DE1 Projekt – Daniel Dajka

\*\*

Data source:

<https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95>

first 1000 rows are used to make the sql lighter

\*\*

create schema termprojekt;

use termprojekt;

drop table motorcrashes;

drop table factors;

drop table streets;

create table motorcrashes

(collision\_id Integer not null,

crash\_date DATE,

crash\_time TIME,

borough VARCHAR(20),

zip\_code VARCHAR(10),

latitude varchar(30),

longitude varchar(30),

vehicle\_type VARCHAR (20),

number\_of\_persons\_injured integer,

number\_of\_persons\_killed integer,

number\_of\_pedestrians\_injured integer,

number\_of\_pedestrians\_killed integer,

number\_of\_cyclist\_injured integer,

number\_of\_cyclist\_killed integer,

number\_of\_motorist\_injured integer,

number\_of\_motorist\_killed integer,

Primary key (collision\_id)

);

create table factors

(factor\_id integer not null auto\_increment,

collision\_id integer not null,

contributing\_factor\_vehicle1 varchar(30),

primary key (factor\_id),

foreign key (collision\_id) references motorcrashes (collision\_id)

);

Create table streets

(street\_id integer not null auto\_increment,

collision\_id Integer not null,

on\_street\_name Varchar(25),

cross\_street\_name Varchar(25),

off\_street\_name Varchar(25),

primary key (street\_id),

foreign key (collision\_id) references motorcrashes (collision\_id)

);

**SET GLOBAL local\_infile=1; /\*hogy tudjak beolvasni CSV a szamitogépemröl\*/**

LOAD DATA Local INFILE '/Users/Dani/Desktop/first10000motorcrash.csv'

INTO TABLE motorcrashes

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

IGNORE 1 LINES

(@collision\_id,

crash\_date,

crash\_time,

borough,

zip\_code,

latitude,

longitude,

on\_street\_name,

vehicle\_type);

* Nem sikerült kóddal, ezért mysql ablakon beimportáltam
* Utána jött a dump, hogy replikálható legyen -> 190 sortól indul az analytics

**Analytics Plan:**

**How many total deaths were there in 2020 until the day?**

select sum(number\_of\_persons\_killed) as total\_killed

from motorcrashes

where crash\_date between '2019-12-31'and '2020-12-31';->30

2020 -> 30 ->last date is 2020.10.23 -> download of the database

**Are there more pedestrian, cyclist or motorist injuries?**

select sum(number\_of\_pedestrians\_injured), sum(number\_of\_cyclist\_injured), sum(number\_of\_motorist\_injured)

from motorcrashes;

****

select sum(number\_of\_pedestrians\_killed), sum(number\_of\_cyclist\_killed), sum(number\_of\_motorist\_killed)

from motorcrashes; 

select

sum(number\_of\_pedestrians\_injured)/sum(number\_of\_pedestrians\_killed) as death\_ratio\_pedestrians,

sum(number\_of\_cyclist\_injured)/sum(number\_of\_cyclist\_killed) as death\_ratio\_cyclist,

sum(number\_of\_motorist\_injured)/sum(number\_of\_motorist\_killed) as death\_ratio\_motorist

from motorcrashes;



On average from every 91 accidents one ends with death for pedestrians. There are 5 times as many accidents for motorist than pedestrians, nevertheless, the death ratio is considerably higher, there have to be on average 185 accidents for one deadly accident.

**Is there pattern in crash time? Morning, afternoon, night?**

alter table motorcrashes

add column crash\_time\_t time after crash\_time;

update motorcrashes

set crash\_time\_t = str\_to\_date(crash\_time, '%H:%i')

where collision\_id > 0;

alter table motorcrashes

drop column crash\_time;

alter table motorcrashes

change crash\_time\_t crash\_time time;

A picture containing table

Description automatically generated

select sum(number\_of\_persons\_killed) as total\_killed,

Case

when crash\_time between '06:00:00' and '12:00:00' then 'Morning'

when crash\_time between '12:00:00' and '20:00:00' then 'Afternoon'

Else 'Night'

END

as period\_of\_day

from motorcrashes

group by period\_of\_day;

Table

Description automatically generated

Not surprisingly most accidents are in the afternoon, however, more than double as many people are killed at night in road accidents then in the afternoon or in the morning.

**Which borough had the biggest number accidents? Which had the least accidents?**

select count(borough), borough

from motorcrashes

where borough not like ''

group by borough;

**Table

Description automatically generated**

**Is there a street that is particularly prone to accidents? Or a junction (cross-street)? Join table**

select count(on\_street\_name), on\_street\_name

from streets

left join motorcrashes

on streets.collision\_id=motorcrashes.collision\_id

where on\_street\_name not like ''

group by on\_street\_name

order by count(on\_street\_name) desc

limit 10;

top 10 most dangerous streets in New York:

Table

Description automatically generated

select count(cross\_street\_name), cross\_street\_name as junction

from streets

left join motorcrashes

on streets.collision\_id=motorcrashes.collision\_id

where cross\_street\_name not like ''

group by cross\_street\_name

order by count(cross\_street\_name) desc;

Table

Description automatically generated

**What was the main contributing factor for the injuries or deaths?**

select sum(number\_of\_persons\_injured), sum(number\_of\_persons\_killed), count(contributing\_factor\_vehicle1), contributing\_factor\_vehicle1

from motorcrashes

inner join factors

on factors.collision\_id=motorcrashes.collision\_id

where number\_of\_persons\_injured >0 or number\_of\_persons\_killed != 0

group by contributing\_factor\_vehicle1

having contributing\_factor\_vehicle1 not like ''

order by count(contributing\_factor\_vehicle1) desc;

Table

Description automatically generated

Driver Inattention is the primary contributing factor to accidents with 969 accidents attributed to this reason. 681 accidents are unspecified out of which accidents 10 people have lost their lives.

**View**

**How many accidents happened under the influence of alcohol?**

DROP VIEW IF EXISTS alcohol\_influenced\_accidents;

create view alcohol\_influenced\_accidents as

select sum(number\_of\_persons\_injured), sum(number\_of\_persons\_killed), count(contributing\_factor\_vehicle1), contributing\_factor\_vehicle1

from motorcrashes

inner join factors

on factors.collision\_id=motorcrashes.collision\_id

group by contributing\_factor\_vehicle1

having contributing\_factor\_vehicle1 like '%alcohol%'

order by count(contributing\_factor\_vehicle1) desc;

select \* from **alcohol\_influenced\_accidents**;

****

**Stored procedure:**

**Is there a vehicle type that causes more accidents than the other?**

DROP PROCEDURE IF EXISTS accidents\_by\_vehicle\_type;

DELIMITER //

CREATE PROCEDURE accidents\_by\_vehicle\_type(

In kocsibemegy Varchar(45)

)

BEGIN

Select \*

from motorcrashes

where vehicle\_type like kocsibemegy;

END //

DELIMITER ;

call accidents\_by\_vehicle\_type('sedan');

Table

Description automatically generated

This is nice, but it ignores other types of sedan from the database like: Sedan + 2 dr sedan, 4dr sedan

* call accidents\_by\_vehicle\_type('sedan%'); -> solves it but the user might not use the %

another solution:

DROP PROCEDURE IF EXISTS accidents\_by\_vehicle\_type;

DELIMITER //

CREATE PROCEDURE accidents\_by\_vehicle\_type(

In kocsibemegy Varchar(45)

)

BEGIN

Select \*

from motorcrashes

where vehicle\_type like concat('%',kocsibemegy, '%');

END //

DELIMITER ;

call accidents\_by\_vehicle\_type('sedan');

Table

Description automatically generated

call accidents\_by\_vehicle\_type('truck');

Table

Description automatically generated

**Trigger -> if a new accident happens?**

drop table log;

create table log

(id integer not null auto\_increment,

message varchar(39),

log\_date DATE,

PRIMARY KEY (id)

);

DROP TRIGGER IF EXISTS logger\_trigger;

CREATE TRIGGER logger\_trigger

AFTER insert

ON motorcrashes FOR EACH ROW

insert into log(message, log\_date) values ("motorcrashes insertion", now());

insert into motorcrashes (collision\_id,borough)

values (10000, 'Yozsefvaros');

select \* from log;

Graphical user interface, application, table

Description automatically generated

* after insert : trigger activates after the new row is added and shows ‚ motorcrashes insertion‘
* log (message, log\_date) -> id is not needed since autoincrement is activated