

NYC Motor Crash Report



Analyze Topic

what factor will lead to higher risk of death from motor vehicle crash?

Why

To find out high risk causes , so we can pay more attention to these causes and better allocate resources



Hypothesis

Multiple car crashes should cause the most death

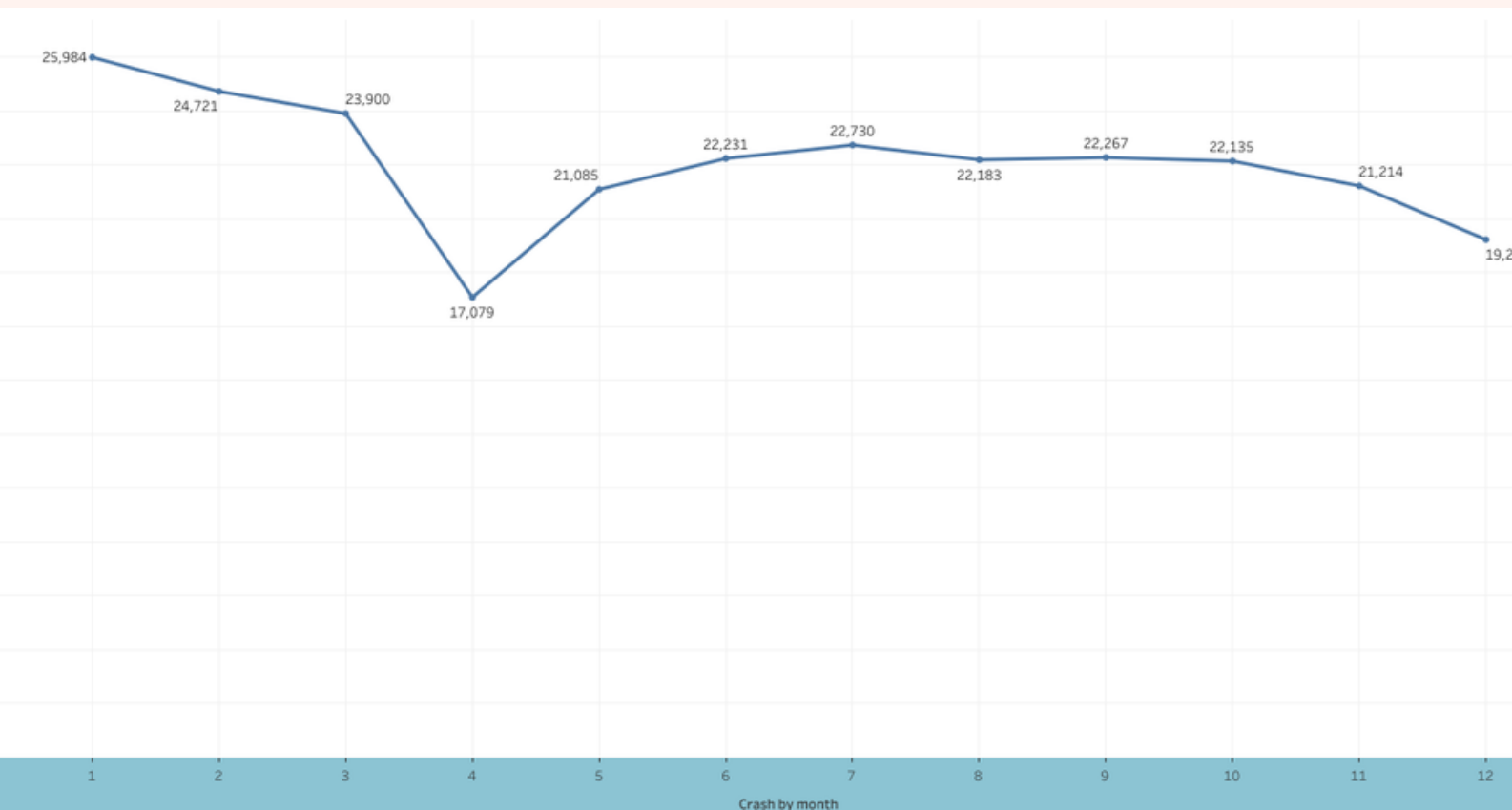


Data

DV: People die in crashes (dummy variable)
Independent variable :Crash_Time,
Crashes_by_month, BOROUGH factor,
Ten_Most_Common_Factor,
,Ten_Most_Vehicle_Type



This graph show the total number of crashes by month



How does it relate to the risk of dying in a car crash?

WHAT DOES LINEAR REGRESSION TELL US

Only September is consider statistically significant, every increase in Steptember, the DV will be increase by $4.727e-04$.

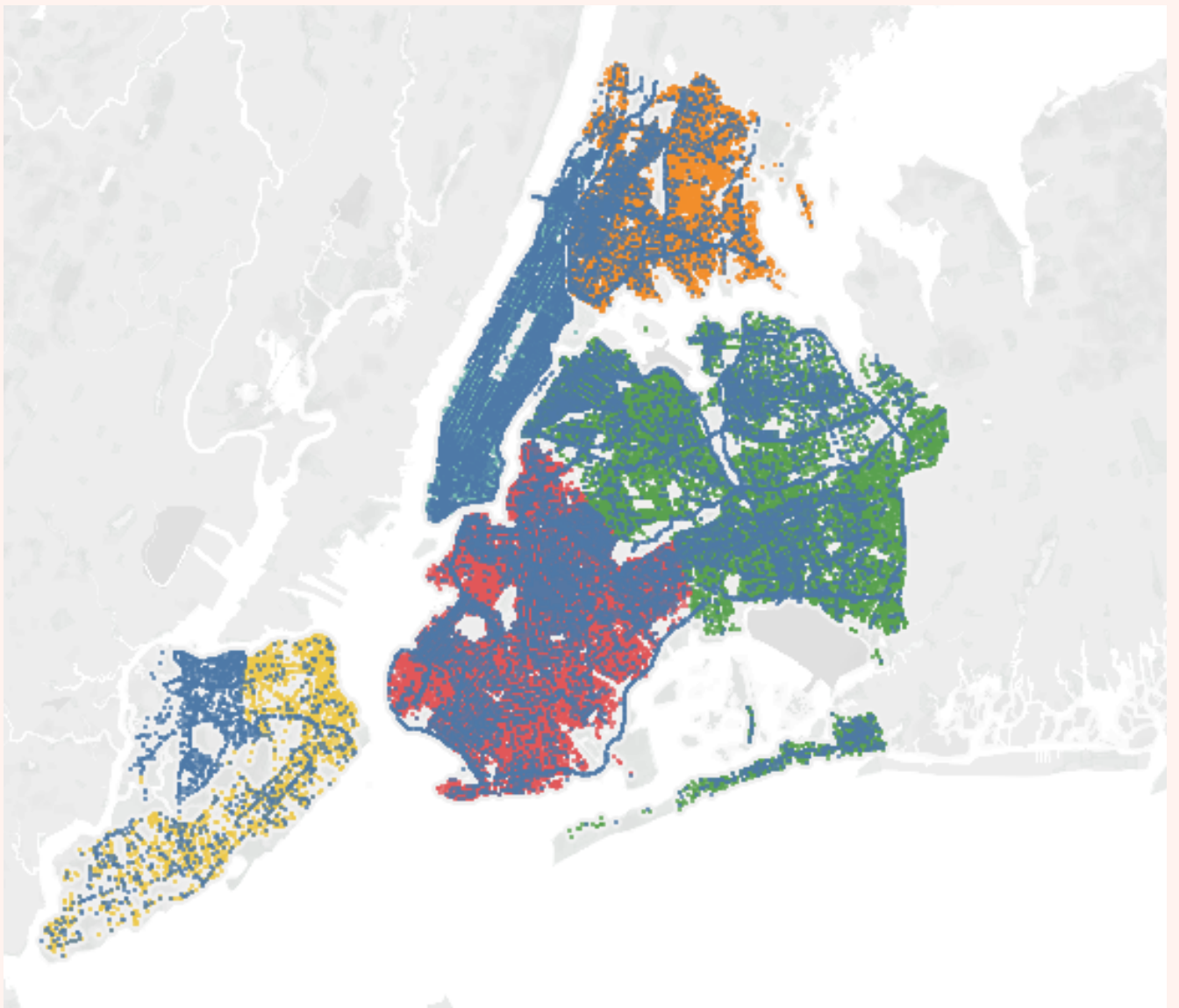
factor(Crashes_by_month)2	5.627e-05	2.620e-04	0.215	0.829959
factor(Crashes_by_month)3	-2.713e-04	2.643e-04	-1.026	0.304737
factor(Crashes_by_month)4	-1.438e-04	2.906e-04	-0.495	0.620596
factor(Crashes_by_month)5	9.681e-06	2.735e-04	0.035	0.971758
factor(Crashes_by_month)6	2.295e-04	2.696e-04	0.851	0.394760
factor(Crashes_by_month)7	-1.129e-05	2.681e-04	-0.042	0.966418
factor(Crashes_by_month)8	1.408e-04	2.699e-04	0.522	0.601791
factor(Crashes_by_month)9	4.727e-04	2.696e-04	1.753	0.079523
factor(Crashes_by_month)10	-5.835e-05	2.699e-04	-0.216	0.828864
factor(Crashes_by_month)11	3.634e-04	2.730e-04	1.331	0.183176
factor(Crashes_by_month)12	2.224e-04	2.807e-04	0.792	0.428105

9

WHAT DOES LOGISTIC REGRESSION TELL US

Logistic regression give the same month. Only September has a p-value less than .05. Increasing value of September will increase the odd of dying car crashes by 0.51243

DOES BOROUGH AFFECT THE DEATH RATE?



WHAT DOES LINEAR REGRESSION TELL US

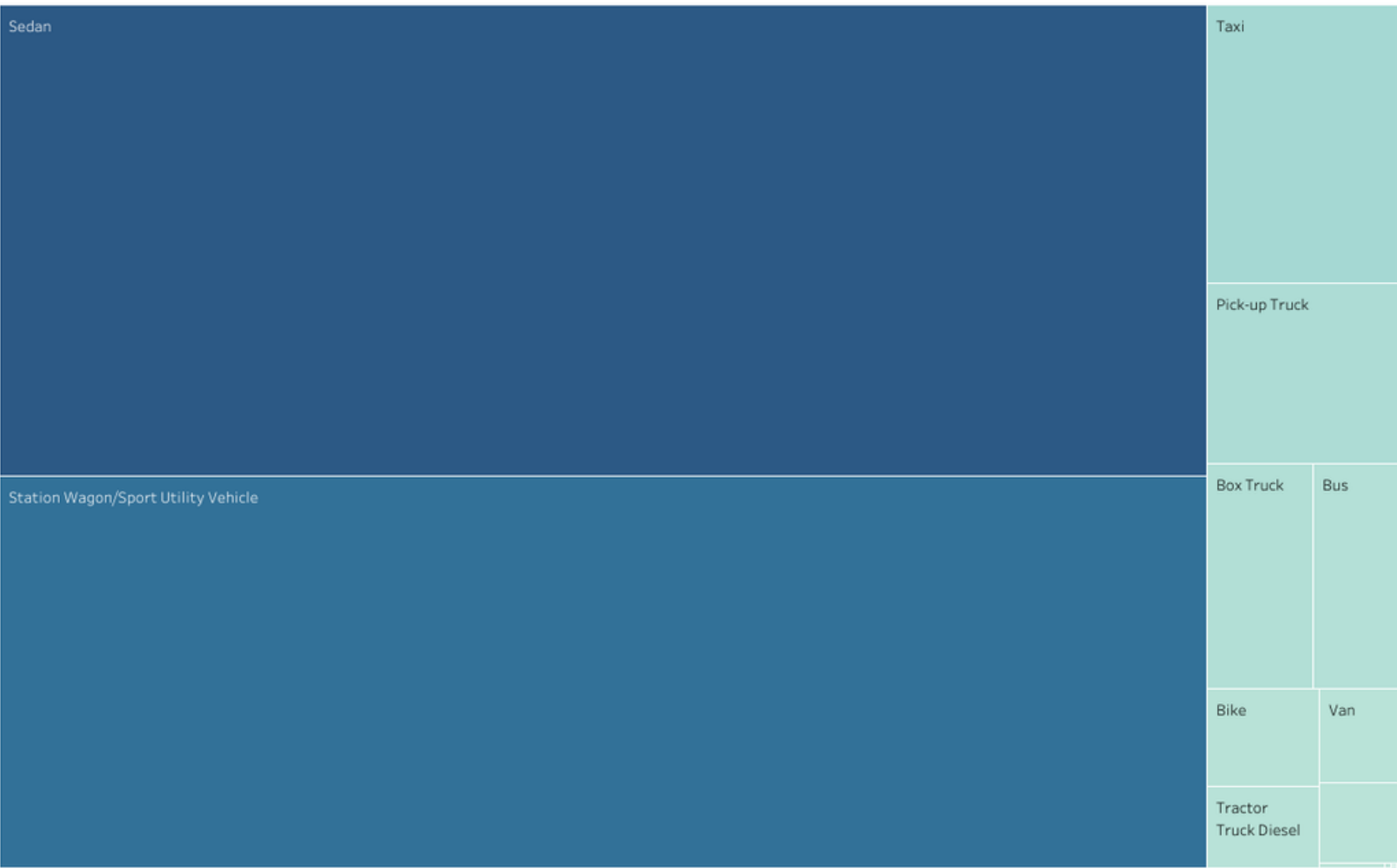
All 5 borough has positive relationship with the DV but only Queens borough has p-value less than .05 which is 0.038139. Increasing of Queens will also increase the DV by $4.506e-04$.

WHAT DOES LOGISTIC REGRESSION TELL US

Borough also has positive relationship with the odd of dying in crash. But only the Queens borough is consider significant. Every increase of Queen will increase the odd of dying by 0.52472

WHAT IS THE MOST DEADLY VEHICLE TYPE?

Sheet 3



WHAT DOES LINEAR REGRESSION TELL US

Even sedan and Station Wagon/Sport Utility Vehicle has way more crash than the other vehicle types, but Motorcycle is the most dangerous vehicle during car crashes. It's p-value also much smaller than .05 which indicate it is statically significant



WHAT DOES LOGISTIC REGRESSION TELL US

We use the 10 most common vehicle type as variable to find the relationship with the DV, not surprise the motorcycle also come out at top. Every increase in motorcycle will increase the odd of dying by 0.136024.

WHAT ARE THE MOST DEADLY CONTRIBUTING FACTORS?



WHAT DOES LINEAR REGRESSION TELL US

Failure to Yield Right-of-Way has the strongest positive relationship and smallest p-value that less than .05 (3.39e-05)

WHAT DOES LOGISTIC REGRESSION TELL US

Same as the linear regression, Yield Right-of-Way(factor 2) also the most serious factor to cause death in crashes

factor(Ten_Most_Common_Factor)2	0.82030	0.21601	3.797	0.000146	*
factor(Ten_Most_Common_Factor)3	-2.43066	1.01270	-2.400	0.016387	*
factor(Ten_Most_Common_Factor)4	-0.95327	0.59528	-1.601	0.109294	
factor(Ten_Most_Common_Factor)5	0.02299	0.43448	0.053	0.957803	
factor(Ten_Most_Common_Factor)6	-13.77037	251.47063	-0.055	0.956330	
factor(Ten_Most_Common_Factor)7	-0.11262	0.59649	-0.189	0.850246	
factor(Ten_Most_Common_Factor)8	-0.21596	0.59900	-0.361	0.718450	
factor(Ten_Most_Common_Factor)9	-0.12193	0.18260	-0.668	0.504287	
factor(Ten_Most_Common_Factor)10	0.07899	0.36488	0.216	0.828617	

MULTIPLE CARS IN ONE CRASH?



WHAT DOES LINEAR REGRESSION TELL US

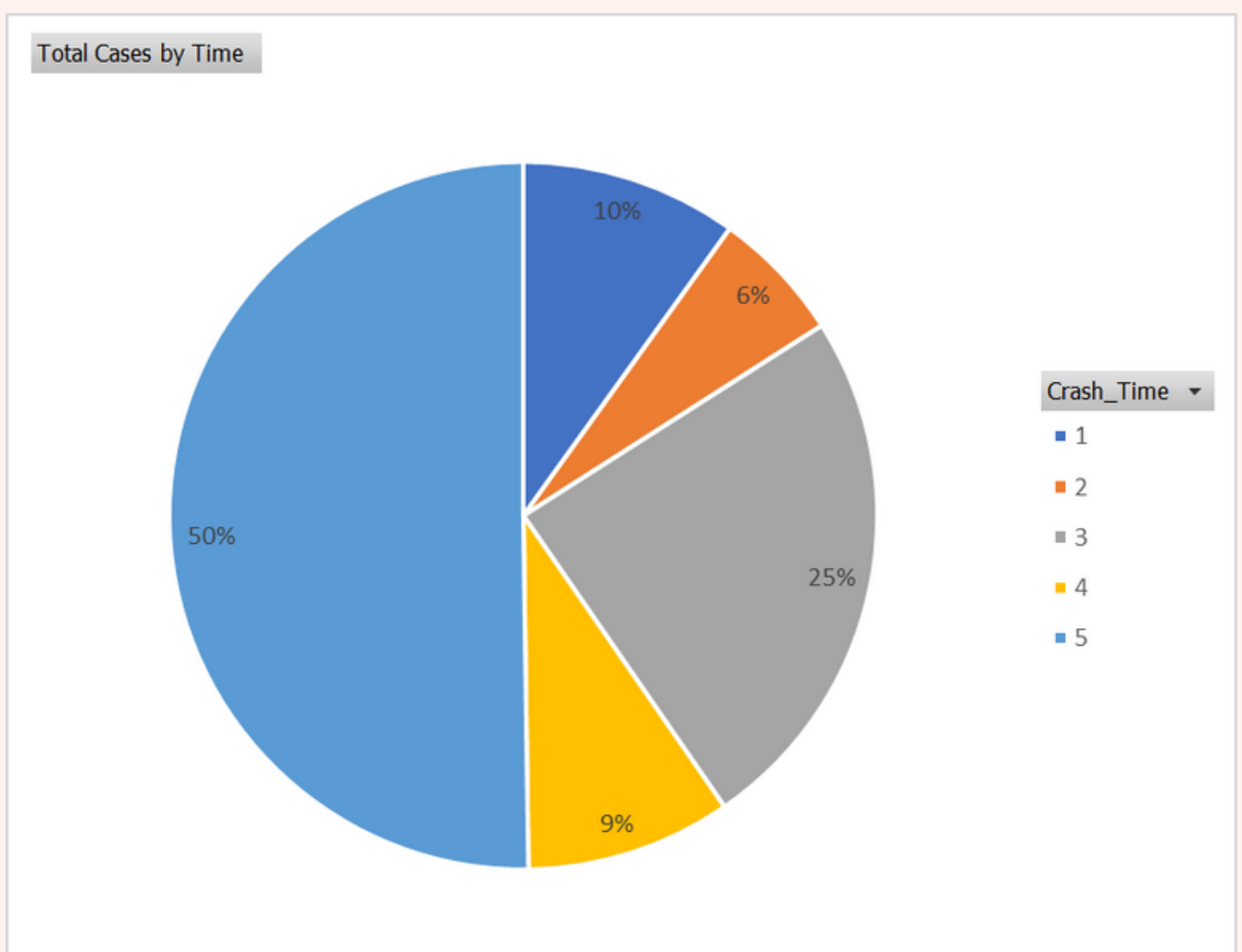
The relationship is negative when it is only involved with 2 cars, then turn into positive as more cars involved. The interesting point here is that more car involved in a crash, the less Est Std it goes.

WHAT DOES LOGISTIC REGRESSION TELL US

The overall p-values are larger than .05. So, multiple cars variables are consider as insignificant in logistic regression

Two_cars	-0.13457	0.30689	-0.439	0.661012
Three_Cars	1.34244	1.31254	1.023	0.306413
Four_Cars	0.75858	2.12415	0.357	0.721003
Five_Cars	1.36381	2.73464	0.499	0.617979

DOES TIME MATTER?



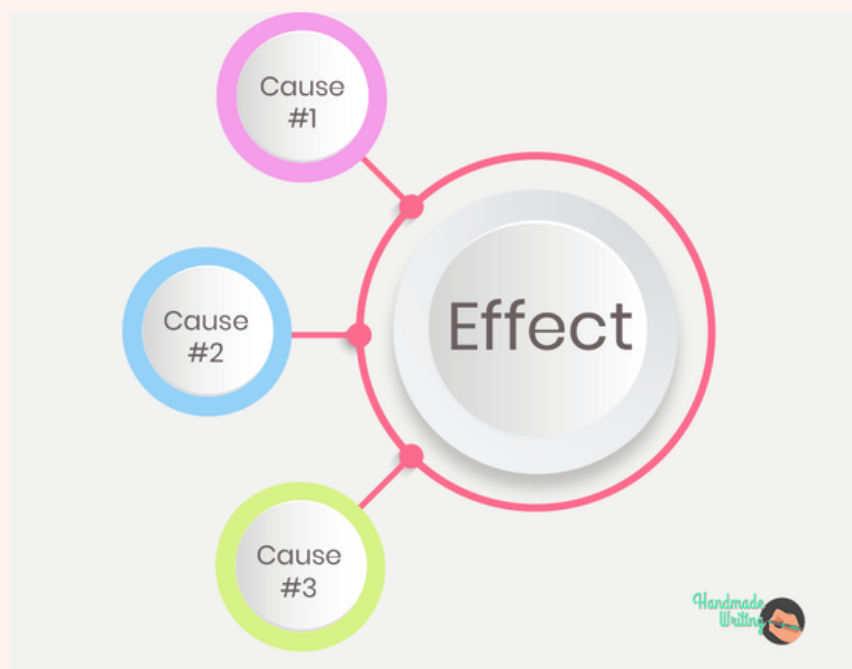
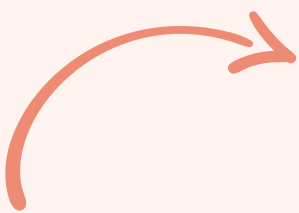
WHAT DOES LINEAR REGRESSION TELL US

Base on the result, midnight (12 AM to 6 AM) has the strongest relationship with DV, the remaining 4 time slots are all negative when compare to the midnight. All their p-value are smaller than .05. So this time slot variables are consider statically significant.

WHAT DOES LOGISTIC REGRESSION TELL US

The Est Std of crash times is -0.11286 , meaning increasing of time will decrease the odd of dying in cashes. This make sense because the the peak Est Std is midnight, and then it getting smaller as time goes by base on the linear regression.

MULTIPLE FACTORS IN ONE CRASH?



WHAT DOES LINEAR REGRESSION TELL US

Intesting Point: when it invloved more than 2 factors, the relationship with DV is negative.

The p-value of $2e-16$ also prove it is statistically significant. The p-value of 3 ,4,5 factors are bigger than .05, so they are insignificant.



WHAT DOES LOGISTIC REGRESSION TELL US

The result is same as the linear regression. When it is two factor, the odd of dying will decrease.



SOLUTION



- Apply a strict license exam for motorcycle
- Hire more NYPD to handle car accidents and enforce traffic laws
- Increase the amount of night shifts
- Create an app and input the data that indicate area that frequently happen car crashes for the policers to portal

