Name: K.Praveen kumar

Assignment-7(11-12-2023)

**Data cleaning:**

To perform some of the opeations in the data cleaning we need to create a table.

Create a table:

create table companies( id int,name varchar(50),industry varchar(30),year\_founded int,city varchar(30));

Inserting values :

insert into companies values(1,'overhex','software',2006,'Franklin'),

(2,'unimattax','itservices',2009,'newtonsquare'),

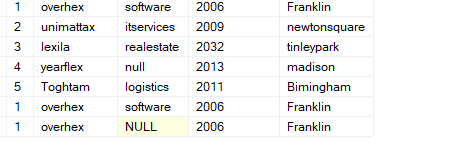
(3,'lexila','realestate',2032,'tinleypark'),

(4,'yearflex','null',2013,'madison')

(5,'Toghtam','logistics',2011,'Bimingham');

insert into companies values(1,'overhex','software',2006,'Franklin');

insert into companies values(1,'overhex',null,2006,'Franklin');



Deleting duplicate data:

select id,count(name) as count

from companies

group by id

having(id)>1;

From the above query we know the duplicate value. So to remove the duplicates we will use a combination of ROW\_NUMBER() and delete.

WITH cte AS (SELECT name,ROW\_NUMBER() OVER (PARTITION BY name ORDER BY name ASC) AS rn

FROM companies)

DELETE FROM cte

WHERE rn > 1;



**Inserted another value into the table:**

insert into companies values(6,'Quotelane',null,null,'Greenville');

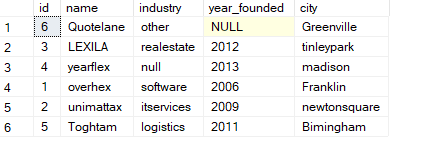


**To update data:**

update companies

set industry='other'

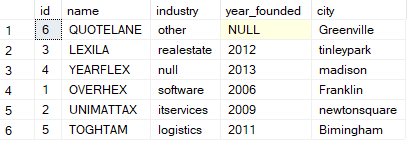
where industry is null;



**TO uppercase:**

update companies

set name =upper(name);

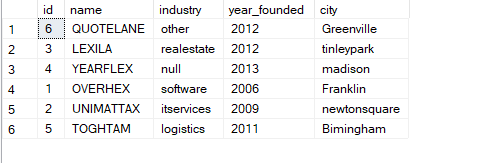


update companies

set year\_founded=2012

from companies

where id='6';



**DATABASE SCHEMA DIAGRAMS:**



**PARTITION BY:**

create table car\_prices(company varchar(50),model varchar(50),car\_type varchar(50),price int);

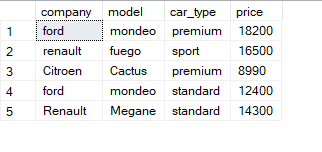
insert into car\_prices values('ford','mondeo','premium',18200),

('renault','fuego','sport',16500),

('Citroen','Cactus','premium',8990),

('ford','mondeo','standard',12400),

('Renault','Megane','standard',14300);



select company,model,price,

avg(price) over() as 'overall avg price',

avg(price) over(partition by car\_type) as 'car type avg price'

from car\_prices;

