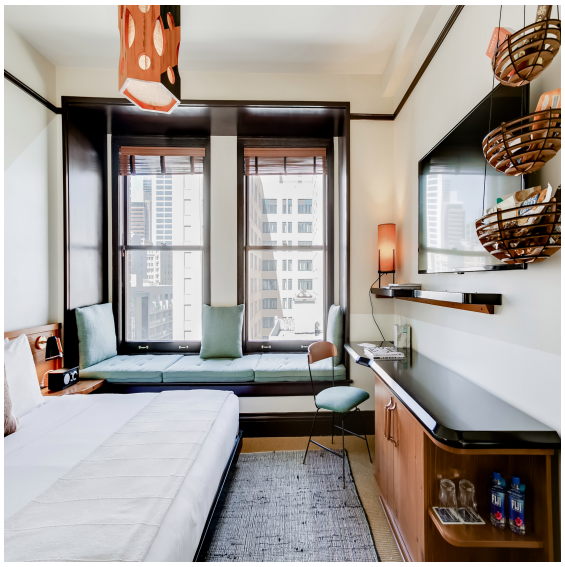
laSalle
UNIVERSITAT RAMON LLULL

Database Systems



Practical Assignment 1 Phase 2
LS Eirbianbi

December 22, 2019

Contents

1 Queries

1.1 Query 1

Show the top 3 cities in which the average savings percentage between staying 7 individual days or staying a week on the apartments with a verified host is bigger.

	name	savings_percentage
1	Cranbourne South	79.22077922077922
2	Wattle Glen	64.28571428571429
3	Doncaster East	58.38509316770186

1.2 Query 2

Show the Guesthouse that have the most expensive price per square feet and also has at least 200 reviews.

	name	price_m2	reviews
1	The Stables, Island of Richmond	$0.37	538

1.3 Query 3

A group of 6 (accommodates = 6) friends want to rent an apartment for 5 days **(and 5 nights)** on Porth Phillip neighbourhood. The apartment must meet the following conditions:

- It must have a balcony.
- It must have more than 1.5 bathrooms.
- The host response rate must be more than 90%.

Find the apartment with the cheapest total price. The total price is rent price multiplied by the number of people plus the cleaning fee and 10% of the deposit.

	name	url	price
1	Spacious Designer Apartment near Albert Park	https://www.airbnb.com/rooms/12528535	$10,050.00

1.4 Query 4

The policy of superhosts has changed on Eirbienbi. Update the database by promoting all users that were registered 5 or more years ago to superhost and demote all the others.

	superhosts
1	4957

	normal_hosts
1	10028

1.5 Query 5

Find the 3 streets that have the biggest number of apartments as long as their average price is lower than 100$.

	street	num	price
1	Brunswick, VIC, Australia	437	$94.80
2	Footscray, VIC, Australia	159	$90.69
3	Preston, VIC, Australia	140	$94.29

1.6 Query 6

Eirbienbi wants to erradicate fake reviews. To do so, they want us to find the 3 people that have done more reviews on the same apartment. Check the reviews on the link of the apartment and choose one of the three users who you think is creating fake reviews and why.

	name_reviewer	url	num_reviews
1	Cameron	https://www.airbnb.com/rooms/7842455	81
2	Therese	https://www.airbnb.com/rooms/1995746	24
3	Michael	https://www.airbnb.com/rooms/9263414	23

1.7 Query 7

Two friends have a budget of 5000$ for a trip of 2 days. Show the most expensive apartments they are able to rent taking into account that:

- The apartment must be for two people (for 2 or more accommodates) and at least it must have 2 beds.
- It must be on Saint Kilda.
- It must have kitchen.
- The host must have the phone verification.

The total price is rent price multiplied by the number of people plus the cleaning fee and 10% of the deposit.

	id	name	price
1	19111	⋆ Beach at your Doorstep: Enjoy Summer in St Kilda	$4,166.00
2	4417	Premium property in great location	$2,000.00
3	20481	Luxury Vintage Melbourne Loft	$1,807.00
4	16656	Gorgeous Stkilda Apartments Sleeps 14 parking	$1,805.00
5	1597	Spacious Designer Apartment near Albert Park	$1,600.00
6	5610	Design House near Saint Kilda	$1,550.00
7	5124	Huge St Kilda Beach PentHouse Beauty	$1,540.00
8	7569	Stylish and Spacious Contemporary Apartment in St Kilda	$1,500.00
9	3260	Family and Large Group Parking WIFI Exc Location	$1,436.00
10	16086	Modern Apart Opposite Albert Park with Breakfast	$1,360.00

From now on, you will realize that you are not able to perform the queries only with one statement. To do so, you will probably need to create auxiliary tables that may help you.

1.8 Query 8

Show the top 3 users calculating his score with the next formula:

$$\sum \frac{1}{apartment\_price_i} * (1 + is\_superhost) * num\_of\_verifications * num\_of\_apartments$$

	name	score
1	Valeria	1146.059229792009
2	Jared SSP	1105.9727272741109
3	Adji	947.2764303036748

1.9 Query 9

Show the top 10 reviewers with the most points taking into account that each review shorter than 100 characters counts for 10 points and each review longer (or equal) than 100 characters counts for 15 points.

	name	points
1	Laurie	1505
2	Michael	715
3	Thanh	540
4	Andrew	520
5	Suzanne	475
6	Anders	475
7	John	435
8	Andrew	425
9	Jon	425
10	Richard	410

1.10 Query 10

You can build up any query you think it would be interesting to analyze, there are no restrictions. The mark of this query will be proportional to the importance of the analyzed aspect.

2 Assignment

The principal task is to structure the data provided in a rudimentary way into a relational model, intuitive, scalable and normalized.
You will need to develop:

- **A conceptual model**
- **A relational model**
- **A physical model**
- **A script to import the data provided into the constructed model**
- **A script with all demanded queries. The output columns must be the same as the ones that appear on the images.**

The deadline of the assignment will be on January 10th. The delivery name will follow the next format:

Fase2_login1_login2.zip

The .zip must contain the documentation of the project, a script to create the structure of the relational database, a script that imports all the data to this structure and another script with the queries.
The documentation must contain:

- **Cover with the name and login of the members of the team**
- **Table of contents**
- **Conceptual model**
- **Relational model**
- **Physical model**
- **Importation verification. Prove that the information has been sucsessfully imported into your proposed database.**
- **Queries and their verification. Add complementary queries to verify the result of the first one and add screen-shots of the result of each query.**
- **Conclusions**

In the queries verification section you must prove that the queries are correct. To do so, you will probably need auxiliar queries that support your results.

3 Considerations and mandatory requirements

Here is a list of the things you must accomplish to opt to pass.

- The model must be normalized
- You must justify each table of your design
- You must justify the chosen decisions while importing the data to your structure. (Ex: At this point we could choose between importing all the data even though it was incomplete or importing only the data that was complete. In this query you can see we chose the second option as we are skipping the incomplete rows.)
- The models must be presented in a digital manner.
- The presented document must be in a PDF format, written in a formal language with no spelling errors
- The delivered .ZIP must contain all the mentioned aspects. If there is a missing document or a missing section the student will not pass the assignment.

Tip: Consider investigating what *regexp_split_to_table* tool does.