A Letter with Questions

From: Raj Kumar(Leader of of the team)

To: Database Team

Hello DBA team,

Happy new year! I hope all of you have enjoyed your holiday and have taken a good rest. And thank you for all your efforts and help in the previous year. Now there comes the first task of the new year. I have just received an email from my manager which asks our team to provide some analysis to the data of the shooting incidents that occurred in NYC in the first half year of 2019. This will help the NY Police Department to explore the nature of shooting activity. And I have some problems about the data which need your help.

In the database 'NY shooting' which include the information shooting incidents of the first half year of 2019:

1. In which month in the first half year of 2019 there happens the most shooting incidents?
2. on which day of the month there happens the most shooting incidents?
3. During which time period of a day are the shooting incidents most likely to happen from this dataset?
4. Which borough is the most dangerous, and which one is the least dangerous only concerning the number of shooting incidents?
5. In the borough which is considered as the most dangerous from the previous problem, in what location (place) there happens the most shooting incidents?
6. How many incidents are given a flag of murder?(Attention: There may be multiple pairs of perpetrators in a single incident where some of which are considered as murder while others are not. As long as a pair of perpetrator and victim in an incident is given a murder flag, this incident should be considered in the result.)
7. In which age group does most perpetrators lie in? And in which age group does most victims lie in?
8. What is the latitude and the longitude of the shooting incidents where there are most number of pairs of perpetrators and victims(i,e. the most serious)?
Thank you for your help. And I am looking forward to seeing your results!
Best,
Raj

A Reply Letter with Answers

From: Qi Sun

To: Raj Kumar

Hello Kumar,

Happy New year! Concerning your questions in the email, I am very willing to answer them on behalf of my teammates.

1.
```sql
select month(occur_time), count(1) from unique_shooting group by month(occur_time) order by count(1) desc limit 1;
```

	`month(occur_time)`	`count(1)`
1	6	89

In the June there happens the most shooting events.

2.
```sql
select day(occur_time), count(1) from unique_shooting group by day(occur_time) order by count(1) desc limit 1;
```

	`day(occur_time)`	`count(1)`
1	23	18

On the 23rd of the month.

3.
```sql
select hour(occur_time), count(1) from unique_shooting group by hour(occur_time)  order by

count(1) desc;
```

	`hour(occur_time)`	`count(1)`
1	0	33
2	22	33
3	23	25
4	1	22
5	21	22
6	19	21
7	2	21
8	18	20

It can be seen from the result that most shooting events happened at midnight (0am, 1am, 22pm, 23pm).

4.

```sql
select b.borough, count(1) from unique_shooting join borough b on
unique_shooting.borough_id = b.id group by b.borough order by count(1) desc;
```

	borough	`count(1)`
1	BROOKLYN	141
2	BRONX	96
3	MANHATTAN	59
4	QUEENS	57
5	STATEN ISLAND	9

Brooklyn is the dangerous, while Staten Island is the least dangerous if only considering the number of shooting events in the first half year of 2019.

5.

```sql
select location, count(1) from unique_shooting join location l on
unique_shooting.location_id = l.id join borough b on unique_shooting.borough_id =
b.id where borough='BROOKLYN' group by location order by count(1) desc;
```

location	`count(1)`
1 MULTI DWELL - PUBLIC HOUS	32
2 MULTI DWELL - APT BUILD	10
3 PVT HOUSE	6
4 GROCERY/BODEGA	5
5 FAST FOOD	1
6 HOSPITAL	1
7 RESTAURANT/DINER	1
8 GAS STATION	1

It can be seen from the result the public house is the place where most shooting events happens in Brooklyn.

6.

```sql
select count(distinct event_id) from perp_and_vict where is_murder=TRUE;
```

	`count(distinct event_id)`
1	63

There are 63 events in the first half year of 2019 which are given a flag of murder.

7.

```sql
select age_group, count(1) from perp_and_vict join age_group ag on
perp_and_vict.perp_age_id = ag.id group by age_group order by count(1) desc;
```

age_group	`count(1)`
1 25-44	111
2 18-24	95
3 <18	29
4 45-64	10
5 UNKNOWN	3
6 65+	2

Because the information of some perpetrators of some events is not known, roughly speaking, 25-44 and 18-24 are the age groups where the most perpetrators lie in.

```sql
select age_group, count(1) from perp_and_vict join age_group ag on
perp_and_vict.vict_age_id = ag.id group by age_group order by count(1) desc;
```

	age_group	`count(1)`
1	25-44	227
2	18-24	125
3	45-64	51
4	<18	35
5	UNKNOWN	3
6	65+	1

25-44 and 18-24 are also the age groups where the most victims lie in.

8.

```
select event_id, count(1), latitude, longitude from perp_and_vict join unique_shooting
us on perp_and_vict.event_id = us.id group by event_id order by count(1) desc;
```

	event_id	`count(1)`	latitude	longitude
1	17	6	40.6665	-73.8878
2	69	6	40.6792	-73.928
3	23	4	40.8554	-73.9047
4	39	4	40.7867	-73.9411
5	53	4	40.8694	-73.8504
6	67	4	40.8341	-73.9429
7	100	4	40.863	-73.9243
8	101	4	40.5792	-74.0004

Their latitudes and longitudes are respectively (40.6665, -73.8878) and (40.6792, -73.928). There are 6 pairs of perpetrators and victims in these two incidents.

I hope the following answers can help you.

Best,

Qi Sun