Data Questions

Import the data into a sqllite3

First, we need to create a sqlite3 file. We open a terminal and run the next command:

```
sqlite3 airbnb.db
```

After, inside of sqlite3, we need to import our csv files into SQL. To achieve this, we can use the following commands from sqlite3:

```
.mode csv
.import dataset/calendar.csv calendar
.import dataset/listings.csv listings
.import dataset/reviews.csv reviews
```

Questions

- What is the total count of rental units, segmented by home type (entire home, private room, etc)?

A) Query:

```
SELECT room_type AS 'Home type',
    COUNT(*) AS 'Rental Units'
FROM listings
GROUP BY room_type
ORDER BY 2 DESC;
```

B) Answer:

Home type	Rental Units
Entire home/apt	2541
Private room	1160

Home type	Rental Units
Shared room	117

Note: to validate this results we can run the next query

```
SELECT DISTINCT COUNT(*) AS 'Total listings'
FROM listings;
```

With this we obtain the total of listings in the database:

Total listings
3818

- Which neighborhoods of Seattle are the cheapest? Most expensive? Do the ranking change when rental units are segmented by home type?

A) Query to obtain all the list (from the cheapest to the most expensive):

```
SELECT neighbourhood_cleansed AS 'Neighborhood',
    AVG(SUBSTR(price, 2)) AS 'Average price ($)'
FROM listings
WHERE market = 'Seattle'
GROUP BY 1
ORDER BY 2;
```

Query to obtain the cheapest:

```
SELECT neighbourhood_cleansed AS 'Neighborhood',
    AVG(SUBSTR(price, 2)) AS 'Average price ($)'
FROM listings
WHERE market = 'Seattle'
GROUP BY 1
ORDER BY 2
LIMIT 1;
```

Query to obtain the most expensive:

```
SELECT neighbourhood_cleansed AS 'Neighborhood',
    AVG(SUBSTR(price, 2)) AS 'Average price ($)'
FROM listings
WHERE market = 'Seattle'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 1;
```

Query to obtain the rental units are segmented by home type:

```
SELECT room_type AS 'Home type',
    neighbourhood_cleansed AS 'Neighborhood',
    AVG(SUBSTR(price, 2)) AS 'Average price ($)'
FROM listings
WHERE market = 'Seattle'
GROUP BY 1,2
ORDER BY 3;
```

Cheapest:

```
SELECT room_type AS 'Home type',
    neighbourhood_cleansed AS 'Neighborhood',
    AVG(SUBSTR(price, 2)) AS 'Average price ($)'
FROM listings
WHERE market = 'Seattle'
GROUP BY 1,2
ORDER BY 3
LIMIT 2;
```

Most expensive:

```
SELECT room_type AS 'Home type',
    neighbourhood_cleansed AS 'Neighborhood',
    AVG(SUBSTR(price, 2)) AS 'Average price ($)'
FROM listings
WHERE market = 'Seattle'
GROUP BY 1,2
ORDER BY 3 DESC
LIMIT 1;
```

B) Answers.

The cheapest:

Neighborhood	Average price ($)
Rainier Beach	68.5555555555556

The most expensive:

Neighborhood	Average price ($)
Southeast Magnolia	231.705882352941

The cheapest segmented by home type:

Home type	Neighborhood	Average price ($)
Shared room	Greenwood	25.0
Shared room	Industrial District	25.0

The most expensive:

Home type	Neighborhood	Average price ($)
Entire home/apt	Sunset Hill	304.375

- What percentage of hosts have multiple listings?

A) Query:

```
WITH multiple AS (SELECT host_id,
    COUNT(*) AS number_of_listings
FROM listings
GROUP BY 1),

counts AS (SELECT SUM(CASE WHEN number_of_listings > 0 THEN 1 ELSE 0 END) AS total_hosts,
    SUM(CASE WHEN number_of_listings > 1 THEN 1 ELSE 0 END) AS hosts_multiple_listings,
    SUM(CASE WHEN number_of_listings = 1 THEN 1 ELSE 0 END) AS hosts_single_listing
FROM multiple)

SELECT 100.0 * hosts_single_listing/total_hosts AS 'Hosts with a single property (%)',
    100.0 * hosts_multiple_listings/total_hosts AS 'Hosts with multiple properties (%)'
FROM counts
```

B) Answer:

Hosts with a single property (%)	Hosts with multiple properties (%)
83.0607051981098	16.9392948018902

- Who is the host with the greatest number of available options throughout 2016? (across all listings)

A) Query:

```
SELECT host_id AS 'Host ID',
    COUNT(has_availability) AS 'Number of listings'
FROM listings
WHERE has_availability IS 't'
GROUP BY host_id
ORDER BY 2 DESC
LIMIT 1;
```

B) Answer:

Host ID	Number of listings
8534462	46