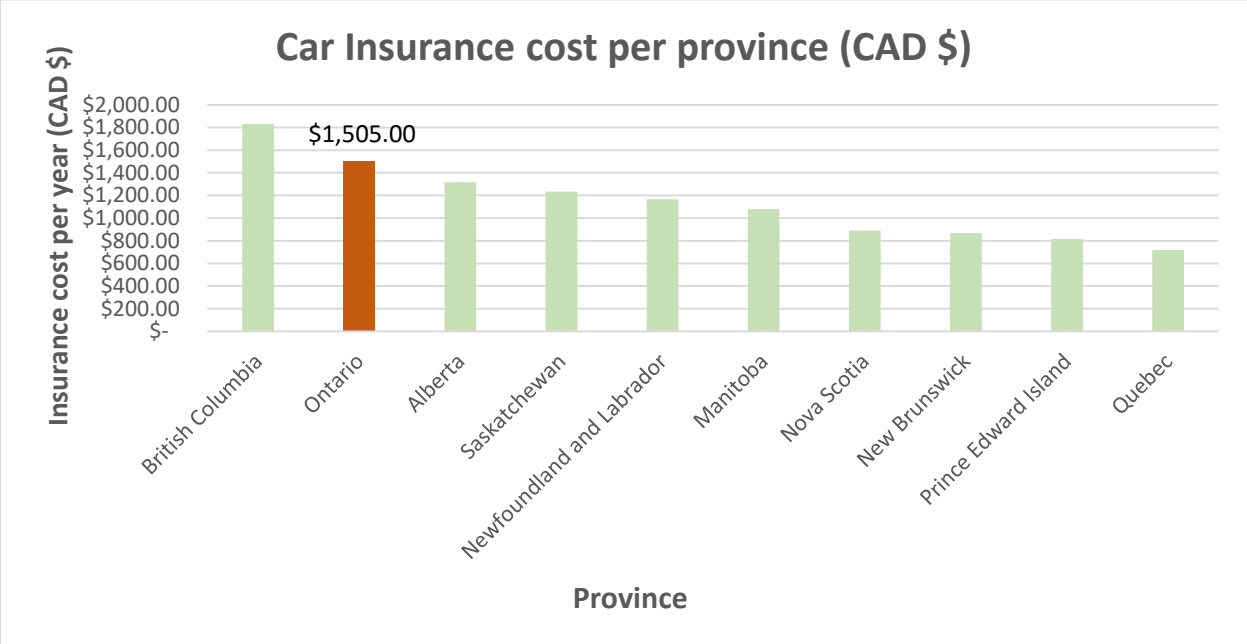
FINDING DATA THAT SUPPORT THE HIGH CAR INSURANCE COST IN TORONTO

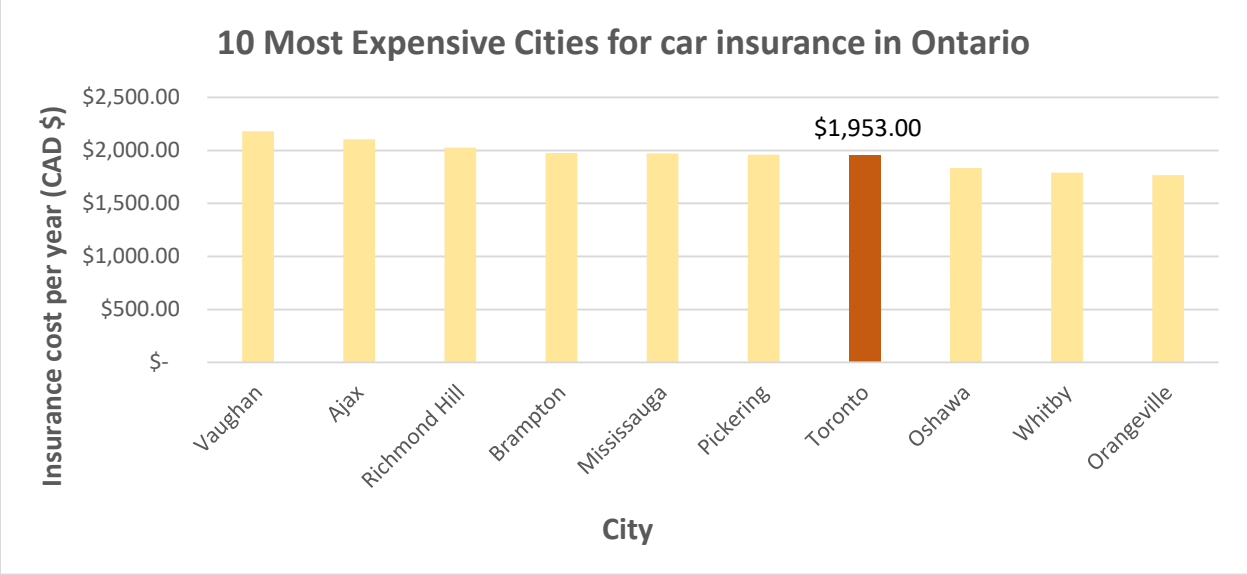
JDataone

May 2022

INTRODUCTION



Source: https://rates.ca/resources/why-ontario-car-insurance-so-expensive



Source: https://rates.ca/resources/these-10-cities-have-highest-car-insurance-rates-ontario

Ontario is the second most expensive province for car insurance in Canada, and Toronto is paying an average of **$ 1953** per year for standard auto policy.This amount of money might represent between 5% to 8% of the family budget. According to ***RATESDOTCA,*** there are some factors that has an impact on the cost of insurance:

- Extreme weather
- Distract driving
- The cost to repair vehicles

Cost of living in Canada could vary between provinces and that could be another factor that boost the insurance cost, as we can see on the graph, the difference between the province with the most expensive insurance car that is British Columbia, and the less expensive (Quebec), is more than a half of the price.

The present analysis we will figure out what is the impact of the extreme weather and distract driving on the insurance cost. We are going to analyse them between others factors to validate if RATESDOTCA is correct on its affirmation or on the other hand, check if there is another factors to take into account that could play an important role on the insurance cost structure.

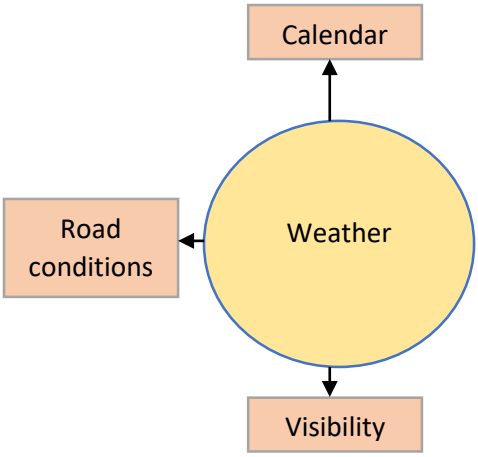
Identifying the most recurrent factors in a car collision, we could develop strategies in order to minimize its impact and help families to save money and most important, **SAVING LIVES**.

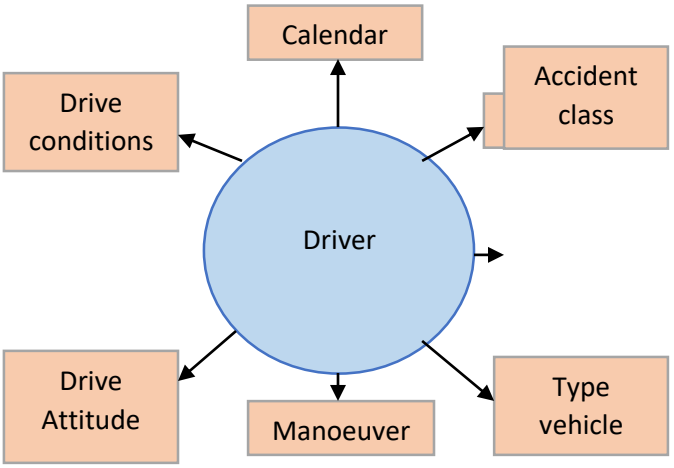
OBJECTIVES

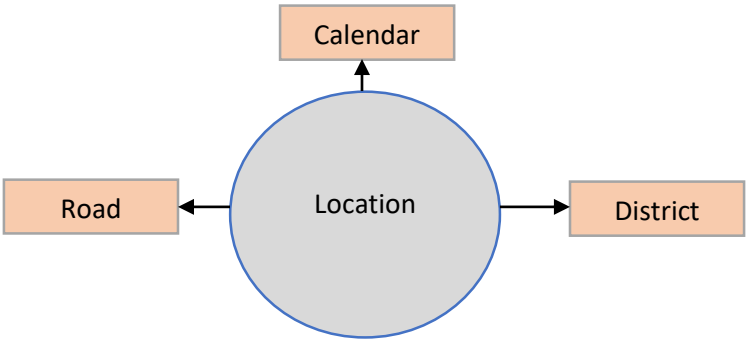
- Identify the season of the year where the car accidents are the most.
- Determine the road and visibility conditions where the car accidents occurred.
- Recognize the Toronto's District with high report of car accidents.
- Observe what kind of type accident is caused in majority for a car accident.
- Identify the different vehicle type with most recurrent car accidents.
- Figure out the driver's conditions and behavior that could trigger out a car accident.

METHODOLOGY

We are going to extrapolate a secondary data from Toronto Open Data website update until 2019. In order to achieve the objectives for this analysis, we will apply Descriptive Statistics to draw conclusions based on the historical data.







The present graph shows how the Data is structured, we have three main parts: weather, driver and location.

Each of these parts consists of different features and sub features as we can see on the next table.

	WEATHER		DRIVER						LOCATION	
Calendar	**Road Conditions**	**Visibility**	**Drive conditions**	**Age**	**Accident class**	**Type vehicle**	**Manoeuver**	**Drive attitude**	**Road**	**District**
	Dry	Clear	Alcohol		Fatal	Automobile	Changing lines	Disobeyed traffic control	Collector	Etobicoke York
	Ice	Drifting snow	Drugs		No fatal	Bicycle	Disable	Driving properly	Highway	North York
	Gravel	Fog	Fatigue			Bus	Going ahead	speed	Laneway	Scarborough
	Loose snow	Freezing rain	Had been drinking			Construction Equipment	Making U Turn	Following too close	Local	Toronto and East York
	Packed snow	rain	Inattentive			Delivery van	Merging	Improper lane change	Major Arterial	
	Slush	snow	Physical disability			Fire vehicle	Overtaking	Improper passing	Minor Arterial	
	Wet	Strong wind	Normal			Intercity bus	Reversing	Improper turn		
						Motorcycle	Turning left	Lost control		
						Municipal bus	Turning right	Speed too fast for conditions		
						Emergency vehicle		Speed too slow		
						Pick up truck		Wrong way in one way road		
						School bus				
						Street car				
						Taxi				

						Tow truck				
						Truck carrier				
						Truck dump				
						Truck tank				
						Truck tractor				

We are going to determine whish of the factors has the most accidents frequency to corroborate the initial hypothesis that is extreme weather and distract driven has the most impact on the car insurance cost structure.

For Extreme weather we are going to use two variables from the data:

Weather

Road Conditions

- ***Ice***
- ***Packed snow***

For Distract Driven:

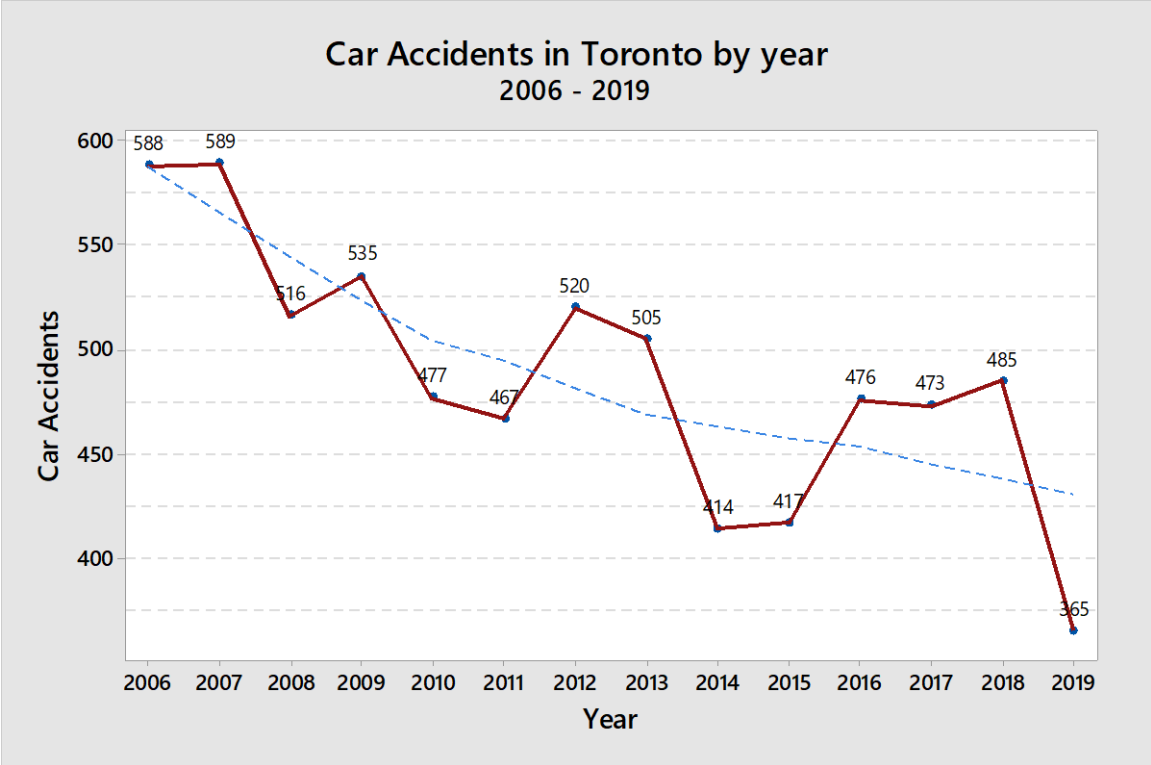
Driver

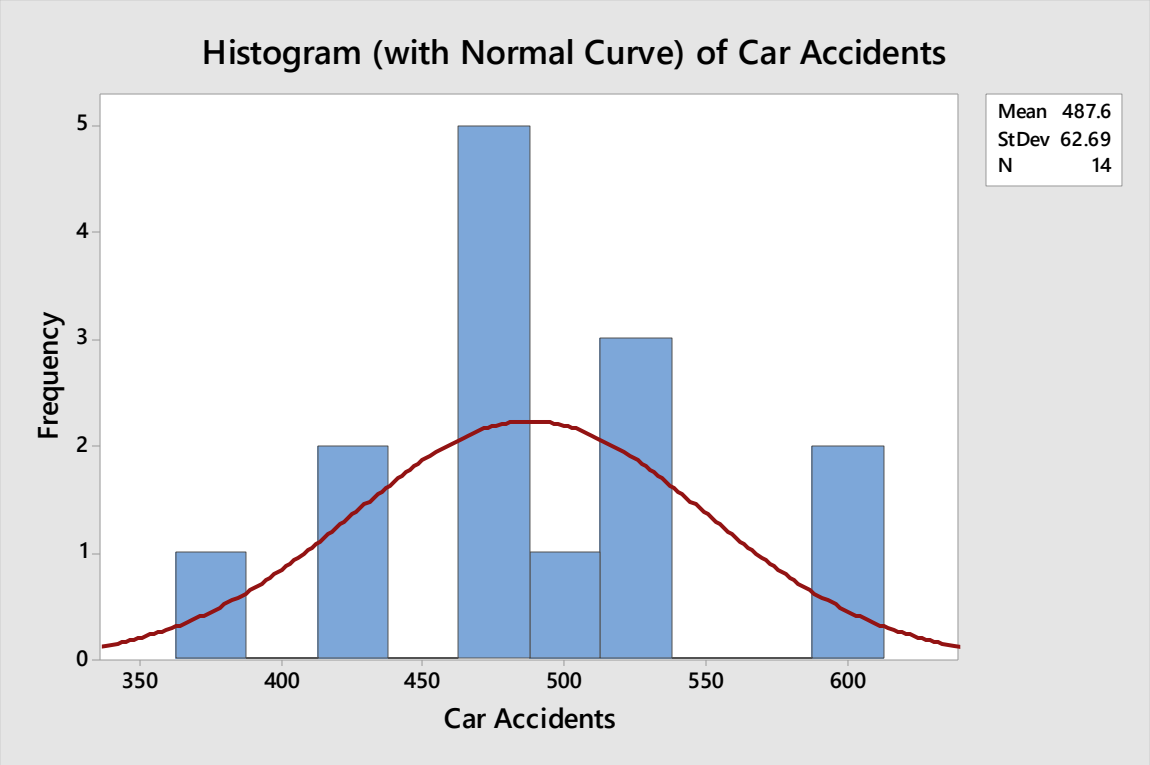
Drive Conditions

- ***Inattentive***

Ice, Packed Snow and Inattentive will be the variables to compare with the rest of variables on the data. This comparison will give us a clue to determine if the initial hypothesis is correct or not.

ANALYSIS


Car Accidents in Toronto by year
2006 - 2019
Car Accidents
Year
600
550
500
450
400
588
589
516
535
477
467
520
505
414
417
476
473
485
365
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

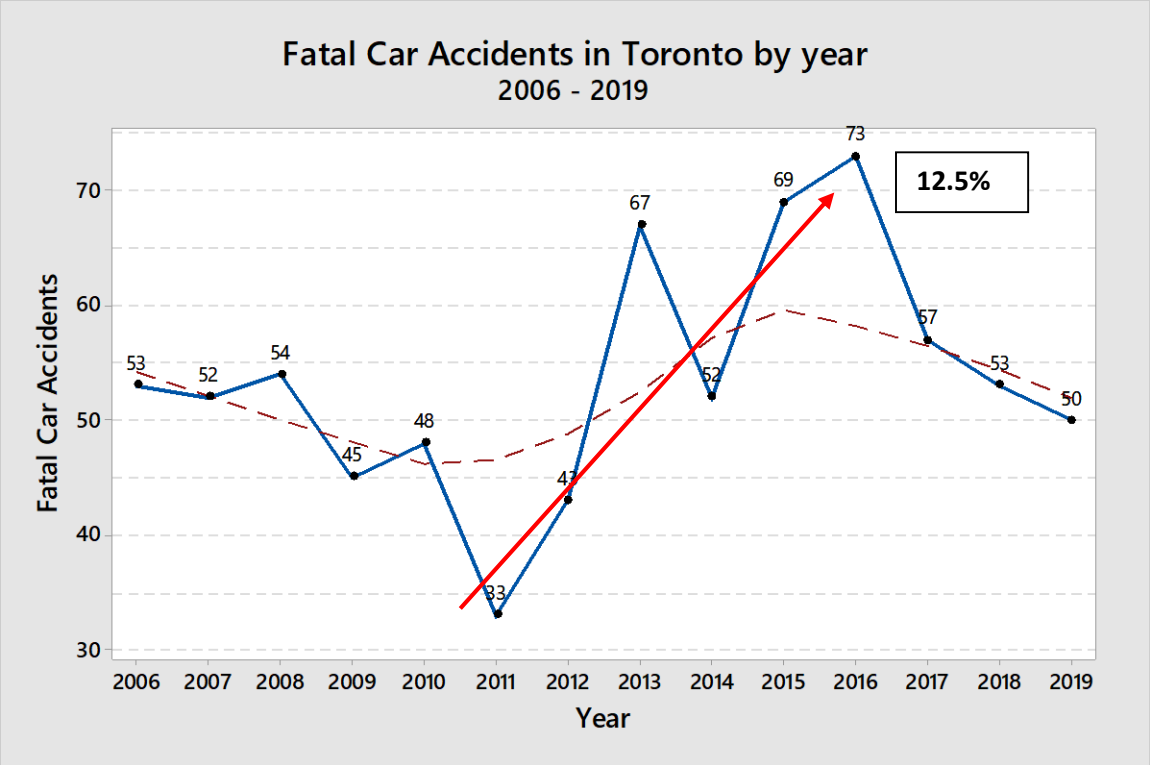


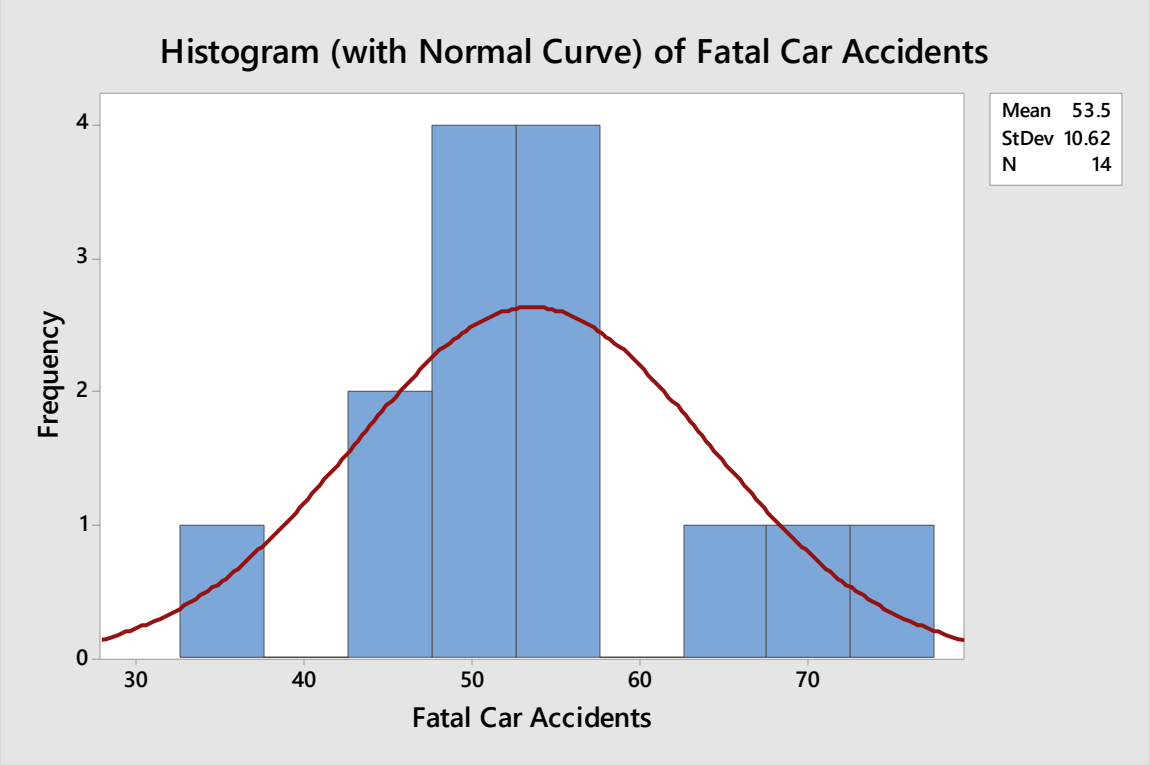
Statistics

Variable	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3
Car Accidents	487.6	16.8	62.7	3930.1	12.86	365.0	454.5	481.0	523.8

Variable	Maximum	Skewness	Kurtosis
Car Accidents	589.0	-0.14	0.08

The number of car accidents in Toronto has been decreased during the las 12 years on a rate of 6.4 %, that is a good sign, however we have an average of 488 per year, it is at least one car accidents per day.

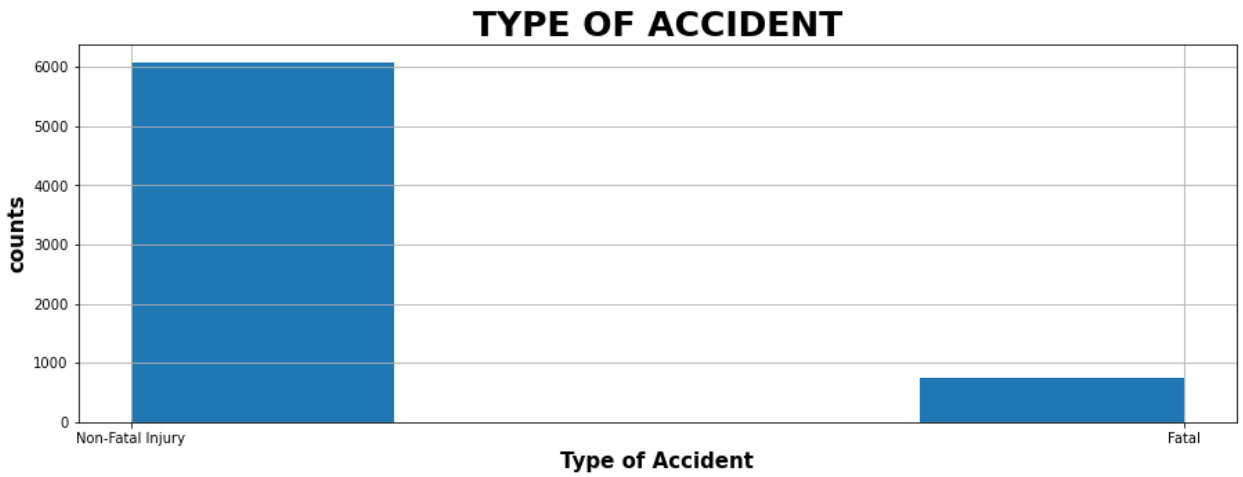

Fatal Car Accidents in Toronto by year
2006 - 2019
Fatal Car Accidents
Year
12.5%
53
52
54
45
48
33
43
67
52
69
73
57
53
50
30
40
50
60
70
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019


Histogram (with Normal Curve) of Fatal Car Accidents
Mean 53.5
StDev 10.62
N 14
Frequency
Fatal Car Accidents
0
1
2
3
4
30
40
50
60
70

Statistics

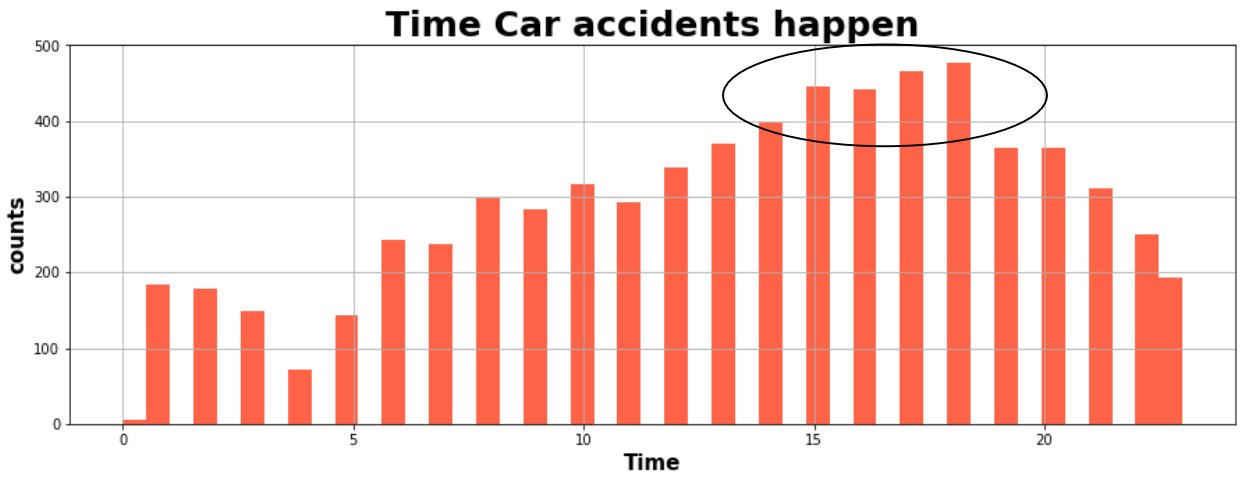
Variable	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3
Car Accidents_1	53.50	2.84	10.62	112.73	19.85	33.00	47.25	52.50	59.50

Variable	Maximum	Skewness	Kurtosis
Car Accidents_1	73.00	0.23	0.31

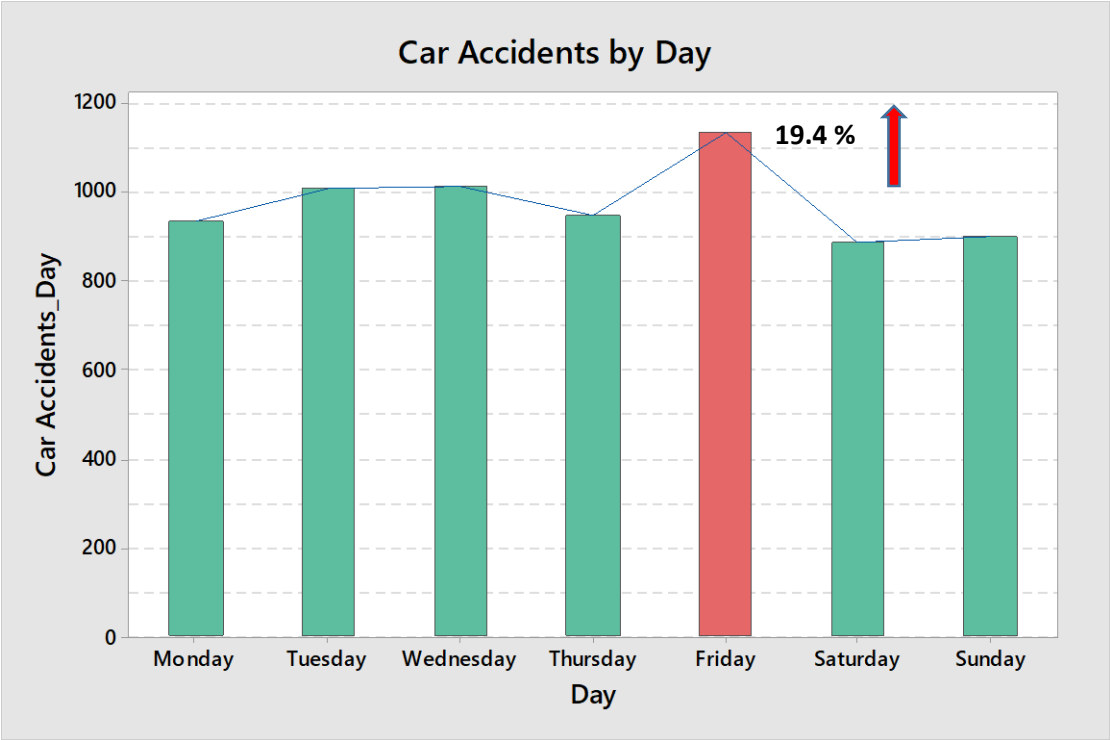


On the contrary, the number of Fatal car accidents present a positive increase trend for 5 consecutive years from 2011 to 2016 with an increase rate of 12.5%, and after that, drop again.

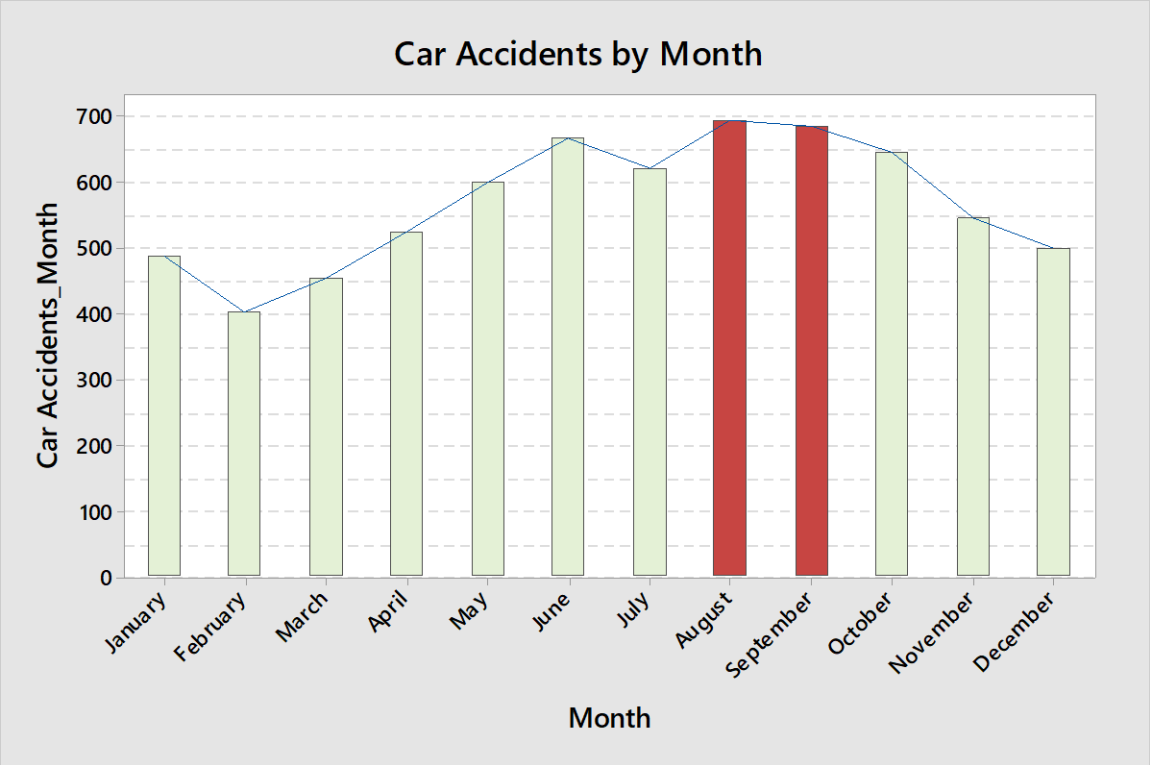
The number of Fatal accidents represent around 12% of the Total with an annual average of 54, that represent one fatal car accident every week.



This graph shows the time at which the accidents happen. We can see that the most of them occurred in Rush time from 15:00 to 18:00.

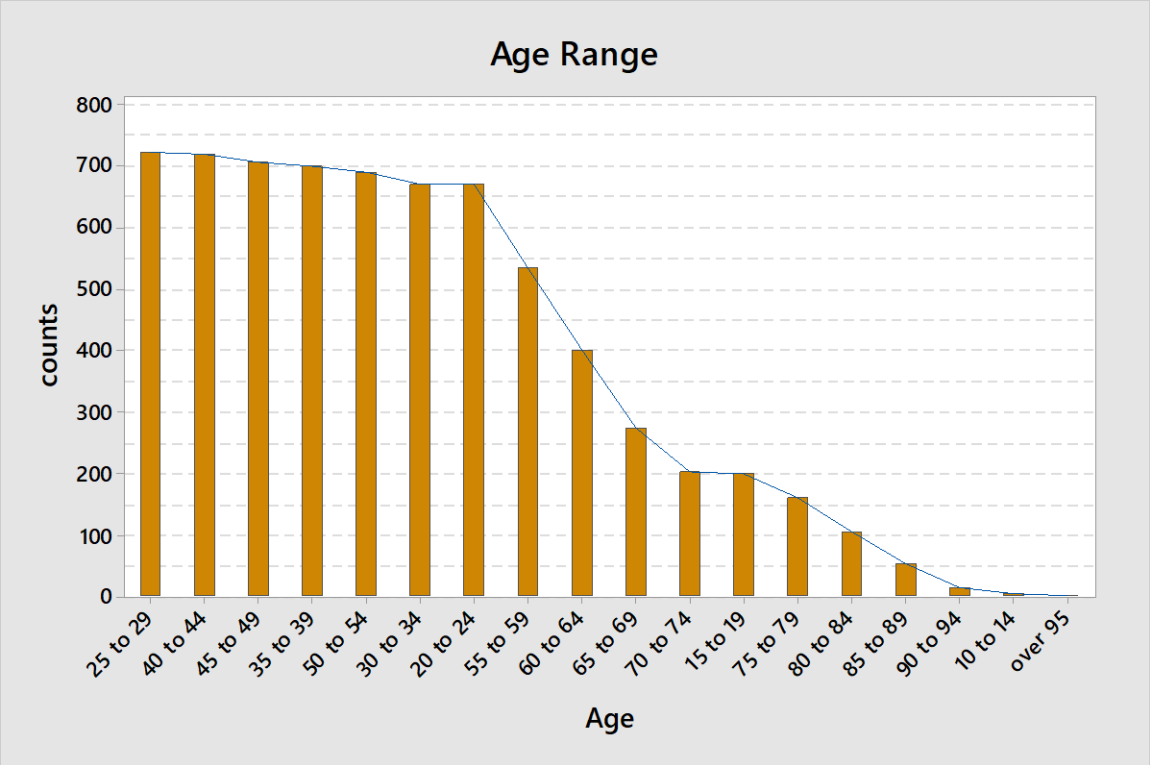


Friday is the day with the most car accidents with a **19.4%** more than the average.

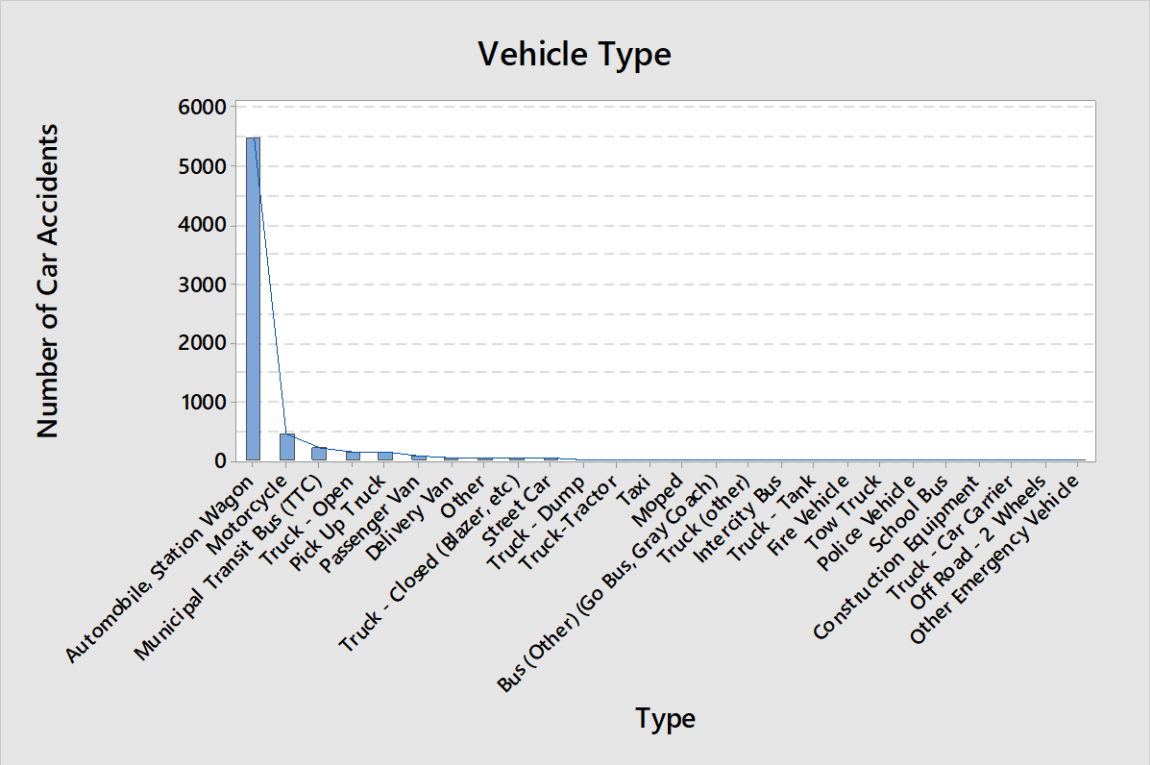


Paradoxically we can think that the largest number of accidents could happen in winter due to weather conditions, but when looking at the graph that shows data of more than ten years, we can see that the largest number of car accidents occur in summer between the months of August and September.

One reason may be that in winter there are fewer cars on the roads, many people prefer to use public transport to avoid accidents due to bad weather.

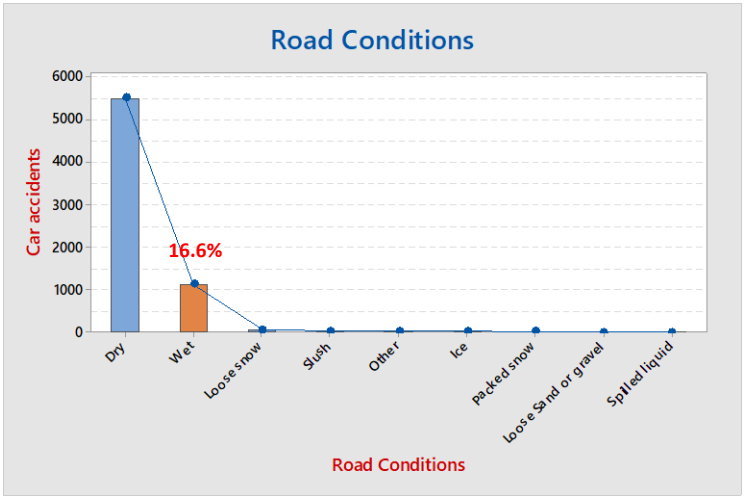


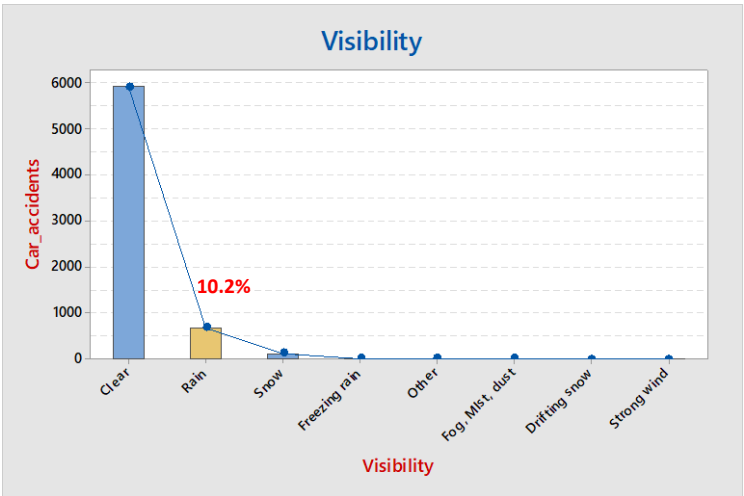
According to the preceding graph, it can be seen that the age range for groups between 25 and 50 years old, group the largest number of car accidents. It should be noted within this group, the age range between 25 and 29 years are those with the most accumulated cases.



In the previous graph it can be seen that the type of vehicle with the highest incidence of traffic accidents falls on automobiles and SW as private cars. They represent **79.86%** of all accidents, a fairly high number, we could say that out of every four vehicles involved in an accident, three belong to this group.

WEATHER



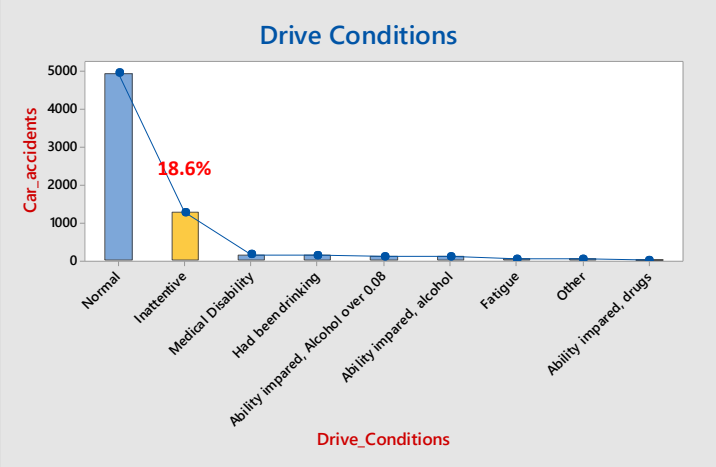


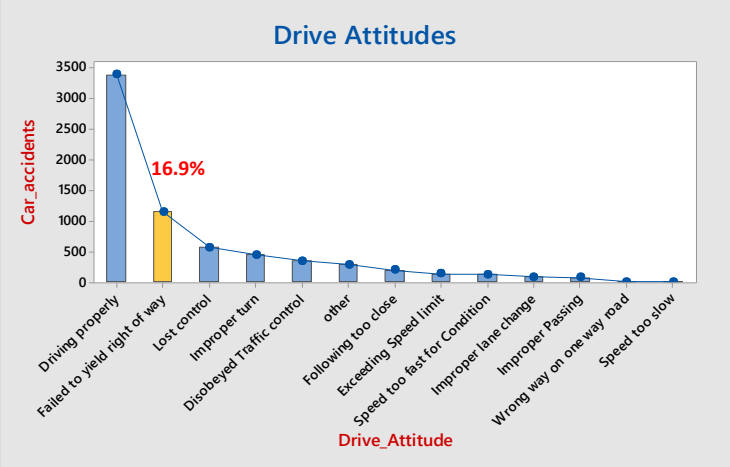
Both graphs show a clear trend where the driving conditions for both, Road Conditions and Visibility are "normal conditions", so we could assume that most accidents occur under these "normal conditions" (Dry and Clear roads), we leave that interpretation to the reader. However, we are going to put emphasis on the second conditions with the highest incidence, which in this case would be ***Wet for Road Conditions***, which represents ***16.6%*** of the total number of accidents, the other conditions are practically null with no greater incidence in the total.

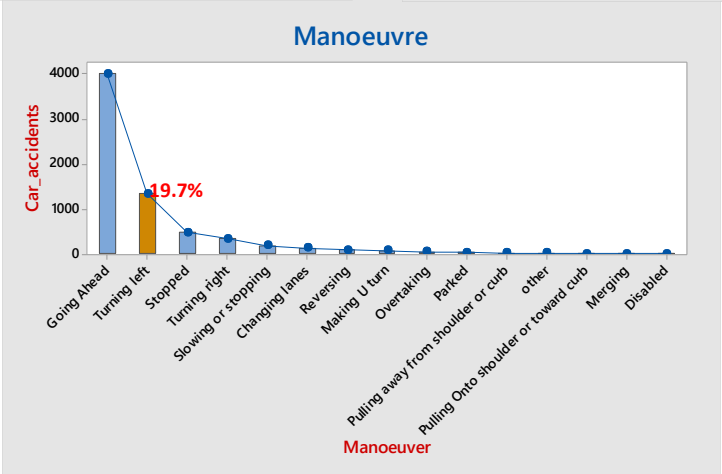
Driving under ***rain conditions represents 10.2%*** of the total cases while the other conditions have an insignificant incidence in the total.

Without taking into account the normal driving conditions, we could conclude that a rainy and wet track affects the driving conditions a lot, which can end in a traffic accident.

DRIVER



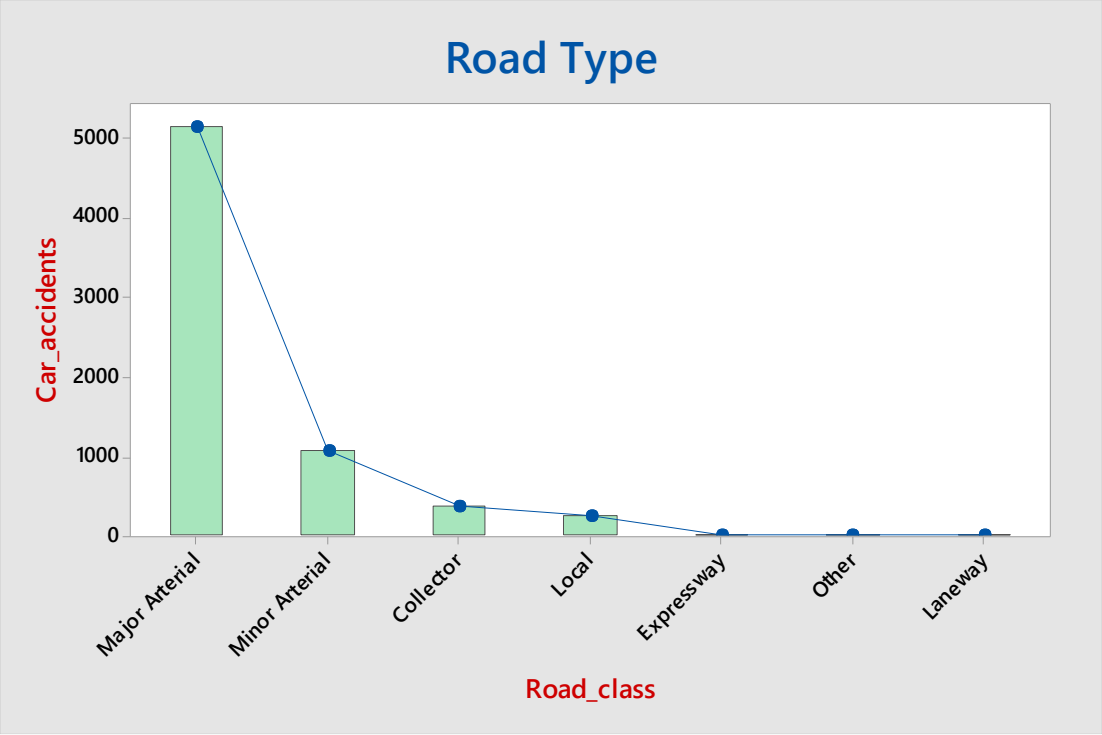


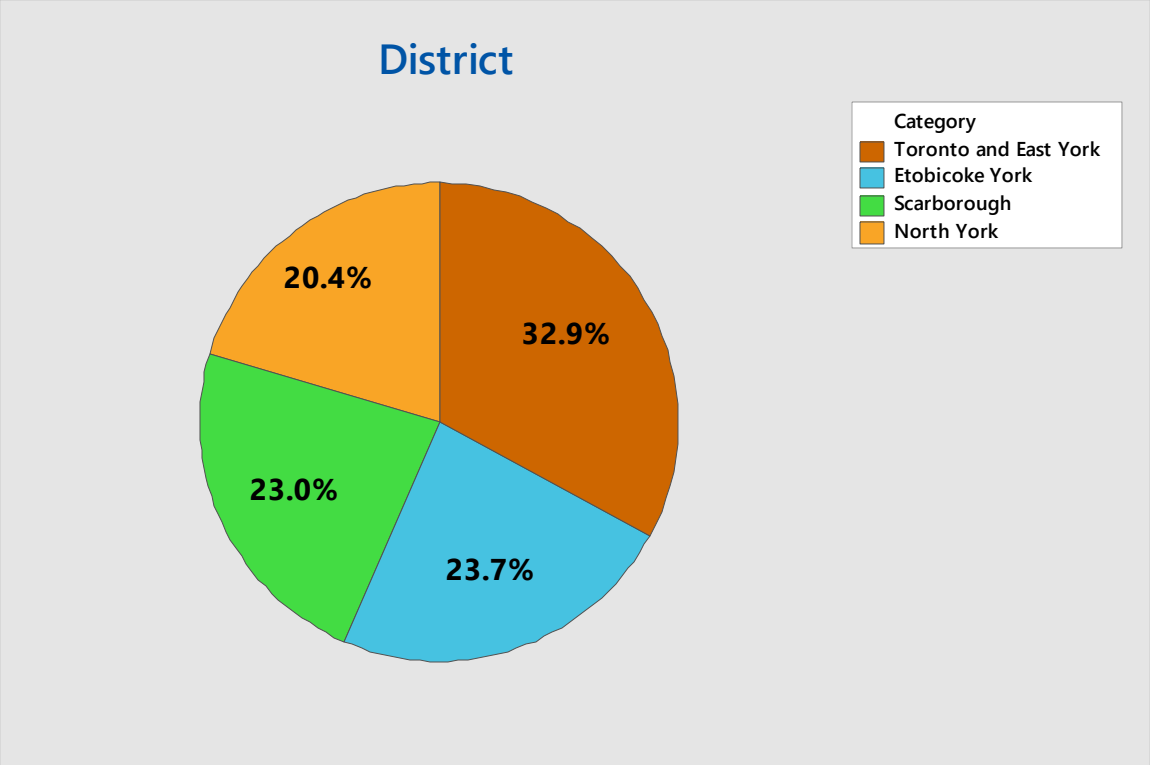


Like the previous "normal" conditions, here we can also see that the same pattern of normality is repeated for the driver's driving conditions, which would be mostly those with the highest number of cases. We leave the interpretation to the reader, but we could assume that there were other reasons that caused the accident under normal driving conditions. We are going to put the lights on the second reasons, which will allow us to have an idea if those reasons affect the cost structure of car insurance.

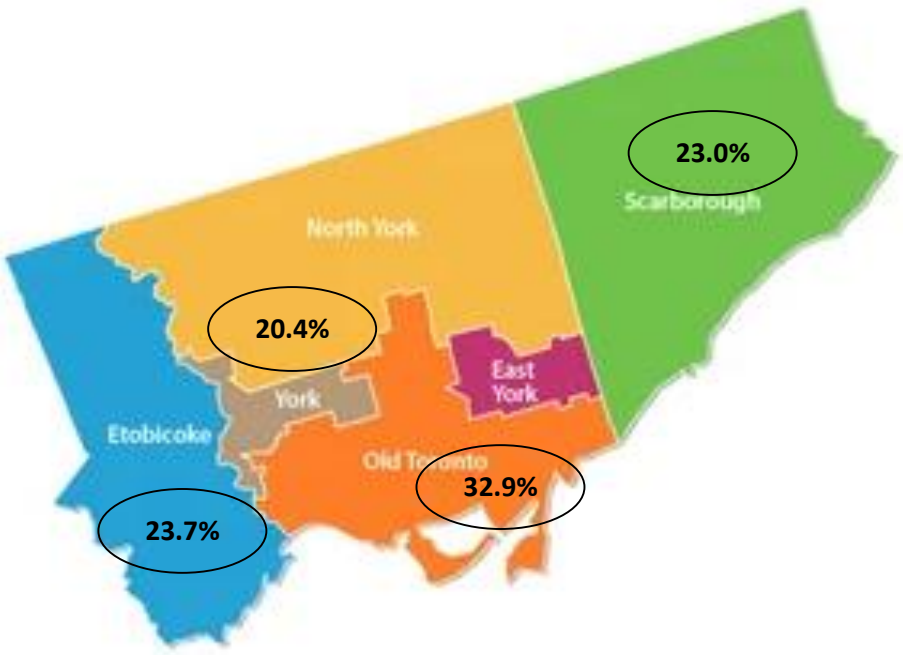
For Drive Conditions we have that Inattentive is in second position with 18.6% of the total cases, which supports the initial hypothesis that this condition does influence the cost structure for car insurance.

LOCATION




District
Category
Toronto and East York
Etobicoke York
Scarborough
North York
20.4%
32.9%
23.0%
23.7%

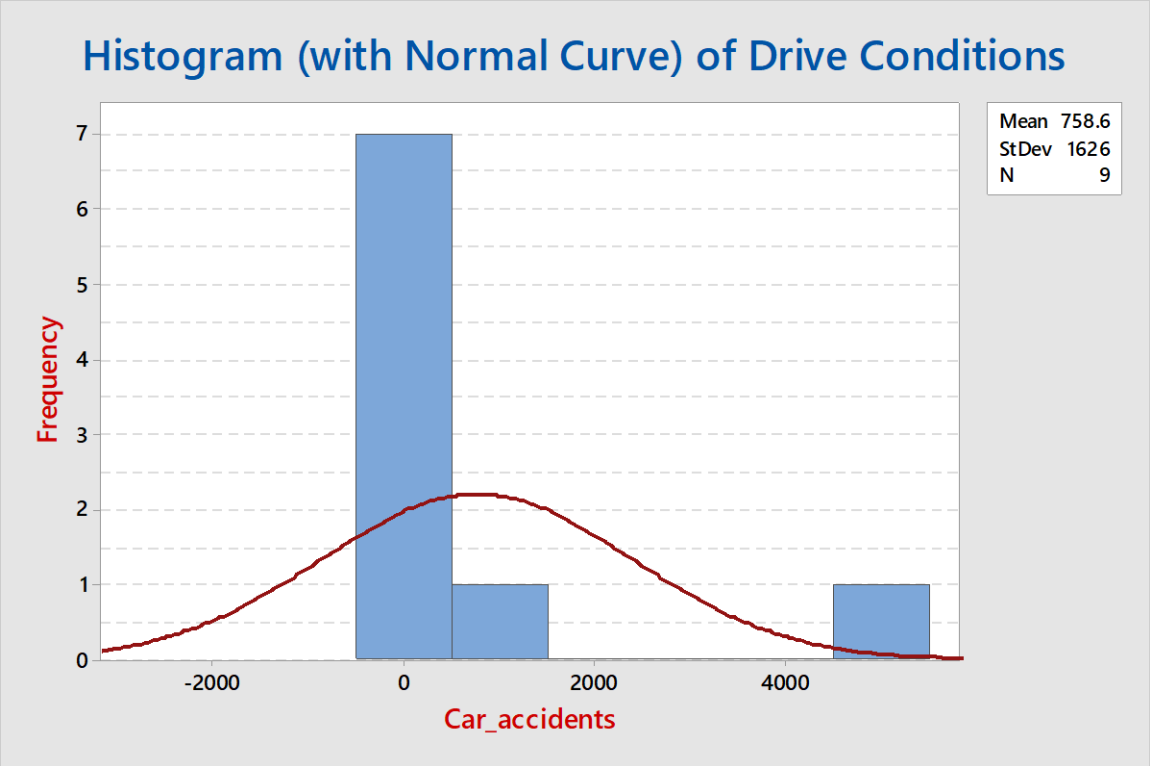
TORONTO


23.0%
Scarborough
North York
20.4%
East York
York
Etobicoke
32.9%
23.7%

Major arterial is the most common type road for car accidents. We are talking about 75.2% of traffic accidents that occur on these roads. This data can make sense if we look at the map of Toronto, we can see that downtown Toronto is the area with the highest number of accidents with 32.9%. Many people travel in their particular vehicles to their workplaces, which generates more traffic that would be the breeding ground for potential subsequent accidents, in addition to that, we can say the greatest number of accidents are recorded in summer and specifically on Fridays at rush hour in Toronto downtown as a concentration point.

HYPOTHESIS TESTING

HYPOTHESIS TESTING: Inattentive Driving has an Impact on the car insurance cost



Let's first see the statistical data of Drive Conditions for the 14 years of study. We can see in that period of time, the variable Driven Conditions has an average of 759 car accidents, so it could be deduced that there would be an average of ***59 accidents per year***

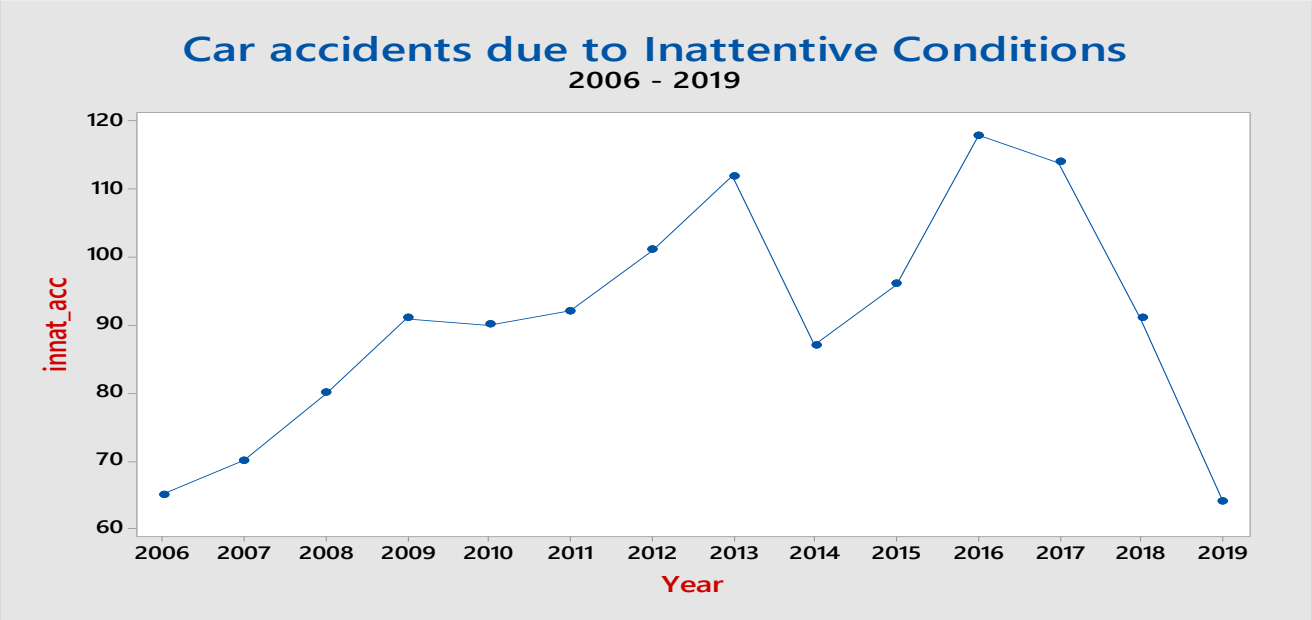
Statistics

Variable	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
Car_accidents_5	759	542	1626	2643138	214.32	15	41	108	713	4964

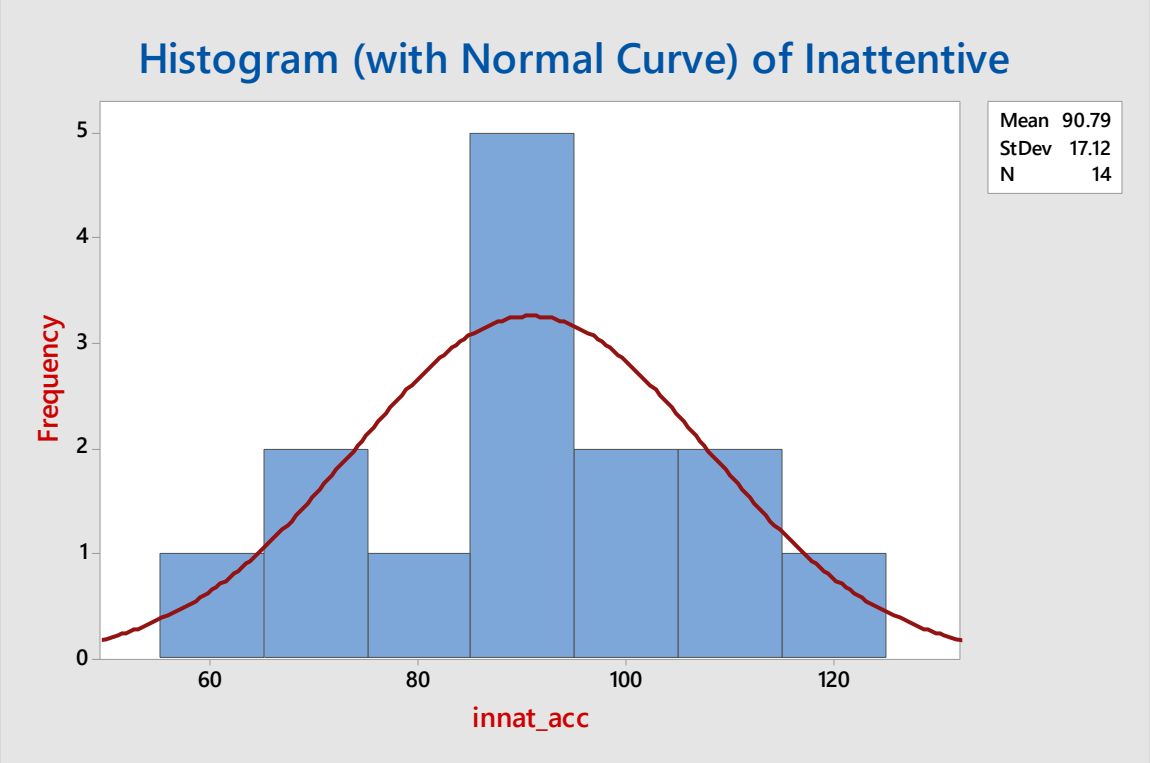
Variable	Skewness	Kurtosis
Car_accidents_5	2.71	7.48

Drive Conditions - Inattentive: Number of car accidents per year due to this variable

2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
65	70	80	91	90	92	101	112	87	96	118	114	91	64



Now let's analyze the variable Inattentive which has an average of **91 accidents** following a normal distribution with Skewness and Kurtosis values acceptable for normal distributions as we can see in the following graph.



Statistics

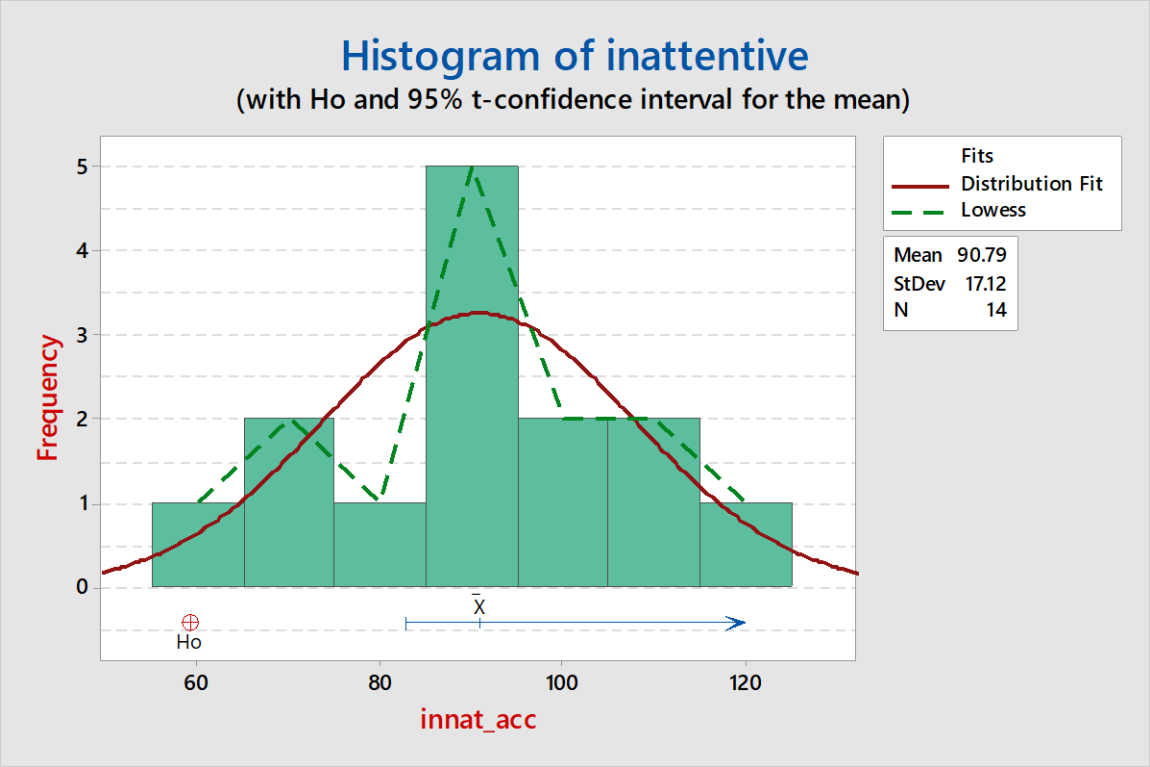
Variable	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
innat_acc	90.79	4.57	17.12	292.95	18.85	64.00	77.50	91.00	103.75	118.00

Variable	Skewness	Kurtosis
innat_acc	-0.06	-0.69

To analyze if the Inattentive variable has an impact on the distribution of car insurance costs, we will use One-Sample hypothesis testing. Test for the mean: population Variance Unknown that involves a single population parameter. As a population Mean we are going to use the number of car accidents of Drive Conditions as a population that is 59.

H0: Inattentive has no Impact on the cost structure.

H1: Inattentive has an Impact on the cost structure.



One-Sample T: innat_acc

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
14	90.79	17.12	4.57	(80.90, 100.67)

μ: mean of innat_acc

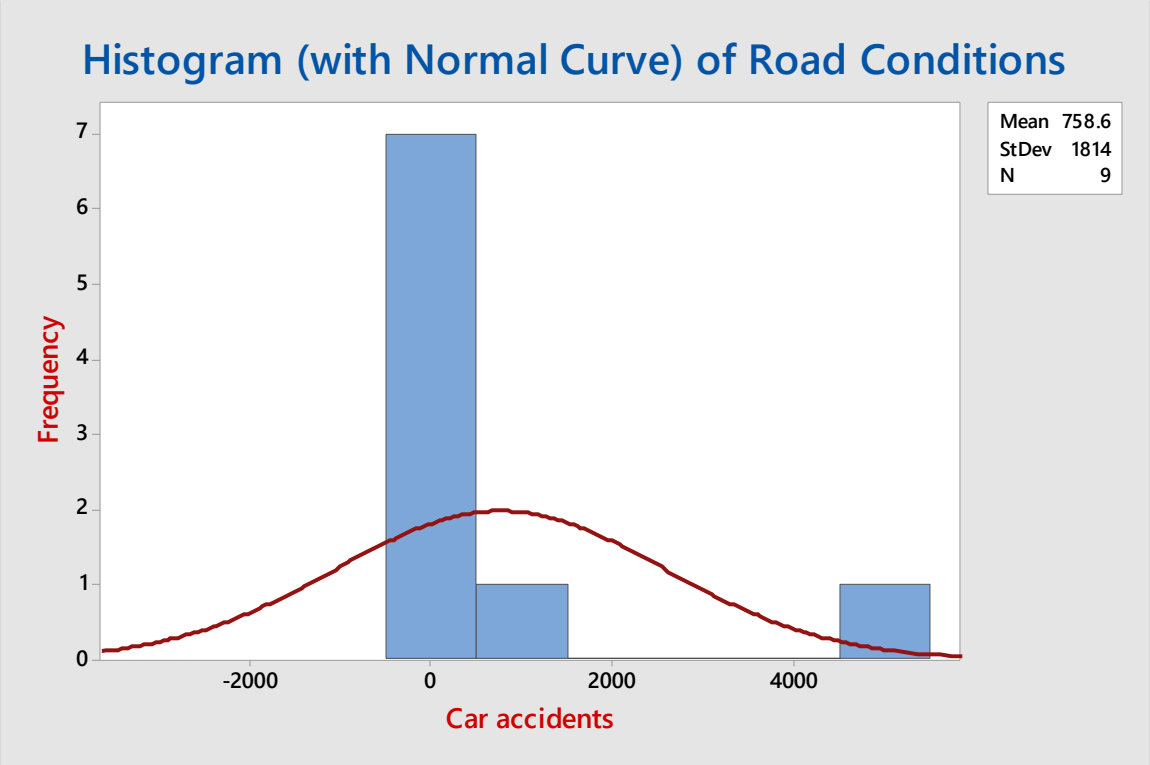
Test

Null hypothesis	H_0: μ = 59
Alternative hypothesis	H_1: μ ≠ 59

T-Value	P-Value
6.95	0.000

After running the Test in Minitab, **we got a P-Value of zero**, which forces us to reject the null hypothesis and accept the alternate hypothesis which indicates that the Inattentive variable does have implications for the cost structure. This due to the number of car accidents is greater than the annual average of the population taken.

HYPOTHESIS TESTING: Ice on the Road has an Impact on the car insurance cost



Let's first look at the Statistical conditions for the variable Road Conditions, which presents an average of 759 accidents for 14 years of studies, which would give us an annuity of **59 accidents per year**.

Descriptive Statistics: Car accidents_3

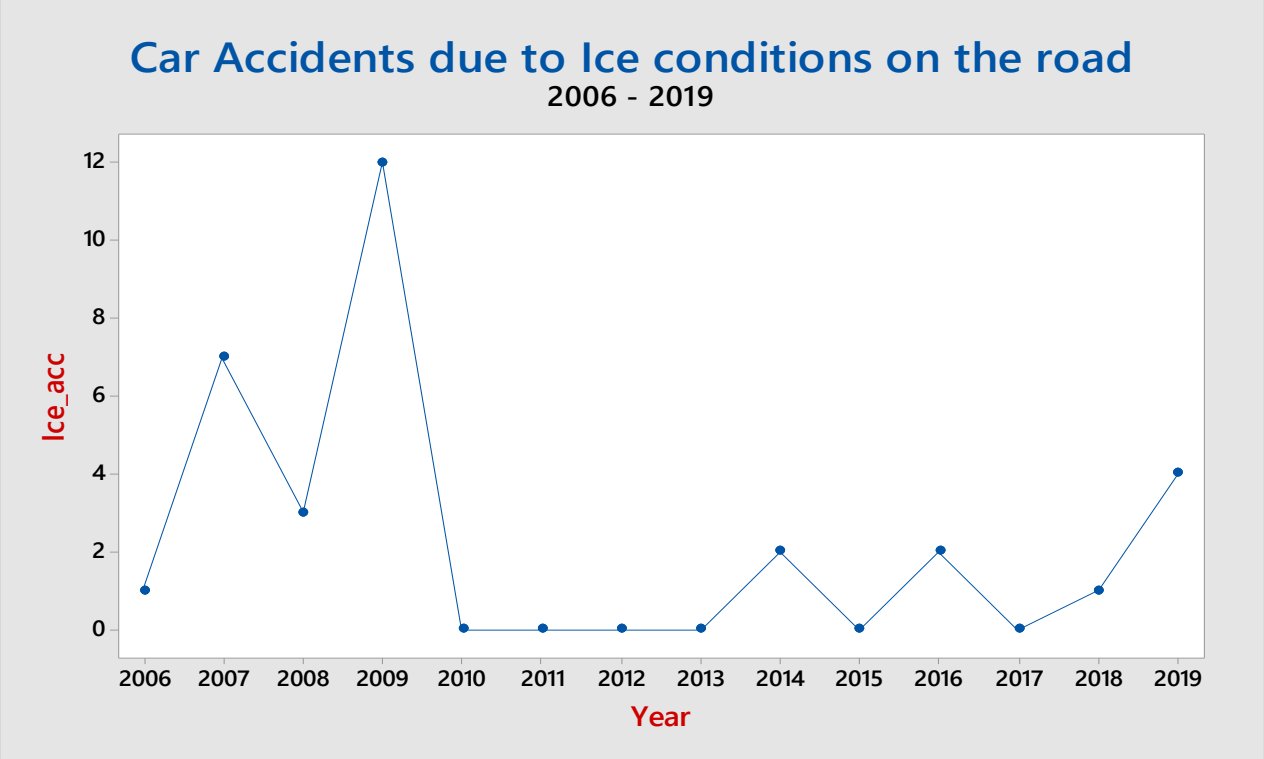
Statistics

Variable	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
Car accidents_3	759	605	1814	3291840	239.18	1	12	37	598	5497

Variable	Skewness	Kurtosis
Car accidents_3	2.79	7.94

Road Conditions - Ice: Number of car accidents per year due to this variable

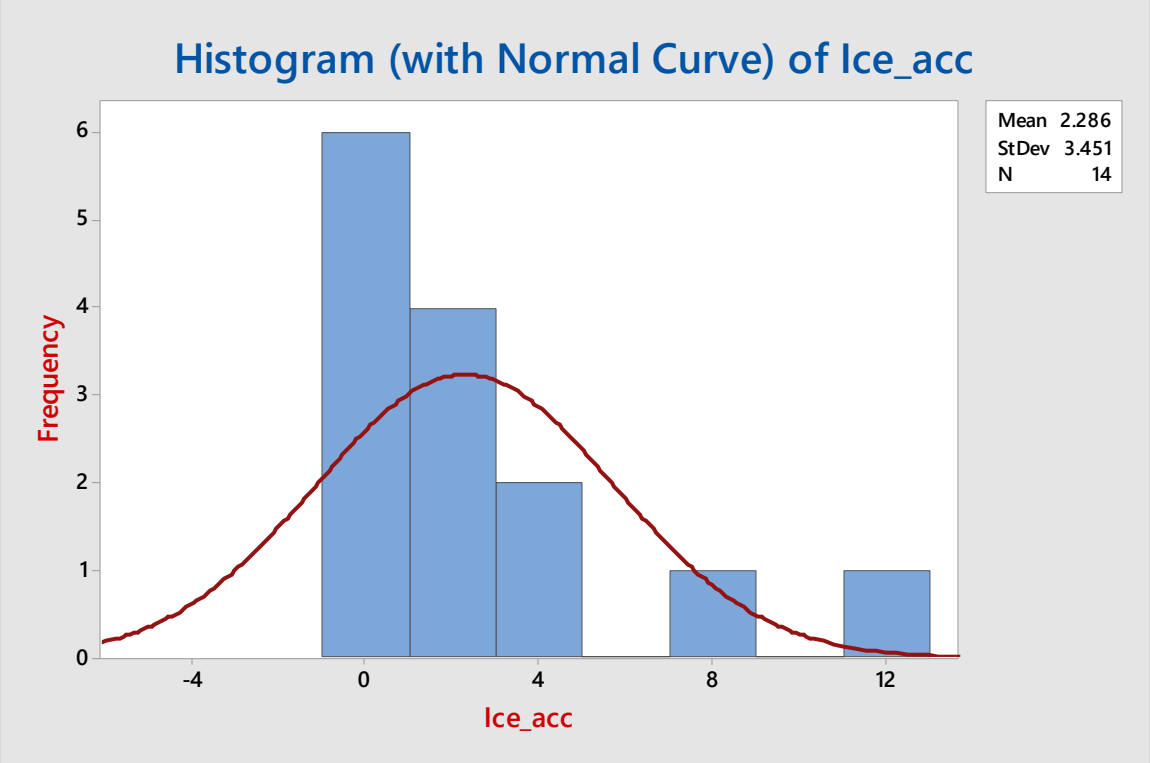
2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	7	3	12	0	0	0	0	2	0	2	0	1	4



As in the previous analysis, we are going to use One-Simple hypothesis testing. Test for the mean: Population Variance Unknown that involves a single population parameter. As a population Mean we are going to use the number of car accidents of Road Conditions as a population that is 59.

H0: Ice Conditions has no Impact on the cost structure.

H1: Ice Conditions has an Impact on the cost structure.

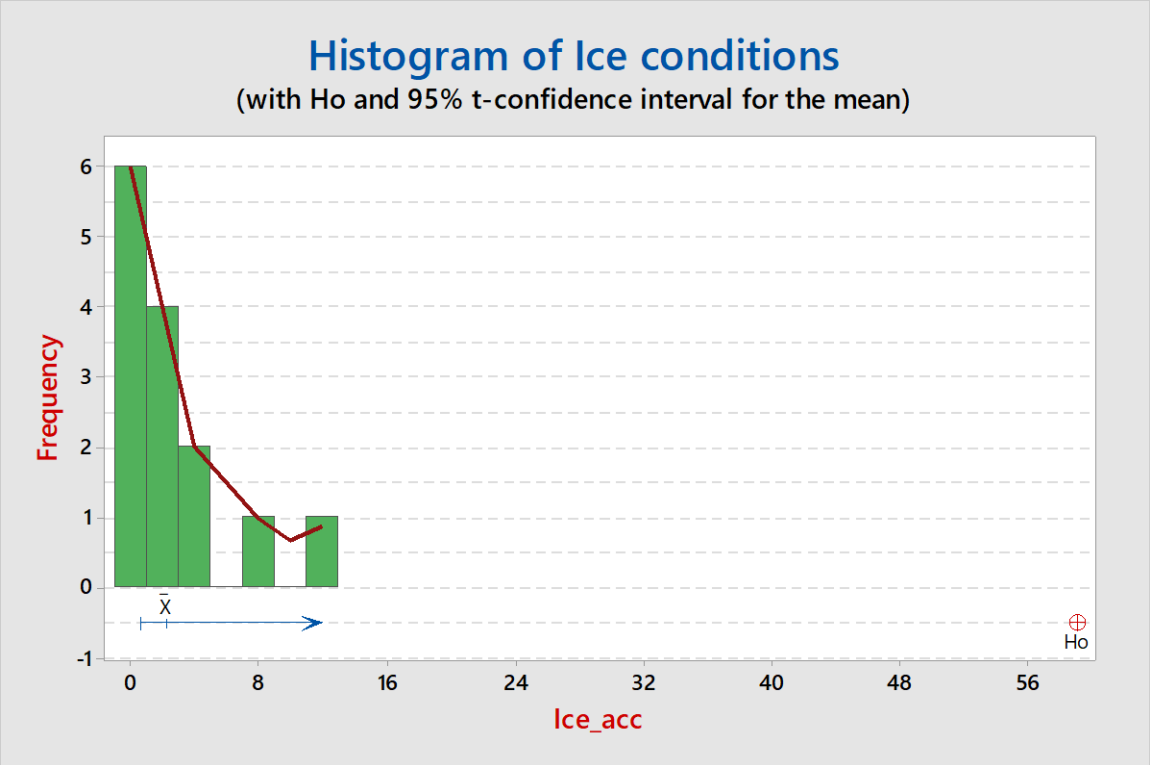


Descriptive Statistics: Ice_acc

Statistics

Variable	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
Ice_acc	2.286	0.922	3.451	11.912	151.00	0.000	0.000	1.000	3.250	12.000

Variable	Skewness	Kurtosis
Ice_acc	2.08	4.42



One-Sample T: Ice_acc

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for μ
14	2.286	3.451	0.922	0.652

μ: mean of Ice_acc

Test

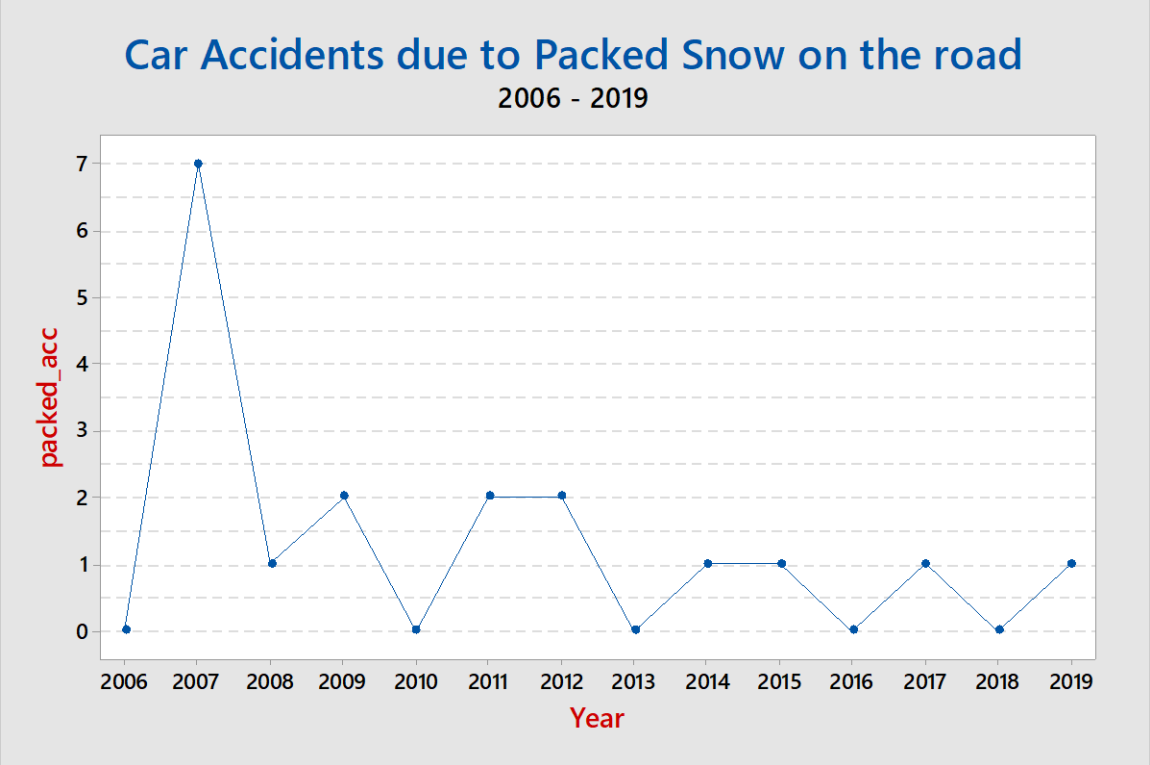
Null hypothesis	H_0: μ = 59
Alternative hypothesis	H_1: μ > 59

T-Value	P-Value
-61.48	1.000

After running the Test in Minitab, **we got a P-Value of one**, which forces us to accept the null hypothesis and reject the alternate hypothesis which indicates that the Ice Road Conditions does not have implications for the cost structure. This due to the number of car accidents is lower than the annual average of the population taken.

HYPOTHESIS TESTING: Packed Snow on the Road has an Impact on the car insurance cost

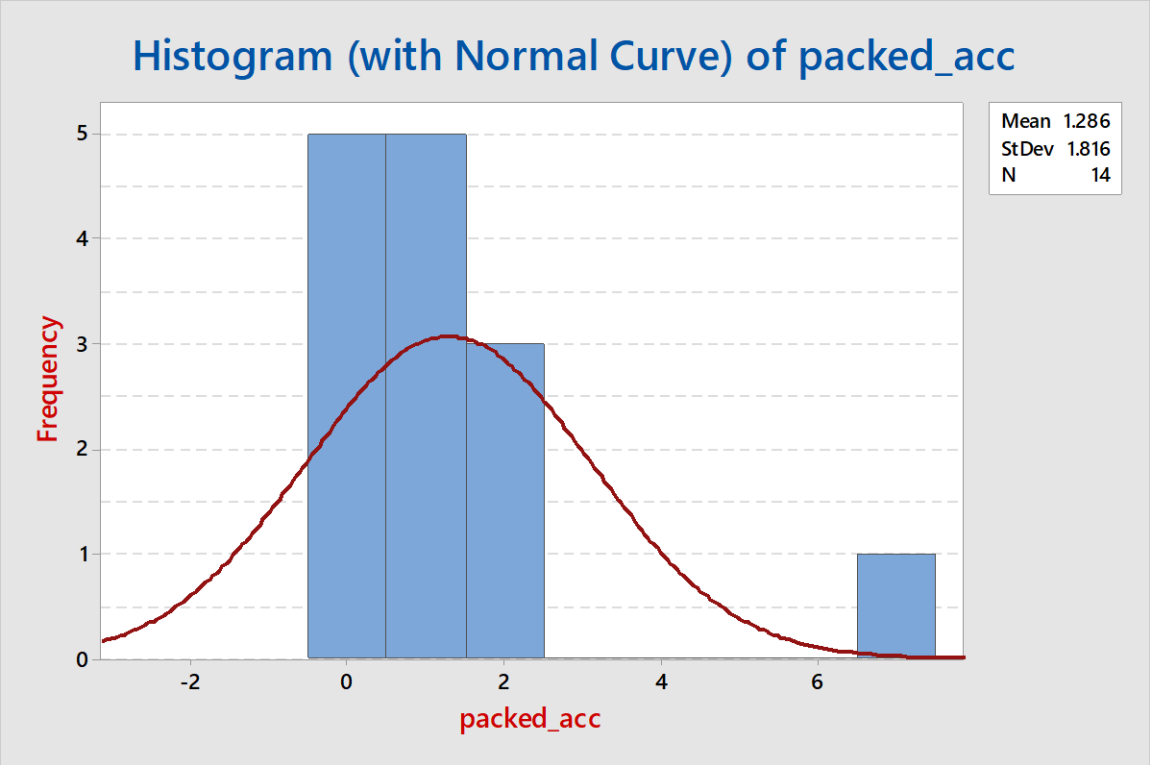
2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	7	1	2	0	2	2	0	1	1	0	1	0	1



We are going to use One-Simple hypothesis testing. Test for the mean: Population Variance Unknown that involves a single population parameter. As a population Mean we are going to use the number of car accidents of Road Conditions as a population that is 59.

H0: Packed Snow Conditions has no Impact on the cost structure.

H1: Packed Snow Conditions has an Impact on the cost structure.

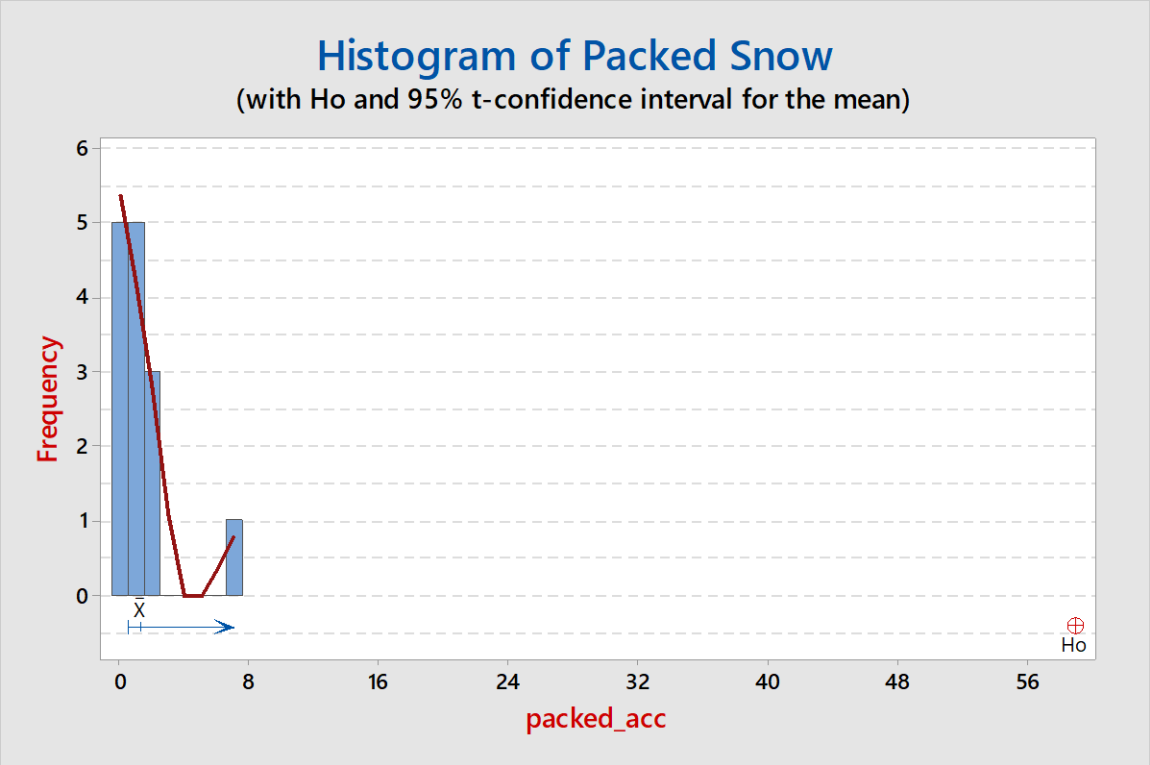


Descriptive Statistics: packed_acc

Statistics

Variable	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
packed_acc	1.286	0.485	1.816	3.297	141.22	0.000	0.000	1.000	2.000	7.000

Variable	Skewness	Kurtosis
packed_acc	2.65	8.33



One-Sample T: packed_acc

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for μ
14	1.286	1.816	0.485	0.426

μ: mean of packed_acc

Test

Null hypothesis	H_0: μ = 59
Alternative hypothesis	H_1: μ > 59

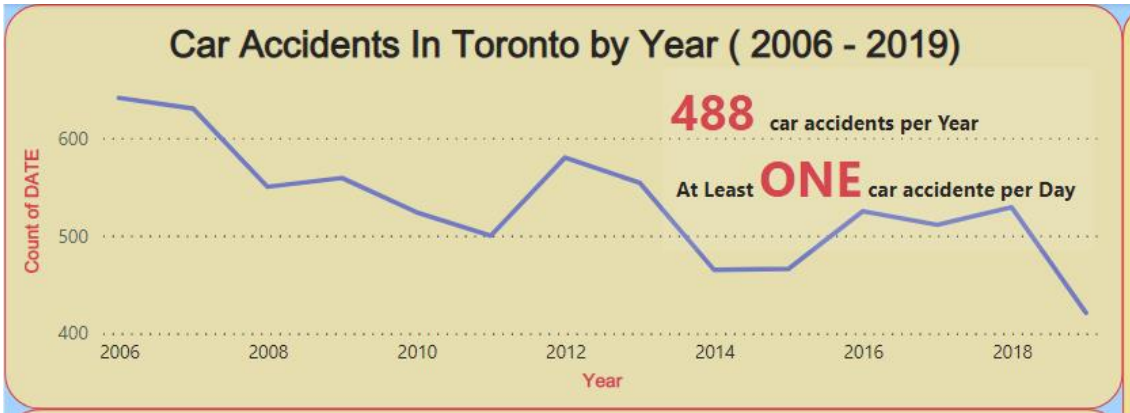
T-Value	P-Value
-118.93	1.000

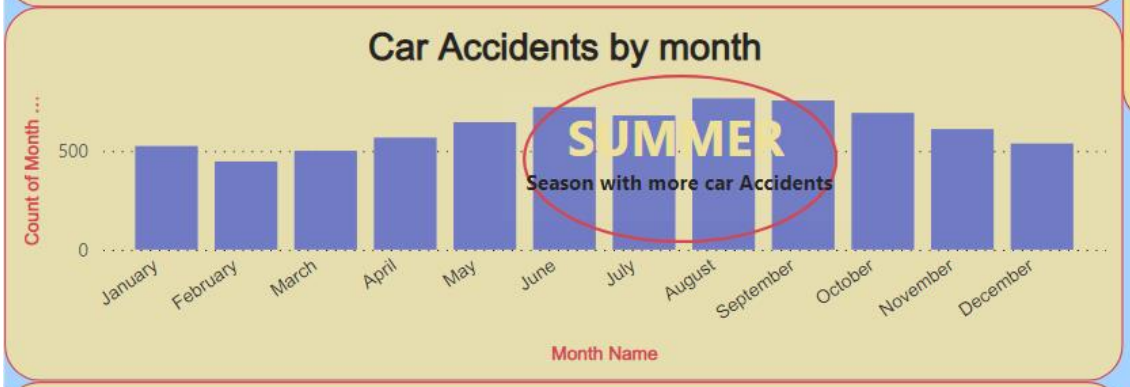
After running the Test in Minitab, **we got a P-Value of one**, which forces us to accept the null hypothesis and reject the alternate hypothesis which indicates that the Packed Snow does not have implications for the cost structure. This due to the number of car accidents is lower than the annual average of the population taken.

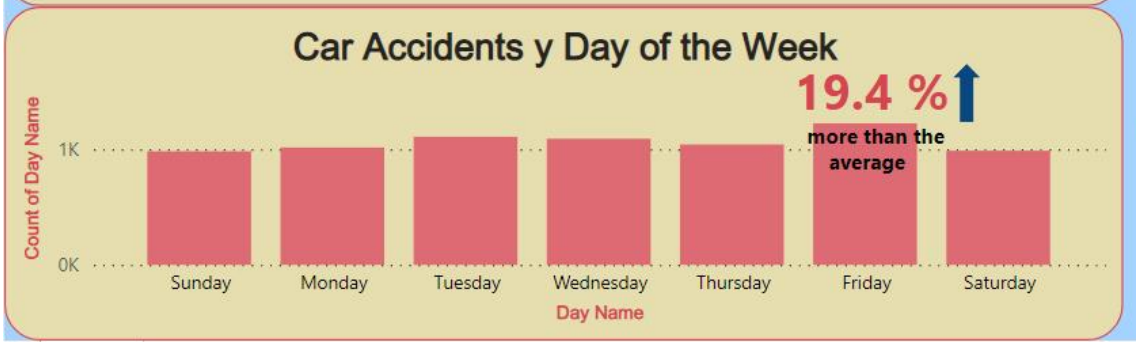
CONCLUSIONS

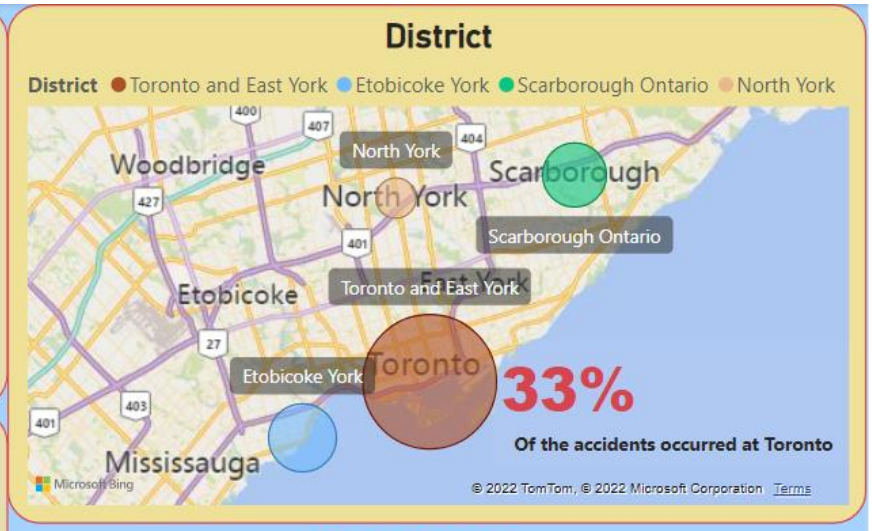
- The annual budget required for car insurance in Toronto is in the range of 5% to 8% of the family budget.
- Insurance companies claim that extreme weather and distracted driving influence the final cost of insurance, which according to the analysis carried out is not entirely true.
- The average of 488 per year, it is at least one car accidents per day.
- the average number of accidents per year is 488, from which it can be deduced that we could easily speak of at least one accident per day.
- The number of car accidents in Toronto has been decreased during the las 12 years on a rate of 6.4 %.
- The number of Fatal accidents represent around 12% of the Total with an annual average of 54, that represent one fatal car accident every week.
- According to the analysis carried out, most accidents occur in the months of August and September, that is, in Summer. We could assume that in Summer there is more vehicular load than in Winter and in the month of September the school year begins, which leads to an increase in traffic. Continuing with more detailed analysis, we find that on Fridays at rush hour the number of accidents increases in 19.4%.
- The age group with the largest number of cases is in the range of 25 to 50 years.
- In 79.86% of car accidents, private cars and SW (Station Wagon) are involved.
- Wet and Rain conditions represent 27% of traffic accidents.
- Inattentive count with 18.6% of the total cases, which supports the initial hypothesis that this condition does influence the cost structure for car insurance.
- We are talking about 75.2% of traffic accidents that occur on these roads. This data can make sense if we look at the map of Toronto, we can see that downtown Toronto is the area with the highest number of accidents with 32.9%.
- The largest number of traffic accidents are concentrated in downtown Toronto with 39% of cases, of which 75.2% occur in Major Arterial.

- The hypothesis test results in a P-Value of zero for the Inattentive variable, rejecting the null hypothesis and accepting the alternative, which indicates that the inattentive variable has an direct impact on the car insurance cost structure.
- The P-value for Ice Road and Packed snow is one, so the null hypothesis is accepted, which indicates that both variables have no significance in the cost structure of car insurance.









P-VALUE

INATTENTIVE	ICE ON THE ROAD	PACKED SNOW ON THE ROAD
0.0000	1.0000	1.0000
Has an Impact on the Car Insurance Cost.	Has not an Impact on the Car Insurance Cost.	

JDataone

REFERENCE

- *Toronto Open Data: https://open.toronto.ca/dataset/bicycle-thefts/*
- *An Introduction to Management Science: Anderson, Sweeney.*
- *The Analytics Life Cycle Tool Kit: Gregory S. Nelson.*
- *Managerial Statistics: Gerald Keller.*
- *Business Analytics: James R. Evans*

Jaime Alvarado

Data Analyst