**PREDICTIVE MODELLING**

**PROJECT**

**(Project Proposal)**

(By Pratik More)

* **What data you have chosen, including a link to where you found the data.**

We used Dataset search option provided by the Google itself. The link of the dataset is mentioned below:

Please click on blue coloured tab ‘Zenorows.com’ to navigate forward to download the dataset.

<https://datasetsearch.research.google.com/search?src=0&query=US%20real%20estate&docid=L2cvMTFueDdtaDJzXw%3D%3D>  
  
**OR,** Please click on below link to download the dataset.

<https://www.zenrows.com/datasets/us-real-estate>

**Steps to download the Dataset:**

* You will need to enter your email ID.
* You will receive a confirmation link in your Email inbox.
* Click on the confirmation link to download the dataset.
* **What business value there might be from your findings and prediction. What business questions are you hoping to answer in your modeling work?**

First we will talk about the **Business Value** that it will bring,

**Informed Decision Making:** With a reliable model, stakeholders can make decisions based on data-driven insights rather than just intuition or surface-level market trends. This can lead to better returns on investment and more successful transactions.

**Dynamic Pricing Strategy:** Real estate agents and sellers can develop a more dynamic pricing strategy based on the model's predictions. It can help them understand when to raise or lower property prices depending on the features and the current market conditions.

**Risk Reduction:** Investors and lenders can use the model to evaluate the risk associated with properties. By understanding the potential market value of a property, stakeholders can make safer investments and lending decisions.

**Tailored Marketing:** If certain features are identified as significantly impacting property prices, agents can emphasize these in marketing efforts. This helps in attracting potential buyers more efficiently.

**Improved Customer Experience:** Buyers often struggle with understanding if a property is priced fairly. A transparent model can offer buyers more confidence in their purchasing decisions, leading to improved customer satisfaction.

**Business Questions Addressed:**

1. Which property features have the most significant influence on its price?

* Understanding the hierarchy of features in terms of their influence on price can guide both sellers in emphasizing certain features and buyers in understanding what they're paying for.

1. How does geographical location (latitude and longitude or city/state) impact the price?

* This will help in understanding location premiums and identifying potential high-growth areas.

1. How accurate are tools like Zestimate in predicting actual market prices?

* Comparing our model's predictions against tools like Zestimate can identify discrepancies and offer insights into the accuracy of such tools.

1. How do additional attributes like having a 3D model, video, or open houses influence the property's perceived value?

* This can guide marketing strategies for sellers and agents.

1. What role do brokers play in property valuation? Are properties associated with certain brokers priced higher or lower on average?

* Helps in understanding the market influence of specific brokers or agencies.

1. Can we predict future real estate market trends based on the historical and current data available?

* Useful for investors looking for long-term investments and for agents to guide their clients.

1. How does the listing's duration (time since it's been listed) impact its price or the likelihood of its sale?

* Helps in understanding market liquidity and devising sales strategies.
* **What you plan to predict from the data. Be specific about your target variable.**

**Target Variable:**

Price: This is the listed price of the property. It's a continuous numerical variable, which makes this a regression problem. Our main goal will be to predict this value as accurately as possible given the attributes of a property.

**Rationale for Choosing Price as the Target Variable:**

The real estate industry is fundamentally centred around the valuation of properties. Knowing the potential price of a property allows stakeholders—be it buyers, sellers, agents, or investors—to make informed decisions. Whether it's about setting a competitive listing price, deciding on a buying price, understanding market trends, or evaluating property for investment purposes, the price of a property is a crucial piece of information. By building a predictive model that can estimate property prices based on certain features, we can offer valuable insights to these stakeholders and contribute to more transparent, data-driven decisions in the real estate market.

Furthermore, understanding how different features and attributes impact the price allows for better marketing, investment strategies, and more informed negotiations during property transactions.

In summary, by focusing on predicting the 'Price' as the target variable, we aim to provide a tool that can assist in various aspects of real estate transactions and investments, ultimately adding value to the industry and its participants.

* **How many rows and how many columns are in the dataset. Give a brief description of what each row represents and what type of information is included in the dataset.**

**Dataset Dimensions:**

* **Rows:** 10,000
* **Columns:** 46

**Row Description:**

Each row in the dataset represents a unique property listing in the US real estate market. It provides detailed attributes of that property, including its pricing, location, features, and other associated metadata.

**Brief Description of Dataset Columns:**

1. **zpid:** A unique identifier for the property.
2. **id:** Another identifier, potentially for internal tracking.
3. **providerListingId:** The ID given by the listing provider.
4. **imgSrc:** The source URL for the property's image.
5. **hasImage:** Indicates if the property listing has an image (binary: yes/no or true/false).
6. **detailUrl:** URL for detailed property information.
7. **statusType:** Type of listing status (e.g., for sale, sold, etc.).
8. **statusText:** Descriptive text of the property's status.
9. **countryCurrency:** The currency in which the price is listed, presumably USD.
10. **price:** The listed price of the property (our target variable).
11. **unformattedPrice:** Price without any formatting (e.g., without commas).
12. **address:** Complete address of the property.
13. **addressStreet:** Street address.
14. **addressCity:** City where the property is located.
15. **addressState:** State of the property location.
16. **addressZipcode:** Zip code of the property.
17. **isUndisclosedAddress:** Indicates if the address has been hidden or undisclosed.
18. **beds:** Number of bedrooms in the property.
19. **baths:** Number of bathrooms.
20. **area:** Total area of the property (likely in square feet or square meters).
21. **latitude:** Geographical latitude of the property.
22. **longitude:** Geographical longitude of the property.
23. **isZillowOwned:** Indicates if the property is owned by Zillow.
24. **variableDataType:** Type of data in the variableDataText column (could be categorical or descriptive).
25. **variableDataText:** Contains varying data, possibly special features or comments.
26. **variableDataIsFresh:** Indicates if the variable data is recent or outdated.
27. **badgeInfo:** Likely information on any badges or accolades the property has.
28. **pgapt:** Unclear without more context; potentially related to property type or features.
29. **sgapt:** Similar to pgapt, might need more context.
30. **zestimate:** Zillow's estimated market value for the property.
31. **shouldShowZestimateAsPrice:** Indicates if the zestimate should be displayed as the property price.
32. **has3DModel:** Indicates if the property listing has a 3D model.
33. **hasVideo:** Indicates if there is a video tour available for the property.
34. **isHomeRec:** Possibly indicates if the property is recommended or featured.
35. **info2String:** Unclear without more context; potentially additional information or comments.
36. **info3String:** Similar to info2String.
37. **brokerName:** Name of the broker or agency listing the property.
38. **hasAdditionalAttributions:** Indicates if there are additional attributions or credits required for the listing.
39. **isFeaturedListing:** Indicates if the listing is prominently featured.
40. **list:** Unclear without context; could be related to a type of listing or categorization.
41. **relaxed:** Unclear without context; might need more details.
42. **hasOpenHouse:** Indicates if there's an open house scheduled for the property.
43. **openHouseStartDate:** Start date for the open house.
44. **openHouseEndDate:** End date for the open house.
45. **openHouseDescription:** Description or details about the open house.
46. **info6String:** Similar to info2String and info3String, potentially more additional information.

In summary, this dataset provides a comprehensive overview of properties listed in the US real estate market, covering a wide range of attributes from basic features like bed and bath counts to more detailed information such as open house schedules and multimedia availability.