

# Data Understanding & Preparation

## Final Project




# Sri Lanka Tourism through Airbnb

### Project Group 4 Team

1. Prabakaran V
2. Ullas D P
3. Venkatesh D
4. Vignesh J

# Business Objective

---

-  To understand the availability and preferences in Airbnb properties across all important locations in Sri Lanka
-  To compare the price range and customer satisfaction level across all kinds of properties and locations
-  To evaluate and find the ideal investment opportunity in an Airbnb property

# About Airbnb

---

Airbnb operates an online marketplace and hospitality service for homeowners to list/rent short-term lodging for tourists and vacationers



Holiday cottages



Apartments



Homestays



Hostel beds



Hotel rooms

The company has over 5 million lodging listed across  
81,000 cities and 191 countries.

# Dataset & Data Dictionary

---

Data Source - <http://tomslee.net/airbnb-data-collection-get-the-data>

Variable Name	Description
room_id	Unique ID of the rooms available
host_id	Unique ID of the host who have put up their space for rent in Airbnb
room_type	Category of rooms available
borough	An administrative division of a state
neighbourhood	Nearby area/close proximity to the borough
reviews	No. of reviews given by customers
overall_satisfaction	Average satisfaction level on a scale of 0-5
accommodates	No. of guests accommodatable in the rooms/room_type
bedrooms	No. of bedrooms available in the rooms/room_type
price	Price of the rooms/room_type
Minstay	Minimum no. of days a customer has to stay to avail a particular room/room_type
Latitude	Location of the room in latitude
Longitude	Location of the room in longitude

# Normalization Process

## 1 NF

room_id	host_id	room_type	bedrooms	accommodates	minstay	price	reviews	overall_satisfaction	neighborhood	borough	latitude	longitude
Primary Key												

## 2 NF

tbl_room												
room_id	host_id	room_type	neighborhood	borough	latitude	longitude						
Primary Key												
tbl_reviews												
review_id	reviews	Overall_satisfaction	room_id									
Primary Key			Foreign Key									
tbl_bedroom												
bedroom_id	bedrooms	accommodates	minstay	price	room_id							
Primary Key					Foreign Key							

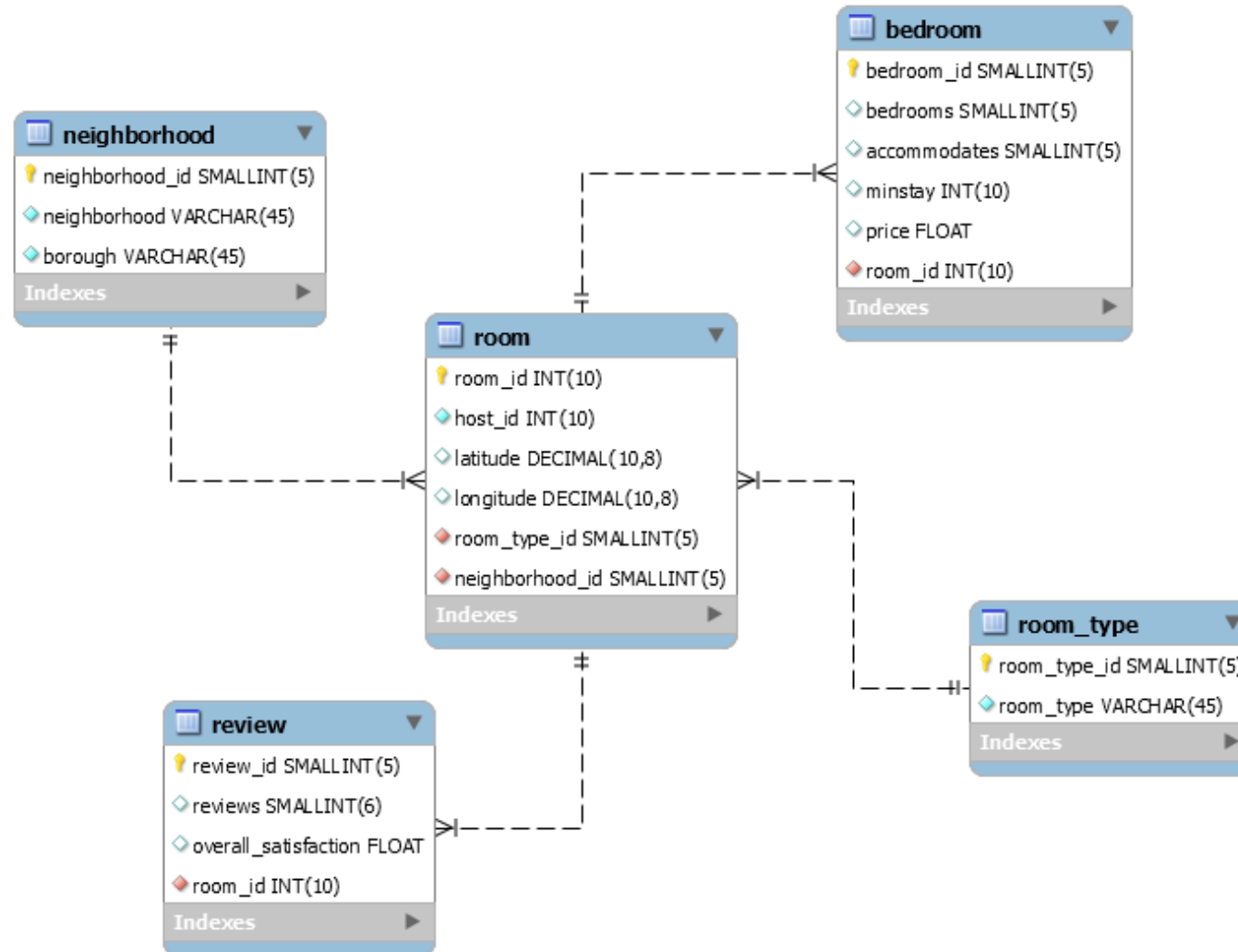


# Normalization Process - Continued

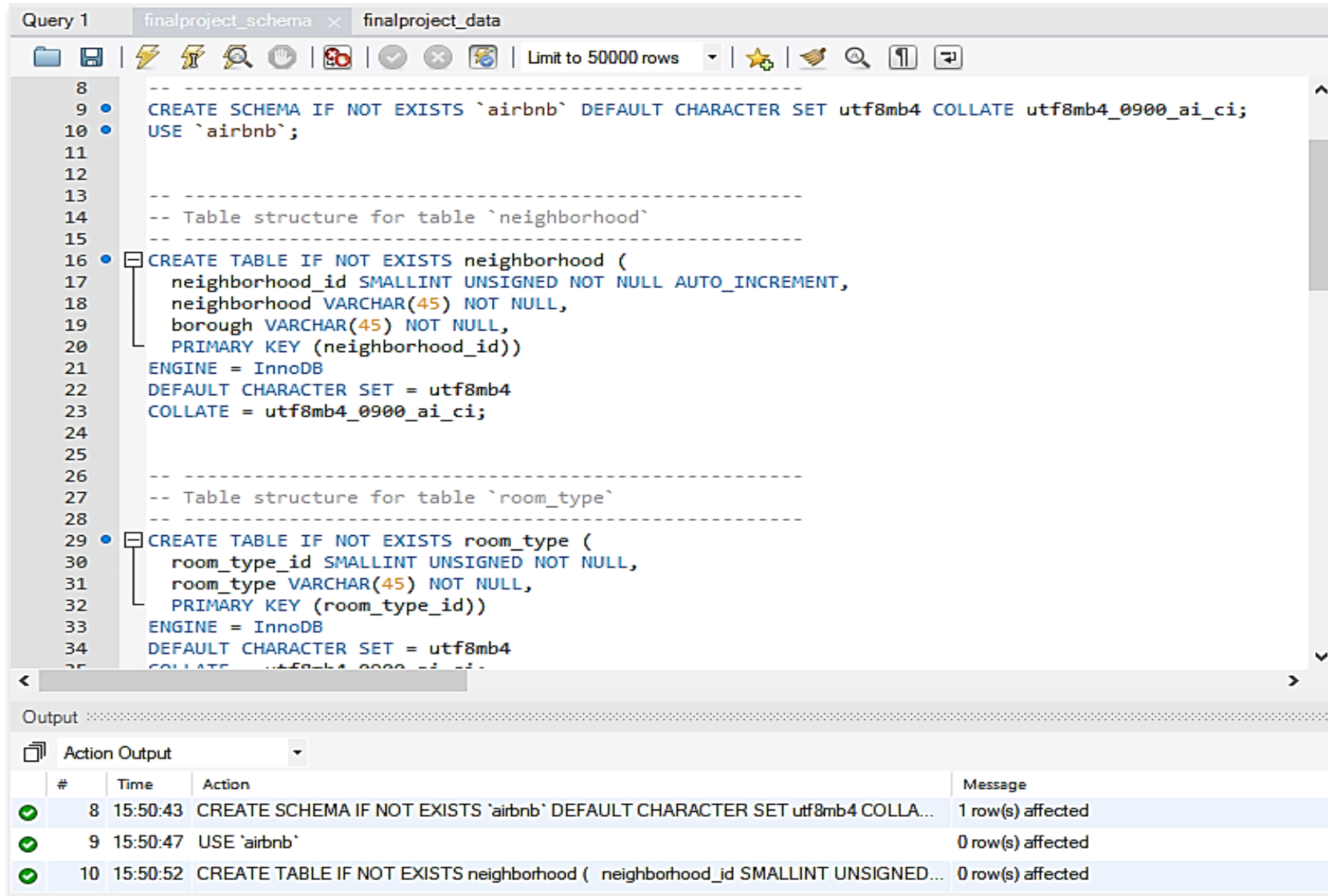
## 3 NF

<b>tbl_room</b>					
room_id	host_id	room_type_id	neighborhood_id	latitude	longitude
Primary Key		Foreign Key	Foreign Key		
<b>tbl_bedroom</b>					
bedroom_id	bedrooms	accommodates	minstay	price	room_id
Primary Key					Foreign Key
<b>tbl_review</b>					
review_id	reviews	overall_satisfaction	room_id		
Primary Key			Foreign Key		
<b>tbl_room_type</b>					
room_type_id	room_type				
Primary Key					
<b>tbl_neighborhood</b>					
neighborhood_id	neighborhood	borough			
Primary Key					

# Entity Relationship Diagram



# Building Table



The screenshot shows a database query editor with a toolbar at the top. The main text area contains SQL code for creating a schema and two tables. The code is as follows:

```
8
9 CREATE SCHEMA IF NOT EXISTS `airbnb` DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci;
10 USE `airbnb`;
11
12
13 -----
14 -- Table structure for table `neighborhood`
15 -----
16 CREATE TABLE IF NOT EXISTS neighborhood (
17     neighborhood_id SMALLINT UNSIGNED NOT NULL AUTO_INCREMENT,
18     neighborhood VARCHAR(45) NOT NULL,
19     borough VARCHAR(45) NOT NULL,
20     PRIMARY KEY (neighborhood_id))
21 ENGINE = InnoDB
22 DEFAULT CHARACTER SET = utf8mb4
23 COLLATE = utf8mb4_0900_ai_ci;
24
25 -----
26 -- Table structure for table `room_type`
27 -----
28
29 CREATE TABLE IF NOT EXISTS room_type (
30     room_type_id SMALLINT UNSIGNED NOT NULL,
31     room_type VARCHAR(45) NOT NULL,
32     PRIMARY KEY (room_type_id))
33 ENGINE = InnoDB
34 DEFAULT CHARACTER SET = utf8mb4
35 COLLATE = utf8mb4_0900_ai_ci;
```

Below the query editor is an "Output" section with a dropdown menu set to "Action Output". It displays a table of execution results:

#	Time	Action	Message
8	15:50:43	CREATE SCHEMA IF NOT EXISTS `airbnb` DEFAULT CHARACTER SET utf8mb4 COLLA...	1 row(s) affected
9	15:50:47	USE `airbnb`	0 row(s) affected
10	15:50:52	CREATE TABLE IF NOT EXISTS neighborhood ( neighborhood_id SMALLINT UNSIGNED...	0 row(s) affected



# Uploading Data

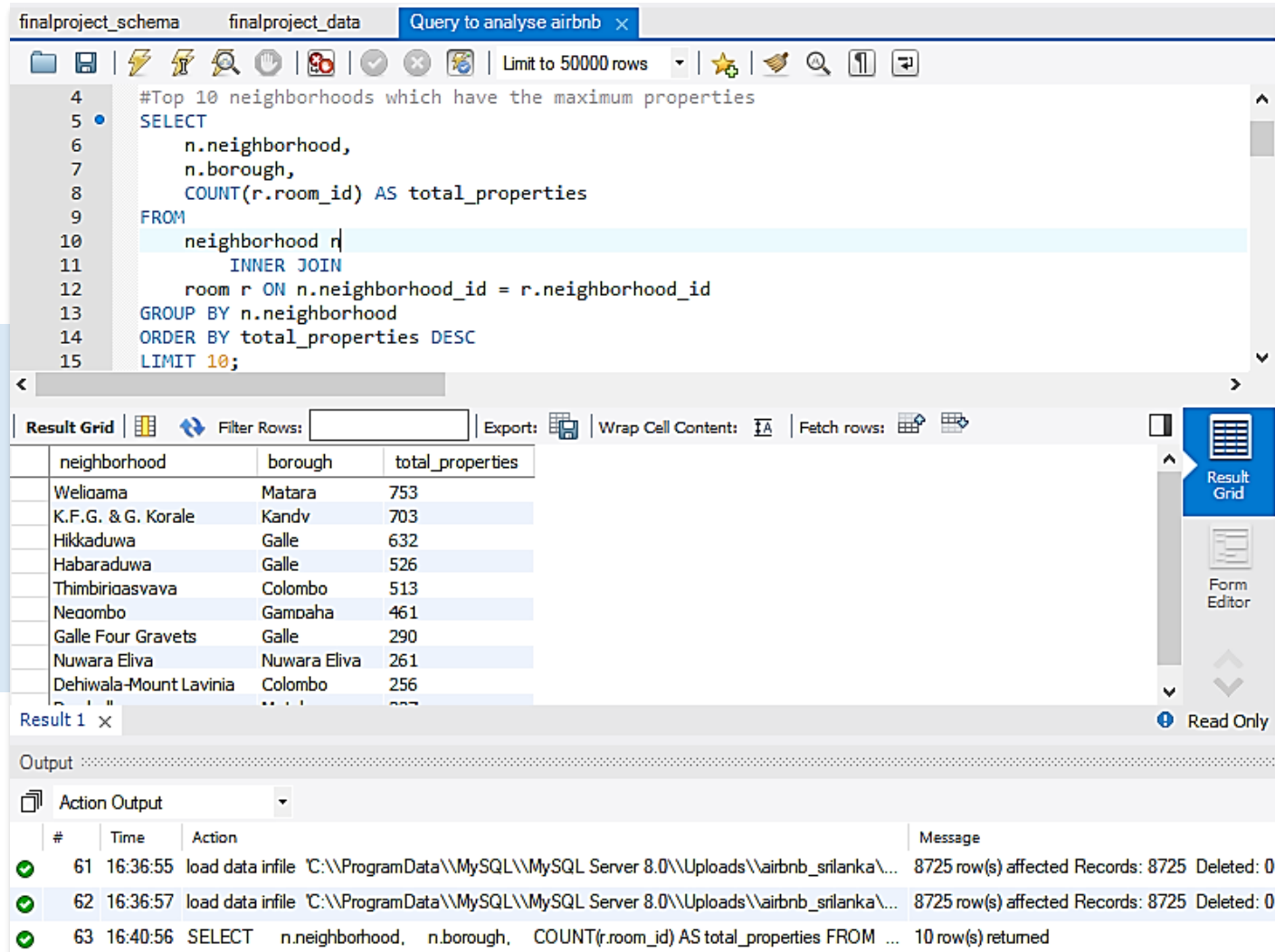
The screenshot displays the MySQL Workbench interface. The top toolbar includes icons for file operations, execution, and navigation. The main editor window shows a SQL script with the following content:

```
1 • USE airbnb;
2
3 • SHOW VARIABLES LIKE "secure_file_priv";
4
5 #Loading the values in 'neighborhood' table
6 • load data infile
7 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\airbnb_srilanka\\neighborhood.csv'
8 into table neighborhood
9 fields terminated by ','
10 lines terminated by '\\n'
11 ignore 1 rows;
12
13
14 #Loading the values in 'room_type' table
15 • load data infile
16 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\airbnb_srilanka\\room_type.csv'
17 into table room_type
18 fields terminated by ','
19 lines terminated by '\\n'
20 ignore 1 rows;
21
22
23 #Loading the values in 'room' table
24 • load data infile
25 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\airbnb_srilanka\\room.csv'
26 into table room
```

The right sidebar contains the 'SQLAdditions' panel with a search bar and a list of snippets. The bottom panel shows the 'Output' window with the 'Action Output' tab selected, displaying a log of successful data loading operations.

#	Time	Action	Message
60	16:36:51	load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\airbnb_srilanka\\...	8725 row(s) affected Records: 8725 Deleted: 0 Skipped: 0 Warnings: 0
61	16:36:55	load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\airbnb_srilanka\\...	8725 row(s) affected Records: 8725 Deleted: 0 Skipped: 0 Warnings: 0
62	16:36:57	load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\airbnb_srilanka\\...	8725 row(s) affected Records: 8725 Deleted: 0 Skipped: 0 Warnings: 0

# Query 1: Top 10 neighborhoods which have the maximum properties



The screenshot shows a database query editor with a SQL query and its results. The query is as follows:

```
#Top 10 neighborhoods which have the maximum properties
SELECT
  n.neighborhood,
  n.borough,
  COUNT(r.room_id) AS total_properties
FROM
  neighborhood n
  INNER JOIN
  room r ON n.neighborhood_id = r.neighborhood_id
GROUP BY n.neighborhood
ORDER BY total_properties DESC
LIMIT 10;
```

The results are displayed in a table with the following columns: neighborhood, borough, and total\_properties. The top 10 neighborhoods are listed below:

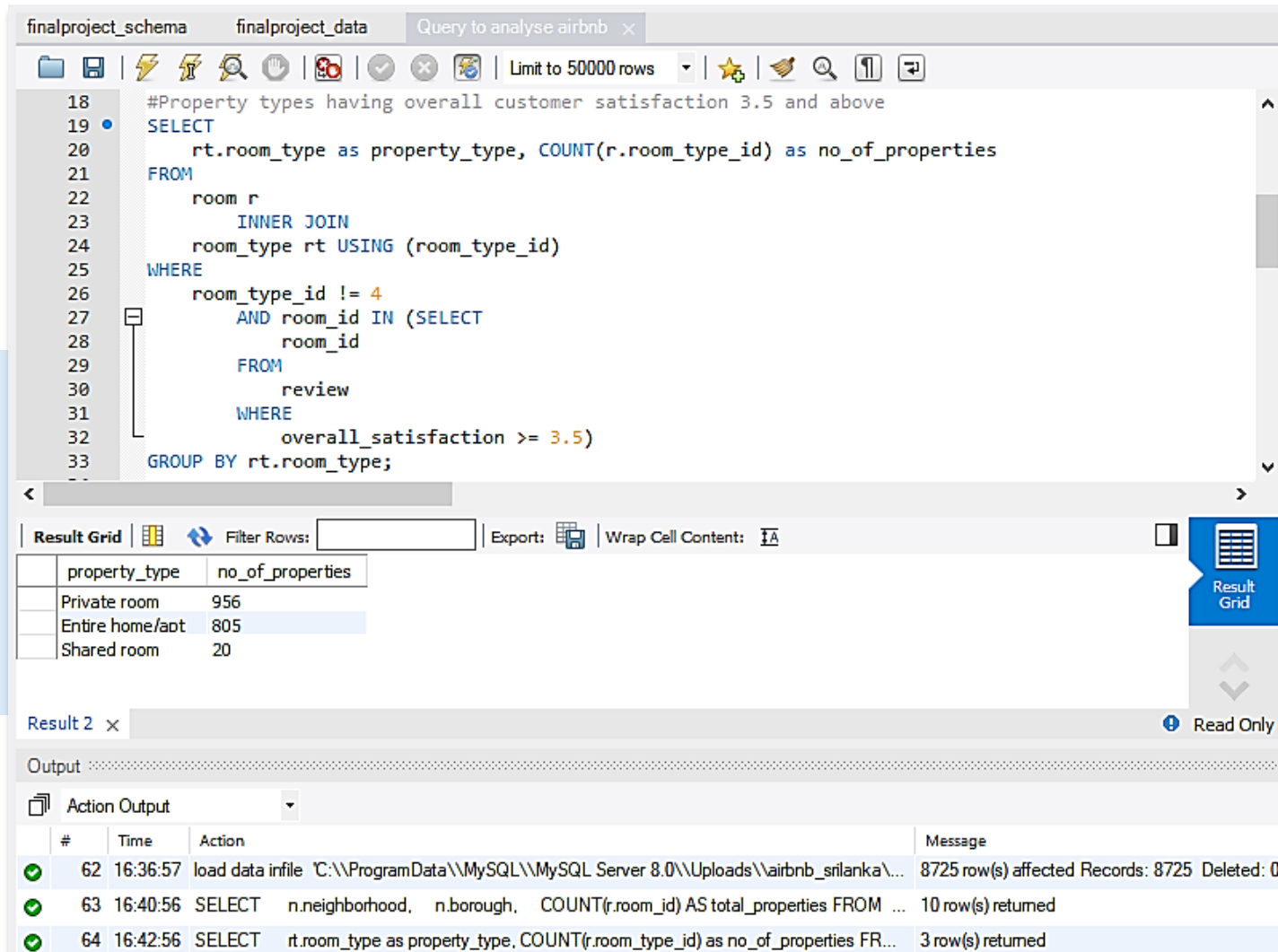
neighborhood	borough	total_properties
Weligama	Matara	753
K.F.G. & G. Korale	Kandy	703
Hikkaduwa	Galle	632
Habaraduwa	Galle	526
Thimbiriaasvava	Colombo	513
Neoombo	Gampaha	461
Galle Four Gravets	Galle	290
Nuwara Eliva	Nuwara Eliva	261
Dehiwala-Mount Lavinia	Colombo	256
...	...	...

The bottom of the screenshot shows the 'Output' section with 'Action Output' and a table of actions and messages:

#	Time	Action	Message
61	16:36:55	load data infile 'C:\ProgramData\MySQL\MySQL Server 8.0\Uploads\airbnb_srilanka\...	8725 row(s) affected Records: 8725 Deleted: 0
62	16:36:57	load data infile 'C:\ProgramData\MySQL\MySQL Server 8.0\Uploads\airbnb_srilanka\...	8725 row(s) affected Records: 8725 Deleted: 0
63	16:40:56	SELECT n.neighborhood, n.borough, COUNT(r.room_id) AS total_properties FROM ...	10 row(s) returned

The maximum no. of properties are found in Matara, Kandy, Galle, Colombo and Gampaha.

## Query 2: Property types having overall customer satisfaction 3.5 and above



The screenshot shows a MySQL query editor window with the following SQL query:

```
18 #Property types having overall customer satisfaction 3.5 and above
19 SELECT
20     rt.room_type as property_type, COUNT(r.room_type_id) as no_of_properties
21 FROM
22     room r
23     INNER JOIN
24     room_type rt USING (room_type_id)
25 WHERE
26     room_type_id != 4
27     AND room_id IN (SELECT
28                     room_id
29                     FROM
30                     review
31                     WHERE
32                     overall_satisfaction >= 3.5)
33 GROUP BY rt.room_type;
```

Below the query editor, the 'Result Grid' shows the following data:

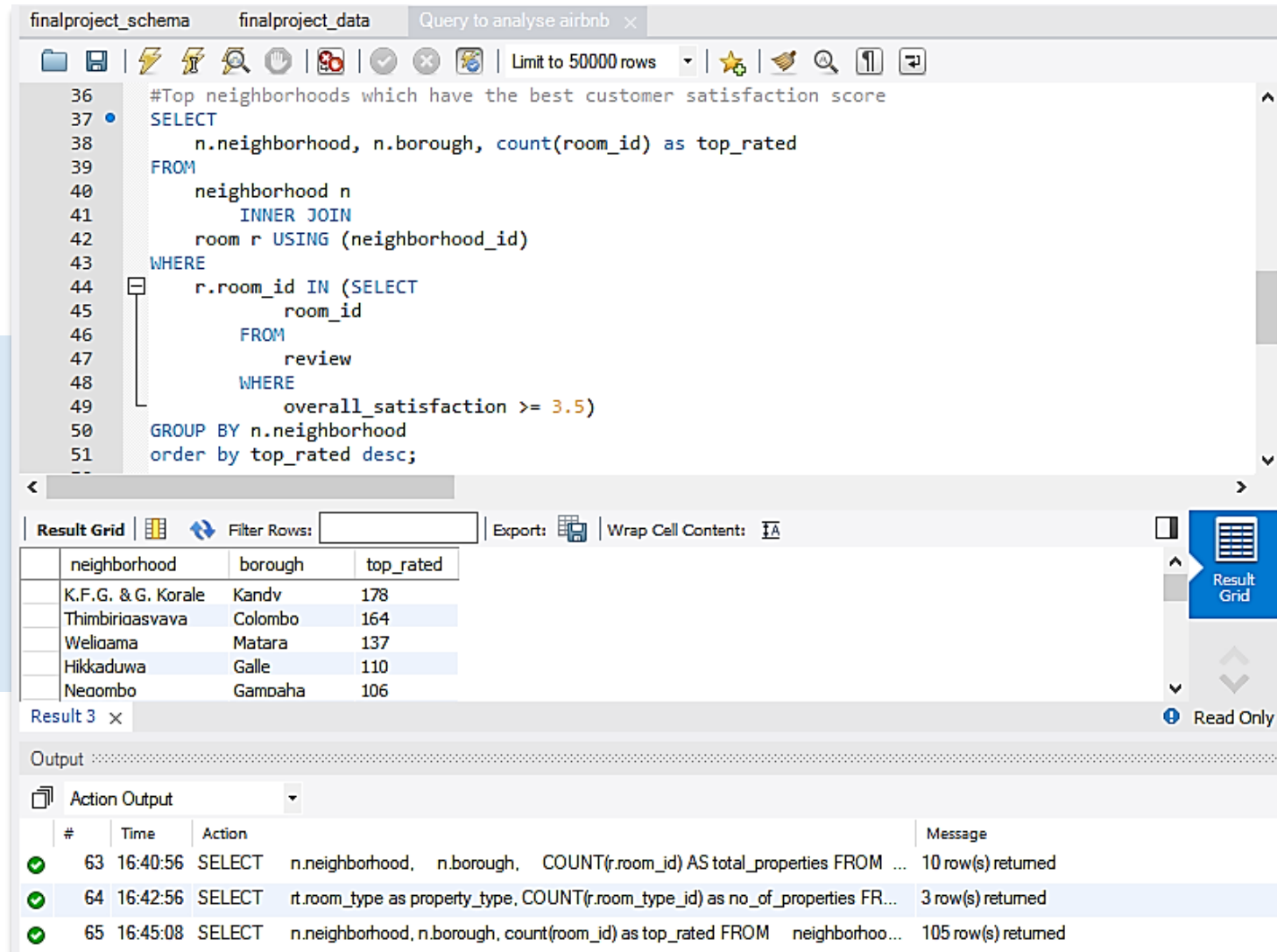
property_type	no_of_properties
Private room	956
Entire home/apt	805
Shared room	20

The 'Output' pane at the bottom shows the 'Action Output' table:

#	Time	Action	Message
62	16:36:57	load data infile 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\airbnb_srilanka\\...	8725 row(s) affected Records: 8725 Deleted: 0
63	16:40:56	SELECT n.neighborhood, n.borough, COUNT(r.room_id) AS total_properties FROM ...	10 row(s) returned
64	16:42:56	SELECT rt.room_type as property_type, COUNT(r.room_type_id) as no_of_properties FR...	3 row(s) returned

Property types 'Private room' and 'Entire home/apt' are highly preferred by the guests

# Query 3: Top neighborhoods which have the best customer satisfaction score



The screenshot shows a SQL query editor with the following query:

```
#Top neighborhoods which have the best customer satisfaction score
SELECT
  n.neighborhood, n.borough, count(room_id) as top Rated
FROM
  neighborhood n
  INNER JOIN
    room r USING (neighborhood_id)
WHERE
  r.room_id IN (SELECT
    room_id
    FROM
      review
    WHERE
      overall_satisfaction >= 3.5)
GROUP BY n.neighborhood
order by top Rated desc;
```

The result grid shows the following data:

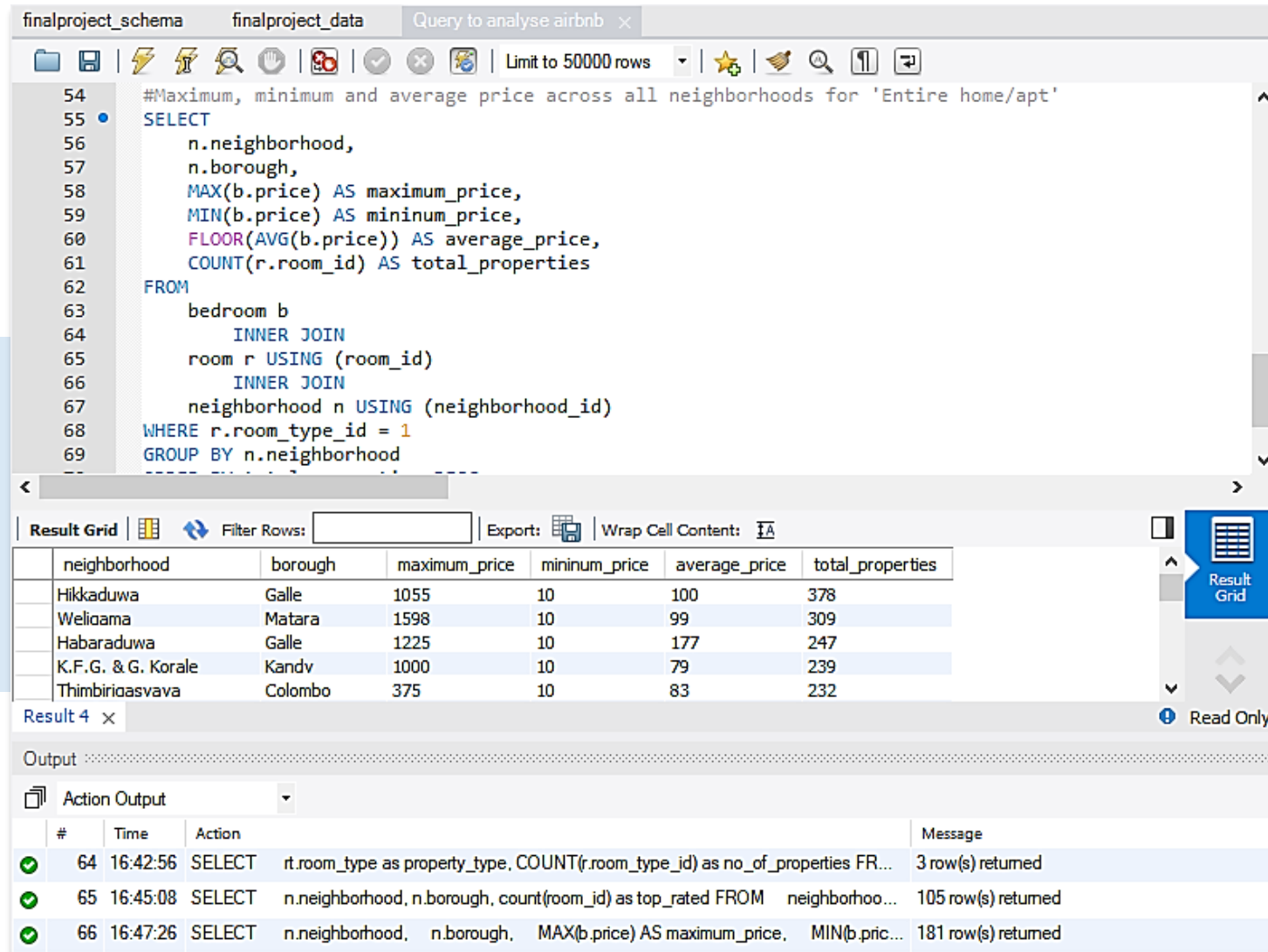
neighborhood	borough	top Rated
K.F.G. & G. Korale	Kandy	178
Thimbiriaasvava	Colombo	164
Weligama	Matara	137
Hikkaduwa	Galle	110
Negombo	Gampaha	106

The action output shows the following results:

#	Time	Action	Message
63	16:40:56	SELECT n.neighborhood, n.borough, COUNT(r.room_id) AS total_properties FROM ...	10 row(s) returned
64	16:42:56	SELECT rt.room_type as property_type, COUNT(r.room_type_id) as no_of_properties FR...	3 row(s) returned
65	16:45:08	SELECT n.neighborhood, n.borough, count(room_id) as top Rated FROM neighborhoo...	105 row(s) returned

Guests highly prefer properties located in Kandy, Colombo, Matara, Galle and Gampaha.

## Query 4: Maximum, minimum and average price across all neighborhoods for 'Entire home/apt'



The screenshot shows a database query editor with a SQL query and its results. The query is as follows:

```
#Maximum, minimum and average price across all neighborhoods for 'Entire home/apt'
SELECT
  n.neighborhood,
  n.borough,
  MAX(b.price) AS maximum_price,
  MIN(b.price) AS minimum_price,
  FLOOR(AVG(b.price)) AS average_price,
  COUNT(r.room_id) AS total_properties
FROM
  bedroom b
  INNER JOIN
  room r USING (room_id)
  INNER JOIN
  neighborhood n USING (neighborhood_id)
WHERE r.room_type_id = 1
GROUP BY n.neighborhood
```

The results are displayed in a table with the following columns: neighborhood, borough, maximum\_price, minimum\_price, average\_price, and total\_properties. The top 5 rows are highlighted in blue.

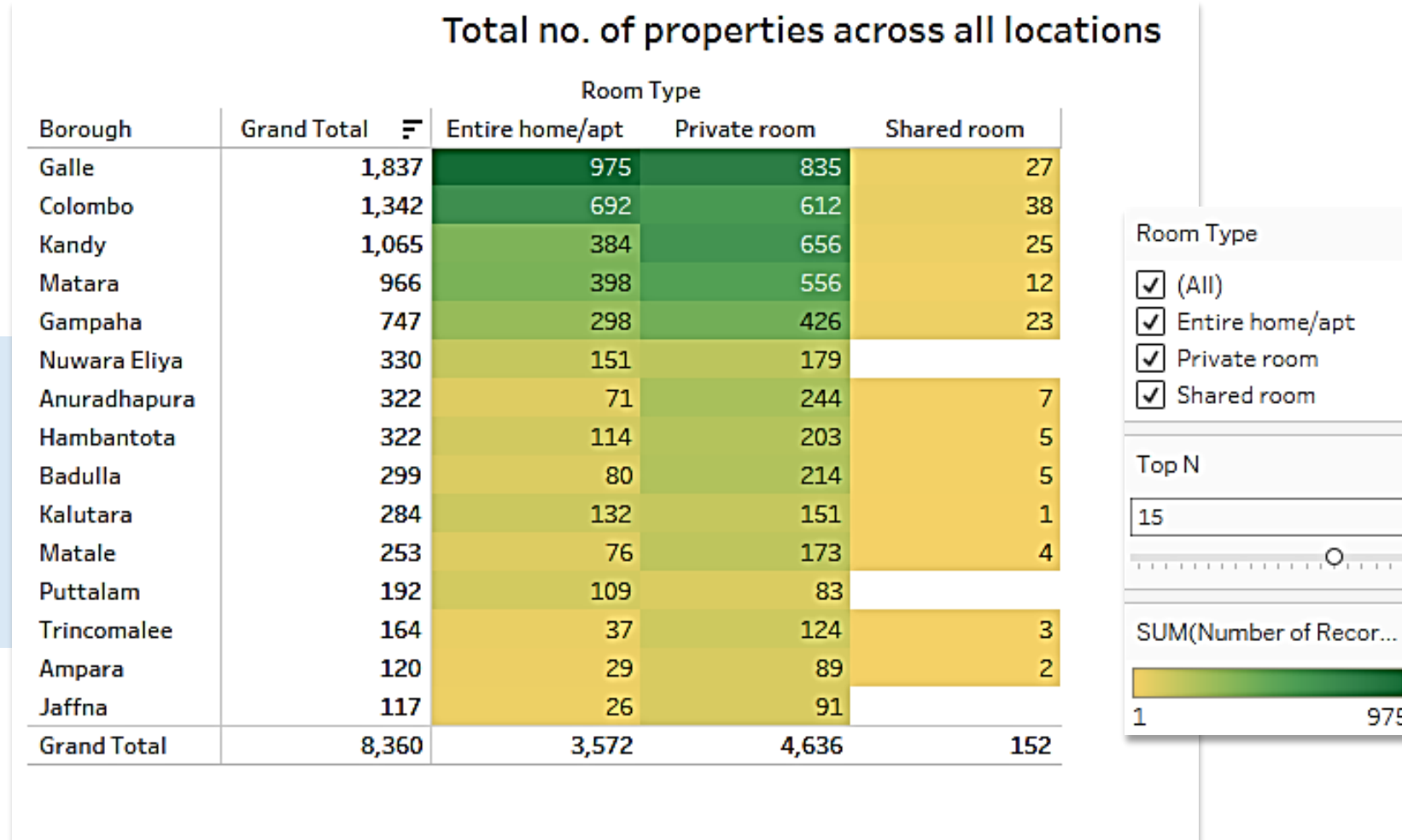
neighborhood	borough	maximum_price	minimum_price	average_price	total_properties
Hikkaduwa	Galle	1055	10	100	378
Welioama	Matara	1598	10	99	309
Habaraduwa	Galle	1225	10	177	247
K.F.G. & G. Korale	Kandv	1000	10	79	239
Thimbiriaasvava	Colombo	375	10	83	232

The bottom section of the screenshot shows the 'Action Output' table, which contains the following rows:

#	Time	Action	Message
64	16:42:56	SELECT rt.room_type as property_type, COUNT(r.room_type_id) as no_of_properties FR...	3 row(s) returned
65	16:45:08	SELECT n.neighborhood, n.borough, count(room_id) as top Rated FROM neighboorhoo...	105 row(s) returned
66	16:47:26	SELECT n.neighborhood, n.borough, MAX(b.price) AS maximum_price, MIN(b.pric...	181 row(s) returned

The average price is around 80-170 for the top 5 preferred locations, with the maximum price can be as high as 1000

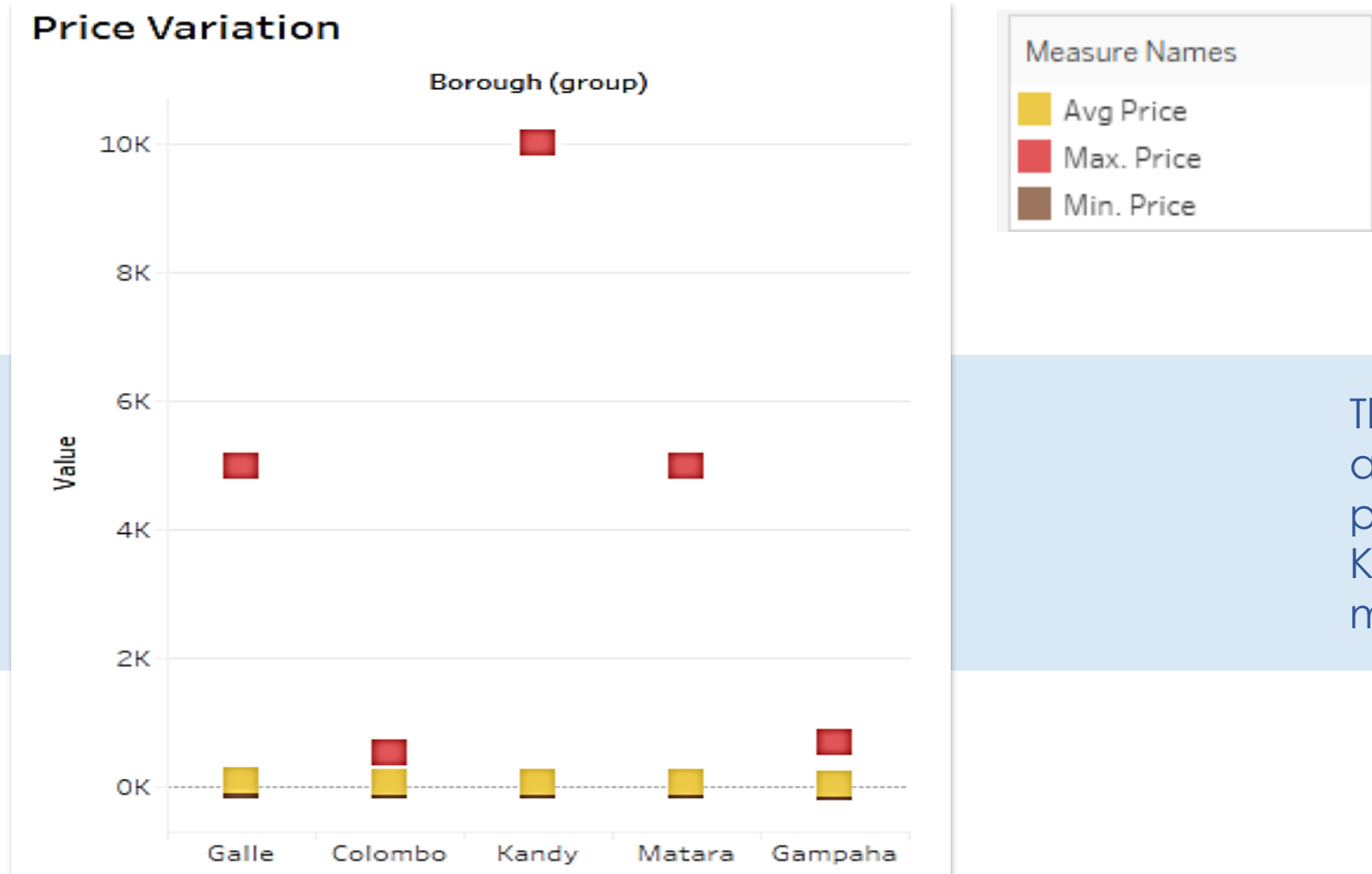
# Tableau Insights 1:



Cross tabulation view:  
Maximum properties are  
located in Galle,  
Colombo, Kandy,  
Matara and Gampaha.



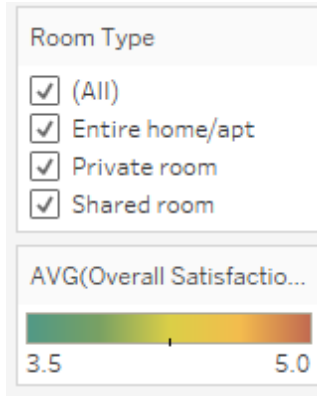
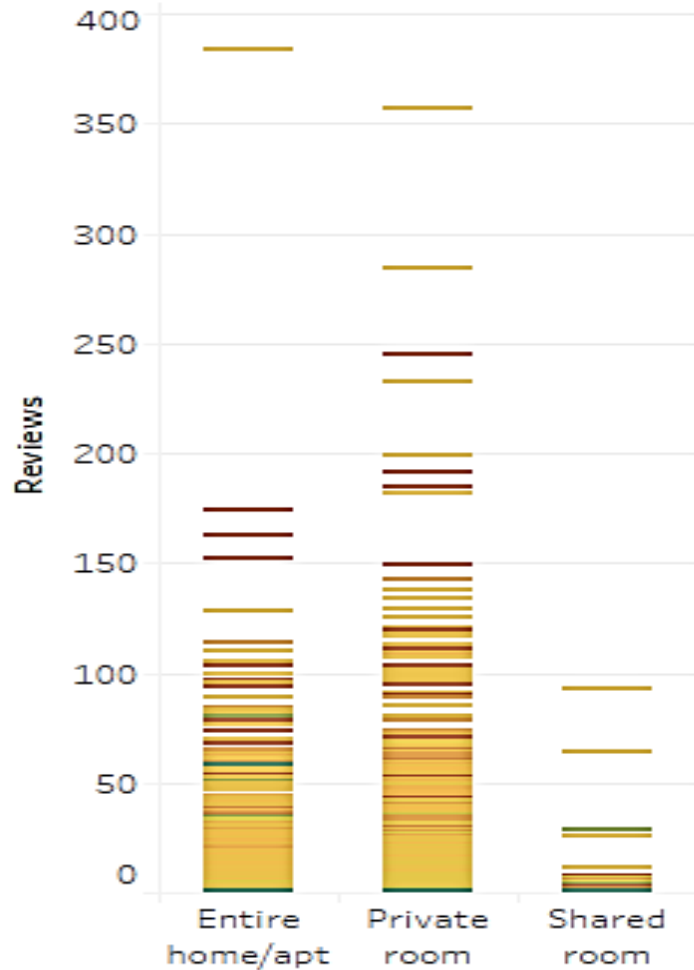
# Tableau Insights 2:



The average price hovers around 100 for the top 5 preferred locations, with Galle, Kandy and Matara having maximum price in 1000s.

# Tableau Insights 3:

**Guest Reviews**



Maximum high guest satisfaction scores are found in 'Private room' followed by 'Entire home/apt'

# Storyboard in Tableau

## Sri Lanka Tourism through Airbnb

### Top Investment Destination

An holistic view on the top 5 investment destination in Sri Lanka that has potential for growth and high returns

#### Top 5 Locations

Borough (g..	Grand Total	Entire home/apt	Private room	Shared room
Colombo	1,342	692	612	38
Galle	1,837	975	835	27
Gampaha	747	298	426	23
Kandy	1,065	384	656	25
Matara	966	398	556	12
Grand Total	5,957	2,747	3,085	125

#### Borough (group)

- ☒ Colombo
- ☒ Galle
- ☒ Gampaha
- ☒ Kandy
- ☒ Matara
- ☐ Others

#### Room Type

All

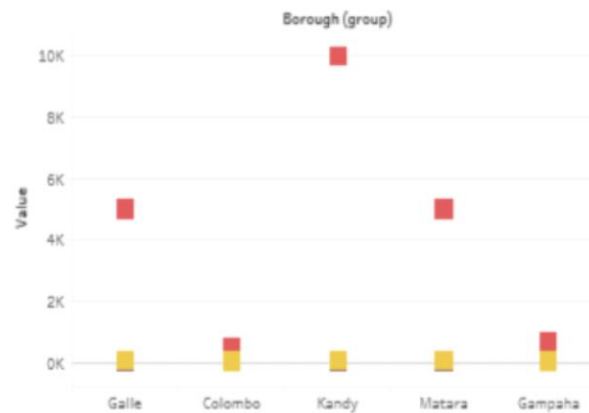
#### Measure Names

- ☒ Avg. Price
- ☒ Max. Price
- ☒ Min. Price

#### Borough



#### Price Variation

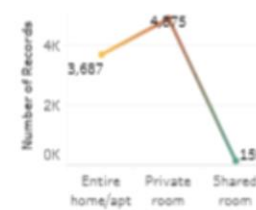


## Sri Lanka Tourism through Airbnb

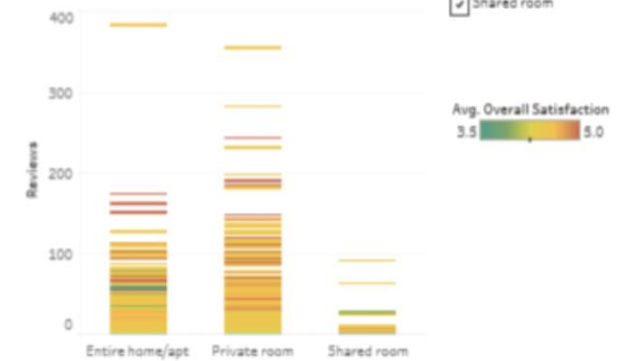
### Preferred Room Type

An holistic view on the preferred room type by tourists and vacationers

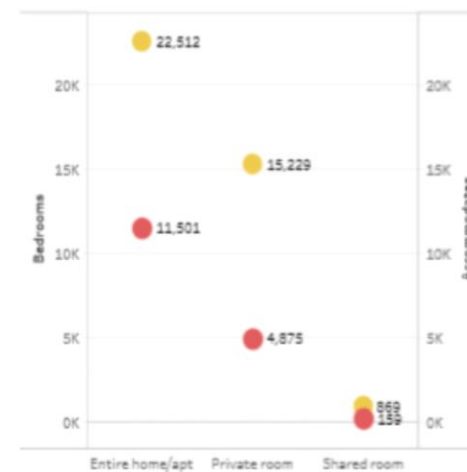
#### Property Type



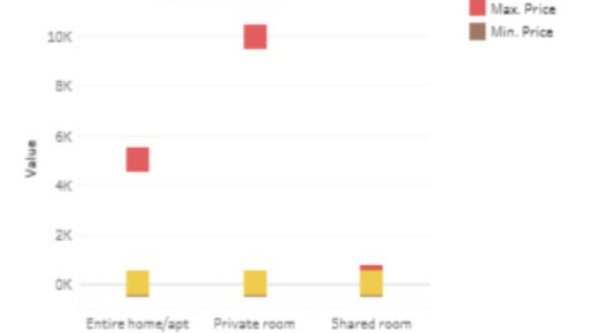
#### Guest Reviews



#### Accommodation Availability Bedroom Vs Accommodates






#### Price Variation



# Recommendations

---

## **To invest in an Airbnb property,**

-  Galle and Colombo looks saturated market. Kandy, Matara and Gampaha locations are developing and has more potential growth.
-  The preferred room type by guests is 'Private room' or 'Entire home/apt'
-  The starting price range can be 100 and above depending on the property type and location

# Limitations

- ⌚ No data available on booking order details to evaluate the ROI of the properties
- ⌚ Seasonal factors, local events and economic outlook need to be considered to estimate the accurate demand
- ⌚ Data on historical property values would aid in better investment destination and predicting future market potential

Thank You

---