

Data Warehousing And Data Mining

Major Project - Phase 1

INSTRUCTOR: Dr. Kamalakar Karlapalem

- Gunjan Karamchandani (20161086), Neal Karpe (20161159)

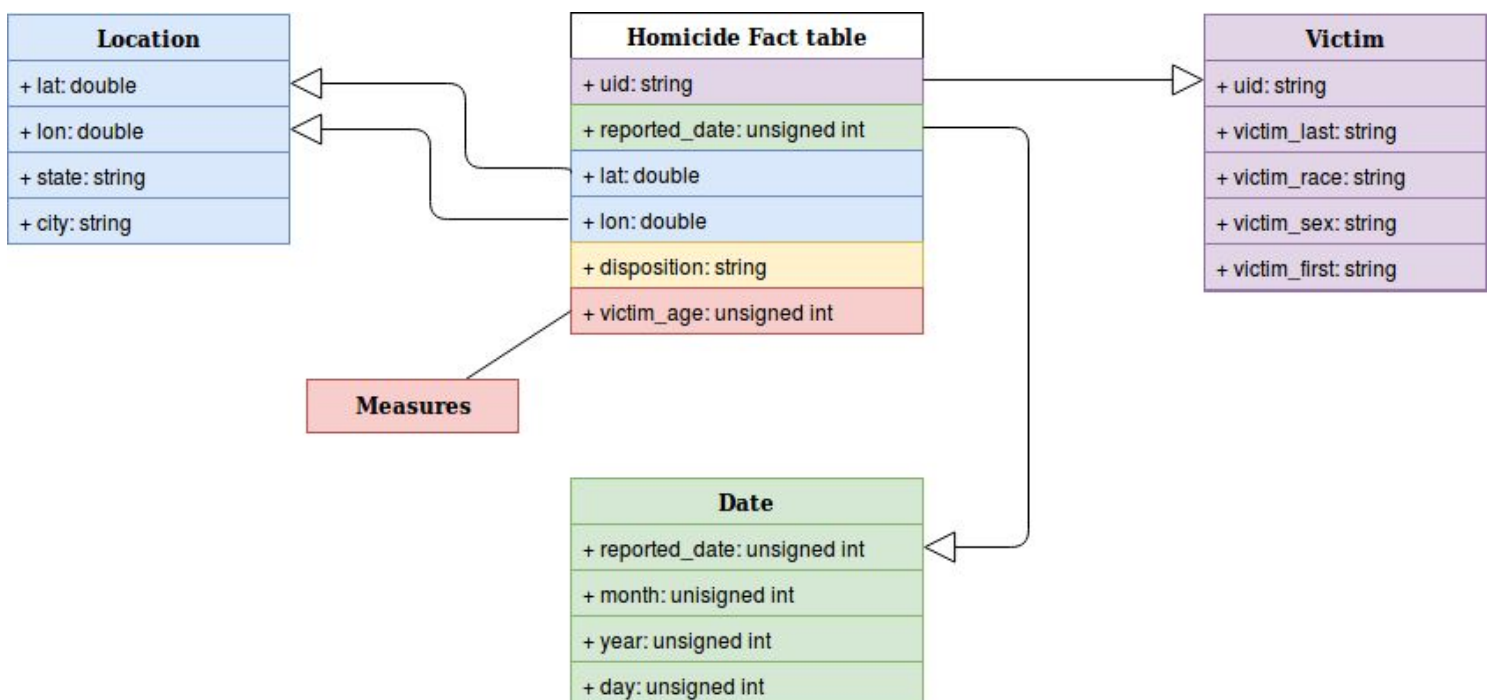
Dataset

We have chosen the crime data set available [here](#). The data set contains homicides recorded in the United States.

Cleaning the Dataset

Since we are using “victim_age” as an Algebraic Measure, we need to clean all cells that have “Unknown” value for the “victim_age” attribute. In MySQL, if “victim_age” is classified as INT type, then all such cells get typecasted to 0. This would obviously lead to false averages. Hence, we marked all such cells with -1, and don’t consider such cells while performing Aggregate Queries.

Star Schema



Interesting Results from the Data Cube

1. When we perform Aggregate (*average*) Query of the Measure (*victim_age*) for 0-D (*apex*) Cuboid, we derive that:
 - a. Average age of the victims is 31.8012.
2. When performing Aggregate (*count*, *average victim_age*) Query for the 1-D Cuboid **{victim_sex}**, we derive the following:
 - a. It was interesting to note that out 52179 homicide incidents, 40739 of them had male victims. This percentage amounts to almost 78%.

<u>Sex</u>	<u>Count</u>	<u>Avg. age</u>
Male	40739	31.39
Female	7209	34.09
Unknown	4231	32.07

- b. Another interesting result is that female victims on an average are 3 years older than male victims.
3. Aggregate (*count*, *average victim_age*) Query for the 1-D Cuboid **{victim_race}**:

<u>Race</u>	<u>Count</u>	<u>Avg. age</u>
Asian	685	37.73
Black	33361	30.43
Hispanic	6901	30.16
White	6333	39.83
Unknown	4199	32.17
Other	700	34.00

- a. 63.9% of the victims belong to black race.
 - b. Among all the races, the oldest victims belong to White.
4. Aggregate (*count*) Query for the 1-D Cuboid **{disposition}**:

<u>Disposition</u>	<u>Count</u>
Closed by arrest	25674
Closed without arrest	2922
Open/no arrest	23583

- a. Most of the cases are either open or closed by arrest (approximately 50-50). Very few cases (~5%) have been closed without arrest.

5. Aggregate (*count, average victim_age*) Query for the 2-D Cuboid {**victim_sex, victim_race**}:

<u>Sex</u>	<u>Race</u>	<u>Count</u>	<u>Avg. age</u>
Male	Black	29256	30.27
Male	Hispanic	5901	30.09
Male	White	4453	39.51
Female	Black	4066	31.54
Female	White	1838	40.64

- a. 29256 victims, which forms a whopping 56% of all records, are male and black. Even out of female victims, the number of records for black females is much higher than that of other race.
- b. For the women and men belonging to same race, the average age of the victim was always more for females than males.
- c. White Females were the oldest victims out of all (sex,race) combinations.

6. Aggregate (*count, average victim_age*) Query for the 2-D Cuboid {**victim_sex, disposition**}:

<u>Sex</u>	<u>Disposition</u>	<u>Count</u>	<u>Avg. age</u>
Female	Closed by arrest	4415	33.44
Female	Closed without arrest	832	38.79
Female	Open	1962	33.52
Male	Closed by arrest	19092	31.72

- a. On an average, out of the homicide cases with Female victim, the cases which are closed without arresting the killer involve the oldest group of victims. This shows a tendency to let killers of older Females go.
- b. When the victim is female, 61% of the cases are closed by arrest. However when the victim is male, only 46% of the cases are closed by arrest. This shows that on an average the female murders were considered more criminal than the male murders.

7. Aggregate (*count*) Query for the 2-D Cuboid **{victim_race, disposition}**:

<u>Race</u>	<u>Disposition</u>	<u>Count</u>
White	Closed by arrest	3959
White	Open + Closed without arrest	2374
Black	Closed by arrest	15462
Black	Open + Closed without arrest	17899

- a. Out of all White homicides, 62.5% of the cases were closed by arrest. Whereas in Black homicides, only 46.3% of the cases were closed by arrest. This shows that the overall consequences were more serious when there was a white victim.

8. Aggregate (*count*) Query for the 3-D Cuboid **{victim_race, victim_race, disposition}**:

- a. Since we saw high arrest rates for White & Female victims, and low arrest rates for Black & Male victims, we will view the disposition patterns for both these race & sex combinations to get our overall understanding of this dataset.

<u>Race</u>	<u>Sex</u>	<u>Disposition</u>	<u>Count</u>
White	Female	Closed by arrest	1148
White	Female	Open + Closed without arrest	690
Black	Male	Closed by arrest	12956
Black	Male	Open + Closed without arrest	16300

- b. As the data clearly shows, when the victim was a White Female, the percentage of cases closed by arrest is 62%. Whereas when the victim was a Black Male, the percentage of cases closed by arrest is just 44%. This shows a pattern of more stringent action for cases involving White & Female victims.