

## **APPENDIX 1**

# **CASE STUDY 1**

## **NYPD Data Set**

COLLEGE PROJECT REPORT

*by*

**Himanshu Singh**

Department of  
COMPUTER APPLICATION  
UNITED UNIVERSITY  
PRAYAGRAJ  
April 2023

**APPENDIX 2**

**BONAFIDE CERTIFICATE**

This is to certify that this project report entitled "**CASE STUDY 1 on NYPD DATASET**" submitted to **UNITED UNIVERSITY PRAYAGRAJ**, is a bonafide record of work done by "**Himanshu Singh**" under my supervision from "**10<sup>th</sup> APRIL**" to "**20<sup>th</sup> APRIL**".

Place:

Date:

**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

### **APPENDIX 3**

been duly acknowledged. I/We agree that if any part of the report is found to be plagiarized, I/we shall take full responsibility for it.

Himanshu Singh

Place:

Date:

### **TABLE OF CONTENTS**

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
<b>1.</b>	<b>Introduction .....</b>	<b>1</b>
	<b>1.1 Introduction to Dataset .....</b>	<b>1 - 4</b>
<b>2.</b>	<b>Tools we will use .....</b>	<b>5</b>
	<b>2.1 Introduction to tools .....</b>	<b>5 - 8</b>

## **APPENDIX 4**

<b>2.2 Steps to create Dashboard .....</b>	<b>9 - 10</b>
<b>2.3 Steps to import Dataset .....</b>	<b>11 - 12</b>
<b>3. Queries to Perform .....</b>	<b>13</b>
<b>3.1 Query 1 .....</b>	<b>15</b>
<b>3.2 Query 2 .....</b>	<b>16</b>
<b>3.3 Query 3 .....</b>	<b>17 - 18</b>
<b>3.4 Query 4 .....</b>	<b>19 - 21</b>
<b>3.5 Query 5 .....</b>	<b>22 - 24</b>
<b>3.6 Query 6 .....</b>	<b>25 - 27</b>
<b>3.7 Query 7 .....</b>	<b>28 - 30</b>



**3.8** Query 8 ..... 31 - 32

**3.9** Query 9 ..... 33 - 35

## Explain the "Data" you are using.

The "NYPD motor vehicle collision" data module in Cognos is a pre-built data set that contains information on motor vehicle collisions that were reported to the New York Police Department (NYPD). This data is publicly available and is updated on a daily basis.

The data module consists of several data tables that are joined together to provide a comprehensive view of each collision. Some of the key tables included in this module are:

- 1. Collision Data Table:** This table contains general information about each collision, including the date and time of the collision, and the location. The key fields in this table include:

**CRASH DATE:** The date on which the collision occurred.

**CRASH TIME:** The time at which the collision occurred.

**ZIP CODE:** The zip code in which the collision occurred.

**LATITUDE:** The latitude of the collision location.

**LONGITUDE:** The longitude of the collision location.

**BOROUGH:** The borough in which the collision occurred (e.g. Manhattan, Brooklyn, etc.).

**STREET NAME:** The names of the street in which collision took place.

Date	Year	Month	Day	Time	Hour	ZIP Code	Latitude	Longitude	Borough	On Street Name
2017-12-30	2017	12	30	05:00:00	5	10471	40.91013	-73.90322	BRONX	RIVERDALE AVENUE
2017-12-30	2017	12	30	05:00:00	5	10032	40.84173	-73.94434	MANHATTAN	Null
2017-12-30	2017	12	30	05:00:00	5	11432	40.710133	-73.78333	QUEENS	Null
2017-12-30	2017	12	30	05:00:00	5	Null	40.602673	-73.89933	Null	BELT PARKWAY
2017-12-30	2017	12	30	05:00:00	5	Null	40.884014	-73.8791	Null	EAST 212 STREET
2017-12-30	2017	12	30	05:00:00	5	Null	40.8436	-73.94511	Null	HENRY HUDSON PARKWAY
2017-12-30	2017	12	30	05:01:00	5	10456	40.82839	-73.916916	BRONX	MORRIS AVENUE
2017-12-30	2017	12	30	05:04:00	5	11231	40.6742	-73.99984	BROOKLYN	HAMILTON AVENUE

- 2. Vehicle Data Table:** This table contains information about the vehicles involved in the collision, including the type of vehicle (e.g. car, truck, motorcycle, cycle). The key fields in this table include:

**VEHICLE TYPE 1:** The type of the first vehicle involved in the collision.

**VEHICLE TYPE 2:** The type of the second vehicle involved in the collision.

**VEHICLE TYPE 3:** The type of the third vehicle involved in the collision.

Vehicle 1 Type	Vehicle 2 Type	Vehicle 3 Type
SPORT UTILITY / STATION WAGON	Null	Null
PASSENGER VEHICLE	Null	Null
SPORT UTILITY / STATION WAGON	SPORT UTILITY / STATION WAGON	Null
PASSENGER VEHICLE	SPORT UTILITY / STATION WAGON	Null
PASSENGER VEHICLE	Null	Null
PASSENGER VEHICLE	Null	Null

3. **Contributing Factor Data Table:** This table contains information about the contributing factors to the collision, such as driver inattention, speeding, or alcohol involvement. The key fields in this table include:

**CONTRIBUTING FACTOR VEHICLE 1:** The primary contributing factor related to the first vehicle involved in the collision.

**CONTRIBUTING FACTOR VEHICLE 2:** The primary contributing factor related to the second vehicle involved in the collision.

**CONTRIBUTING FACTOR VEHICLE 3:** The primary contributing factor related to the third vehicle involved in the collision.

Contributing Factor Vehicle 1	Contributing Factor Vehicle 2	Contributing Factor Vehicle 3
Unspecified	Null	Null
Unspecified	Null	Null
Reaction to Other Uninvolved Vehicle	Unspecified	Null
Driver Inexperience	Unspecified	Null
Unspecified	Null	Null
Unspecified	Null	Null
Outside Car Distraction	Passing or Lane Usage Improper	Null
Unspecified	Unspecified	Null

4. **Person Data Table:** This table contains information about the individuals involved in the collision, whether they were cyclist, motorist, or pedestrian. The key fields in this table include:

**TOTAL INJURED:** This data shows the total number of injuries.

**PEDESTRIAN INJURED:** This data shows the number of pedestrians that were involved in the collision.

**CYCLISTS INJURED:** This data shows the number of cyclists that were involved in the collision.

**MOTORISTS INJURED:** This data shows the number of motorists that were involved in the collision.

Total Injured	Total Killed	Pedestrians Injured	Pedestrians Killed	Cyclists Injured	Cyclists Killed	Motorists Injured
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

By analyzing the data in these key tables, analysts can gain insights into the causes and circumstances of motor vehicle collisions in New York City.

### **Explain the Tool which you are going to use.**

**COGNOS:** IBM Cognos is a business intelligence and performance management software suite that helps organizations analyze 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.

**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 userfriendly 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.

A data module is designed to integrate with a wide range of data sources, including relational databases, flat files, and big data platforms. It provides a flexible and scalable approach to data integration, allowing users to connect to multiple data sources and combine them into a single view. This eliminates the need for users to have deep technical knowledge of the underlying data sources and makes it easier to work with data from different systems.

**To create a data module in IBM Cognos, follow these steps:**

- Open the IBM Cognos Analytics interface and navigate to the Data module option on the home screen.
  - Select the data source that you want to use as the basis for your data module. This can be a relational database, flat file, or big data platform.
  - Define the relationships between your data sources by dragging and dropping the relevant fields onto the canvas.
- 
- Create any necessary calculations and aggregations by selecting the relevant fields and applying the appropriate mathematical or statistical functions.
  - Define any data governance rules and policies that you want to apply to your data module.

Save your data module and use it as the basis for creating reports and dashboards.

Overall, a data module provides a user-friendly and flexible approach to data integration, allowing users to work with data from multiple sources without requiring specialized technical skills. By simplifying the process of data integration and governance, data modules can help organizations to make more informed decisions and gain insights into their data.

**DASHBOARD:** In Cognos, a dashboard is a visual representation of data that provides a quick and easy way to monitor and analyze key performance indicators (KPIs) and metrics. A dashboard typically consists of a set of graphical charts, tables, and other visualizations that help users to quickly understand the data and make informed decisions.

Cognos dashboards can be customized to fit the needs of specific users or groups. They can also be interactive, allowing users to drill down into the data to get more detailed information or to filter the data to focus on specific areas of interest.

Dashboards in Cognos can be created using a drag-and-drop interface, making them easy to build and customize. They can also be shared with others in the organization, making it easy to collaborate and make informed decisions based on the data.

### **Tools In Dashboard:**

**Summary:** A summary is a type of visualization that displays a condensed version of data in a dashboard. It typically shows aggregated data in a simple format, such as a total or an average.

**Dropdown List:** A dropdown list is a user interface element that allows users to select an item from a list of options. It is often used in Cognos Dashboard to filter data based on user selection.

**Heat Map:** A heat map is a type of visualization that displays data in a matrix format where each cell is shaded according to its value. It is often used to represent the density of data points in a particular area.

**Pie Chart:** A pie chart is a circular chart that is divided into slices to represent numerical proportions. It is often used in Cognos Dashboard to show the relative proportions of different data points.

**Crosstab:** A crosstab is a table that displays data in a grid format where each row represents a category and each column represents a different variable. It is often used in Cognos Dashboard to compare data across different categories and variables.

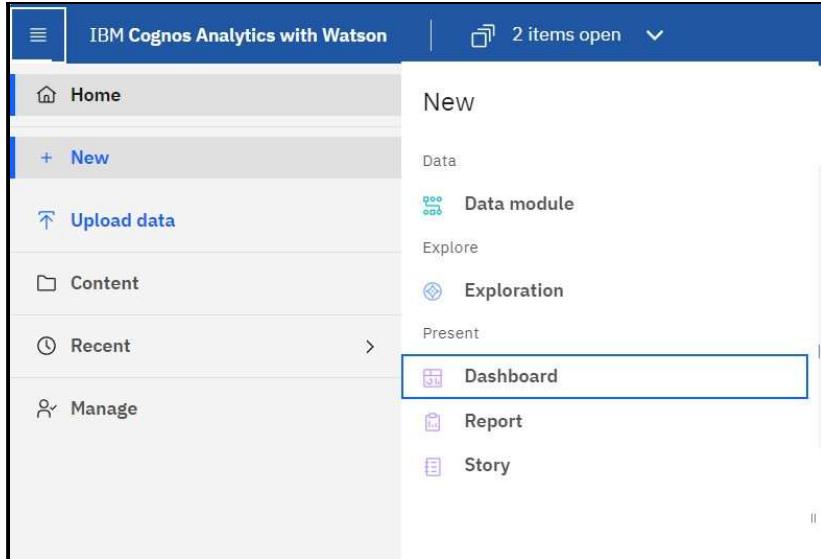
**Map:** A map is a type of visualization that displays data geographically. It is often used in Cognos Dashboard to show the distribution of data across different regions.

**Network:** A network is a type of visualization that displays data as nodes connected by edges. It is often used in Cognos Dashboard to show relationships between different data points.

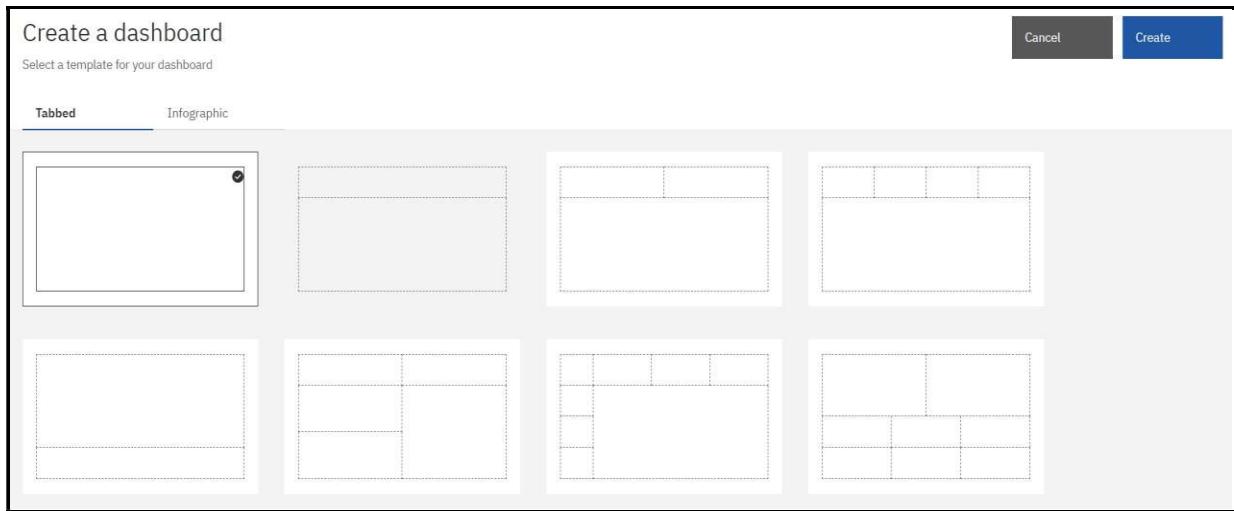
**Column:** A column is a type of visualization that displays data as vertical bars. It is often used in Cognos Dashboard to compare data across different categories.

## **Steps to create Dashboard.**

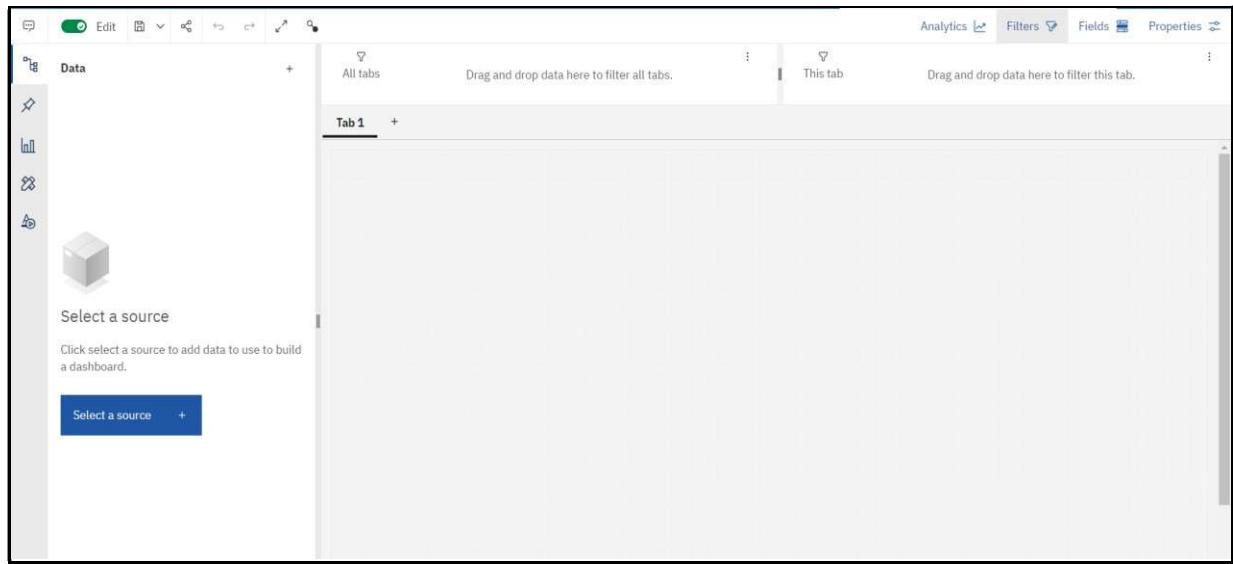
**STEP 1:** On the home page of IBM Cognos there is a Hamburger menu on the top left of the screen, by clicking on that and then new the user can see the option to create a new Dashboard



**STEP 2:** By clicking on the Dashboard icon a new window will show the user some predefined templates for the Dashboard

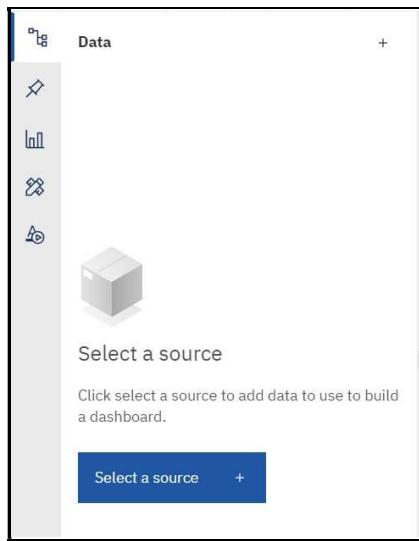


**STEP 3:** After choosing the template the user can click on the “CREATE” button on the top right of the screen, this will then create the dashboard and open the empty dashboard page



## **Steps to import the DATA.**

**STEP 1:** To import the dataset that the user will use on the Dashboard, click on the “Select a source” button on the left side of the screen



**STEP 2:** After clicking on the button a window will pop showing the location that user can import data from (Make sure to import data modules as Cognos tools only support data modules.)

A screenshot of a modal window titled 'Select a source'. It has tabs for 'My content' and 'Team content', with 'Team content' being active. The main area displays a table with columns: Name, Type, and Last Accessed. The data includes several folders: 'Calendars', 'Get started', 'Honda', 'Samples', 'Templates', and 'The Weather Company, an IBM Business'. At the bottom are 'Cancel' and 'Add' buttons.

We are using “NYPD motor vehicle collision” data for which we have to follow the following path: **Team content / Samples / By feature / Core / Data / NYPD motor vehicle collisions**

Name	Type	Last Accessed
Source files	Folder	5/16/2022, 12:05 PM
Boston 311 calls	Data module	2/3/2022, 12:03 PM
California website visits	Data module	2/3/2022, 12:02 PM
Great outdoors data module	Data module	2/3/2022, 12:03 PM
NYPD motor vehicle collisions	Data module	2/3/2022, 12:02 PM
Sample data module	Data module	5/9/2022, 11:33 AM
Storm events 2015	Data module	2/3/2022, 12:02 PM
Weather analytics	Data module	2/3/2022, 12:02 PM

**STEP 3:** Click on “Add” button to import the data into Dashboard.

**Queries you have to resolve using Dashboard Components:**

- You have to show the total number of "**total injury/total killed**"
- Show me the following Summary from total injury number: pedestrian injured, cyclist and motorists injured.
- Show me the injuries on the basis of year.
- Show me the number of injuries on the basis on "vehicle 1 contribution factor", "vehicle 2 contribution factor" and "vehicle 3 contribution factors"
- New York animal welfare department wants to subdue 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.
- Give me the number of pedestrian / cyclist / motorists injured for on the basis on any particular city, month and year.
- Give me 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.
- 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 Committee auth. wants to create some new street lane for more than 6 tiers vehicles. So, suggest me any top 7 locations in New York in which we can build a new lane for heavy vehicles and why?

**NOTE:** In all the problems below we will use this dashboard as a default dashboard, and as the queries require different tools to solve the problems, we will create different dashboards for different queries using this dashboard.

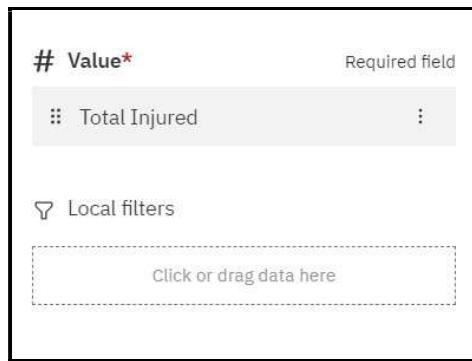


**QUERY 1:** You have to show the total number of "total injury/total killed"

**STEP 1:** Click on the “Visualisation” icon on the left side of the screen. It will show you all the visualisation tools that you can use for your dashboard.

**STEP 2:** Drag and drop the “Summary” tool on the dashboard page.

**STEP 3:** On the right side of the screen there will appear an option to add data to the Summary element.



**STEP 4:** Click on the required field named “# Value” and select the dataset named “Total Injured”.

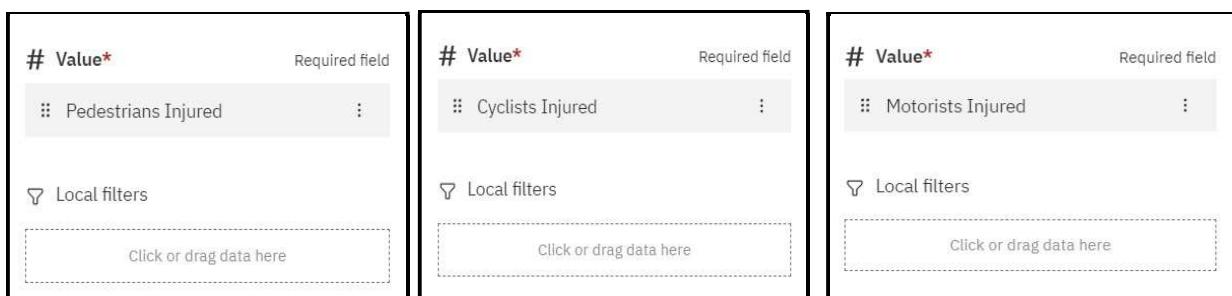
**NOTE:** The Dashboard shows the total number of injuries.



**QUERY 2:** Show me the summary for: Total injured, Pedestrian Injured, Cyclists injured, Motorists Injured.

**STEP 1:** From the visualisation tab drag and drop three more Summary tool on the dashboard.

**STEP 2:** Then like previous Query we will add three datasets to the Summary tools that we drag dropped in Dashboard.



**STEP 3:** The data sets for these summary tools will be Pedestrian Injured, Cyclists Injured, Motorists Injured.



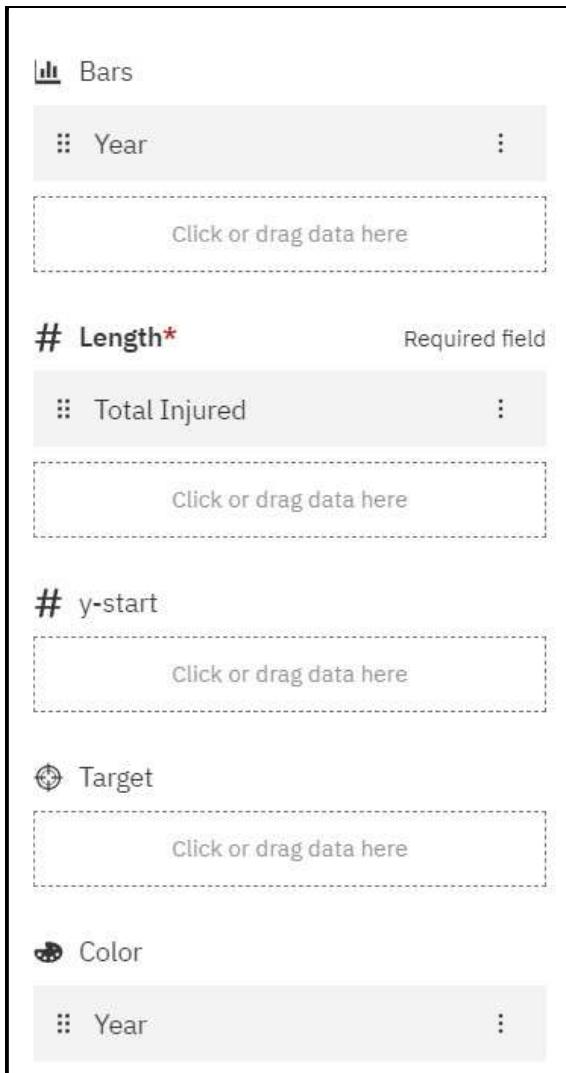
**NOTE:** The Dashboard shows the total number of injuries for Pedestrians, cyclists, motorists.

**QUERY 3:** Show the injuries on the basis of a Year.

**STEP 1:** From the visualisation tab drag and drop a Column chart on the dashboard.

**STEP 2:** On the right side on the screen, we have fields named ‘Bars’, ‘Length’, ‘Color’.

**STEP 3:** In Bars field we will add ‘Year’ data, in length field we will add ‘Total Injured’ and in Color field we will add ‘Year’ again so that each year is distinct from one another by different colors.



**STEP 4:** After adding these data in their fields we can see a Column chart in the dashboard.

**STEP 5:** Now to show a particular data by a particular year we can select that year's bar in the chart and all the data in the dashboard will show data related to that year.



**NOTE:** The dashboard shows the data according to selected year which is 2015.

**QUERY 4:** Show the number of injuries in the basis of:

**1. Vehicle 1 contribution factor.**

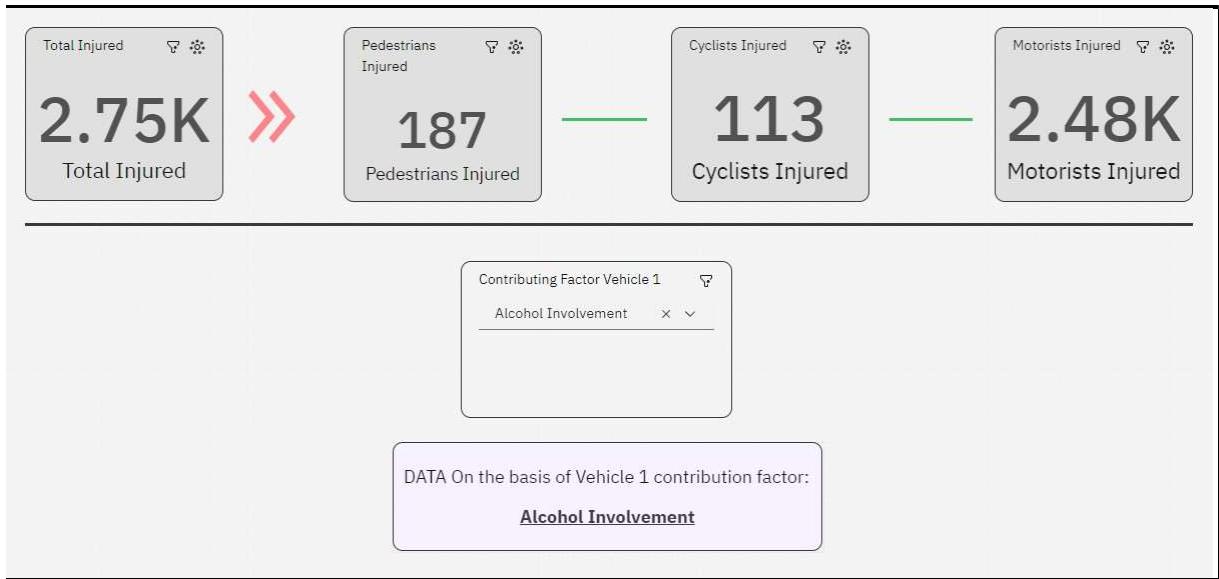
**STEP 1:** From the visualisation tab drag and drop a “Drop Down List” tool.

**STEP 2:** On the right side on the screen There is a field named ‘Drop-down selection’.

**STEP 3:** Add the data set named ‘Contributing Factor Vehicle 1’ in that field.

The interface shows a dropdown menu labeled "Contributing Factor Vehicle 1". Below it is a section for "Local filters" with a placeholder "Click or drag data here".

**STEP 4:** In the dashboard click on the drop-down list and select any contribution factor.



**NOTE:** Dashboard shows data on basis of Alcohol Involvement contributing factor for Vehicle type 1.

## 2. Vehicle 2 contribution factor.

**STEP 1:** From the visualisation tab drag and drop a “Drop Down List” tool.

**STEP 2:** On the right side on the screen There is a field named ‘Drop-down selection’.

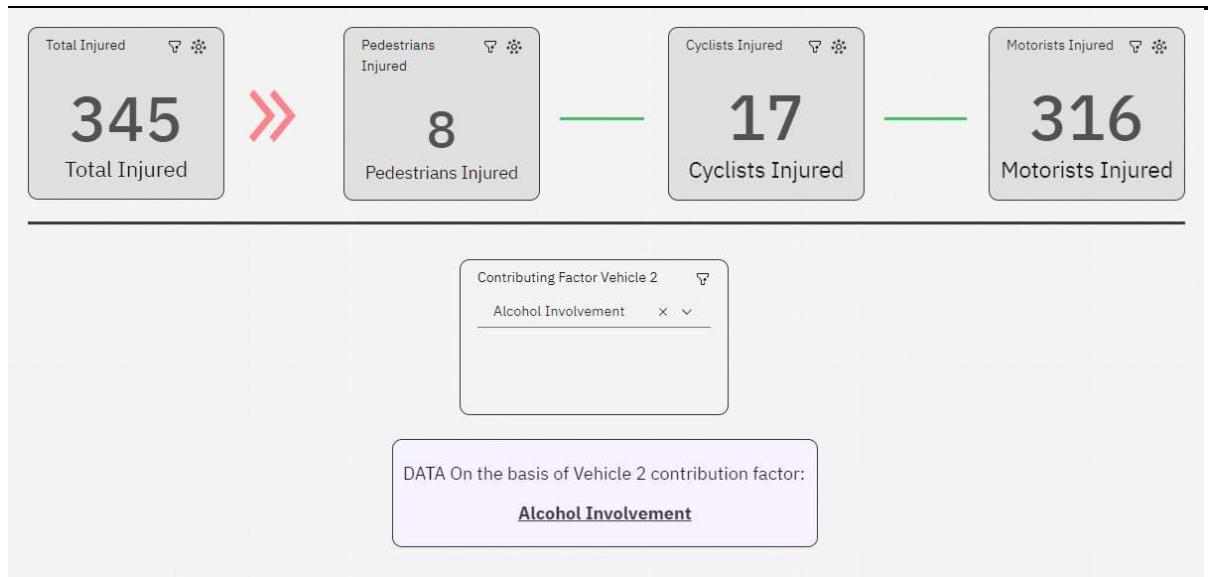
**STEP 3:** Add the data set named ‘Contributing Factor Vehicle 2’ in that field.

The figure shows a configuration interface for a "Drop-down selections" tool. The title bar says "abc Drop-down selections\* Required field".

The main area contains a section titled "Contribution Factor Vehicle 2".

Below this, there is a "Local filters" section with a "Click or drag data here" placeholder.

**STEP 4:** In the dashboard click on the drop-down list and select any contribution factor.



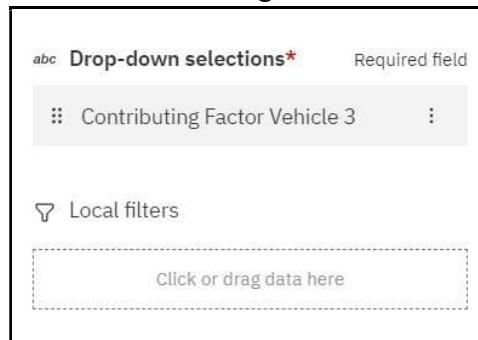
**NOTE:** Dashboard shows data on basis of Alcohol Involvement contributing factor for Vehicle type 2.

### 3. Vehicle 3 contribution factor.

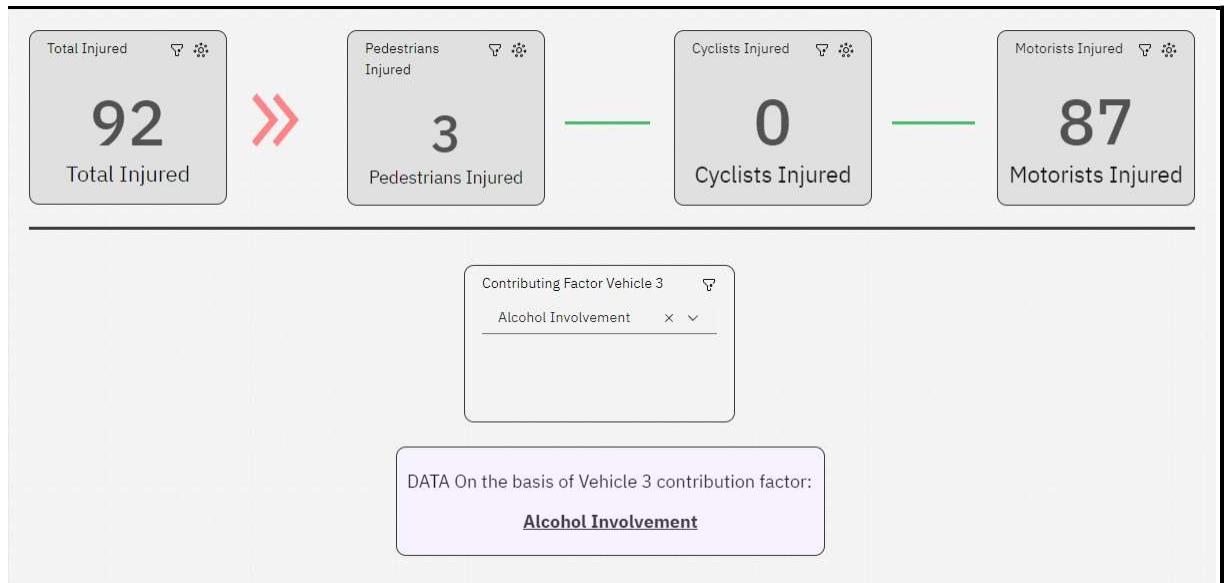
**STEP 1:** From the visualisation tab drag and drop a “Drop Down List” tool.

**STEP 2:** On the right side on the screen There is a field named ‘Drop-down selection’.

**STEP 3:** Add the data set named ‘Contributing Factor Vehicle 3’ in that field.



**STEP 4:** In the dashboard click on the drop-down list and select any contribution factor.



**NOTE:** Dashboard shows data on basis of Alcohol Involvement contributing factor for Vehicle type 3.

**QUERY 5:** New York animal welfare department wants to subdue 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.

**NOTE:** To solve this problem with the help of dashboard we will use a tool called Heat map.

**STEP 1:** First let's drag and drop a Drop-down list on the **Default dashboard** and in the field add contributing factor vehicle 1 dataset.

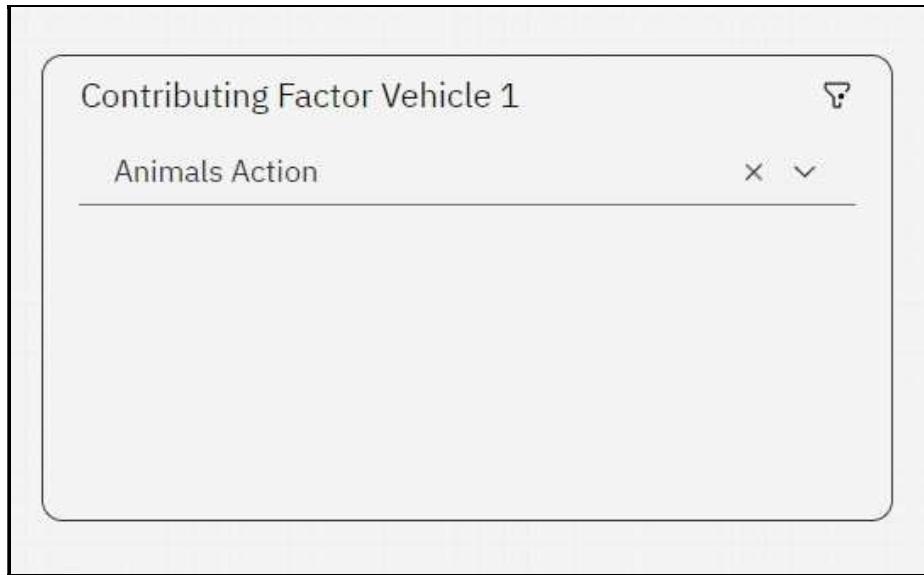
abc **Drop-down selections\*** Required field

⋮ Contributing Factor Vehicle 1 ⋮

▽ Local filters

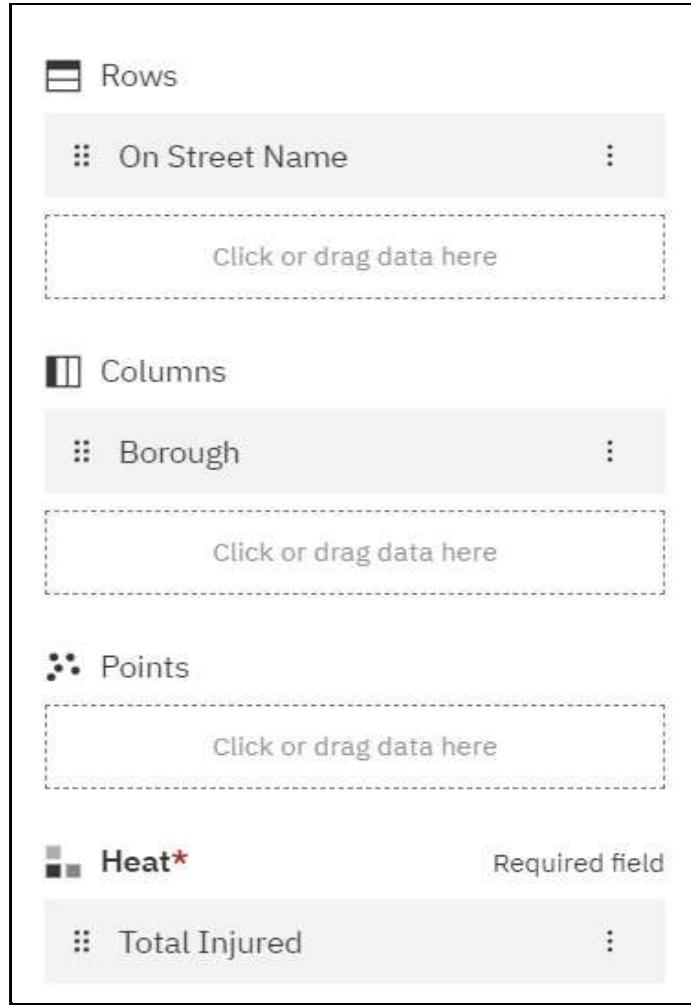
Click or drag data here

**STEP 2:** In the drop-down list we select the Contribution factor as Animal Action. This will show us the data for injuries that are caused factors that include animals.

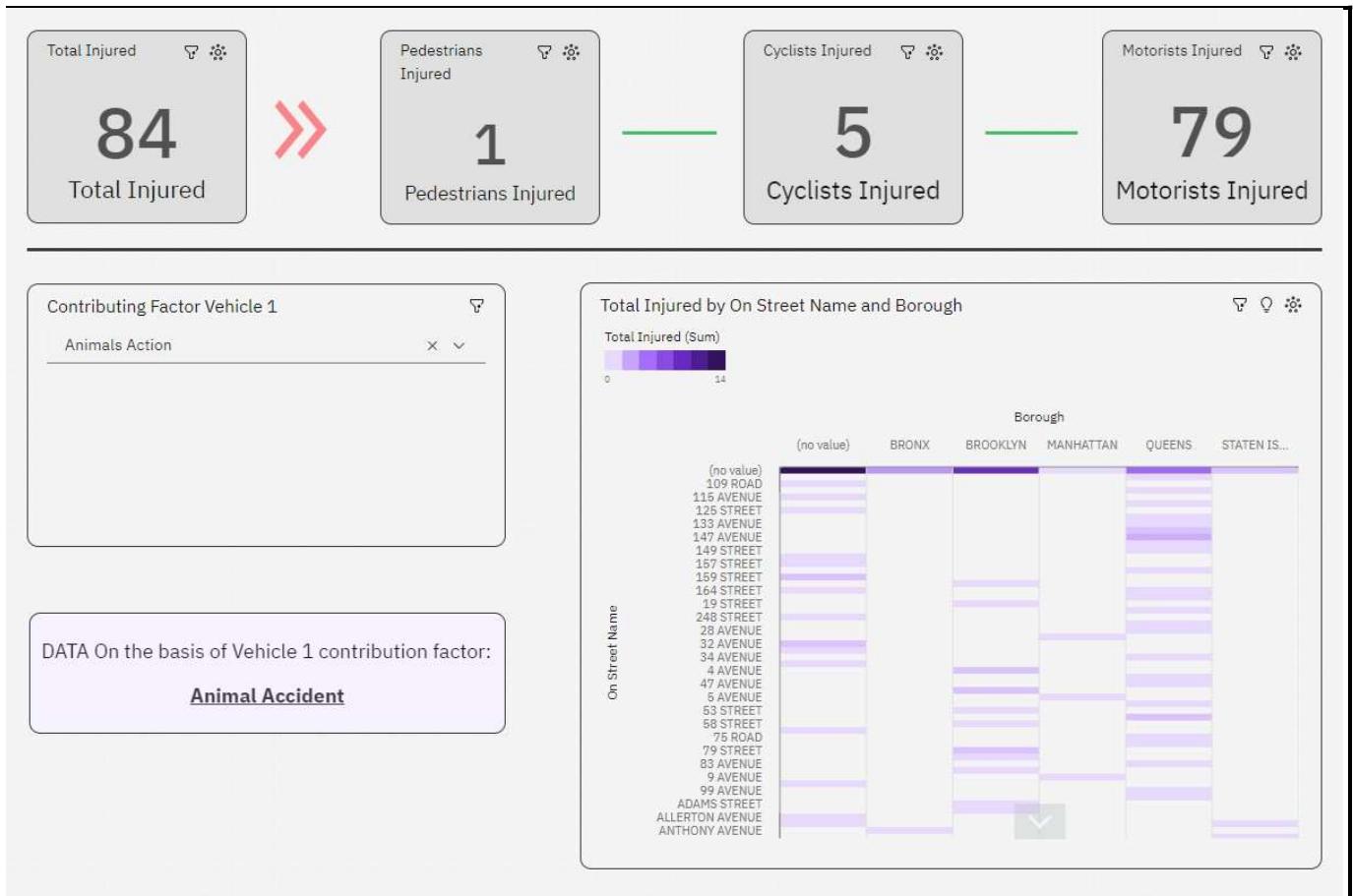


**STEP 3:** Now we will drag and drop a 'Heat map' tool from the visualisation tab.

**STEP 4:** In the fields of this heat map Rows, Columns, Heat. We will add Data sets. In Rows we will add 'On Street Name', in Columns we will add 'Borough' and in Heat we will add 'Total Injured'.



**STEP 5:** And now the Dashboard will show the specific street and city name in which the most injuries happened due to animal incidents.



**NOTE:** So, we can show the New York animal welfare department that the most accidents that took place due to presence of animals were on a Street named (no value) in the city Brooklyn.

**QUERY 6:** Give me the number of pedestrian / cyclist / motorists injured for on the basis on any particular city, month and year.

**NOTE:** First let's create the dashboard.

**STEP 1:** In the **Default dashboard** drag and drop a Crosstab which we will use for showing data according to Year and month.

**STEP 2:** In the fields of crosstab, we have Columns, Rows, #Values. We will add datasets. In the Columns field we will add Year and Month data sets, in the Rows field we will add Borough data set and in the #Values we will add Total Injured.

The screenshot shows a data visualization interface with three main sections: Columns, Rows, and # Values. The Columns section contains 'Year' and 'Month'. The Rows section contains 'Borough'. The # Values section contains 'Total Injured'. Each section has a 'Click or drag data here' placeholder area below it. A 'Required field' label is positioned next to the '# Values' section.

**STEP 3:** Now let's add a Pie chart to show the data city wise. Drag and drop a Pie Chart from visualisation tab.

**STEP 4:** In the fields of Pie Chart, in field Segments we will add Borough data and, in the field, #Size we will add Total Injured.

**Segments\*** Required field

⋮ Borough ⋮

Click or drag data here

**# Size\*** Required field

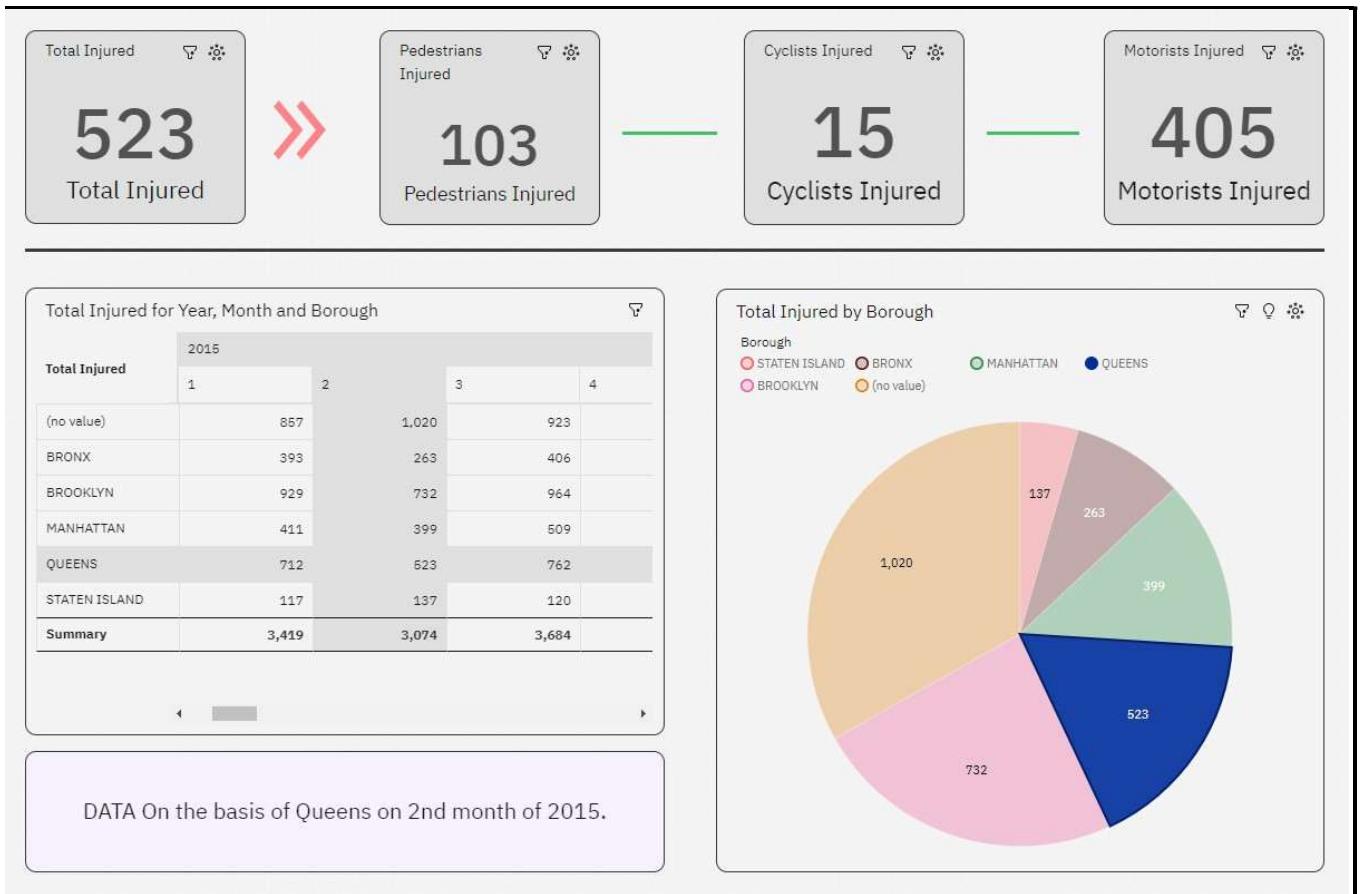
⋮ Total Injured ⋮

**STEP 5:** Now that we created the dashboard, we can select specific data to show.

(Next Page)

**STEP 6:** Click on the Month column on the Crosstab to show data according to that month and year. (Click on 2<sup>nd</sup> month of 2015).

**STEP 7:** Click on the part of the Pie chart that specifies the city we want to show data for (Click on blue part for city QUEENS).



**NOTE:** Dashboard shows the Injuries in QUEENS on 2<sup>nd</sup> month of 2015.

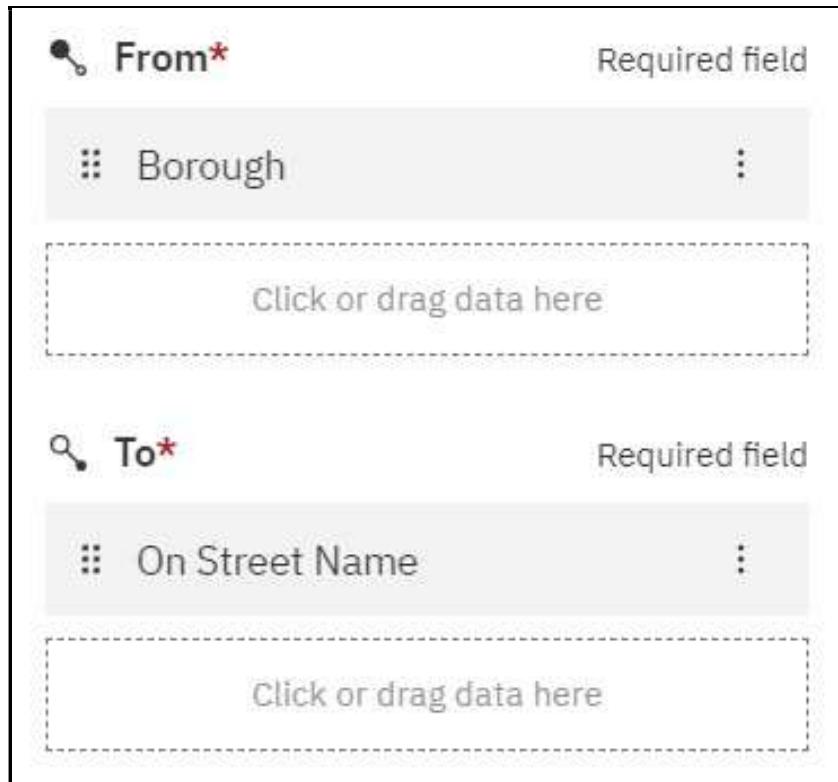
(PS I know that the crosstab shows the city data as well, but I still added city in a pie chart for more visual aid)

**QUERY 7:** Give me the map view of any city street where accident injury rate is high.

**NOTE:** To get the exact street view on map first we need to select that street on Network Tool.

**STEP 1:** Drag and drop a ‘Network’ tool from the visualisation tab on the Default Dashboard.

**STEP 2:** On the right side of the screen in the fields of Network tool add following Data Sets In  
‘From’ field add Borough data set and in ‘To’ field add On Street Name.



**STEP 3:** Now drag down a ‘Map’ tool from visualisation tab into dashboard.

**STEP 4:** In the fields of Map tool, add data set Borough and in Location color add data set  
borough.

 **Locations\*** Required field

⋮ Borough ⋮

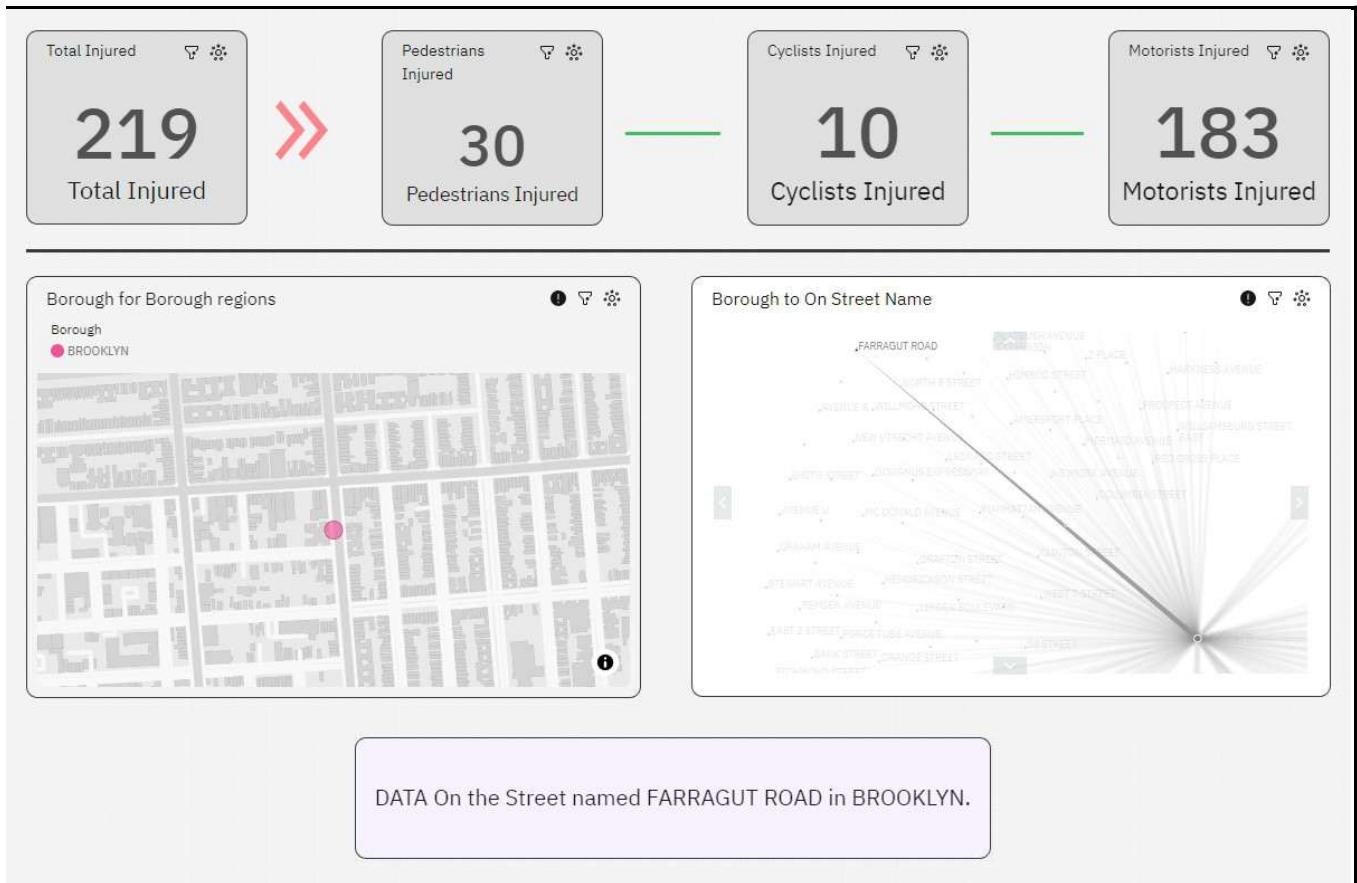
Click or drag data here

 Location color

⋮ Borough ⋮

Click or drag data here

**STEP 5:** In the Network tool you can zoom and select a particular street in a particular city.  
(We will select the street named FARRAGUT ROAD in the city BROOKLYN).



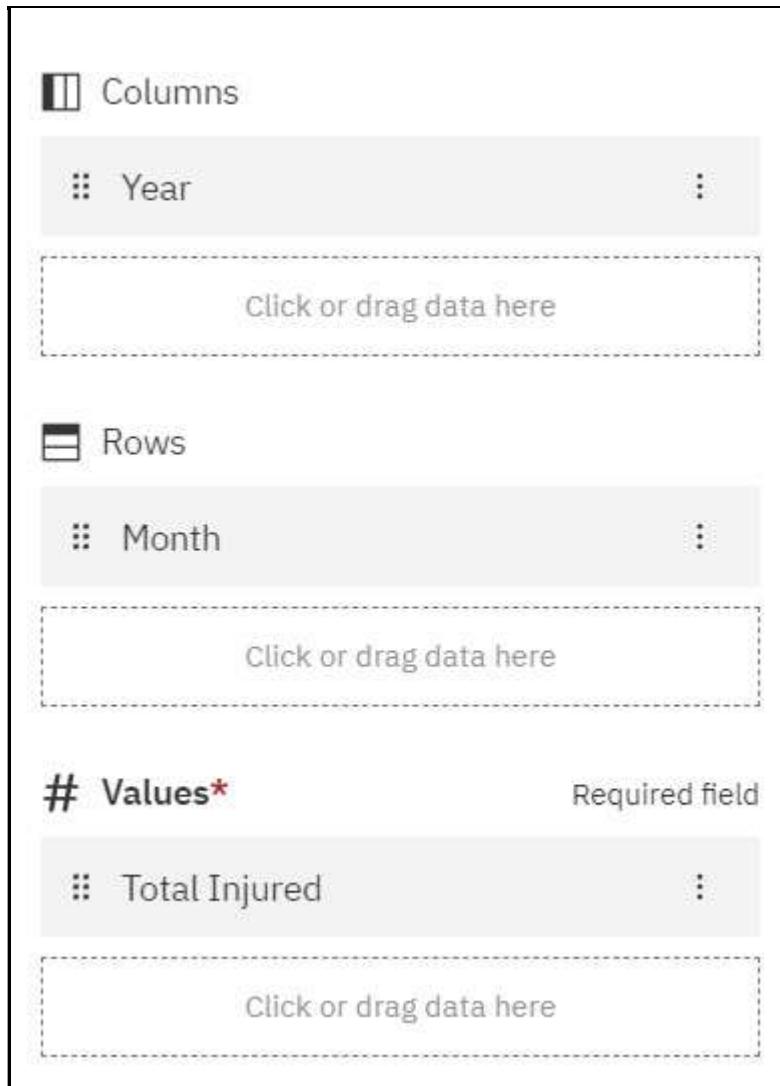
**NOTE:** Now after selecting the street Farragut Road in Brooklyn we can see the Map tool also changes and gives a map view of that street.

**QUERY 8:** Give the total number of all kind of injuries on the basis on month in a crosstab.

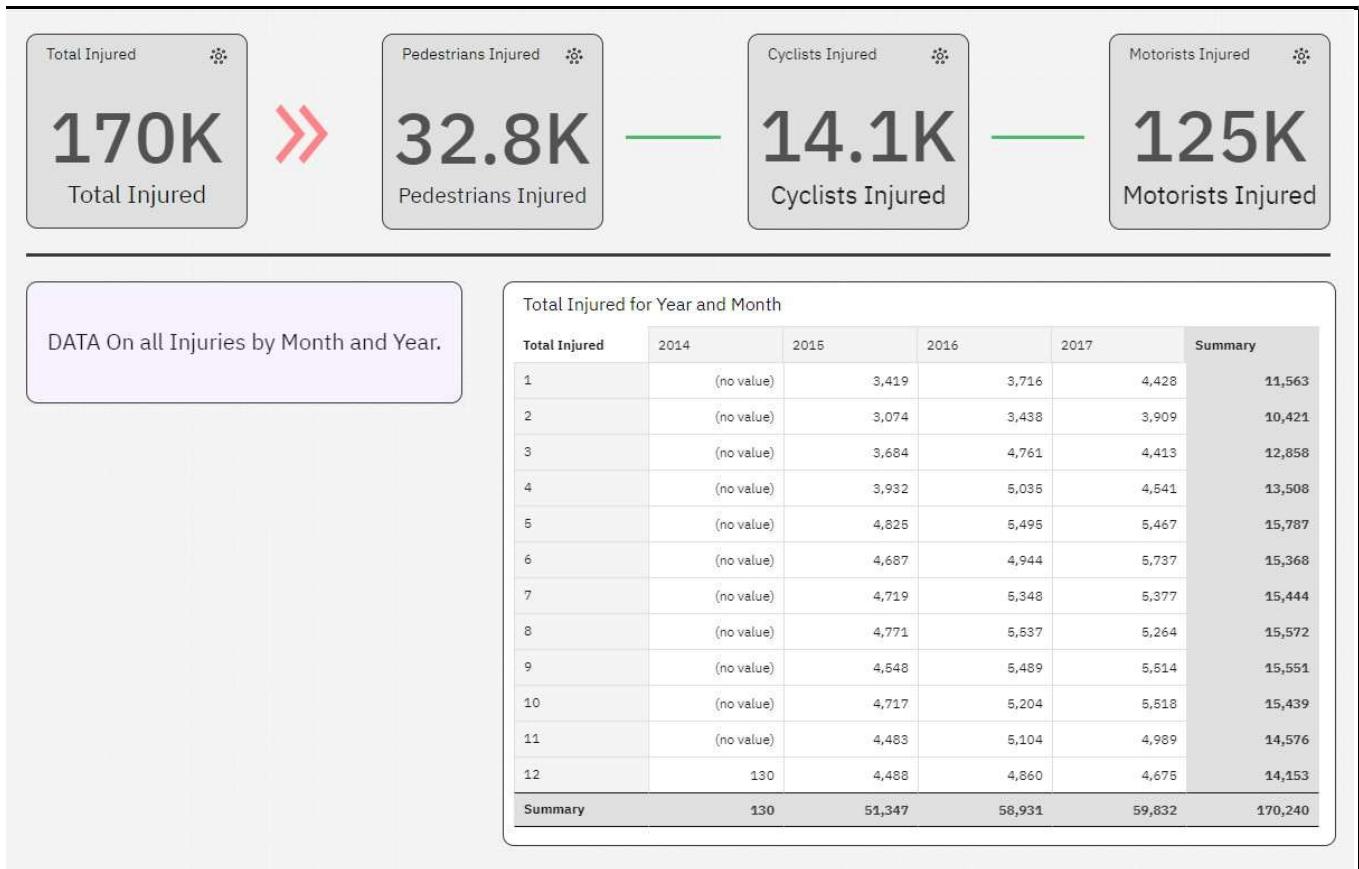
**NOTE:** For this question we will be using a Crosstab tool from visualisation tab.

**STEP 1:** Drag and drop Crosstab from visualisation tab into Default dashboard.

**STEP 2:** Now to show data on the basis of month using Crosstab we can add the “Month” data set in the **Rows** field of the Crosstab, then we add “Total Injured” data set in the **#Values** field. (We can also add “Year” data set in the column field).



**STEP 3:** Now the dashboard will show data on the basis of Month and Year.



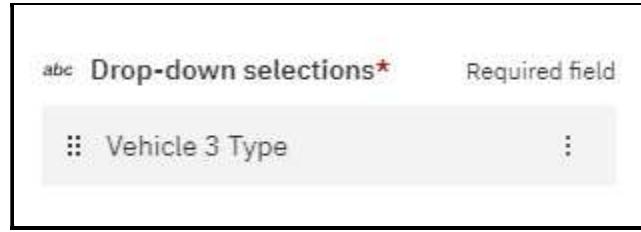
**NOTE:** In the Crosstab we can see the total injuries in each month for their respective years.

**QUERY 9:** NY Road Committee auth. wants to create some new street lane for more than 6 tire vehicles. So, suggest me any top 7 locations in New York in which we can build a new lane for heavy vehicles and why?

**NOTE:** To solve this problem we will use Pie Chart to show data visually.

**STEP 1:** On a Default Dashboard drag and drop a Drop-down list tool.

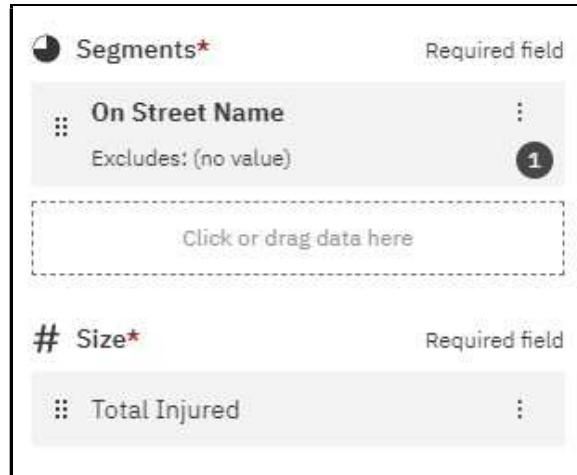
**STEP 2:** In the field ‘Drop-Down selection’ add “Vehicle 3 type”. (We are using Vehicle 3 type data because vehicles with 6 or more tires come under this data set.)



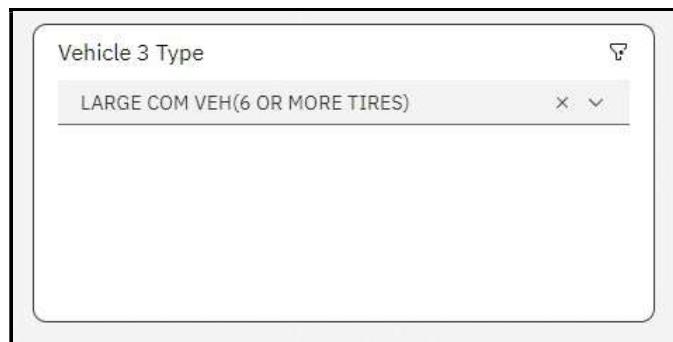
**STEP 3:** Now that we have selected the vehicle type, we can visualize the data using Pie Chart.

**STEP 4:** Drag and drop a Pie Chart tool into the dashboard.

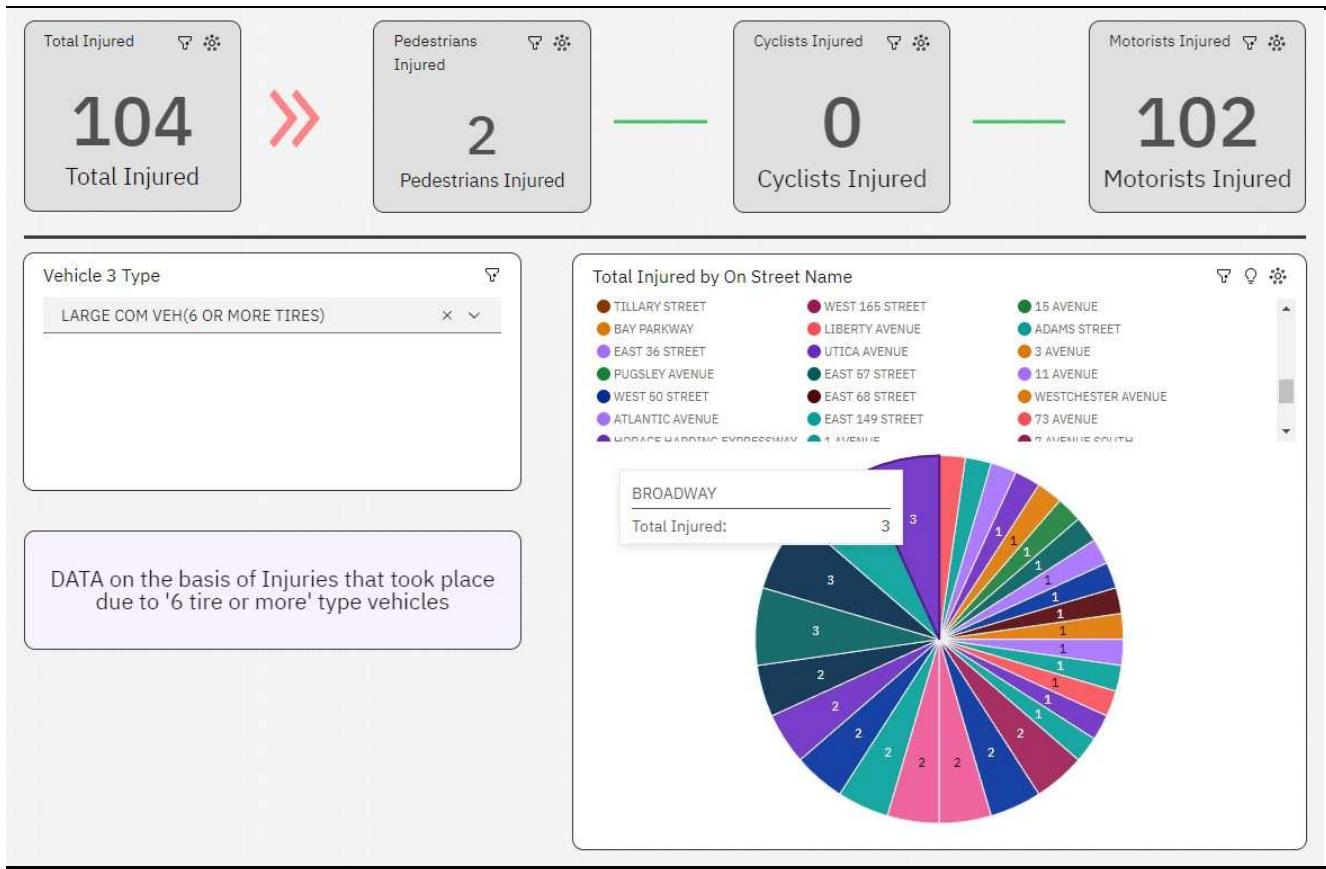
**STEP 5:** In the “Segment” field of the Pie chart add “On street name” data set and in the “#Size” field add “Total Injured” data set.



**STEP 6:** Now we can select the vehicle with “6 or more tire” type on the Drop-Down list tool.



**NOTE:** So, the dashboard now shows all the injuries that took place due to large vehicles that have 6 or more tires. But to find out exact streets in which these accidents took place we can look at the pie chart.



**NOTE:** So, as we can see on the dashboard, if we hover over the segments of the pie chart, we can see the Street name on which the injuries happened.

By hovering over the Pie chart segments we found out, the top 7 streets on which injuries took place by a vehicle with 6 or more tires are:

1. BROADWAY
2. WHITE PLAINS ROAD
3. 108 STREET
4. CONDUIT BOULEVARD
5. 9th AVENUE
6. EAST 49 STREET
7. ROCKAWAY AVENUE

We excluded (no value) street from the pie chart because it doesn't specify a specific street.



