**Analysis of NYPD Vehicle Collision: Cause and Impact**

*by*

Rishika kesharwani

(Roll Number(s): 56)

Department of Computer Science

United University

Allahabad

April 2023

BONAFIDE CERTIFICATE

This is to certify that this project report entitled**”** Analysis of NYPD Vehicle Collision: Cause and Impact**” submitted** to **United University Allahabad**,is a bonafide record of work done by “ **RISHIKA KESHARWANI** **”** under my supervision from **“17/4/2023** to “ **23/4/2023 ”**

Kuldeep Gupta

**Department of computer application**

Date:- 25/04/2023

## Declaration by Author(s)

This is to declare that this report has been written by me/us. No part of the report is plagiarized from other sources. All information included from other sources have been duly acknowledged. I/We aver that if any part of the report is found to be plagiarized, I/we are shall take full responsibility for it.

**Rishika kesharwani**

Roll no :-56

Date: 25/04/2023

### APPENDIX 4

| **1. Introduction ..........................................................................** |  |  |  | **1** |
| --- | --- | --- | --- | --- |
| **1.1 About Data report** ....................................................... |  |  |  | 1 |
| **2. Tools ...........................................................** |  |  |  | 7 |
| **2.1 What is Data Analytics..................................** |  |  |  | **7** |
| **2.2** **What is IBM Cognos tool ………..….** |  |  |  | **9** |
| **2.3** **What is Data Module?......................................** |  |  |  |  |
|  |  |  |  |  |
| **3.1 standard based problem………………**  **3.2 Scenario based problem……………………..** | | | | | | | | |
| **2.5 What is IBM Cognos Dashboard?............................** |  |  |  |  |
|  |  |  |  | **18** |
|  | | | | | |  |
|  | | | | | |  |
|  | | | | | |  |  |  |
|  | | | | | | | | |  |  |  |  |
|  | | | | | | | | |  |  |  |  |
|  | | | | | |  |  |  |
|  | | | | | |  |  |  |
|  | | | | | |  |  |  |

**INTRODUCTION**

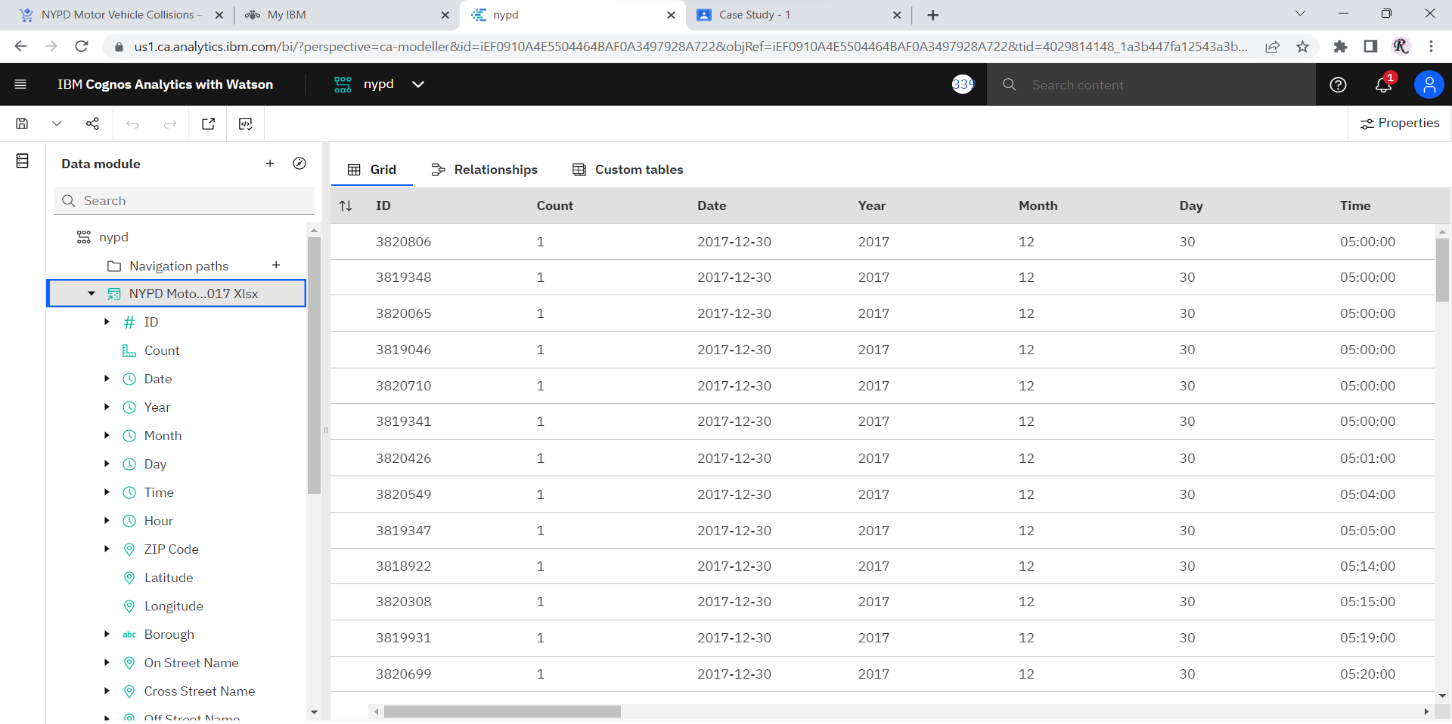
This data includes details of Motor Vehicle Collisions in New York City provided by the Police Department (NYPD) of 3 years (2015-17) . This data set contains all the information regarding to the motor vehicle.

This data shows the causes of those accident and also contain the name of those streets, which help the NYPD to easily understand the cause of these accidents and take action accordingly.

It contains all the data of total injured and killed people which are mainly :-

* Pedestrians injured
* Cyclists injured
* Motorists injured

It contains all the information about the location , place , street name and type of vehicle through which accident occur and many more .



Tools through which we are going to analysis the NYPD dataset

* IBM COGNOS:- IBM Cognos is a business intelligence and performance management software suite that helps organizations analyse and make informed decisions based on their data. The software provides tools for reporting, analysis, score carding, and monitoring, as well as planning, budgeting, and forecasting.

*This software provides tools for reporting, analysis, score carding, and monitoring, as well as planning, budgeting, and forecasting. And the reason behind using IBM Cognos is that it is a BI tool that is designed for enterprise-level organizations. It provides a comprehensive suite of capabilities for reporting, analysis, and planning. Cognos allows users to create dashboards and reports that can be accessed by stakeholders across the organization. It also offers advanced analytics capabilities, such as predictive analytics and data mining.*

* DATA MODULE;- Data module is a tool that represent simplified and unified view of data sources that are used to create reports and dashboard . it provide user friendly interface for creating relationship between data source. IBM cognos uses data module in a place raw data.
* DATA DASHBOARD :- A data dashboard is an interactive tool that allows you to track, analyze, and display KPIs and metrics. Modern dashboards allow you to combine real-time data from multiple sources and provide you AI-assisted data preparation, chart creation, and analysis.
* **IBM Cognos Dashboard:-**

*Dashboard is a place where you can explain your data visually and make the communication easily to understand. It is Data Visualizing board for Cognos.*

*There are many tools which help you to make data more attractive. Here are some tools which we are going to use or some which are present in Dashboard.*

1. *Column*

*Use a column visualization to compare values by one or more columns, such as sales*

*for products or sales for products each month.*

1. *Stacked column*

*Use a stacked column visualization to compare the proportional contributions for each*

*item to the total, such as sales for products and sales for products each month.*

1. *Bar*

*Use a bar visualization to compare values by one or more columns, such as sales for*

*products or sales for products each month.*

1. *Stacked bar*

*Use a stacked bar visualization to compare the proportional contributions for each item*

*to the total, such as sales for products and sales for products each month.*

1. *Bubble*

*Use a bubble visualization to show relationships among columns that contain numeric*

*values, such as revenue and profit.*

1. *Packed bubble*

*Use a packed bubble visualization when you want to show relationships among*

*columns that contain numeric values, such as revenue. It is similar to the bubble*

*visualization but the bubbles are tightly packed instead of spread over a grid. A packed*

*bubble visualization shows a large amount of data in a small space.*

1. *Line*

*Use a line visualization to show trends over time.*

1. *Line and column*

*Use a line and column visualization to highlight relationships between multiple data*

*series by combining bars and lines with one visualization.*

1. *List*

*Use a list visualization to create an overview the data in a hierarchical way.*

1. *Point*

*Use a point visualization to show trends over time.*

1. *Area*

*Use an area visualization to emphasize the magnitude of change over time.*

1. *Pie*

*Use a pie visualization to highlight proportions. Each slice shows the relative*

*relationship of each part to the whole.*

1. *Tree map*

*Use a tree map visualization to identify patterns and exceptions in a large, complex data asset.*

1. *Table*

*Use a table to show detailed information from your database, such as product lists and*

*customer lists. A table shows data in rows and columns. Each column shows all the*

*values for a data item in the database or a calculation based on data items in the*

*database.*

1. *Hierarchy*

*Use a hierarchy when you want to see the data in rows and columns.*

1. *Summary*

*Use a summary visualization when you want to see the total for a measure or the count*

*for a categorical column.*

1. *Radial bar*

*In a radial bar visualization, each bar appears in a circle with longer bars that represent*

*larger values. Hover over a bar to see the details about it, such as the exact value*

*represented by the bar. Each bar starts at 12 noon and goes in a clockwise direction for*

*positive values and counterclockwise for negative values.*

1. *Scatter*

*Scatter visualizations use data points to plot two measures anywhere along a scale, not*

*only at regular tick marks.*

1. *Word cloud*

*Use a word cloud visualization when you want to see a text-based visualization of a*

*given column. The text height represents the scale. The name itself is the different*

*members of the column.*

1. *Network*

*Use a network visualization when you want to see the connections among columns in*

*your data asset. A network visualization is a good choice to show connections,*

*networks, and points of intersection.*

1. *Heatmap*

*Use a heat map visualization to visualize the relationship between columns and you*

*want it to be represented in a matrix type view.*

1. *Data player*

*Use a data player to see an animation of the impact of a column on the other*

*visualizations.*

**In this case study we are going to divide the data into two problem set:-**

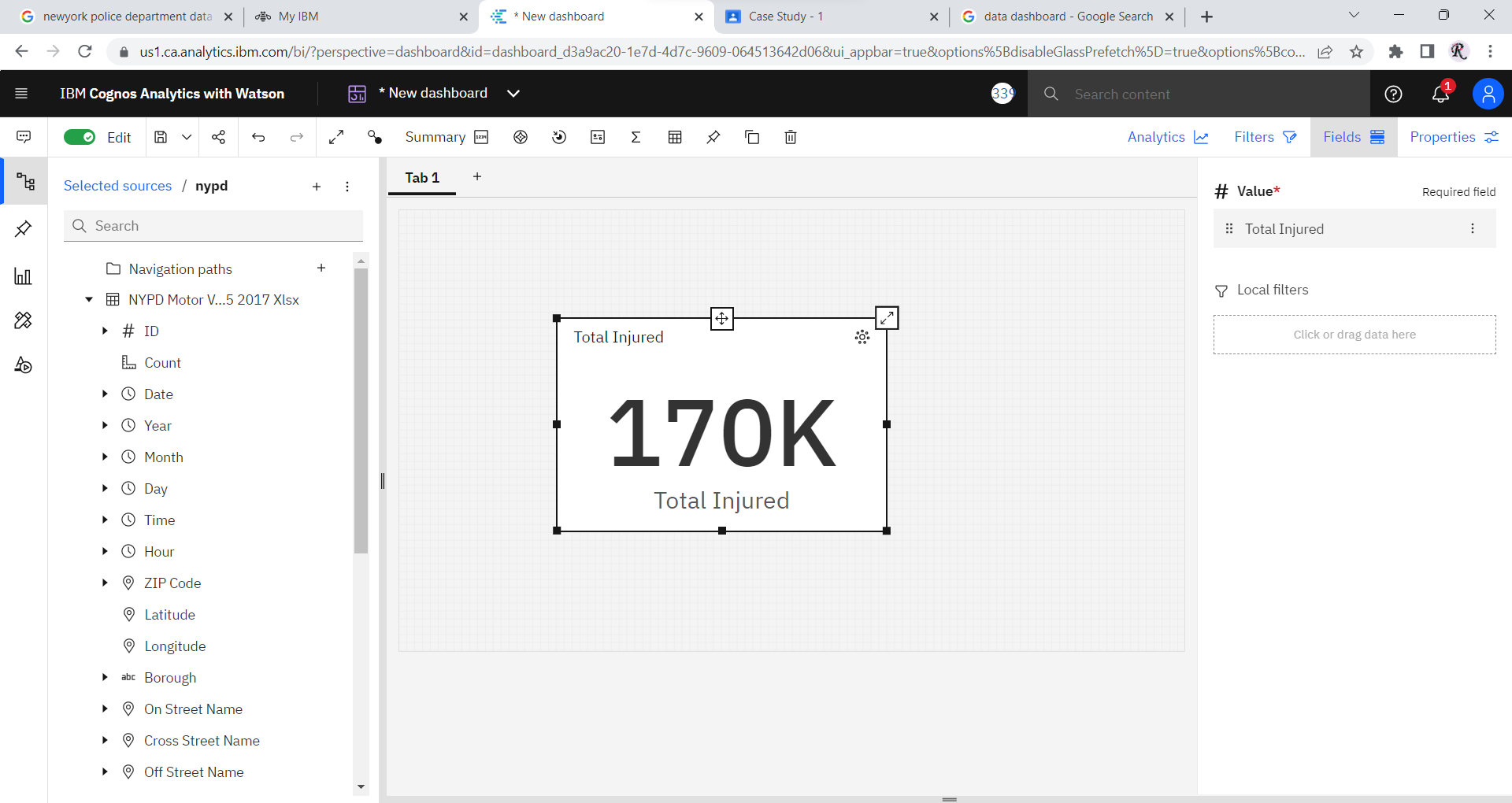
1. **Standard problem**:- In data visualization, a standard problem refers to a common question or issue that can be addressed using well-established techniques or standard visualizations. For example, if you want to compare the sales of different products over time, you might use a line chart to show the trend for each product.
2. **Scenario based problem :-** A scenario-based problem refers to a specific issue or question that requires a tailored approach or a combination of techniques to address. For example, if you want to understand the factors that influence customer satisfaction, you might use a combination of scatterplots, heatmaps, and regression analysis to identify the key drivers.

**Standard problem**

* You have to show the total number of "*total injury/total killed".*
* Show the following Summary from total injury number: Pedestrians injured, cyclist and motorcyclist injured.
* show the injuries on the basis of year.
* show the number of injuries on the basis on "vehicle 1 contribution factor" .
* show the number of injuries on the basis on "vehicle 2 contribution factor" .
* show the number of injuries on the basis on "vehicle 3 contribution factor"
* Give the number of pedestrian/cyclist/motorist injured for on the basis on any particular city, month and year.
* Give the map view of any city street where accident injury rate is high.
* Give the total number of all kind of injuries on the basis on month in a crosstab

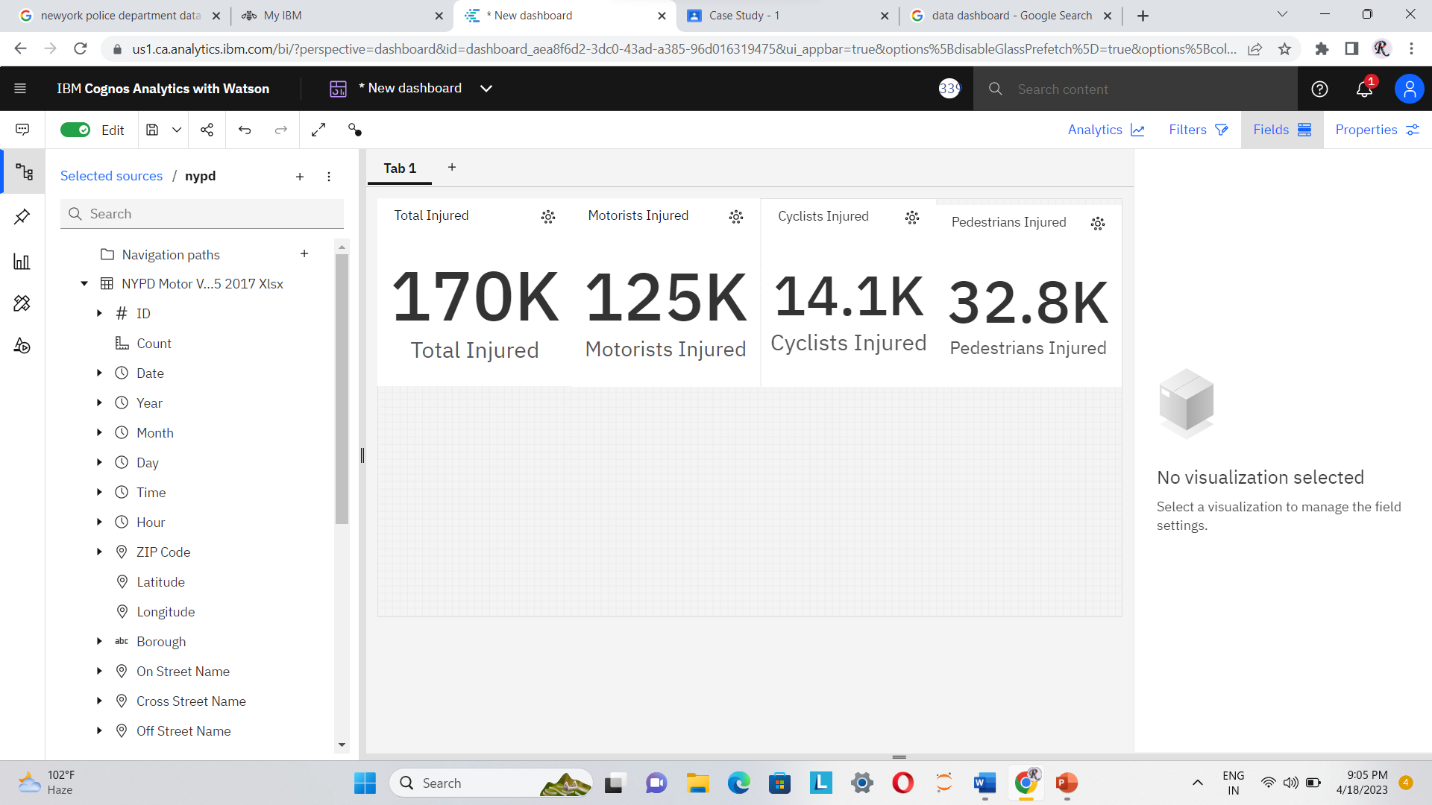
**(1)**

Total injury/total killed are shown by using **summary** tool.



**(2)**

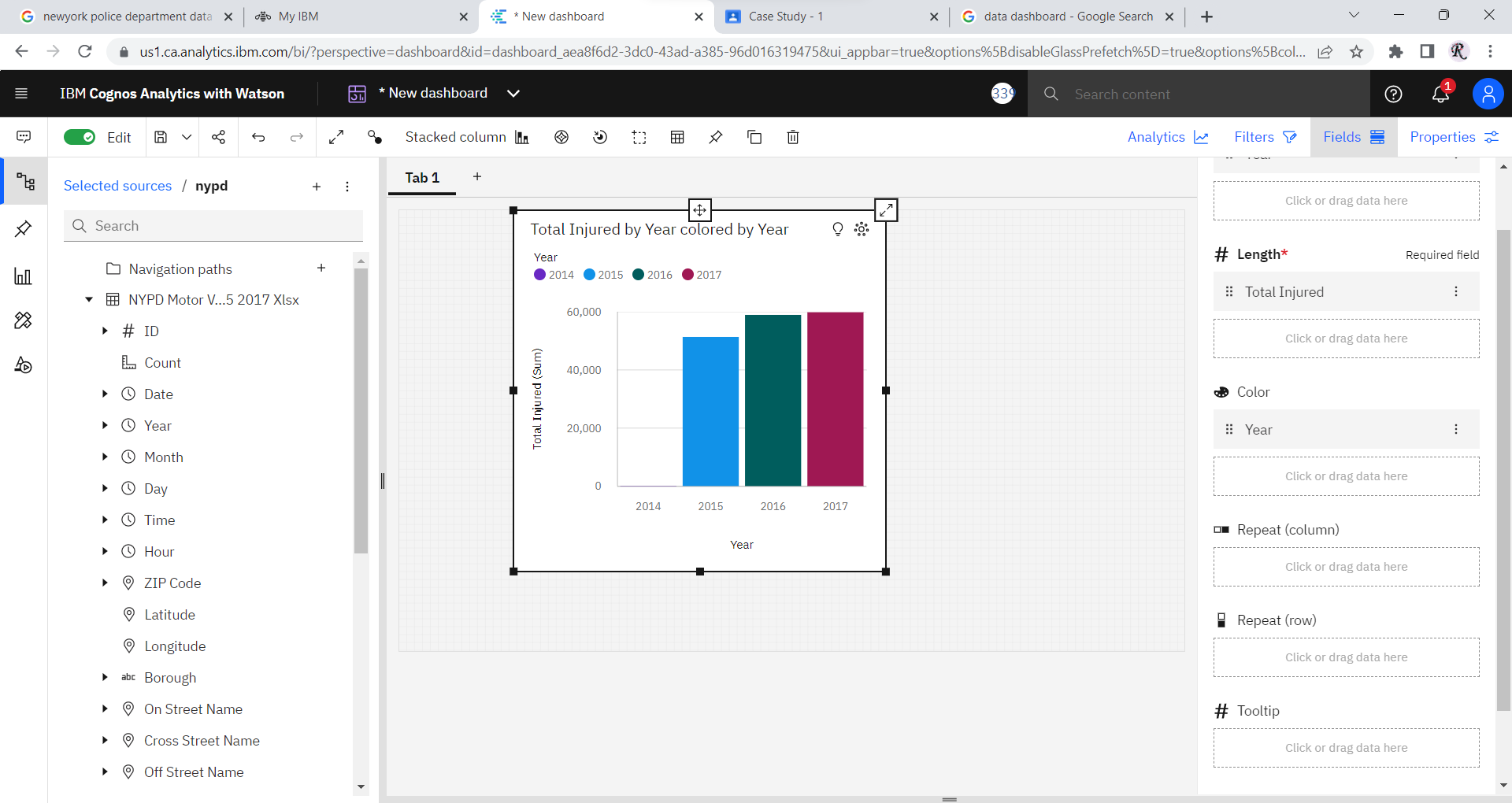
We are using the summary tool to show all the summary of data.



**(3)**

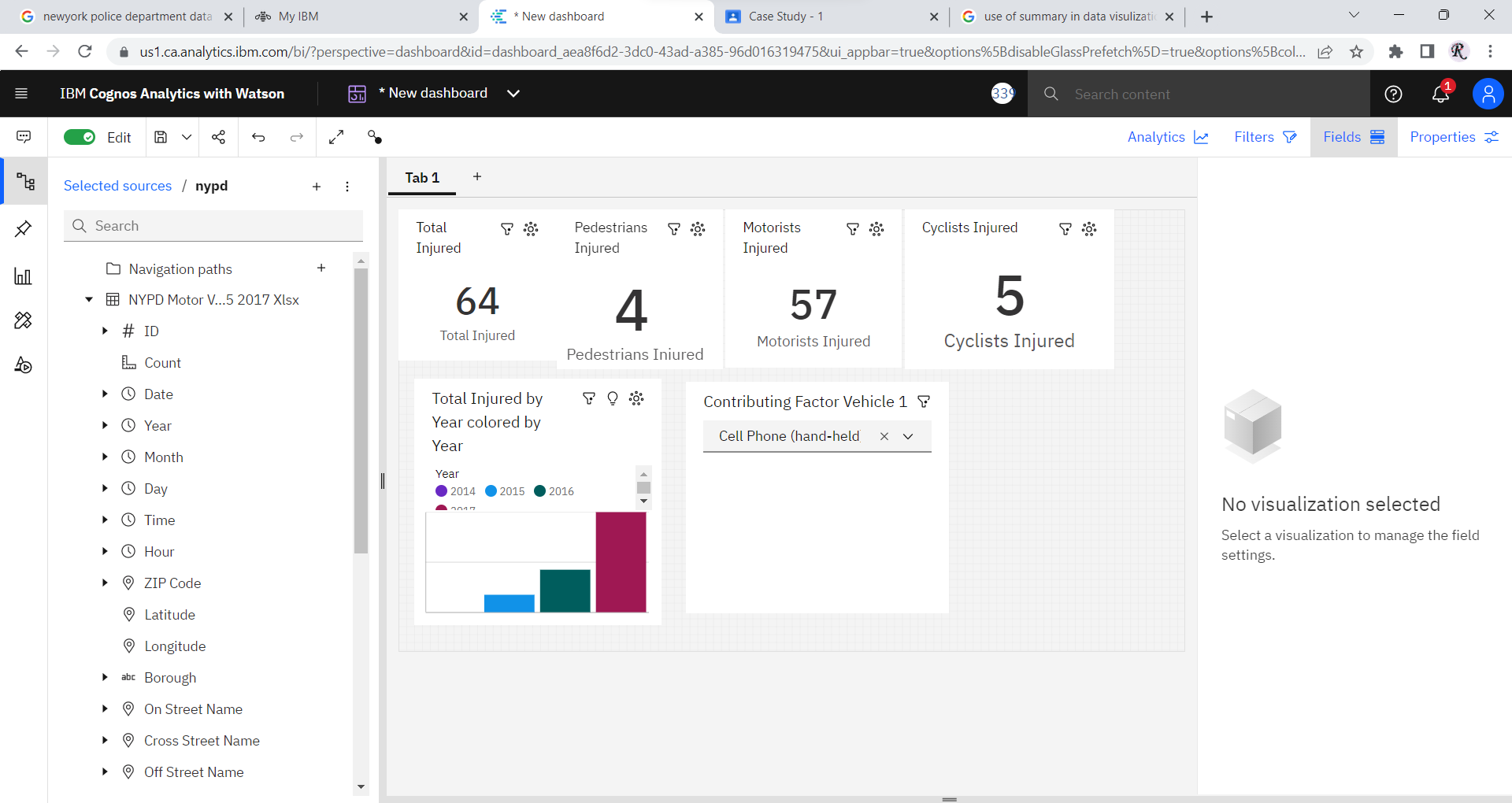
we use stacked column chart to show the injuries based on year

Stacked Column Chart comprises several column series stacked vertically, one on another. The length of each series is determined by the value in each data point. Stacked Column Charts are a great option if you need to simultaneously observe how each of several variables and their sum change.



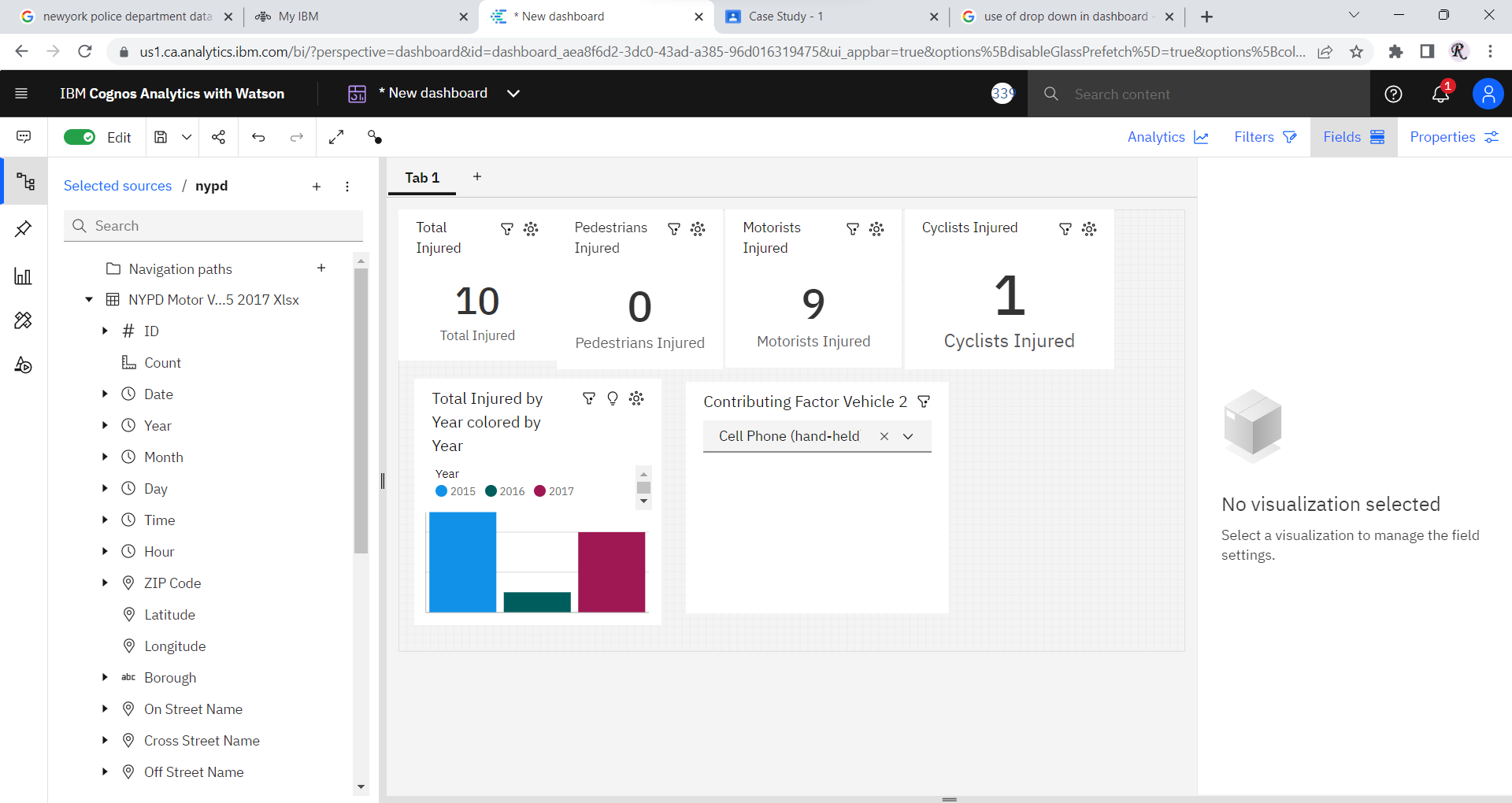
**(4)**

we use drop down to show the vehicle 1 contribution factor

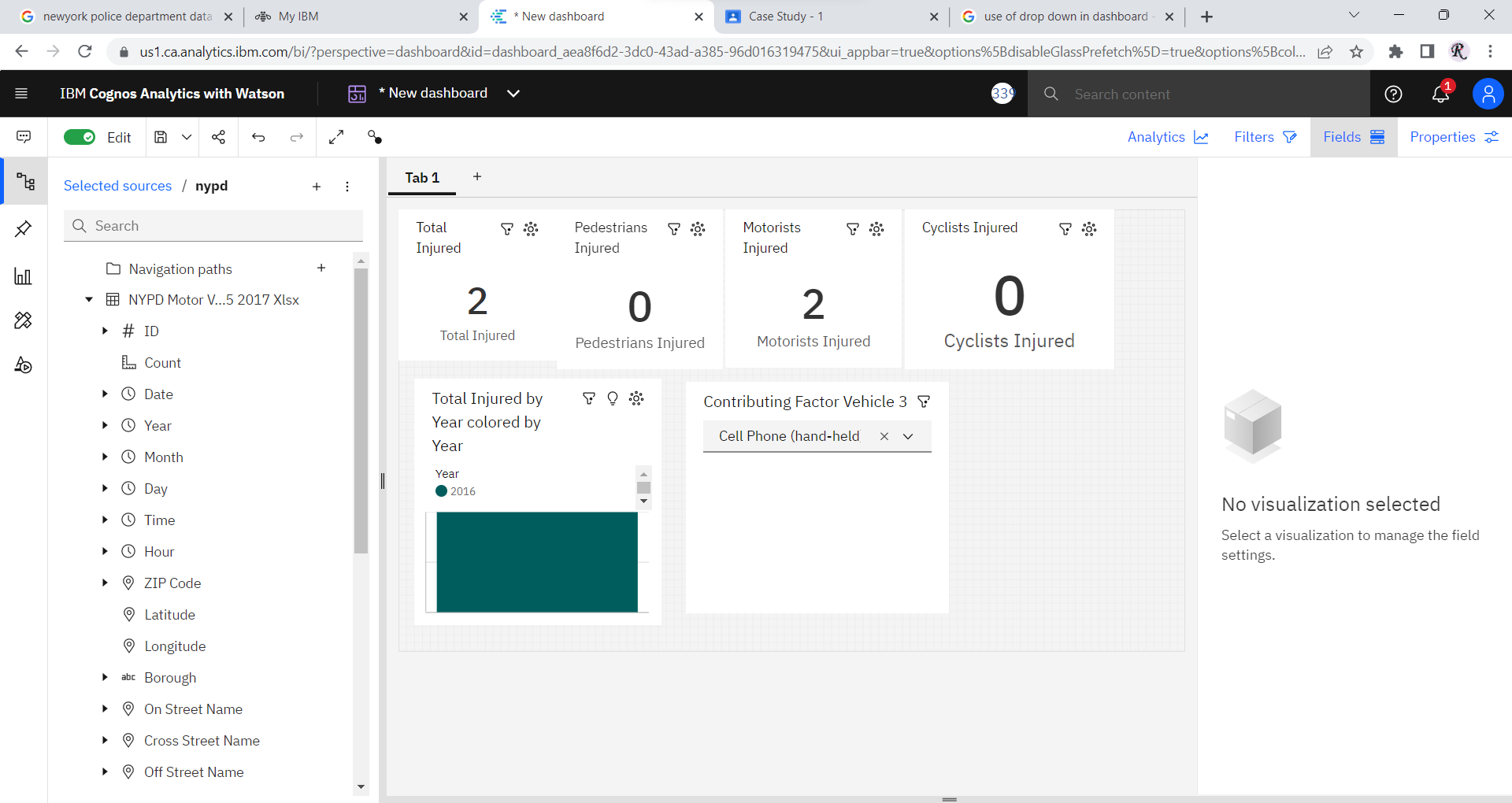


A drop-down list is a graphical control element, similar to a list box, that allows the user to choose one value from a list.

**(5)** Data of vehicle 2 contribution factor with cell phone (hand-held).



**(6)** Data of vehicle 3 contribution factor with cell phone (hand-held).

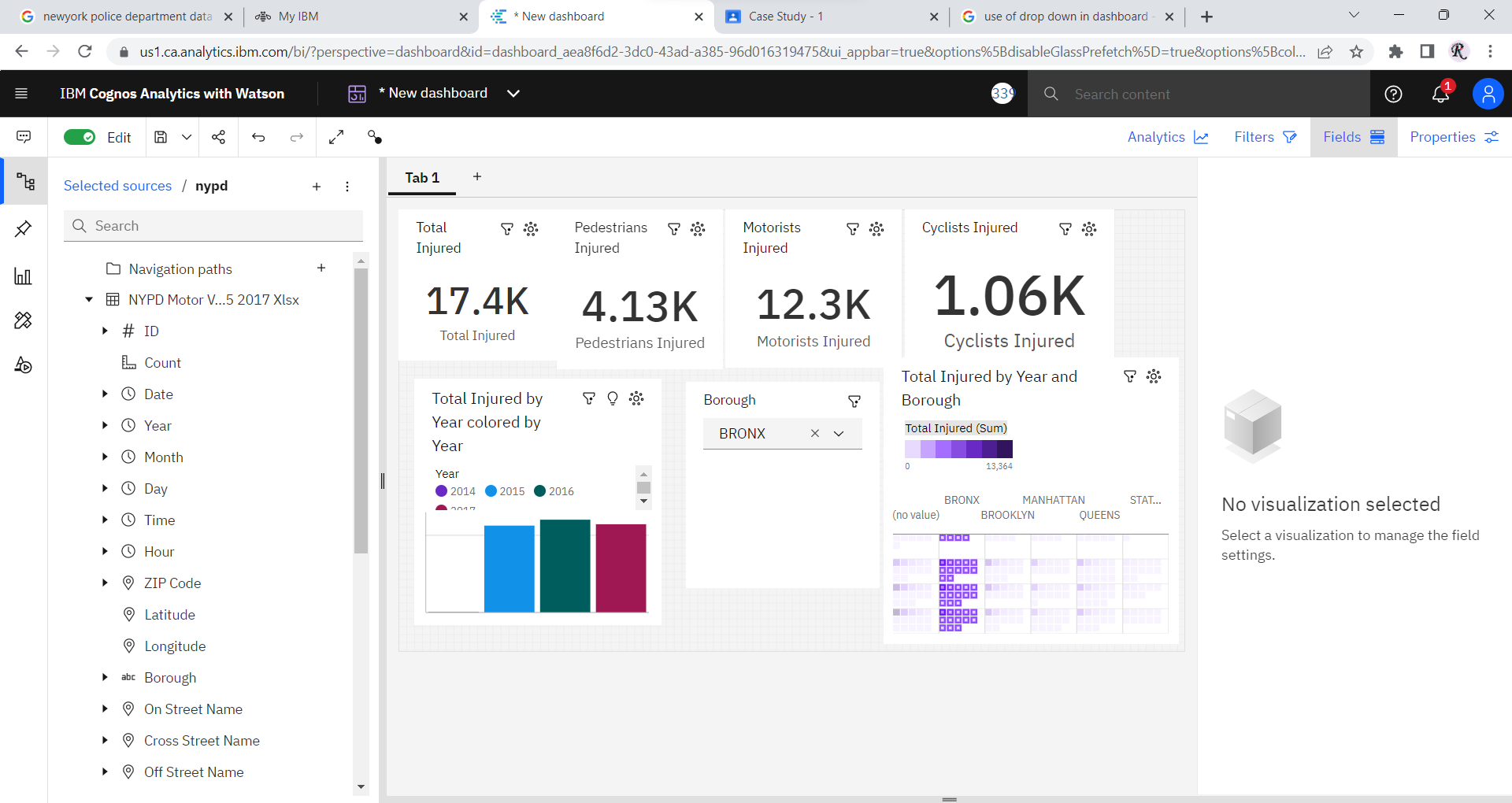
****

**This Data show that in 2016, Total 2 people get into accident because of Cell-Phone(hand-held).**

**(7)** The number of pedestrian/cyclist/motorist injured for on the basis on any particular city, month and year.

we use drop down tool for city (borough)

we use heatmap for showing the data of month and we use column stacked for showing year.



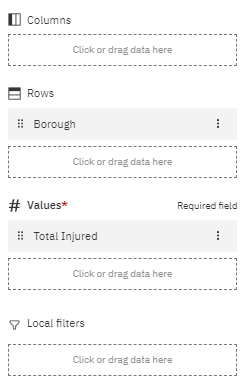
we use crosstab to show the total number of injured on the basis of month.

(8) For this query we are going to use multiple tools:

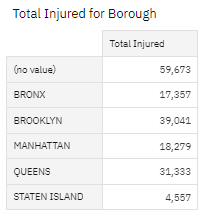
* Cross-Tab
* Network

Now, let’s create the solution of this query:

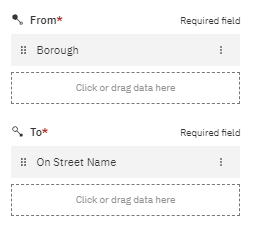
1. So, first we are going to visualization tab and select **cross-tab.**
2. After that we will select the data for values and rows in field:



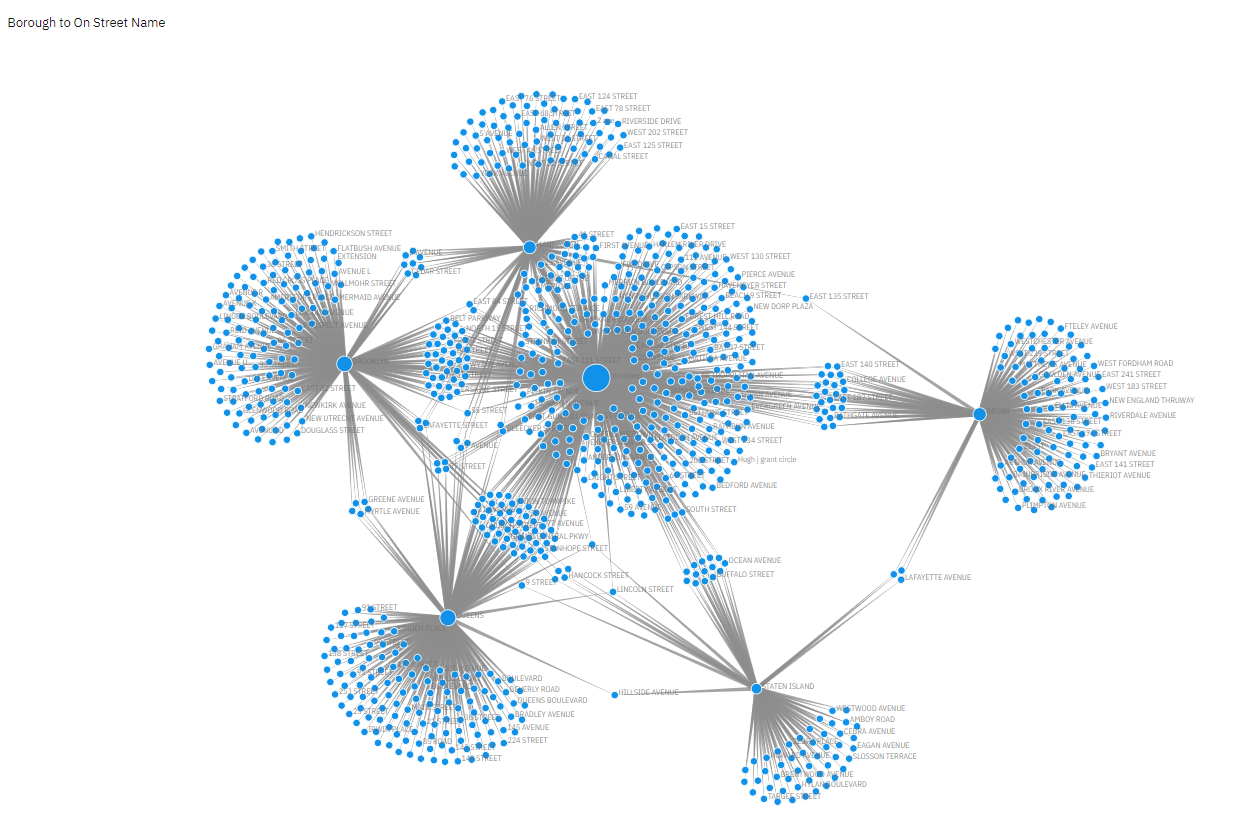
1. After that you will get the result of **cross-tab**:

****

1. Now, we will select **network** **tool** from visualization tab drag-drop or double click and open it.
2. After that select these data for **from** and **to** in field:



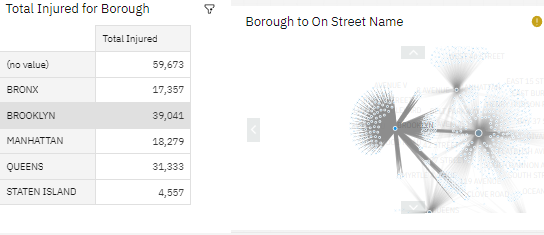
1. Now, the result of this will be shown like this:



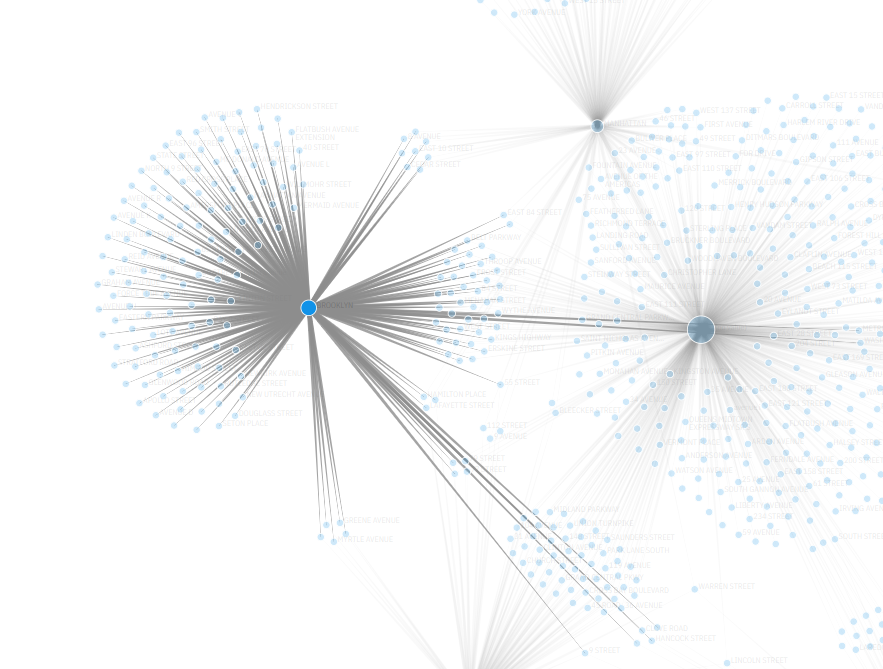
1. Now, again go to the **cross-tab tool** and select the city with high rate of injuries. (For example, I am selecting Brooklyn because it has high rate of injured)

[Note- Here, in cities there a category no value (which is made because of data error)]

1. After that in **network tool** the result will be like this:

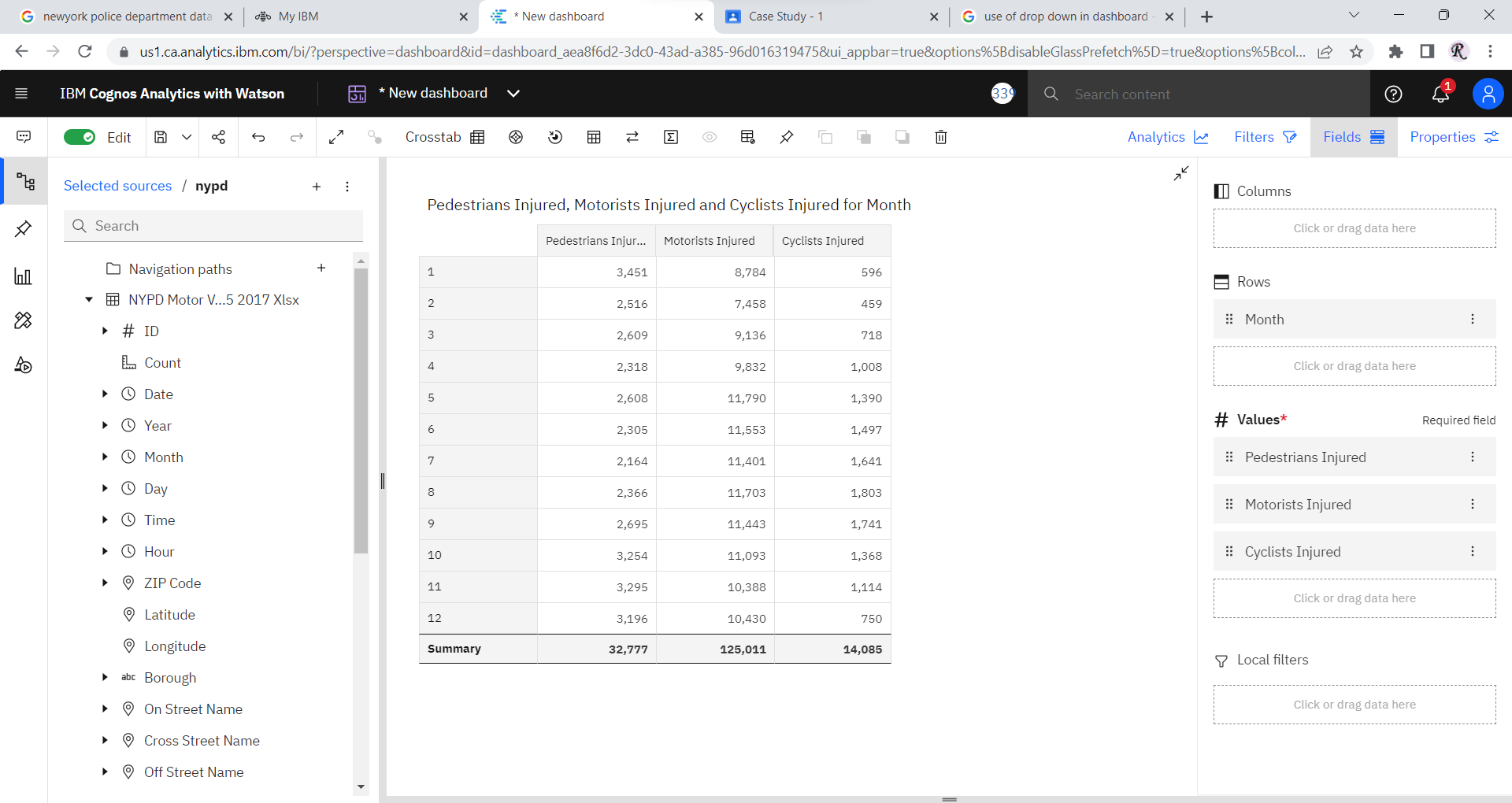


*Here, it is showing that we select ‘Brooklyn’ city and the* ***network tool*** *change its value.*



*Here is the close look of the network graph which is showing all the streets which are connected to ‘Brooklyn’ city, with the help of this we can simply view those streets.*

**9.** The total number of all kind of injuries on the basis on month in a crosstab

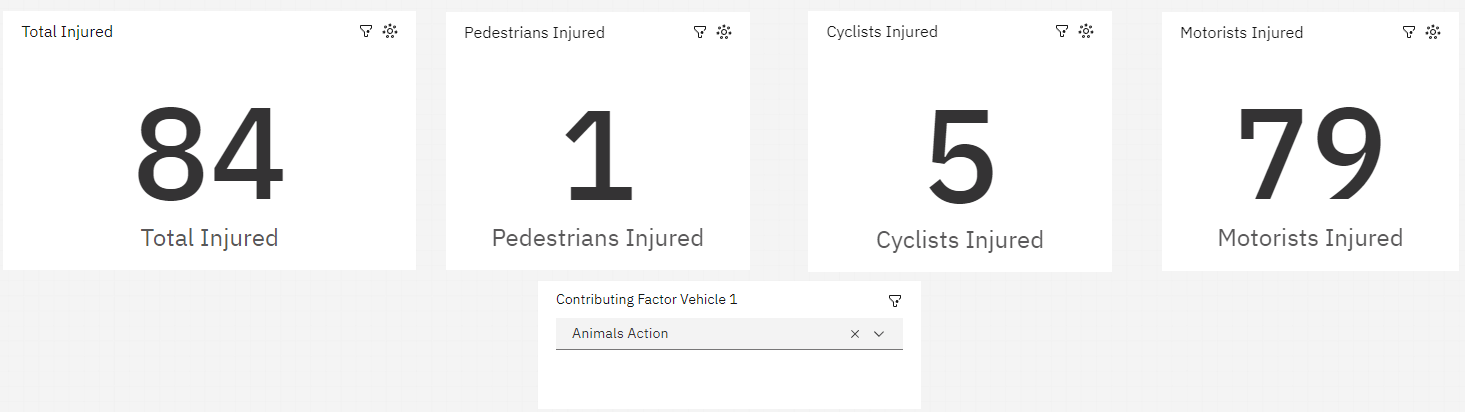


**Scenario based problem**

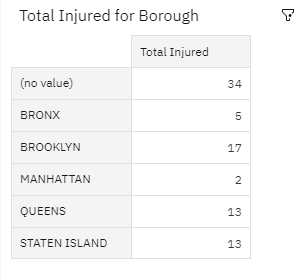
* New york animal welfare department wants to sub due the rate of accident due to Animal accident factor. please tell us which particular city and on which particular street they have to focus on.
* NY traffic department wants to spread the traffic awareness knowledge to some local citizens who are not properly aware about the rules. so please help us from which city and street we start and why?
* NY Road community auth. want to create some new street lane for  more than 6 tiers vehicle. so suggest me any top 7 locations in new york in which we can build a new lane for heavy vehicles and why?



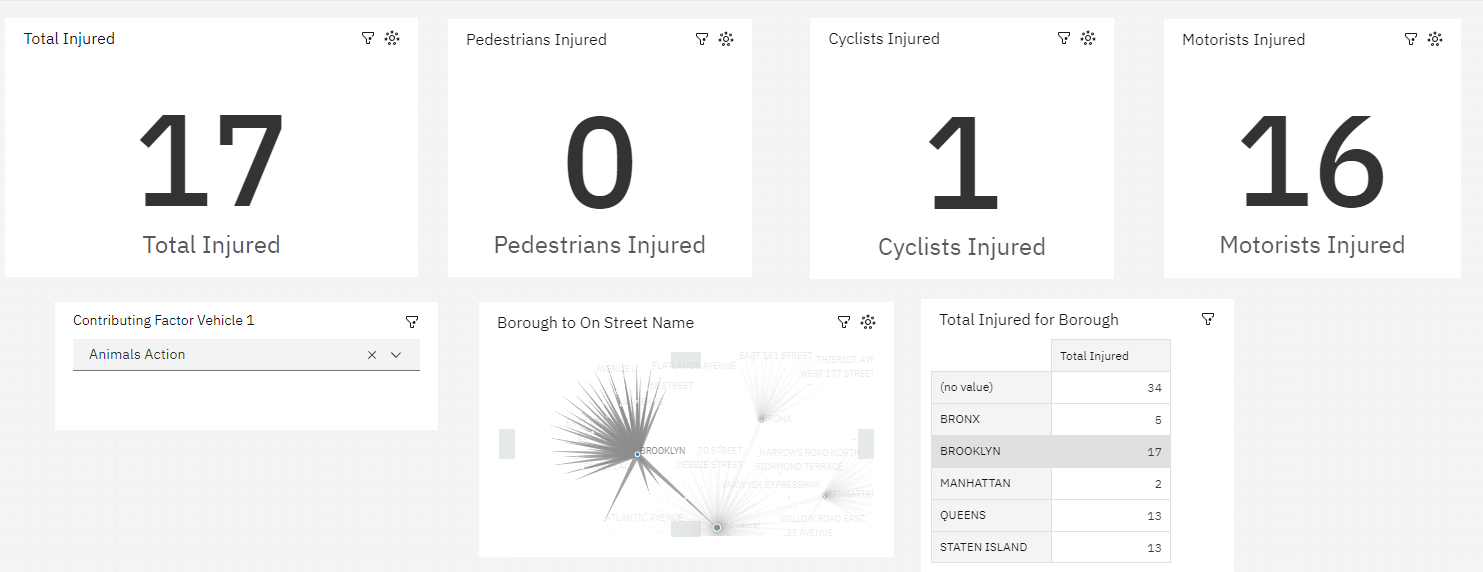
.First let us understand the number of injuries which are cause by Animals Action:



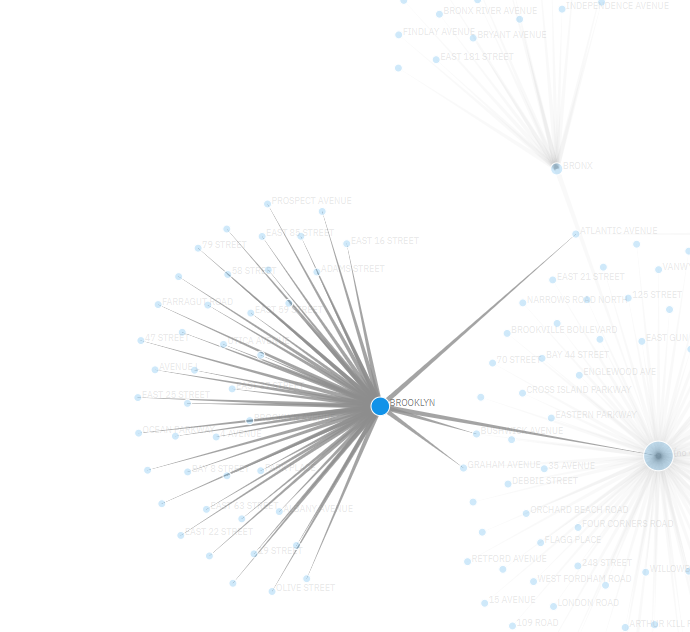
According to this data we can clearly see that total 84 people are injured from which 1 were pedestrian, 5 were cyclists and mostly people are injured is Motorists. From this we can understand that accident is mostly cause due to speeding of vehicle.



*Here is another data which is showing where the most accident cause due to animal action in these cities. So, we can select Brooklyn city where most accident cause due to animal action.*

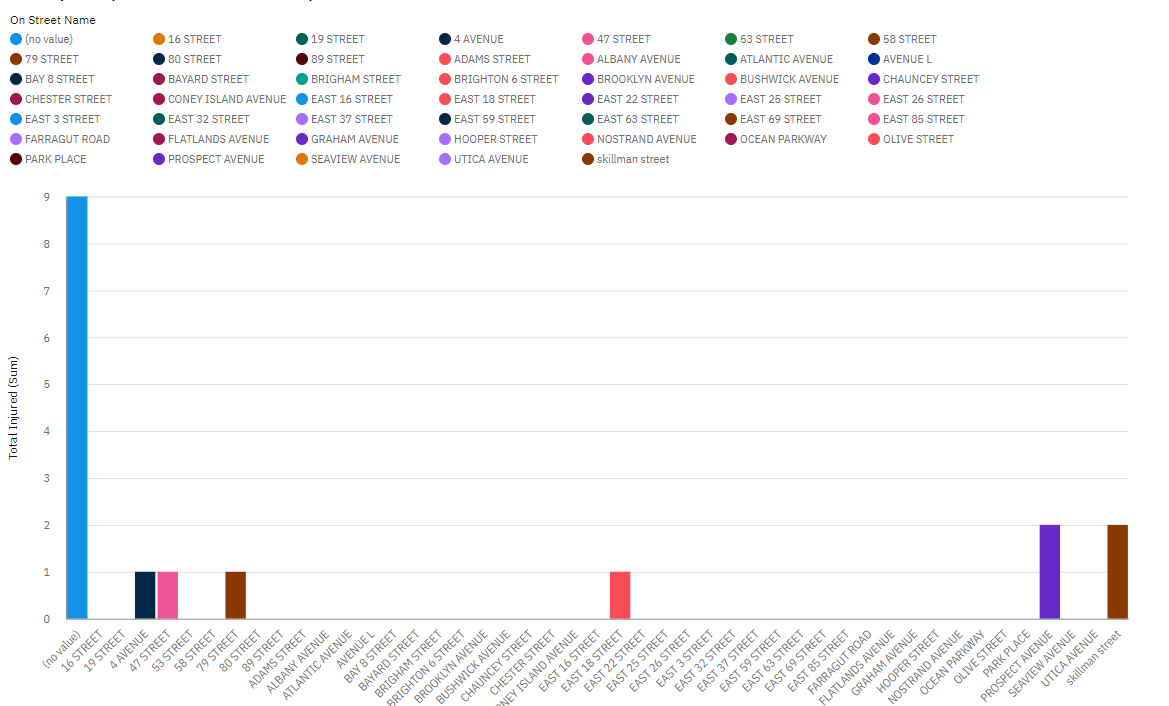


*As we can see here is the data which is showing out of 17 injured 16 were motorists and 1 is cyclist.*



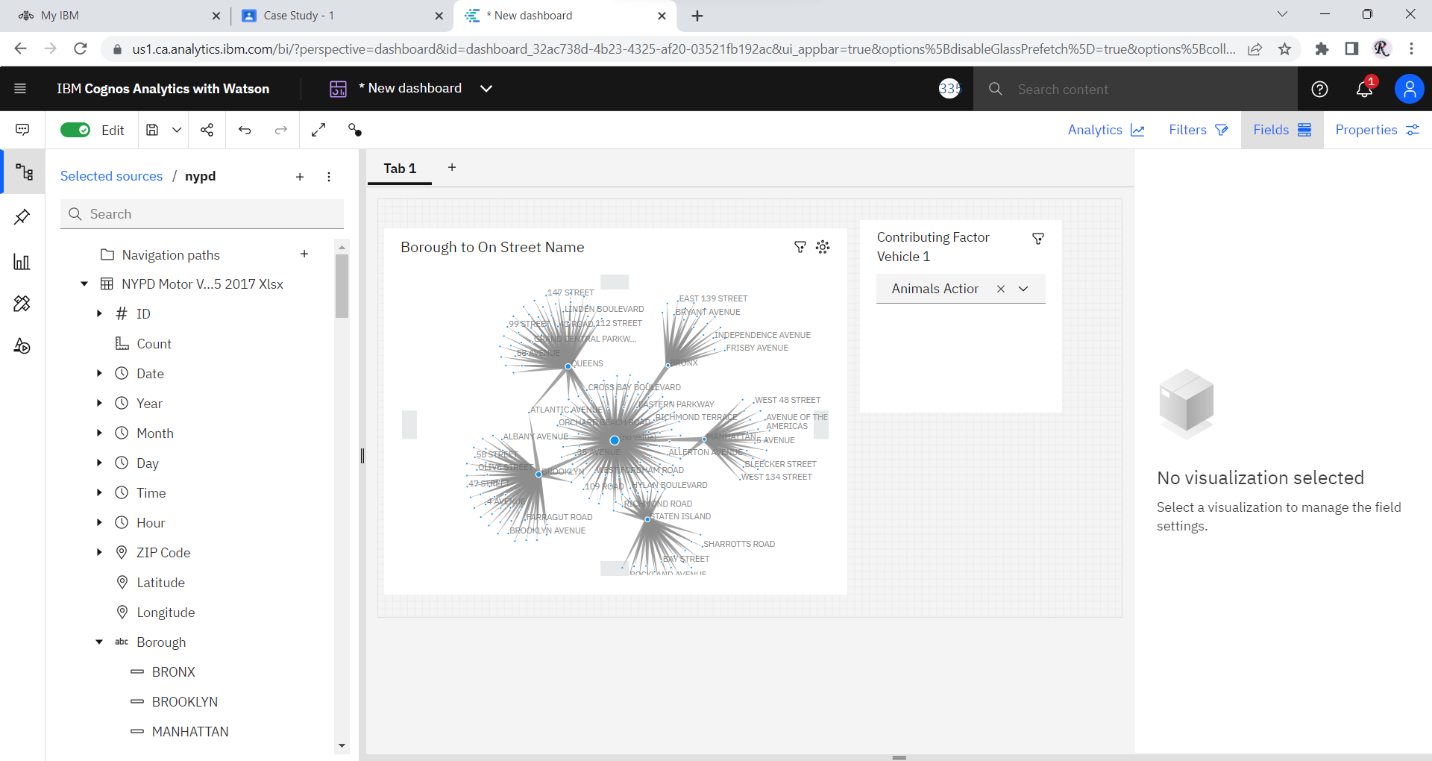
As you can see these are the streets where accident rate is high in Brooklyn city.

With this we data we come to the conclusion that New York Animal department should start with Brooklyn City and Prospect Avenue & Skillman streets (Both these streets contain same number of injured).



Here are some steps which can be taken by New York Animal department to decreases this accident:

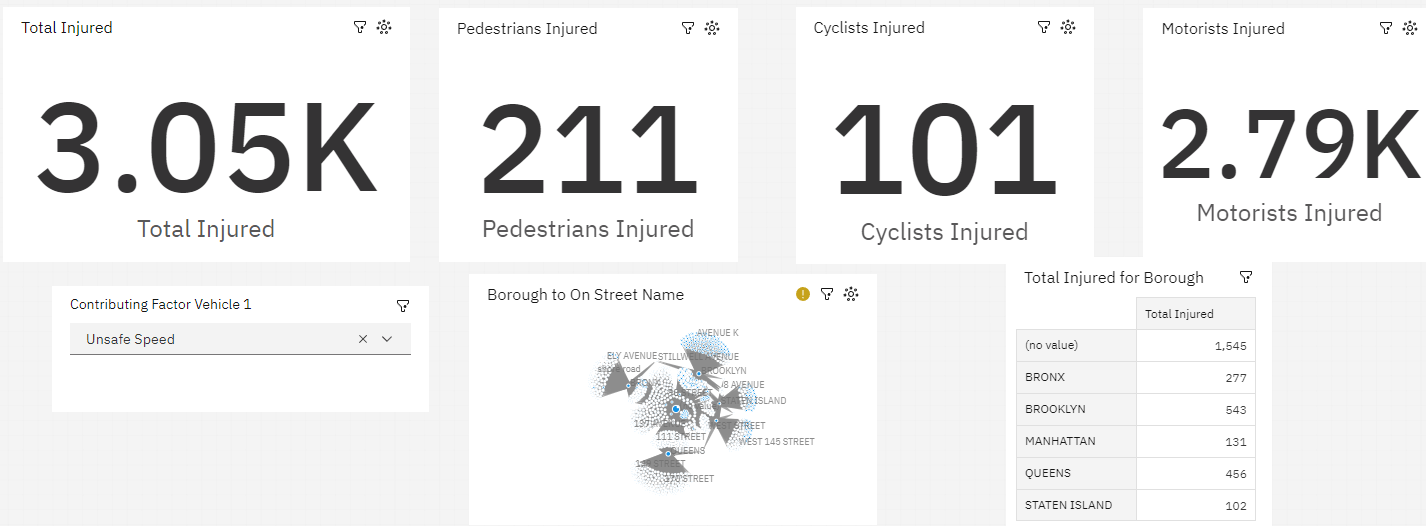
1. Sign board: They can use speed sign boards or Animal sign board so that people get alert.
2. Speed breakers: This can be use to reduce the speed of Motorcyclist as they most injured rate are cause by them.
3. Fencing: They can use fencing to stop the animal to come on these roads.

****

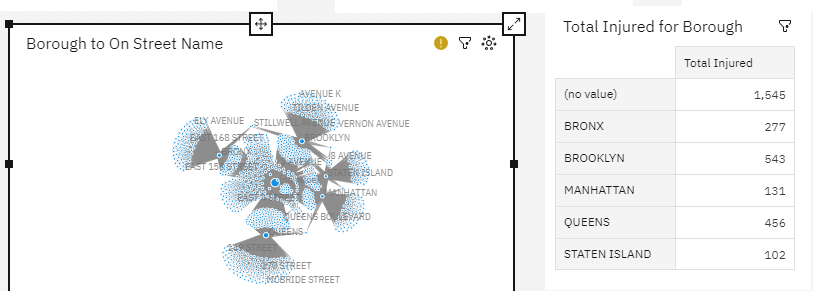
1. Firstly let understand what are the some of the traffic rules at New York:
2. Speed Limits
3. Traffic Signals and signs
4. Driving under the influence: It is illegal to drive in the influence of Drugs and alcohol.
5. Cell phone use
6. Right of way: Drivers must yield to pedestrians and other vehicles that have the right-of-way.
7. Passing and changing lane: Drivers must use turn signals when changing lanes or passing other vehicles.

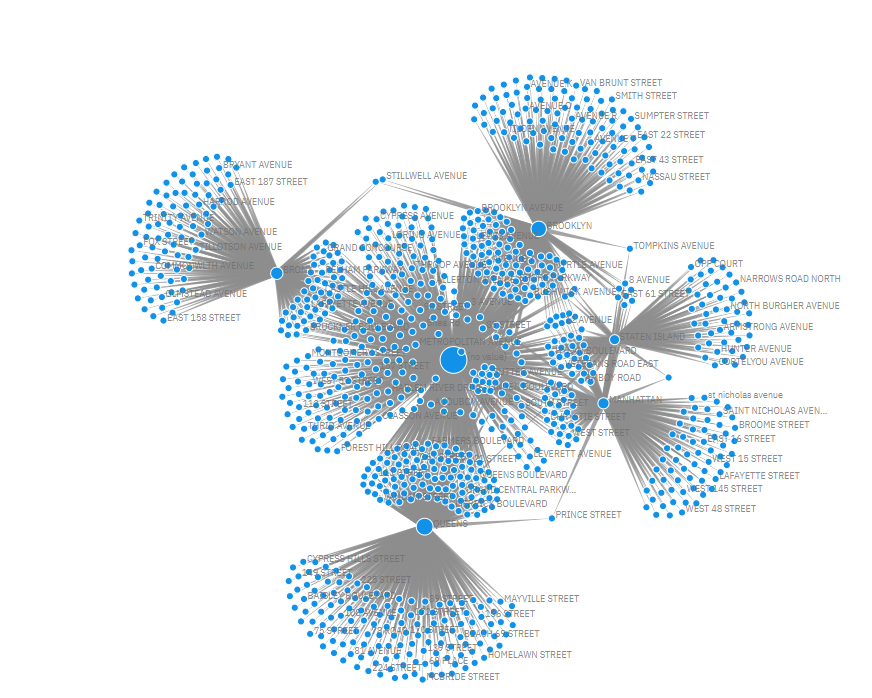
These are the some of the Traffic rules you must follow while driving, now let start our analysis on data according to the information we retrieve:

So, Let start with Injuries that cause due to over speeding-



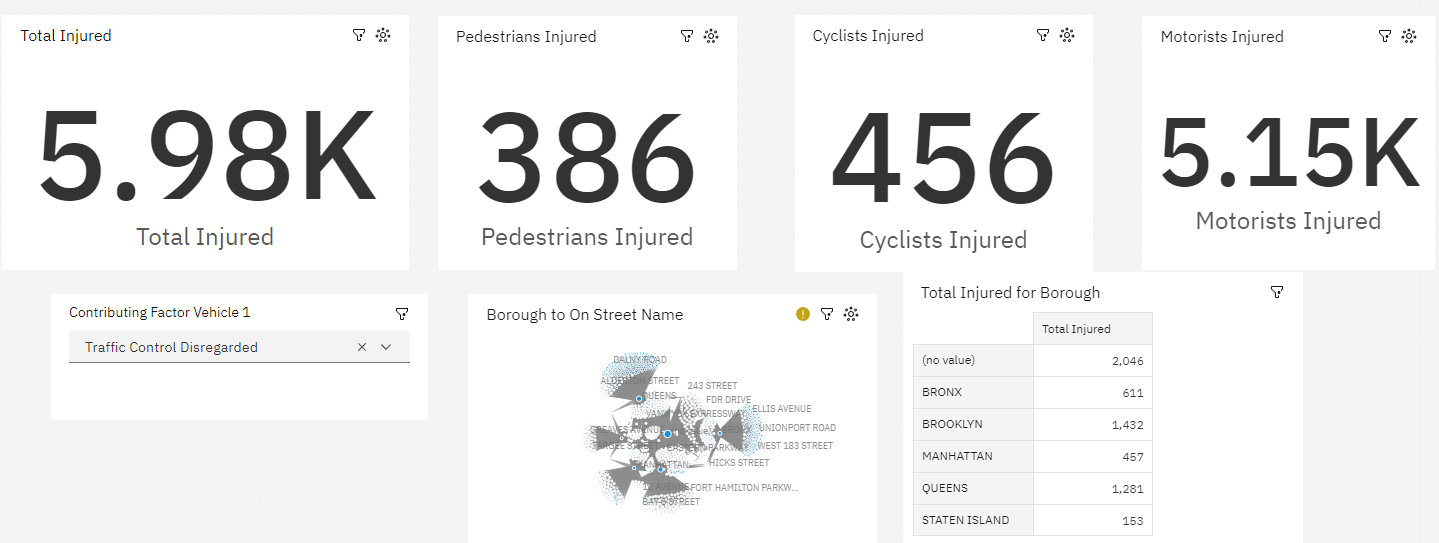
*Here, you can see these are the accident cause due to Over speeding. About 3,050 people are injured from which 211 are Pedestrians, 101 are cyclists and 2790 are Motorists. (There is some error because of un-enter data so the sum might be different).*



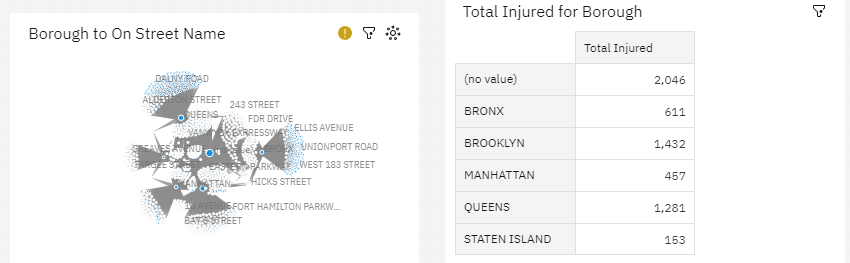


*Here is network map that is showing the streets of New York where the accident causes due to Over speeding.*

Now, we are going to show the injuries which is cause due to Traffic signals and signs-



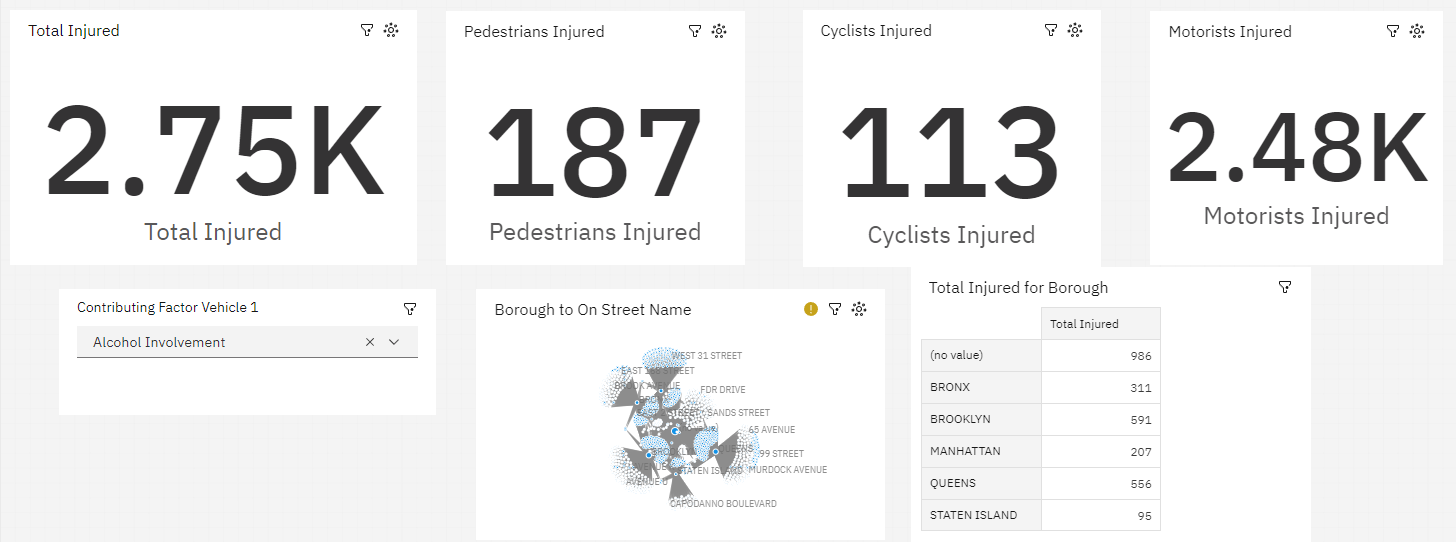
*Here, you can see that these are the injuries which are cause due to Traffic signal and sign. With this we can derived that total 5980 people are injured due to this, from which 386 are pedestrians, 456 are cyclists and 5150 are motorists’ injuries.*



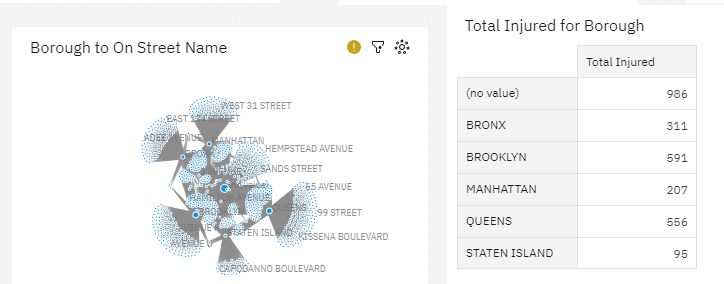


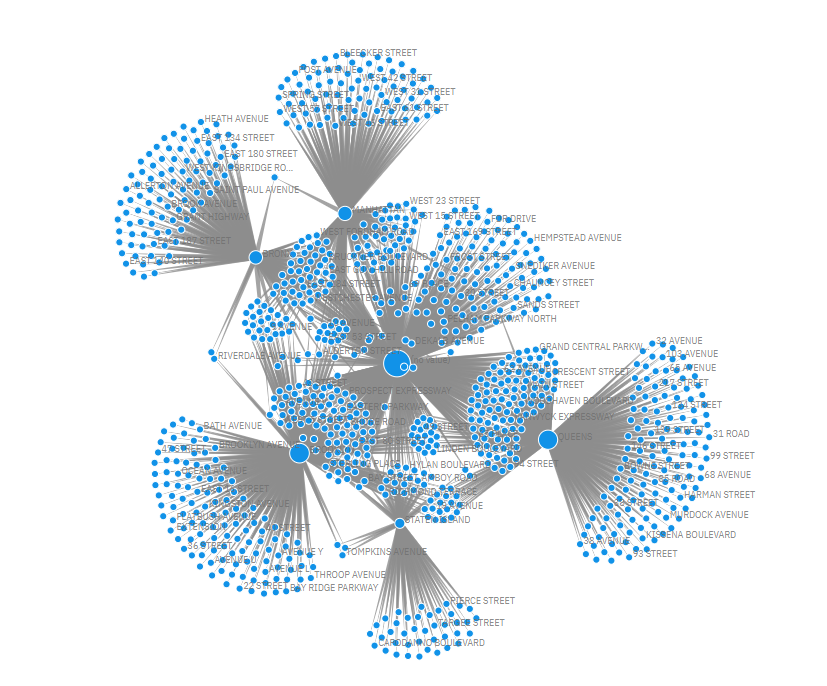
*Here, this map shows the street in which accident cause due to this situation.*

Now, we are showing accident cause due to driving under the influence-



*Here, you can see number of people get injured because of this cause. About 2750 people are total injured from which 187 are pedestrians, 113 are cyclists, and 2480 are Motorists.*

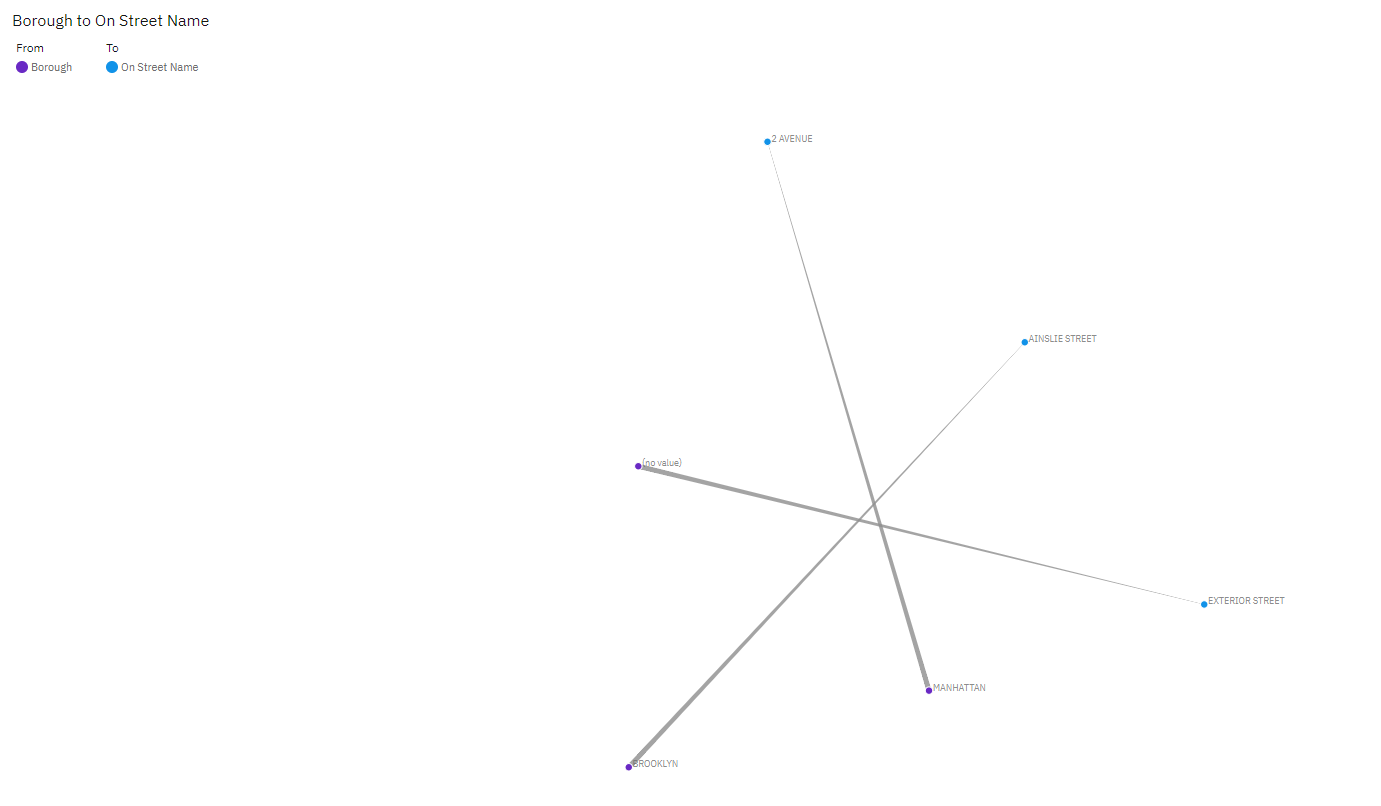




*Here, is the map view which is showing the injuries cause by this situation .*

(3) **.**For this query we going to select- 1st Type of Vehicle and 2nd Contribution factor for in proper lane.

These data will help us to gain the answer for the query. Here, we are getting 5 streets location (not including no value).

****

Streets names are –

1. 2 Avenue
2. Ainslie Street
3. Exterior Street
4. Manhattan
5. Brooklyn

Reason behind making a new lane for 6 tiers vehicle is-

1. Increase Capacity: New lanes for 6-tiers vehicle will help in increasing the capacity of the road network, allow them to transport more good eaisly and efficently.This can help in improving the overall flow of vehicles.
2. Safety: As 6-tiers vehicle contain heavy loads that increase the load on tiers, there stopping disctance and also have slower accelaration. Making the different lane will deacrase the cases of injuries and death beause of these vehicle.

Overall, creating new lanes for 6-tier vehicles can help to improve the efficiency and safety of the transportation system, reduce congestion, and have a positive impact on the environment.

**Conclusion**

The data collected and analyzed by NYPD Traffic department can provide us with valuable insides for traffic, Accidents location, safety masseurs, and other trend related to the Transportation of New York City.

By using this 3-year data (2015-2017), we can predict the cause of the accident and take measures to stop them. This will also help Government to make new traffic laws according to the problem facing, this will help in improving the traffic safety, reduce traffic and create a more effective and sustainable transportation system for the city. Also, with this data we can take measures for accident in these cities and reduce the number of death and injuries for the people.