Final Project Phase-2: Database Design

D-532 Applied Database Final Project

Team: Amit Banerjee, Pravallika Pentapati, Sahil Dhingra.

(Summer 2022)

Table of Contents

Project Title: D-532 Applied Database US Real Estate Listing Site. 3

Team Name: Data Wizards. 3

Project Summary: 3

Schema Design: 3

- Conceptual Data model: 3
- Schema Explanation: 4
- Physical Model: 4

Database Objects: 5

- Constraints: 5
- Additional Objects: 5

Code: 5

- Database Object Code: 5
- Data Load Queries: 5
- Webapp Queries: 5

Assessment Table: 6

Project Title: D-532 Applied Database US Real Estate Listing Site.

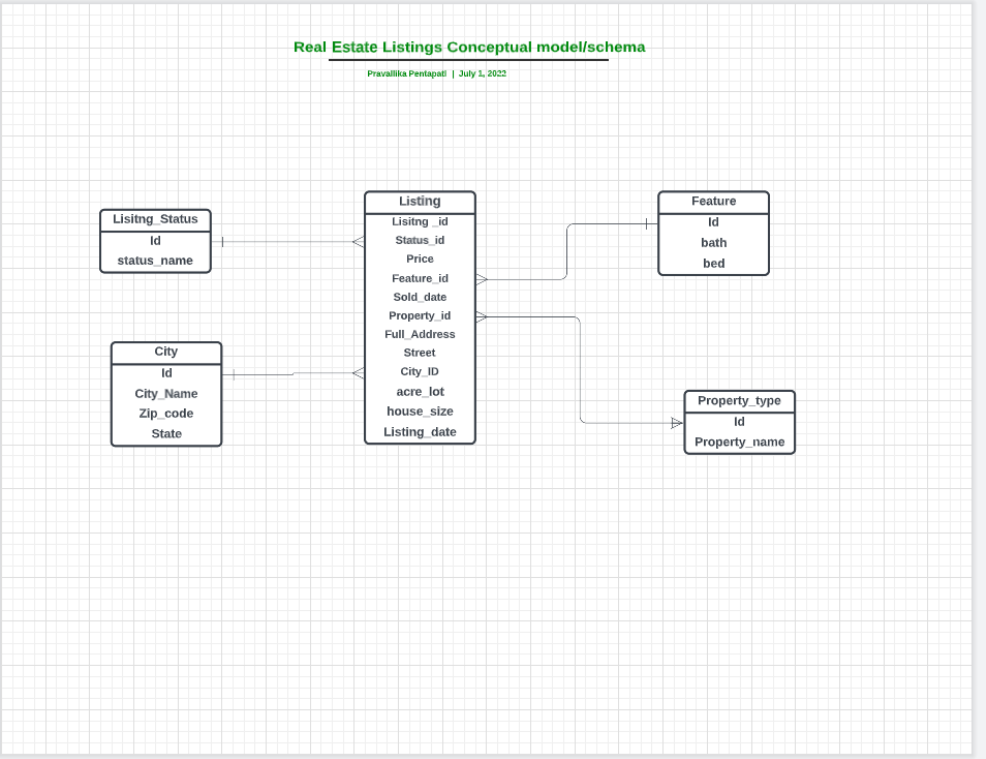
Team Name: Data Wizards.

Project Summary:

As per the statistics from **review42**, 50% of buyers find their new home online. In addition, with the record number of tech-savvy millennials jumping into real estate, this number will only grow. Looking at this interesting fact and using real estate data from Kaggle, our application will showcase the house listings in the US. Users will be able to find their preferred home based on multiple search criteria. In addition, we will provide potential home buyers with relevant statistical visualizations to make an informed decision.

Schema Design:

Conceptual Data model:



Schema Explanation:

The dataset used in this project consists of real estate listings in the US collected from the following Kaggle location https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset. The denormalized dataset contains the following information:

	A	B	C	D	E	F	G	H	I	J	K	L
1	status	price	bed	bath	acre_lot	full_address	street	city	state	zip_code	house_size	sold_date
2	for_sale	105000	3	2	0.12	Sector Yahuecas T	Sector Yal	Adjuntas	Puerto Ric	601	920	
3	for_sale	80000	4	2	0.08	Km 78 9 Carr # 135	Km 78 9 C	Adjuntas	Puerto Ric	601	1527	
4	for_sale	67000	2	1	0.15	556G 556-G 16 St,	556G 556-	Juana Diaz	Puerto Ric	795	748	
5	for_sale	145000	4	2	0.1	R5 Comunidad El	R5 Comun	Ponce	Puerto Ric	731	1800	
6	for_sale	65000	6	2	0.05	14 Navarro, Maya	14 Navarr	Mayaguez	Puerto Ric	680		

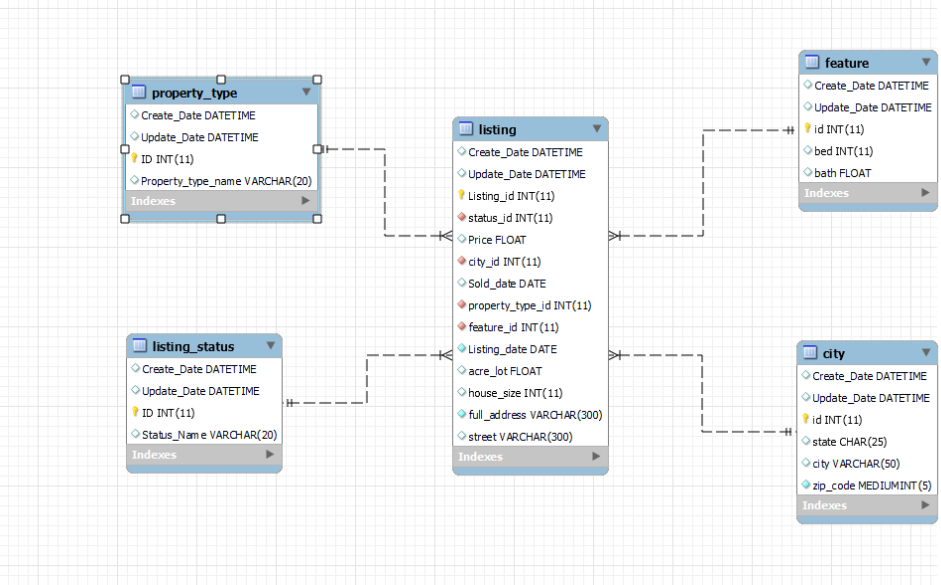
The natural key of this data is the complete address (full_address + street + city + zip_code) and status. To ensure 1NF we make sure that there are no duplicate records for this key. We also add the listing date and type of property to enrich the dataset.

To convert this data to 2NF and 3NF we break down the above dataset into: a Listing dataset, Listing Status dataset, Listing Type (property type) dataset, Feature dataset and City dataset.

The final entities and relationships in the ER model are:

- *Listing* - Identified uniquely by listing_id.
- *Listing Status* – Identified uniquely by status_id and has a 1:N relationship to Listing.
- *Listing Type* - Identified uniquely by listing_type_id and has a 1:N relationship to Listing.
- *Feature* - Identified uniquely by feature_id and has a 1:N relationship to Listing.
- *City* - Identified uniquely by city_id and has a 1:N relationship to Listing.

Physical Model:



Database Objects:

Constraints:

We would define and enable the following constraints in the entities/tables:

Entity/Table	Column(s)	Constraint	Comments
Listing	Listing_id	Primary Key	
	Status_id, City_id, Listing_type_id, Feature_id	Foreign Keys	Additional NOT NULL constraints since all listings should have these.
	Price	Check Constraint	Checks to ensure positive value
	Full_address, Street_address	NOT NULL	
Property Type	Id	Primary Key	
	Property_type_name	Check Constraint	Permits only valid values
Listing Status	Id	Primary Key	
	Status_name	Check Constraint	Permits only valid values
City	Id	Primary Key	
	City, state, zip_code	NOT NULL	
Feature	Id	Primary Key	

Additional Objects:

We are planning to create procedures for the insert, update and delete modules. These procedures would mainly take care of the guardrails around the CRUD process to ensure correctness of the operations.

Code:

Database Object Code:

Create table statements have been prepared by Pravallika Pentapati and available in SECTION 1: CREATE STATEMENTS of the attached file.

Data Load Queries:

Data cleanup and load queries have been prepared by Sahil Dhingra and available in SECTION 3: DATA CLEANUP AND LOAD of the attached file.

Webapp Queries:

Webapp queries have been prepared by Amit Banerjee and available in SECTION 2: WEBAPP QUERIES of the attached file.

All Queries are added in one SQL file which is Final_Project_realtor_data_Part2.sql

Assessment Table:

Team member	Questions	Comments
Pravallika Pentapati	**Are you satisfied with the task completion (scale 1-10)**	Yes, I am satisfied with the task completion. (10)
	Team Work	The three members of our team are collaboratively working together and share the work portions equally. we often connect via Zoom and Team calls whenever required.
	Time Commitment	We worked a good number of hours towards the project deliverables
	what could be done better:	
	Conceptual Schema:	I feel we have designed well our schema diagram with additional attributes.
	Database:	I think we have designed SQL scripts well as per standards like specifying primary keys, and set up the table relationships and constraints.
	code	I think we have designed SQL scripts relevant to web app functionality.