**Project – Relational Database Model in 3NF and Physical Database Creation**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Activity** | **Member 1** | **Member 2** | **Member 3** | **Member 4** |
| Prepared Data Model and Created Physical DB | Yes | Yes | Yes | Yes |
| Loaded Data into Database | Yes | Yes | Yes | Yes |

**Contents**

Relational Data Model 3

Assumptions/Notes About Data Entities and Relationships 3

Entity-Relationship Diagram 5

Physical MySQL Database 7

Assumptions/Notes About Data Set 7

Screen shot of Physical Database objects 7

Data in the Database 10

# **Relational Data Model**

## Assumptions/Notes About Data Entities and Relationships

* Assumptions/notes about **Entities and data relationship**:
* List Registration Table has listing id, listing\_url and name and each listing id represents a **unique** Airbnb room which is mentioned in listing\_url
* Each listing id has **unique** Description id associated with it.
* So, **one to one relationship** exists between List Registration Table and Description Table **(Cardinality)**
* Every listing id should have a description id **(Modality = 1)**
* All listing id has experience level as 1. So, **many-to-one relationship** exists Description table and Experience table and every record has in description table has an experience Id i.e. **(Modality= 1)**
* Each promotion id is linked to only one Promotion id. So, **one-to-one Relationship** is there between the two tables. **Modality of both tables is 1** and
* Description id**,** neighborhood\_overview, notes, transit, thumbnail\_url, medium\_url in Promotion table have null values
* Each promotion id is linked to only one Host id. So, **one-to-one Relationship** is there between the two tables. **Modality of both tables is 1**
* In **Host AIRBNB table and HOST Table**, **many to one relationship** exists( because same host can post advertisements on different dates). And the **Modality** of host\_id in Host table can be **1**.
* host\_about, host\_response\_time, host\_response\_rate host\_acceptance\_rate can have null values
* Between **Host AIRBNB table and Location verification table – one to one relationship exists. Modality of 1**
* **Location verification** table **and Property and room type** tablehave **many-to-one relationships.** Every Location id have Property and room type so **MODALITY is 1.**
* **Location verification** table **and room\_type** tablehave **many-to-one relationships.** Each Location id have Property and room type so **MODALITY is 1.**
* **Accommodation availability table and Host AIRBNB table** has one to one relationship. And every record of Accommodation availability table has one row connection, so **Modality is 1.**
* **Accommodation availability table and Bed Type** has **many-to-one relationship. And modality is 1.**
* **FEE AND SERVICE TABLE and Host AIRBNB table** has **one to one relationship with modality is 1.**
* weekly\_price, monthly\_price, security\_deposit, cleaning\_fee has null values
* **DAY NIGHT CAPACITY TABLE and Host AIRBNB table** has **one-to-one** relationship and **modality is 1**
* **TIME PERIOD AVAILABILITY TABLE** and **Host AIRBNB table** has **one-to-one** relationship with **Modality** as **1**.
* **Review Score Table** and **Host AIRBNB table** has **one-to-one** relationship with ***Modality*** *as* ***0*** because all multiple records in **Review Score Table** are **null which** doesn’t provide any significance.
* review\_scores\_rating, review\_scores\_accuracy, review\_scores\_cleanliness, review\_scores\_checkin, review\_scores\_communication, review\_scores\_location, review\_scores\_value, reviews\_per\_month can have null values
* **Review frequency Table** and **Host AIRBNB table** has **one-to-one** relationship with **Modality as 1**
* number\_of\_reviews, first\_review, last\_review have null values
* **Strictness Table** and **Host AIRBNB table** has **one-to-one** relationship with **Modality** as **1**.
* **Review\_table and Reviewer\_table** has one to one relationship
* **Review\_table and LR\_table**  has one to one relationship
* **List Registration table and Listing\_detail table** has one-to-many relationship.

**Include reasons why the data model is in 3NF.**

This meets the 3NF requirement because the non-key attributes are not dependent on any other non-key attribute. Hence, overcoming transitive dependency.

## Entity-Relationship Diagram

Graphical user interface

Description automatically generated

**Assumptions/notes about data:**

***Procedure Followed for normalization***

* The raw data provided in Kaggle has many redundant rows and noise. We first convert the raw data into normalized form
* After converting into 1NF, we make sure to decompose the 3 xlsx files into multiple tables to obey the condition for 2NF. Here, we make sure the non-key attributes solely depend on the primary key or composite primary key
* The last step to filtration is checking for 3NF. The non-key attributes must not be dependent on the other non-key attributes.

**Assumptions/notes about data**

* Assuming that the attributes in the entities which are missing implies that user/host isn’t interested in knowing/providing that info. So, those values can be filled with null values in MySQL Workbench.
* As web scrapping tool is used to get the dataset on Kaggle and it’s a SINGLE value so we can eliminate the column’s scrape\_id, last\_scraped.
* Pictures of Airbnb listing are in listing\_url where images of facilities provided can be found so we can eliminate the Datathumbnail\_url, medium\_url, picture\_url, xl\_picture\_url columns.
* host\_response\_time can be removed as host\_response\_rate and host\_acceptance\_rate provides numerical values for the “Response time of host “
* host\_verifications and host\_has\_profile\_pic are not useful to make decisions.
* neighbourhood, neighbourhood\_cleansed, neighbourhood\_group\_cleansed are not relevant to frame the relations between tables because they provide any useful infoormation
* market, smart\_location, country\_code, country attributes are repeated attributes so using those to make a entity isn’t required. An individual would buy an item from its “City market” which is already in host\_location.
* square\_feet attribute is NULL so didn’t considered it for ER Diagram and many more.
* price attribute doesn’t convey what AIRBNB is charging that amount for
* calendar\_updated (NON- NUMERIC) which doesn’t imply numerical values for How many days before calendar was last updated
* calendar\_last\_scraped isn’t needed as it’s related to web scraping date
* requires\_license, license, jurisdiction\_names, instant\_bookable attributes aren’t useful

# Physical MySQL Database

## Assumptions/Notes About Data Set

Include any assumptions made about data such as empty fields, sparse data, bad data, etc.

## Screen shot of Physical Database objects

The Seattle database has the following Entities:

Graphical user interface, application, Word

Description automatically generated

Data inside the entities

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| Listing Table | listing id | host\_id | 1000 |
| Description Table | Description id | listing id | 1000 |
| Promotion Table | Promotion id | Description id | 1000 |
| Host AIRBNB Table | Host\_ No | host\_id | 1000 |
| Host Table | host\_id | N.A | 769 |
| Location Verification Table | Location no. | Host\_ No, P Id, R Id | 1000 |
| Property and room type table | P. Id | N.A | 12 |
| Room table | R. Id | N.A | 3 |
| Time\_Period\_ availability table | A.Id | Host\_ No, B. Id | 1000 |
| Fee AND Service Table | FS.Id | Host\_ No | 1000 |
| Day Night Capacity Table | DN Id | Host\_ No | 1000 |
| Time Period Availability Table | TP Id | Host\_ No | 1000 |
| Review Score Table | Score Id | Host\_No | 1000 |
| Review frequency Table | RF id | Host\_ No | 1000 |
| Strictness Table | Strictness id | host no. | 1000 |
| listing\_Detail\_table | Listing\_No | Listing\_id | 1000 |
| Review\_table | Review\_No | Reviewer\_id | 1000 |
| Reviewer\_table | Reviewer\_id | Host\_ No | 1000 |
| LR Table | Id | Review\_No | 1000 |
|  |  |  |  |