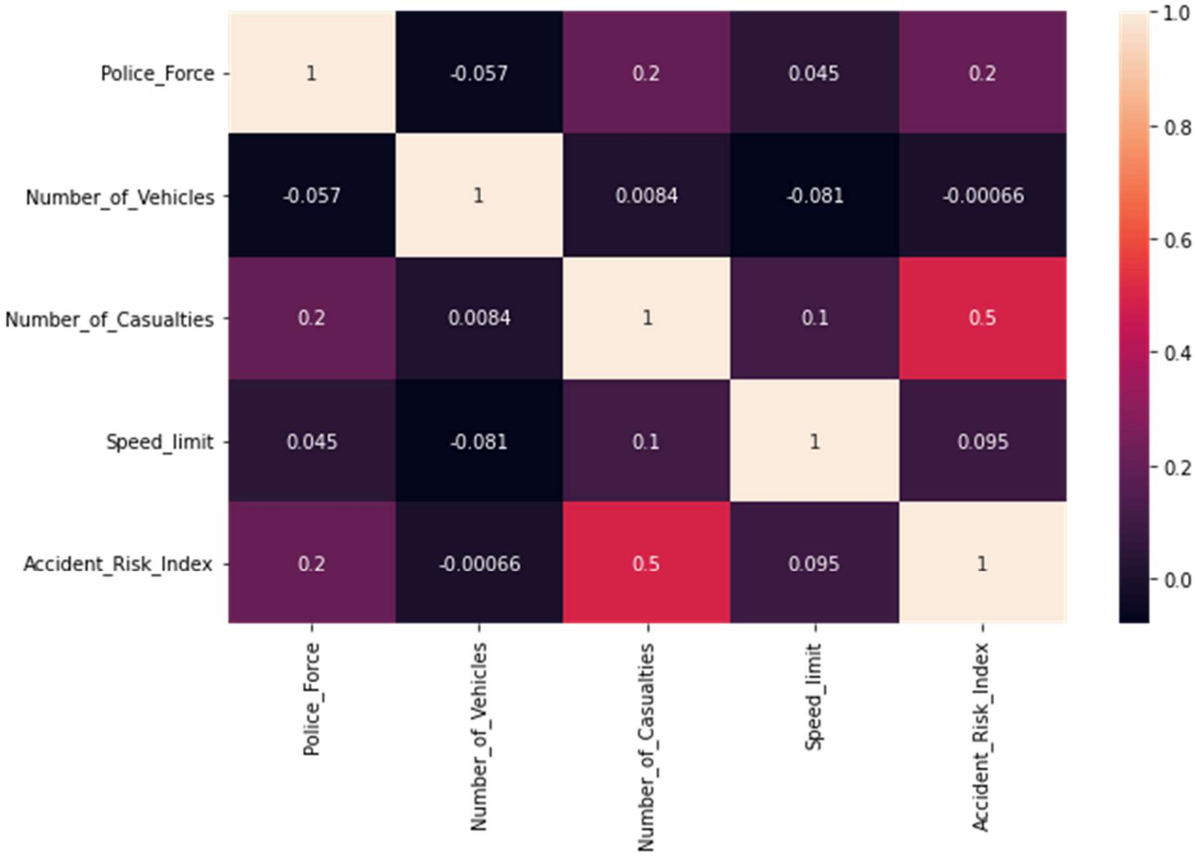
- Road_Type: Dual carriageway has higher average risk index than single carriageway
- 2nd_Road_Class: 6th road class has higher average risk index than 1st road calss
- Light_Conditions: Darkness has lower risk index than daylight
- Weather_Conditions: Fine without high winds has higer average accident risk index than raining without high winds.
- Road_Surface_Conditions: Dry condition has higher average risk index than Wet/Damp.
- Special_Conditions_at_Site: Both the conditions has almost equal average risk index.
- Urban_of_Rural_Area: Although Rural area is seen least in the data average risk index is higher for Rural area
- Day: All the day has almost identical average risk index. 17and 28th being slightly higher. Lowest average risk index is on the 15th.
- Month: From 6th month average risk index is almost identical. 2nd month is highest risk index which is slightly higher than 1st month.
- Year: 2012 has higher risk index than 2013
- Local_Authority_Highway_statewise: Wales has higher risk index than other two contries. Even though England's data is seen more average accident risk index is lowest in England

Multivariate Analysis:

After multivariate analysis the inference that can be stated are given below:-

- England's average accident risk index is lower for all time zones
- England seems to be safer place as compared to other two countries. Scotland and Wales seems almost equal.
- Urban area has higher average accident risk index.
- During night time street light is unlit has the higher average accident_risk_index
- In the month of December during christmas holidays average accident_risk index is lower
- There are other dates in each month has seen lower average accident risk index that can be summarized as other holidays

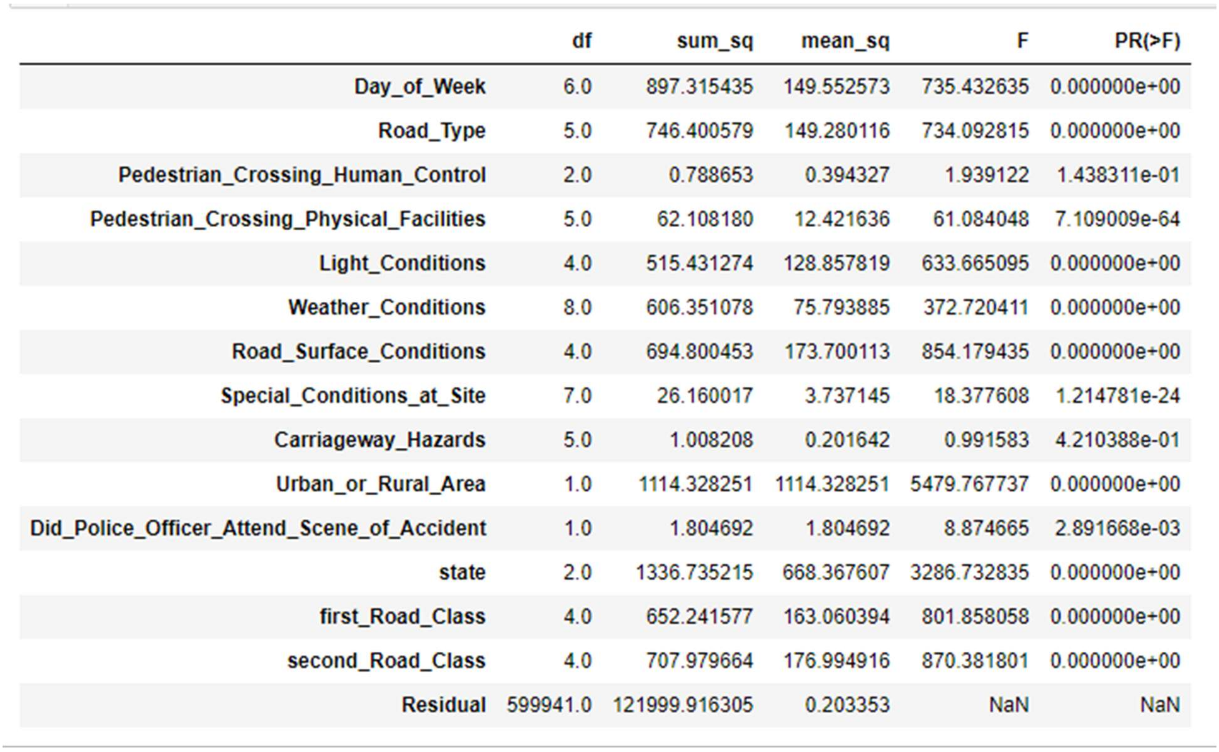
Numerical Feature correlation:



- From the correlation matrix of numerical features, we can say that they don’t have strong relationship with each other.
- Number of causalities and accident risk index is showing relationship but to say that it has strong relationship won’t be a good analysis.

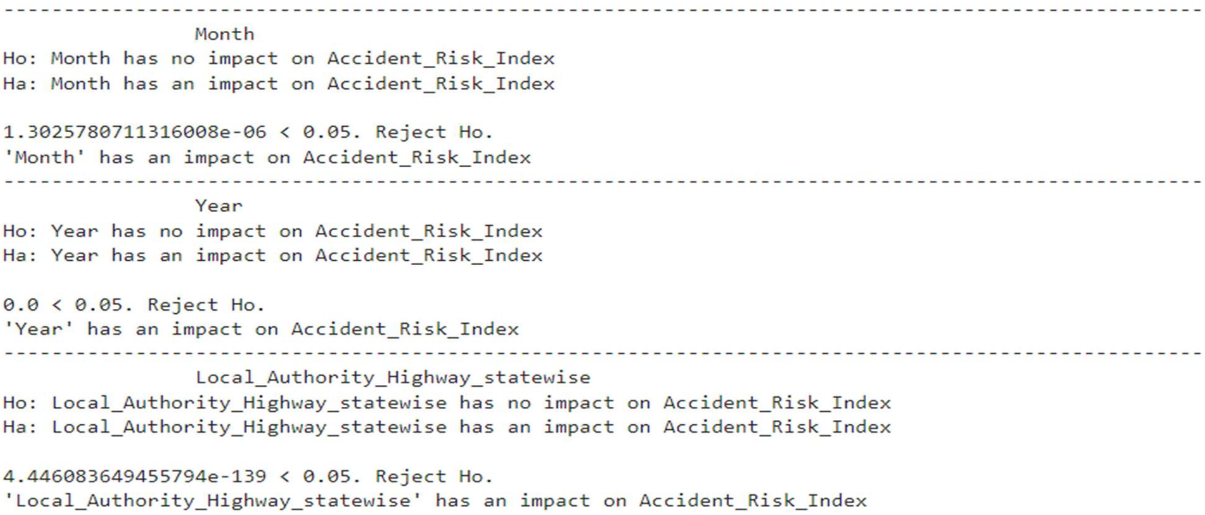
Statistical Analysis:

We used annova test to check the impact of categorical features on the target variable.



	df	sum_sq	mean_sq	F	PR(>F)
Day_of_Week	6.0	897.315435	149.552573	735.432635	0.000000e+00
Road_Type	5.0	746.400579	149.280116	734.092815	0.000000e+00
Pedestrian_Crossing_Human_Control	2.0	0.788653	0.394327	1.939122	1.438311e-01
Pedestrian_Crossing_Physical_Facilities	5.0	62.108180	12.421636	61.084048	7.109009e-64
Light_Conditions	4.0	515.431274	128.857819	633.665095	0.000000e+00
Weather_Conditions	8.0	606.351078	75.793885	372.720411	0.000000e+00
Road_Surface_Conditions	4.0	694.800453	173.700113	854.179435	0.000000e+00
Special_Conditions_at_Site	7.0	26.160017	3.737145	18.377608	1.214781e-24
Carriageway_Hazards	5.0	1.008208	0.201642	0.991583	4.210388e-01
Urban_or_Rural_Area	1.0	1114.328251	1114.328251	5479.767737	0.000000e+00
Did_Police_Officer_Attend_Scene_of_Accident	1.0	1.804692	1.804692	8.874665	2.891668e-03
state	2.0	1336.735215	668.367607	3286.732835	0.000000e+00
first_Road_Class	4.0	652.241577	163.060394	801.858058	0.000000e+00
second_Road_Class	4.0	707.979664	176.994916	870.381801	0.000000e+00
Residual	599941.0	121999.916305	0.203353	NaN	NaN

- From the above diagram we can say that Pedestrian_Crossing_Human_Control, Carriageway_Hazards are not impactful on the target according to the statistical analysis.

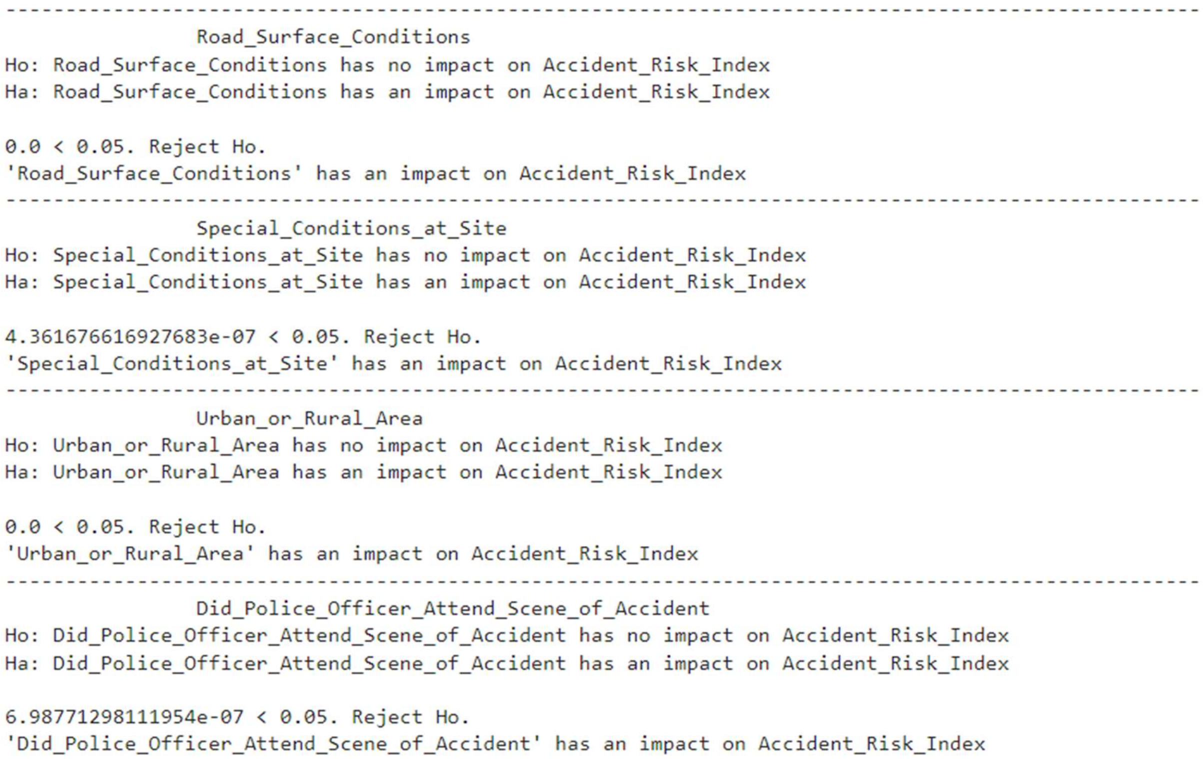


```
----------------------------------------------------------------------------------------------------
                Month
Ho: Month has no impact on Accident_Risk_Index
Ha: Month has an impact on Accident_Risk_Index

1.3025780711316008e-06 < 0.05. Reject Ho.
'Month' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Year
Ho: Year has no impact on Accident_Risk_Index
Ha: Year has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Year' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Local_Authority_Highway_statewise
Ho: Local_Authority_Highway_statewise has no impact on Accident_Risk_Index
Ha: Local_Authority_Highway_statewise has an impact on Accident_Risk_Index

4.446083649455794e-139 < 0.05. Reject Ho.
'Local_Authority_Highway_statewise' has an impact on Accident_Risk_Index
```



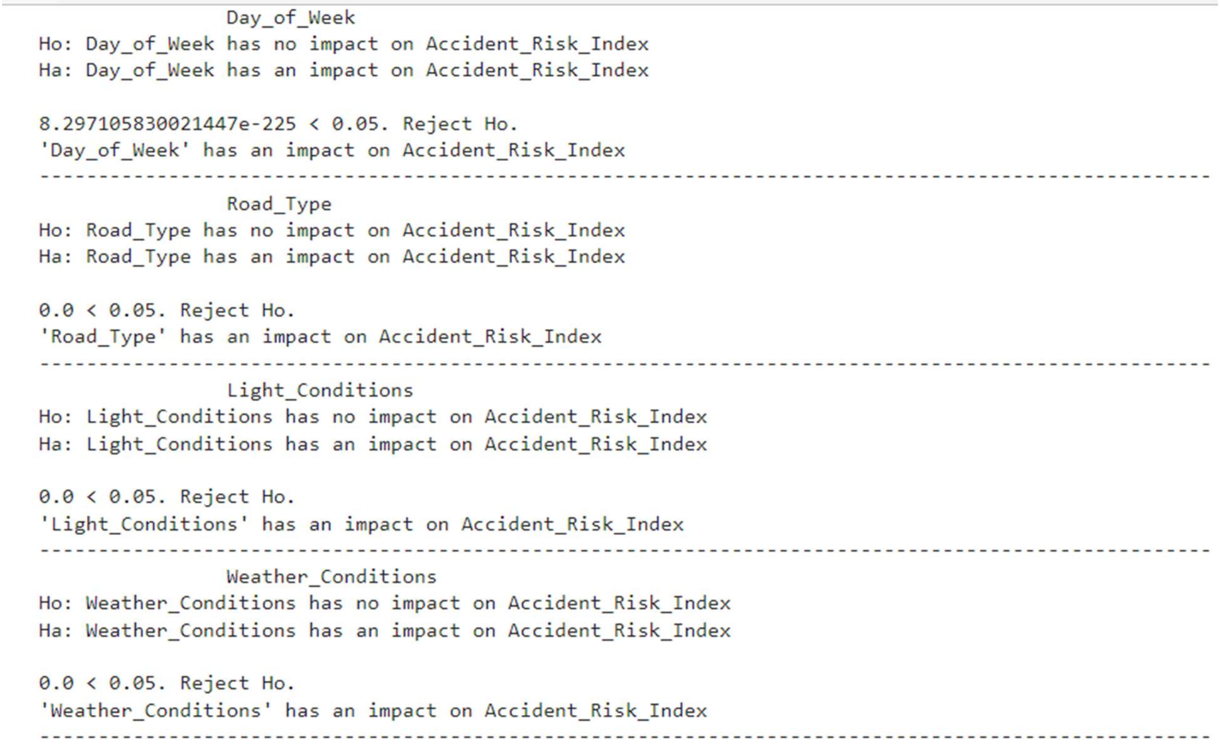
```
----------------------------------------------------------------------------------------------------
                Road_Surface_Conditions
Ho: Road_Surface_Conditions has no impact on Accident_Risk_Index
Ha: Road_Surface_Conditions has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Road_Surface_Conditions' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Special_Conditions_at_Site
Ho: Special_Conditions_at_Site has no impact on Accident_Risk_Index
Ha: Special_Conditions_at_Site has an impact on Accident_Risk_Index

4.361676616927683e-07 < 0.05. Reject Ho.
'Special_Conditions_at_Site' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Urban_or_Rural_Area
Ho: Urban_or_Rural_Area has no impact on Accident_Risk_Index
Ha: Urban_or_Rural_Area has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Urban_or_Rural_Area' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Did_Police_Officer_Attend_Scene_of_Accident
Ho: Did_Police_Officer_Attend_Scene_of_Accident has no impact on Accident_Risk_Index
Ha: Did_Police_Officer_Attend_Scene_of_Accident has an impact on Accident_Risk_Index

6.98771298111954e-07 < 0.05. Reject Ho.
'Did_Police_Officer_Attend_Scene_of_Accident' has an impact on Accident_Risk_Index
```



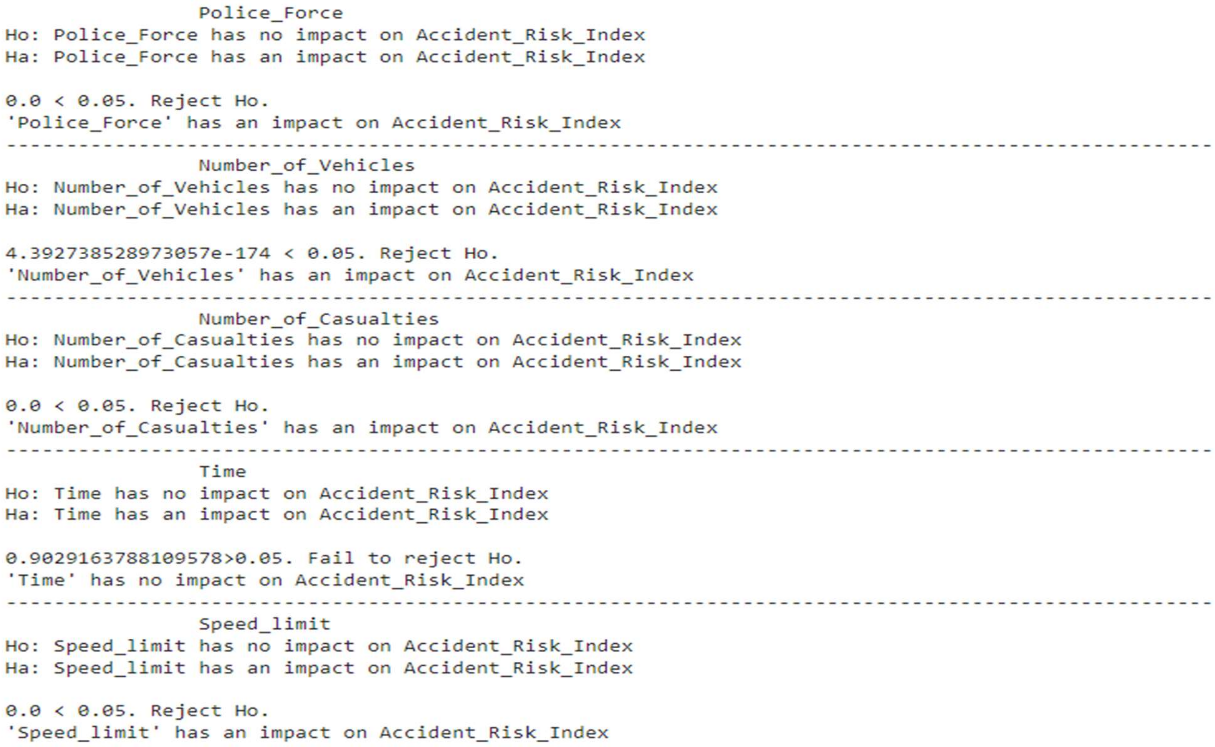
```
                Day_of_Week
Ho: Day_of_Week has no impact on Accident_Risk_Index
Ha: Day_of_Week has an impact on Accident_Risk_Index

8.297105830021447e-225 < 0.05. Reject Ho.
'Day_of_Week' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Road_Type
Ho: Road_Type has no impact on Accident_Risk_Index
Ha: Road_Type has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Road_Type' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Light_Conditions
Ho: Light_Conditions has no impact on Accident_Risk_Index
Ha: Light_Conditions has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Light_Conditions' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Weather_Conditions
Ho: Weather_Conditions has no impact on Accident_Risk_Index
Ha: Weather_Conditions has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Weather_Conditions' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
```



```
                Police_Force
Ho: Police_Force has no impact on Accident_Risk_Index
Ha: Police_Force has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Police_Force' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Number_of_Vehicles
Ho: Number_of_Vehicles has no impact on Accident_Risk_Index
Ha: Number_of_Vehicles has an impact on Accident_Risk_Index

4.392738528973057e-174 < 0.05. Reject Ho.
'Number_of_Vehicles' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Number_of_Casualties
Ho: Number_of_Casualties has no impact on Accident_Risk_Index
Ha: Number_of_Casualties has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Number_of_Casualties' has an impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Time
Ho: Time has no impact on Accident_Risk_Index
Ha: Time has an impact on Accident_Risk_Index

0.9029163788109578>0.05. Fail to reject Ho.
'Time' has no impact on Accident_Risk_Index
----------------------------------------------------------------------------------------------------
                Speed_limit
Ho: Speed_limit has no impact on Accident_Risk_Index
Ha: Speed_limit has an impact on Accident_Risk_Index

0.0 < 0.05. Reject Ho.
'Speed_limit' has an impact on Accident_Risk_Index
```

Variables that have impact on Accident Risk Index.

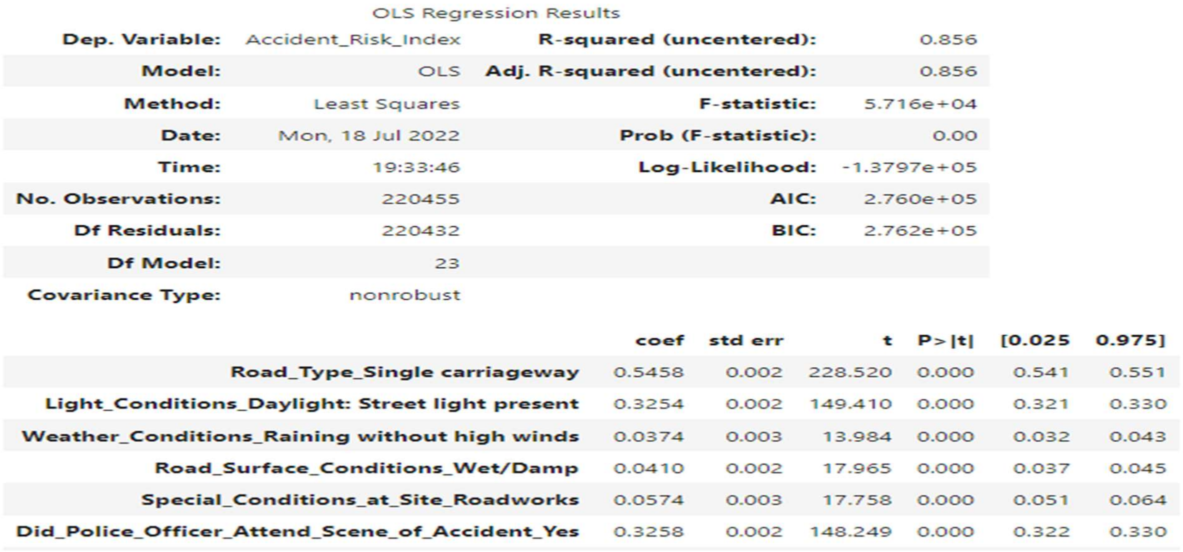
- Day_of_Week
- Road_Type
- Light_Conditions
- Weather_Conditions
- Road_Surface_Conditions
- Special_Conditions_at_Site
- Urban_or_Rural_Area
- Did_Police_Officer_Attend_Scene_of_Accident
- Month
- Year
- Local_Authority_Highway_statewise
- 1st_road_calss
- 2nd_road_calss
- Day_of_Week',
- Police_Force
- Number_of_Vehicles
- Number_of_Casualties
- Speed_limit
- Police_Force
- Number_of_Vehicles
- Number_of_Casualties
- Time
- Speed_limit
- Police_Force
- Number_of_Vehicles
- Number_of_Casualties
- Speed_limit

Variables has no impact on Accident Risk Index:

- Time

Base Model:

A] Linear Regression Model



OLS Regression Results

Dep. Variable:	Accident_Risk_Index	**R-squared (uncentered):**	0.856
Model:	OLS	**Adj. R-squared (uncentered):**	0.856
Method:	Least Squares	**F-statistic:**	5.716e+04
Date:	Mon, 18 Jul 2022	**Prob (F-statistic):**	0.00
Time:	19:33:46	**Log-Likelihood:**	-1.3797e+05
No. Observations:	220455	**AIC:**	2.760e+05
Df Residuals:	220432	**BIC:**	2.762e+05
Df Model:	23		
Covariance Type:	nonrobust		

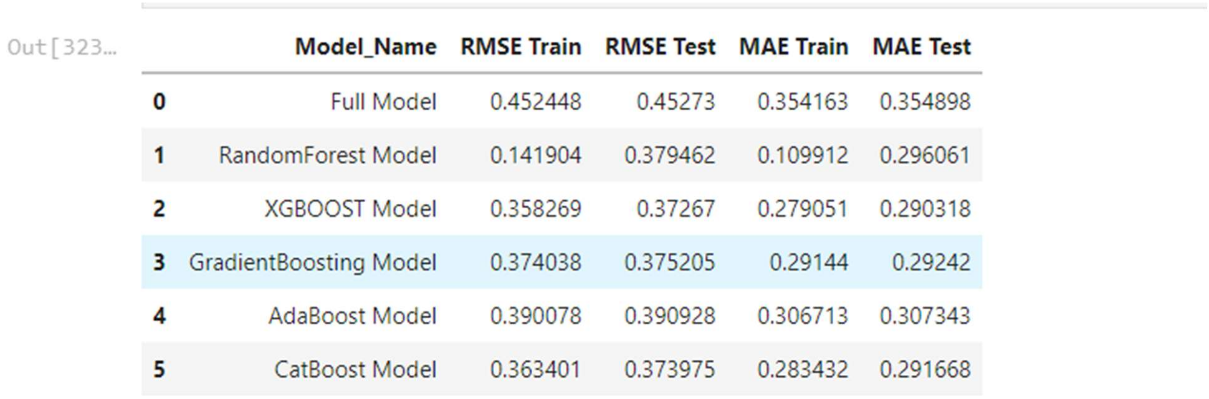
	coef	std err	t	P>|t|	[0.025	0.975]
Road_Type_Single carriageway	0.5458	0.002	228.520	0.000	0.541	0.551
Light_Conditions_Daylight: Street light present	0.3254	0.002	149.410	0.000	0.321	0.330
Weather_Conditions_Raining without high winds	0.0374	0.003	13.984	0.000	0.032	0.043
Road_Surface_Conditions_Wet/Damp	0.0410	0.002	17.965	0.000	0.037	0.045
Special_Conditions_at_Site_Roadworks	0.0574	0.003	17.758	0.000	0.051	0.064
Did_Police_Officer_Attend_Scene_of_Accident_Yes	0.3258	0.002	148.249	0.000	0.322	0.330

In statistics, ordinary least squares (OLS) is a type of linear least squares method for estimating the unknown parameters in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares, that is, minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being predicted) in the given dataset and those predicted by the linear function.

As we can see we have received an R-square value of 0.856 and the same for Adjusted R-square. A low R-squared value indicates that independent variable is not explaining much in the variation of dependent variable .the R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression

An R-Squared value of 0.856 would indicate that 85% of the variance of the dependent variable being studied is explained by the variance of the independent variable. Which means it is a good model but there is still scope to improve the model performance.

B] OTHER MODEL’S



Out[323…

	Model_Name	RMSE Train	RMSE Test	MAE Train	MAE Test
0	Full Model	0.452448	0.45273	0.354163	0.354898
1	RandomForest Model	0.141904	0.379462	0.109912	0.296061
2	XGBOOST Model	0.358269	0.37267	0.279051	0.290318
3	GradientBoosting Model	0.374038	0.375205	0.29144	0.29242
4	AdaBoost Model	0.390078	0.390928	0.306713	0.307343
5	CatBoost Model	0.363401	0.373975	0.283432	0.291668

One way to assess how well a regression model fits a dataset is to calculate the root mean square error, which is a metric that tells us the average distance between the predicted values from the model and the actual values in the dataset. The lower the RMSE, the better a given model is able to “fit” a dataset.
From the above output we can conclude that XGBOOST Model having the least RMSE for the test dataset. So after comparing the all the model XGBOOST model is the best model on the basis RMSE scores.

Future Work:

We will tune the parameters using XGBOOST and then with the help of best parameters we will generate a model using XGBOOST. We will also try a model withtune parameters known as random forest and many more to Improve the Rsquared and RMSE of our model. Following our the steps we would proceed to improve our model.

1. Outliers Treatment and data normalization if needed.

2. With the help of GridSearchCV or Baysian Search we should tune the parameters and find the best one which can improve our RMSE value.

3. We can also do data sampling with the help of stratified sampling method or statistical sampling to come up with good results.

4. Improve Model Benchmark using AutoML.

5. Interpretable Machine Learning technique such as SHAP/LIME/ELIS tools which helps to explain factors(features) influencing the model performance.

Appendix:

- https://machinehack.com/hackathons/predict_accident_risk_score_for_unique_postcode/data
- https://www.google.com/search?client=firefox-b-d&q=EHEATHROW
- https://www.cycling-embassy.org.uk/dictionary/controlled-crossing
- https://www.gov.uk/government/publications/guidance-on-road-classification-and-the-primary-route-network/guidance-on-road-classification-and-the-primary-route-net