**DATA SOURCE:**

American House Prices (kaggle.com)

The dataset comprises of various variables around housing and demographics for the top 50 American cities by population.

**Variables Overview:**

| | |
|---|---|
| Zip Code | Zip code within which the listing is present. |
| Price | Listed price for the property. |
| Beds | Number of beds mentioned in the listing. |
| Baths | Number of baths mentioned in the listing. |
| Living Space | The total size of the living space, in square feet, mentioned in the listing. |
| Address | Street address of the listing. |
| City | City name where the listing is located. |
| State | State name where the listing is located. |
| Zip Code Population | The estimated number of individuals within the zip code. Data from Simplemaps.com. |
| Zip Code Density | The estimated number of individuals per square mile within the zip code. Data from Simplemaps.com. |
| County | County where the listing is located. |
| Median Household income | Estimated median household income. Data from the U.S. Census Bureau. |
| Latitude | Latitude of the zip code. Data from Simplemaps.com. |
| Longitude | Longitude of the zip code. Data from Simplemaps.com. |

The data is obtained from kaggle.com uploaded by an individual. The sourcee mentioned are:

https://simplemaps.com/data/us-zips
https://data.census.gov/profile/United_States?g=010XX00US

which includes a government undertaking. So I believe the data is trustable, and of course no data is 100% trustworthy as there are still chances of bias and manual errors.  The dataset satisfied all the requirements as mentioned in the project brief.

## DATA PROFILE:

| | |
|---|---|
| Zip Code | Discrete, ordinal |
| Price | Continuous, ordinal |
| Beds | Discrete, ordinal |
| Baths | Discrete, ordinal |
| Living Space | Continuous, ordinal |
| Address | Discrete, nominal |
| City | Discrete, nominal |
| State | Discrete, nominal |
| Zip Code Population | Continuous, ordinal |
| Zip Code Density | Continuous, ordinal |
| County | Discrete, nominal |
| Median Household income | Continuous, ordinal |
| Latitude | Continuous, ordinal |
| Longitude | Continuous, ordinal |

| | Zip Code | Price | Beds | Baths | Living Space | Zip Code Population | Zip Code Density | Median Household Income |
|---|---|---|---|---|---|---|---|---|
| **count** | 39981.000000 | 3.998100e+04 | 39981.000000 | 39981.000000 | 39981.000000 | 39981.000000 | 39981.000000 | 39979.000000 |
| **mean** | 64833.391336 | 6.227771e+05 | 3.171682 | 2.466572 | 1901.522723 | 37726.201996 | 2379.412483 | 110837.259861 |
| **std** | 25614.601116 | 9.469793e+05 | 1.308796 | 1.323042 | 1211.307257 | 18672.647445 | 2946.574792 | 47309.055715 |
| **min** | 10013.000000 | 1.800000e+03 | 1.000000 | 1.000000 | 2.000000 | 0.000000 | 0.000000 | 27475.000000 |
| **max** | 98199.000000 | 3.800000e+07 | 54.000000 | 66.000000 | 74340.000000 | 116469.000000 | 58289.600000 | 900203.00000 |

Missing Values**:** Median Household Income 2 values updated to NA.

Duplicates: There are 962 rows duplicated. Removed duplicates.

Shape : (39019, 14)
Mixed data type : No

**QUESTIONS TO EXPLORE:**

1. What are the main factors affecting House pricing?
2. Is the zipcode density and zipcode population play a vital role in determining house price?
3. How does the median household income affects house pricing in a region?
4. Does the price increases with number of beds and bath or with the living space or both?