create database test;

use test;

drop table airbnb;

create table airbnb (

id int,

name varchar(100),

host\_id bigint,

host\_identity\_verified varchar(50),

host\_name varchar(50),

neighbourhood\_group varchar(50),

neighbourhood varchar(50),

latitude double,

longitude double,

country varchar(50),

country\_code varchar(10),

instant\_bookable varchar(10),

cancellation\_policy varchar(10),

room\_type varchar(50),

construction\_year int,

price int,

service\_fee int,

minimum\_nights int,

number\_of\_reviews int,

last\_review varchar(50),

reviews\_per\_month Decimal(5,2),

review\_rate\_number int,

calculated\_host\_listings\_count int,

availability int,

house\_rules text,

license text

);

LOAD DATA INFILE 'C:/Airbnb\_Open\_Data3.csv'

INTO TABLE airbnb

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 lines

;

show warnings;

with cte as(

select \*,

row\_number() over(partition by id order by id)ranked from airbnb)

select count(\*)

from cte

where ranked=2;

# Made Primary key afte removing duplicates

DELETE FROM airbnb

WHERE id IN (

SELECT id FROM (

SELECT id, ROW\_NUMBER() OVER (PARTITION BY name, host\_id ORDER BY id) AS row\_num

FROM airbnb

) temp WHERE row\_num > 1

);

UPDATE airbnb

SET last\_review = STR\_TO\_DATE(last\_review, '%m/%d/%Y')

WHERE last\_review LIKE '%/%/%';

select \* from airbnb;

# Trim

update airbnb

set host\_identity\_verified= trim(host\_identity\_verified);

# Setting primary key

alter table airbnb

add primary key (id);

select count(cancellation\_policy)

from airbnb

where cancellation\_policy="flexible"