NYC Motor Vehicles Collisions crash data published by New York city agencies and partners at NYC Open data which contains details on crash event. Data contains information from all the police reported motor vehicle collisions where MV104-AN form is to be filled which is mandatory form in NY when there is at least 1000$ of damage.

What is Expected in this report / SQL Operations performed ?

To perform comprehensive detailed traffic safety analysis using SQL on the last 10 years data depicting fatalities , deceased count in each borough and to identify significant factors responsible and further sharing feedback to improve on some areas identified.

- Injuries and Causalities Trend (Yearly, Vehcile type)
- Year wise deceased analysis.
- Finding out peak crash hour in a day.
- Significant contributing factor (For each borough, street)
- Top 5 streets with high crash rate(each borough)

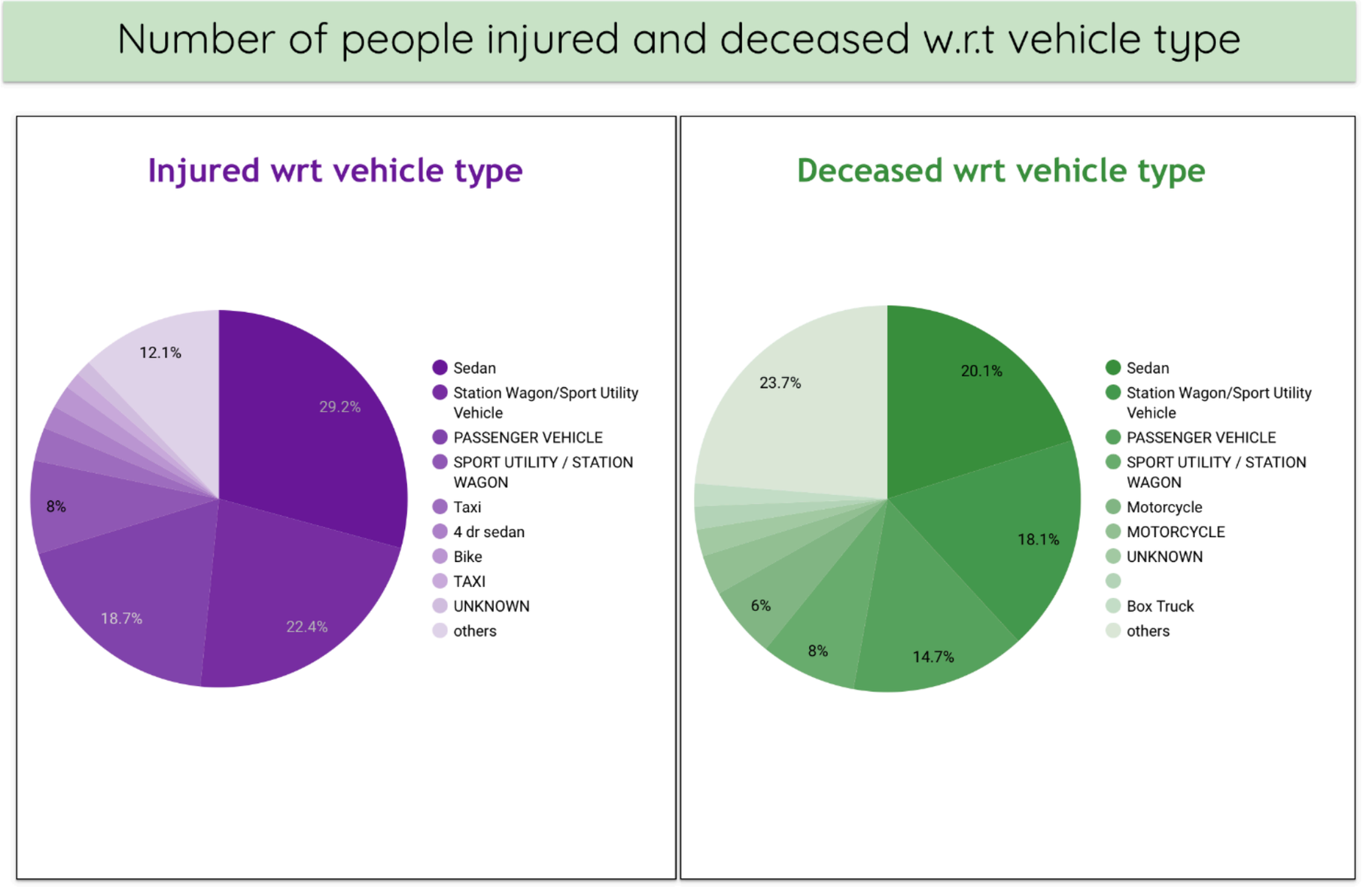
Tools used: Google Big query

Number of people injured and deceased w.r.t vehicle type

SQL Query:

```
select vehicle_type_code1,
number_of_cyclist_injured + number_of_motorist_injured +
number_of_pedestrians_injured + number_of_persons_injured as
Total_injured,
number_of_cyclist_killed+number_of_motorist_killed+
+number_of_pedestrians_killed+ number_of_persons_killed as
Total_deceased
from
(select vehicle_type_code1,
sum(number_of_cyclist_injured) as number_of_cyclist_injured,
sum(number_of_cyclist_killed) as number_of_cyclist_killed,
sum(number_of_motorist_injured) as number_of_motorist_injured,
sum(number_of_motorist_killed) as number_of_motorist_killed,
sum(number_of_pedestrians_injured) as number_of_pedestrians_injured,
sum(number_of_pedestrians_killed) as number_of_pedestrians_killed,
sum(number_of_persons_injured) as number_of_persons_injured,
sum(number_of_persons_killed) as number_of_persons_killed
from `assignment-1-
368502.nypd_motor_vehicle_collisions.nypd_mv_collisions`
group by vehicle_type_code1)
```

Output



Seadan and Sport Utility vehicles have the most number of casualties

Observation: Sedan and Sport Utility vehicles has the highest number of causalities reported

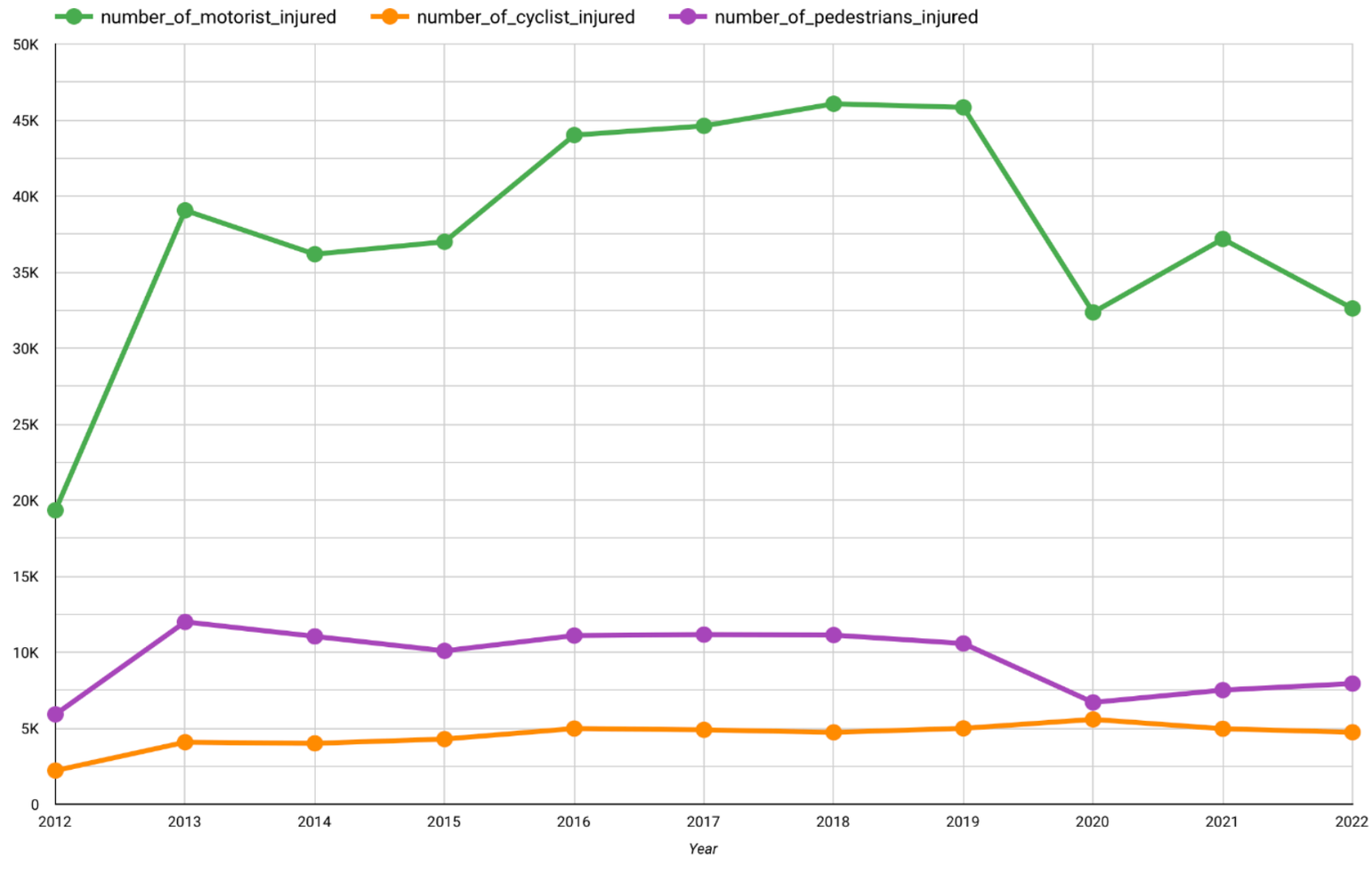
Year on Year injury analysis

SQL Query:

```
with casuality_count as
(select Year,
number_of_cyclist_injured, number_of_cyclist_killed,
number_of_cyclist_injured + number_of_cyclist_killed as
Total_Cyclist_Casualities,
number_of_pedestrians_injured, number_of_pedestrians_killed,
number_of_pedestrians_injured + number_of_pedestrians_killed as
Total_Pedestrians_Casualities,
number_of_motorist_injured, number_of_motorist_killed,
number_of_motorist_injured + number_of_motorist_killed as
Total_Motarist_Casualities,
number_of_persons_injured,number_of_persons_killed,
number_of_persons_injured + number_of_persons_killed as
Total_person_Casualities,
(number_of_cyclist_injured + number_of_cyclist_killed +
number_of_pedestrians_injured + number_of_pedestrians_killed +
number_of_motorist_injured + number_of_motorist_killed +
number_of_persons_injured + number_of_persons_killed ) as
Total_Casualities
from (select
Year,
```

```
sum(number_of_cyclist_injured) as number_of_cyclist_injured,
sum(number_of_cyclist_killed) as number_of_cyclist_killed,
sum(number_of_motorist_injured) as number_of_motorist_injured,
sum(number_of_motorist_killed) as number_of_motorist_killed,
sum(number_of_pedestrians_injured) as number_of_pedestrians_injured,
sum(number_of_pedestrians_killed) as number_of_pedestrians_killed,
sum(number_of_persons_injured) as number_of_persons_injured,
sum(number_of_persons_killed) as number_of_persons_killed,
from (select EXTRACT(YEAR FROM DATETIME(timestamp))
AS Year, *
from `assignment-1-
368502.nypd_motor_vehicle_collisions.nypd_mv_collisions` ) group
by Year
))
select *,(Total_Casualities*100)/ (select sum(Total_Casualities)
from casuality_count) as Total_Casualities_Percentage_Reported from
casuality_count
order by Total_Casualities desc
```

Year on Year injury analysis



Inferences: It can be observed from the plot that, more motorist injured compared to cyclist and pedestrians

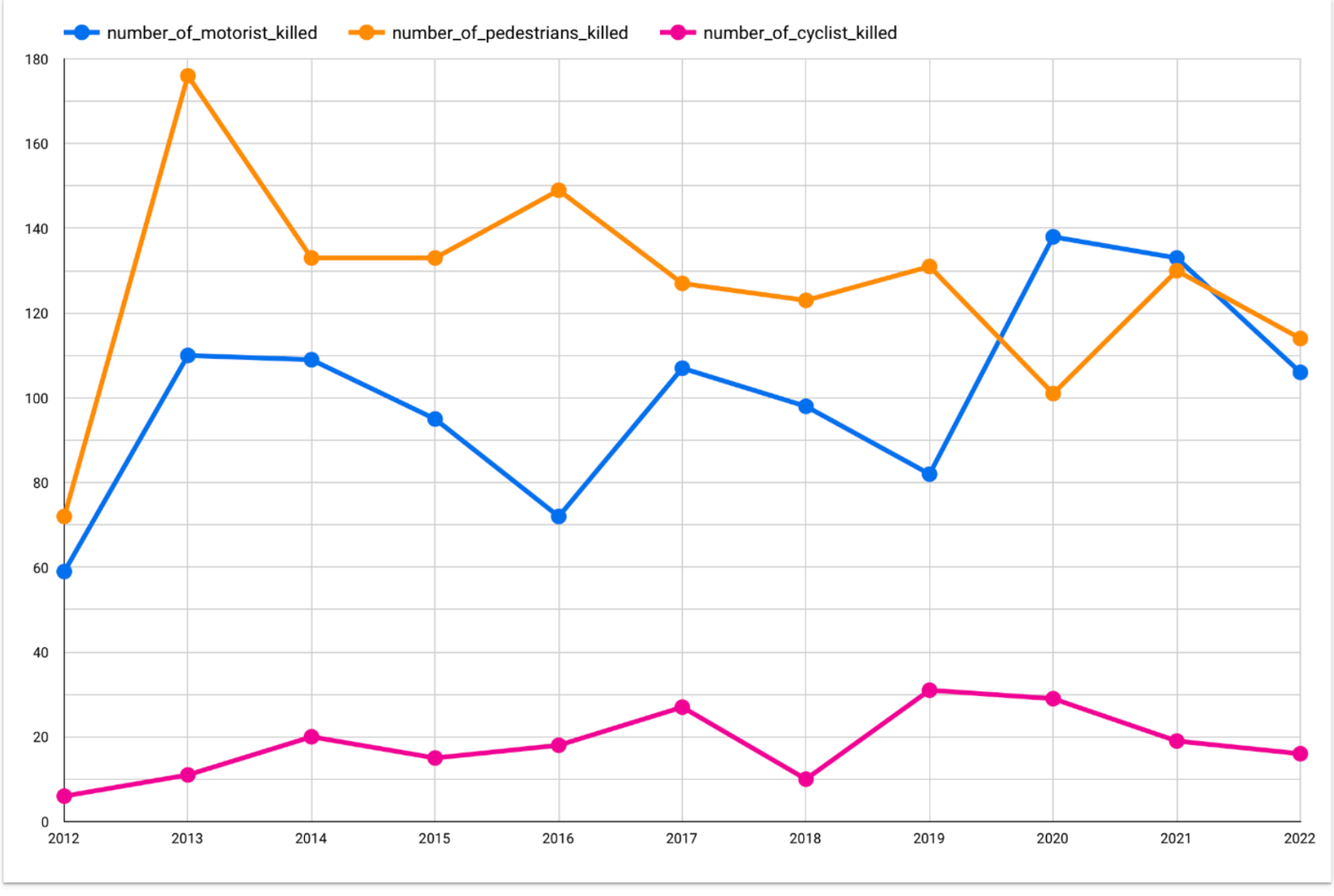
Yearly deceased analysis

SQL Query :

```
with casuality_count as
(select Year,
number_of_cyclist_injured, number_of_cyclist_killed,
number_of_cyclist_injured + number_of_cyclist_killed as
Total_Cyclist_Casualities,
number_of_pedestrians_injured, number_of_pedestrians_killed,
number_of_pedestrians_injured + number_of_pedestrians_killed as
Total_Pedestrians_Casualities,
number_of_motorist_injured, number_of_motorist_killed,
number_of_motorist_injured + number_of_motorist_killed as
Total_Motarist_Casualities,
number_of_persons_injured,number_of_persons_killed,
number_of_persons_injured + number_of_persons_killed as
Total_person_Casualities,
(number_of_cyclist_injured + number_of_cyclist_killed +
number_of_pedestrians_injured + number_of_pedestrians_killed +
number_of_motorist_injured + number_of_motorist_killed +
number_of_persons_injured + number_of_persons_killed ) as
Total_Casualities
from (select
Year,
sum(number_of_cyclist_injured) as number_of_cyclist_injured,
sum(number_of_cyclist_killed) as number_of_cyclist_killed,
sum(number_of_motorist_injured) as number_of_motorist_injured,
```

```
sum(number_of_motorist_killed) as number_of_motorist_killed,
sum(number_of_pedestrians_injured) as number_of_pedestrians_injured,
sum(number_of_pedestrians_killed) as number_of_pedestrians_killed,
sum(number_of_persons_injured) as number_of_persons_injured,
sum(number_of_persons_killed) as number_of_persons_killed,
from (select EXTRACT(YEAR FROM DATETIME(timestamp))
AS Year, *
from `assignment-1-
368502.nypd_motor_vehicle_collisions.nypd_mv_collisions` ) group
by Year
))
select *,(Total_Casualities*100)/ (select sum(Total_Casualities)
from casuality_count) as Total_Casualities_Percentage_Reported from
casuality_count
order by Total_Casualities desc
```

Yearly deceased analysis



Inference: Over the decade, it is evident that more number of pedestrians deceased at start and then these accidents are regulated by bringing the laws to give “right of way” for the pedestrians

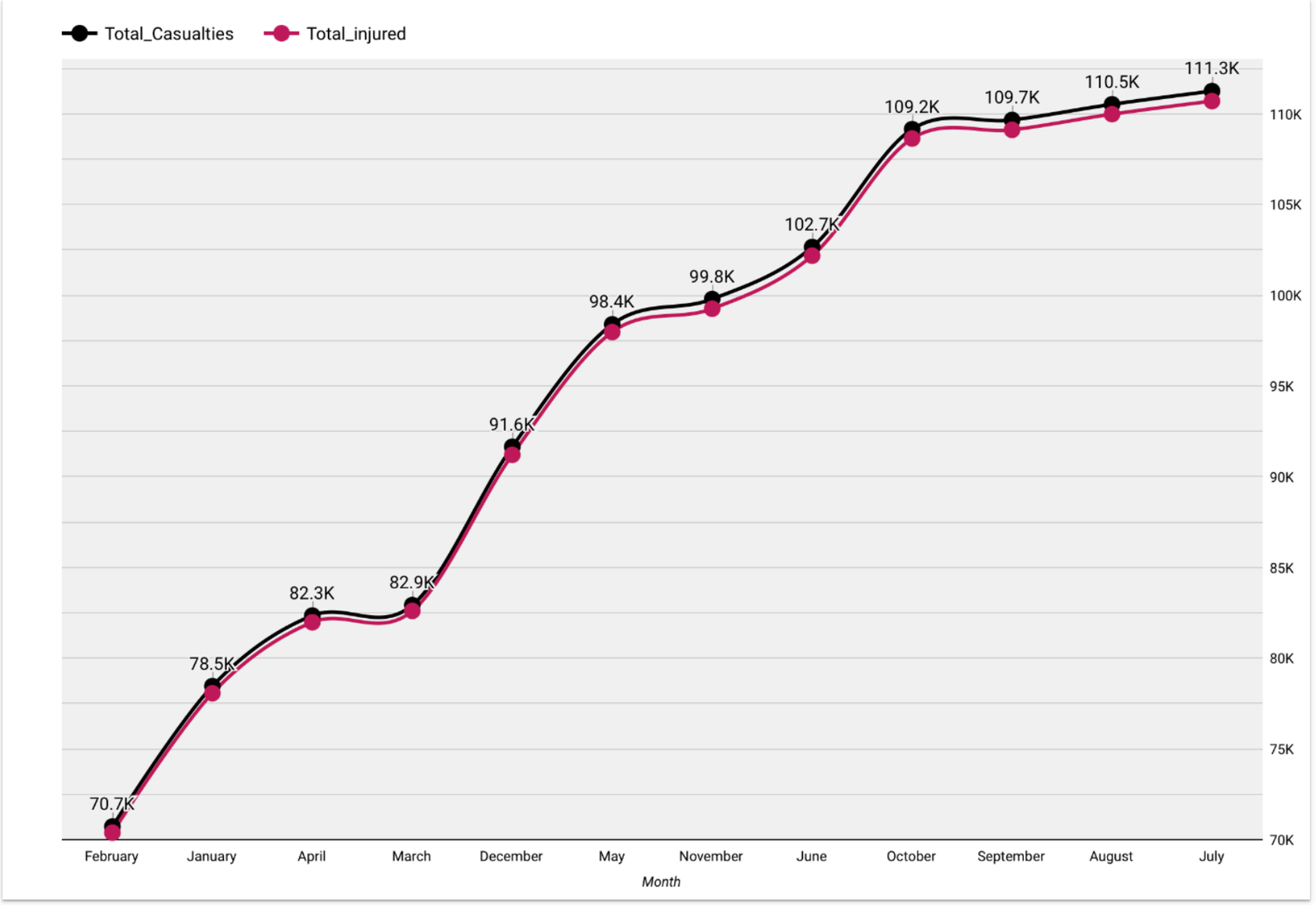
Monthly wise injuries and casualties

SQL Query:

```
with monthly_inj_deceased as
(select Month,
number_of_cyclist_injured + number_of_motorist_injured +
number_of_pedestrians_injured + number_of_persons_injured as
Total_injured,
number_of_cyclist_killed+number_of_motorist_killed+
+number_of_pedestrians_killed+ number_of_persons_killed as
Total_deceased
from
(select Month,
sum(number_of_cyclist_injured) as number_of_cyclist_injured,
sum(number_of_cyclist_killed) as number_of_cyclist_killed,
sum(number_of_motorist_injured) as number_of_motorist_injured,
sum(number_of_motorist_killed) as number_of_motorist_killed,
sum(number_of_pedestrians_injured) as number_of_pedestrians_injured,
sum(number_of_pedestrians_killed) as number_of_pedestrians_killed,
sum(number_of_persons_injured) as number_of_persons_injured,
sum(number_of_persons_killed) as number_of_persons_killed
from (select case EXTRACT(MONTH FROM DATETIME(timestamp))
     when 1 then 'January'
     when 2 then 'February'
     when 3 then 'March'
     when 4 then 'April'
```

```
      when 5 then 'May'
      when 6 then 'June'
      when 7 then 'July'
      when 8 then 'August'
      when 9 then 'September'
      when 10 then 'October'
      when 11 then 'November'
      when 12 then 'December'
      end Month, *
from `assignment-1-
368502.nypd_motor_vehicle_collisions.nypd_mv_collisions` )
group by Month))
select *,(Total_injured+Total_deceased) as Total_Casualties from
monthly_inj_deceased;
```

Monthly wise injuries and casualties



Inference: Causalities are total injured, and deaths combined. In months of October, September, August and July more number of causalities are observed.

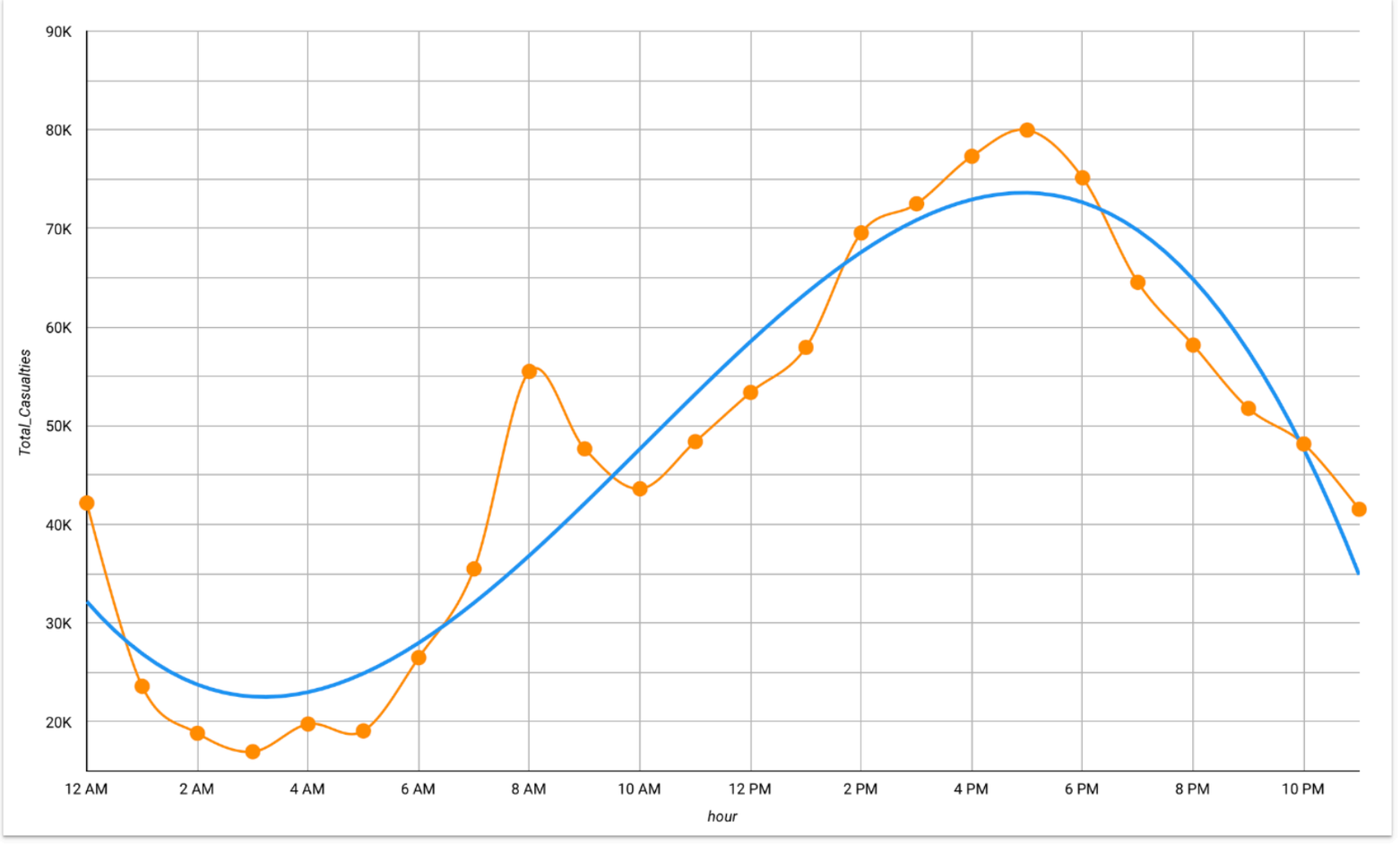
Total Causalities and its approximation hourly

SQL Query:

```
with hour_inj_deceased as
(select *,(Total_injured+Total_deceased) as Total_Casualties from
(select hour,
number_of_cyclist_injured + number_of_motorist_injured +
number_of_pedestrians_injured + number_of_persons_injured as
Total_injured,
number_of_cyclist_killed+number_of_motorist_killed+
+number_of_pedestrians_killed+ number_of_persons_killed as
Total_deceased
from
(select hour,
sum(number_of_cyclist_injured) as number_of_cyclist_injured,
sum(number_of_cyclist_killed) as number_of_cyclist_killed,
sum(number_of_motorist_injured) as number_of_motorist_injured,
sum(number_of_motorist_killed) as number_of_motorist_killed,
sum(number_of_pedestrians_injured) as number_of_pedestrians_injured,
sum(number_of_pedestrians_killed) as number_of_pedestrians_killed,
sum(number_of_persons_injured) as number_of_persons_injured,
sum(number_of_persons_killed) as number_of_persons_killed
from (select EXTRACT(hour FROM DATETIME(timestamp)) as hour, *
from `assignment-1-
368502.nypd_motor_vehicle_collisions.nypd_mv_collisions` )
```

```
        group by hour)))
        select *,(Total_Casualties * 100)/(select sum(Total_Casualties) from
        hour_inj_deceased) as total_percentage_casualty_per_hr
from hour_inj_deceased order by total_percentage_casualty_per_hr desc;
```

Total Casualities and its approximation hourly



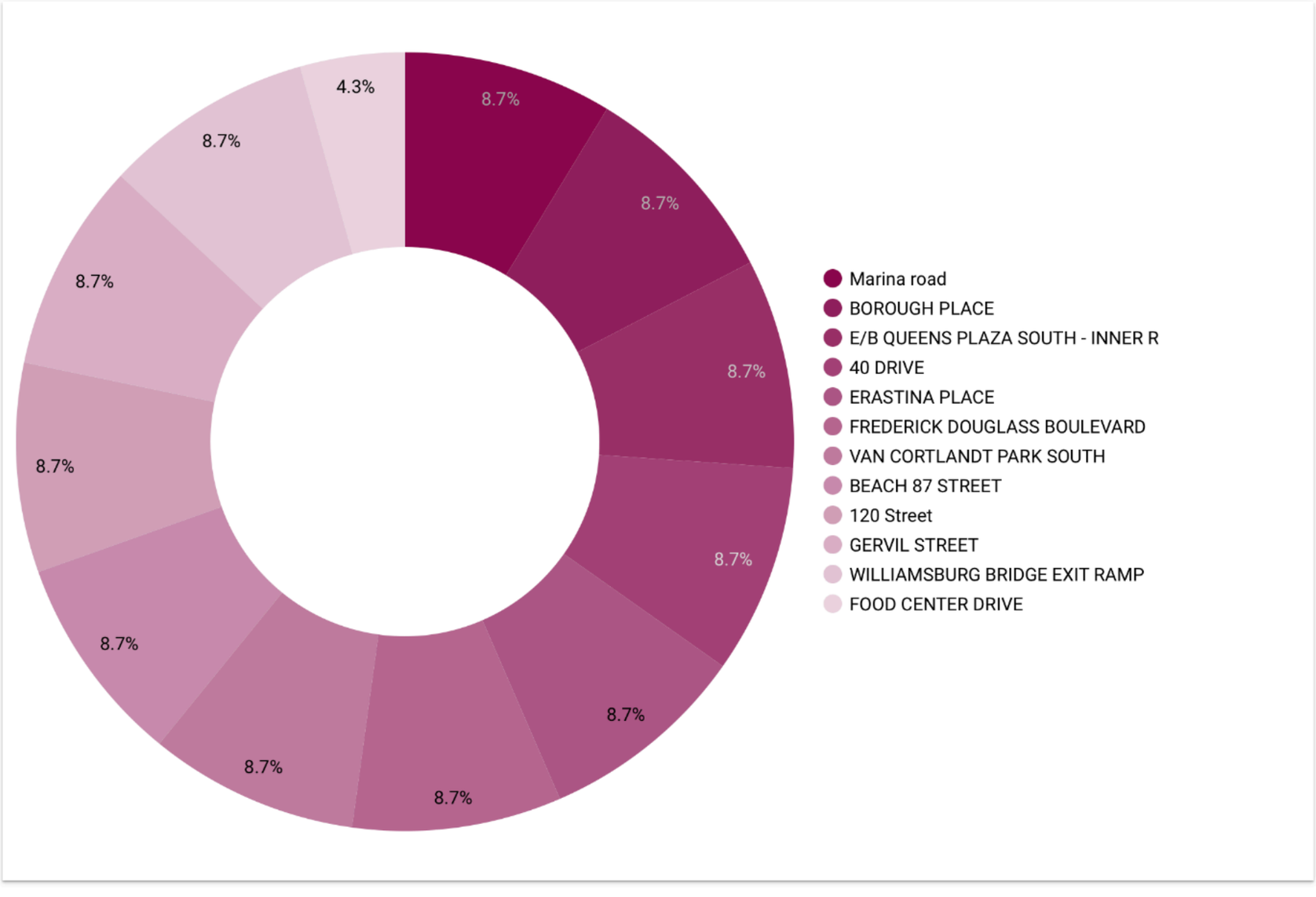
Inference: Highest Number of causalities are observed during afternoons (From 2Pm to 6PM) and later decreases.

Streets with more deceased than injured

SQL Query:

```
with street_analytics as
(select on_street_name,
number_of_cyclist_injured + number_of_motorist_injured +
number_of_pedestrians_injured + number_of_persons_injured as Total_injured,
number_of_cyclist_killed+number_of_motorist_killed+
+number_of_pedestrians_killed+ number_of_persons_killed as Total_deceased
from
(select on_street_name,
sum(number_of_cyclist_injured) as number_of_cyclist_injured,
sum(number_of_cyclist_killed) as number_of_cyclist_killed,
sum(number_of_motorist_injured) as number_of_motorist_injured,
sum(number_of_motorist_killed) as number_of_motorist_killed,
sum(number_of_pedestrians_injured) as number_of_pedestrians_injured,
sum(number_of_pedestrians_killed) as number_of_pedestrians_killed,
sum(number_of_persons_injured) as number_of_persons_injured,
sum(number_of_persons_killed) as number_of_persons_killed
from `assignment-1-368502.nypd_motor_vehicle_collisions.nypd_mv_collisions`
group by on_street_name))
select *,(Total_injured+Total_deceased) as Total_Casualties from
street_analytics where Total_deceased > Total_injured
```

Streets with more deceased than injured



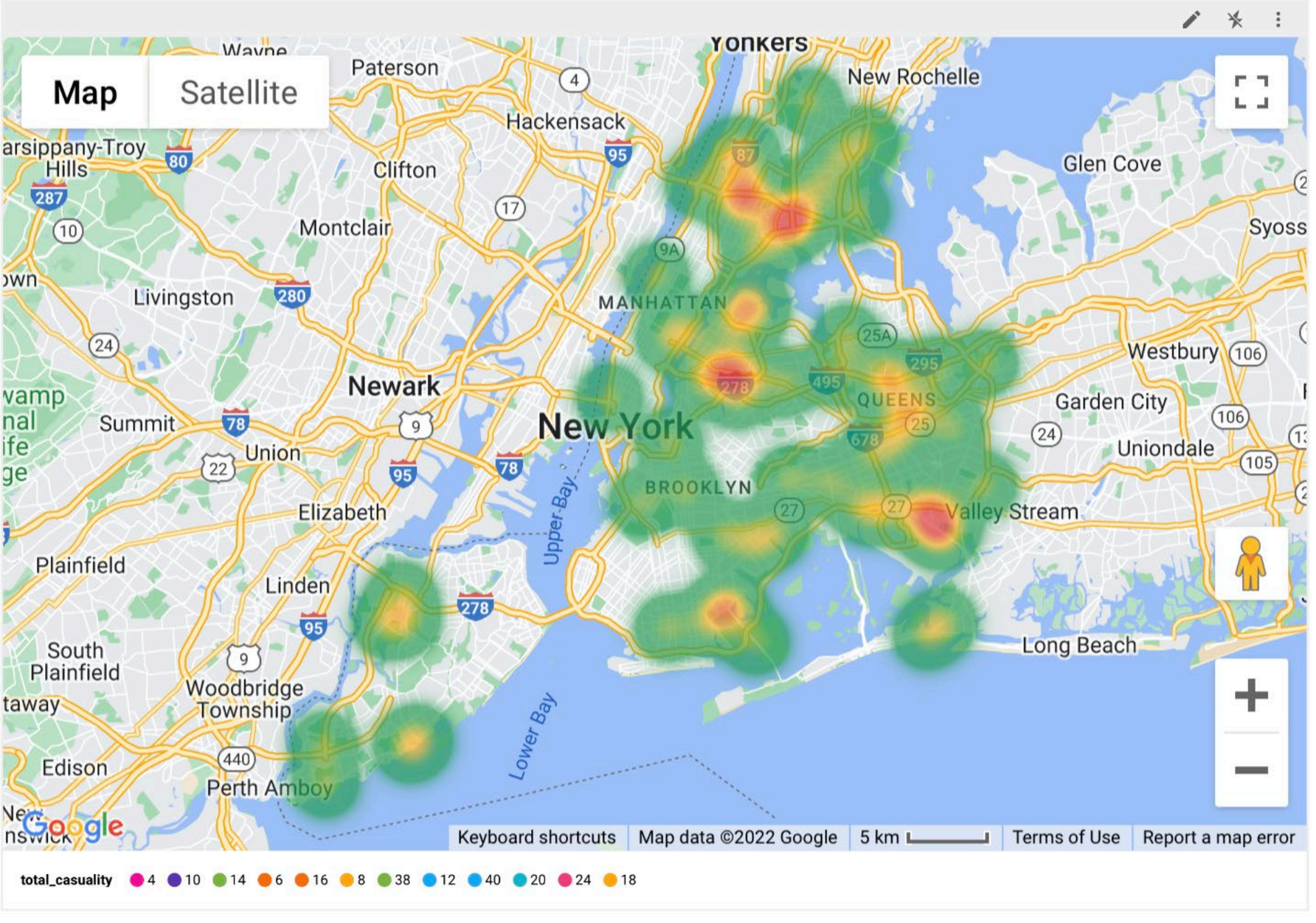
Inference: Mariana Road and borough place are the streets with more deaths than injured people.

Causality Heatmap

SQL Query:

```
SELECT
ST_GEOGPOINT(longitude, latitude) AS
geographic_location,borough,location,number_of_cyclist_injured,number_of_mo
torist_injured,number_of_pedestrians_injured,number_of_cyclist_injured +
number_of_motorist_injured +
number_of_pedestrians_injured + number_of_persons_injured +
number_of_cyclist_killed+number_of_motorist_killed+
+number_of_pedestrians_killed+ number_of_persons_killed as total_casuality
FROM `assignment-1-368502.nypd_motor_vehicle_collisions.nypd_mv_collisions`
WHERE
longitude IS NOT NULL AND latitude IS NOT NULL and number_of_persons_killed
>=2
```

Casuality Heatmap

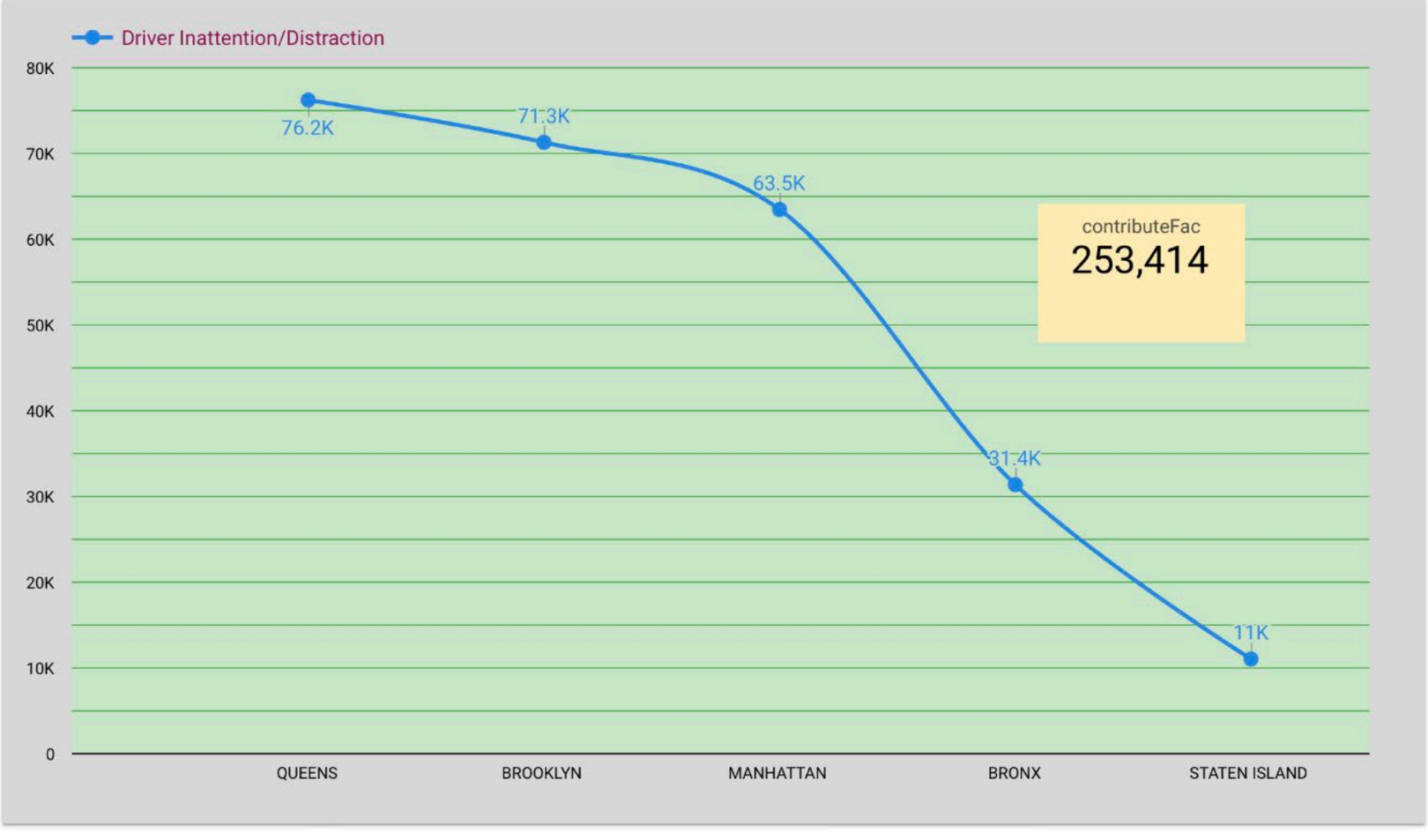


Significant contributing Factor for all counties

SQL Query:

```
SELECT borough, contributing_factor_vehicle_1, contributeFac FROM
(SELECT
borough,
contributing_factor_vehicle_1,
COUNT(contributing_factor_vehicle_1) AS `contributeFac`,
ROW_NUMBER() OVER (PARTITION BY borough order by
COUNT(contributing_factor_vehicle_1)
desc) row_num
FROM
`assignment-1-368502.nypd_motor_vehicle_collisions.nypd_mv_collisions`
WHERE borough is NOT NULL and contributing_factor_vehicle_1 <>
'Unspecified'
GROUP BY
borough,
contributing_factor_vehicle_1
)
Where row_num=1
```

Significant contributing Factor for all counties



Inference: Driver Inattention/Distraction has been identified as the significant contributing factor for Causalities across all counties which is depicted borough wise.

Top 5 streets casualty count for each borough/county

SQL Query:

```
select borough, on_street_name, casualities from (
select
borough,
on_street_name,
count(number_of_persons_injured + number_of_persons_killed) as
casualities,
ROW_NUMBER() OVER (PARTITION BY borough order by
COUNT(number_of_persons_injured + number_of_persons_killed) desc) as
row_num
from
assignment-1-368502.nypd_motor_vehicle_collisions.nypd_mv_collisions
where borough is not null and on_street_name is not null
group by
borough,
on_street_name
)
where row_num <= 5
order by borough asc, casualities desc;
```

Top 5 streets casualty count for each borough/county

	on_street_name / casualities					
borough	ATLANTIC AVENUE	NORTHERN BOULEVARD	LINDEN BOULEVARD	FLATBUSH AVENUE	QUEEN	Grand total
BROOKLYN	3.2K	-	2.3K	2.1K		10.6K
QUEENS	-	2.3K	-	-		8.3K
BRONX	-	-	-	-		5.9K
MANHATTAN	-	-	-	-		5.7K
STATEN ISLA...	-	-	-	-		3.4K
Grand total	**3.2K**	**2.3K**	**2.3K**	**2.1K**		**33.9K**

Inference: Using Hash table we have plotted Top 5 streets with high causalities across all boroughs.

CONCLUSION

1.In Winter season we have noticed low casualty rate when compared Summer or Fall which travel period, for most of the tourists visiting NYC may be attributing to higher causality rate. More staff can be pooled and detailed planning can be adopted during these months.

2.Higher causalities are reported during 2pm to 6pm. so we can have an active live monitoring system to reduce response time there by preventing deaths.

3.More patrolling needed on E/B queens and Beach 87 street, Marina Road, 40 Drive certain specific actions needed on these streets finding out more probable cause from the local patrol officers and municipal officials.

4. Driver inattention/distracted driving is noticed as significant factor across all counties. where mobile phone may be the driving factor, certain awareness campaigns needs to be run with help of local officials to address this issue.

5.Needs to be discussed with patrol officers in these streets which are having high fatalities as these streets needed immediate attention and monthly review can be planned with the officers heading these streets/counties.

Awareness Billboards used across other states show below.