House Price Predictor

1. Introduction:

➤ The real estate market is complex and dynamic, making it challenging for buyers and sellers to determine fair and accurate house prices. In response to this challenge, we embarked on a data science project aimed at predicting house prices using a design thinking approach.

➤ The significance of this project lies in its potential to provide stakeholders, such as real estate agents, buyers, and sellers, with a valuable tool for making informed decisions in the housing market.

➤ This document outlines our journey through the design thinking process, from empathizing with the needs of our users to defining clear objectives and ultimately creating a predictive model for house prices.

2. Empathize:

House Buyer

Hear

- ❖ Advice from real estate agents on pricing and staging.
- ❖ Stories of bidding wars and rising property prices, which add to the pressure of making the right decision.
- ❖ Suggestions from friends and Neighborhood .

See

- ❖ Online listings, open houses, and for-sale signs everywhere.
- ❖ Sees various online listings on real estate websites, showcasing a wide range of properties in different neighborhoods and price ranges.
- ❖ Notices "For Sale" signs in their desired neighborhoods, which raises their hopes of finding the perfect home.
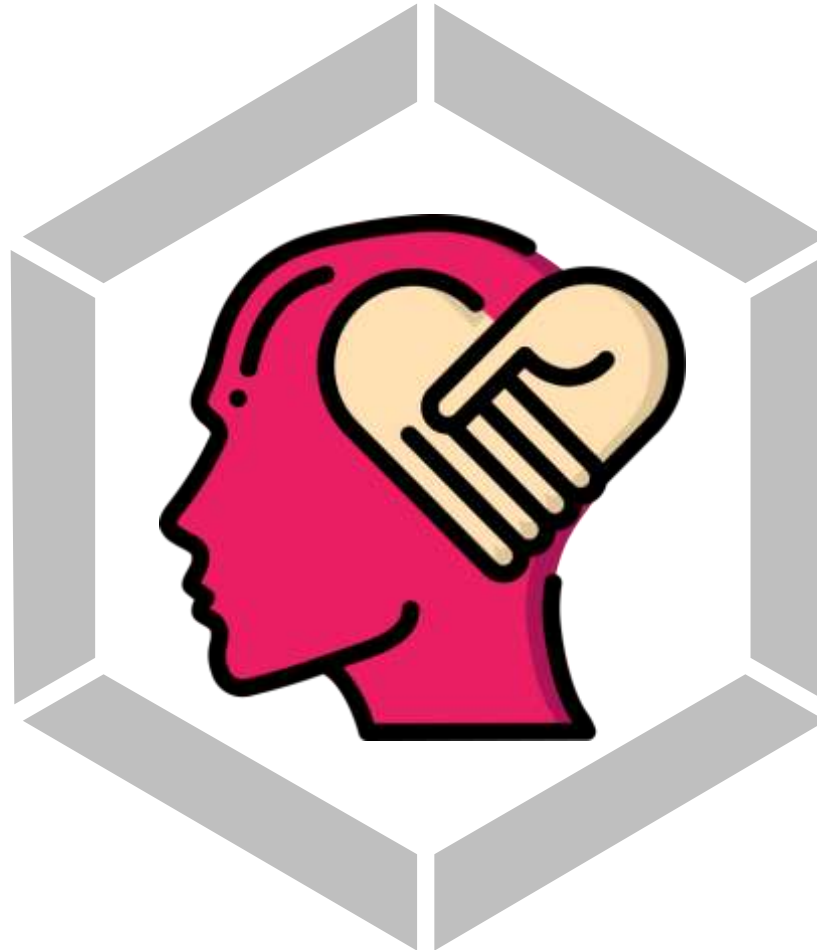
Say and do

- ❖ Expresses their excitement and concerns to friends and family, seeking advice and reassurance during the homebuying process.
- ❖ Asks questions to real estate agents about property histories, potential for appreciation, and negotiation strategies.
- ❖ I need a house that fits my family's needs and lifestyle.

Think and feel

- ❖ I want to find my dream home within my budget.
- ❖ I want to find my dream home within my budget, considering factors like location, size, and amenities.

Real Estate Agent

Think & Feel

- I must provide accurate pricing guidance to clients.

- Feels the pressure of client expectations, knowing that clients rely on them for expert guidance in a complex and competitive market.

- Strategizes on pricing properties competitively to attract buyers while ensuring clients receive fair market value for their homes.

Hear

- Client concerns about market trends and investment potential.

See

- A competitive market with fluctuating property values

- Sees both motivated and hesitant buyers and sellers navigating the market..

Say & Do

- Communicates market updates and property listings to clients, providing valuable insights into pricing, location, and investment potential.

- I can help you find the right property at the right price.

Pain

- Irregular income and high competition.
- Market volatility and client expectations.

Gain

- Flexibility and the potential for a substantial income.
- Income potential and personal growth opportunities.

Hear

❑ Advice from real estate agents on pricing and staging.

Say & Do

❑ I hope to get a fair offer for my property

See

❑ Other houses in the neighborhood going on the market.

Think & Feel

❑ I want to sell my house for the best possible price.

❑ Hopeful about the sale but also uncertain about market conditions

Say & do 2

3 **See**

Hear 1

4 **Think & feel**

House Seller

Define:

Problem Statement:

The goal of this project is to use AI model to guess how much a house should cost. To do this, we'll use information like where the house is, how big it is, how many bedrooms and bathrooms it has, and other important things about it. We'll start by getting the dataset by the source provided by your team or your organization from the kaggle, then we'll pick a good way to make our guess, and finally, we'll train our AI model to predict the price for excepted features of the house given as the input.

Ideate:

- ➢ To predict the house price we use the dataset features provided by your organization for our project from Kaggle (USA_Housing).
- ➢ The price is predicted by using the selected model by us.

- ***Dataset Selection:***

- Dataset Name: USA Housing

- Source:Kaggle. (Dataset Link: https://www.kaggle.com/datasets/vedavyasv/usa-housing) Provided by your organization.

- ***Description:***

- The USA Housing dataset is a collection of data related to housing properties in the United States. It contains information about various features.The dataset contains 5000 rows and 7 columns.

- ***Data Preprocessing:***

- **Handling Missing Values**: Missing Data is a common issue in datasets. Depending on the extent of missing values,it can be following way.

- **Imputation**: Fill in missing values with a suitable measure.

- **Common method includes:** Mean,Median or Mode imputation for numerical features.Most frequent category imputation for categorical features.

Features:

➢   Avg. Area Income

➢   Avg. Area House Age

➢   Avg. Area Number of Rooms

➢   Avg. Area Number of Bedrooms

➢   Area Population

➢   Price

➢   Address

Prototype:

Feature Selection:

To predict the house price by the features provided in dataset . There are seven features are provided in the given dataset such as area income, area house age, number of rooms ,number of bed rooms ,and population at specific area and average price for each categories.

But we can't use all of it's features like address and area income because these data are not essential for predictions .we need only the following features.

To Predict House Prices:

•   Avg. Area House Age

•   Avg. Area Number of Rooms

•   Avg. Area Number of Bedrooms

•   Area Population

Model Selection:

Model selection is the phase that to select a appropriate model for project to give a better performance and to guess the price nearly approximate of real price. There are many model such as liner regression and random forest regression. For this problem multiple linear regression is a efficient and easy way to predict.

▶ Test:

Model Training:

- ▶ • Our model can be trained by splitting the preprocessed dataset into training and validation set.

- ▶ • The common split ratio is 80% for training and 20% for testing .

- ▶ • And by changing the ratio multiple times to improve accuracy of prediction.

- ▶ • And then train our selected model using training data by regression algorithm.

Evaluation:

• By using some error finding regression methods such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) to measure prediction accuracy.

• By ploting the multiple prediction with corresponding inputs in graph to find the distinct value to identify the error.

• And then change test and train ration to improve our model prediction.

• And the repeat these above steps until the model give higher accuracy with less error.

Conclusion:

Here in this document we give that what we understand in the problem statement and how we going to complete our project.

And provided the design thinking for our project.