**Analysis of NYPD Vehicle Collision: Case Study**

CLASS REPORT

***by***

AARTI AGRAHARI

(Roll Number(s): 22201020001)

Department of Computer Science

United University

Allahabad

**BONAFIDE CERTIFICATE**

This is to certify that this project report entitled **“Analysis of NYPD Vehicle collision: Case Study”** submitted to **United University Allahabad**, is a bonafide record of work done by “**Aarti Agrahari”** under my supervision from **“13April,2023**” to “**25April,2023” .**

MR. KULDEEP GUPTA

DATA ENGINEER

Prayagraj, India

25 April 2023

**Declaration by Author(s)**

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

Aarti Agrahari

Roll Number: 22201020001

Prayagraj, India

25 April 2023

**TABLE OF CONTENTS**

| **Content** | **Page No.** |
| --- | --- |
| Introduction to IBM Cognos and Data Analytics | 1-2 |
| IBM Cognos Dashboard | 3-4 |
| Standard Problems | 5-13 |
| Scenarion Based problems | 14-17 |
| Conclusion | 18 |

**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.

**6. Investigation outcome:**

The dataset includes information on the outcome of the investigation, such as whether an arrest was made or the case was closed due to lack of evidence.

**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.

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. Bubble:

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

values, such as revenue and profit.

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. 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. Summary:

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

for a categorical column.

1. Scatter:

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

only at regular tick marks.

.

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. Dataplayer:

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

visualizations.

**Standard Problems**

**Some of the Standard Problem queries which were asked:**

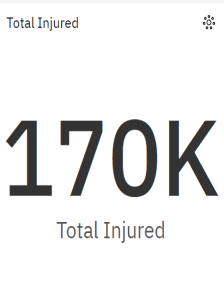
1. You have to show the total number of "***total injury/total killed"***
2. Show the Summary from total injury number: Pedestrian injured, Cyclist and Motorcyclist injured.
3. Show me the injuries on the basis of year.
4. Show me the number of injuries on the basis on "vehicle 1 contribution factor”, “vehicle 2 contribution factor" and “vehicle 3 contribution factor"
5. Give me the number of pedestrian/cyclist/motorcyclists injured on the basis on any particular city, month and year.
6. Give me the map view of any city street where accident injury rate is high.
7. Give the total number of all kind of injuries on the basis on month in a crosstab.

**Interpretation of these queries are:**

1. **For showing the total number of "*total injury*” we will simply use summary tool**

**STEP:**

* First open a Dashboard in an IBM Cognos and go to Visualization tab on left.
* Then select the **Summary Tool** and drag-drop.
* Select ‘total injured’ from the NYPD dataset in Value.
* Then Result will be shown in this way.

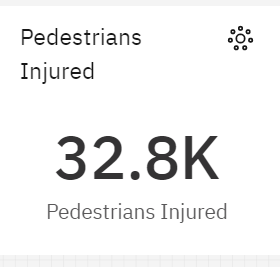


*From this data we get information that in 3 years (2015-17) total number of people injured are 170,000.*

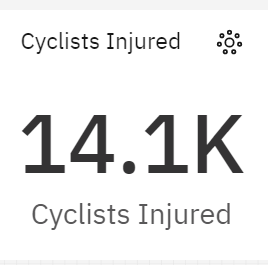
1. **For showing “pedestrian injured, cyclist and motorcyclist injured” from total injured we will simply use summary tool.**

**STEP:**

* First open a Dashboard in an IBM Cognos and go to Visualization tab on left.
* Then select the **summary tool** and drag-drop.
* We select ‘pedestrian injured’ from the NYPD dataset in Value.
* Finally total pedestrian injured data is showing in following set.



* Copy the summary tool and select the cyclists injured data is showing in following set.



* Copy the summary tool and select the motorcyclist injured data is showing in following set.

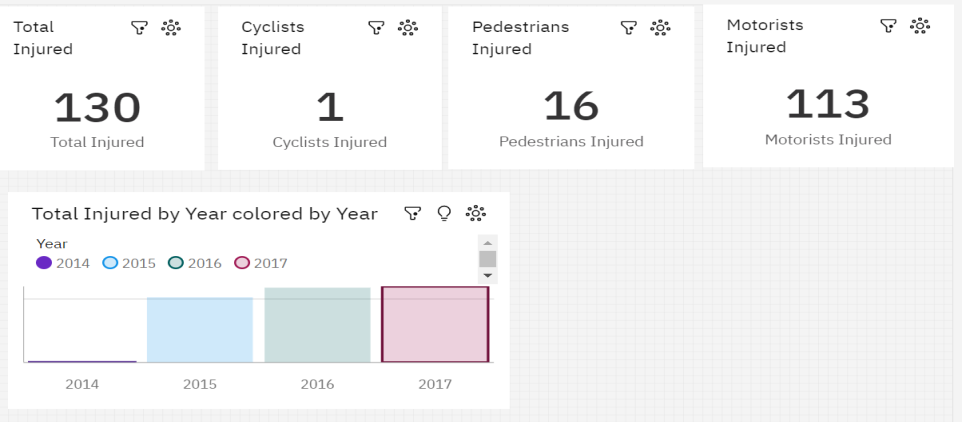


*From this data we get that different number of people get injured by pedestrian, cyclistand motorcyclist injuries*. *So according to this data from total 170,000 people injured 32,800 are Pedestrians, 14,100 are Cyclists and 125,000 are Motorists injured.*

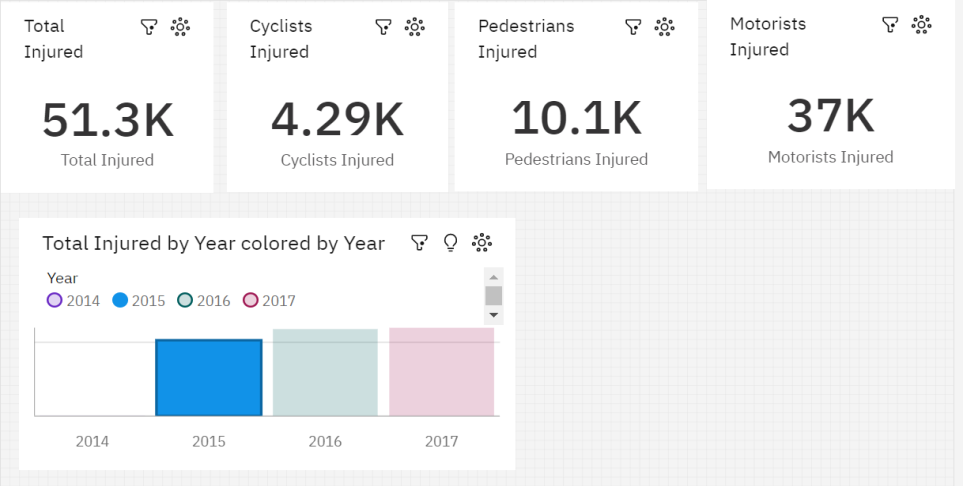
**3. For showing the injuries on the basis of year we will use column tool.**

**STEP:**

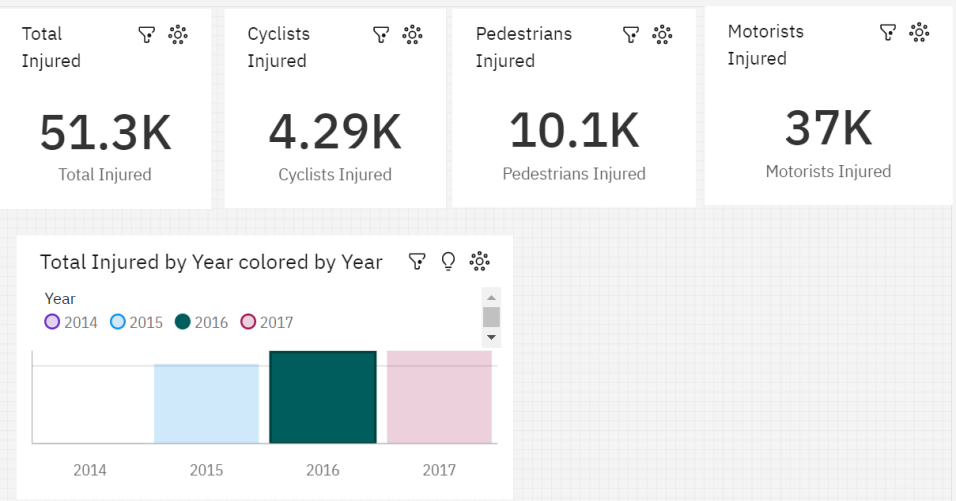
* First open a Dashboard in an IBM Cognos and go to Visualization tab on left.
* Then select the **column tool** and drag-drop.
* We select ‘total injured’ in Length, ’year’ in Bars and ‘years’ in color from NYPD dataset.
* Finally the injuries on the basis of year is showing in following set.



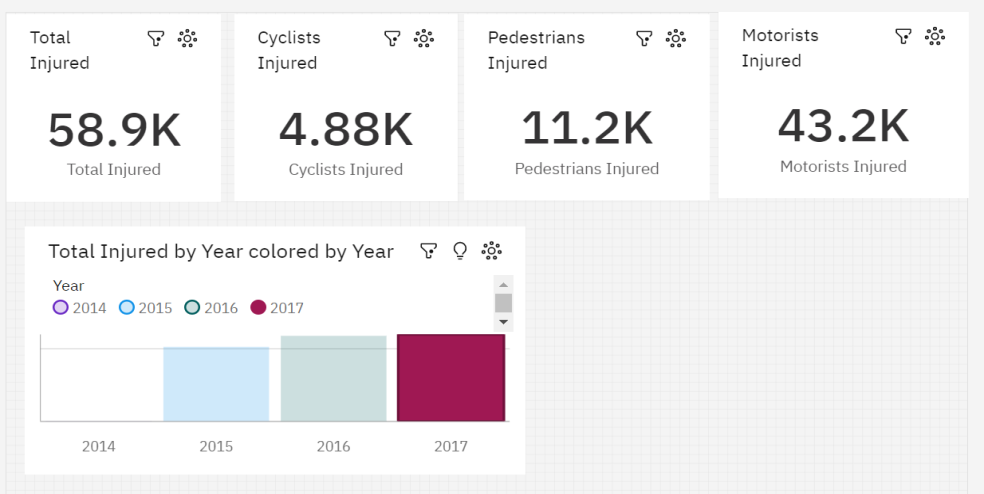
**Data of injuries in 2014**



**Data of injuries in 2015**



**Data of injuries in 2016**



**Data of injuries in 2017**

*From this data we get that different number of people get injured by pedestrian, cyclist and motorcyclist injuries on the basis of year*. *So according to this data,*

*In 2014 total 130 people injured,16 are Pedestrians, 1 is Cyclists and 113 are Motorists injured.*

*In 2015 total 51.3K people injured,10.1K are Pedestrians, 4.29 K are Cyclists and 37K are Motorists injured.*

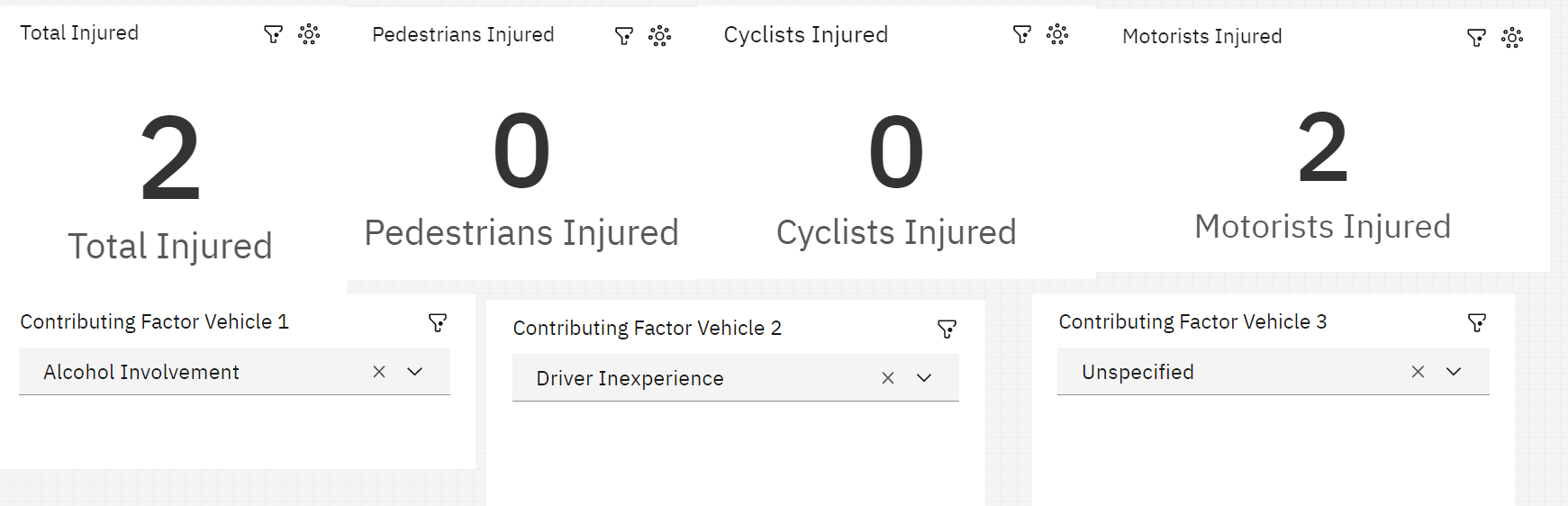
*In 2016 total 51.3K people injured,10.1K are Pedestrians, 4.29K are Cyclists and 37K are Motorists injured.*

*In 2017 total 58.9K people injured,11.2K are Pedestrians, 4.88K are Cyclists and 43.2K are Motorists injured.*

**4. For showing the number of injuries on the basis on “ Vehicle 1 contribution factor”, “vehicle 2contribution factor" and “vehicle 3 contribution factor" we will use Drop Down menu**

**STEP**:

* First open a Dashboard in an IBM Cognos and go to Visualization tab on left.
* Then select the **Drop Down menu** and drag-drop in three times.
* We select any factor in vehicle 1 contribution factor, vehicle2 contribution factor , vehicle3 contribution factor and it will be reflected on Summary scorecard.
* Finally number of injuries is showing in following set.

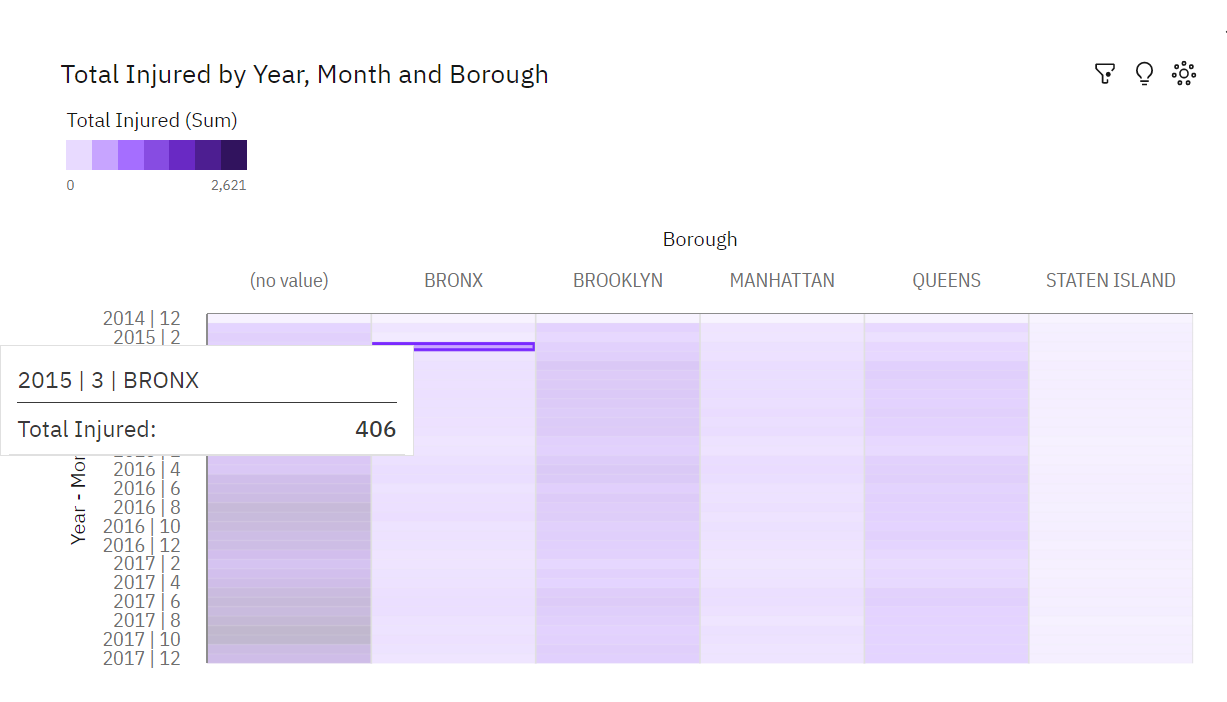
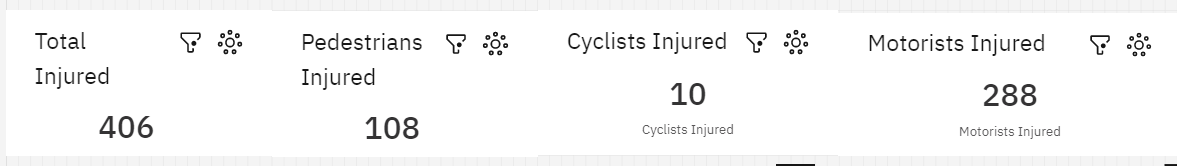


*From this data we get that different number of people get injured by pedestrian, cyclist and motorcyclist injuries on the basis on “ Vehicle 1 contribution factor”, “vehicle 2 contribution factor" and “vehicle 3 contribution factor". So according to this data from total 2 people injured 0 are Pedestrians,0 are Cyclists and 0 are Motorists injured.*

**5. For showing the number of pedestrian/cyclist/motorcyclists injured on the basis on any particular city, month and year.**

**STEP:**

* First open a Dashboard in an IBM Cognos and go to Visualization tab on left.
* Then select the **Heat Map tool** and drag-drop.
* We select ‘Borough’ in columns, ‘Year and Month’ in Rows and ‘Total injured’ in Heat from NYPD dataset.
* Finally the number of pedestrian/cyclist/motorcyclists injured on the basis on any particular city, month and year are reflected back in Summary Scorecard is showing in following set.

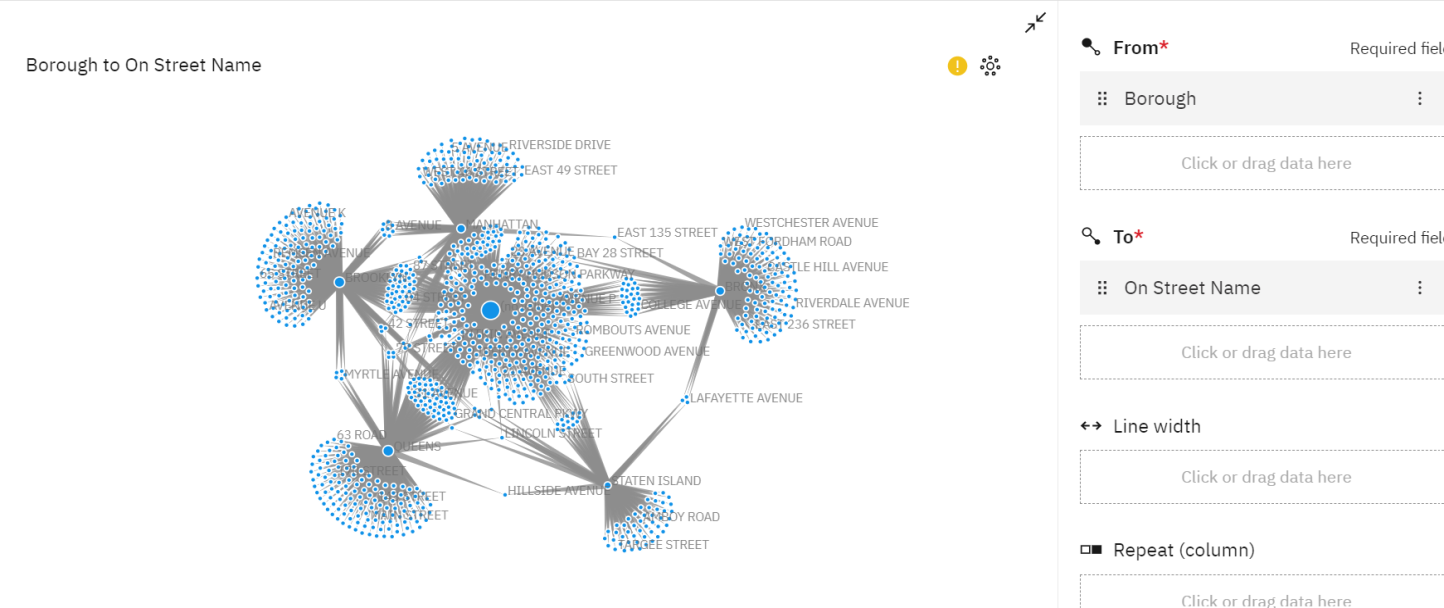


*From this data we get that different number of people get injured by pedestrian, cyclist and motorcyclist injuries on the basis on “*2015,3th month of year and Bronx Borough(City)*". So according to this data from total 406 people injured 108 are Pedestrians,10 are Cyclists and 288are Motorists injured.*

**6. For showing the map view of any city street where accident injury rate is is high.**

**STEP:**

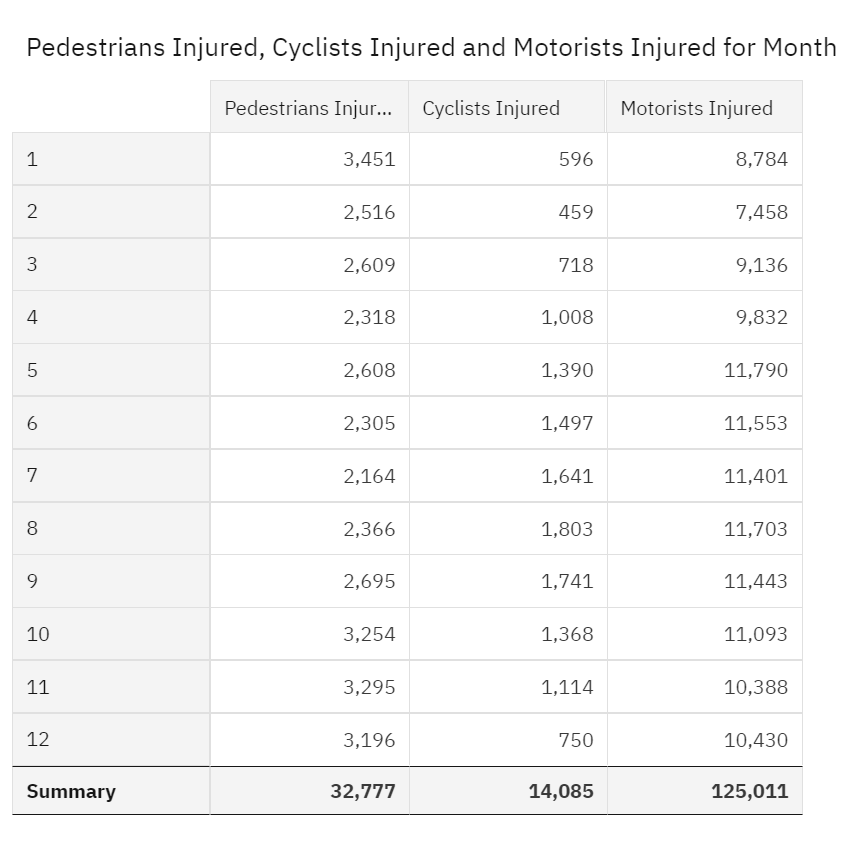
* First open a Dashboard in an IBM Cognos and go to Visualization tab on left.
* Then select the **Network** and drag-drop.
* We select ‘Borough’ in From and ‘On Street Name’ in To Value from NYPD dataset.
* Finally get the map view of any city street where accident injury rate is is high.

**

**7.For showing the total number of all kind of injuries on the basis on month in a crosstab.**

**STEP:**

* First open a Dashboard in an IBM Congnos and go to Visualization tab on left.
* Then select the **Crosstab tool** and drag-drop.
* We select ‘Month’ in Rows and ‘pedestrian injured, cyclist injured, motorcyclists injured’ in Value from NYPD dataset.
* Finally the total number of all kind of all kind of injuries on the basis on month in a crosstab is showing in following set.



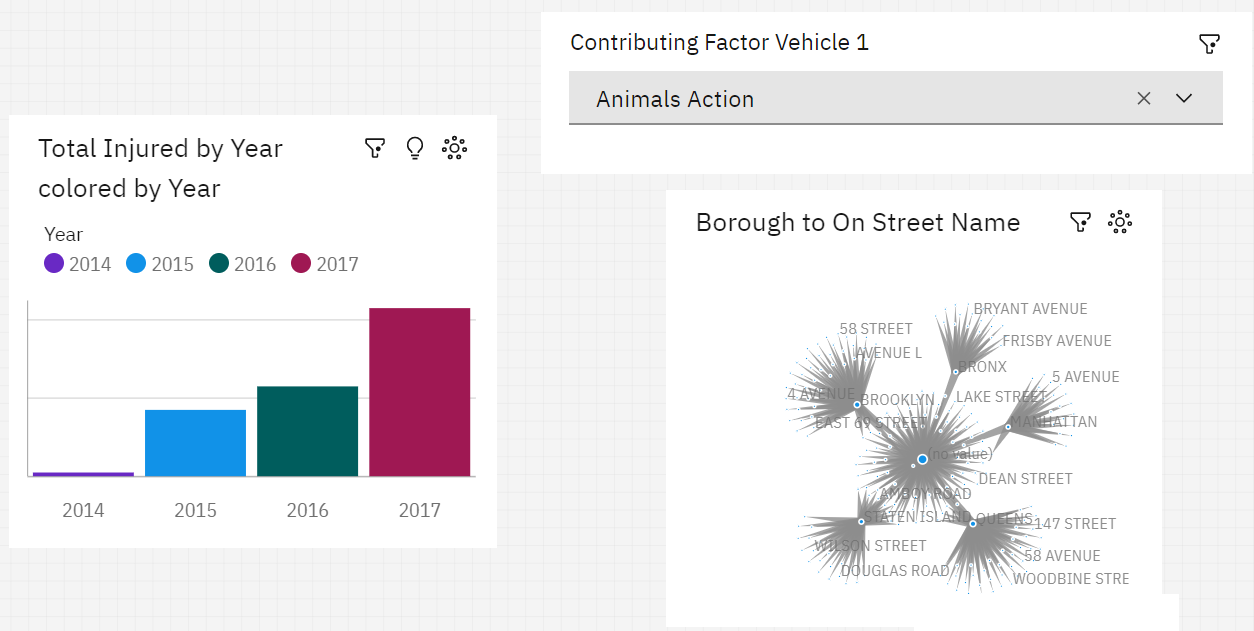
*From this data we get the information about the number of people injured on the basis of types of injury and month.*

**Scenarion Based problems**

1. New York animal welfare department wants to sbdue 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.
2. 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?
3. NY Road Commity auth. want to create some new stree 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?
4. **For showing New york animal welfare department wants to sbdue the rate of accident due to Animal accident factor on the basis of particular city and on which particular street they have to focus on.**

**STEP:**

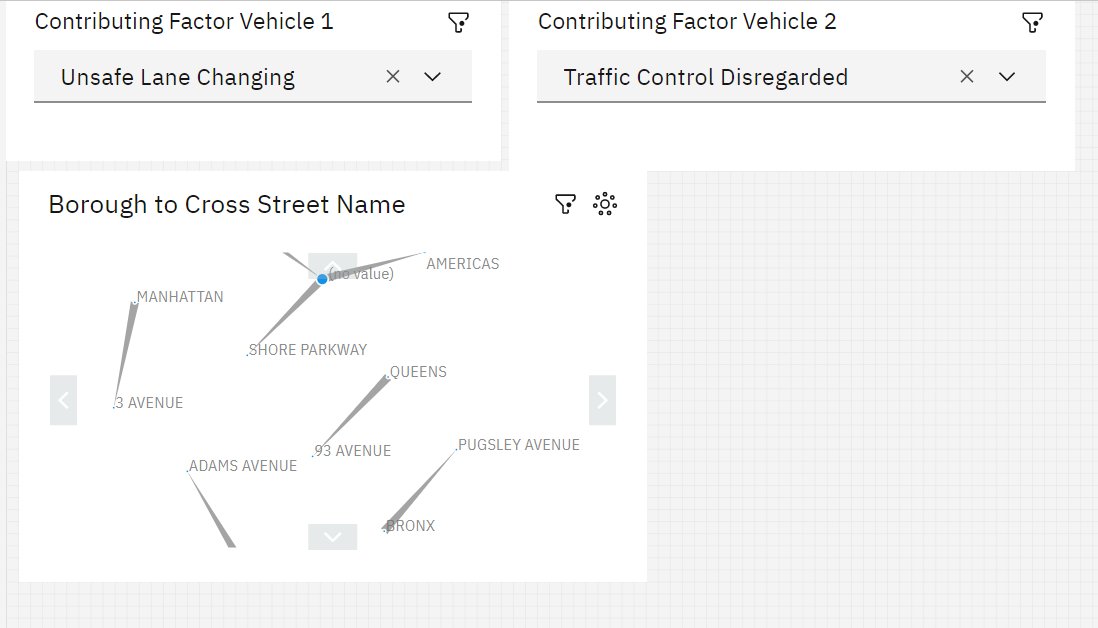
* First open a Dashboard in an IBM Cogons and go to Visualization tab on left.
* Then select the **Column tool** and drag-drop.
* We select ‘Years’ in Bar and ‘Total injured’ in Length and ‘year’ in Color from NYPD dataset.
* Select the  **Drop down Menu** and drag-drop.
* We select ‘Contributing Factor Vehicle 1’ in Drop Down selection from NYPD dataset.
* Select the  **Network** and drag-drop.
* We select ‘Borough’ in From and ‘On Street Name’ in To selection from NYPD dataset.
* Finally get the information about the rate of accident due to Animal accident factor on the basis of particular city and on which particular street they have to focus on.

****

1. **NY Traffic Department wants to spread the traffic awareness knowledge to some local citizens who are not properly aware about the rules.**

**STEP:**

* First open a Dashboard in an IBM Cogons and go to Visualization tab on left.
* Select the  **Drop down Menu** and drag-drop.
* We select ‘Contributing Factor Vehicle 1’ in Drop Down selection from NYPD dataset.
* Select the  **Drop down Menu** and drag-drop.
* We select ‘Contributing Factor Vehicle 2’ in Drop Down selection from NYPD dataset.
* Select the  **Network** and drag-drop.
* We select ‘Borough’ in From and ‘Cross Street Name’ in To selection from NYPD dataset.

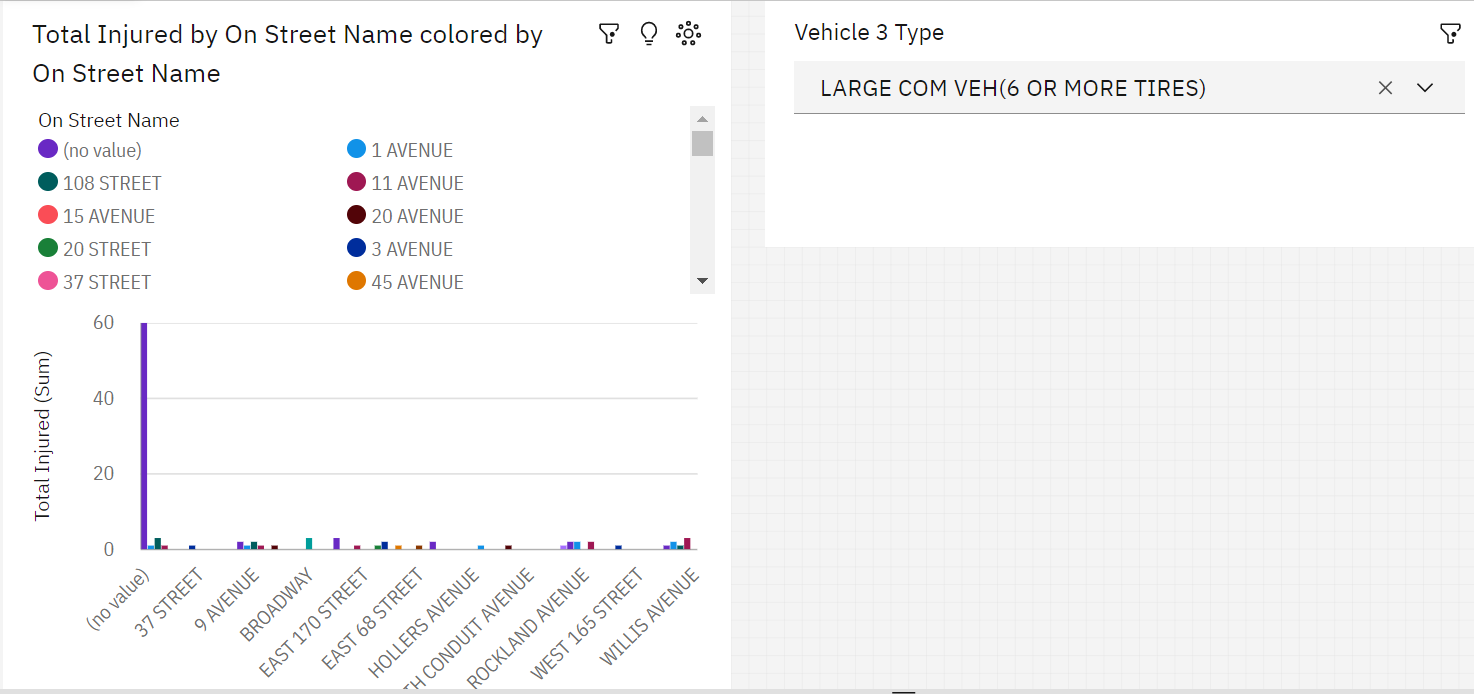


*From this data we get the information about injuries which are happening due to lack of knowledge of traffic rules by the help of Network Tool.*

1. **NY Road Commity auth. want to create some new stree lane for  more than 6 tiers vehicle. Suggest any top 7 locations in New York in which we can build a new lane for heavy vehicles .**

**STEP:**

* First open a Dashboard in an IBM Cogons and go to Visualization tab on left.
* Then select the **Column tool** and drag-drop.
* We select ‘On Street Name’ in Bar and ‘Total injured’ in Length and ‘On Street Name’ in Color from NYPD dataset
* Select the  **Drop down Menu** and drag-drop.
* We select ‘ Vehicle 3’ in Drop Down selection from NYPD dataset.
* In field of Drop down Tool fil Large Com Veh(6 or more tires).

****

*These are the location in New York which is mention in screenshot where most of the people got injured due to the heavy traffic by large com vehicles and short roads.*

**Conclusion**

The NYPD Motor Vehicle Collisions dataset in IBM Cognos is a collection of data related to motor vehicle collisions that occurred in New York City. The dataset contains detailed information about each collision, including the date and time, location, vehicle types, contributing factors, and injuries.

The dataset is maintained by the New York City Police Department (NYPD) and is publicly available for analysis and research purposes.

The dataset is an important resource for researchers, policymakers, and law enforcement officials who are interested in understanding the factors that contribute to motor vehicle collisions and developing strategies to reduce their occurrence and severity.

IBM Cognos is a business intelligence platform that provides tools for data visualization, reporting, and analysis. The NYPD Motor Vehicle Collisions dataset can be loaded into IBM Cognos to conduct data analysis and create reports and visualizations that can help identify trends and patterns in the data.

Overall, the NYPD Motor Vehicle Collisions dataset in IBM Cognos is a valuable resource for anyone interested in studying motor vehicle collisions in New York City and improving road safety in the city.