2. LOGISTIC REGRESSION

Logistic Regression is famously utilized as a Classification Algorithm, categorized under the Supervised learning technique. This regression technique is utilized to anticipate the likelihood of the event of the observation values into one of the two categories of the dichotomous Dependent variable (i.e., two dependent values). Logistic regression analysis is a method to determine the result-reason relationship of independent variables with dependent variable.

OBJECTIVE:

The purpose of this analysis is to find the how well logistic regression method can predict whether or not an accident is serious or slight. (Accident severity)

DATASET:

This Road Accident dataset is obtained by merging two different datasets from a depositary of official UK government statistics, as various factors leading to accident. This dataset has information about Accidents and vehicles for last 5 years (2017 -2021) in UK region. road_surface_conditions, Speed Limit, number_of_casualties, vehicle_type & vehicle manoeuvre are considered as the independent variables in this analysis and the Accident Severity of the injury of the accident victim is considered as the Dependent variable.

Data pre-processing and transformation has been done with the help of python language. Datasets were downloaded in .csv format, which are then imported in python language. With help of panda's library values from accidents csv and vehicles csv were merged with inner join method with "**accident index"** as a primary key column. Null values were checked using "**Isnull (). sum ()"** function and later Null values were dropped using DROP () function.

Also based on correlation matrix features which are not correlated are dropped from the dataframe.

Link to Data Source:

Independent Variables:

day_of_week
junction_detail
number_of_vehicles
road_surface_conditions
special_conditions_at_site
time
weather_conditions
hit_object_in_carriageway
hit_object_off_carriageway
sex_of_driver
skidding_and_overturning
vehicle_manoeuvre
vehicle_type

Dependent Variable:

Accident Severity

ASSUMPTIONS:

1. **Dependent variable is dichotomous:**
   Here dependent variable is Accident Severity which is 1 Serious and 2 is Slight in terms of severity.
2. **Collinearity and Multi-Collinearity**
   This can be checked with correlation matrix:

		accident_index	first_road_class	day_of_week	junction_detail	light_conditions	number_of_casualties	number_of_vehicles	road_surface_condition	road_type	special_conditions_at_site	speed_limit	time	urban_or_rural_area	weather_conditions	age_band_of_driver	age_of_vehicle	hit_object_in_carriageway	hit_object_off_carriageway	engine_capacity_cc	sex_of_driver	skidding_and_overturning	vehicle_manoeuvre	vehicle_type	Acci_severity
Acci_severity	Pearson Correlation	0.000	.005*	.006**	.036**	-.038**	-.068**	.021**	.033**	-.033**	.036**	-.074**	0.000	-.080**	.028**	-.050**	-0.001	.044**	.033**	.007**	.054**	.006**	0.004	.006*	1
	Sig. (2-tailed)	1.000	0.028	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.902	0.000	0.000	0.000	0.688	0.000	0.000	0.002	0.000	0.008	0.087	0.011	
	N	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654

Correlations

		Acci_severity	accident_index	first_road_class	day_of_week	junction_detail	light_conditions	number_of_casualties	number_of_vehicles	road_surface_conditions	road_type	special_conditions_at_site
Acci_severity	Pearson Correlation	1	.000	.005*	.006**	.036**	-.038**	-.068**	.021**	.033**	-.033**	.036**
	Sig. (2-tailed)		1.000	.028	.004	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
accident_index	Pearson Correlation	.000	1	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Sig. (2-tailed)	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	N	160398	160398	160398	160398	160398	160398	160398	160398	160398	160398	160398
first_road_class	Pearson Correlation	.005*	.000	1	.009**	.045**	-.031**	-.111**	-.143**	.000	.297**	-.023**
	Sig. (2-tailed)	.028	1.000		<.001	<.001	<.001	.000	.000	.920	.000	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
day_of_week	Pearson Correlation	.006**	.000	.009**	1	.004	.009**	-.006*	-.001	.000	.003	-.001
	Sig. (2-tailed)	.004	1.000	<.001		.058	<.001	.013	.589	.887	.188	.639
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
junction_detail	Pearson Correlation	.036**	.000	.045**	.004	1	.000	-.033**	-.040**	.224**	.129**	.280**
	Sig. (2-tailed)	<.001	1.000	<.001	.058		.969	<.001	<.001	.000	.000	.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
light_conditions	Pearson Correlation	-.038**	.000	-.031**	.009**	.000	1	.034**	-.013**	.113**	-.015**	.015**
	Sig. (2-tailed)	<.001	1.000	<.001	<.001	.969		<.001	<.001	.000	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
number_of_casualties	Pearson Correlation	-.068**	.000	-.111**	-.006*	-.033**	.034**	1	.264**	-.025**	-.032**	-.036**
	Sig. (2-tailed)	<.001	1.000	.000	.013	<.001	<.001		.000	<.001	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
number_of_vehicles	Pearson Correlation	.021**	.000	-.143**	-.001	-.040**	-.013**	.264**	1	-.044**	-.078**	-.043**
	Sig. (2-tailed)	<.001	1.000	.000	.589	<.001	<.001	.000		<.001	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
road_surface_conditions	Pearson Correlation	.033**	.000	.000	.000	.224**	.113**	-.025**	-.044**	1	.113**	.510**
	Sig. (2-tailed)	<.001	1.000	.920	.887	.000	.000	<.001	<.001		.000	.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
road_type	Pearson Correlation	-.033**	.000	.297**	.003	.129**	-.015**	-.032**	-.078**	.113**	1	.120**
	Sig. (2-tailed)	<.001	1.000	.000	.188	.000	<.001	<.001	<.001	.000		.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
special_conditions_at_site	Pearson Correlation	.036**	.000	-.023**	-.001	.280**	.015**	-.036**	-.043**	.510**	.120**	1
	Sig. (2-tailed)	<.001	1.000	<.001	.639	.000	<.001	<.001	<.001	.000	.000	
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
speed_limit	Pearson Correlation	-.074**	.000	-.396**	-.018**	-.121**	.048**	.185**	.164**	-.013**	-.210**	-.037**
	Sig. (2-tailed)	<.001	1.000	.000	<.001	.000	<.001	.000	.000	<.001	.000	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
time	Pearson Correlation	.000	.000	.029**	.032**	.016**	.226**	.022**	-.014**	-.024**	.020**	-.001
	Sig. (2-tailed)	.902	1.000	<.001	<.001	<.001	.000	<.001	<.001	<.001	<.001	.589
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
urban_or_rural_area	Pearson Correlation	-.080**	.000	-.224**	-.016**	-.114**	.017**	.139**	.089**	-.017**	-.057**	-.041**
	Sig. (2-tailed)	<.001	1.000	.000	<.001	.000	<.001	.000	.000	<.001	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
weather_conditions	Pearson Correlation	.028**	.000	.019**	.001	.123**	.090**	-.029**	-.038**	.428**	.066**	.273**
	Sig. (2-tailed)	<.001	1.000	<.001	.535	.000	.000	<.001	<.001	.000	<.001	.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
age_band_of_driver	Pearson Correlation	-.050**	.000	-.092**	.001	-.051**	-.076**	.072**	-.019**	-.073**	-.046**	-.083**
	Sig. (2-tailed)	<.001	1.000	.000	.679	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
age_of_vehicle	Pearson Correlation	-.001	.000	.004	.000	-.016**	.027**	.065**	.006**	.001	.015**	-.020**
	Sig. (2-tailed)	.688	1.000	.105	.943	<.001	<.001	<.001	.009	.635	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
hit_object_in_carriageway	Pearson Correlation	.044**	.000	-.006*	.000	.312**	.023**	-.051**	-.055**	.430**	.124**	.559**
	Sig. (2-tailed)	<.001	1.000	.014	.828	.000	<.001	<.001	<.001	.000	.000	.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
hit_object_off_carriageway	Pearson Correlation	.033**	.000	-.007**	.000	.291**	.028**	-.042**	-.061**	.412**	.129**	.518**
	Sig. (2-tailed)	<.001	1.000	.002	.955	.000	<.001	<.001	<.001	.000	.000	.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
engine_capacity_cc	Pearson Correlation	.007**	.000	-.098**	.009**	-.023**	-.001	.078**	.037**	-.012**	-.052**	-.011**
	Sig. (2-tailed)	.002	1.000	.000	<.001	<.001	.674	<.001	<.001	<.001	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
sex_of_driver	Pearson Correlation	.054**	.000	.080**	-.003	.014**	-.007**	-.030**	.041**	.025**	.039**	.022**
	Sig. (2-tailed)	<.001	1.000	<.001	.240	<.001	.004	<.001	<.001	<.001	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
skidding_and_overturning	Pearson Correlation	.006**	.000	-.016**	-.005*	.228**	.041**	-.042**	-.088**	.387**	.100**	.454**
	Sig. (2-tailed)	.008	1.000	<.001	.024	.000	<.001	<.001	.000	.000	.000	.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
vehicle_manoeuvre	Pearson Correlation	.004	.000	-.012**	-.001	.250**	.038**	-.038**	-.104**	.339**	.100**	.401**
	Sig. (2-tailed)	.087	1.000	<.001	.566	.000	<.001	<.001	.000	.000	.000	.000
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654
vehicle_type	Pearson Correlation	.006*	.000	-.041**	.005*	-.019**	-.013**	.029**	.026**	-.007**	-.023**	-.009**
	Sig. (2-tailed)	.011	1.000	<.001	.032	<.001	<.001	<.001	<.001	.003	<.001	<.001
	N	192654	160398	192654	192654	192654	192654	192654	192654	192654	192654	192654

		speed_limit	time	urban_or_rural_area	weather_conditions	age_band_of_driver	age_of_vehicle	hit_object_in_carriageway	hit_object_off_carriageway	engine_capacity_cc	sex_of_driver	skidding_and_overturning	vehicle_manoeuvre	vehicle_type
Acci_severity	Pearson Correlation	-.074**	.000	-.080**	.028**	-.050**	-.001	.044**	.033**	.007**	.054**	.006**	.004	.006*
	Sig. (2-tailed)	<.001	.902	<.001	<.001	<.001	.688	<.001	<.001	.002	<.001	.008	.087	.011
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
accident_index	Pearson Correlation	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Sig. (2-tailed)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	N	160398	160398	160398	160398	160398	160398	160398	160398	160398	160398	160398	160398	160398
first_road_class	Pearson Correlation	-.396**	.029**	-.224**	.019**	-.092**	.004	-.006*	-.007**	-.098**	.080**	-.016**	-.012**	-.041**
	Sig. (2-tailed)	.000	<.001	.000	<.001	.000	.105	.014	.002	.000	<.001	<.001	<.001	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
day_of_week	Pearson Correlation	-.018**	.032**	-.016**	.001	.001	.000	.000	.000	.009**	-.003	-.005*	-.001	.005*
	Sig. (2-tailed)	<.001	<.001	<.001	.535	.679	.943	.828	.955	<.001	.240	.024	.566	.032
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
junction_detail	Pearson Correlation	-.121**	.016**	-.114**	.123**	-.051**	-.016**	.312**	.291**	-.023**	.014**	.228**	.250**	-.019**
	Sig. (2-tailed)	.000	<.001	.000	.000	<.001	<.001	.000	.000	<.001	<.001	.000	.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
light_conditions	Pearson Correlation	.048**	.226**	.017**	.090**	-.076**	.027**	.023**	.028**	-.001	-.007**	.041**	.038**	-.013**
	Sig. (2-tailed)	<.001	.000	<.001	.000	<.001	<.001	<.001	<.001	.674	.004	<.001	<.001	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
number_of_casualties	Pearson Correlation	.185**	.022**	.139**	-.029**	.072**	.065**	-.051**	-.042**	.078**	-.030**	-.042**	-.038**	.029**
	Sig. (2-tailed)	.000	<.001	.000	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
number_of_vehicles	Pearson Correlation	.164**	-.014**	.089**	-.038**	-.019**	.006**	-.055**	-.061**	.037**	.041**	-.088**	-.104**	.026**
	Sig. (2-tailed)	.000	<.001	.000	<.001	<.001	.009	<.001	<.001	<.001	<.001	.000	.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
road_surface_conditions	Pearson Correlation	-.013**	-.024**	-.017**	.428**	-.073**	.001	.430**	.412**	-.012**	.025**	.387**	.339**	-.007**
	Sig. (2-tailed)	<.001	<.001	<.001	.000	<.001	.635	.000	.000	<.001	<.001	.000	.000	.003
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
road_type	Pearson Correlation	-.210**	.020**	-.057**	.066**	-.046**	.015**	.124**	.129**	-.052**	.039**	.100**	.100**	-.023**
	Sig. (2-tailed)	.000	<.001	<.001	<.001	<.001	<.001	.000	.000	<.001	<.001	.000	.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
special_conditions_at_site	Pearson Correlation	-.037**	-.001	-.041**	.273**	-.083**	-.020**	.559**	.518**	-.011**	.022**	.454**	.401**	-.009**
	Sig. (2-tailed)	<.001	.589	<.001	.000	<.001	<.001	.000	.000	<.001	<.001	.000	.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
speed_limit	Pearson Correlation	1	-.045**	.667**	-.031**	.128**	.043**	-.089**	-.056**	.107**	-.071**	-.016**	-.027**	.085**
	Sig. (2-tailed)		<.001	.000	<.001	.000	<.001	.000	<.001	.000	<.001	<.001	<.001	.000
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
time	Pearson Correlation	-.045**	1	-.047**	-.010**	-.029**	.022**	-.001	-.001	-.036**	.002	-.007**	.003	-.031**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001	.804	.524	<.001	.315	.002	.175	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
urban_or_rural_area	Pearson Correlation	.667**	-.047**	1	-.032**	.132**	.040**	-.089**	-.056**	.067**	-.056**	-.017**	-.026**	.066**
	Sig. (2-tailed)	.000	<.001		<.001	.000	<.001	.000	.000	<.001	<.001	<.001	<.001	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
weather_conditions	Pearson Correlation	-.031**	-.010**	-.032**	1	-.077**	-.015**	.233**	.222**	-.017**	.031**	.207**	.186**	.000
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001	.000	.000	<.001	<.001	.000	.000	.856
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
age_band_of_driver	Pearson Correlation	.128**	-.029**	.132**	-.077**	1	.125**	-.127**	-.108**	.153**	-.446**	-.097**	-.089**	-.042**
	Sig. (2-tailed)	.000	<.001	.000	<.001		.000	.000	.000	.000	.000	.000	.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
age_of_vehicle	Pearson Correlation	.043**	.022**	.040**	-.015**	.125**	1	-.031**	-.021**	.266**	-.062**	-.017**	-.041**	-.054**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	.000		<.001	<.001	.000	<.001	<.001	<.001	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
hit_object_in_carriageway	Pearson Correlation	-.089**	-.001	-.089**	.233**	-.127**	-.031**	1	.844**	-.030**	.039**	.653**	.657**	-.011**
	Sig. (2-tailed)	.000	.804	.000	.000	.000	<.001		.000	<.001	<.001	.000	.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
hit_object_off_carriageway	Pearson Correlation	-.056**	-.001	-.056**	.222**	-.108**	-.021**	.844**	1	-.025**	.033**	.670**	.659**	-.008**
	Sig. (2-tailed)	<.001	.524	.000	.000	.000	<.001	.000		<.001	<.001	.000	.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
engine_capacity_cc	Pearson Correlation	.107**	-.036**	.067**	-.017**	.153**	.266**	-.030**	-.025**	1	-.120**	-.049**	-.055**	.118**
	Sig. (2-tailed)	.000	<.001	<.001	<.001	.000	.000	<.001	<.001		.000	<.001	<.001	.000
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
sex_of_driver	Pearson Correlation	-.071**	.002	-.056**	.031**	-.446**	-.062**	.039**	.033**	-.120**	1	.001	-.018**	.034**
	Sig. (2-tailed)	<.001	.315	<.001	<.001	.000	<.001	<.001	<.001	.000		.813	<.001	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
skidding_and_overturning	Pearson Correlation	-.016**	-.007**	-.017**	.207**	-.097**	-.017**	.653**	.670**	-.049**	.001	1	.553**	-.026**
	Sig. (2-tailed)	<.001	.002	<.001	.000	.000	<.001	.000	.000	<.001	.813		.000	<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
vehicle_manoeuvre	Pearson Correlation	-.027**	.003	-.026**	.186**	-.089**	-.041**	.657**	.659**	-.055**	-.018**	.553**	1	-.027**
	Sig. (2-tailed)	<.001	.175	<.001	.000	.000	<.001	.000	.000	<.001	<.001	.000		<.001
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654
vehicle_type	Pearson Correlation	.085**	-.031**	.066**	.000	-.042**	-.054**	-.011**	-.008**	.118**	.034**	-.026**	-.027**	1
	Sig. (2-tailed)	.000	<.001	<.001	.856	<.001	<.001	<.001	<.001	.000	<.001	<.001	<.001	
	N	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654	192654

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

   Independent factors have Pearson Relationship less than 0.7 inside themselves, this fulfils the condition of Multicollinearity.

3. **Sample Size:** Logistic Regression needs large number of records with high number of values to classify the Output. (Minimum 50 cases per prediction) Taking 192654 records in this analysis satisfies this assumption.

Case Processing Summary			
Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	192654	100.0
	Missing Cases	0	.0
	Total	192654	100.0
Unselected Cases		0	.0
Total		192654	100.0

a. If weight is in effect, see classification table for the total number of cases.

4. **Outliers:** This dataset has Exceptions which impacts less to classify the Dependent variable; subsequently this presumption is confirmed.
5. **Goodness-of-fit:** This dataset is analysed and has found to be having Goodness-of-fit.

ANALYSIS OF LOGISTIC REGRESSION MODEL:

This analysis has been conducted within the IBM SPSS Statistics software. Here the factors First Road Class, Number of casualties, Day of week, Junction Detail, Number of vehicles, Road surface conditions, Urban or Rural Area, Weather Conditions, Age of Driver, sex of Driver and Hit object in carriageways has been given as the independent variable and Accident Severity has been given as Dependent variable.

Then under Analyse->Regression->Binary Logistic ->under Options ->Statistics and Plots -> the Classification plots, Hosmer-Lemeshow goodness-of-fit, Case wise listing of residuals has been chosen and CI for exp(B) is kept in 95%.

Omnibus Test:

Omnibus test is used to test the performance of model. If model fit is significant this shows that there is significant improvement in fit as compared to null model.in this case as significant value is less than 0.05 which is 0.027, which satisfies the condition.

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	4500.764	22	.000
	Block	4500.764	22	.000
	Model	4500.764	22	.000

Model Summary:

This Model Summary table makes a difference to discover the esteem of distraction in dependent variable in foreseeing the output. This could be done with the assistance of Cox & Snell R Square and Nagelkarke R Square values, here the values are observed as 0.023 and 0.038 respectively. This output has been taken to prove that the predicted value has distraction somewhere between 23% to 38% from the actual value.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	172056.101[a]	.023	.038

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test:

This Hosmer and Lemeshow table offer assistance to demonstrate that the presumption of Goodness-of-fit has been fulfilled. The Significance value watched within the analysis must be more than 0.05, which is 0.583 here. This clarifies the presence of relationship between indicator variable and dependent variable.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	231.603	8	.583

In below figure, difference between observed and expected values are approximately equal so dataset fits the chosen model.

Contingency Table for Hosmer and Lemeshow Test

		Acci_severity = 1.00		Acci_severity = 2.00		
		Observed	Expected	Observed	Expected	Total
Step 1	1	6102	5568.538	13163	13696.462	19265
	2	4209	4357.959	15056	14907.041	19265
	3	3791	3861.958	15474	15403.042	19265
	4	3299	3536.892	15966	15728.108	19265
	5	3225	3290.174	16040	15974.826	19265
	6	2982	3068.161	16283	16196.839	19265
	7	2702	2848.249	16563	16416.751	19265
	8	2434	2601.656	16831	16663.344	19265
	9	2287	2293.102	16978	16971.898	19265
	10	2005	1609.315	17264	17659.685	19269

Classification Table:

This table shown in Classification table is Confusion Matrix, which is used to check for the accuracy of the output after applying the model, which is here 82.9%.

Classification Table[a]

			Predicted		
			Acci_severity		Percentage Correct
Observed			1.00	2.00	
Step 1	Acci_severity	1.00	48	32988	.1
		2.00	23	159595	100.0
	Overall Percentage				82.9

a. The cut value is .500

Variables In the Equation:

The B value in this table signifies the contribution of the independent variable in anticipating the value of output variable.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	first_road_class	.079	.082	.921	1	.337	1.082	.921	1.271
	day_of_week	-.078	.053	2.192	1	.139	.925	.834	1.026
	junction_detail	-.039	.034	1.329	1	.249	.961	.899	1.028
	number_of_vehicles	.414	.155	7.159	1	.007	1.513	1.117	2.048
	road_surface_conditions	-.122	.118	1.078	1	.299	.885	.703	1.114
	special_conditions_at_site	-.139	.099	1.956	1	.162	.870	.717	1.057
	time	.000	.000	.187	1	.665	1.000	1.000	1.000
	weather_conditions	.093	.059	2.454	1	.117	1.097	.977	1.232
	hit_object_in_carriageway	.025	.016	2.614	1	.106	1.025	.995	1.057
	hit_object_off_carriageway	.056	.042	1.788	1	.181	1.057	.974	1.147
	sex_of_driver	.107	.165	.418	1	.518	1.113	.805	1.537
	skidding_and_overturning	-.037	.085	.193	1	.661	.963	.815	1.138
	vehicle_manoeuvre	-.014	.010	1.784	1	.182	.986	.966	1.006
	vehicle_type	.003	.013	.059	1	.809	1.003	.978	1.029
	Constant	1.193	.633	3.554	1	.059	3.297		

a. Variable(s) entered on step 1: first_road_class, day_of_week, junction_detail, number_of_vehicles, road_surface_conditions, special_conditions_at_site, time, weather_conditions, hit_object_in_carriageway, hit_object_off_carriageway, sex_of_driver, skidding_and_overturning, vehicle_manoeuvre, vehicle_type.

In this case number of vehicles will increase 0.414 log odds of output variable.

log(p/1-p) = b0 + b1*x1 + b2*x2 + b3*x3 + b3*x3+…Bn*Xn

Substituting B values in equation, to derive the equation for this Logistic Regression as shown below.

log(p/1-p) = 1.193 + 0.079*First_road_class - 0.078* day_of_week - 0.039*junction detail + 0.414*number_of_vehicles - 0.122*road_surface_conditions - 0.139*Special_conditions_at_site+ 0.093*weather_conditions + 0.025*hit_object_in_carriageway +0.056*hit_object_off_carriageway + 0.107*sex_of_driver - 0.037*skidding_and_overturning -0.014*vehicle_manoeuvre + 0.003*vehicle_type

The independent variables have significance value more than 0.05 which specifies that these variables are less contributing for the prediction of output, while Number of vehicles involved in causing accident contributes on higher side for the prediction.

Exp(B) is the exponential of the Coefficients, this provides us the odd’s ratio of the predictor. In this analysis, the odds of getting severely injured with Vehicles (Car’s has higher number) is 1.513 higher than the opposite.

Also, other factors such as first_raod_class, time, weather conditions, hit object in carriageway, hit object off carriageway, sex of driver (male or female driver) and type of vehicle increases the probability of getting severely injured in accident.

CONCLUSION:

This study has been conducted to analyse the Severity of road accidents of the victims from UK. This resulted that this Binary logistic model could analyse with accuracy of **82.9%** of this case across various scenarios.