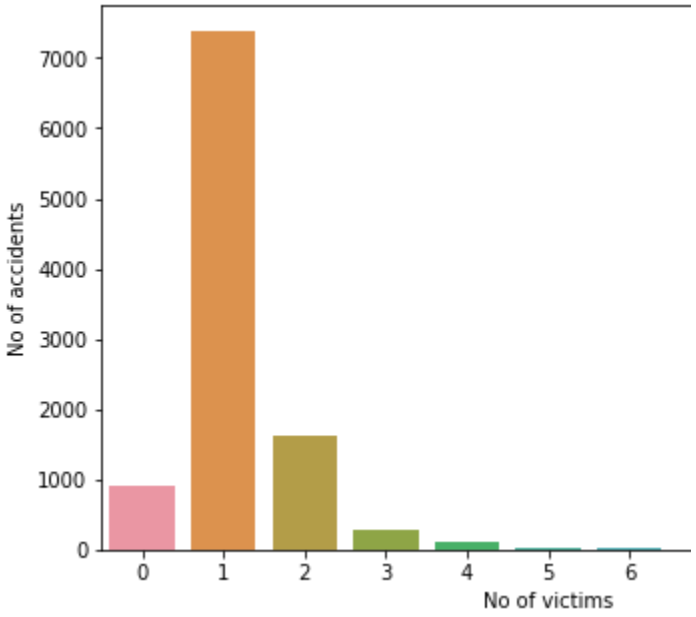
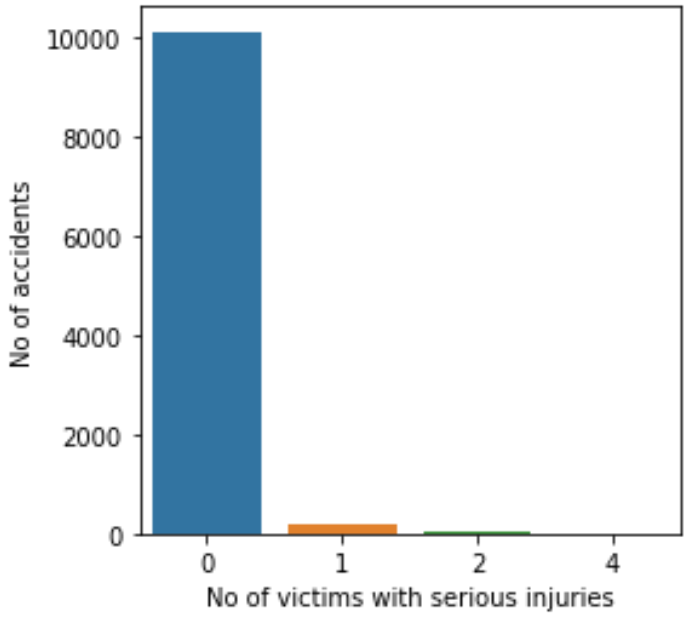
Exploratory Data Analysis on Barcelona 2017 Accidents Dataset as Assignment for DAV Head Position Application

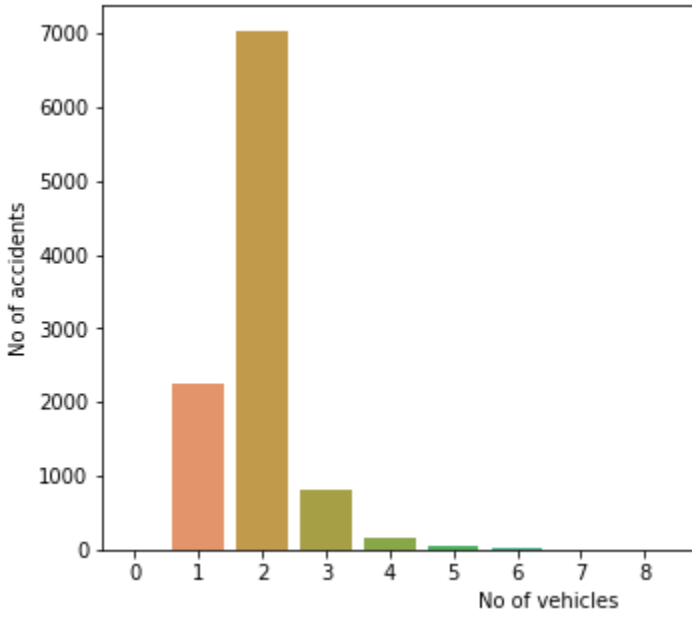
By Shivprasad Kathane (C-MInDS IDDDP Student)

	Id	District Name	Neighborhood Name	Street	Weekday	Month	Day	Hour	Part of the day	Mild injuries	Serious injuries	Victims	Vehicles involved	Longitude	Latitude
0	2017S008429	Unknown	Unknown	Número 27 ...	Friday	October	13	8	Morning	2	0	2	2	2.125624	41.340045
1	2017S007316	Unknown	Unknown	Número 3 Zona Franca / Número 50 Zona Franca ...	Friday	September	1	13	Morning	2	0	2	2	2.120452	41.339426
2	2017S010210	Unknown	Unknown	Litoral (Besòs) ...	Friday	December	8	21	Afternoon	5	0	5	2	2.167356	41.360886
3	2017S006364	Unknown	Unknown	Número 3 Zona Franca ...	Friday	July	21	2	Night	1	0	1	2	2.124529	41.337668
4	2017S004615	Sant Martí	el Camp de l'Arpa del Clot	Las Navas de Tolosa ...	Thursday	May	25	14	Afternoon	1	0	1	3	2.185272	41.416365

What are the characteristics of a typical accident?

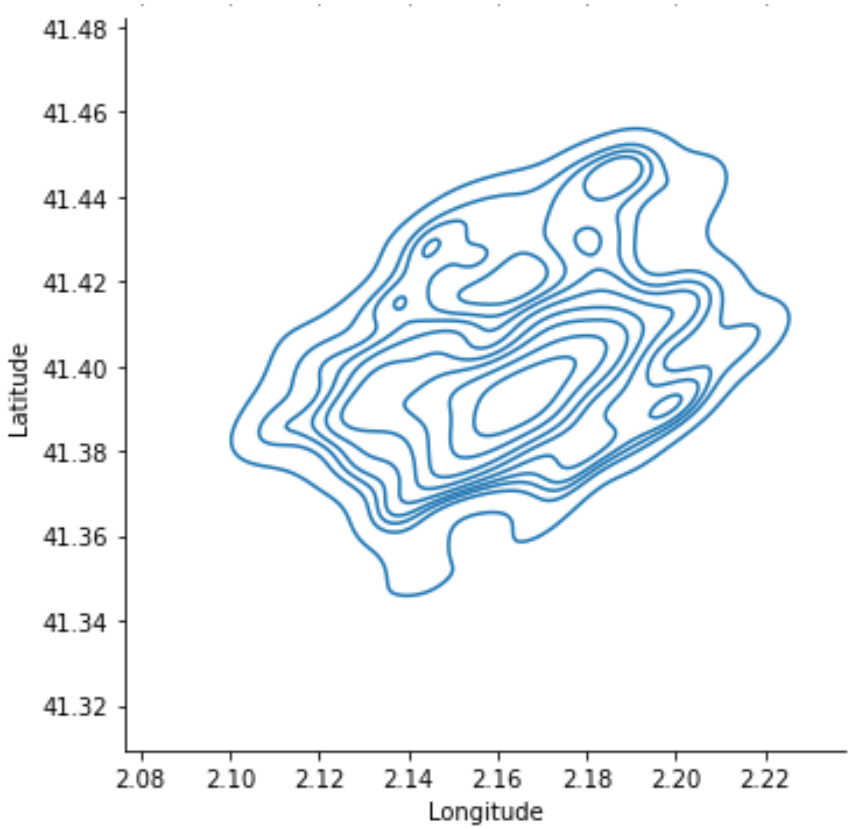




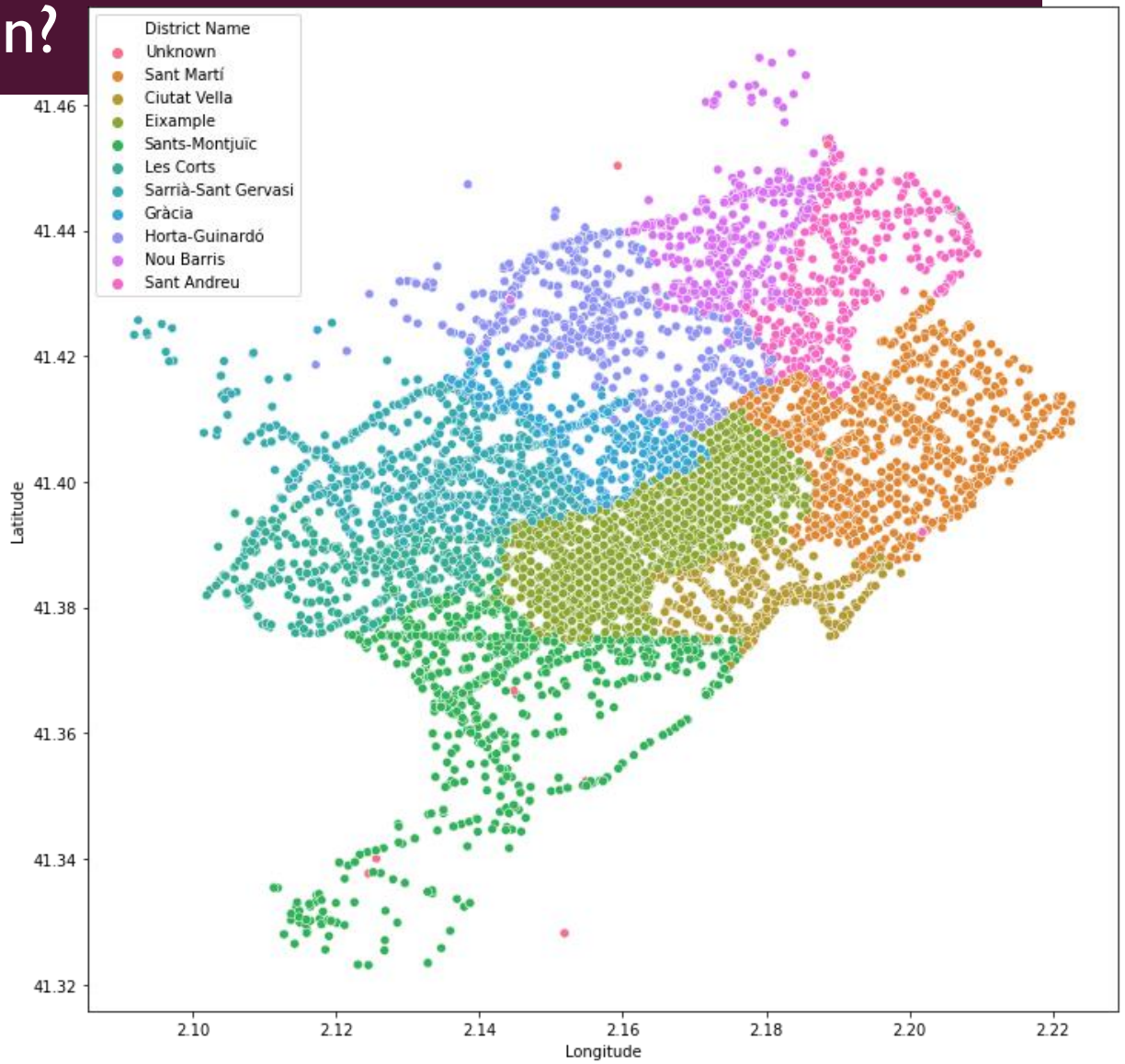


A typical accident involves collision between 2 vehicles and leads to one victim who most often has just mild injuries.

How do the accidents vary with location?

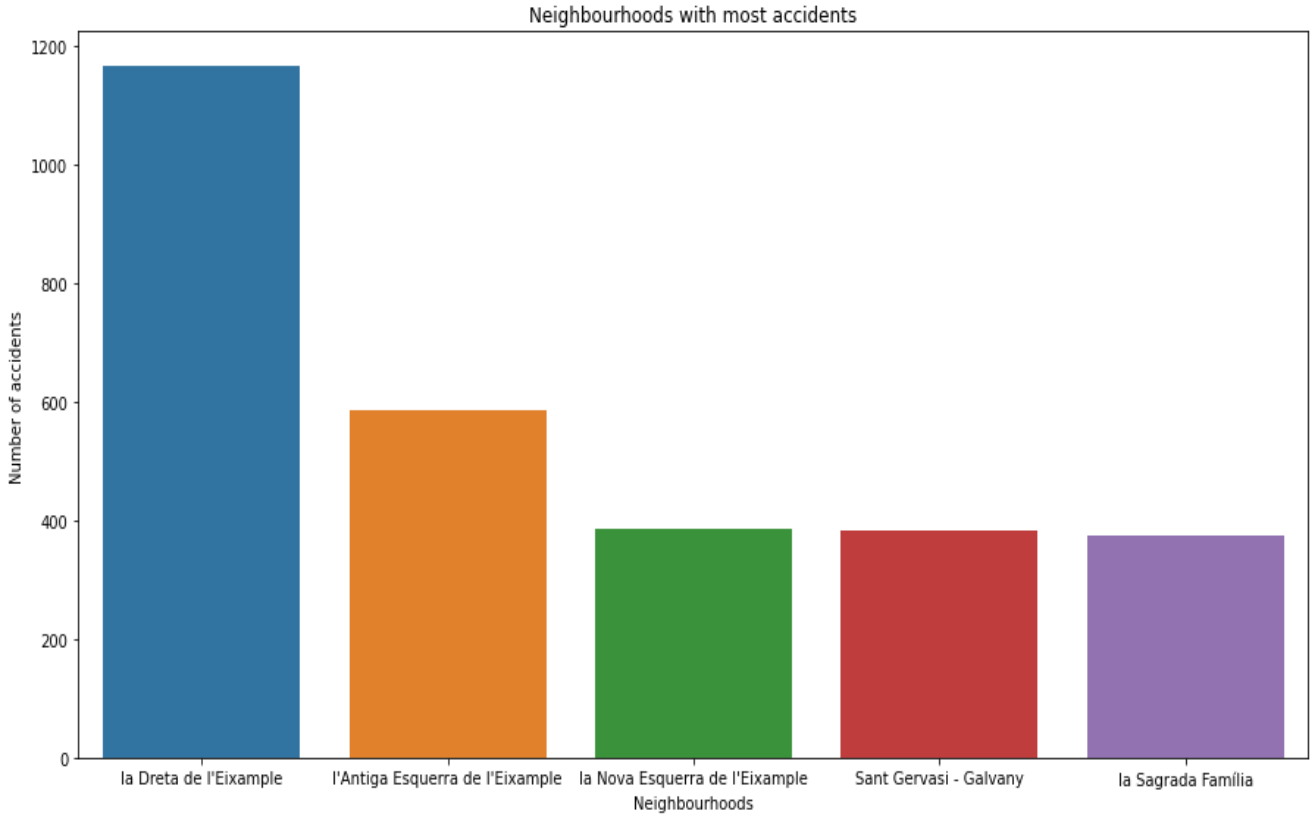


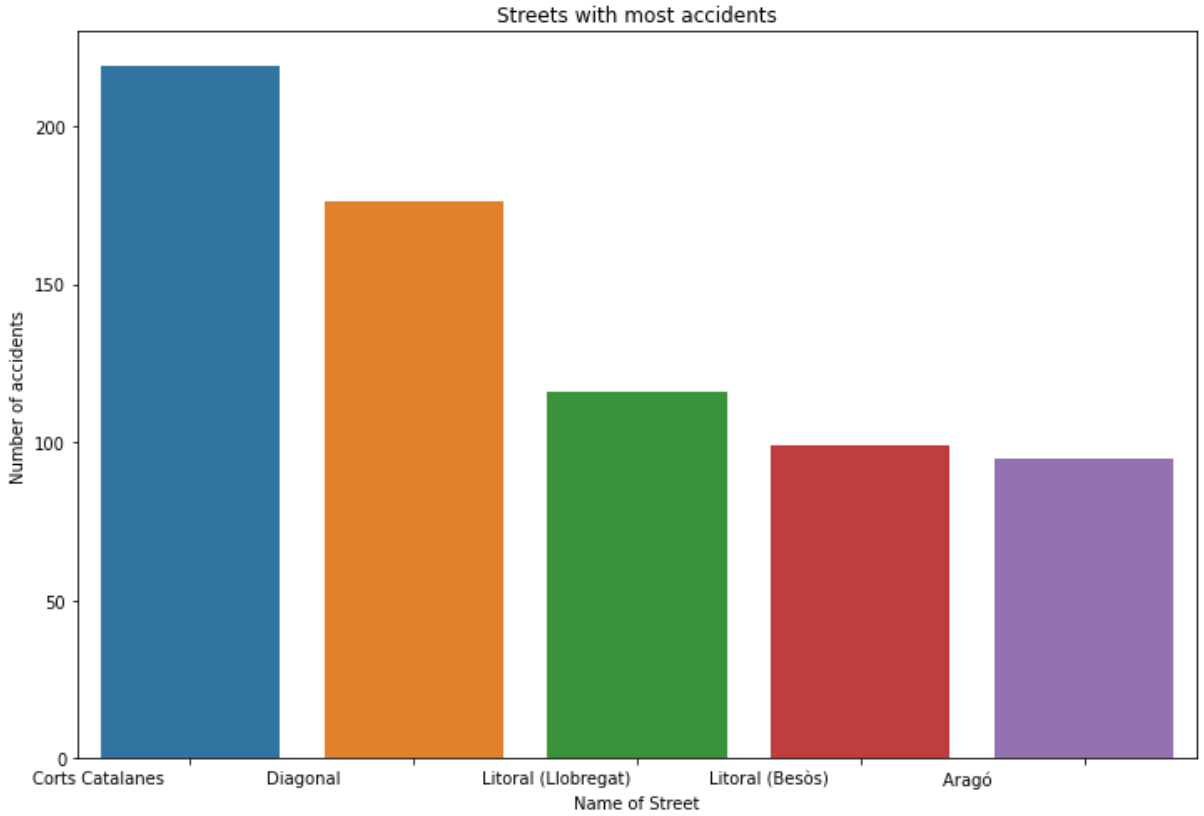
From both plots as well as data, it can be concluded that 'Eixample' is the district with maximum number and density of accidents



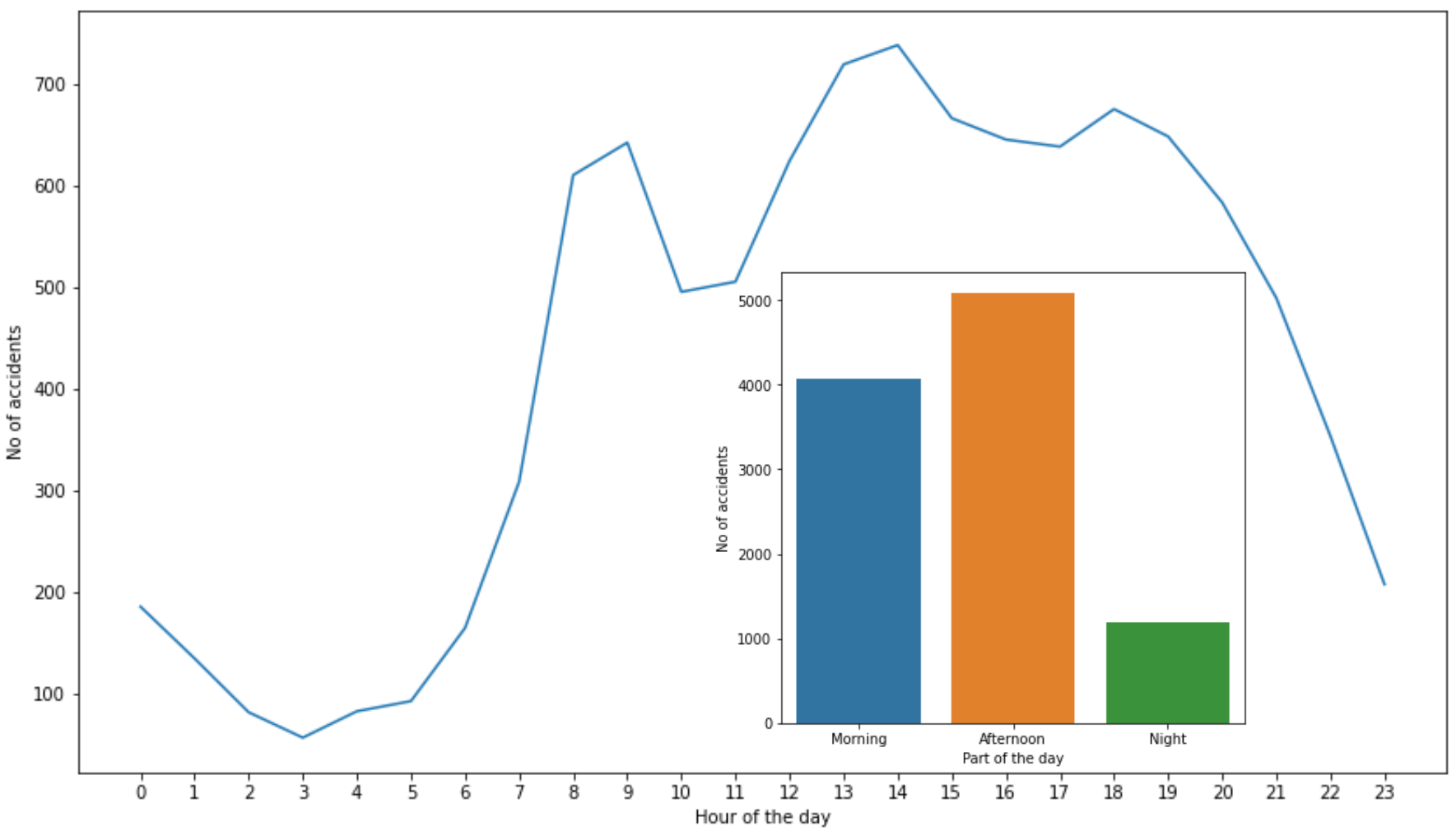
The 2D map on right shows the location of the accidents with the color representing the district. The plot on its left shows density curves indicating max density of accidents around (2.17, 41.39).

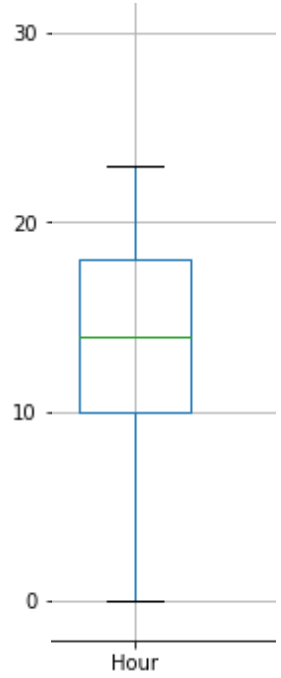
What are the top neighbourhoods and streets prone to accidents?





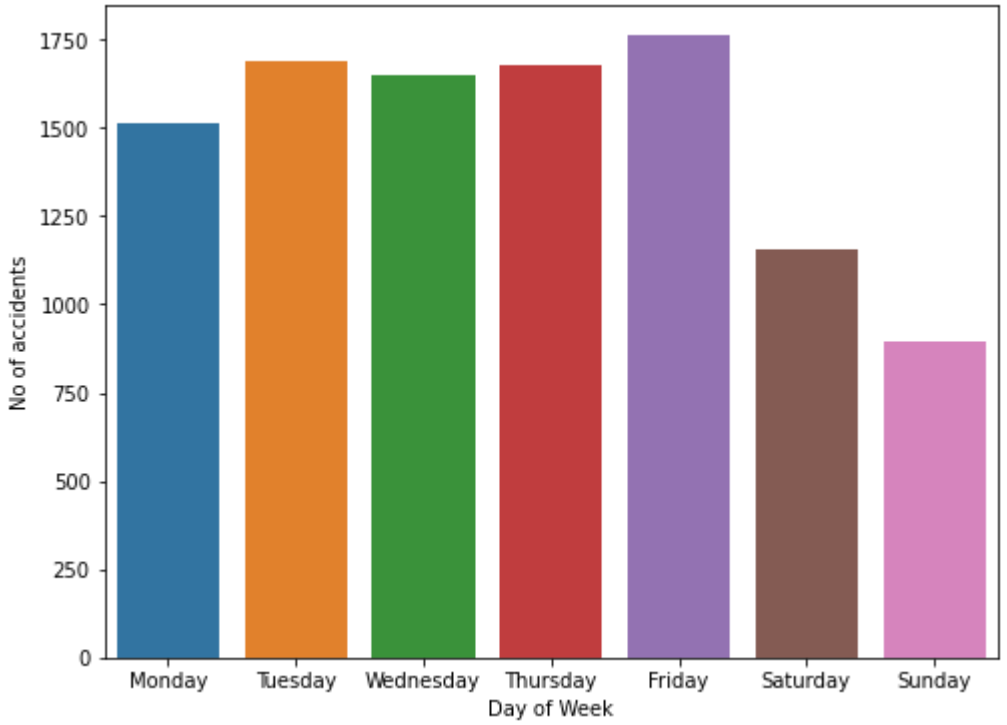
How do the accidents vary with time of day?

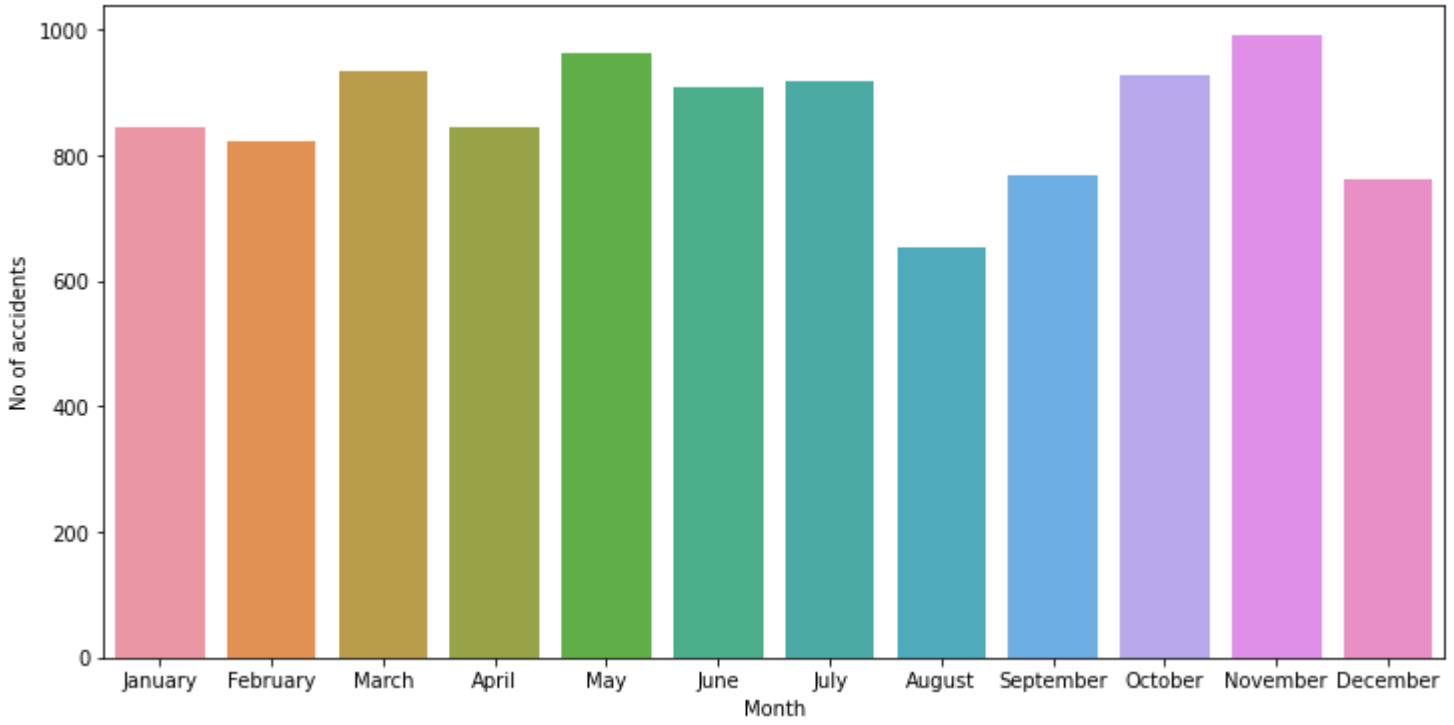




Central 50 % of the accidents take place between 10th and 18th hour of the day which also corresponds to usual working hours. Very few occur at night.

How do accidents vary with the day of week and with month?

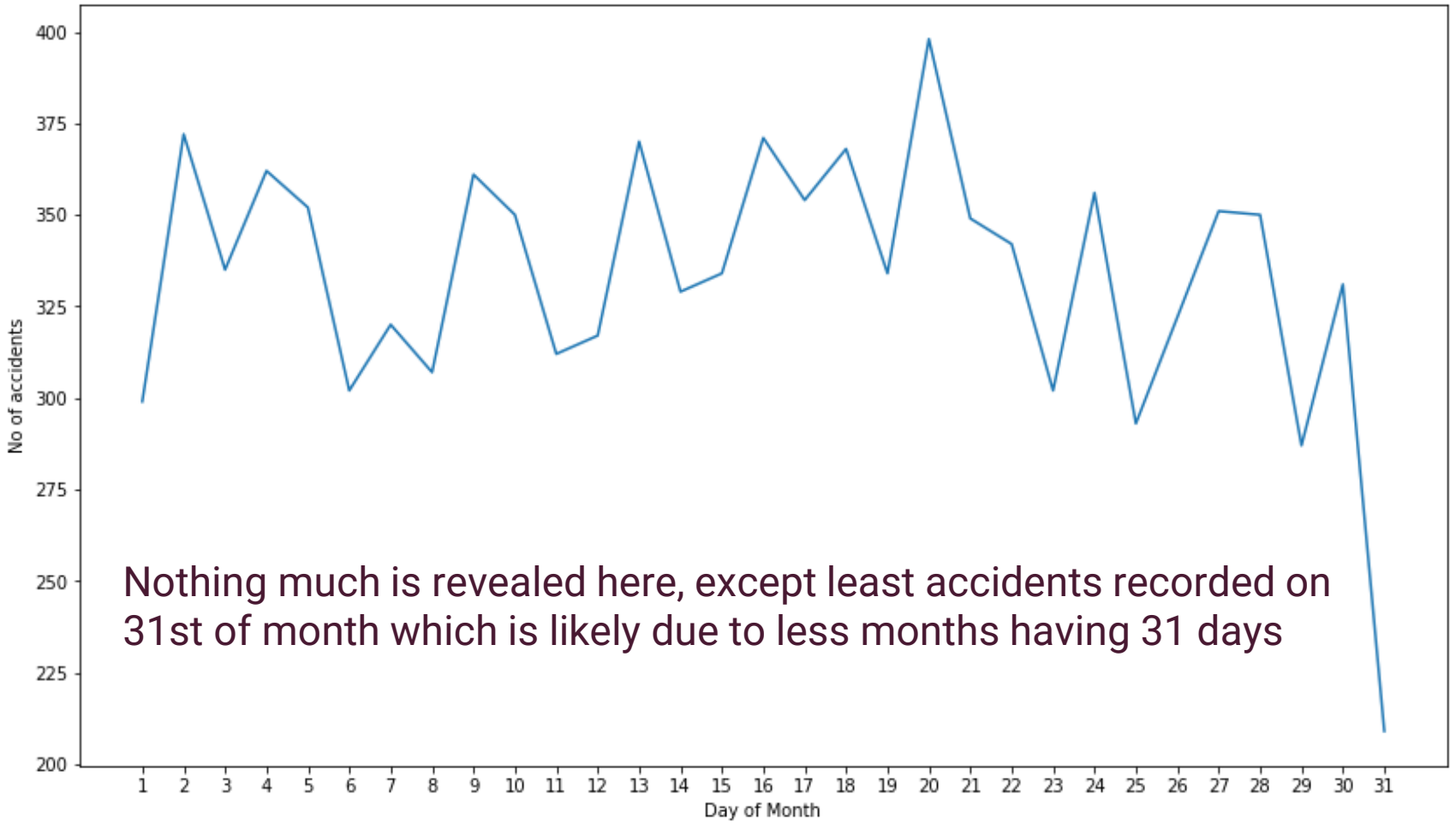




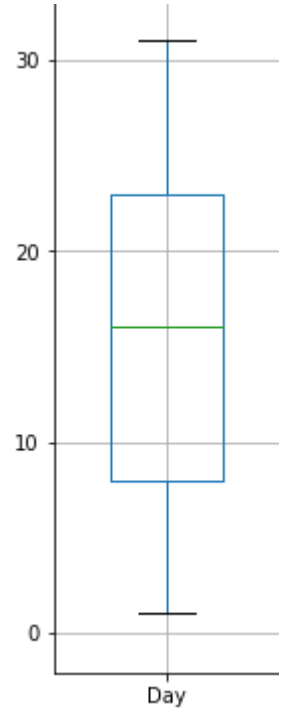
Clearly less number of accidents on weekend i.e. Saturday and Sunday compared to workdays

Highest number of accidents recorded in November while lowest in August

How do the accidents vary with the day of month?



Nothing much is revealed here, except least accidents recorded on 31st of month which is likely due to less months having 31 days

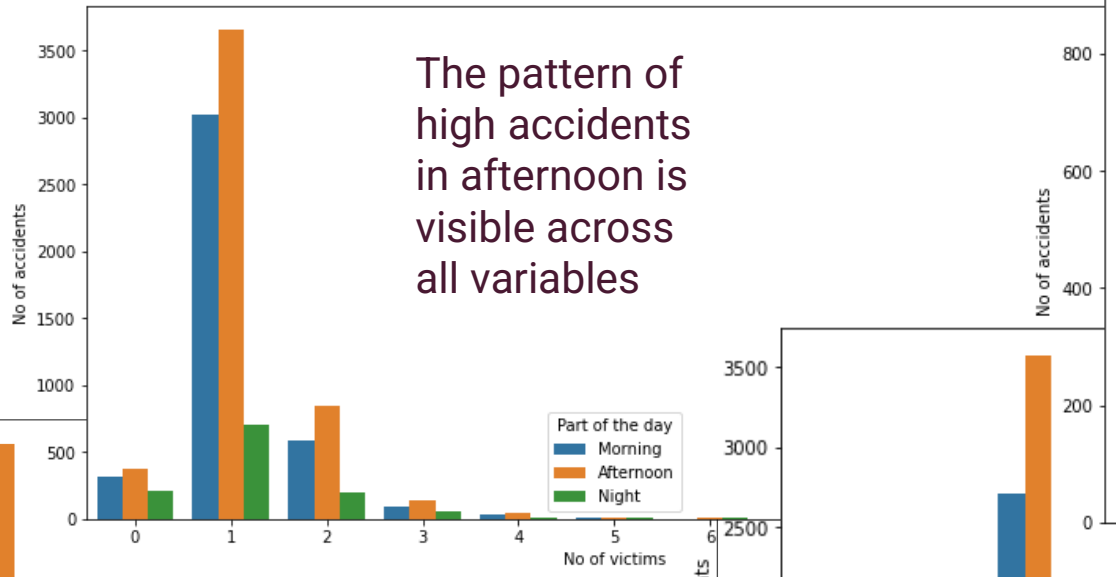


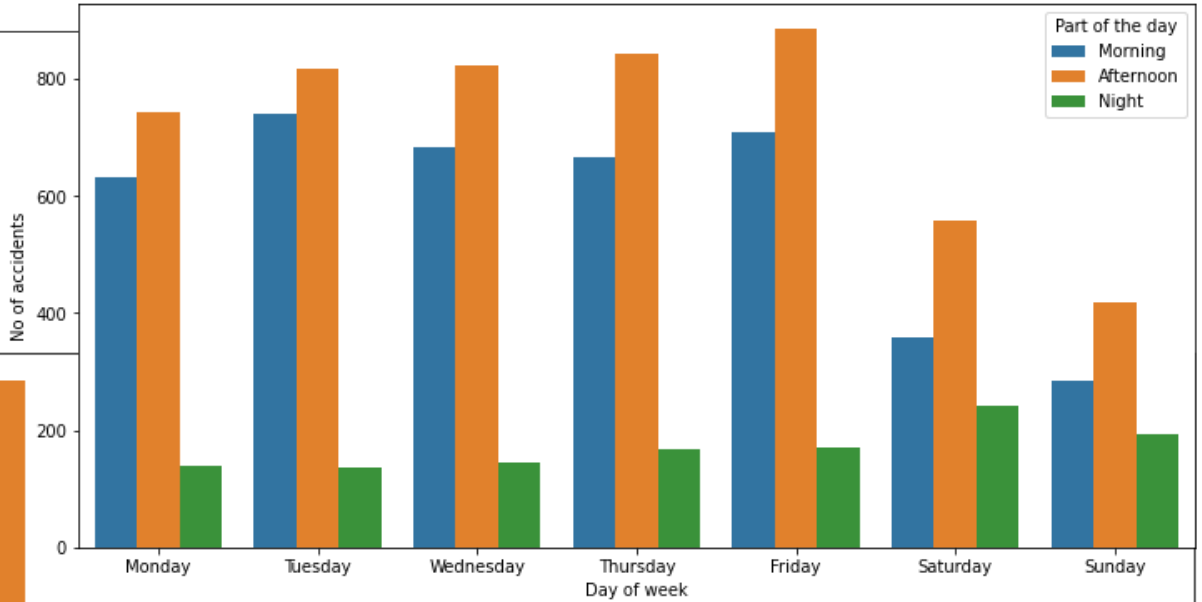
Box plot showing central 50 % of the accidents taking place between 8th and 23rd day of the month

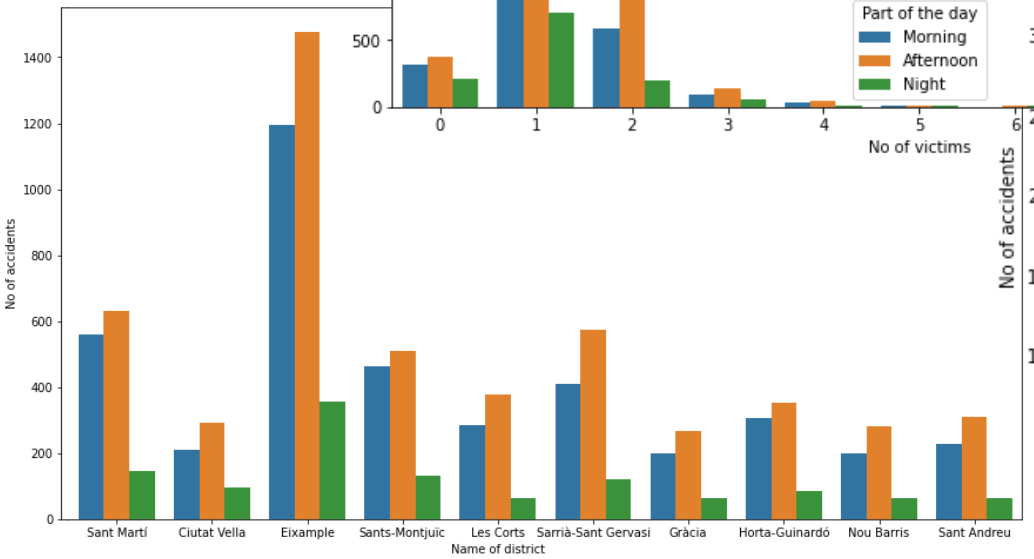
Does the part of day have special influence on any other variable?

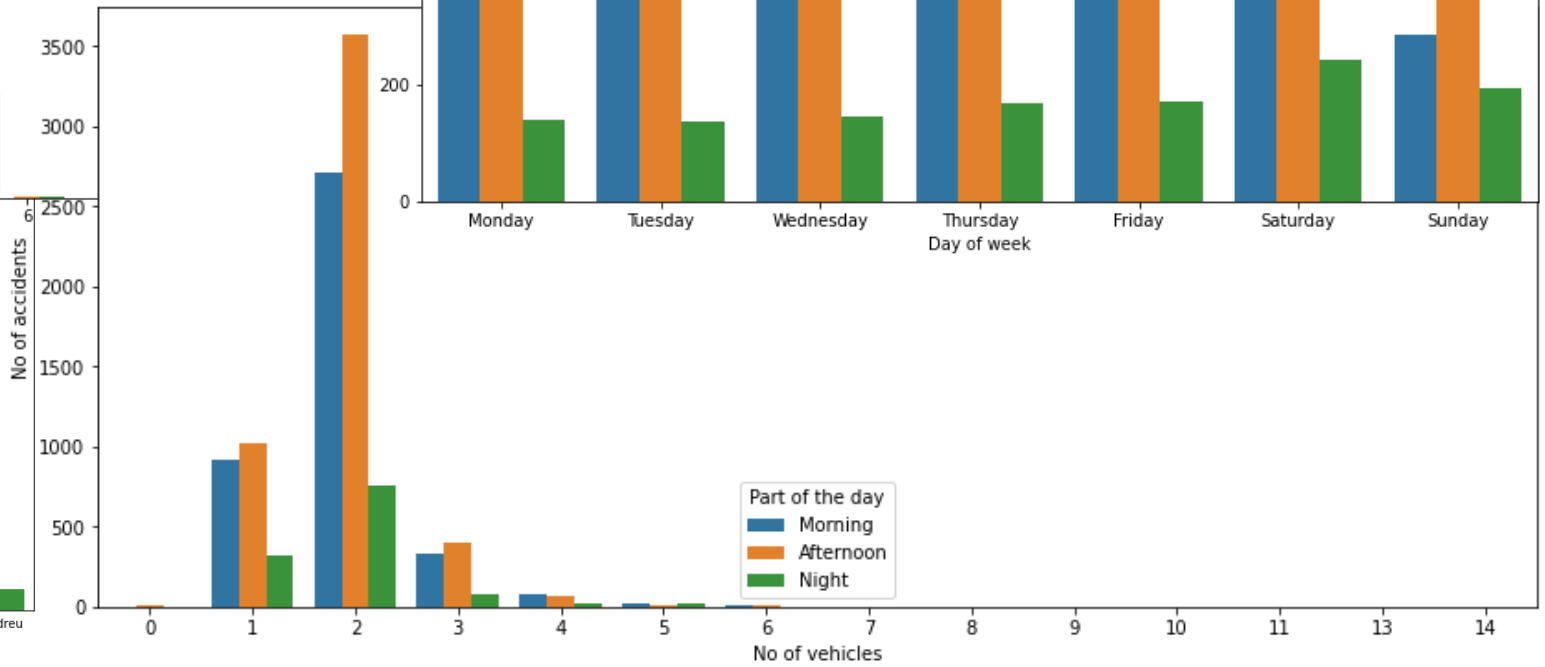
No!

The pattern of high accidents in afternoon is visible across all variables

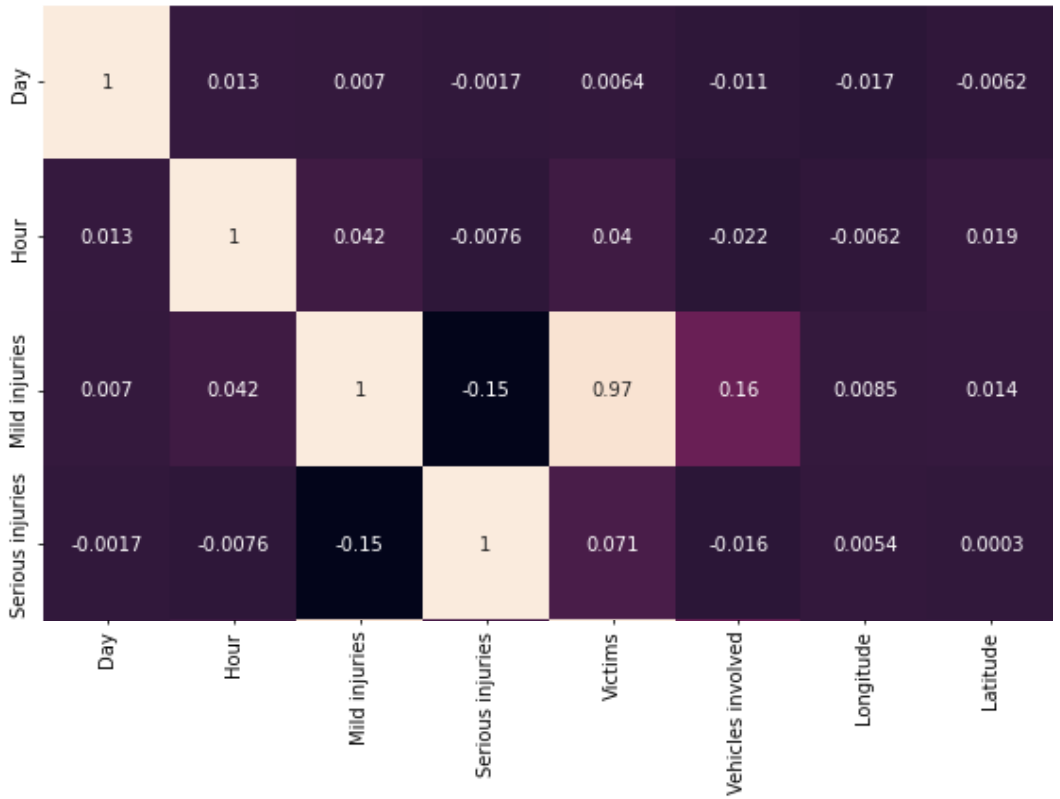


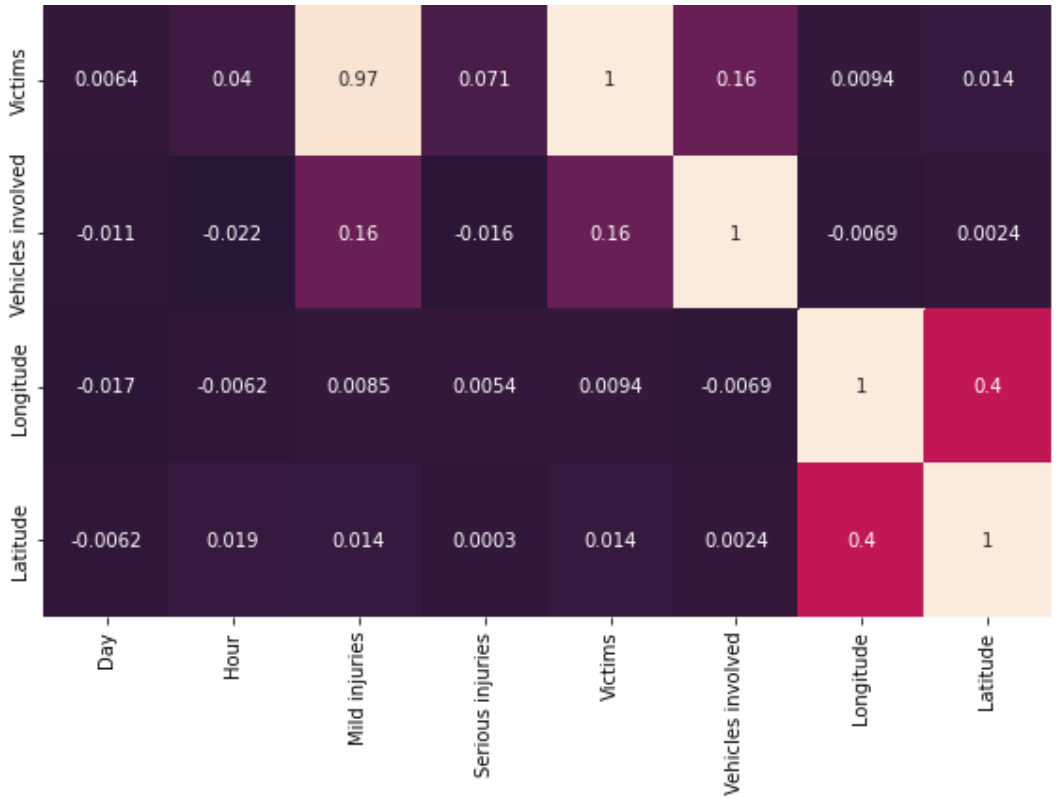






Do the numerical variables influence each other (i.e. any correlation) ?





Strong positive correlation between victims and mild injuries. Here, this implies most of the victims have mild injuries.
Weak positive correlation between latitude and longitude. This is observed on map as most accidents fall in a inclined rectangular region.

Thank You