**Final Project Phase-1**

**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](#_Toc737002474)

[Team Name: Data Wizards. 3](#_Toc59339955)

[Project Summary: 3](#_Toc869488264)

[Project description: 3](#_Toc1209010156)

[Team: 3](#_Toc1518856808)

[Objective: 3](#_Toc602032555)

[Usefulness: 3](#_Toc1533422665)

[Dataset: 4](#_Toc1496153812)

[Data Preprocessing: 4](#_Toc668916937)

[Description of the functionalities: 5](#_Toc807852012)

[Communication and Sharing: 6](#_Toc1520374830)

[Milestones: 6](#_Toc242325948)

[Contributions: 6](#_Toc2069545337)

[Appendix A: 7](#_Toc1680958203)

[Additional data observations 7](#_Toc1236011336)

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

# Team Name: Data Wizards.

# Project Summary:

As per the statistics from [review42](https://review42.com/resources/real-estate-statistics/), 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.

# Project description:

Team: Amit Banerjee, Pravallika Pentapati, Sahil Dhingra.

## Objective:

The objective of our project is to create a web app for US Real Estate listings. The users would be welcomed by a page listing the top/latest listing and would have the capability to-

* Search listing by filters like zip code, price range, etc.
* Add a new listing.
* Edit/Remove an existing Listing.
* Research relevant statistics about the real estate market.

Real Estate websites have a lot of information and it’s often difficult to browse through all of that easily. We are trying to build a simple website with essential information in the listings so it could be an easy starting point for house hunters.

## Usefulness:

The real estate market in the US is at an all-time high and people are searching for houses every minute. There are many real estate listing and search services available today, some are listed below-

* Zillow
* Redfin
* Open-door
* Realtor

All the above-listed websites have comprehensive real estate listings with a lot of details including pictures, thus generally requiring more time to browse through many listings. We aim to keep listings limited to essential information so multiple listings could be easily scrolled and browsed. We are also going to add, modify and remove listings.

The users that would benefit from this web app are the ones looking to do initial research of listings within an area, within a price range, or by some other criteria. This site would also be helpful for people adding, modifying, or deleting their personal listings.

## Dataset:

This dataset contains a list of house sale prices in the USA region. Data was scraped from a real estate listings website operated by the News Corp subsidiary Move, Inc. based in Santa Clara, California. It is the second most visited real estate listings website in the United States as of 2021, with over 100 million monthly active users, then hosted data on Kaggle in CSV format for performing the statistical analysis and predicting the house prices. We used this dataset for building the web app in the final project. This dataset in Kaggle will be updated frequently. There are 203,216 entries and 12 columns. The dataset in Kaggle could be found at <https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset>.

## Data Preprocessing:

Below are the Data cleaning and enhancing pointers that we are planning to undertake as part of data preprocessing:

#### Data Cleansing

* Duplicate records: Based on our keys, we will delete the duplicate records. Only 1 listing should be there for 1 property at a given point, although historical data for a house sold can exist.
* Status correction – If the status is for sale, and we have sold date as NOT NULL, the status will be updated to SOLD.
* ZIP Codes - There are 195 records with Null zip codes. We will update those records based on city and state data.
* Cities - 1 record for the city for the state of New York is NULL. As per address, the same will be updated with correct city data - “ELIZAVILLE”
* Street: 25 records have the value of 9:00 AM in street. Looking at the address field, it looks like 9 A from the address is incorrectly set to 9:00 AM or a listing agent might have put that incorrectly when it might be the time of the tour. We will update this to 9 to match the address field.

#### Data Enhancement

* Adding Attributes: We would add a Listing Date, and a Property Type (House, land, etc.) to enrich the data.
* Adding Master Data Table: We would gather the Master Data for City, State, and Zip codes.

Please find below detailed information for column attributes:

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Comments |
| **status** | Varchar | Housing Status (on sale or other option) |
| **price** | Number | Price in USD. |
| **bed** | varchar | Bedroom count |
| **bath** | Number | Bathroom count |
| acre\_lot | Number | Acre lot |
| full\_address | Varchar | Full address |
| street | Varchar | Street name |
| city | Varchar | City name |
| state | Varchar | State name |
| zip code | number | Zip Code |
| house\_size | Varchar | House size in sqft (square feet) |
| Sold\_date | Varchar | The date when the house is sold |

## Description of the functionalities:

Basic functionalities:

We will have basic CRUD functionalities which are limited to:

* Read listings: Users can view and search all available listings by using any filter criteria like address, zip code, city, etc.
* Create listings: The user can create a new property listing.
* Update listings: The user can search and edit any listing(s).
* Delete listings: The user can search and delete any listing(s).

Advanced functionalities:

We will create compound queries and do statistical analysis of our data to provide data insights. In addition, we will display visual charts based on a statistical analysis of the application.

Here is the list of advanced functionalities which we are planning but are not limited to:

* Visualize no. of listings with appropriate filter conditions like State, Zip, status, time on the market, etc.
* Off Market in Two Weeks as a Percentage of Pending Sales.
* Average sale price per Sqft.
* Time-series visuals for: Median sale price, new listings, etc.

## Communication and Sharing:

Team communication related to project work and discussions would be facilitated via MS Teams.

Documents and code would be collaboratively worked on and shared with OneDrive and GitHub.

## Milestones:

* Week 6 - project description
* Week 8 - database
* Week 10 - web app mock
* Week 11 - short video presentation
* Week 12 - full demo

## Contributions:

* Amit: Project Planning, Database design, web app development, Participating discussions, report preparation based on contributions.
* Pravallika: Project Planning, Database design, web app development, participating in discussions, report preparation based on contributions.
* Sahil: Project Planning, Database design, web app development, participating in discussions, report preparation based on contributions.

GitHub link: <https://github.iu.edu/prpent/D-532-Applied-Database-Final-Project>

# Appendix A:

## Additional data observations

Cities:

51 records for cities of Virgin Island are null. Being an Island, there might not be a requirement for the city on Virgin Island city. But, to maintain the data sanity, we will update these records manually.

House Size:

There are around 30K records that have null house size. Out of these only 421 records have null in acre\_lot as well. House size and acre\_lot can be used as per a particular city or state. We will not remove these listings as other appropriate columns have enough relevant information for a user to have a good idea.

Acre Lot:

320 records have null values for acre\_lot as well as house\_size for mainland US. Looking at a few records online looks like they are land listings and many have NULL values on listing sites. Therefore, we will not update these records and let them be.

107 records for Puerto Rico and Virgin Island have no data for acre lot and house size. These are islands and listings differ when one sells in the mainland vs island. Also, since both the fields are nullable, we will let them be. User has access to update these fields if they wish to.

Bed/Bath:

Bed and bath have null values as well. This is understandable as it depends on the listing type. For example, if the listing is for land, these values will be null.