BARCELONA ACCIDENTS DATA

Brief Description of Data:

Rows : 10339
Columns : 15

Features :

['Id', 'District_Name', 'Neighborhood_Name', 'Street', 'Weekday', 'Month', 'Day', 'Hour', 'Part_of_the_day', 'Mild_injuries', 'Serious_injuries', 'Victims', 'Vehicles_involved', 'Longitude', 'Latitude']

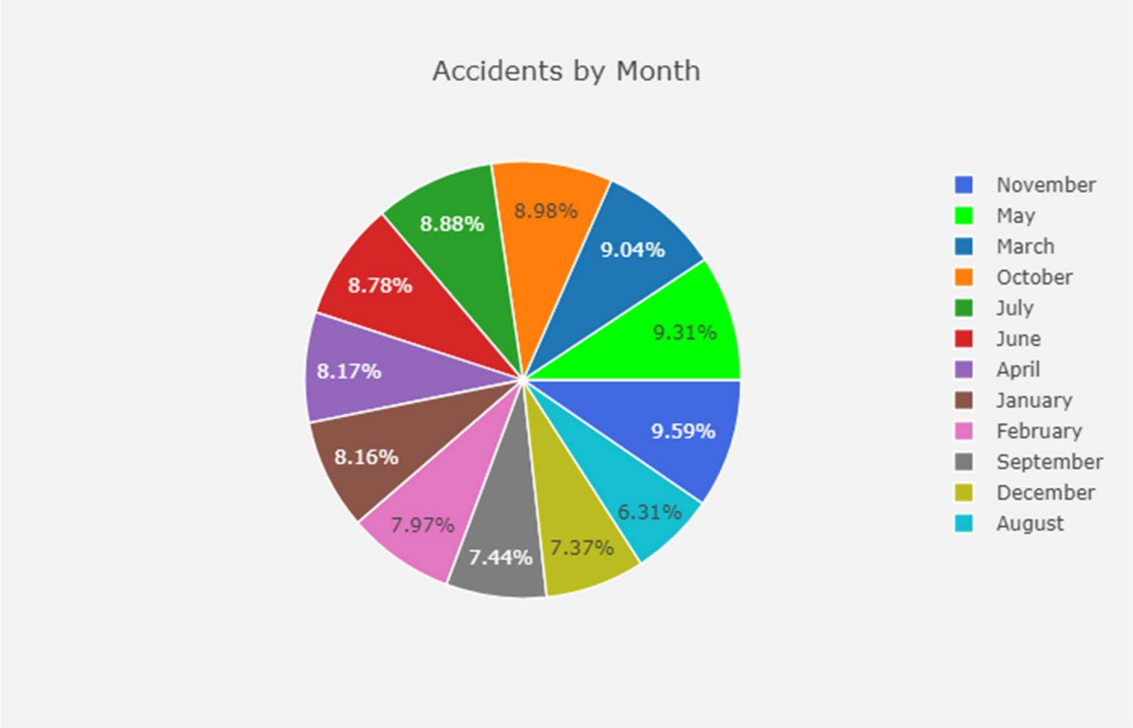
Missing Values: 0

Unique Values:

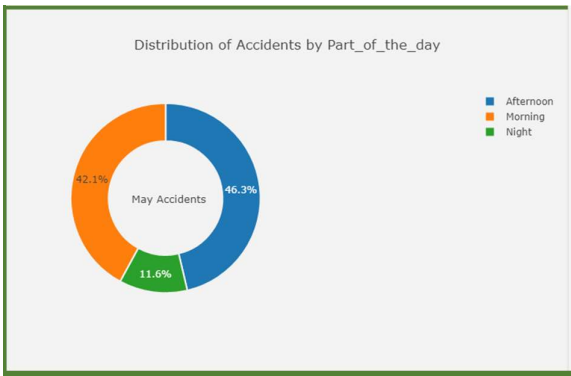
Id	10335
District_Name	11
Neighborhood_Name	74
Street	4253
Weekday	7
Month	12
Day	31
Hour	24
Part_of_the_day	3
Mild_injuries	11
Serious_injuries	4
Victims	11
Vehicles_involved	14
Longitude	5492
Latitude	5442

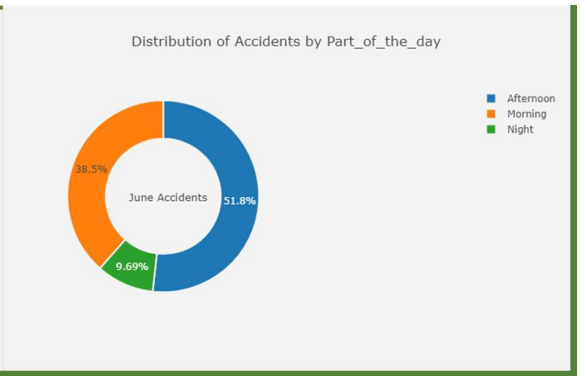
dtype: int64

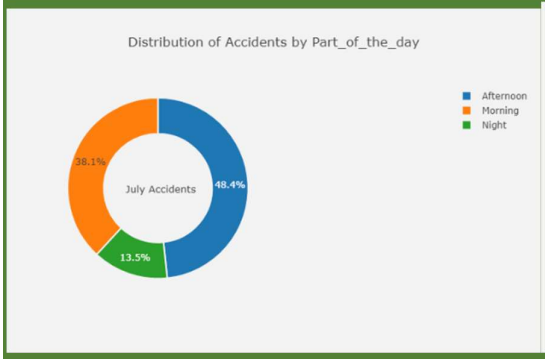
DATA VISUALIZATION

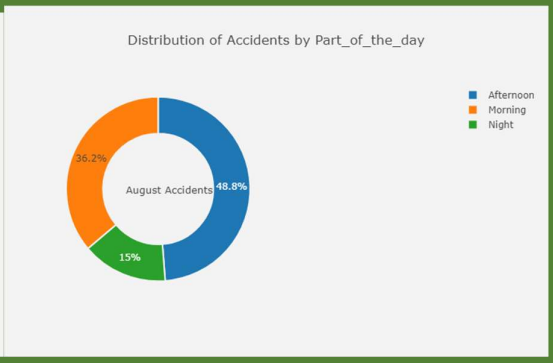


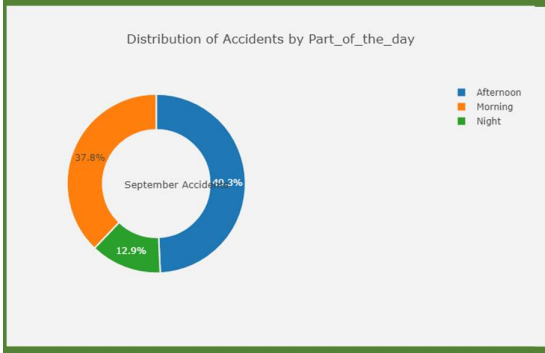
MONTHLY DISTRIBUTION OF ACCIDENTS BY PART_OF_DAY

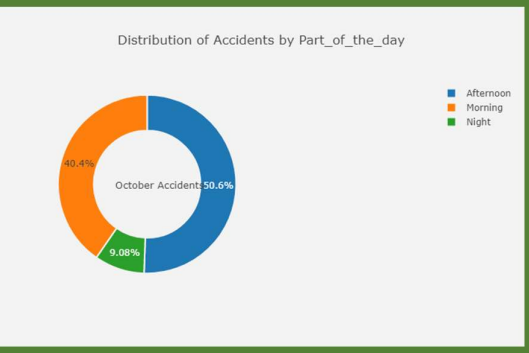

Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
May Accidents
46.3%
42.1%
11.6%

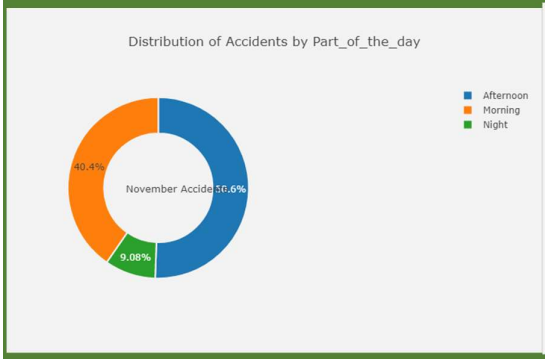

Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
June Accidents
51.8%
38.5%
9.69%

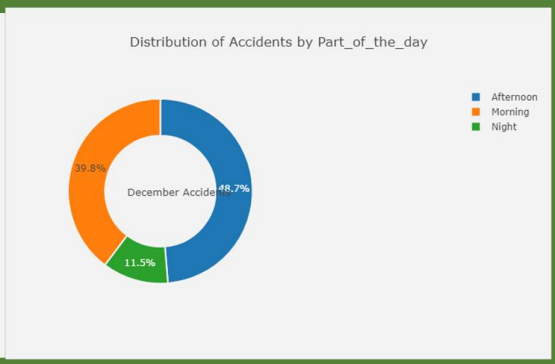

Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
July Accidents
48.4%
38.1%
13.5%


Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
August Accidents
48.8%
36.2%
15%


Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
September Accide
40.3%
37.8%
12.9%

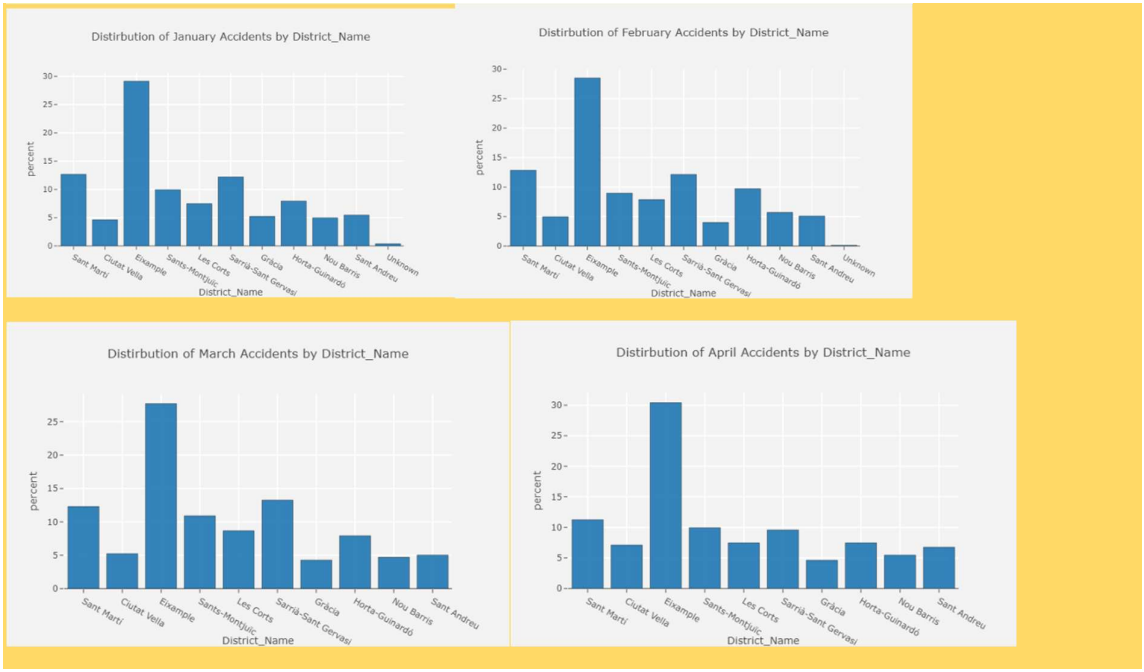

Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
October Accident
50.6%
40.4%
9.08%

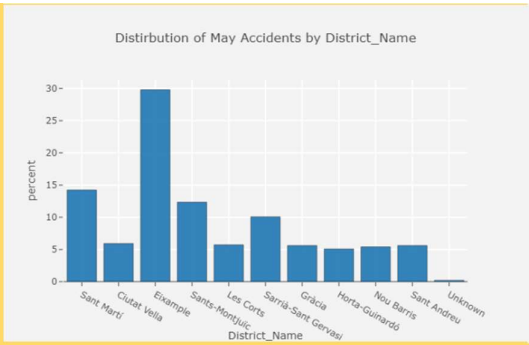

Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
November Accide
50.6%
40.4%
9.08%

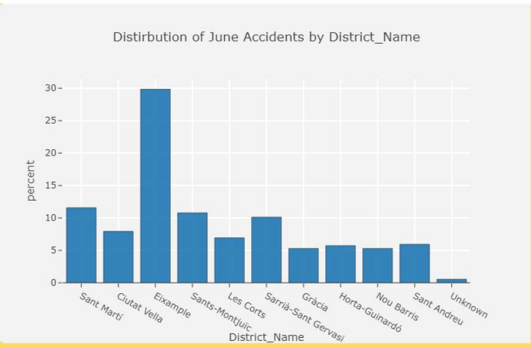

Distribution of Accidents by Part_of_the_day
Afternoon
Morning
Night
December Accide
48.7%
39.8%
11.5%

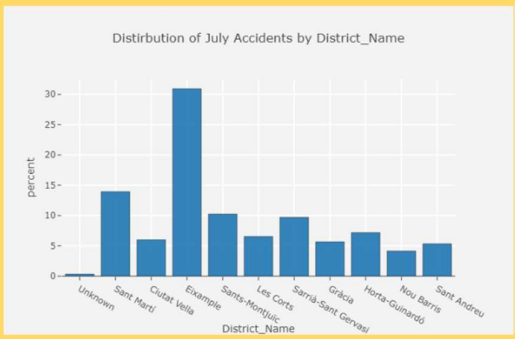
MONTHLY DISTRIBUTION OF ACCIDENTS BY SERIOUS INJURIES

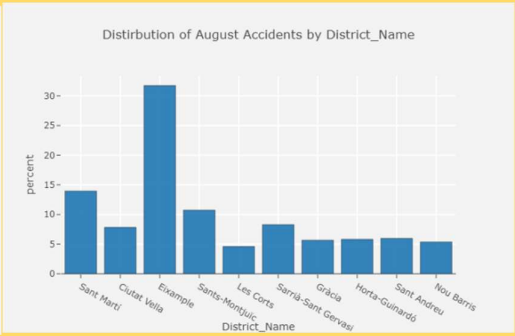
MONTHLY DISTRIBUTION OF ACCIDENTS BY DISTRICT NAME

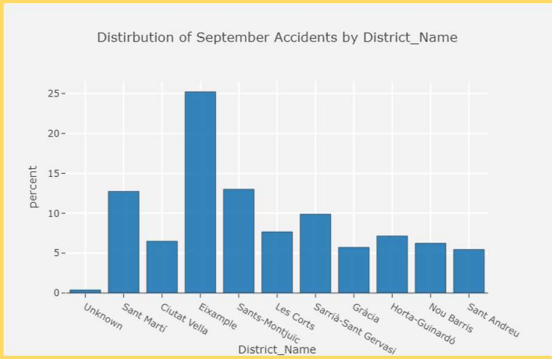


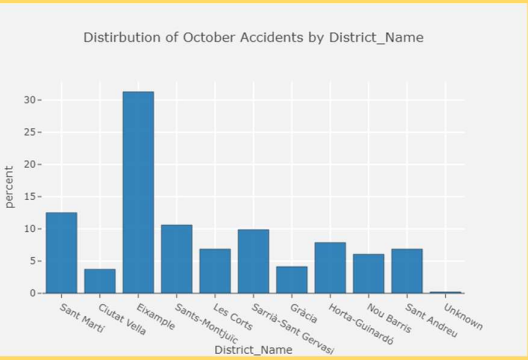

Distirbution of May Accidents by District_Name
percent
District_Name
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Nou Barris
Sant Andreu
Unknown

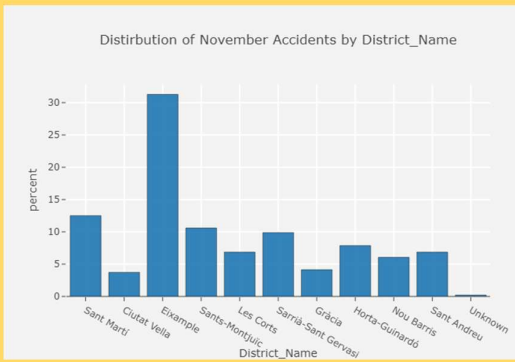

Distirbution of June Accidents by District_Name
percent
District_Name
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Nou Barris
Sant Andreu
Unknown

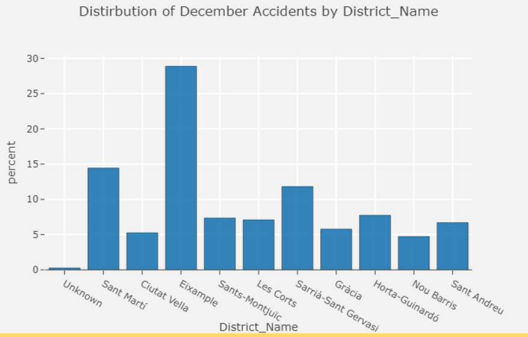

Distirbution of July Accidents by District_Name
percent
District_Name
Unknown
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Nou Barris
Sant Andreu


Distirbution of August Accidents by District_Name
percent
District_Name
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Sant Andreu
Nou Barris

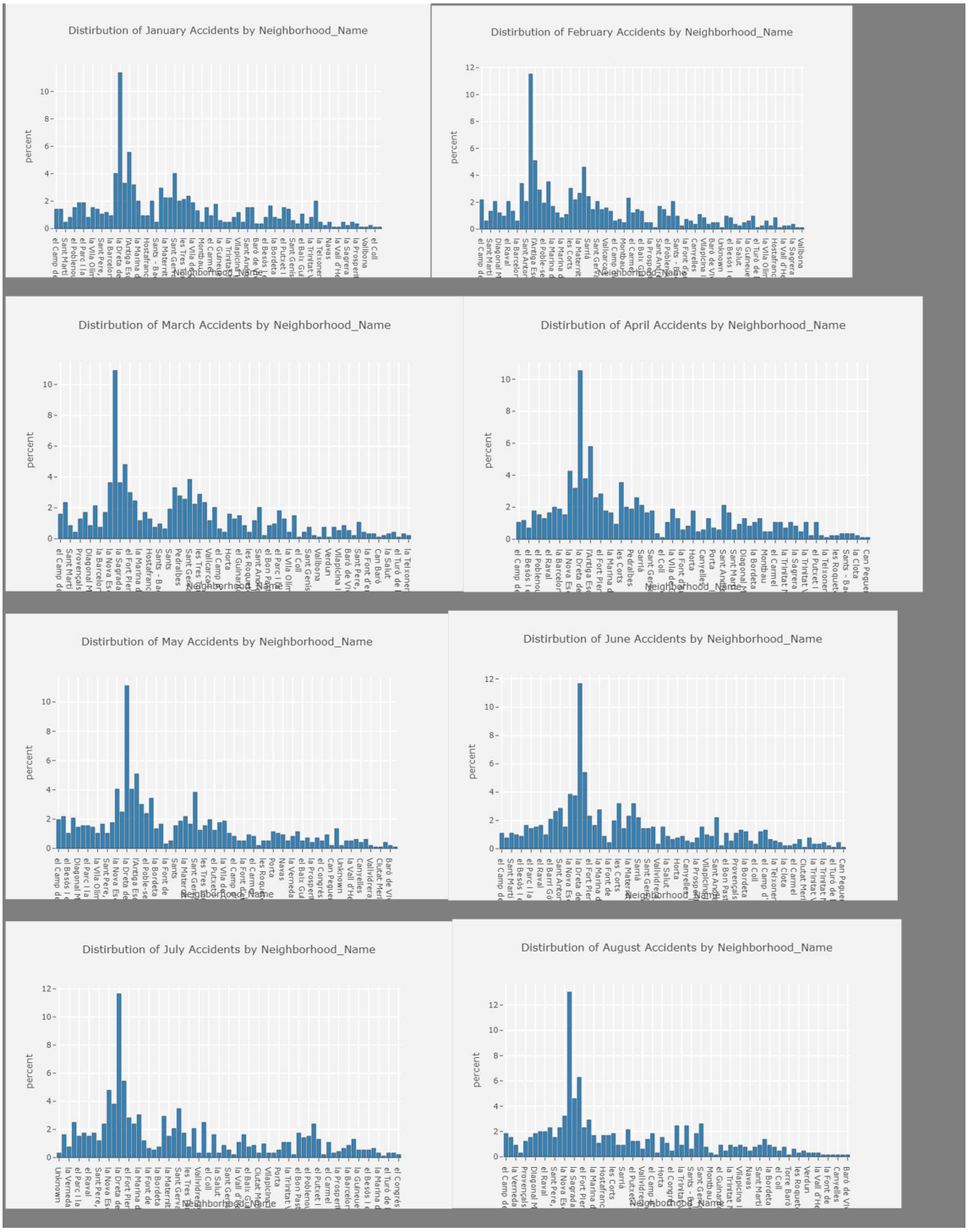

Distirbution of September Accidents by District_Name
percent
District_Name
Unknown
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Nou Barris
Sant Andreu

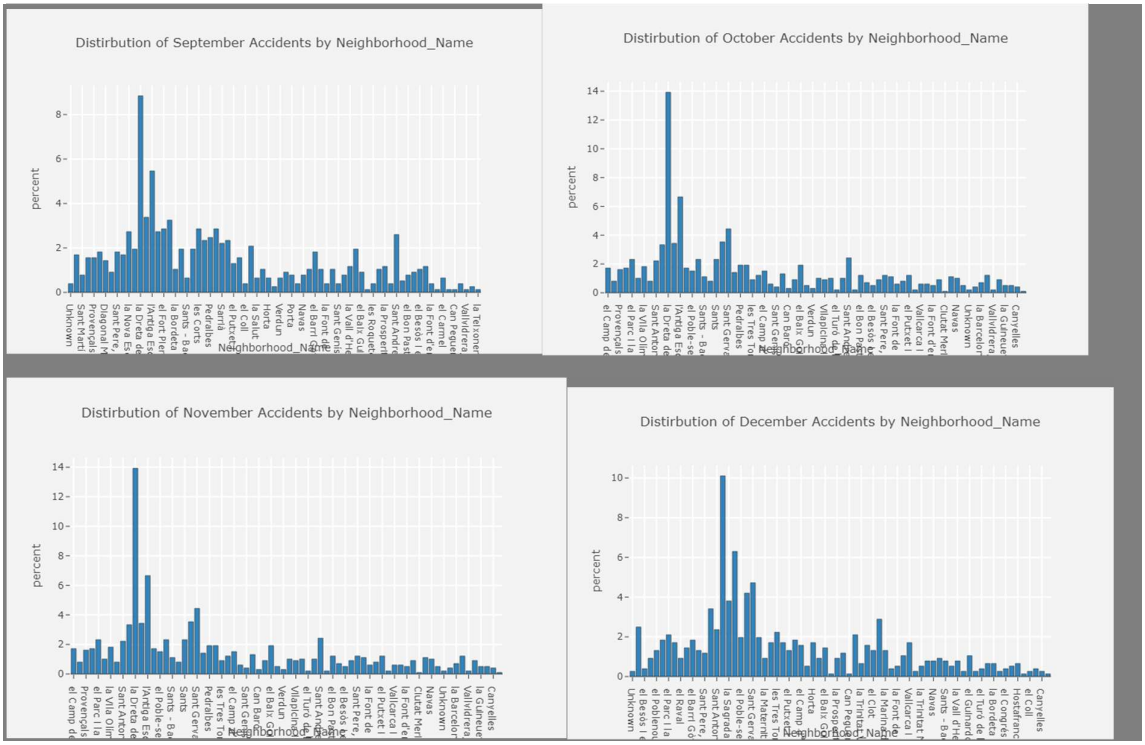

Distirbution of October Accidents by District_Name
percent
District_Name
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Nou Barris
Sant Andreu
Unknown


Distirbution of November Accidents by District_Name
percent
District_Name
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Nou Barris
Sant Andreu
Unknown

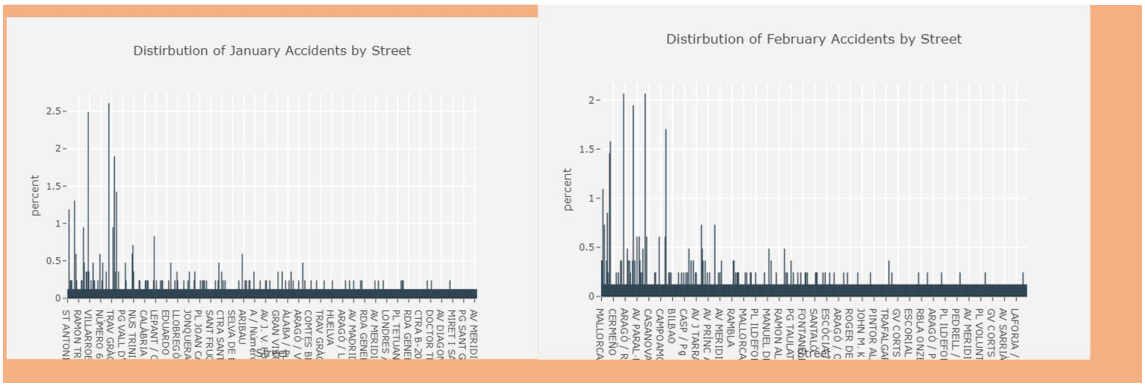

Distirbution of December Accidents by District_Name
percent
District_Name
Unknown
Sant Martí
Ciutat Vella
Eixample
Sants-Montjuïc
Les Corts
Sarrià-Sant Gervasi
Gràcia
Horta-Guinardó
Nou Barris
Sant Andreu

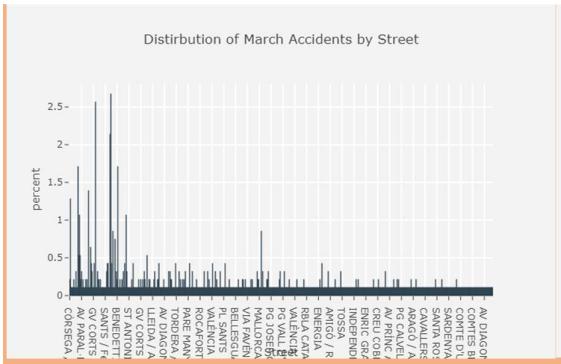
MONTHLY DISTRIBUTION OF ACCIDENTS BY NEIGHBOURHOOD

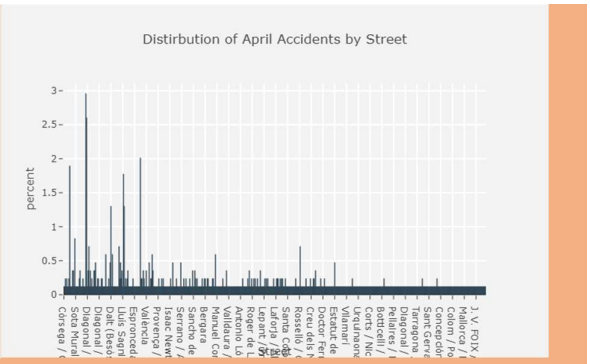


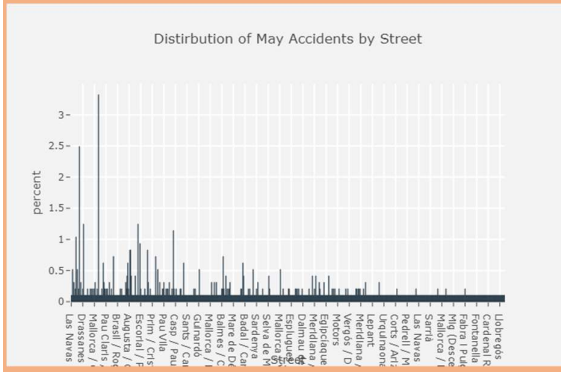


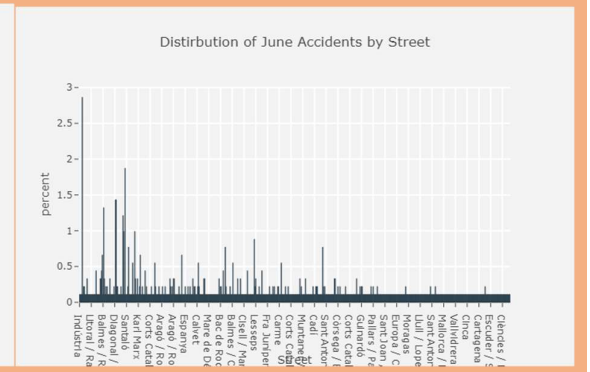
DISTRIBUTION OF MONTLY ACCIDENTS BY STREET

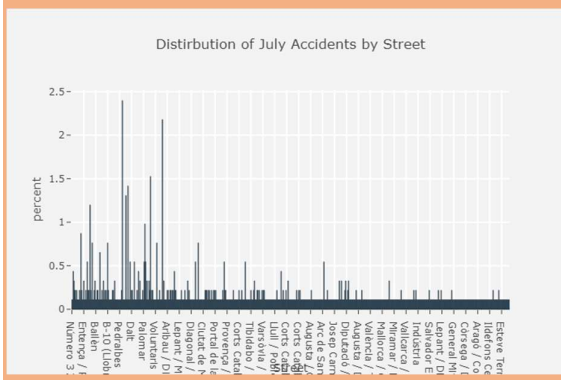


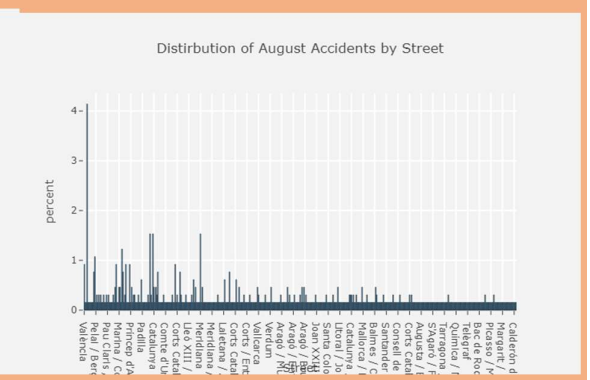

Distirbution of March Accidents by Street
percent

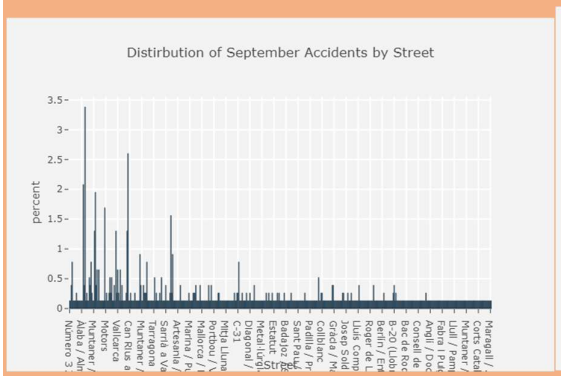

Distirbution of April Accidents by Street
percent

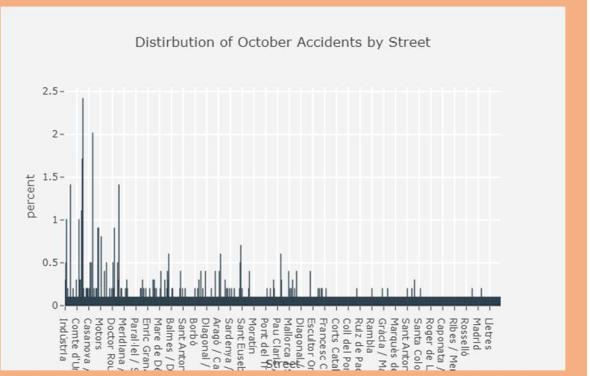

Distirbution of May Accidents by Street
percent

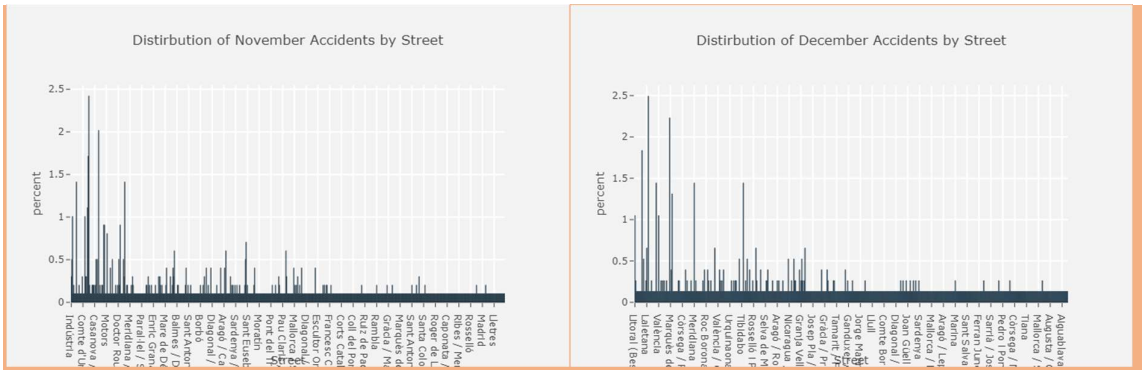

Distirbution of June Accidents by Street
percent


Distirbution of July Accidents by Street
percent


Distirbution of August Accidents by Street
percent


Distirbution of September Accidents by Street
percent

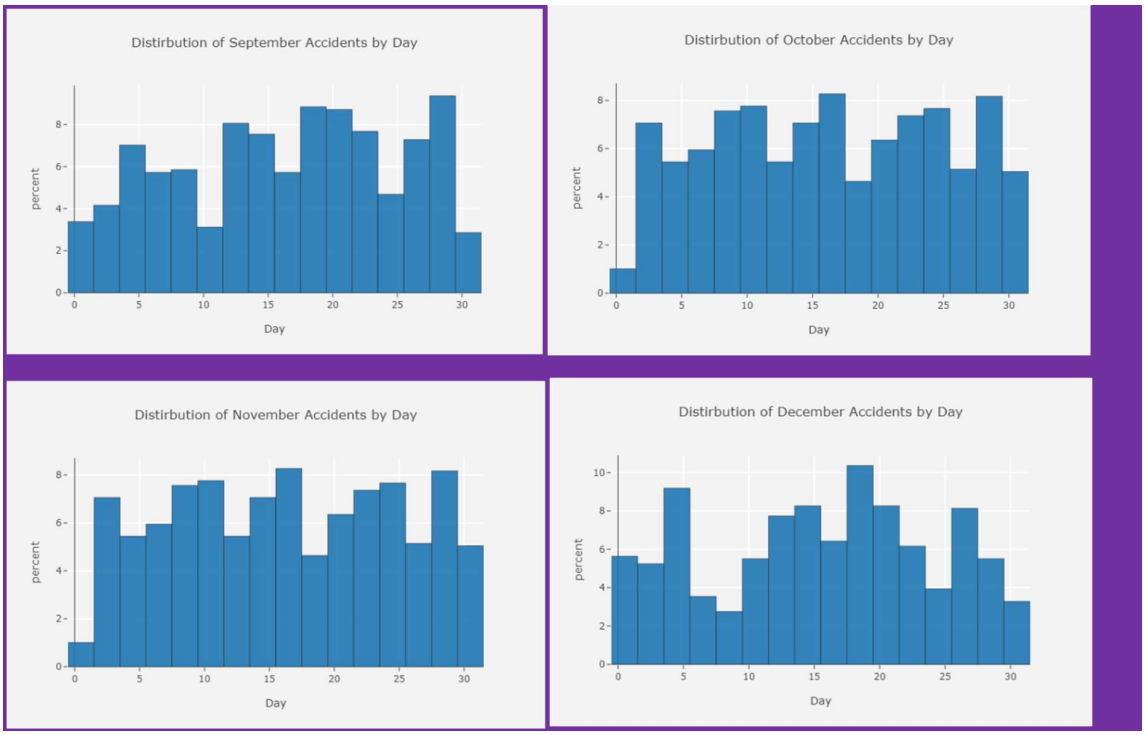

Distirbution of October Accidents by Street
percent



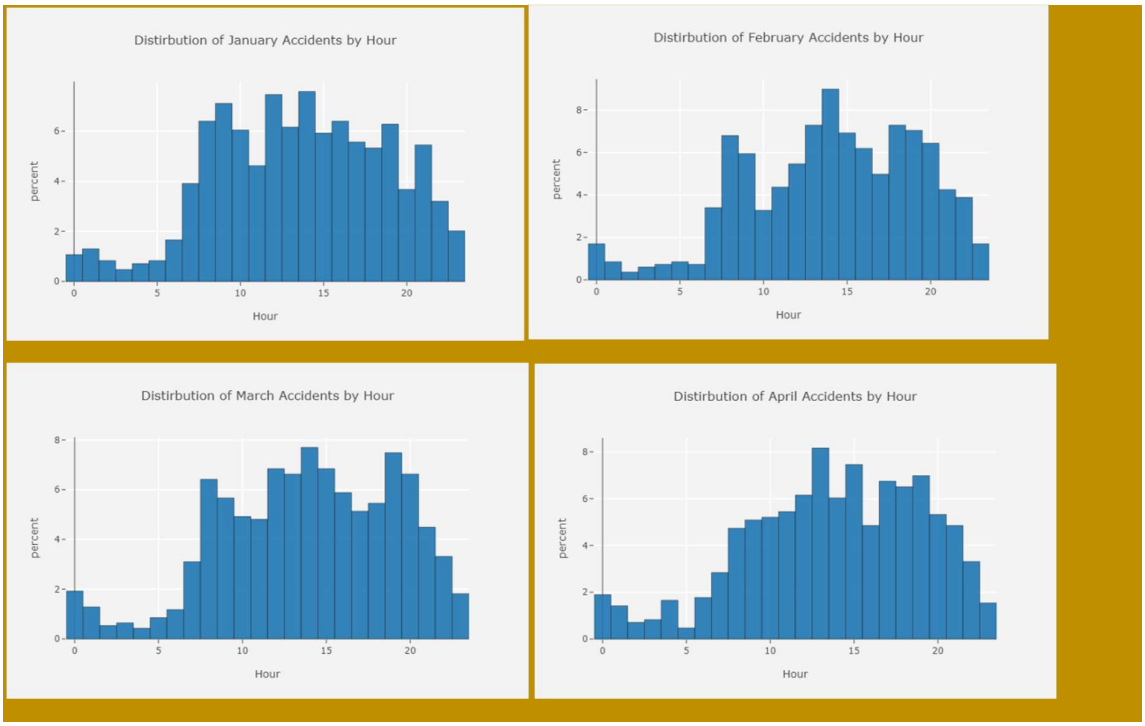
DISTRIBUTION OF MONTLY ACCIDENTS BY WEEKDAY

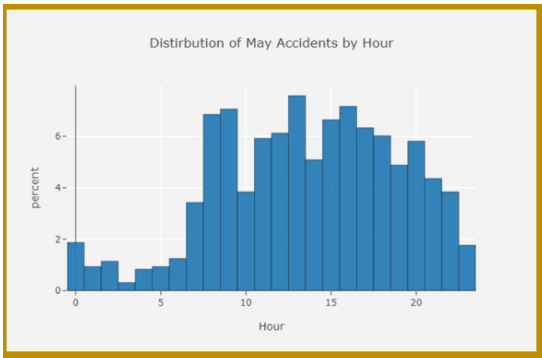

Distirbution of July Accidents by Weekday
percent
0
5
10
15
Friday
Monday
Wednesday
Tuesday
Saturday
Thursday
Sunday
Weekday
Distirbution of August Accidents by Weekday
percent
0
5
10
15
20
Thursday
Tuesday
Wednesday
Friday
Saturday
Sunday
Monday
Weekday
Distirbution of September Accidents by Weekday
percent
0
5
10
15
20
Friday
Wednesday
Monday
Saturday
Thursday
Tuesday
Sunday
Weekday
Distirbution of October Accidents by Weekday
percent
0
5
10
15
20
Thursday
Wednesday
Friday
Tuesday
Saturday
Monday
Sunday
Weekday
Distirbution of November Accidents by Weekday
percent
0
5
10
15
20
Thursday
Wednesday
Friday
Tuesday
Saturday
Monday
Sunday
Weekday
Distirbution of December Accidents by Weekday
percent
0
5
10
15
Friday
Tuesday
Thursday
Monday
Saturday
Wednesday
Sunday
Weekday

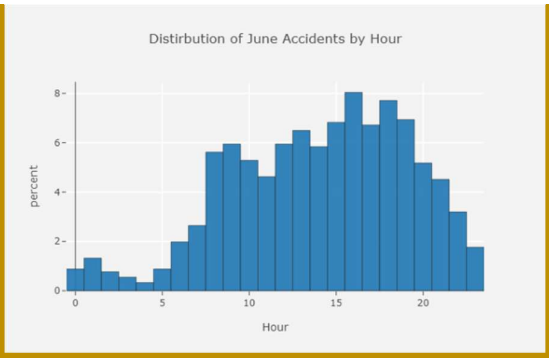
DISTRIBUTION OF MONTLY ACCIDENTS BY DAY

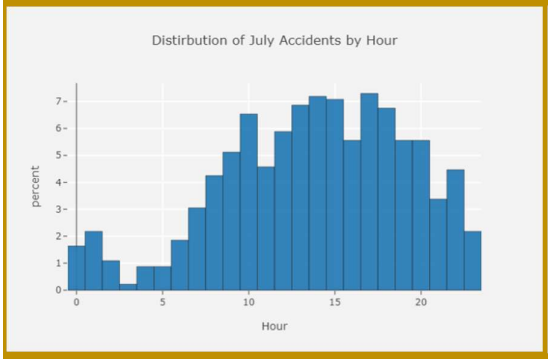


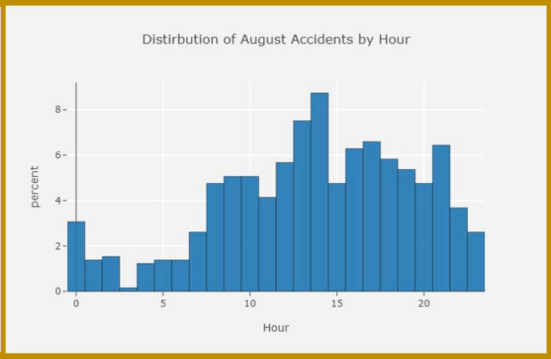
DISTRIBUTION OF MONTLY ACCIDENTS BY HOUR

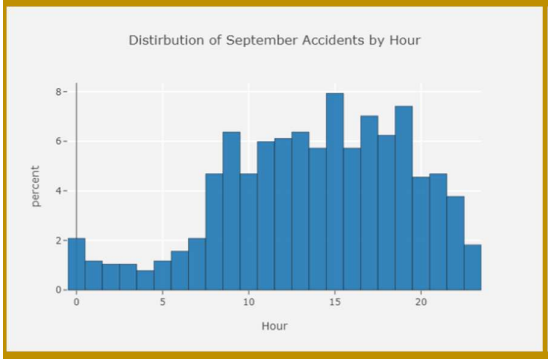


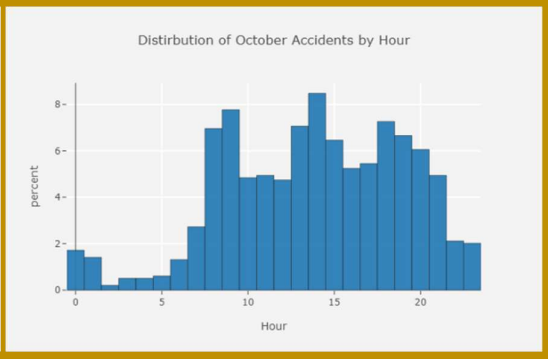

Distirbution of May Accidents by Hour
percent
Hour

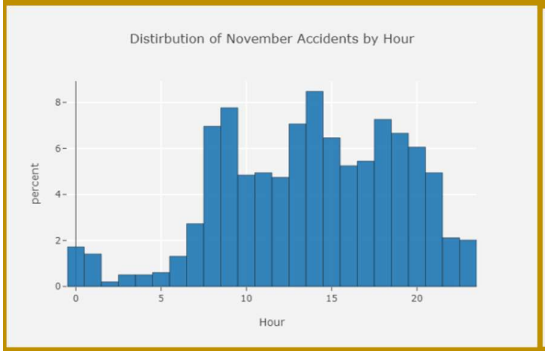

Distirbution of June Accidents by Hour
percent
Hour

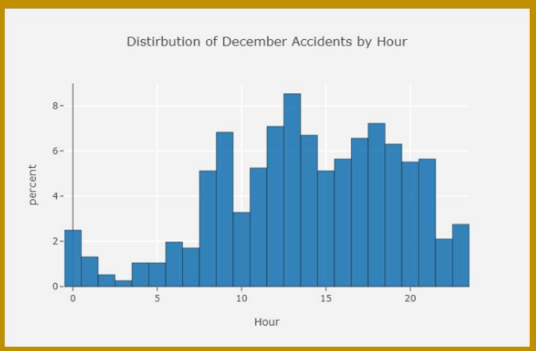

Distirbution of July Accidents by Hour
percent
Hour


Distirbution of August Accidents by Hour
percent
Hour

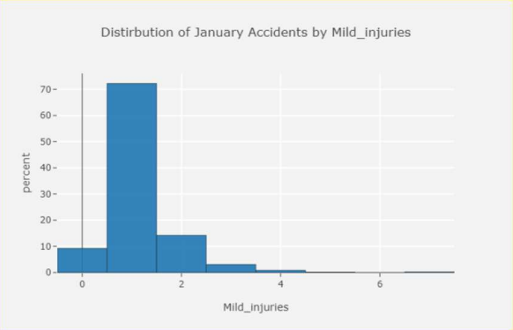

Distirbution of September Accidents by Hour
percent
Hour

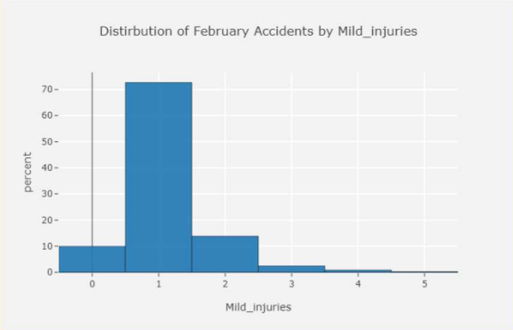

Distirbution of October Accidents by Hour
percent
Hour

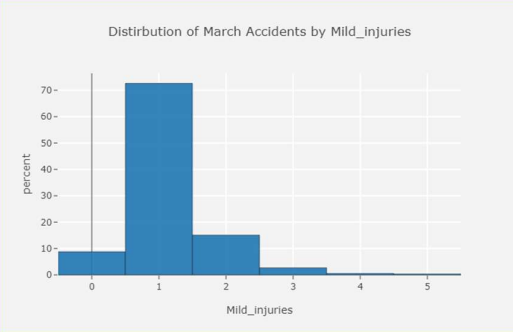

Distirbution of November Accidents by Hour
percent
Hour

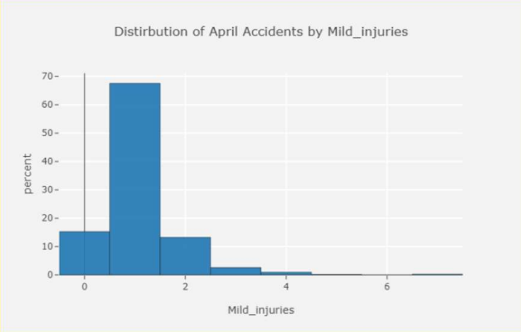

Distirbution of December Accidents by Hour
percent
Hour

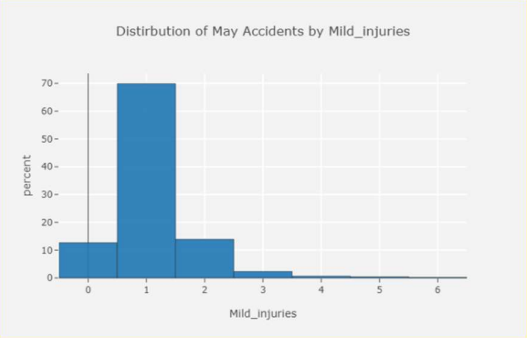
DISTRIBUTION OF MONTLY ACCIDENTS BY MILD-INJURY

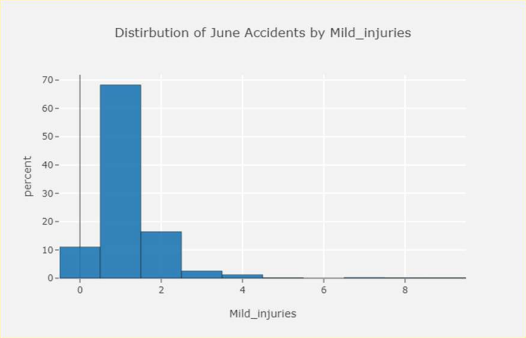


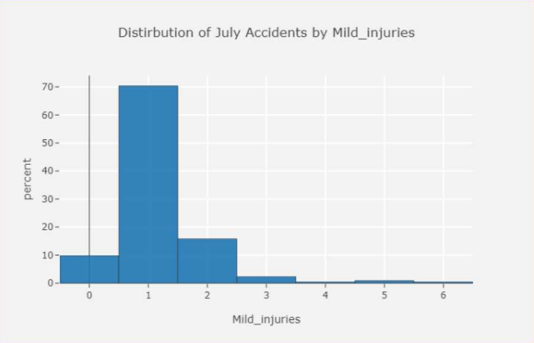


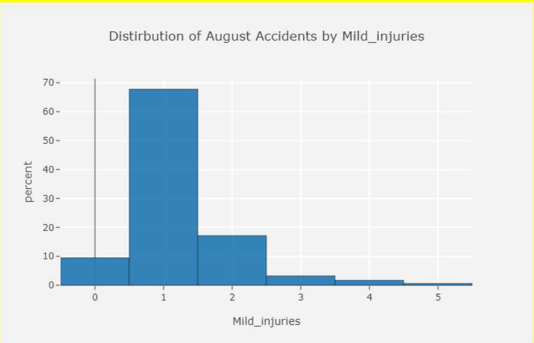


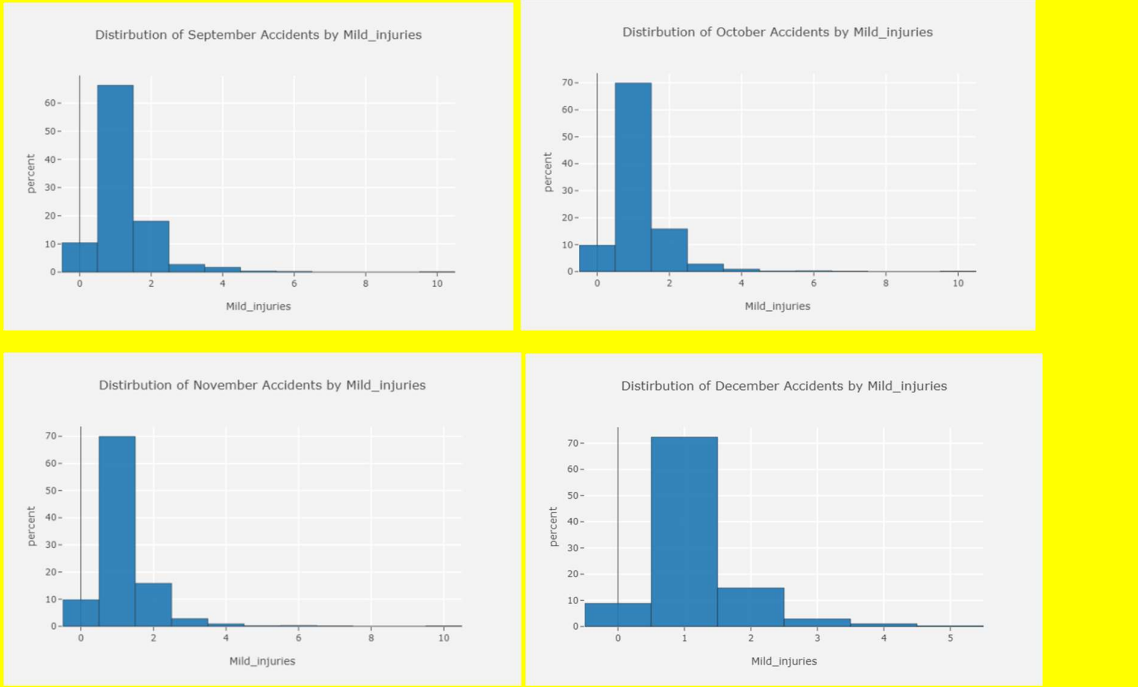




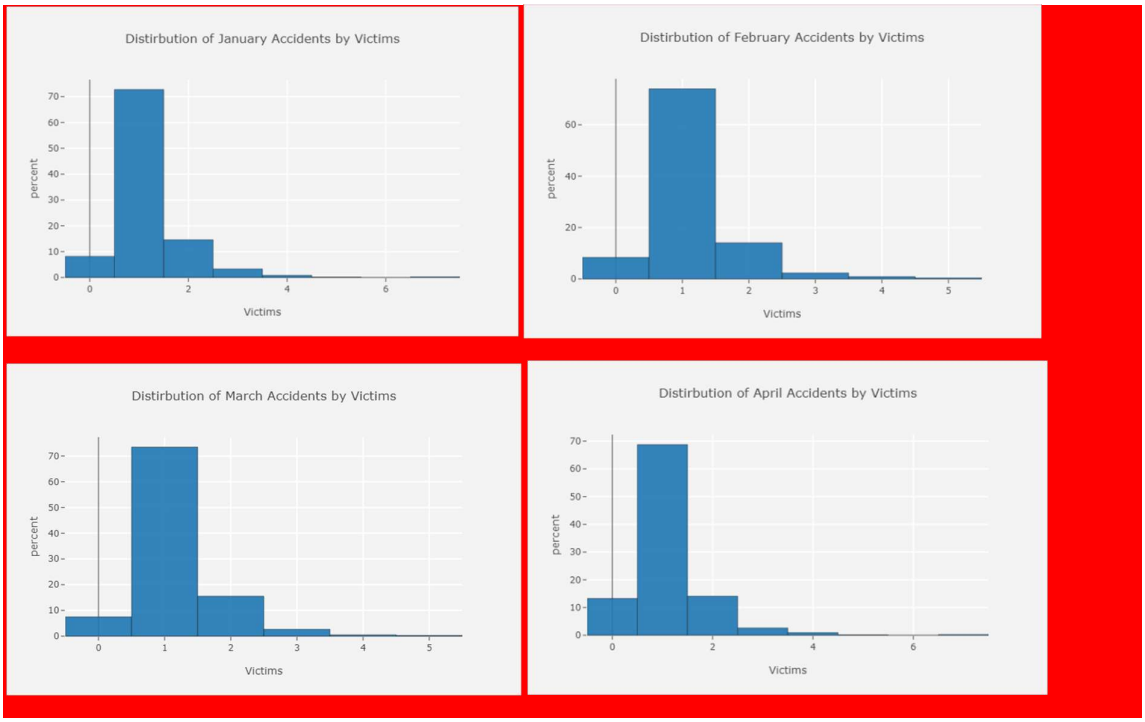


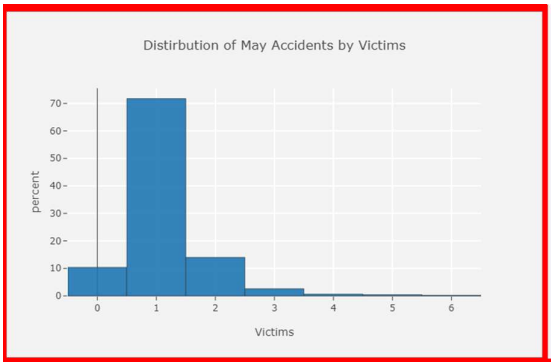


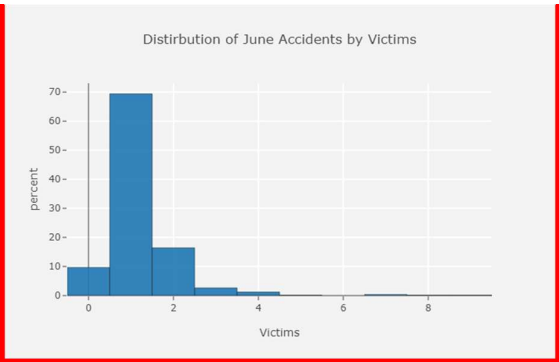


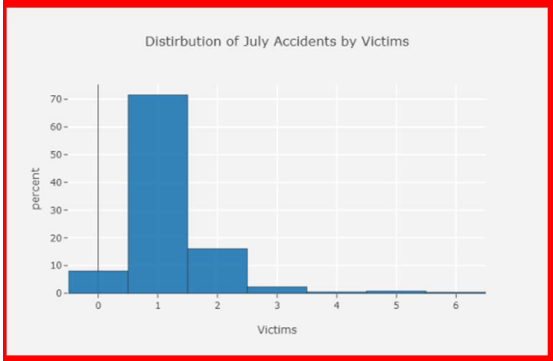


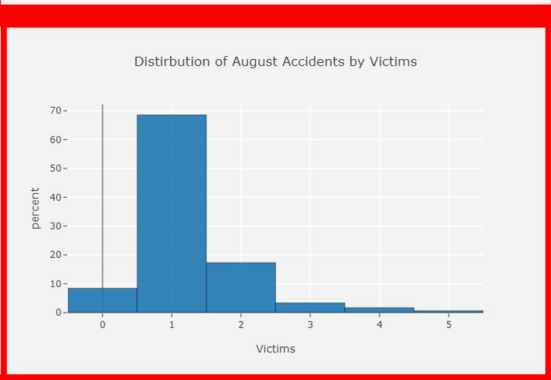
DISTRIBUTION OF MONTLY ACCIDENTS BY VICTIMS

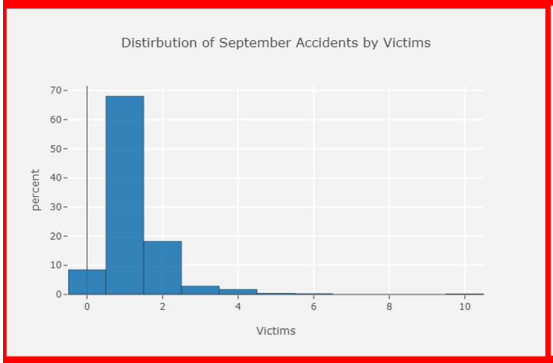


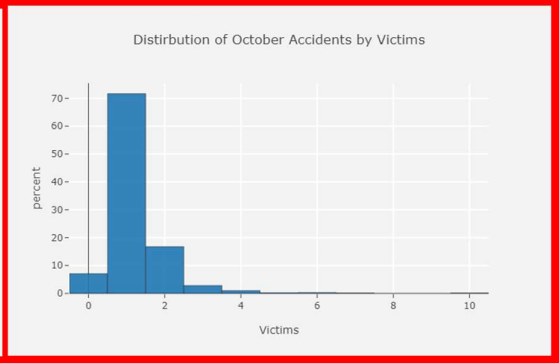

Distirbution of May Accidents by Victims
percent
Victims

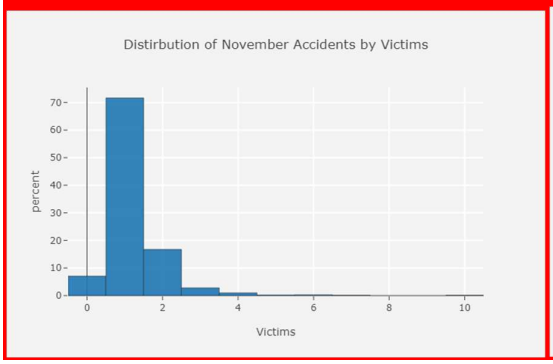

Distirbution of June Accidents by Victims
percent
Victims

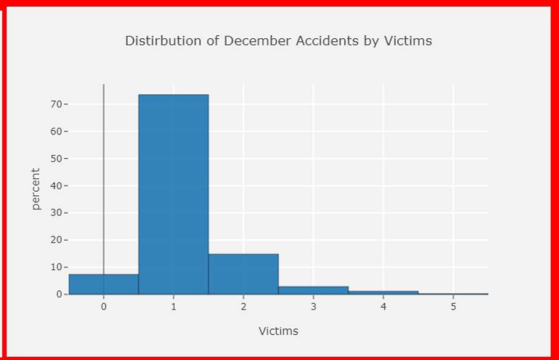

Distirbution of July Accidents by Victims
percent
Victims


Distirbution of August Accidents by Victims
percent
Victims

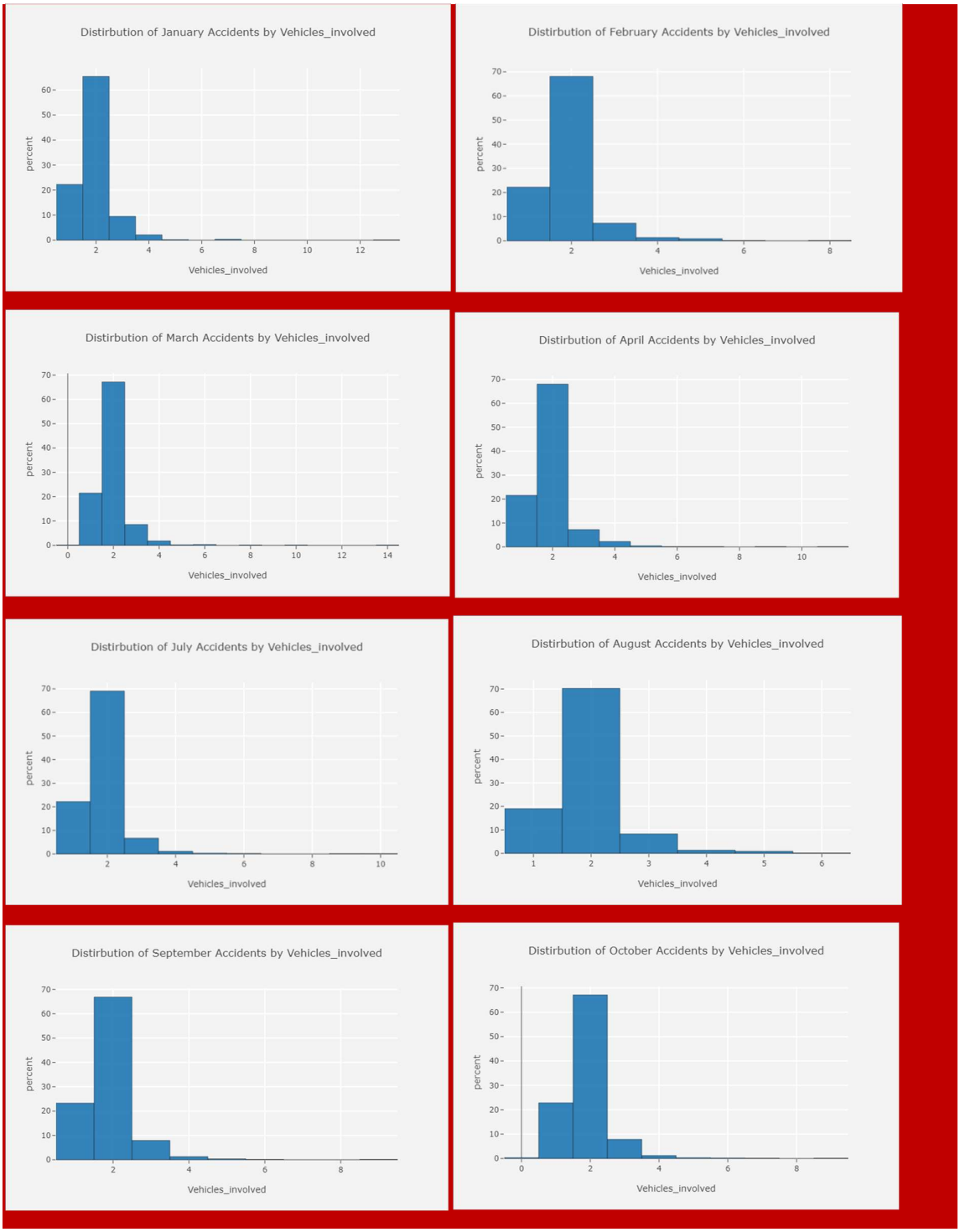

Distirbution of September Accidents by Victims
percent
Victims

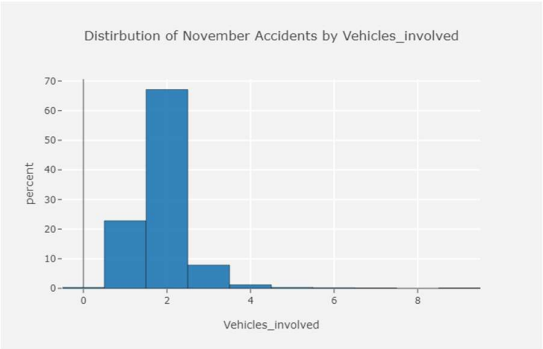

Distirbution of October Accidents by Victims
percent
Victims

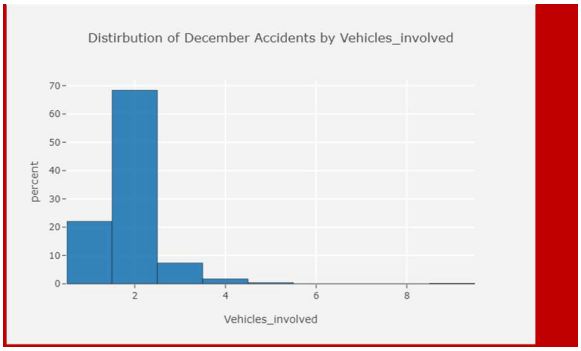

Distirbution of November Accidents by Victims
percent
Victims


Distirbution of December Accidents by Victims
percent
Victims

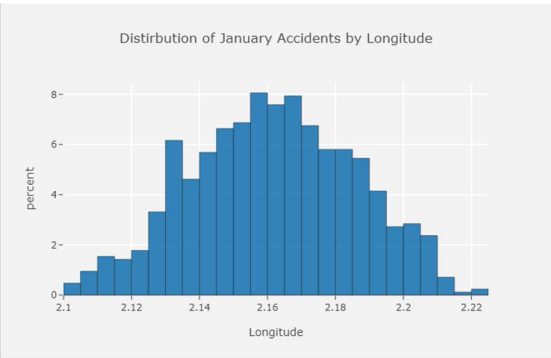
DISTRIBUTION OF MONTLY ACCIDENTS BY VEHICLES INVOLVED

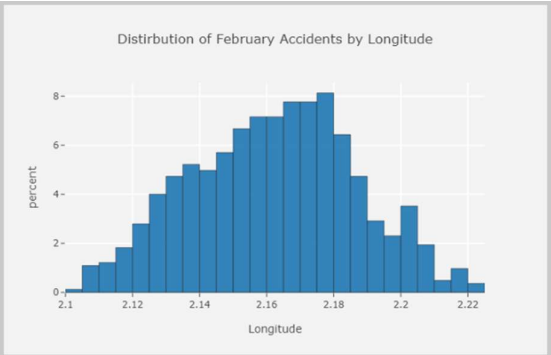


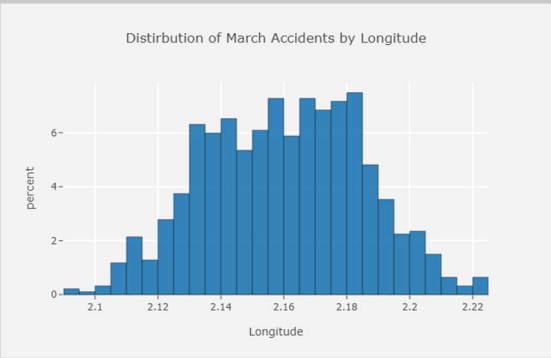


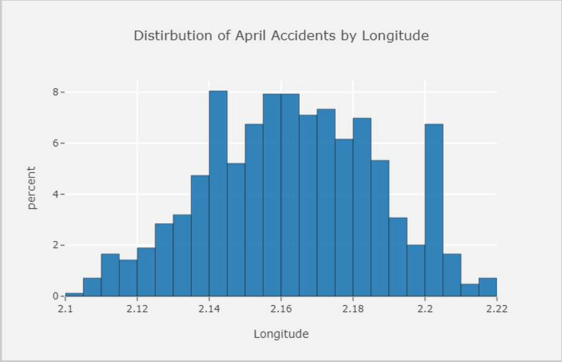


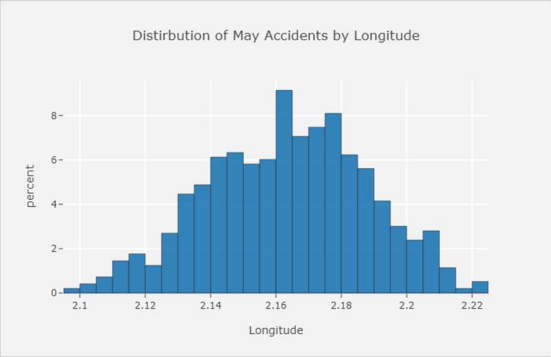
DISTRIBUTION OF MONTLY ACCIDENTS BY LONGITUDE

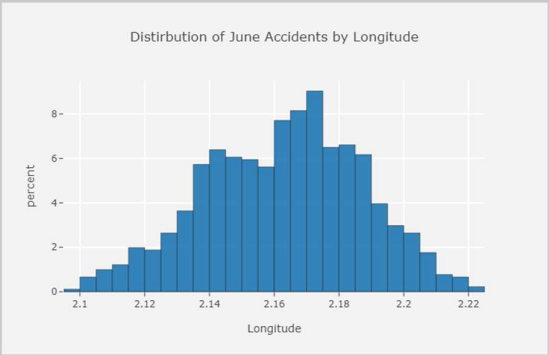


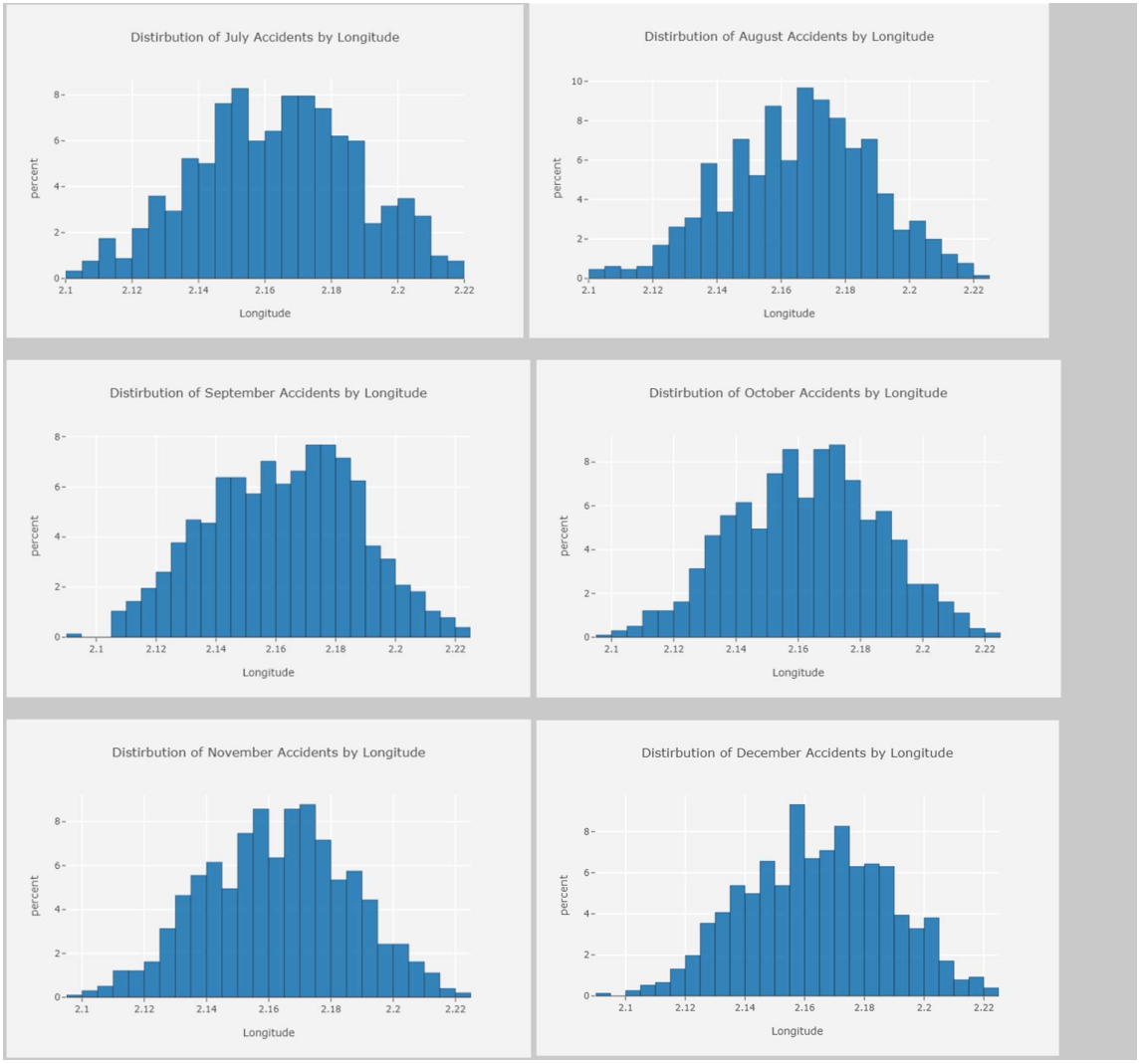




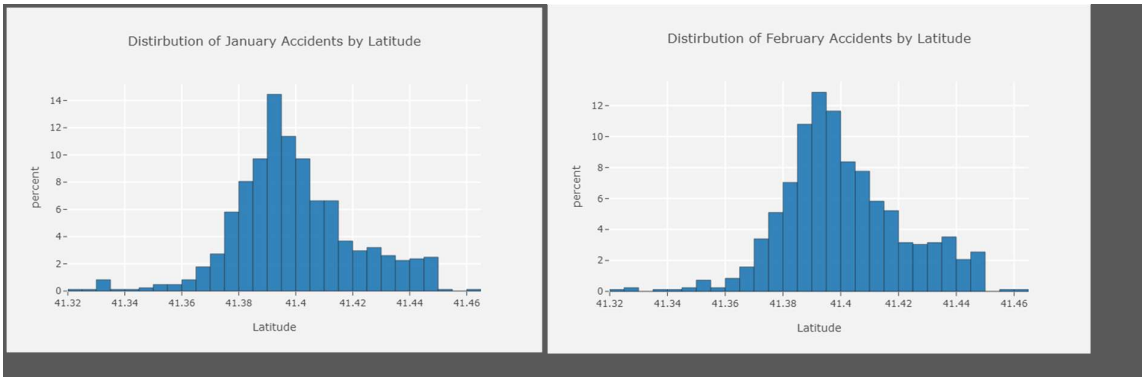


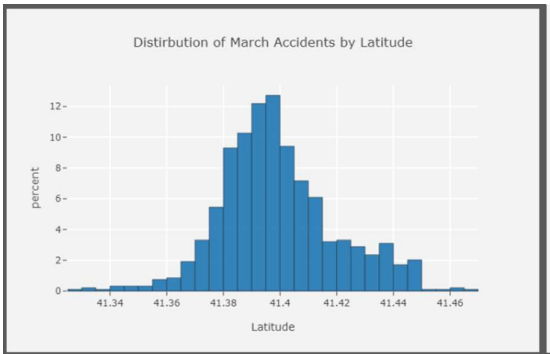


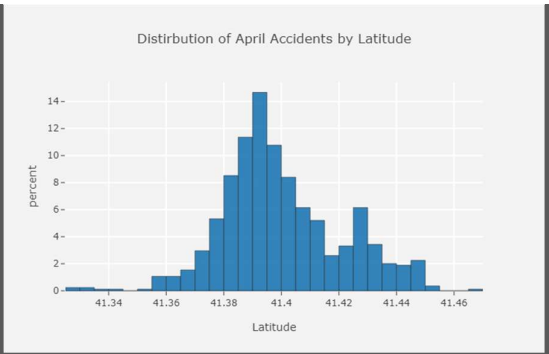


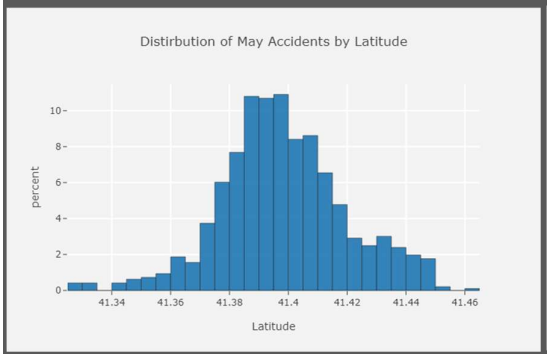


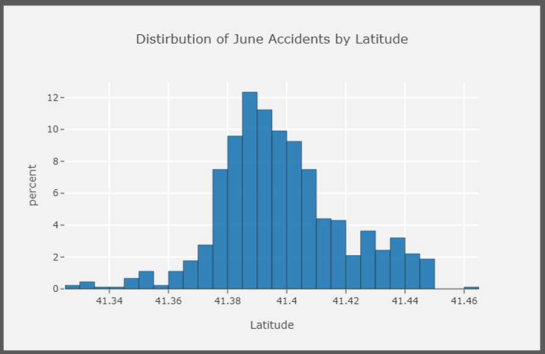
DISTRIBUTION OF MONTLY ACCIDENTS BY LATITUDE

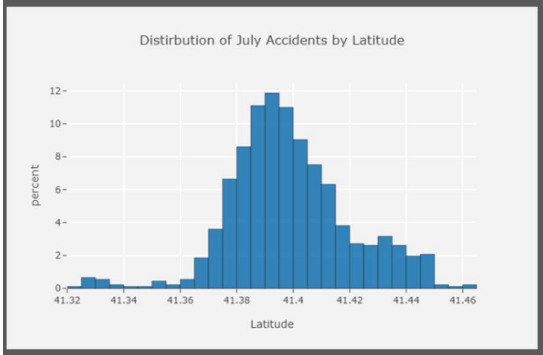


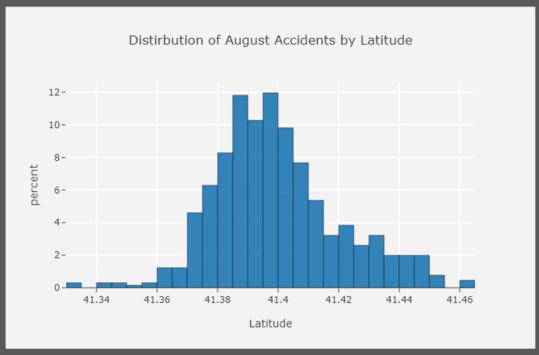

Distirbution of March Accidents by Latitude
percent
Latitude

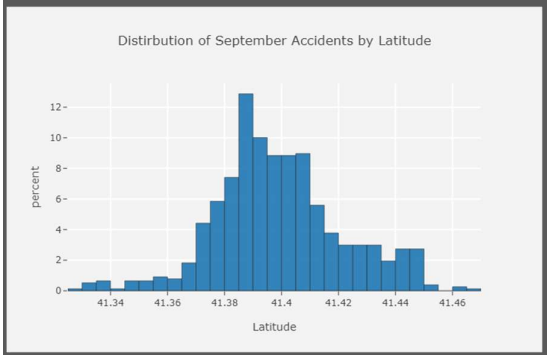

Distirbution of April Accidents by Latitude
percent
Latitude

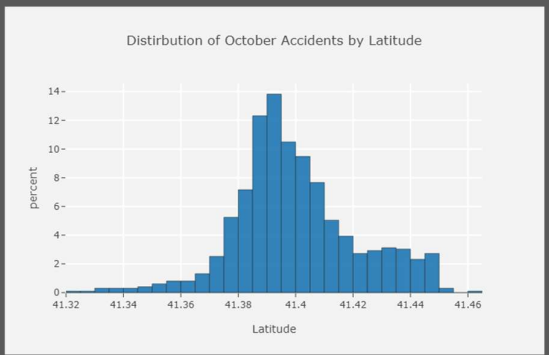

Distirbution of May Accidents by Latitude
percent
Latitude

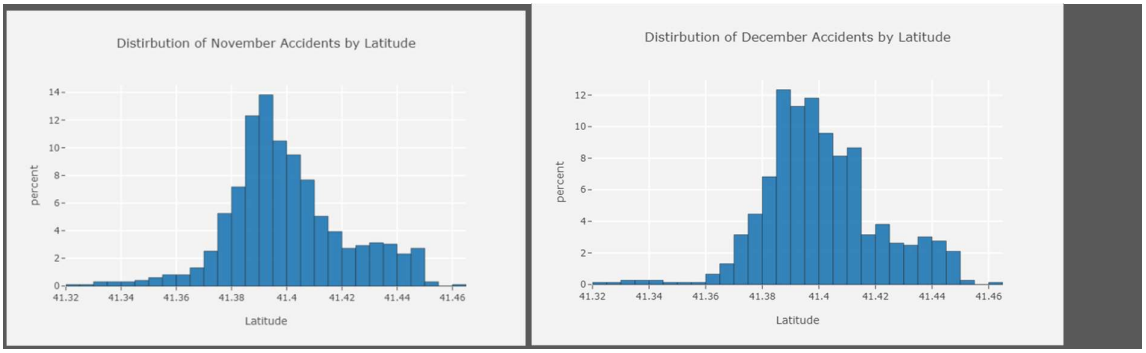

Distirbution of June Accidents by Latitude
percent
Latitude

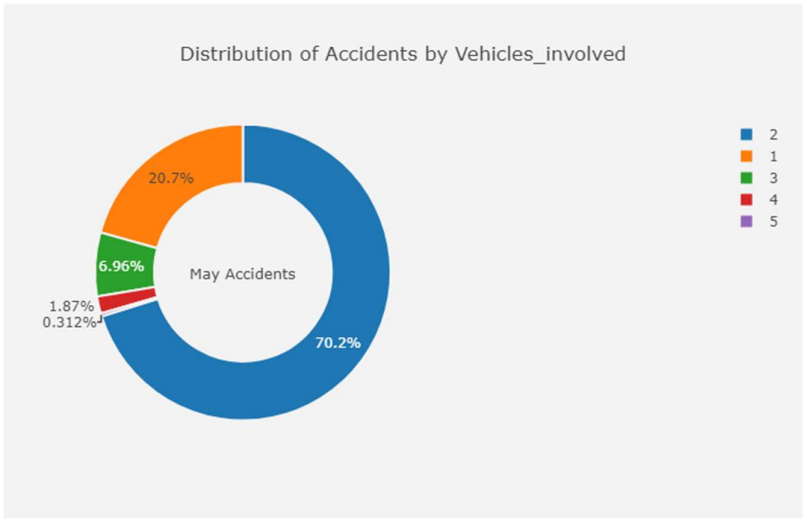

Distirbution of July Accidents by Latitude
percent
Latitude

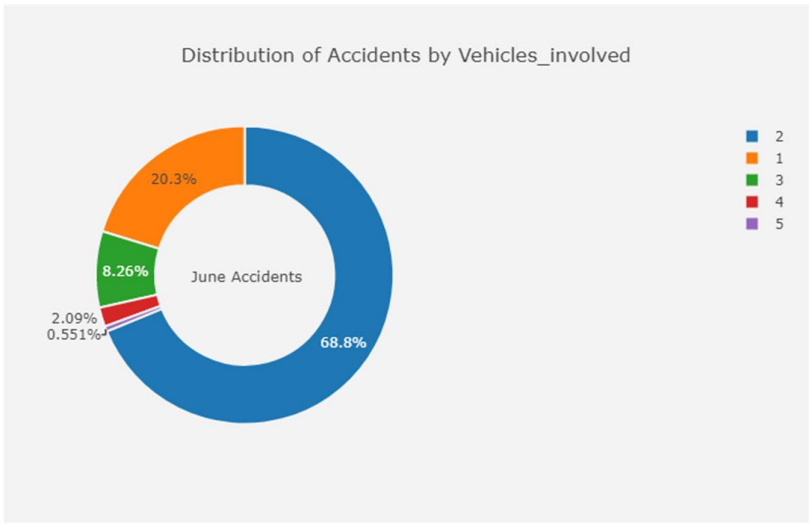

Distirbution of August Accidents by Latitude
percent
Latitude


Distirbution of September Accidents by Latitude
percent
Latitude


Distirbution of October Accidents by Latitude
percent
Latitude


Distirbution of November Accidents by Latitude
percent
Latitude
Distirbution of December Accidents by Latitude
percent
Latitude

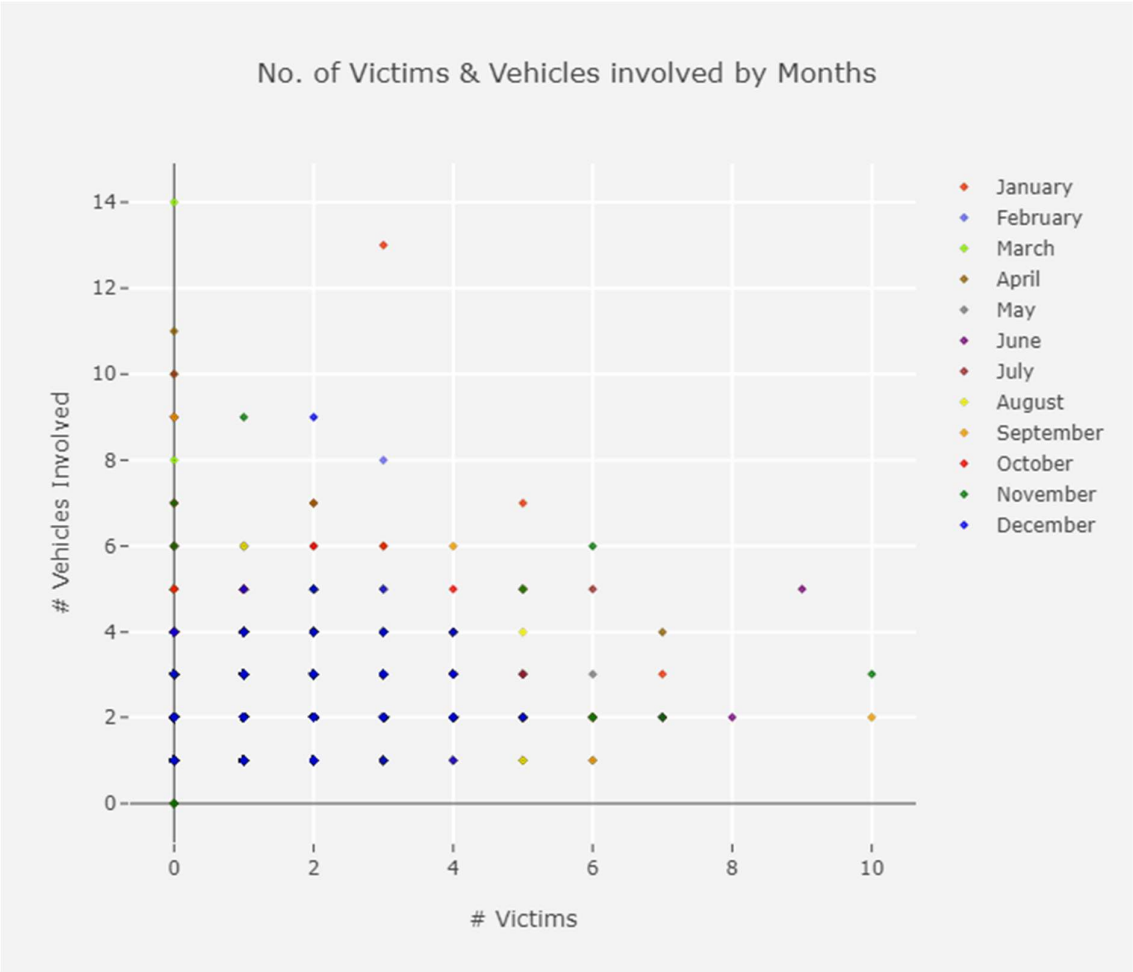

Distribution of Accidents by Vehicles_involved
May Accidents
20.7%
6.96%
1.87%
0.312%
70.2%
2
1
3
4
5

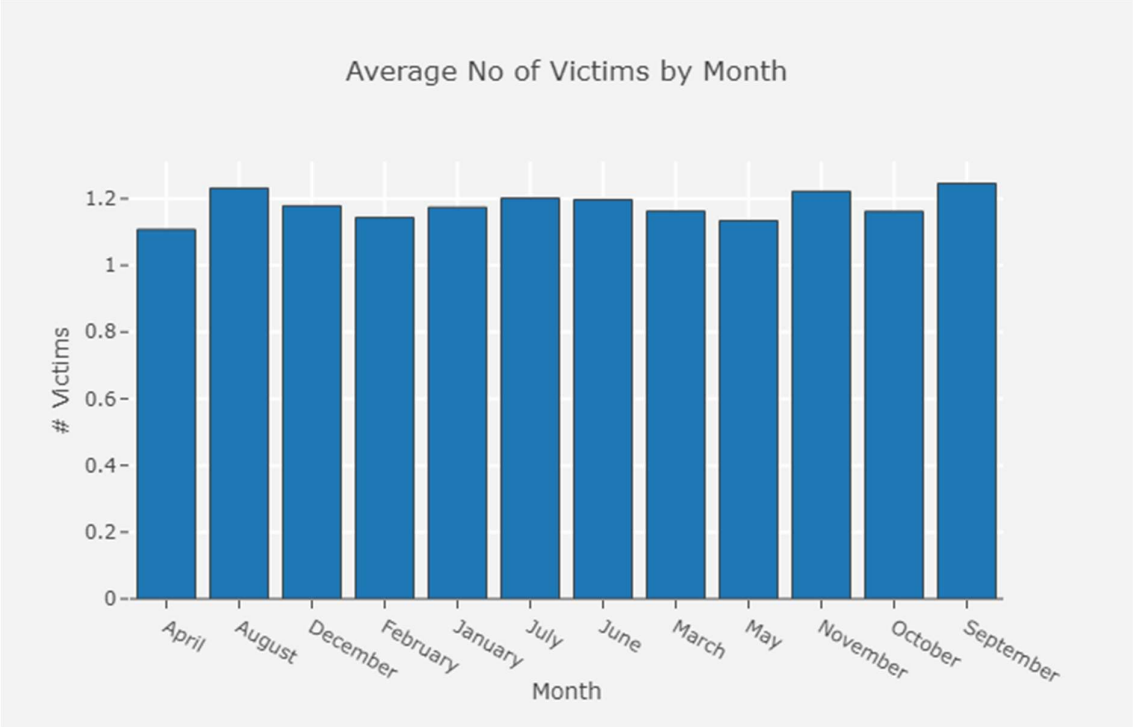

Distribution of Accidents by Vehicles_involved
June Accidents
20.3%
8.26%
2.09%
0.551%
68.8%
2
1
3
4
5

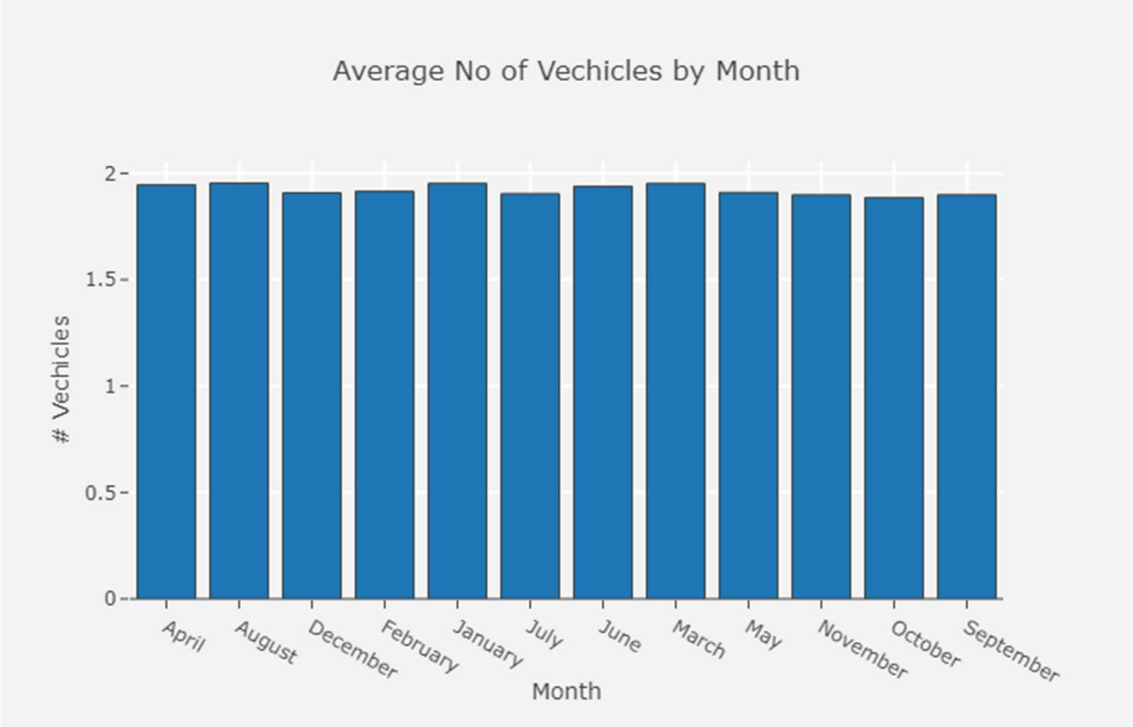
Correlation Coefficients between features:

	Day	Hour	Mild_injuries	Serious_injuries	Victims	Vehicles_involved	Longitude	Latitude	
Day	1.000000	0.012877	0.006975	-0.001686	0.006407	-0.010914	-0.017236	-0.006184	
Hour	0.012877	1.000000	0.042420	-0.007614	0.040177	-0.022223	-0.006215	0.018699	
Mild_injuries	0.006975	0.042420	1.000000	-0.150482	0.974272	0.160052	0.008475	0.014091	
Serious_injuries	-0.001686	-0.007614	-0.150482	1.000000	0.071450	-0.015678	0.005372	0.000299	
Victims	0.006407	0.040177	0.974272	0.071450	1.000000	0.157185	0.009355	0.013520	
Vehicles_involved	-0.010914	-0.022223	0.160052	-0.015678	0.157185	1.000000	-0.006864	0.002440	
Longitude	-0.017236	-0.006215	0.008475	0.005372	0.009355	-0.006864	1.000000	0.396106	
Latitude	-0.006184	0.018699	0.014091	0.000299	0.013520	0.002440	0.396106	1.000000	

Features to be selected for regression analysis: Mid-injuries, vehicles involved, victims, serious injuries, longitude, latitude

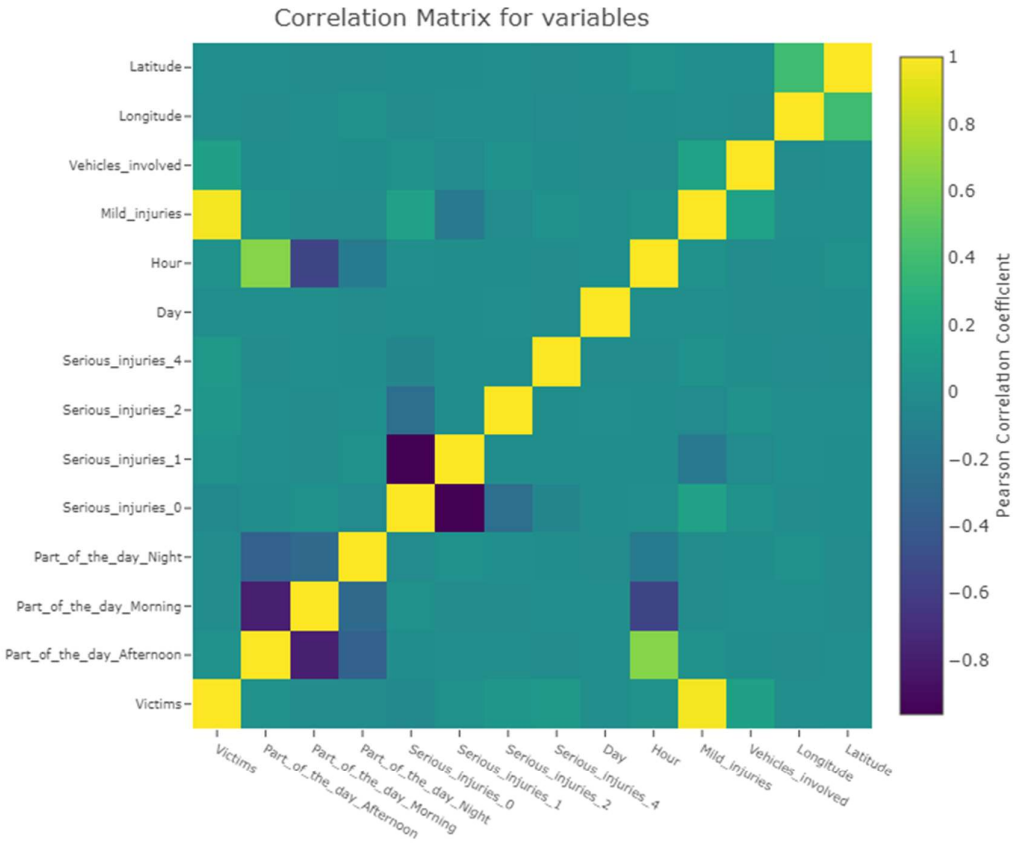

No. of Victims & Vehicles involved by Months
Vehicles Involved
14
12
10
8
6
4
2
0
0
2
4
6
8
10
Victims
January
February
March
April
May
June
July
August
September
October
November
December


Average No of Victims by Month
Victims
1.2
1
0.8
0.6
0.4
0.2
0
April
August
December
February
January
July
June
March
May
November
October
September
Month


Average No of Vechicles by Month
Vechicles
2
1.5
1
0.5
0
April
August
December
February
January
July
June
March
May
November
October
September
Month

Variable Summary

feature	count	mean	std	min	25%	50%	75%	max
Victims	10339	1.179	0.735	0	1	1	1	10
t_of_the_day_Afternc	10339	0.492	0.5	0	0	0	1	1
rt_of_the_day_Morni	10339	0.393	0.489	0	0	0	1	1
Part_of_the_day_Nigh	10339	0.115	0.319	0	0	0	0	1
Serious_injuries_0	10339	0.978	0.145	0	1	1	1	1
Serious_injuries_1	10339	0.02	0.14	0	0	0	0	1
Serious_injuries_2	10339	0.001	0.038	0	0	0	0	1
Serious_injuries_4	10339	0	0.01	0	0	0	0	1
Day	10339	0	1	-1.686	-0.887	0.026	0.824	1.737
Hour	10339	0	1	-2.598	-0.717	0.035	0.788	1.728
Mild_injuries	10339	0	1	-1.555	-0.208	-0.208	-0.208	11.917
Vehicles_involved	10339	0	1	-2.691	0.111	0.111	0.111	16.918



Prediction Models:

KNN-Classifier

Accuracy Score

```
k=1, accuracy=95.96%
k=3, accuracy=96.35%
k=5, accuracy=96.13%
k=7, accuracy=95.88%
k=9, accuracy=95.47%
k=11, accuracy=94.84%
k=13, accuracy=94.33%
k=15, accuracy=93.88%
k=17, accuracy=93.51%
k=19, accuracy=92.95%
k=21, accuracy=92.41%
k=23, accuracy=91.97%
k=25, accuracy=91.50%
k=27, accuracy=91.00%
k=29, accuracy=90.62%
k=31, accuracy=90.22%
k=33, accuracy=89.80%
k=35, accuracy=89.49%
```

```
k=37, accuracy=89.11%
k=39, accuracy=88.68%
k=3, achieved highest accuracy of 96.35%
```

Logistic Regression

```
Accuracy of logistic regression classifier on test set: 0.810928
10-fold cross validation average accuracy:0.813
```

Confusion Matrix:

```
[[ 152    1    0    0    0    0    0    0]
 [   2 1498    0    0    0    0    0    0]
 [   1  295   27    0    0    0    0    0]
 [   0    0   61    0    0    0    0    0]
 [   0    0   19    0    0    0    0    0]
 [   0    0    8    0    0    0    0    0]
 [   0    0    3    0    0    0    0    0]
 [   0    0    1    0    0    0    0    0]]
```

Prediction is not possible using Logistic Regression for this data without any Boolean data in dataset

Let's explore Decision Tree now:

```
Predictions using GINI index:
Predicted Values:
[1 1 1 ... 1 1 1]

Confusion Matrix:
[[ 153    0    0    0    0    0    0    0]
 [   2 1498    0    0    0    0    0    0]
 [   0    1  322    0    0    0    0    0]
 [   0    0    1   60    0    0    0    0]
 [   0    0    0    0   19    0    0    0]
 [   0    0    0    0    8    0    0    0]
 [   0    0    0    0    3    0    0    0]
 [   0    0    0    0    1    0    0    0]]

Accuracy:
99.22630560928434

Detailed Report using GINI Index:
             precision    recall  f1-score   support

          0       0.99      1.00      0.99       153
          1       1.00      1.00      1.00      1500
          2       1.00      1.00      1.00       323
          3       1.00      0.98      0.99        61
          4       0.61      1.00      0.76        19
          5       0.00      0.00      0.00         8
          6       0.00      0.00      0.00         3
          8       0.00      0.00      0.00         1

avg / total       0.99      0.99      0.99      2068
```

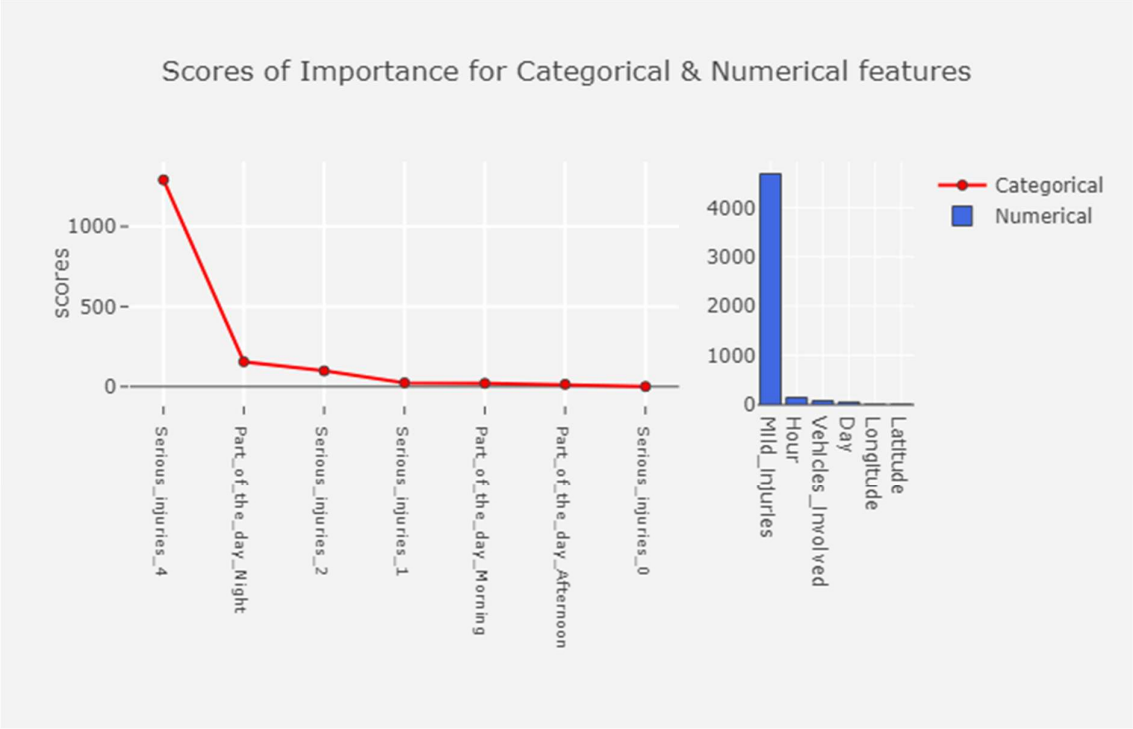
Predicted values have been saved in excel file (attached here)

Sample Decision Tree using GINI is attached here:

Gini Tree.pdf

```
Predictions using ENTROPY index:
Predicted Values:
[1 1 1 ... 1 1 1]

Confusion Matrix:
[[ 153    0    0    0    0    0    0    0]
 [   2 1498    0    0    0    0    0    0]
 [   0    1  322    0    0    0    0    0]
 [   0    0    1   60    0    0    0    0]
 [   0    0    0    0   19    0    0    0]
 [   0    0    0    0    8    0    0    0]
 [   0    0    0    0    3    0    0    0]
 [   0    0    0    0    1    0    0    0]]

Accuracy:
99.22630560928434

Detailed Report using ENTROPY Index:
             precision    recall  f1-score   support

          0       0.99      1.00      0.99       153
          1       1.00      1.00      1.00      1500
          2       1.00      1.00      1.00       323
          3       1.00      0.98      0.99        61
          4       0.61      1.00      0.76        19
          5       0.00      0.00      0.00         8
          6       0.00      0.00      0.00         3
          8       0.00      0.00      0.00         1

avg / total       0.99      0.99      0.99      2068
```

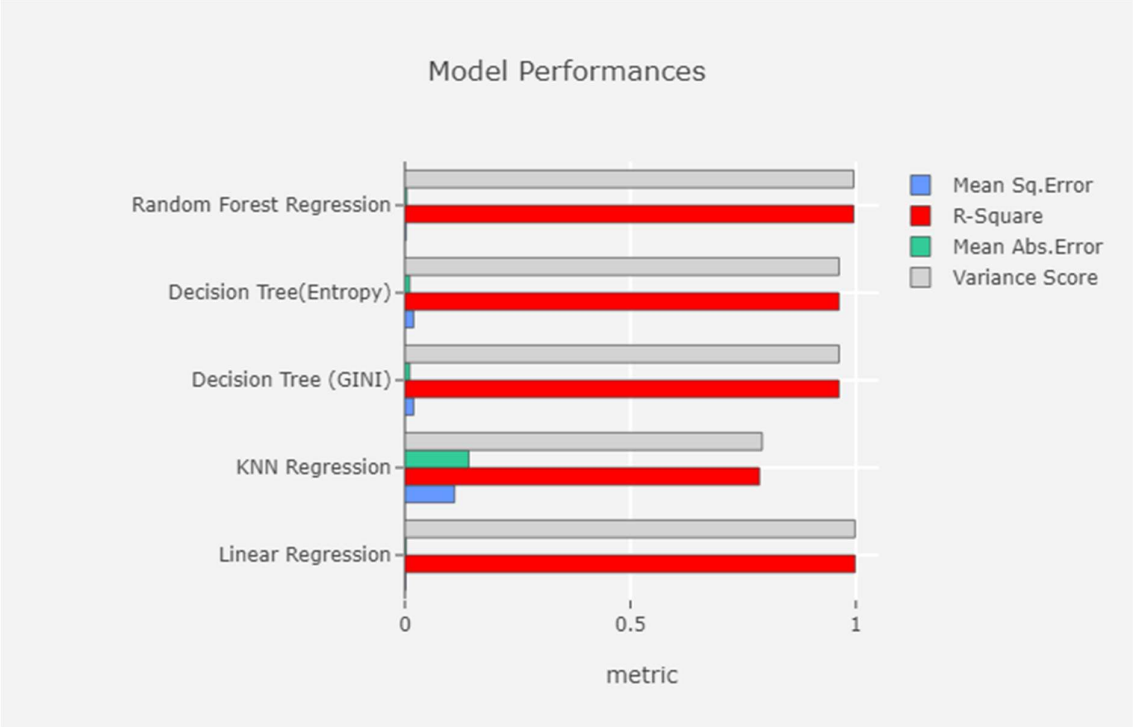
Features-scores-p-value Matrix:



Predicted value for all regressions is attached herewith:

victim_prediction.xlsx

Model	Mean Sq.Error	R-Square	Mean Abs.Error	Variance Score
Linear Regression	0.001	0.9981	0.0026	0.9981
KNN Regression	0.1101	0.786	0.1415	0.7921
Decision Tree (GINI)	0.0193	0.9624	0.0106	0.9626
Decision Tree(Entropy)	0.0193	0.9624	0.0106	0.9626
Random Forest Regression	0.0025	0.995	0.0039	0.995



All the Regressions have produced almost same results with very high accuracy which supports the fact that prediction is almost accurate based on test data. The sample of test data vs. predicted number supports this conclusion:

Test Data	1.002297	1.004458	1.003465	0.00377	2.000085	1.000839	0.00111	1.000334	1.002
Predicted Value(LR)	1.002297	1.004458	1.003465	0.00377	2.000085	1.000839	0.00111	1.000334	1.002
Predicted Value (KNN)	1	1	1	0	2	1	0	1	
Predicted Value(GINI)	1	1	1	0	2	1	0	1	
Predicted Value (RF)	1	1	1	0	2	1	0	1	