**Analyzing Dataset of Traffic Crashes in Las Vegas**

**Final Project**

Sha Liu

Luddy School, Indiana University

SQL AND NOSQL

July 28, 2020

# ­­­­­­

**Part One**

The dataset is called Traffic Crashes and I pulled it from Data World. City of Las Vegas published this dataset on its city website (2016). The dataset contains the traffic crashes occurred in Las Vegas in 2016. The size of the original csv file is 5.17 MB. It has 61 columns and 12,720 rows.

In order to make it easy to import the file into Neo4j, I have removed some variables to make the dataset more concise. I also only selected the top 1000 rows so it will not slow down the importing process. The final dataset for importing contains 19 columns and 1000 records. I have reviewed data and identified the Neo4j schema. Please see my figure schema from Arrows as follows:

A close up of a map

Description automatically generated

In order to analyze the traffic crashes, I have used the most seven relevant variables as my nodes and they are Crash, FirstDriver, SecondDriver, FirstVehicle, SecondVehicle, Environment and Agency. The relationships are as follows: Crash is caused by FirstDriver and SecondDriver; Crash occurs with FirstVehicle and SecondVehicle; Crash occurs under specific Environment; Crash is reported to Agency.

The unique constraint of Crash is FileID. The unique constraint of FirstDriver and SecondDriver is age. The unique constraint for FirstVehicle and SecondVehicle is vehicle type. The unique constraint for Environment is distance. The unique constraint for Agency is agency name.

The properties of Crash are FileID(integer), CrashType(string), CrashServerity(string), PrimaryStreet(string), County(string), year(date), month(date), day(date), NumberOfInjured(integer), and NumberOfFatalities(integer). The properties of FirstDriver and SecondDriver are Age(integer) and Condition(string). The properties of FirstVehicle and SecondVehicle are Type(string), Direction(string), Lane(integer), Action(string) and Causation(string). The properties of Environment are Distance(integer), Weather(string), Lighting(string), RoadCondition(string) and PedestrainInvolved(boolean). The property of Agency is Name(string).

**Part Two**

Run cypher code in neo4j

**Part Three**

***Data Structure***

On part two, I ran the cypher codes in neo4j and created 7 nodes and 4 types of relationships. Figure 1 shows the number of each node label. We can see that the 7 nodes are Crash, FirstDriver, SecondDriver, FirstVehicle, SecondVehicle, Environment and Agency. The relationships are as follows: Crash is caused by FirstDriver and SecondDriver; Crash occurs with FirstVehicle and SecondVehicle; Crash occurs under specific Environment; Crash is reported to Agency. Please see Figure 2 for value count of each relationship type.

A screenshot of a cell phone

Description automatically generated

Figure 1: Value count of each node label

A screenshot of a cell phone

Description automatically generated

Figure 2: Value count of each relationship

In order to visualize the graph structure, I have run the schema visualization in Figure 3 below.

A screenshot of a cell phone

Description automatically generated

Figure 3: Schema Visualization

This type of graph is simple graph as node pairs only have one relationship between them. The graph is in scale-free network structure. Also, the graph is connected, unweighted, directed, acyclic and in rooted tree shape.

***Data Analysis***

In order to understand the damage caused by each crash type, I need to know the number of people injured and died in each type of car crash. I have run the following query in Figure 4. From the query results, angle crash type is the most common type of crash will cause people get injured then followed by rear-end crash. I am actually surprised to find out that the highest number of fatalities is caused by non-collision crash.

A screenshot of a cell phone

Description automatically generated

Figure 4: Number of Injured and Fatalities by Crash Type

Usually when we think of traffic crashes, we think of car collision. However, the query results indicate there are 210 non-collision crashes out of 1000 crashes. 21% is a high percentage for me as I thought most of the crashes happen with collision. Please see Figure 5 for query results.

A screenshot of a cell phone

Description automatically generated

Figure 5: Number of Non-Collision Crashes

In order to understand what the average age of the drivers in the traffic crash is, I ran the query shown on Figure 6, the average group for the first driver and second driver involved in the traffic crash are 50.2 and 49.6 years old respectively. Please see Figure 6 for query results.

A screenshot of a cell phone

Description automatically generated

Figure 6: Average age of the drivers

To understand what the drivers weather condition are when the crashes occurred. I analyzed how many of crashes occurs under raining. The query shows that there are 224 out 1000 crashes are occurred in raining environment, which is reasonable to me. Please see Figure 7 for the query results.

A screenshot of a cell phone

Description automatically generated

Figure 7: Number of Crashes Under Raining

I also interested to know: How the traffic crashes rates change by vehicle type? I have run query with the first vehicle and second vehicle shown on Figure 8 and 9. Both queries show that a 4-door sedan has the highest number of crashes and number of injuries among all vehicle types. I guess this make sense to me as most of the car owner population is 4-door sedan.

A picture containing computer

Description automatically generated

Figure 8: Crash by Vehicle Type per First Vehicle

A screen shot of a social media post

Description automatically generated

Figure 9: Crash by Vehicle Type per Second Vehicle

Please note that all the query results above only resulted from the first 1000 records of the original dataset for assignment purpose. If I ran the whole City of Las Vegas Traffic Crashes dataset (Web 2016), the query results might differ.

Citation:

City of Las Vegas. (2016). Traffic Crashes [Data file]. Retrieved from  <https://opendata.lasvegasnevada.gov/d/w2h7-67ye>