Name - Shivam Raj Employee ID - TAS047

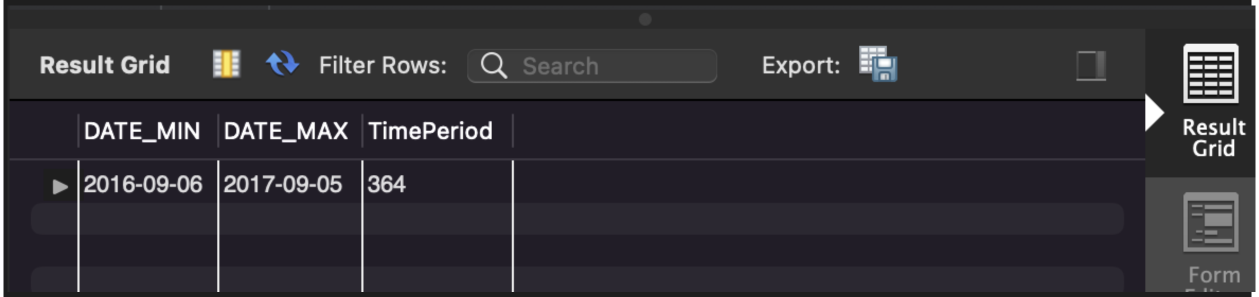
```
-- Creating the database
create database shivam;
use shivam;
-- Creating the Table in the Database
CREATE TABLE `shivam`.`airbnb` (
   `listing_id` INT NOT NULL,
   `date` DATE NULL,
   `available` VARCHAR(1) NULL,
   `price` VARCHAR(10) NULL
);

desc airbnb;

drop table airbnb;

SELECT
   *
FROM
   airbnb;


SHOW VARIABLES LIKE "secure_file_priv";

-- Importing the CSV FILE

load data infile "/tmp/airbnb_calendar.csv"
into table airbnb
fields terminated by ','
enclosed by '"'
lines terminated by '\n'
```

```
ignore 1 rows;

SELECT
   *
FROM
   airbnb
ORDER BY listing_id;


-- Cleaning the Data

SET SQL_SAFE_UPDATES = 0;
UPDATE airbnb
SET
   price = NULL
WHERE
   price = '';
UPDATE airbnb
SET
   price = TRIM(LEADING '$' FROM price);
```

```
/*============Q1==================*/
```

Q1. What is the time period used?

```
(SELECT
   date
FROM
   airbnb
ORDER BY date
LIMIT 1) UNION (SELECT
   date
FROM
   airbnb
ORDER BY date DESC
LIMIT 1);
```

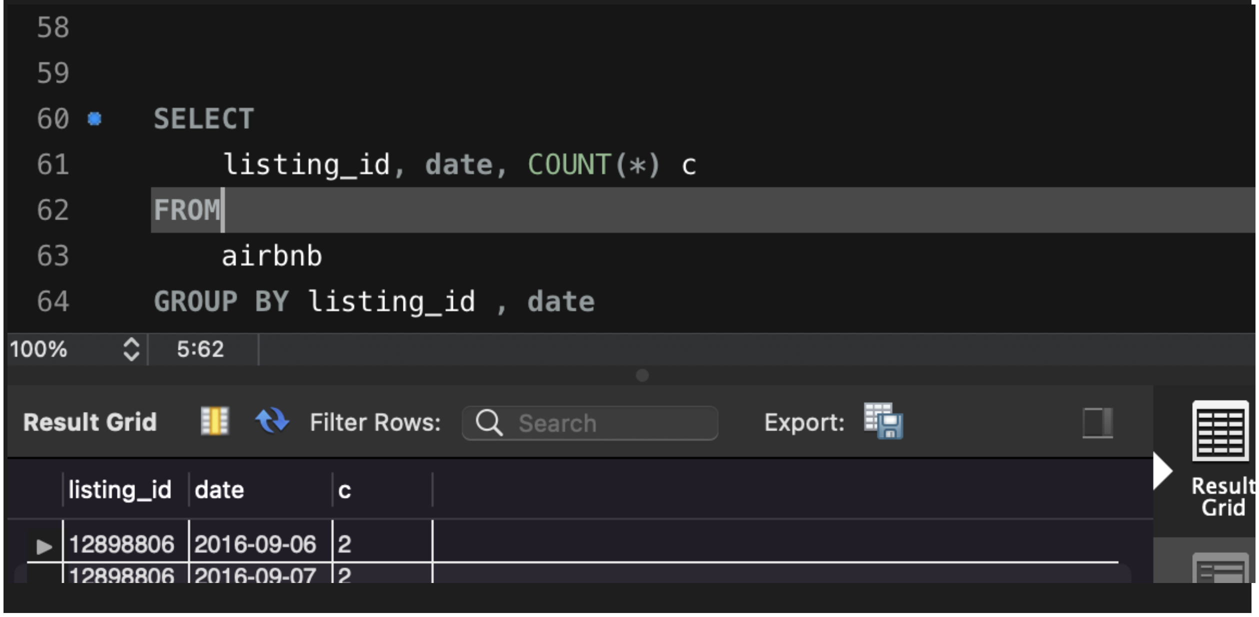


DATE_MIN	DATE_MAX	TimePeriod
2016-09-06	2017-09-05	364

```
/*=============Q2===================*/
```

Q2. How many properties have duplicate entries? Remove duplicate rows (say a row appears 3 times, remove 2 and keep 1)

```
SELECT
   listing_id, date, COUNT(*) c
FROM
   airbnb
GROUP BY listing_id , date
HAVING COUNT(*) > 1;

DELETE t1 FROM (SELECT *, ROW_NUMBER () OVER(listing_id ) as
rownum FROM airbnb) t1
INNER JOIN (SELECT *, ROW_NUMBER() OVER(listing_id) as rownum
FROM airbnb) t2
WHERE
t1.date= t2.date AND
t1.listing_id = t2.listing_id AND
ti.rownum < t2.rownum;
```



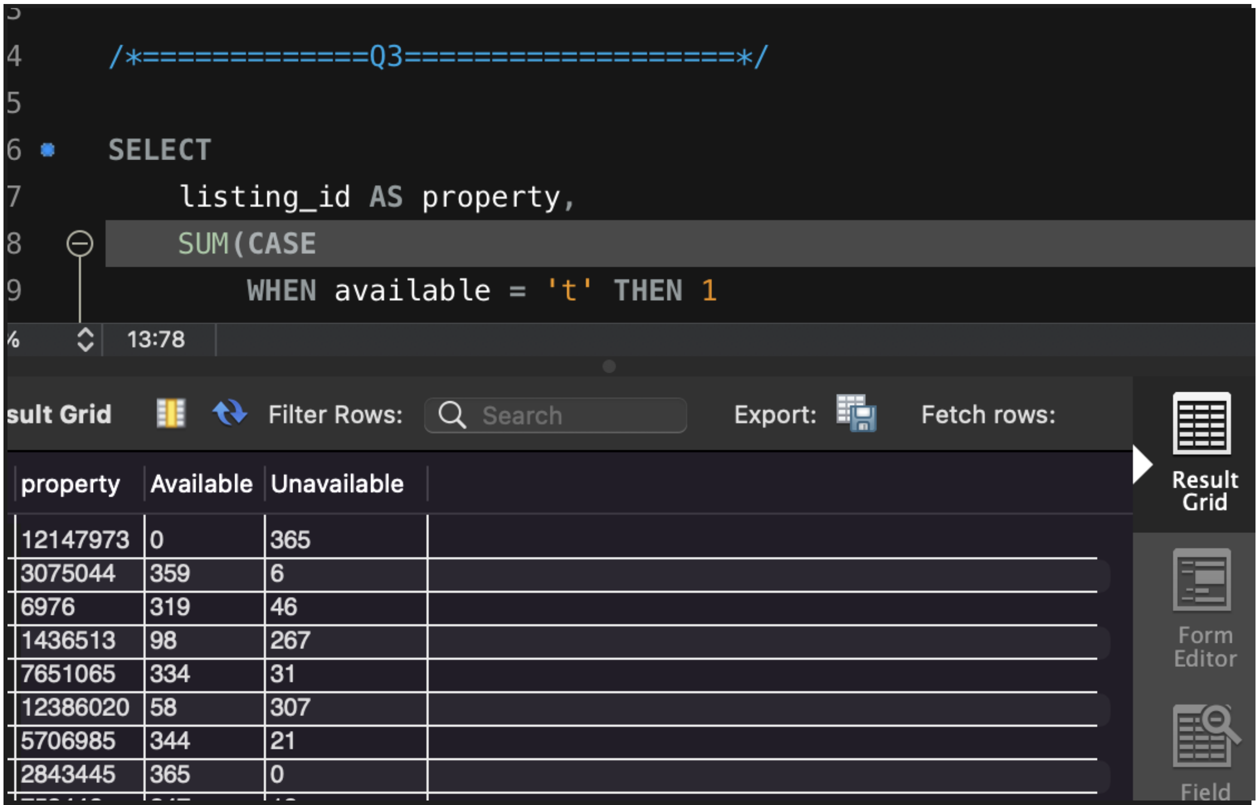
```
SELECT
    listing_id, date, COUNT(*) c
FROM
    airbnb
GROUP BY listing_id , date
```

```
  listing_id date       c
▶ 12898806   2016-09-06 2
  12898806   2016-09-07 2
```

```
/*=============Q3===================*/
```

Q3. For each property, find out the number of days the property was available and not available (create a table with listing_id, available days, unavailable days and available days as a fraction of total days)

```
SELECT
   listing_id AS property,
   SUM(CASE
       WHEN available = 't' THEN 1
       ELSE 0
   END) AS Available,
   SUM(CASE
       WHEN available = 'f' THEN 1
       ELSE 0
   END) AS Unavailable
FROM
   airbnb
GROUP BY listing_id;
```



```
/*=============Q3===================*/

SELECT
    listing_id AS property,
    SUM(CASE
        WHEN available = 't' THEN 1
```

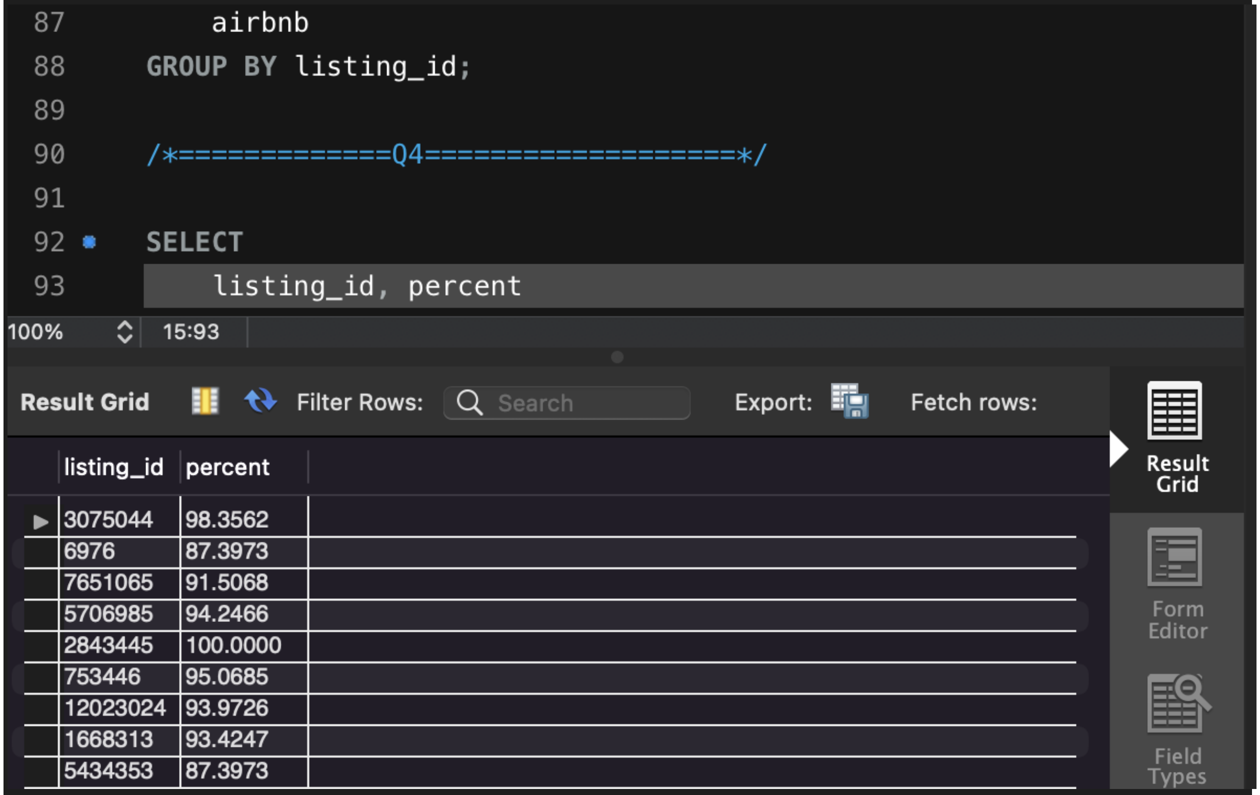
property	Available	Unavailable
12147973	0	365
3075044	359	6
6976	319	46
1436513	98	267
7651065	334	31
12386020	58	307
5706985	344	21
2843445	365	0

```
/*=============Q4===================*/
```

Q4. How many properties were available on more than 50% of the days? How many properties were available on more than 75% of the days?

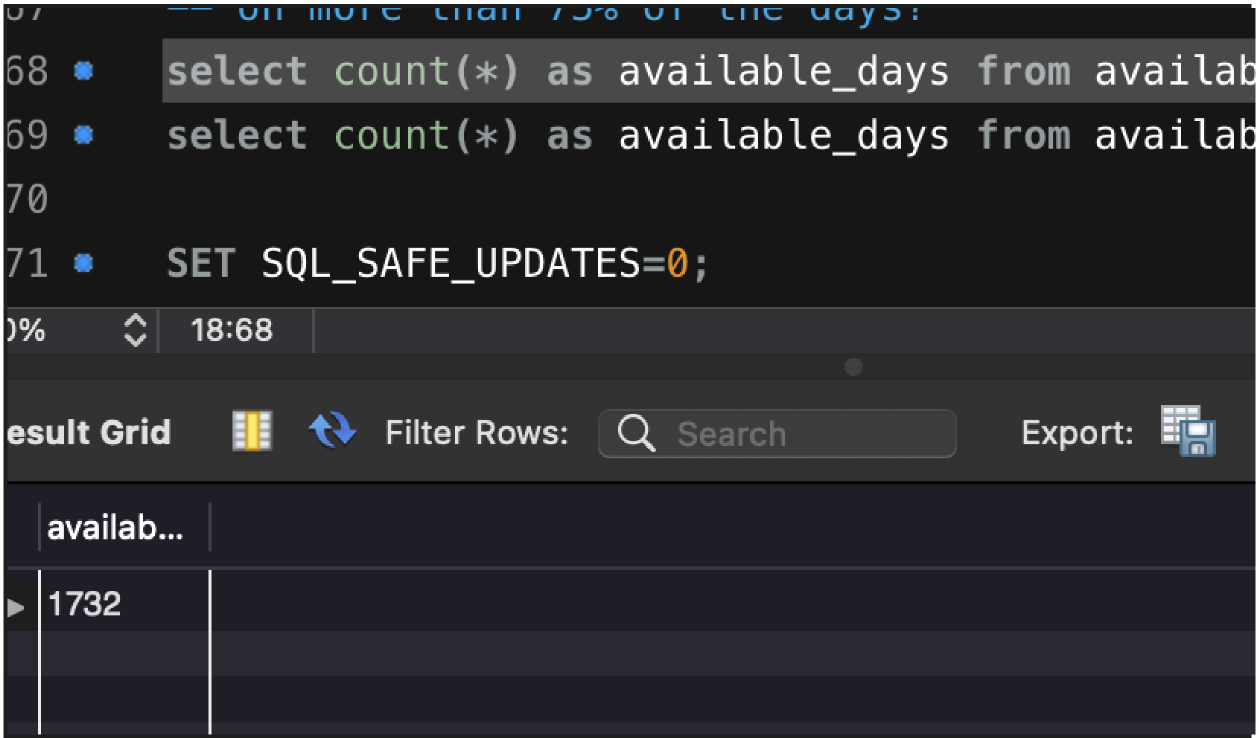
```
SELECT
   listing_id, percent
FROM
   (SELECT
       listing_id,
           (COUNT(CASE
               WHEN available = 't' THEN 1
           END) / COUNT(*)) * 100 AS percent
   FROM
       airbnb
   GROUP BY listing_id) AS temp
WHERE
   percent > 50;
SELECT
```

```
   listing_id, percent
FROM
   (SELECT
       listing_id,
           (COUNT(CASE
               WHEN available = 't' THEN 1
           END) / COUNT(*)) * 100 AS percent
   FROM
       airbnb
   GROUP BY listing_id) AS temp
WHERE
   percent > 75;
```



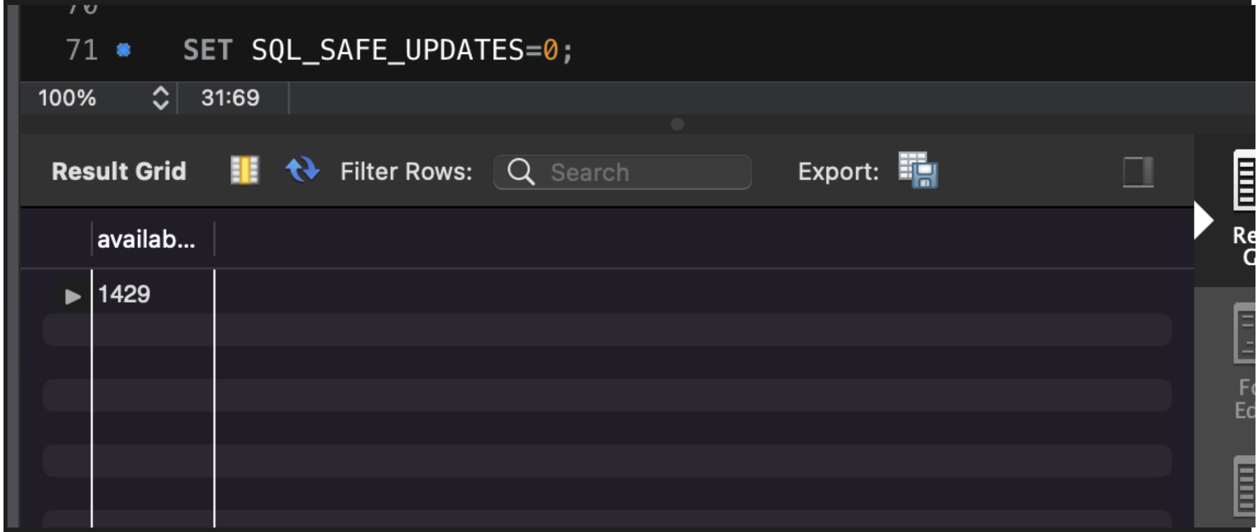
```
        airbnb
    GROUP BY listing_id;

    /*=============Q4===================*/

    SELECT
        listing_id, percent
```

listing_id	percent
3075044	98.3562
6976	87.3973
7651065	91.5068
5706985	94.2466
2843445	100.0000
753446	95.0685
12023024	93.9726
1668313	93.4247
5434353	87.3973



```
-- on more than 75% of the days?
select count(*) as available_days from availab
select count(*) as available_days from availab

SET SQL_SAFE_UPDATES=0;
```

```
availab...
1732
```



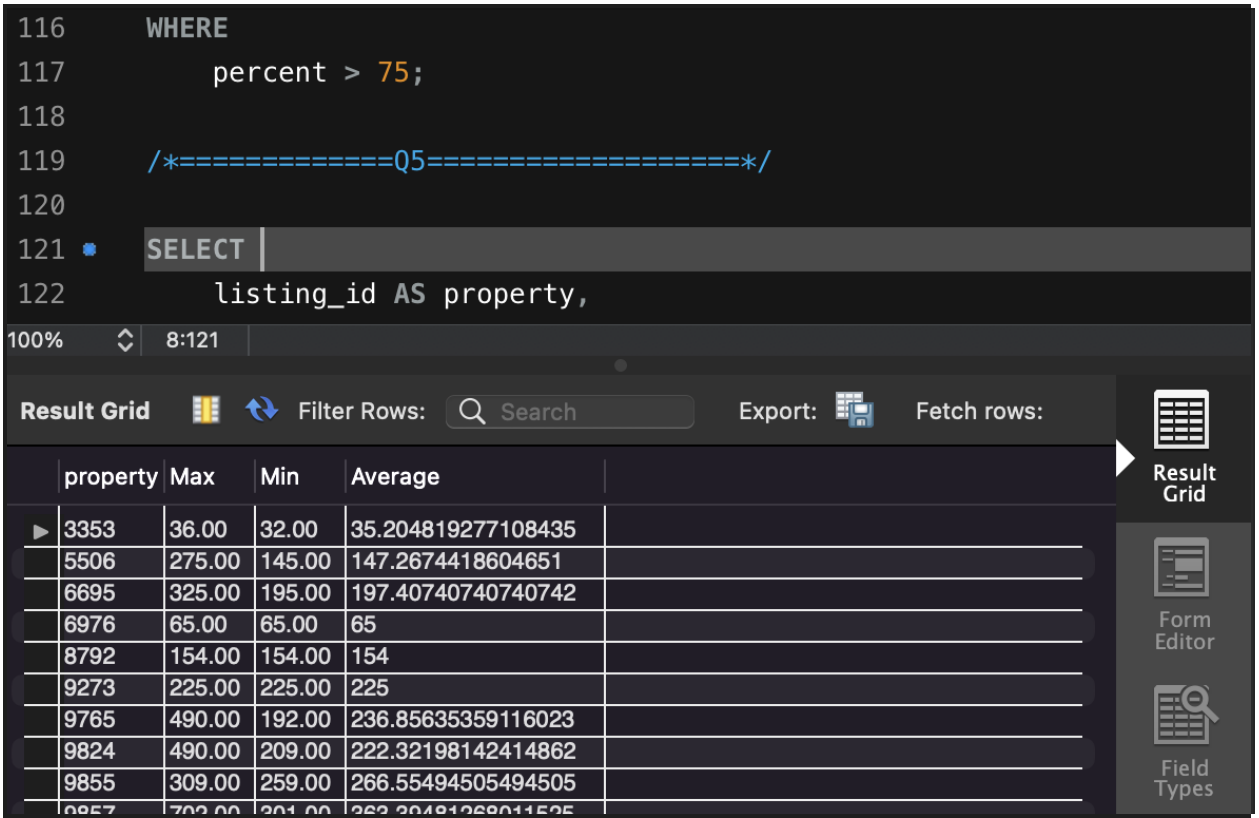
```
SET SQL_SAFE_UPDATES=0;
```

```
availab...
1429
```

```
/*=============Q5===================*/
```

Q5. Create a table with max, min and average price of each property.

```
SELECT
   listing_id AS property,
   MAX(price) AS Max,
   MIN(price) AS Min,
   AVG(price) AS Average
FROM
   airbnb
GROUP BY listing_id;
```



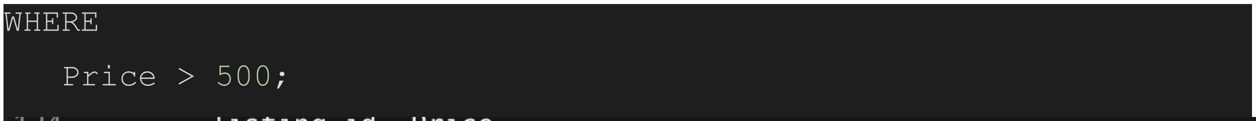
```
    WHERE
        percent > 75;

    /*============Q5==================*/

    SELECT
        listing_id AS property,
```

property	Max	Min	Average
3353	36.00	32.00	35.204819277108435
5506	275.00	145.00	147.2674418604651
6695	325.00	195.00	197.40740740740742
6976	65.00	65.00	65
8792	154.00	154.00	154
9273	225.00	225.00	225
9765	490.00	192.00	236.85635359116023
9824	490.00	209.00	222.32198142414862
9855	309.00	259.00	266.55494505494505
9857	702.00	301.00	363.39481268011525

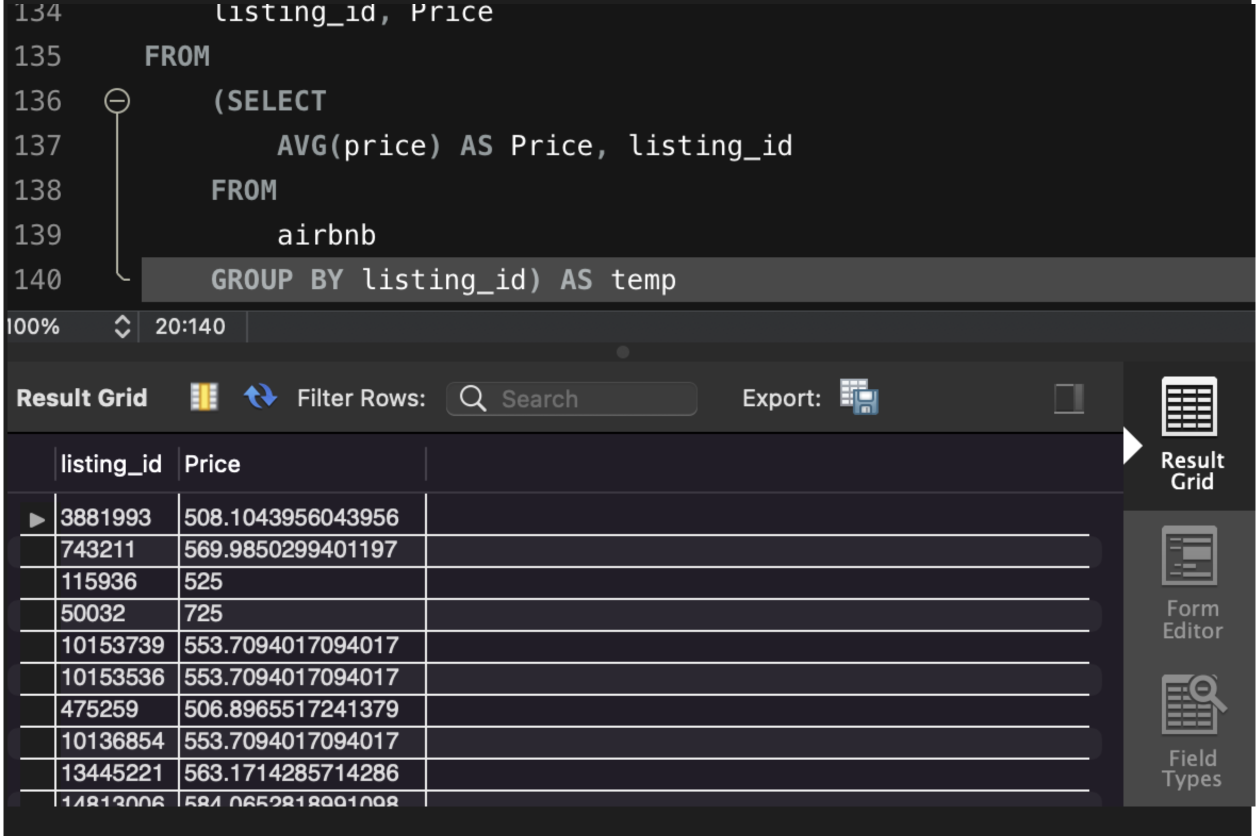
```
/*============Q6==================*/
```

Q6. Extract properties with an average price of more than $500.

```
SELECT
   listing_id, Price
FROM
   (SELECT
       AVG(price) AS Price, listing_id
   FROM
       airbnb
   GROUP BY listing_id) AS temp
```



```
WHERE
   Price > 500;
```



```
        listing_id, Price
    FROM
        (SELECT
            AVG(price) AS Price, listing_id
        FROM
            airbnb
        GROUP BY listing_id) AS temp
```

```
listing_id  Price
3881993     508.1043956043956
743211      569.9850299401197
115936      525
50032       725
10153739    553.7094017094017
10153536    553.7094017094017
475259      506.8965517241379
10136854    553.7094017094017
13445221    563.1714285714286
14813006    584.0652818991098
```