Conception Phase

Concerning the design of a database for Airbnb use case, the object of the portfolio examination requested by the course **DLBDSPBDM01**.

Student:

Abdelali BARIR

Matriculation Nº92127958

1. Introduction:

1.1 Overview:

Developing a database is always a challenging task that requires careful planning and execution, particularly when you're trying to simulate the database design of a well-known platform like Airbnb. With over 6 million active listings as of late 2022 (Airbnb, 2023). This platform has become the number one choice for property owners seeking to increase their income by offering short-term rentals. The platform has indeed transformed the travel industry by creating a mutually beneficial environment for both hosts and guests. What sets it apart is the privacy it provides, allowing hosts and guests to interact without any external intervention. Furthermore, the platform ensures the security of its users by holding off payment to the host until 24 hours after the guest's arrival, effectively eliminating the risk of fraudulent practices and deceitful hosts. The amount of data generated by the platform is certainly much bigger than the number of active listings we mentioned beforehand.

1.2 Objective:

In this project, we'll try to design a database schema that simulates the one being used by Airbnb.

1.3 Scope of the data:

The data supposed to be stored in this database will be used for both transactional and analytical purposes but since this is an academic project, the only data that will be stored in this database is dummy data generated randomly for testing reasons only.

1.4 Requirements:

1.4.1 Users:

There are two categories of users in the system: administrators and users. Guests and hosts fall under the category of users, while administrators are employees of the IT department. Users have varying levels of access depending on their category. For example, hosts can edit the price of their listing, whereas guests cannot. Figure n°1 establishes a comparison of these access levels.

Administrators have full access to read any information in the database, except for conversations between users. Access to conversations requires written consent from both parties involved, and only for legitimate reasons. Administrators are also authorized to modify user data upon request if the user is unable to do so on their own and after following identity confirmation protocols. The term "data" here refers to any information that the user is authorized to modify based on their category.

1.4.2 Roles

The database should store all property details, specifications, and images, along with the account details of both hosts and guests. When guests search for available stays, it should be clear if a property is already reserved for the specified dates. Additionally, all fees associated with the property, including any additional costs must be showed to the guest. All conversations between hosts and guests must be saved, as they are in contact throughout the booking process. Guests are responsible for payment and are entitled to a refund if the property does not meet the specifications presented on the platform. After the stay, both guests and hosts should write reviews to help the community make informed choices about their bookings, guests, and hosts. To generate revenue, the platform charges a transaction fee of 3% to hosts and 10% to guests (excluding VAT). An archive of past bookings is necessary to keep track of all transactions. At any given time, both users are encouraged to seek the administrators for help or to report some inappropriate activities. This is to keep the community safe and secure place for everyone.

2. **Data modelling**:

2.1 Conceptual data model:

Based on the requirements section, Figure n°2 illustrates the conceptual data model for this database and highlight the cardinalities between its entities.

2.2 Normalization the data model:

The model was improved to meet the requirements of the first normal form (1NF) ensuring fast information retrieval and query execution. Figure n°3 demonstrates the entities of our database along with their attributes and corresponding datatypes and short description.

2.3 Logical data model:

This model comes to summarize the previous efforts. Figure n°4 demonstrates the logical data model of the database.

3. Conclusion:

The conception phase has been successful in laying the foundation for the development of the database, by covering the basics aspects. the conception data model helped shape the form of the database, the normalization process helped creating convenient attributes and assigning the primary and foreign keys, which lead eventually to creating the logical data model.

In the next phase, the testing data will be used to monitor the behaviour of the database to adjust or further improve the design.

Figure n°1: Access levels of users.

	Guest	*Host*
Confirm their booking and access their booking history.	X	
Check their spending on bookings.	X	
Make, edit, or even cancel their reservations.	X	
Modify information of properties.		X
Check revenue generated by their properties.		X
Check booking history of their properties.		X
Edit his personal information.	X	X
Check their conversations, start new ones, or even delete their copy of it.	X	X
Read information of properties.	X	X
Read public information about guests.	X	X
Read public information about hosts.	X	X

Figure n°2: Conceptual data model of the database.

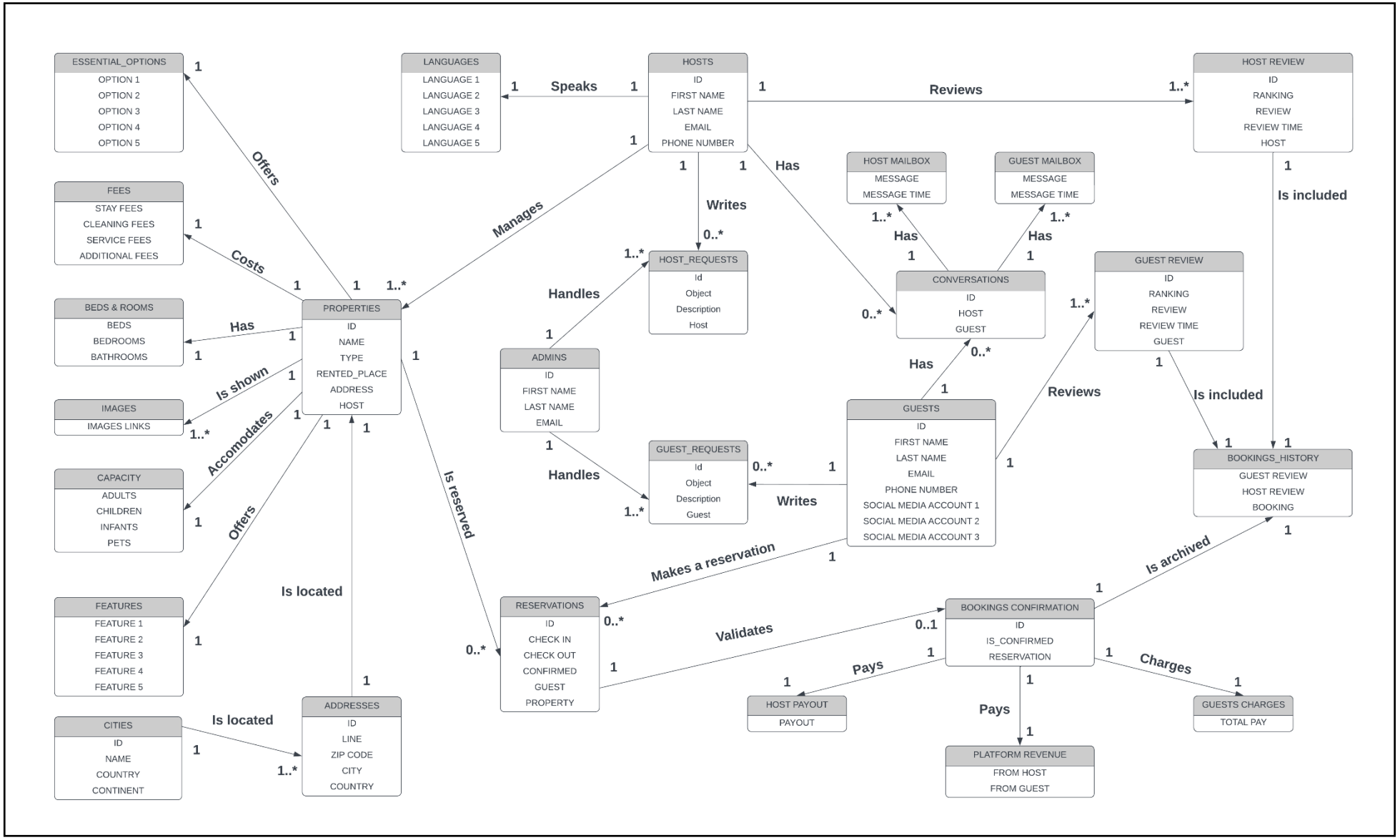


Figure 3: Table of the entities along their attributes, datatypes, primary and foreign keys.

Entity	Column	Datatype	Description
Hosts	HOST_ID (*)	SERIAL	Unique identifier for every host.
	FIRST_NAME	VARCHAR	The first name of the host.
	LAST_NAME	VARCHAR	The last name of the host.
	EMAIL	VARCHAR	The email of the host.
	PHONE_NUMBER	VARCHAR	The phone number of the host.
	HOST_TYPE	VARCHAR	The host can be a particular/professional.
	HOST_STATUS	VARCHAR	The host can be regular, super, or plus.
Guests	GUEST_ID (*)	SERIAL	Unique identifier for every guest.
	FIRST_NAME	VARCHAR	The first name of the guest.
	LAST_NAME	VARCHAR	The last name of the guest.
	EMAIL	VARCHAR	The email of the guest.
	PHONE_NUMBER	VARCHAR	The phone number of the guest.
	FACEBOOK_ACCOUNT	VARCHAR	Link to the Facebook account of the guest, could be left empty.
	GOOGLE_ACCOUNT	VARCHAR	Link to the Google account of the guest, could be left empty.
	INSTAGRAM_ACCOUNT	VARCHAR	Link to the Instagram account of the guest, could be left empty.
Hosts_languages	ID (*)	SERIAL	Unique identifier of every record in the entity.
	FIRST_LANGUAGE	VARCHAR	The main language spoken by the host.
	SECOND_LANGUAGE	VARCHAR	Secondary language spoken by the host, it could be left empty.
	THIRD_LANGUAGE	VARCHAR	Secondary language spoken by the host, it could be left empty.
	FOURTH_LANGUAGE	VARCHAR	Secondary language spoken by the host, it could be left empty.
	FIFTH_LANGUAGE	VARCHAR	Secondary language spoken by the host, it could be left empty.
	HOST_ID (**)	SMALLINT	Identifier referring to the host.
Hosts_reviews	HOST_REVIEW_ID (*)	SERIAL	Unique identifier of a host's review.
	RANKING	SMALLINT	On a scale from 1 to 10, 1 stands for poor and 10 stands for excellent.
	REVIEW	TEXT	The host's thoughts about the guest
	REVIEW_DATE	TIMESTAMP	The date when the review was submitted
	HOST_ID (**)	SMALLINT	Identifier referring to the host.

Entity	Column	Datatype	Description
Guests_reviews	GUEST_REVIEW_ID (*)	SERIAL	Unique identifier of a guest's review.
	RANKING	SMALLINT	On a scale from 1 to 10, 1 stands for poor and 10 stands for excellent.
	REVIEW	TEXT	The host's thoughts about the guest
	REVIEW_DATE	TIMESTAMP	The date when the review was submitted
	GUEST_ID (**)	SMALLINT	Identifier referring to the guest.
Admins	ADMIN_ID (*)	SERIAL	Unique identifier for an administrator.
	FIRST_NAME	VARCHAR	The first name of the administrator.
	LAST_NAME	VARCHAR	The last name of the administrator.
	EMAIL	VARCHAR	The email of the administrator.
Guests_support_requests	GUEST_REQUEST_ID (*)	SERIAL	Unique identifier of a support request made by guests.
	REQUEST_OBJECT	VARCHAR	The object of the request.
	DESCRIPTION	TEXT	Full description of the request.
	GUEST_ID (**)	SMALLINT	Identifier referring to the guest.
	ADMIN_ID (**)	SMALLINT	Identifier referring to the administrator handling the request.
Hosts_support_requests	HOST_REQUEST_ID (*)	SERIAL	Unique identifier of a support request made by hosts.
	REQUEST_OBJECT	VARCHAR	The object of the request.
	DESCRIPTION	TEXT	Full description of the request.
	HOST_ID (**)	SMALLINT	Identifier referring to the host.
	ADMIN_ID (**)	SMALLINT	Identifier referring to the administrator handling the request.
Conversations	CONVERSATION_ID (*)	SERIAL	Unique identifier of a conversation.
	GUEST_ID (**)	SMALLINT	Identifier referring to the guest.
	HOST_ID (**)	SMALLINT	Identifier referring to the host.
Guests_mailbox	GUEST_MESSAGE_ID (*)	SERIAL	Unique identifier of a message.
	MESSAGE	TEXT	The message sent to a host from a guest.
	MESSAGE_TIMESTAMP	TIMESTAMP	The timestamp of the message.
	CONVERSATION_ID (**)	SMALLINT	Identifier referring to the conversation.
Hosts_mailbox	HOST_MESSAGE_ID (*)	SERIAL	Unique identifier of a message.
	MESSAGE	TEXT	The message sent to a guest from a host.
	MESSAGE_TIMESTAMP	TIMESTAMP	The timestamp of the message.
	CONVERSATION_ID (**)	SMALLINT	Identifier referring to the conversation.

Entity	Column	Datatype	Description
Cities	CITY_ID (*)	SMALLINT	Unique identifier of every city.
	CITY	VARCHAR	Name of the city.
	COUNTRY	VARCHAR	Name of the country.
	CONTINENT	VARCHAR	Name of the continent.
Addresses	ADDRESS_ID (*)	SERIAL	Unique identifier for every address.
	ADDRESS_LINE	VARCHAR	Contains the number and street of the property.
	ZIP_CODE	VARCHAR	Zip code of the property.
	CITY_ID (**)	SMALLINT	Identifier referring to the city.
Properties	PROPERTY_ID (*)	SERIAL	Unique identifier for every property.
	PROPERTY_NAME	VARCHAR	The name of the property.
	PROPERTY_TYPE	VARCHAR	The type of the property.
	RENTED_PLACE	VARCHAR	The place offered for rent.
	ADDRESS_ID (**)	SMALLINT	Identifier referring to the address.
	HOST_ID (**)	SMALLINT	Identifier referring to the host.
Properties_capacities	ID (*)	SERIAL	Unique identifier for every specification.
	ADULTS	SMALLINT	Number of adults to be accommodated.
	CHILDREN	SMALLINT	Number of children to be accommodated.
	INFANTS	SMALLINT	Number of infants to be accommodated.
	PETS	SMALLINT	Number of pets to be accommodated.
	PROPERTY_ID (**)	SMALLINT	Identifier referring to the property.
Properties_essentials	ESSENTIAL_ID (*)	SERIAL	Unique identifier for every specification.
	WIFI	BOOLEAN	True if the property includes the option, otherwise, False.
	WASHER	BOOLEAN	True if the property includes the option, otherwise, False.
	AIR_CONDITIONING	BOOLEAN	True if the property includes the option, otherwise, False.
	DEDICATED_WORKPLACE	BOOLEAN	True if the property includes the option, otherwise, False.
	IRON	BOOLEAN	True if the property includes the option, otherwise, False.
	KITCHEN	BOOLEAN	True if the property includes the option, otherwise, False.
	TV	BOOLEAN	True if the property includes the option, otherwise, False.
	DRYER	BOOLEAN	True if the property includes the option, otherwise, False.
	HEATING	BOOLEAN	True if the property includes the option, otherwise, False.
	PROPERTY_ID (**)	SMALLINT	Identifier referring to the property.

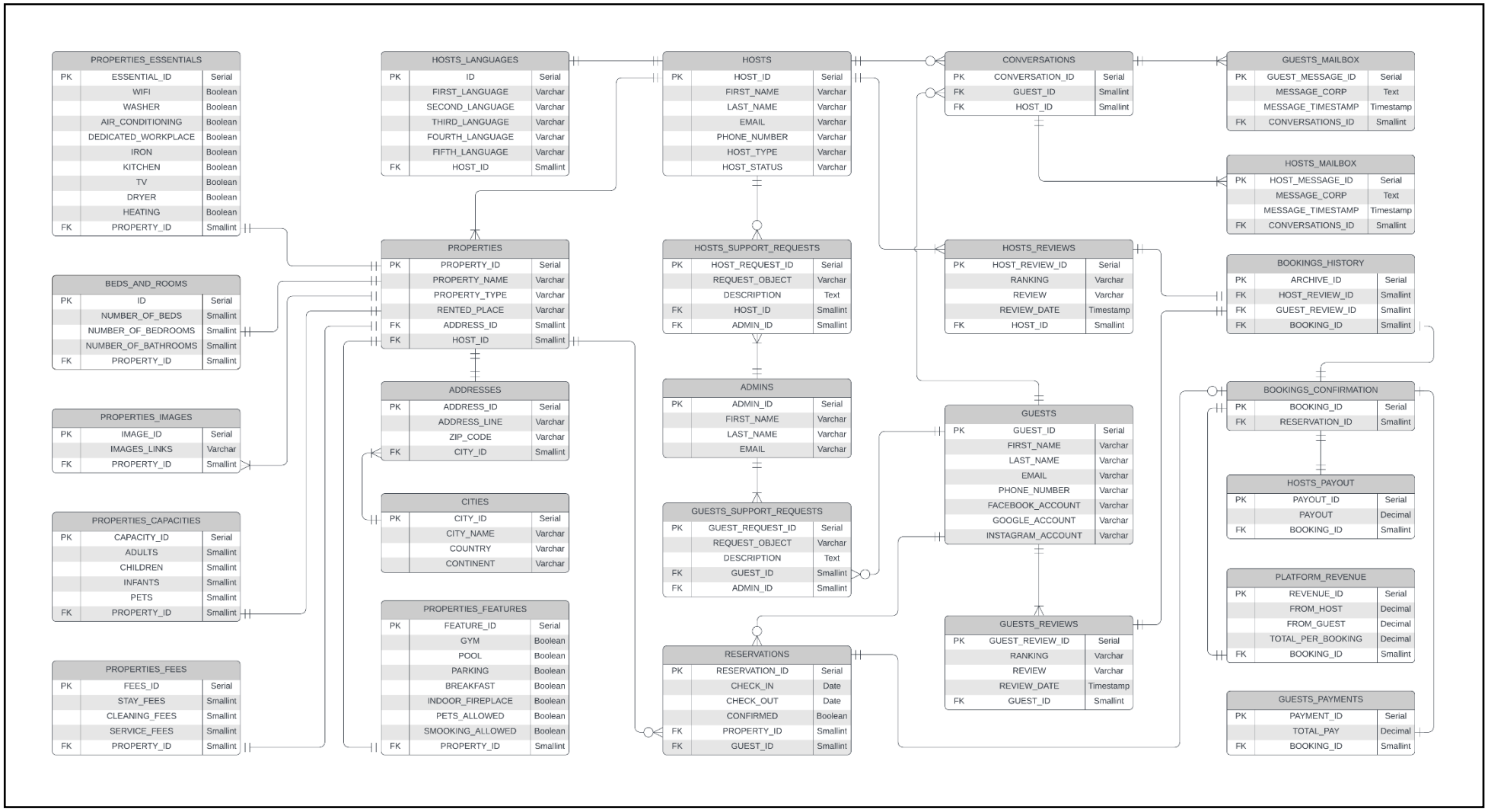
Entity	Column	Datatype	Description
Beds_and_rooms	ID (*)	SERIAL	Unique identifier for every specification.
	NUMBER_OF_BEDS	SMALLINT	Number of beds in the property.
	NUMBER_OF_BEDROOMS	SMALLINT	Number of bedrooms in the property.
	NUMBER_OF_BATHROOMS	SMALLINT	Number of bathrooms in the property.
	PROPERTY_ID (**)	SMALLINT	Identifier referring to the property.
Properties_features	FEATURE_ID (*)	SERIAL	Unique identifier for every specification.
	GYM	BOOLEAN	True if the property includes the option, otherwise, False.
	POOL	BOOLEAN	True if the property includes the option, otherwise, False.
	PARKING	BOOLEAN	True if the property includes the option, otherwise, False.
	BREAKFAST	BOOLEAN	True if the property includes the option, otherwise, False.
	INDOOR_FIREPLACE	BOOLEAN	True if the property includes the option, otherwise, False.
	PETS_ALLOWED	BOOLEAN	True if the property includes the option, otherwise, False.
	SMOKING_ALLOWED	BOOLEAN	True if the property includes the option, otherwise, False.
	PROPERTY_ID (**)	SMALLINT	Identifier referring to the property.
Properties_fees	FEES_ID (*)	SERIAL	Unique identifier for every specification.
	STAY_FEES	SMALLINT	The fees for a one-night stay.
	CLEANING_FEES	SMALLINT	The fees for cleaning services.
	SERVICE_FEES	SMALLINT	The fees for the service provided.
	PROPERTY_ID (**)	SMALLINT	Identifier referring to the property.
Property_images	IMAGE_ID (*)	SERIAL	Unique identifier for every image.
	IMAGE_LINK	VARCHAR	Links to the images located in servers.
	PROPERTY_ID (**)	SMALLINT	Identifier referring to the property.
Reservations	RESERVATION_ID (*)	SERIAL	Unique identifier of a reservation.
	CHECK_IN	DATE	The date of the guest's arrival.
	CHECK_OUT	DATE	The date of the guest's departure.
	CONFIRMED	BOOLEAN	True if the guest finds the accommodation as described, otherwise, False.
	PROPERTY_ID (**)	SMALLINT	Identifier referring to the property.
	GUEST_ID (**)	SMALLINT	Identifier referring to the guest.
Bookings_confirmation	BOOKING_ID (*)	SERIAL	Unique identifier of a booking.
	RESERVATION_ID (**)	SMALLINT	Identifier referring to the reservation.

Entity	Column	Datatype	Description
Platform_revenue	REVENUE_ID (*)	SERIAL	Unique identifier for every revenue.
	FROM_HOST	DECIMAL	The fees paid from the host for the service
	FROM_GUEST	DECIMAL	The fees paid from the guest for the service.
	BOOKING_ID (**)	SMALLINT	Identifier referring to the booking.
Hosts_payout	PAYOUT_ID (*)	SERIAL	Unique identifier for every pay-out.
	PAYOUT	DECIMAL	The fees paid from the platform for the stay, the service fees are deducted.
	BOOKING_ID (**)	SMALLINT	Identifier referring to the booking.
Guests_payments	PAYMENT_ID (*)	SERIAL	Unique identifier for every payment.
	TOTAL_PAY	DECIMAL	The total fees, the guest paid for the stay.
	BOOKING_ID (**)	SMALLINT	Identifier referring to the booking
Bookings_archive	ARCHIVE_ID (*)	SERIAL	Unique identifier for every past booking.
	GUEST_REVIEW_ID (**)	SMALLINT	Identifier referring to the guest's review.
	HOST_REVIEW_ID (**)	SMALLINT	Identifier referring to the host's review.
	BOOKING_ID (**)	SMALLINT	Identifier referring to the booking.

(*) : Primary keys.

(**) : Foreign keys.

Figure 4: Logical data model of the database.



List of figures:

References:

Airbnb. (2023, February 15). Airbnb Q4 2022 and full-year financial results. *Airbnb Newsroom*. https://news.airbnb.com/en-uk/airbnb-q4-2022-and-full-year-financial-results/