

# **CASE STUDY 1**

## **NYPD DATA SET**

### **COLLEGE PROJECT REPORT**

*by*

**AMIT**

ROLL.NO 14

Department of Computer Science

United University

Allahabad

2023

## **BONAFIDE CERTIFICATE**

This is to certify that this project report entitled "**Analysis NYPD Vehicle collision: Cause and Impact**" submitted to **United University Allahabad**, is a Bonafide record of work done by "**AMIT**" under my supervision from "**13April,2023**" to "**25April,2023**"

**MR. Kuldeep Gupta**

IBM Trainer

**MR. Prashant Shukla**

Head of Department  
Computer science

Place: United University

Date: 25/04/2023

**Declaration by Author**

This is to declare that this report has been written by **AMIT**. 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.

**AMIT**  
Roll no.14

Place: United University

Date: 25/04/2023

## **TABLE OF CONTENTS**

### **CHAPTER NO. TITLE PAGE NO.**

**ABSTRACT iii NOMENCLATURE xvii CONTENTS xx**

<b>1. Introduction .....</b>	<b>1</b>
<b>1.1 Problem addressed .....</b>	<b>1</b>
<b>1.1.1 .....</b>	<b>3</b>
<b>1.1.2 .....</b>	<b>5</b>
<b>1.2 Related Queries to Perform .....</b>	<b>13</b>
<b>1.1 Query 1 .....</b>	<b>5-7</b>
<b>1.2 Query 2 .....</b>	<b>7-9</b>
<b>1.3 Query 3 .....</b>	<b>9-11</b>
<b>1.4 Query 4 .....</b>	<b>11-12</b>
<b>1.5 Query 5 .....</b>	<b>12-13</b>
<b>1.6 Query6.....</b>	<b>13-15</b>
<b>1.7 Query7.....</b>	<b>15-17</b>
<b>1.8 Query8.....</b>	<b>17-18</b>

**1.**

## Introduction

### About the Data

The NYPD dataset provides information on crimes reported to the New York City Police Department. The dataset includes information on the date, time, location, and type of crimes reported, as well as additional details such as the victim and suspect demographics, weapon used, and the outcome of the investigation. Some of the specific information that can be found in the NYPD dataset includes:

**1. Crime type:**

The dataset includes information on various types of crimes, such as murder, rape, robbery, burglary, and auto theft.

**2. Crime location:**

The dataset includes information on the location of crimes, such as the borough, precinct, and specific address where the crime occurred.

**3. Date and time:**

The dataset includes information on the date and time when the crime was reported to the police.

**4. Victim and suspect demographics:**

The dataset includes information on the age, gender, and race/ethnicity of both the victim and the suspect.

**5. Weapon used:**

The dataset includes information on the type of weapon used in the crime, if applicable.

## TOOL-

We are using IBM Cognos for data analytics. So, first question arises it that-

### **What is Data Analytics?**

Data Analytics is the process of collecting, processing, and analyzing data to uncover insights, trends, and patterns that can be used to make informed decisions. It involves using various tools and techniques to extract valuable insights from large and complex datasets.

### **What is IBM Cognos tool and why we are using it?**

IBM Cognos is a business intelligence and performance management software suite that is designed to help organizations make better decisions by providing a comprehensive view of their data. The tool allows users to access, analyze, and report on data from various sources, such as databases, spreadsheets, and other business systems.

IBM Cognos can be used for a variety of purposes, such as creating dashboards, reports, and scorecards to track key performance indicators (KPIs) and monitor business performance. The tool can also be used for ad hoc data analysis, to identify trends, and to make predictions based on historical data.

IBM Cognos is popular among businesses because it offers a range of features and capabilities that make it easy to use and highly customizable.

### **What is Data Module?**

Data Module is a metadata layer that provides a simplified and unified view of data sources that are used to create reports and dashboards. The data module provides a user-friendly interface for creating relationships between data sources, defining calculations and aggregations, and specifying data governance rules. This makes it easier for business users to access and work with data, without requiring specialized technical skills.

### **What is Data Visualization?**

Data visualization is the representation of data in a graphical or visual form. It involves

creating visual representations of data to communicate complex information clearly and effectively. Data visualization is an essential aspect of data analytics, as it helps to convey

insights and trends in a more intuitive and understandable way.

## What is IBM Cognos Dashboard?

IBM Cognos Dashboard is a business intelligence tool that allows users to create custom dashboards and visualizations to analyze and present data from various sources. The tool is part of the IBM Cognos Analytics suite and provides users with an easy-to-use interface for designing and sharing dashboards with colleagues.

With IBM Cognos Dashboard, users can drag and drop different widgets and charts onto their dashboard, connect to multiple data sources, and configure various settings to create a customized view of their data. The tool supports a wide range of data visualization options, including tables, charts, maps, and gauges.

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.
2. 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.
3. Bar:  
Use a bar visualization to compare values by one or more columns, such as sales for products or sales for products each month.
4. 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.
5. Bubble:  
Use a bubble visualization to show relationships among columns that contain numeric values, such as revenue and profit.

6. 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.
7. Line:  
Use a line visualization to show trends over time.
8. Lined column:  
Use a line and column visualization to highlight relationships between multiple data series by combining bars and lines with one visualization.
9. List:  
Use a list visualization to create an overview the data in a hierarchical way.
10. Point:  
Use a point visualization to show trends over time.
11. Area:  
Use an area visualization to emphasize the magnitude of change over time.
12. Pie:  
Use a pie visualization to highlight proportions. Each slice shows the relative relationship of each part to the whole.
13. Tree map:  
Use a tree map visualization to identify patterns and exceptions in a large, complex data set.
14. 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.
15. Hierarchy:  
Use a hierarchy when you want to see the data in rows and columns.



16. Summary:

Use a summary visualization when you want to see the total for a measure or the count for a categorical column.

17. 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.

18. Scatter:

Scatter visualizations use data points to plot two measures anywhere along a scale, not only at regular tick marks.

19. 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.

20. 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.

21. Heatmap:

Use a heat map visualization to visualize the relationship between columns and you want it to be represented in a matrix type view.

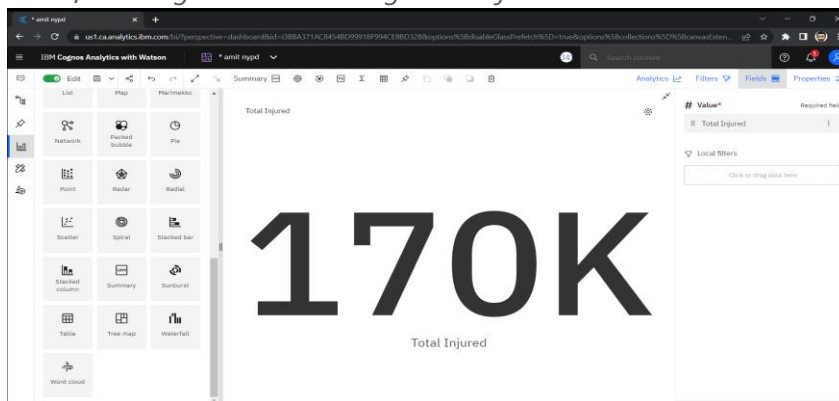
22. Data player:

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

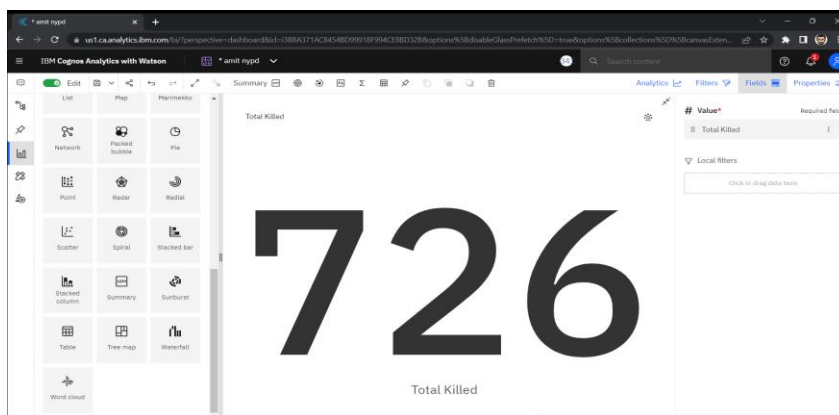
## 1. Show the total number of total Injury and total killed?

To see the total number of total injuries and total killed we have to navigate in dashboard section In IBM Cognos.

1. Navigate to dashboard and select blank page.
2. Navigate to visualization tool and drag drop summary tool in canvas.
3. Select total injured in value field.
4. Now following table is showing total injured numbers.



5. To see total killed data
6. Copy summary tool and select total killed in value field.
7. Now following table is showing total killed data.



## 2. Show the following summary from total injury number: pedestrian injured, cyclist and motorcyclist injured?

**To check summery from total injury number follow these steps:**

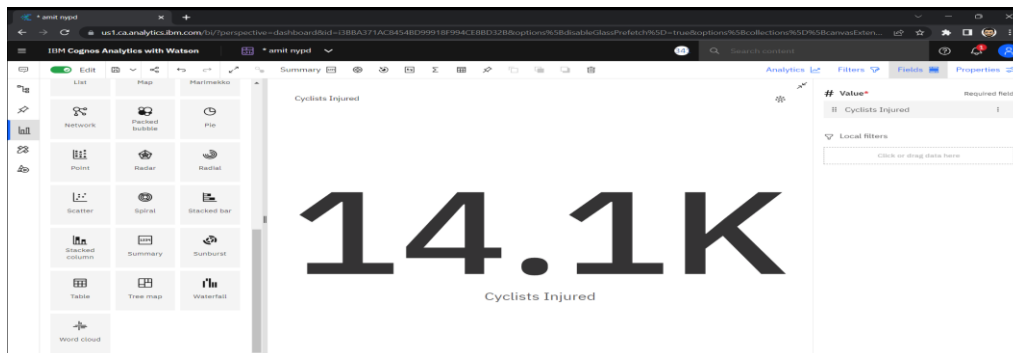
**To check total pedestrian injured.**

1. Navigate to visualization tool and select summery tool and drag drop in canvas.
2. After that select pedestrian injured in value field.
3. Finally total pedestrian injured data is showing in following set.



**To check total cyclist injured.**

1. Navigate to visualization tool and select summery tool and drag drop in canvas.
2. After that select total cyclist injured in value field.
3. Finally total cyclist injured data is showing in following set.

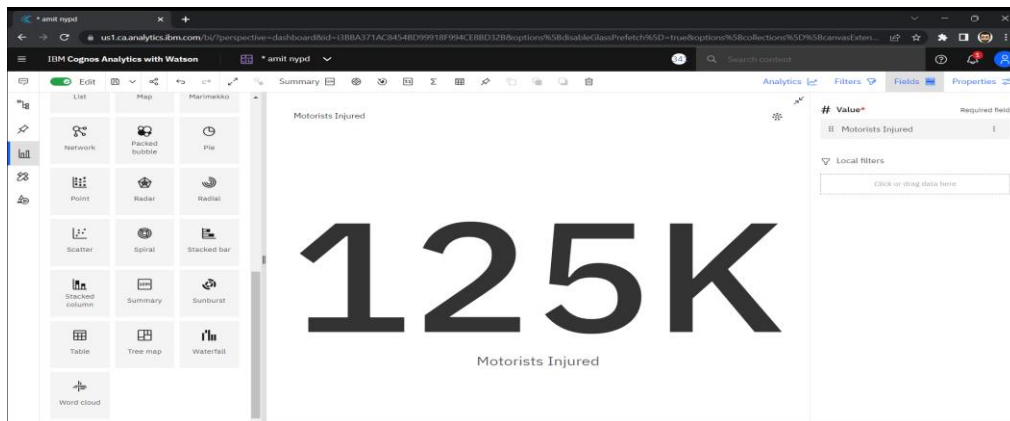


**To check total motorist injured.**

Drag drop summery tool from visualization tool in canvas

Select total motorist injured in value field.

Finally total cyclist injured data is showing in following set.



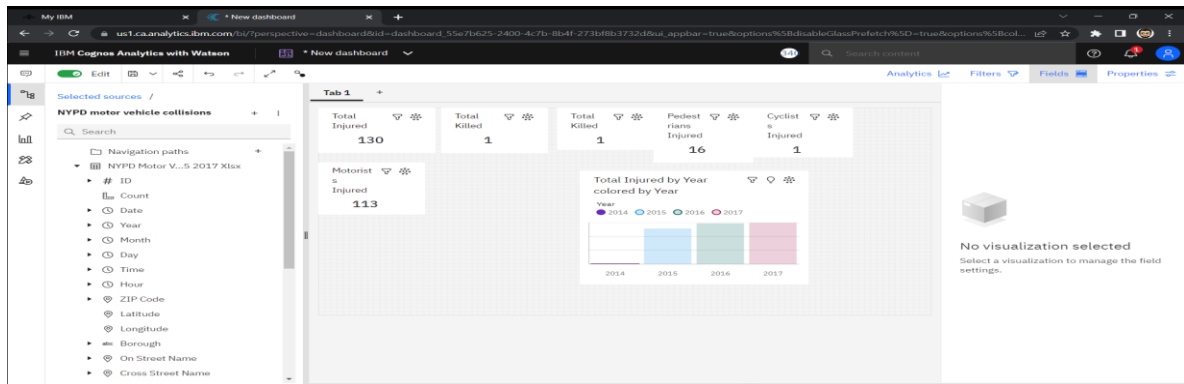
### 3. Show the injuries on the basis of the year?

To check injuries on the basis of the year follow these steps:

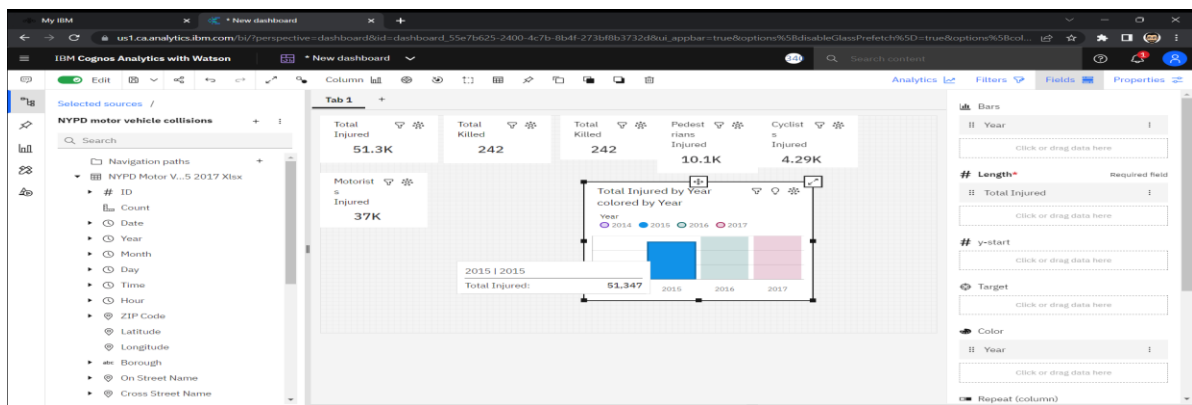
1. Navigate to visualization tool and select column tool.

2. Drag drop column tool in canvas.
3. Select year in bars field.
4. Select total injured in length field.
5. Select year in color field.
6. Now the injuries will show according the year.

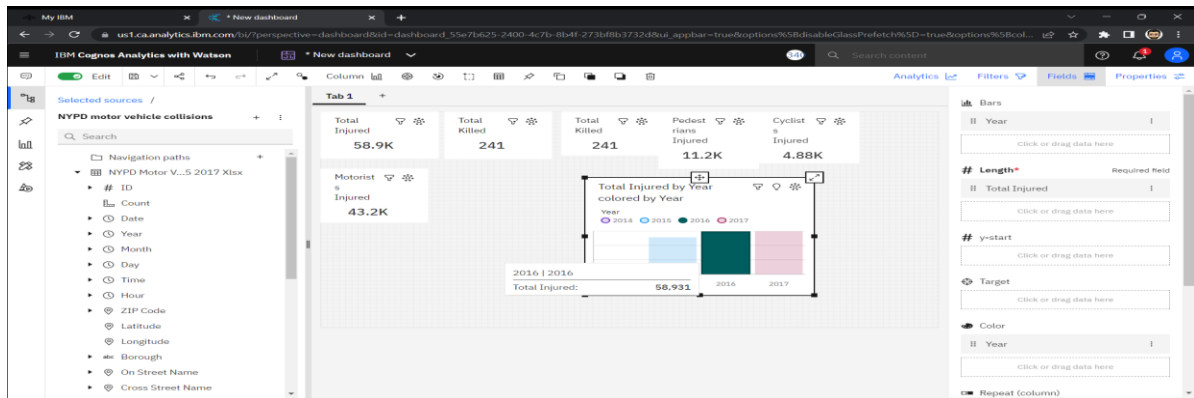
## Data of total injuries in 2014.



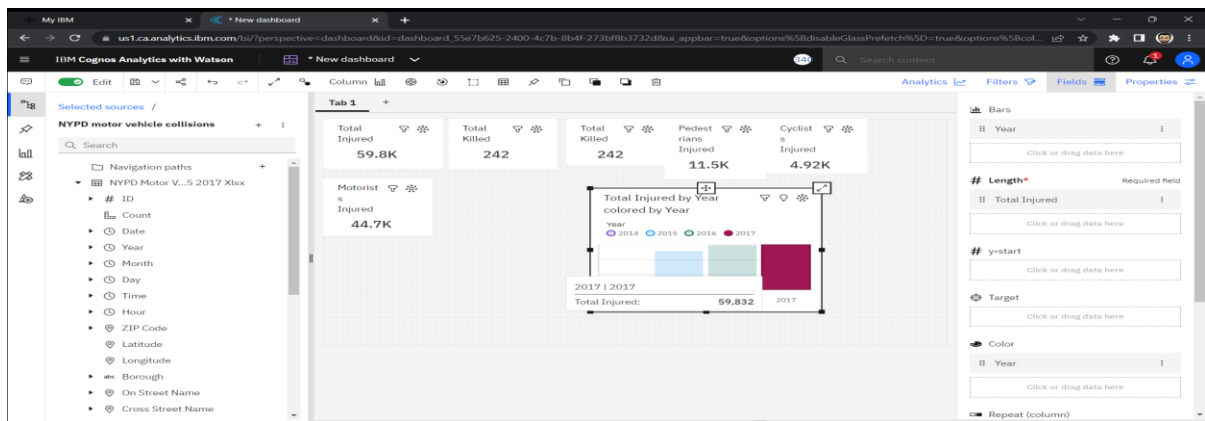
## Data of total injuries in 2015.



## Data of total injuries in 2016.



**Data of total injuries in 2017.**

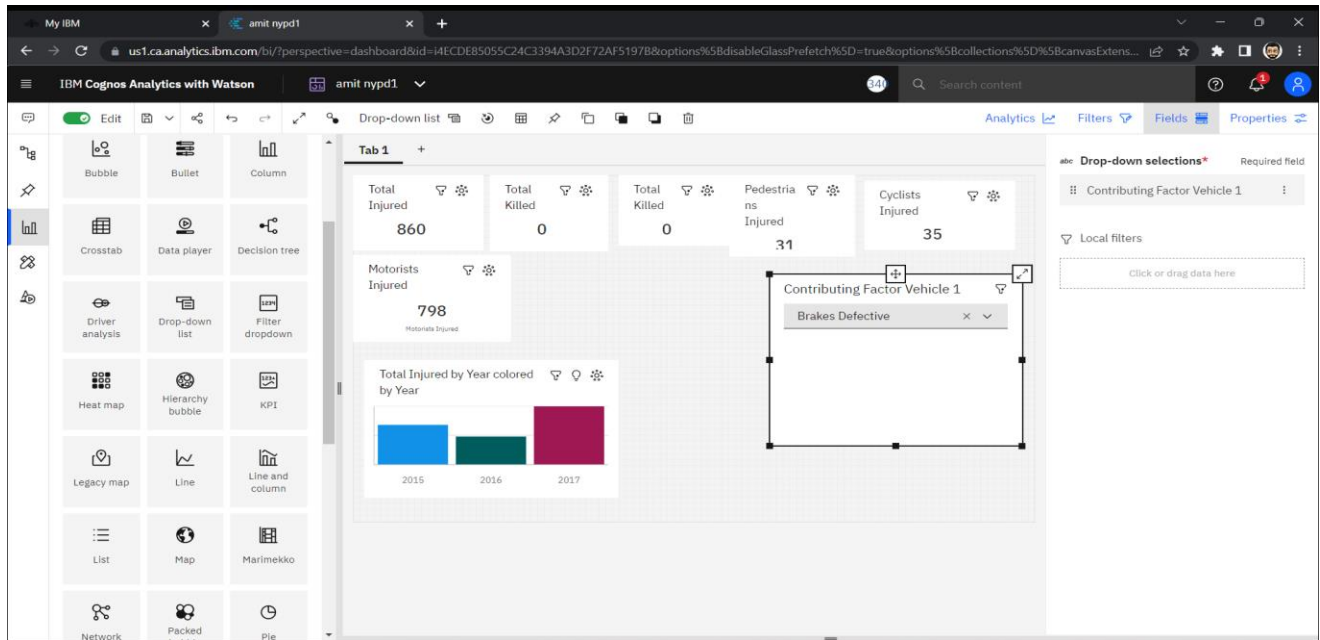


**In above examples total injuries are shown on the basis of year on different conditions.**

4. Show the number of injuries on the basis of.
  - “Vechile1 contribution factor”
  - “Vechile2 contribution factor”
  - “Vechile3 contribution factor”.

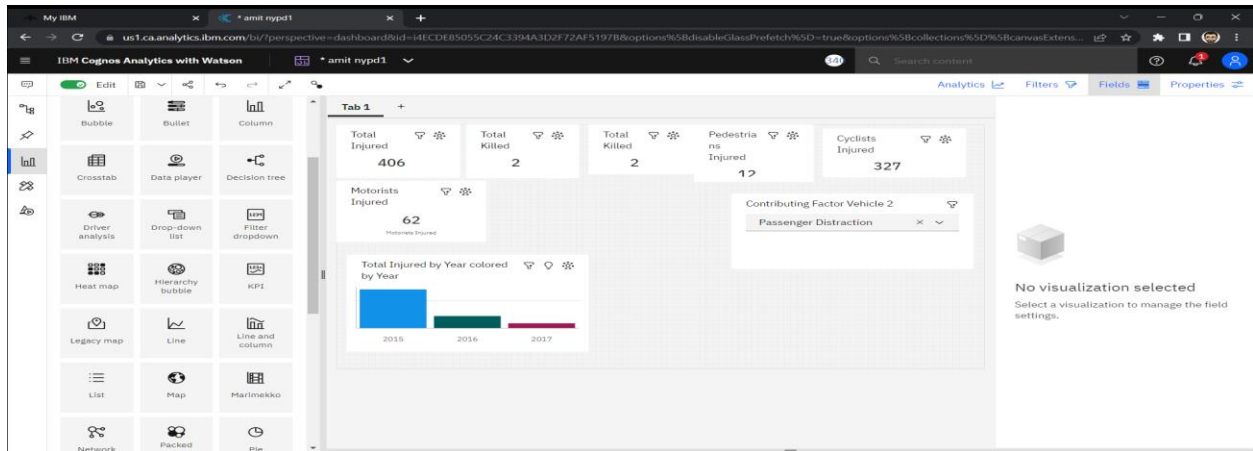
**To show the injuries on the basis of vehicle1 contribution factor follow these steps:**

1. Navigate to visualization tool.
2. Select Drop-Down list and drag drop in canvas.
3. Select contributing factor vehicle1.
4. Open menu and select factors according to need.
5. Now the data will show on basis of vehicle1 contribution factor [Brake Defects].



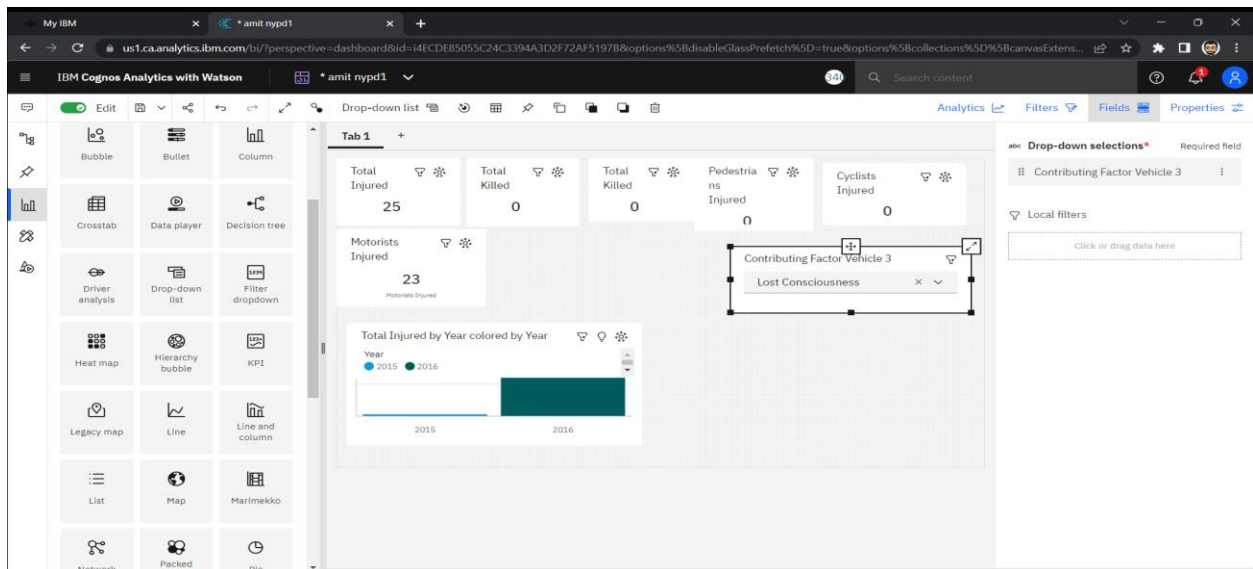
**To show the injuries on the basis of vehicle2 contribution factor follow these steps:**

1. Navigate to visualization tool.
2. Select Drop-Down list and drag drop in canvas.
3. Select contributing factor vehicle2.
4. Open menu and select factors according to need.
5. Now the data will show on basis of vehicle2 contribution factor [Passenger Distraction].



**To show the injuries on the basis of vehicle3 contribution factor follow these steps:**

1. Navigate to visualization tool.
2. Select Drop-Down list and drag drop in canvas.
3. Select contributing factor vehicle3 from field.
4. Open menu and select factors according to need.
5. Now the data will show on basis of vehicle3 contribution factor [Lost Consciousness].





**5. New York animal welfare department wants to check the rate of accident due to 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?**

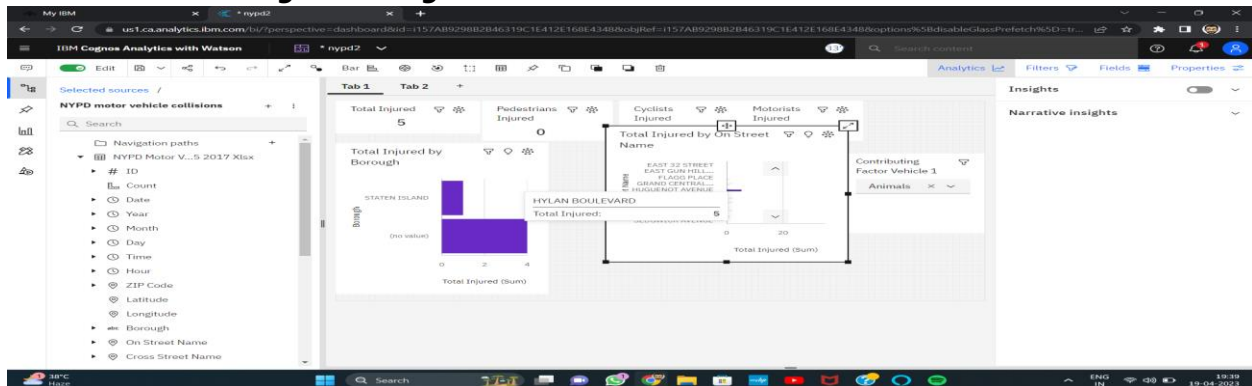
**To show the total injuries on the basis of animal accident factor.**

1. Navigate to visualization tool.
2. Select drop-down list and drag in canvas.
3. Select contributing factor vehicle1 from field.
4. Open menu and select animal accident factor.

**To show the total injuries on the basis of borough and street.**

1. Navigate to visualization tool.
2. Select bar tool from tool box.
3. Select borough and on street name in bar from field.
4. Also select total injured in the length field.

**Now data is showing according to the above conditions.**



**Conclusion:** New York police department have to focus on 'STATEN ICELAND' city and 'HYLAN BOULEVARD' street because these places are showing number of injuries due to Animal Accident factor.

## **6. Give the number of Pedestrian, cyclist, motorist injured for on the basis on any particular city, month and year.**

### **To check injuries on basis of pedestrian, cyclist, motorist.**

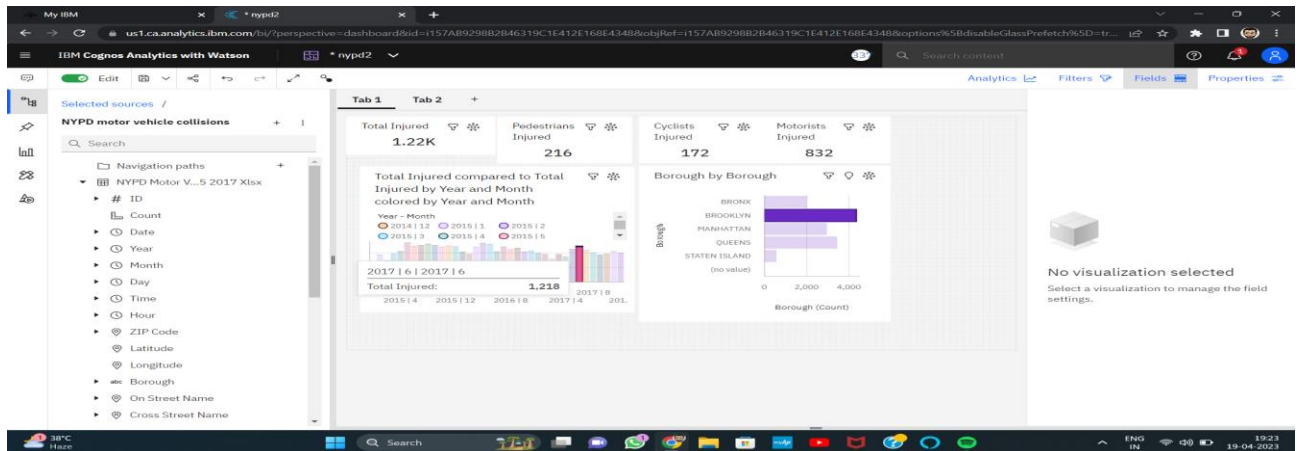
1. Navigate to visualization tool.
2. Select summery tool from visualization tool.
3. Select pedestrian injured from field section.
4. Copy summery tool and select cyclist injured from field section.
5. Again copy summery tool and select motorist injured from field section.

### **To show the injuries on the basis of month and year.**

1. Navigate to visualization tool.
2. Select column tool from visualization tool.
3. And select year and month in x & y field.
4. Select total injured in length field.
5. Also select total injured in color field to differentiate between color month and year.

### **To show the injuries on the basis of particular city.**

1. Navigate to visualization tool.
2. Select bar from tool section.
3. Select borough in in field section.
4. Select borough in color section.

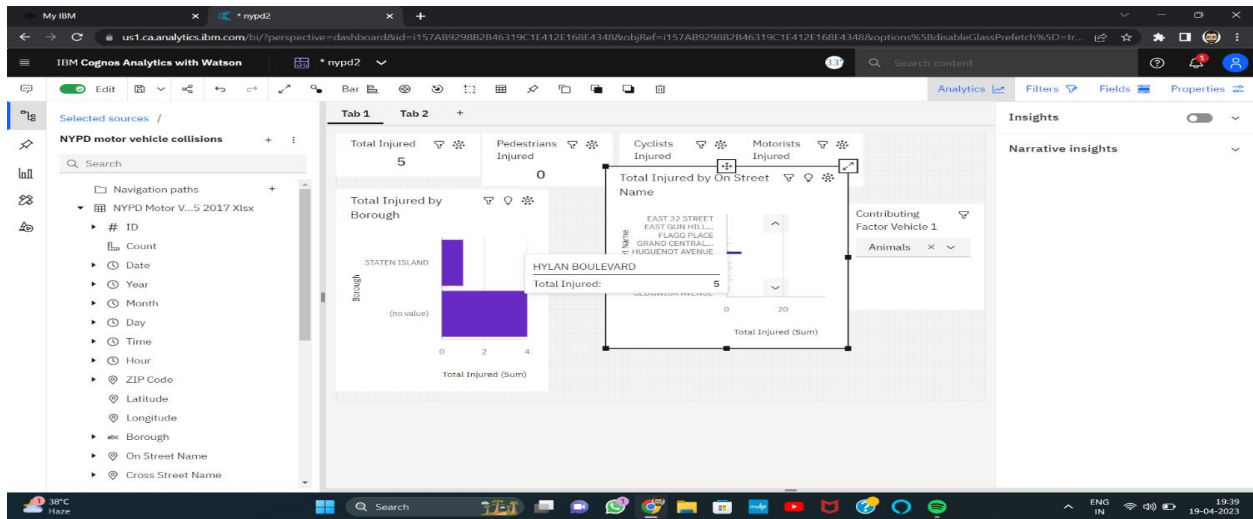


The above steps is showing the data according to the query.

*7. New York animal welfare department wants to check the rate of accident due to Animal accident factor. Please tell us which particular city and on which particular street they have to focus.*

1. To show the total injuries on the basis of animal accident factor.
2. Navigate to visualization tool.
3. Select drop-down list and drag in canvas.
4. Select contributing factor vehicle1 from field.
5. Open menu and select animal accident factor.
6. To show the total injuries on the basis of borough and street.
7. Navigate to visualization tool.
8. Select bar tool from tool box.
9. Select borough and on street name in bar from field.
10. Also select total injured in the length field

Now the data will show according to the above conditions.



**Conclusion:** New York police department have to focus on 'STATEN ICELAND' city and 'HYLAN BOULEVARD' street because these places are showing number of injuries due to Animal Accident factor.