***Road Safety Index calculation for a route in NYC***

*By*

*Sai Raghunath Kolusu*

*Data Science & Business Analytics, Saint Peter’s University, Jersey City, NJ*

****

**ABSTRACT**

An interactive Google map has been designed. Using this map, multiple routes between any 2 given points in NYC are plotted. Then, Accident points along the routes between 2 specific points in New York city are identified programmatically based on 2016 NYC road accidents data and plotted on the map. The Accident points is used to subset the Accident data. Accident Density at time t is calculated as function of number of accidents and mean volume. Principal Component Analysis is conducted to obtain weights for traffic volume w.r.t time and type of crash w.r.t time. A mathematical equation is designed to calculate the Road Safety Index as a function of Accident Density and weighted mean of indicators.

*Keywords: Road Safety Index, Principal Component Analysis, Weighted Mean*

**INTRODUCTION**

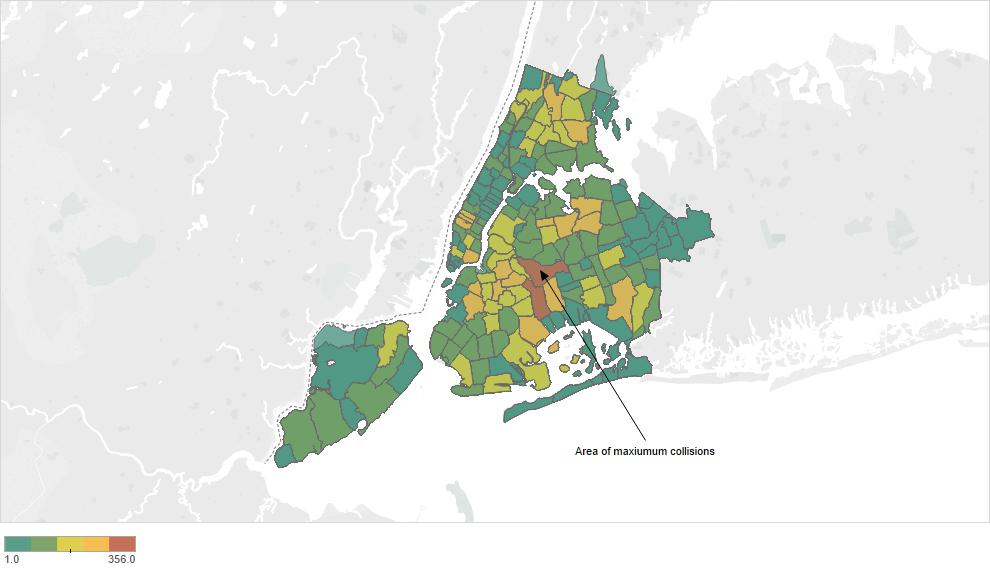
Road accidents are every Nation’s economic and social problem as proven by studies on loss of lives and properties in many countries around the world. The World Health Organization (WHO) estimated that the number of fatalities that result from road accidents per year are about 1.24 million and injuries record up to 30 million were with an average of 3000 deaths/day and 30,000 injuries/day("Global status report road safety," 2013). Moreover, its consequences have a devastating impact on economic and social conditions w.r.t health care costs of disabilities and injuries, in addition to government compensations. According to WHO, the economic costs procured from road accidents touched 518 billion USD per year in high income countries and 65 billion USD per year in low and medium income countries("Global status report road safety," 2013). USA is worst affected with road accidents with reports revealing a staggering fact; more than 30,000 are killed every year in road accidents ("List of countries by traffic-related death rate," 2014)

New York city in placed at 5th position in the top 10 cities of United States with worst traffic ("New York City Ranks #5 in the Top 10 Worst Traffic Cities," 2013)Commuters experience a 20-25 percent increase in their overall commute time every day, due to busy traffic. Also, the fatalities in road accidents in NYC accounts for 25 percent of national rate and cost an average of 4.3 billion dollars to the economy of the city(Peden et al., 2004).

A recent study tried to identify the reasons for the fatal accidents in New York city (McCarroll & Haddon, 1962) . The study compares the drivers of automobiles under crash and the non- involved drivers, who were passing through the accident sites at the same time. Some of their findings were

* 73 % of the drivers were under the DUI effect to some extent with 46% of them having the alcohol percentage more than the threshold i.e. 250 mg % and over in their blood.
* Also there were drivers fleeing in stolen cars, and they were drivers under medication effect
* The people who were fatally injured were closer to home and none of the fatally injured lived outside the city
* The drivers were mostly males and although were involved in crashes, it was men who were involved in late nights and early morning crashes.
* There was no significant difference in age among the drivers involved in road accidents.
* Finally, alcoholism was the main reason for the fatal accidents than mere social drinking.

A close look at the concentration of accidents in New York city, tells us that most accidents are concentrated in the central part of the city i.e. in Brooklyn and Queens area. Also major accidents occur around Manhattan, J.F.K airport and. Understandably, the outskirts of the city have significantly lower accidents as the traffic density over there is very less.



**Figure 1**: Accident density in NYC by Location. The central NYC has most no of accidents reported

Road Safety Index(SI) is a composed index which accident related factors are weighed to a single measure. SI has a great influence on the insurance premium. Lower the SI, higher the

insurance premium and vice-versa. Previous studies focused more on the influence of Indicators on the safety of a (Anstey, Wood, Lord, & Walker, 2005). or suggested the change of road design (Elvik, Vaa, Erke, & Sorensen, 2009). Other studies tried to develop and share best practices for road safety. Some recent examples are (Aarts & Van Schagen, 2006), providing best practices in road safety measures in general, and on best practices for the safety of powered-two-wheelers(Haworth, 2012). (Elvik, 2007) studied the best practices on black spot management and safety analysis of road networks. Moreover, previous studies include the study between the accidents and the hourly traffic flow and tried to establish a relationship between them but not taking the other factors which have greater weightage on the accidents. But a concrete metric is not developed which directly indicates the degree of safeness of a route. Although tries to assess SI as an performance index(Hermans, Brijs, Wets, & Vanhoof, 2009), it fails to take into account the external factors influencing the accident. (Yau, 2004) studied factors causing accident of 3 vehicle types in Hong Kong including goods vehicles, private vehicles, public transport and motorcycles via logistic regression models. The study found that each vehicle type significantly demonstrated different surviving chances. While the surviving chances were mainly influenced by age of vehicle, gender of driver, time of the accident and light conditions in case of private vehicles, seat belt usage played important part in case of goods vehicles. The same was the case for motorcycle type, but driver characteristics were of minimal effect. (O'donnell & Connor, 1996) researched the traffic accident data using both binary logit and binary probit models. O'Donnell and Connor found that driver’s age and speed of the vehicle played the most important role in affecting the probabilities of survival. Other factors that might have some affect include vehicle make and type of collision, blood alcohol level seating position and vehicle type. (Kockelman & Kweon, 2002)researched the traffic accident data using ordinal probit models and found out that the vehicles such as Pickup Vans and SUV’s are more dangerous than passenger cars, especially for accidents that involve a single vehicle. However, for accidents that involves two-vehicle collisions, the chances of a driver surviving are more than the occupants. (Al-Ghamdi, 2002)researched the influence of accident factors on fatal and non-fatal accidents in Saudi Arabia. The research findings showed that even a location of the accident and license status of the driver is significantly associated with a fatal accident.

Principal Component analysis is a statistical technique in which a set of variables (correlated & partially correlated) are converted to a set of linear uncorrelated variables. These are called as Principal Components whose number are more less equal to the number of original variables(citation). The importance and uses of Principal component analysis has been explained in previous studies (Jolliffe, 2002). In common terms, PCA uses vector transformation to reduce the dimensionality of large data sets. Mathematical projection of the original dataset, which originally contained a lot of variables, now can be represented in a fewer variables (Principal Components). Using these principal components, the user can plot graphs and spot various trends in the data more easily and comprehensively than by not performing PCA.(Abdi & Williams, 2010).

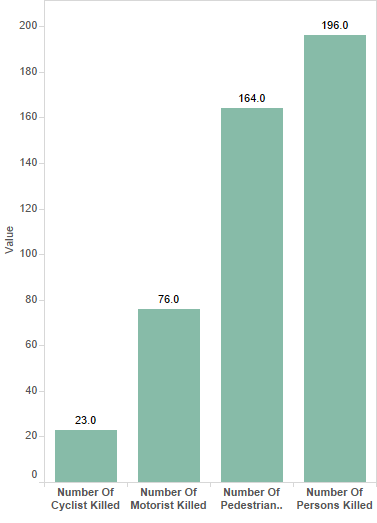
Principal component analysis can also be called as factor analysis. These group variables in the data set which are collinear to each other to form a composite variable which explains the maximum information in the model. Each composite variable contains a group of variables which have highest association within themselves. The whole idea behind doing this is to remove the variation between groups and by grouping into a composite variable, the variance between groups is accounted for. Therefore, the index is independent of dimensionality of the dataset but it rather depends upon the “statistical” dimensions of the data. According to this theory, the weights obtained after the analysis correct the overlapping information of correlated variables only. This is not a value which gives the importance of the variable in the dataset.

In this study, Principal Component Analysis helps to obtain the weightage of each component that is used to calculate the SI. In other words, PCA assigns a numeric constant value to each factor in the equation to calculate the SI. This numeric value would be directly proportional to the importance of that factor in the equation.

**DATA:**

The data considered for this study is the NYPD road crashes data for the year 2016 published online by the NYC government. The dataset contains 130,000 observations. The indicators include the location of the accident, no of fatalities, driver condition, vehicle type among others.

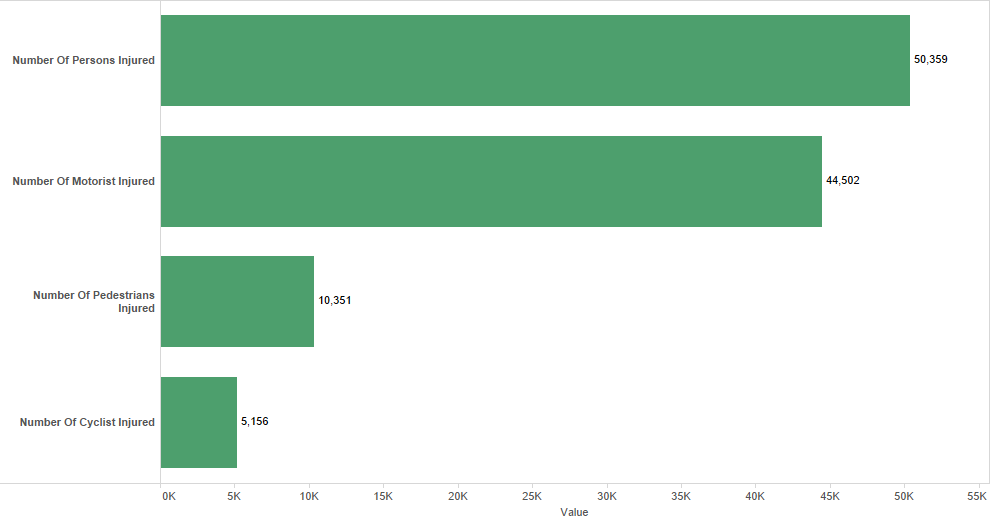
The causalities are categorized into cyclists, motorists, pedestrians and persons. Persons include the people that do qualify as either of the other three mentioned.



**Figure 2**: Number of causalities by categories. Other Persons tops the list

The graph indicates that the Other Persons suffer the most in a road accident. But the surprising thing is the number of pedestrians as causalities are on par with the persons.

However, when it comes to injuries, the motorist’s injuries occupy second place unlike in earlier scenario. Another interesting thing to be observed is that there is a sizeable amount of cyclist casualties. New York is urban city and lot of cyclists

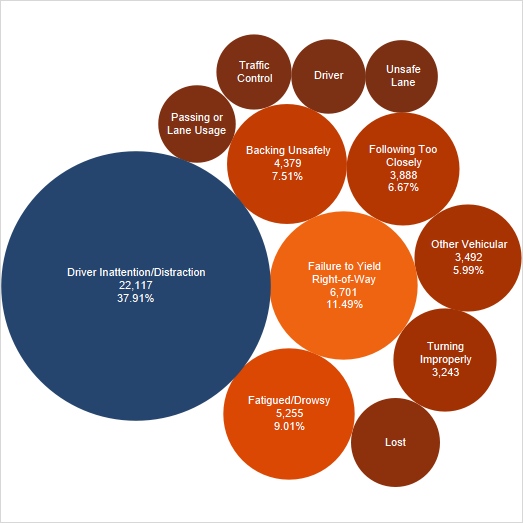


**Figure 3**: Number of injuries by categories. Interestingly, motorists occupy second place

The various indicators in the dataset are

* *Contributing Factor*:

Contributing Factors can be termed as the main event which led to the accident



**Figure 4**: Distribution of Contributing Factor. Driver Inattention has major share in it

Driver Inattention/Distraction : Driver Inattention/Distraction is the cause for most number of accidents .(Wang, Knipling, & Goodman, 1996)

To a larger extent, this distraction or inattention is due to use of new technology like cellular phones, navigation devices and loud audio systems. The NYC road crashes data for year 2016 also indicates the same i.e. Driver Inattention/Distraction is the major contributor accounting for close to 40% of the total crashes.

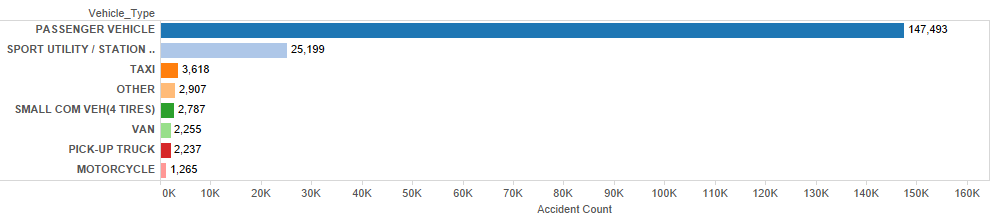
Failure to yield right of way: Failing to yield the right of way is another concern causing road crashes. It is the second major factor causing crashes in NYC (see figure) with 12 %.

According to New York traffic regulations ("Failure To Yield The Right Of Way Has Many Variations," 2015) the driver should yield right of way to pedestrians who is at a crosswalk and in the line of the vehicle under Section 4-04(D), to moving or stopped emergency vehicles(Section 1144- a) , at intersections (Section 1142) and to motorist entering a roadway( Section 1143). So, the driver violating these laws would disrupt and disturb the traffic leading to an accident

Fatigue: Driver Fatigue is not very much discussed but it is very hard to expect a driver to drive when is feeling tired or sleepy. When a driver feels sleepy or tired, the tendency to react and respond to the events on the road tends to decrease to a large extent. In some cases, the driver may actually go into a mini sleep and lose consciousness for some time. This phenomenon is generally observed during late nights and early mornings. This is the 3rd biggest contributing factor in NYC.

* *Vehicle Type:*

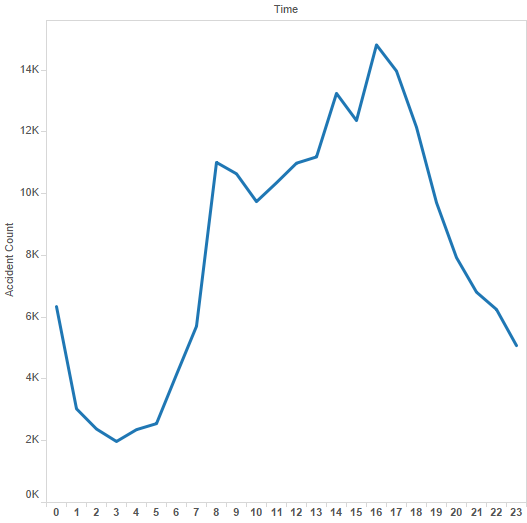
Passenger Vehicles account for nearly 80 % of the accidents and understandably so as they are the most found vehicles on the roads.(Ross, 2002) Generally , sports utility vehicles are considered more risky as it is driven by mostly young people and in this case, this fact is vindicated as it occupies the second place among others. However, there no previous case studies.



**Figure 4**: Distribution of Accidents by Vehicle Type. Passenger Vehicle tops the list by far

* *Time:*

According to (Hänecke, Tiedemann, Nachreiner, & Grzech-Šukalo, 1998), time of the day also plays an important role in road crashes. Previous study indicates that most number of accidents occur at the night time because of drunk and driving, excessive speeds and fatigue. But the data in this study, shows otherwise. It can be inferred that majority of the accidents occur during day time with the count peaking at 8 AM and 4 PM which are actually start and close of business hours.



**Figure 5**: Hourly Variation of Accident count starting from 12 am – 11 am

**METHOD:**

HTML/JavaScript has been used to design the map. Google APIs along with Mapstraction have been used to implement the function calls. Mapstraction is an open source Js library which uses the features of other mapping APIs. The advantage of Mapstraction API is that; it is easy to switch from one type of Map to another type with minimal code changes. R is another open source technology which is extensively used for data analysis on large datasets. Since JavaScript doesn’t allow to read/write data from external files by itself due to security layers, R is used as an intermediate medium to facilitate this process. Data cleaning and preparation is also done using R. An algorithm is developed Principal Component Analysis is done to assign the weights for each factor that contributes towards calculating the Road Safety Index.

*Step 1:*

The Dataset is cleaned removing all the null values and is prepared for analysis.

*Step 2*:

The frequency of each factor for a particular variable is calculated. For example, the Passenger vehicle has a frequency of 0.78 in the variable “Vehicle\_Type”, that means, it occurs 78% times in that dataset.

*Step: 3*

The volume dataset is downloaded and mean volume is calculated w.r.t time.

*Step 4:*

A Basic Map is designed using mapstraction and google apis with HTML and JavaScript.

*Step 5:*

The design is then enhanced to include options to select a start point and an end point of a route. Multiple routes between the two entered points are plotted. Direction service library of google API has been used to plot the multiple routes. Each route is indicated with different color and the total distance between that route and the estimated travel time is also displayed.

*Step 6:*

The first route is selected and the coordinates along the route are extracted.

*Step 7:*

The “location” variable is read from the dataset. These coordinates are then compared with the accident location coordinates extracted from the map, to check if any of the one coincides

with the other. The rationale in doing this, is to check if any of the accident coordinates fall in this specific route.

*Step 8:*

The matching coordinates are then plotted on the map along with the route using a heat

map. This helps us to identify accident prone locations in that route.

*Step 9:*

Using these accident location coordinates, the original dataset is subset to smaller dataset pertaining to the accident data for that particular route.

*Step 10:*

The two datasets are merged by Time & street name

*Step 11:*

The datasets are ordered w.r.t Street name and Hour.

*Step 12:*

Data at time t (7 pm) is retrieved from this dataset)

*Step 13:*

The number of accidents and the mean traffic volume of the route at 7 pm is calculated from this dataset

*Step 14:*

The accident density is calculated as

Accident Density=f (a, v) = (Accidents/Accidentsoverall) ÷ (Mean\_Volt / Mean\_Voloverall)

*Step 15:*

The Contributing Factor and Vehicle Type variables are assigned their individual numerical weights based on their frequency in the original and subset dataset.

*Step 16:*

The Variables Mean Traffic Volume, Contributing Factor and Vehicle Type are then normalized to a standard scale

*Step 17:*

Principal Component Analysis is done on the 3 variables and since we don’t have a target variable to build a model, each component is taken into consideration and the resultant value is calculated.

*Step 18:*

Here we consider the absolute value of Eigen values because the sign indicates their orientation in the space.

*Step 19:*

The resultant sum product of the component and variable value is calculated

*Step 20:* The Road Safety Index of ith value is calculated as

RSIi = f (a, v, c) = Accident Density + ƛ1iv1i+ ƛ2iv2i+ ƛ2iv2i

Where ƛni is the principal component of ith observation of nth variable

Vni is the value of ith observation in nth variable

**RESULTS:**

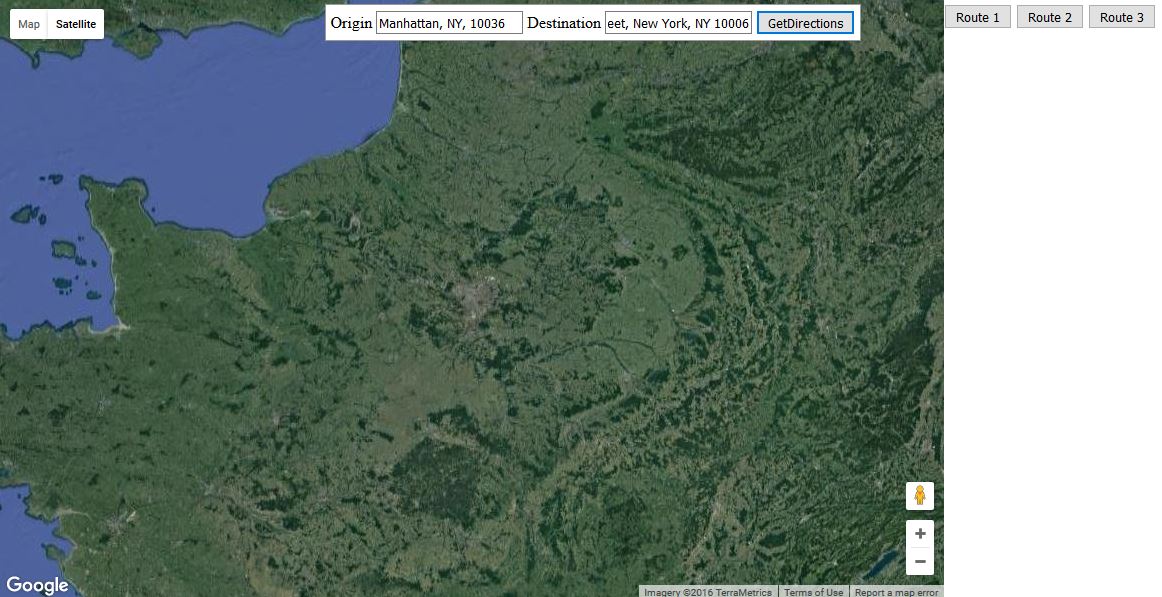
The start and end point are entered as Time Square and World Trade Center in New York

Start: **Manhattan, NY, 10036**

End: **One World Trade Center, 285 Fulton Street, New York, NY 10006**

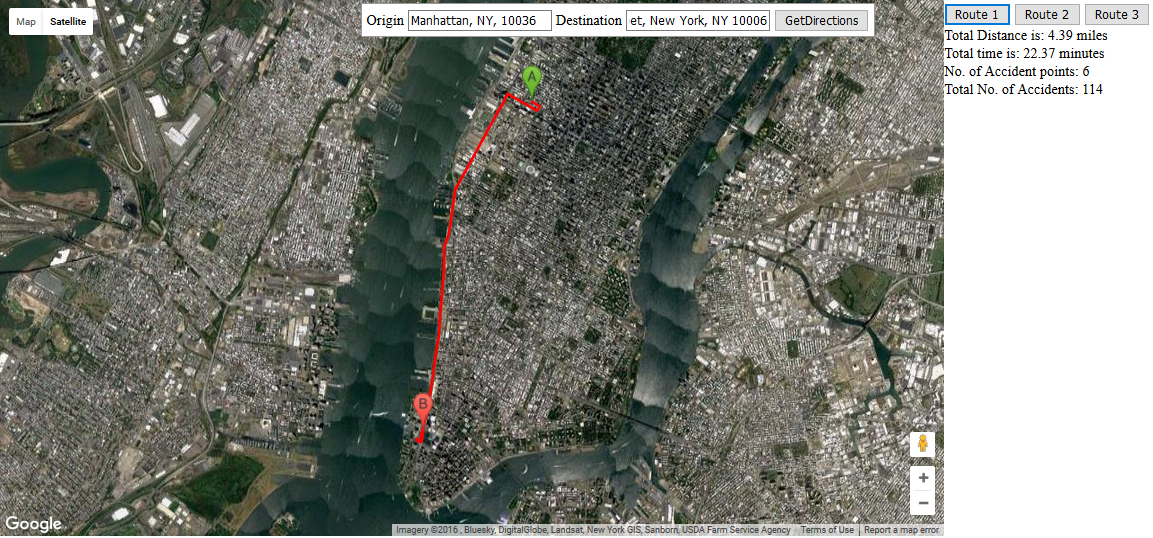
Time: **7 pm**

After clicking on ‘Get Directions’ button, we get different buttons for each route



**Figure 6: shows buttons for each route. 3 buttons displayed, so there are 3 routes.**

***Route 1:***



**Figure 7**: Total Distance, Total Time taken and Location of accident points along the Route 1

|  |  |
| --- | --- |
| Total Distance Taken | 4.39 miles |
| Total Time Taken | 22.37 minutes |

**Table 1**: Total Distance and Total Time for Route 1

|  |  |
| --- | --- |
| **Accident points** | **No. of Accidents** |
| 40.7551, -74.0025 | 9 |
| 40.7577, -74.0006 | 5 |
| 40.7597, -73.9917 | 18 |
| 40.7608, -73.9983 | 53 |
| 40.7615, -73.9978 | 23 |
| Total | 108 |

**Table 2**: Accident points and the number of accidents occurred at those point

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Time** | **Location** | **Street\_Name** | **Contributing\_Factor** | **Vehicle\_Type** |
| 10/5/2016 | 0:10 | 40.7608, -73.9983 | 11 AVENUE | Turning Improperly | Taxi |
| 2/25/2016 | 7:33 | 40.7615, -73.9978 | 11 AVENUE | Other Vehicular | Sport Utility / Station Wagon |
| 1/12/2016 | 10:00 | 40.7551, -74.0025 | 11 AVENUE | Backing Unsafely | Pick-Up Truck |
| 2/12/2016 | 10:02 | 40.7577, -74.0006 | 11 AVENUE | Failure to Yield Right-of-Way | Large Com Veh (6 Or More Tires) |
| 10/6/2016 | 13:10 | 40.7597, -73.9917 | 9 AVENUE | Driver Inattention/Distraction | Sport Utility / Station Wagon |

**Table 3:** A sample of Accident data subset based on accident points obtained

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Street Name | 0:00 | 1:00 | 2:00 | 3:00 | 4:00 |
| 11 AVENUE | 649.00 | 495.06 | 394.00 | 342.00 | 293.88 |
| 9 AVENUE | 1426.88 | 1152.56 | 965.25 | 841 | 691.44 |

**Table 4:** Sample of volume of traffic in the streets at each hour where accidents occurred

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Location** | **Street\_Name** | **Contributing\_Factor\_Vehicle** | **Vehicle\_Type** | **Hour** | **Mean Traffic Vol** | **Total Injured** |
| 40.7608, -73.9983 | 11 Avenue | Turning Improperly | Taxi | 0 | 649 | 0 |
| 40.7615, -73.9978 | 11 Avenue | Unspecified | Passenger Vehicle | 2 | 394 | 0 |
| 40.7551, -74.0025 | 11 Avenue | Backing Unsafely | Pick-Up Truck | 10 | 914.88 | 2 |
| 40.7577, -74.0006 | 11 Avenue | Failure to Yield Right-Of-Way | Large Com Veh | 10 | 914.88 | 0 |
| 40.7597, -73.9917 | 9 Avenue | Unspecified | Passenger Vehicle | 0 | 1426.88 | 0 |

**Table 5:** Sample of resultant dataset after merging accident dataset with mean volume w.r.t time

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Overall** | **At time t** | **Density(xt/xoverall)** | **Accident Density at time t (Acct /Mean\_Volt)** |
| Accidents | 58 | 4 | 0.689 | 0.07 |
| Mean\_Traff\_Vol | 979.56 | 962.65 | 0.982 |

**Table 6:** Accident Density along Route 1 at 7 pm

|  |  |  |
| --- | --- | --- |
| **Mean\_Traffic\_Vol** | **Contributing\_Factor** | **Vehicle\_Type** |
| 0.9599584 | 0.1 | 0.04 |
| 0.9599584 | 0.04 | 0.04 |
| 0.9599584 | 0.04 | 0.04 |
| 0.9599584 | 0.2 | 0.07 |

**Table 7**: Normalized values of the variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Comp.1** | **Comp.2** | **Comp.3** |
| Mean\_Vol | 0 | 0 | 1 |
| Cont\_Vehicle\_Frequency | 0.983279 | 0.1821059 | 0 |
| Veh\_Type\_Frequency | 0.182106 | 0.9832789 | 0 |

**Table 8:** PCA results on the 3 variables

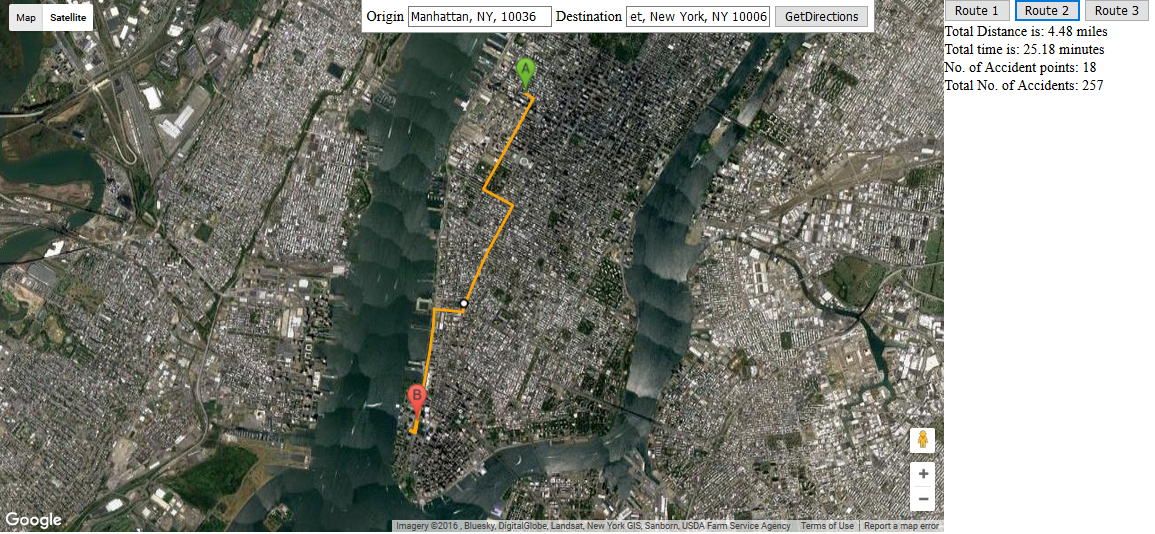
|  |
| --- |
| **RSI** |
| 0.172 |
| 0.134 |
| 1.02 |

**Table 9:** Mean RSI values for 3 components

Considering the minimum of the 3 values

***RSI for Route 1 = 0.172***

***Route 2:***



**Figure 8**: Total Distance, Total Time taken and Location of accident points along the Route 2

|  |  |
| --- | --- |
| Total Distance Taken | 4.30 miles |
| Total Time Taken | 22.52 minutes |

**Table 10:** Total Distance and Total Time for Route 2

|  |  |
| --- | --- |
| **Accident points** | **No. of Accidents** |
| 40.7429, -74.004 | 25 |
| 40.7441, -74.0068 | 9 |
| 40.7452, -74.0023 | 3 |
| 40.7527, -73.9968 | 12 |
| 40.7528, -73.9967 | 1 |
| 40.7541, -73.9958 | 4 |
| 40.7547, -73.9954 | 17 |
| 40.7553, -73.9949 | 21 |
| 40.7559, -73.9945 | 16 |
| 40.7565, -73.994 | 24 |
| 40.7574, -73.9934 | 1 |
| 40.7584, -73.9926 | 44 |
| 40.7597, -73.9917 | 18 |
| **Total** | 195 |

**Table 11**: Accident points and the number of accidents occurred at those point

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Time** | **Location** | **Street\_Name** | **Cont\_Factor\_Vehicle** | **Vehicle\_Type** |
| 3/6/2016 | 0:17 | 40.7429, -74.004 | 9 AVENUE | Outside Car Distraction | PASSENGER VEHICLE |
| 11/1/2016 | 11:11 | 40.7528, -73.9967 |  | Unspecified | PASSENGER VEHICLE |
| 2/25/2016 | 22:40 | 40.7597, -73.9917 | WEST 44 STREET | Backing Unsafely | TAXI |
| 1/9/2016 | 0:50 | 40.7441, -74.0068 | 10 AVENUE | Unspecified | OTHER |
| 3/24/2016 | 13:30 | 40.7574, -73.9934 |  |  | PASSENGER VEHICLE |
| 7/1/2016 | 0:01 | 40.7452, -74.0023 | WEST 21 STREET | Unspecified | PASSENGER VEHICLE |
| 2/24/2016 | 10:50 | 40.7527, -73.9968 | 9 AVENUE | Lost Consciousness | VAN |
| 7/4/2016 | 14:50 | 40.7541, -73.9958 | WEST 35 STREET | Unspecified | PASSENGER VEHICLE |
| 1/27/2016 | 11:40 | 40.7547, -73.9954 | 9 AVENUE | Driver Inattention/Distraction | OTHER |
| 4/16/2016 | 0:30 | 40.7553, -73.9949 | 9 AVENUE | Turning Improperly | PASSENGER VEHICLE |
| 6/23/2016 | 0:15 | 40.7559, -73.9945 | 9 AVENUE | Unspecified | PASSENGER VEHICLE |
| 7/12/2016 | 10:00 | 40.7565, -73.994 | 9 AVENUE | Unspecified | PASSENGER VEHICLE |
| 4/20/2016 | 0:10 | 40.7584, -73.9926 | 9 AVENUE | Turning Improperly | PASSENGER VEHICLE |

**Table 12**: A sample of Accident data subset based on accident points obtained

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Location** | **Street\_Name** | **Cont\_Factor\_Vehicle** | **Vehicle\_Type** | **Hour** | **Mean\_Vol** | **Total Injured** | **Total Killed** |
| 40.7441, -74.0068 | 10 AVENUE | Unspecified | PASSENGER VEHICLE | 19 | 798.97 | 4 | 0 |
| 40.7559, -73.9945 | 9 AVENUE | Unspecified | PASSENGER VEHICLE | 0 | 1426.88 | 0 | 0 |
| 40.7429, -74.004 | 9 AVENUE | Turning Improperly | SPORT UTILITY / STATION WAGON | 1 | 1152.56 | 0 | 0 |
| 40.7584, -73.9926 | 9 AVENUE | Unspecified | PASSENGER VEHICLE | 8 | 1582.12 | 0 | 0 |
| 40.7527, -73.9968 | 9 AVENUE | Lost Consciousness | VAN | 10 | 1764.75 | 0 | 0 |
| 40.7547, -73.9954 | 9 AVENUE | Driver Inattention/Distraction | OTHER | 11 | 1766 | 0 | 0 |
| 40.7452, -74.0023 | 9 AVENUE | Unsafe Lane Changing | PASSENGER VEHICLE | 12 | 1755.62 | 0 | 0 |
| 40.7565, -73.994 | WEST 39 STREET | Failure to Yield Right-of-Way | PASSENGER VEHICLE | 16 | 614.25 | 0 | 0 |
| 40.7553, -73.9949 | 9 AVENUE |  | AMBULANCE | 11 | 1766 | 0 | 0 |
| 40.7584, -73.9926 | 9 AVENUE | Traffic Control Disregarded | UNKNOWN | 22 | 1566 | 0 | 0 |
| 40.7541, -73.9958 | 9 AVENUE | Driver Inattention/Distraction | SPORT UTILITY / STATION WAGON | 19 | 1755.56 | 0 | 0 |

**Table 13**: Sample of resultant dataset after merging accident dataset with mean volume w.r.t time

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Overall** | **At time t** | **Density(xt/xoverall)** | **Accident Density at time t (Acct /Mean\_Volt)** |
| Accidents | 113 | 13 | 0.115 | 0.12 |
| Mean\_Traff\_Vol | 1429.58 | 1345.16 | 0.94 |

**Table 14**: Calculating Accident Density along Route 2 at 7 pm

|  |  |  |
| --- | --- | --- |
| **Mean\_Traffic\_Vol** | **Contributing\_Factor** | **Vehicle\_Type** |
| 0.949972 | 0.11 | 0.08 |
| 0.40508 | 0.18 | 0.08 |
| 0.949972 | 0.08 | 0.08 |
| 0.949972 | 0.12 | 0.08 |
| 0.300219 | 0.5 | 0.08 |
| 0.40508 | 0.05 | 0.08 |
| 0.949972 | 0.05 | 0.08 |
| 0.949972 | 0.05 | 0.08 |
| 0.300219 | 0.05 | 0.08 |
| 0.300219 | 0.05 | 0.08 |
| 0.949972 | 0.11 | 0.12 |
| 0.949972 | 0.18 | 0.12 |
| 0.949972 | 0.06 | 0.08 |

**Table 15**: Normalized values of the variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Comp.1** | **Comp.2** | **Comp.3** |
| Mean\_Vol | 0.989535 | 0.14296791 | 0.019508 |
| Cont\_Vehicle\_Frequency | 0.143445 | 0.98932315 | 0.025754 |
| Veh\_Type\_Frequency | 0.015617 | 0.02828231 | 0.999478 |

**Table 16**: PCA results on the 3 variables

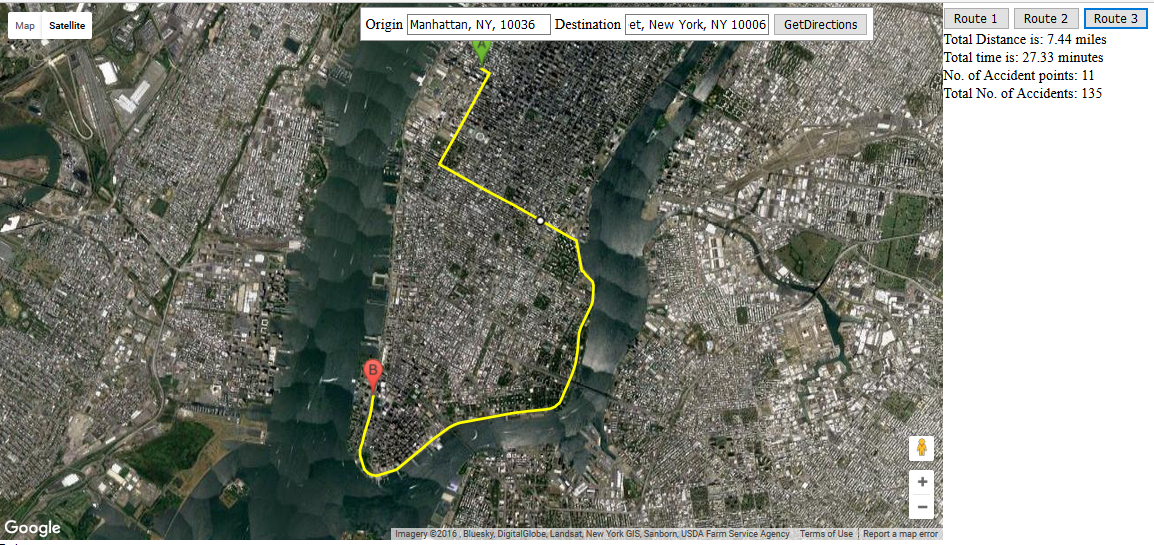
|  |
| --- |
| **RSI** |
| 0.848 |
| 0.346 |
| 0.223 |

**Table 17**: Mean RSI values for 3 components

Considering the minimum of the 3 values

***RSI for Route 2 = 0.223***

***Route 3:***



**Figure 9**: Total Distance, Total Time taken and Location of accident points along the Route 3

|  |  |
| --- | --- |
| Total Distance Taken | 7.33 miles |
| Total Time Taken | 25.12 minutes |

**Table 18**: Total Distance and Total Time for Route 2

|  |  |
| --- | --- |
| **Accident points** | **No. of Accidents** |
| 40.7052, -74.0166 | 1 |
| 40.7094, -73.9943 | 3 |
| 40.7106, -73.9847 | 5 |
| 40.7153, -73.9763 | 9 |
| 40.7295, -73.9721 | 8 |
| 40.7475, -73.9688 | 4 |
| 40.7498, -73.9722 | 37 |
| 40.7528, -73.9793 | 13 |
| 40.7584, -73.9926 | 44 |
| 40.7597, -73.9917 | 18 |
| Total | 142 |

**Table 19:** Accident points and the number of accidents occurred at those point along Route 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Time** | **Location** | **Street\_Name** | **Cont\_Factor\_Vehicle** | **Vehicle\_Type** |
| 10/17/2016 | 0:10 | 40.7498, -73.9722 | 2 AVENUE | Other Vehicular | PICK-UP TRUCK |
| 4/20/2016 | 0:10 | 40.7584, -73.9926 | 9 AVENUE | Turning Improperly | PASSENGER VEHICLE |
| 7/6/2016 | 0:20 | 40.7597, -73.9917 | 9 AVENUE | Unspecified | PASSENGER VEHICLE |
| 3/22/2016 | 8:46 | 40.7295, -73.9721 | EAST 16 STREET | Driver Inattention/Distraction | PASSENGER VEHICLE |
| 5/25/2016 | 22:50 | 40.7475, -73.9688 | EAST 41 STREET | Unspecified | PASSENGER VEHICLE |
| 7/5/2016 | 11:40 | 40.7052, -74.0166 | WEST STREET | Unspecified | PASSENGER VEHICLE |
| 3/31/2016 | 16:20 | 40.7094, -73.9943 | EAST 16 STREET | Driver Inattention/Distraction | PASSENGER VEHICLE |
| 2/3/2016 | 18:30 | 40.7106, -73.9847 | WEST STREET | Turning Improperly | SPORT UTILITY / STATION WAGON |
| 2/18/2016 | 18:30 | 40.7153, -73.9763 | 9 AVENUE | Fatigued/Drowsy | PASSENGER VEHICLE |
| 3/4/2016 | 16:15 | 40.7528, -73.9793 | MADISON AVENUE | Lost Consciousness | BUS |

**Table 20:** A sample of Accident data subset based on accident points obtained

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Location** | **Street\_Name** | **Cont\_Factor\_Vehicle** | **Vehicle\_Type** | **Hour** | **Mean\_Vol** | **Total Injured** | **Total Killed** |
| 40.7153, -73.9763 | GRAND STREET | Lost Consciousness | PASSENGER VEHICLE | 1 | 261 | 0 | 0 |
| 40.7498, -73.9722 | 2 AVENUE | Other Vehicular | PASSENGER VEHICLE | 8 | 2586.79 | 0 | 0 |
| 40.7528, -73.9793 | MADISON AVENUE | Unspecified | PASSENGER VEHICLE | 0 | 742.69 | 4 | 0 |
| 40.7584, -73.9926 | 9 AVENUE | Unspecified | PASSENGER VEHICLE | 13 | 1737.31 | 0 | 0 |
| 40.7597, -73.9917 | 9 AVENUE | Unspecified | PASSENGER VEHICLE | 0 | 1426.88 | 0 | 0 |

**Table 21**: Sample of resultant dataset after merging accident dataset with mean volume w.r.t time

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Contributing\_Factor** | **Vehicle\_Type** | **Hour** | **Mean\_Vol** | **Total\_Injured** | **Total\_Killed** |
| Pedestrian/Bicyclist/Other Pedestrian Error/Confusion | PASSENGER VEHICLE | 19 | 2816.06 | 2 | 0 |
| Unspecified | PASSENGER VEHICLE | 19 | 2816.06 | 2 | 0 |
| Unspecified | PASSENGER VEHICLE | 19 | 1755.56 | 0 | 0 |
| Failure to Yield Right-of-Way | SPORT UTILITY / STATION WAGON | 19 | 1755.56 | 2 | 0 |
| Fatigued/Drowsy | PASSENGER VEHICLE | 19 | 1755.56 | 0 | 0 |
| Driver Inattention/Distraction | PASSENGER VEHICLE | 19 | 1085.81 | 0 | 0 |

**Table 22:** Accident data along Route 3 at 7 pm

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Overall** | **At time t** | **Density(xt/xoverall)** | **Accident Density at time t (Acct /Mean\_Volt)** |
| Accidents | 46 | 6 | 0.13 | 0.11 |
| Mean\_Traff\_Vol | 1673.71 | 1997.44 | 1.193 |

**Table 23:** Calculating Accident Density along Route 3 at 7 pm

|  |  |  |
| --- | --- | --- |
| **Mean\_Traffic\_Vol** | **Contributing\_Factor** | **Vehicle\_Type** |
| 0.3555671 | 0.05 | 0.05 |
| 0.5933994 | 0.17 | 0.05 |
| 0.9699893 | 1 | 0.05 |
| 0.9699893 | 0.03 | 0.05 |
| 0.9699829 | 0.03 | 0.05 |
| 0.9699829 | 0.2 | 0.08 |

**Table 24**: Normalized values of the variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Comp.1** | **Comp.2** | **Comp.3** |
| Mean\_Vol | 0.344863 | 0.93851251 | 0.01624 |
| Cont\_Vehicle\_Frequency | 0.938653 | 0.34482093 | 0.005415 |
| Veh\_Type\_Frequency | 0.000518 | 0.01711112 | 0.999853 |

**Table 25:** PCA results on the 3 variables

|  |
| --- |
| **RSI** |
| 0.619 |
| 0.951 |
| 0.179 |

**Table 26:** Mean RSI values for 3 components

Considering the minimum of the 3

**RSI for Route 3 = 0.179**

***Summary Table***

Below is the summary of the results for Road Safety Index for routes along Time Square to New York at 7 PM

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Route no** | **Total Distance** | **Total Time** | **No of Accident points** | **Total no of Accidents** | **Accident Density** | **Weighted Components Mean** | **RSI** |
| Route 1 | 4.39 | 22.37 | 3 | 6 | 0.07 | 0.1019 | 0.172 |
| Route 2 | 4.30 | 22.52 | 6 | 13 | 0.12 | 0.1027 | 0.223 |
| Route 3 | 7.33 | 25.12 | 3 | 6 | 0.11 |  |  |

**Table 27:** Summary values for all the 3 Routes

**DISCUSSIONS**

*Route 1:*

The Route is 4.39 miles in distance and takes 22.52 minutes to travel and is the most travelled route. There are total of 5 accident points and 108 accidents that took place in this year in this route. The Accidents mostly happened in 11th Avenue and 9th Avenue. The Contributing Factor majorly was ‘Turning Improperly’ and Vehicle Type was mostly ‘Passenger Vehicle’. The Mean Traffic Volume is 979.56 for this route. The total number of people killed were 2 and 10 people got injured. These low number might suggest that this route is safer

We considered the time to be 7 Pm for analysis. At approximately 7 PM, there were total of 6 accidents that occurred in 3 locations in this year. The Mean Volume of Traffic at this time is 962.65. Consequently, the Accident Density is calculated as 0.07.

The component values obtained are mostly 0 indicating the components are independently oriented from each other. The multiple RSI values obtained have significant variation with each other. The RSI of the route is obtained a 0.172

*Route 2:*

This Route is shorter than the first one (4.30) miles in distance but takes slightly longer (22.52 minutes) to travel which indicates that this might a crowded. The mean traffic volume in this route is 1429.58 which adds weight to this argument. Even the number of accident points in this route (11) are double than the route 1 despite being the shorter one. Also the number of accidents occurred in this route are significantly higher (195). None of the people were killed but 130 people got injured. This indicates this accident prone route.

At 7 PM, there were total of 13 accidents that occurred in 6 locations along that route in this year. The Mean Volume of Traffic at this time is 962.65. The Accident Density is obtained as 0.12.

The Eigen values are non-zeroes unlike in Route 1, but are negative in value which indicate they are oriented in the opposite direction. The RSI of this route is obtained a 0.223 which is higher to that of Route 1.

*Route 3:*

This Route is longest one (7.33) miles among all the three and understandably takes longer time to travel (25.12 minutes) to travel. The mean traffic volume in this route is 1673.71 which is understandable considering the distance. The number of accident points in this route (10) is also not unacceptable in this context since it is lesser than Route 2. Also the number of accidents occurred is also lesser than Route 2 (142). Similar trend as Route 2 can be observed here where none of the people were killed but 105 people got injured.

At **7** PM, there were total of 6 accidents that occurred in 3 locations along that route in this year which is reasonable. The Mean Volume of Traffic at this time is 1997.44 which is higher than the average value. This indicates that this route is crowded than average hour at 7 pm. The Accident Density is obtained as 0.11.

The RSI of this route is obtained a 0.179 which is closer to Route 1.

**CONCLUSION**

* An Interactive Map has been designed in which different routes between 2 points can be plotted
* This design would be used to get the accident points along a route between the specified points
* Using this accident points the accident data is subsetted.
* Mean Volume at the time t along the route is calculated from volume dataset
* Accident density and the weights of accident influencing variables at that time is calculated from the subset accident dataset and volume dataset
* RSI is calculated as a function of the accident density and the weights
* The RSI of the Route 1 is the smallest among the three and therefore considered to be safest. The other data and results discussed also vindicates this result

**REFERENCES**

1. Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention, 38*(2), 215-224.
2. Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433-459.
3. Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention, 34*(6), 729-741.
4. Anstey, K. J., Wood, J., Lord, S., & Walker, J. G. (2005). Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical psychology review, 25*(1), 45-65.
5. Elvik, R. (2007). *State-of-the-art approaches to road accident black spot management and safety analysis of road networks*: Transportøkonomisk institutt.
6. Elvik, R., Vaa, T., Erke, A., & Sorensen, M. (2009). *The handbook of road safety measures*: Emerald Group Publishing.
7. Failure To Yield The Right Of Way Has Many Variations. (2015, September 12). *Weiss & Associates*. Retrieved from <http://nytrafficticket.com/failure-to-yield-the-right-of-way-has-many-variations/>
8. Global status report road safety. (2013). *World Health Organization*. Retrieved from <http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/>
9. Hänecke, K., Tiedemann, S., Nachreiner, F., & Grzech-Šukalo, H. (1998). Accident risk as a function of hour at work and time of day as determined from accident data and exposure models for the German working population. *Scandinavian journal of work, environment & health*, 43-48.
10. Haworth, N. (2012). Powered two wheelers in a changing world—Challenges and opportunities. *Accident Analysis & Prevention, 44*(1), 12-18.
11. Hermans, E., Brijs, T., Wets, G., & Vanhoof, K. (2009). Benchmarking road safety: lessons to learn from a data envelopment analysis. *Accident Analysis & Prevention, 41*(1), 174-182.
12. Jolliffe, I. (2002). *Principal component analysis*: Wiley Online Library.
13. Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis & Prevention, 34*(3), 313-321.
14. List of countries by traffic-related death rate. (2014). *Wikipedia*. Retrieved from <https://en.wikipedia.org/wiki/List_of_countries_by_traffic-related_death_rate>
15. McCarroll, J. R., & Haddon, W. (1962). A controlled study of fatal automobile accidents in New York City. *Journal of chronic diseases, 15*(8), 811-826.
16. New York City Ranks #5 in the Top 10 Worst Traffic Cities. (2013). *INTRIX*. Retrieved from <http://inrix.com/new-york-city-ranks-5-in-the-top-10-worst-traffic-cities/>
17. O'donnell, C., & Connor, D. (1996). Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. *Accident Analysis & Prevention, 28*(6), 739-753.
18. Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Mathers, C. D. (2004). World report on road traffic injury prevention: World Health Organization Geneva.
19. Ross, M. (2002). An analysis of traffic deaths by vehicle type and model.
20. Wang, J.-S., Knipling, R. R., & Goodman, M. J. (1996). *The role of driver inattention in crashes: New statistics from the 1995 Crashworthiness Data System.* Paper presented at the 40th annual proceedings of the Association for the Advancement of Automotive Medicine.
21. Yau, K. K. (2004). Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accident Analysis & Prevention, 36*(3), 333-340.