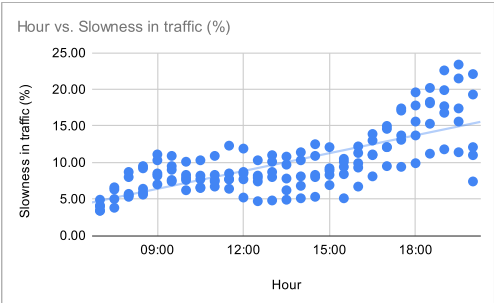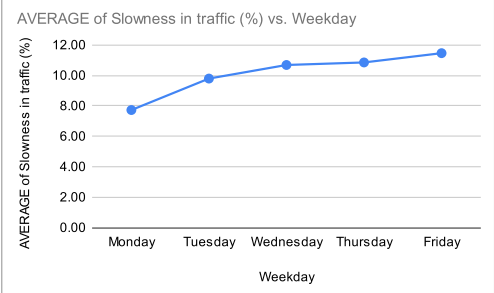## Part 2) Identifying and excluding the outliers!

In part 1, we have established that there are certain outliers/extremes that exist within the data.
In this section, we will closely analyze the outliers and will make a decision if we want to remove/retain them.

Percentile Rule For Detecting Outliers		IQR Rule For Detecting Outliers						
P1 -	3.53	Slowness in traffic	Min	Lower Whisker	Q1	Q3	Upper Whisker	Max
P99 -	22.44	Slowness in traffic	3.40	0.69	7.4	11.875	18.59	23.40

Insight - According to the percentile rule, any value outside the range P1 - P99 is considered to be an outlier.
According to the IQR rule, any value outside the range of Lower Whisker - Upper Whisker is considered to be an outlier.

Part 3) Analyzing traffic patterns by time

Correlation Between Hours and Slowness in Traffic
Finding Correlation On Weekdays

Weekdays	Correlation(Hours vs slown	Correlation Strength
Monday	0.28	Weak
Tuesday	0.76	Strong
Wednesday	0.90	Strong
Thursday	0.92	Strong
Friday	0.83	Strong
Overall	0.73	Strong

Overall Hour vs Slowness Relation (Scatter Plot)

Hour vs. Slowness in traffic (%)

Insight - With the above analysis, we can conclude that Hours and Slowness in Traffic are Positively Correlated
There are multiple peaks within the data with respect to time.
The congestion happens during the peak hours of morning and evening time.
The congestion may happen due to office/school hours during the peak time.

Trend Analysis of Average Slowness In Traffic During Weekdays

Weekday	AVERAGE of Slowness in	
Monday	7.72	< Lowest Avg Congestion Rate
Tuesday	9.79	
Wednesday	10.68	
Thursday	10.85	
Friday	11.46	< Highest Avg Congestion Rate

AVERAGE of Slowness in traffic (%) vs. Weekday

Insight - With the plots defined above, we can clearly state that the congestion happens due to office/school hours aka peak hours.
Usual peak hours: 8:00 - 9:00 AM in the morning and slowly reduce around 10:30-11 am.
The next set of peak hours starts from 4 pm in the evening and goes until the late evening.

Part 4) Identifying the relationship between variables and their contribution towards slowness in traffic

Correlation Matrix In Order To Find Different Variable's Impact On Slowness in Trffic

CORRELATION MATRIX	Slowness in traffic (%)
Fire vehicles	0.13
Broken Truck	0.13
Accident victim	0.12
Immobilized bus	0.10
Occurrence involving freight	0.03
Incident (dangerous freight)	0.00
Running over	0.00
Vehicle excess	-0.05

CORRELATION MATRIX	Slowness in traffic (%)
Lack of electricity	0.44
Point of flooding	0.42
Semaphore off	0.35
Manifestations	0.06
Fire	-0.05
Tree on the road	-0.10
Intermittent Semaphore	-0.12
Defect(trolleybuses network)	-0.14

Insight - The lack of electricity and point of flooding are the 2 variables that highly affect the traffic.
Though not easy, these are certain factors that we can control to a certain extent as we have dedicated electricity departments and municipal corporations in cities.