# Real Estates in Connecticut

Exploratory Data Analysis

Name: Example 1

## **Problem Statement:**

The Real Estate market has been on a constant rise worldwide. You are tasked by your project manager to investigate residential properties in Connecticut and see how the property value has changed over the last 2 decades. Using this dataset, find out property assessment value vs what the property has sold for. The objective is to find out how the Real Estate market changes overtime in Connecticut. With this dataset, you can use the data on the characteristics of each property, to estimate what a hypothetical property might be assessed at, and what a property might sell for.

## Business Impact:

Exploring this data could allow the company to make better decisions in regards to investment properties. It could also allow for more accurate valuations through predictive analytics. This can help sellers know how to accurately price their properties, and it could help buyers know whether the home will meet the appraisal. Analysis could also view trends and assess market forces leading to a rise or dip in property prices.

In support of this analysis, the following dataset is used.

## **General Dataset Information:**

File Name: Real\_Estate\_Sales\_2001-2020\_GL.csv

Description: Property Sales, Assessments, and Trends in Connecticut 2001-2020

**Dataset Details**: 997213 Rows & 14 Columns

**Size**: 107,807KB ( 105MB )

**Source**: Kaggle - [Dataset Link](https://www.kaggle.com/datasets/utkarshx27/real-estate-sales-2001-2020-gl)

## **Target Features:**

The problem statement outlines the essential metrics of finding Property Value changes over the last 2 decades. The key features in question that can be found within the dataset:.

1. **Property Type** - Use to separate Residential from Commercial and Industrial Properties.
2. **Residential Type** - Use to identify the intended purpose of the property.
3. **Date Recorded** - Use to identify when the sales are recorded.
4. **Assessed Value** - Possible use in comparison between property valuation vs property sales.
5. **Sales Amount** - The actual amount was paid for the transaction of the property.
6. **Year Recorded** - Feature Engineered column to isolate the sales recorded year.
7. **Sales Bucket** - Feature Engineered column to group property values.

To answer the problem statement. These seven metrics in combination will be used to conduct the analysis.

# List of Analysis

## **Analysis #1 - Total Records**

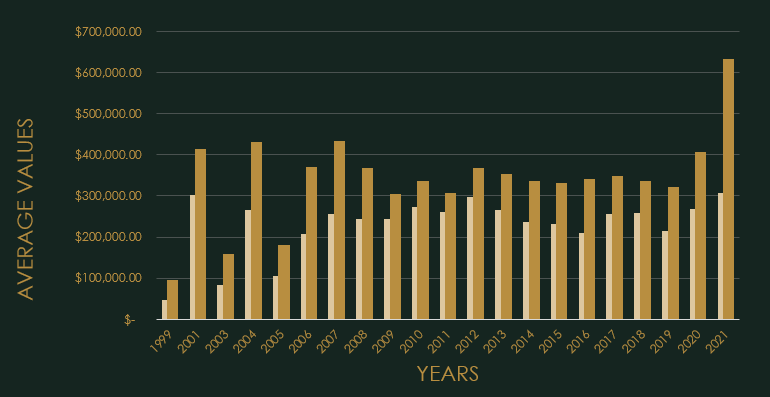
To bring into perspective, the number of records was used to demonstrate the research effort in deriving the conclusion of Real Estate Market Trends. After isolating all Residential Properties, there are 609,328 records that span across 20 years.

## **Analysis #2 - Records by Year**

Using the record, a new table can be tabulated by grouping “Year Recorded” to understand the record distribution by year. The result demonstrates the linear progression of property value that had been sold in each year. In addition, “**Average Assessment**” and “**Average Sales**” can be aggregated to show the overall changes in the market condition. The result of the aggregation can then be visualized to outline trends.

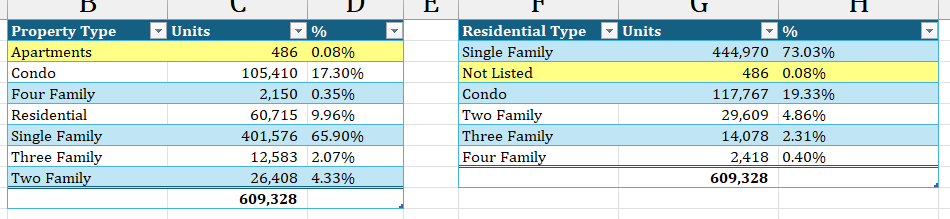
It also reveals irregularities between 1999 - 2005 that appear to have fewer records which disrupts the overall trend.

The chart below shows a clear stable trend over the last 2 decades, which conflicts with the original assumption that property value is rising globally. During the market crash of 2008, it had minimal effect on the Real Estate Market in Connecticut. It is only until the rise of COVID pandemic where property value begins to escalate in 2020 - 2021. Throughout each, there is a consistent separation between “Assessment Value” and “Property Value”. Homes are always sold for more than the assessments.



## **Analysis #3 - Property Type Distribution**

The trendline can be further analyzed by understanding the property distributions. There are two columns with similar values “Property Type” and “Residential Type” that need reconciliation to ensure data congruences.



By tabulating two tables using their respective column, there appears to be disagreement at first glance. “Residential” as property type does not appear in the “Residential Type” column. It is distributed among their “Residential Type”. This discrepancy is caused by the usage of industry terminology. Property Type defines the structure on the property, while Residential Type defines the habitat intended use. Apartments are considered “Not Listed” because apartments can be used as Residential as well as Commercial, which in this case, it is undecided.

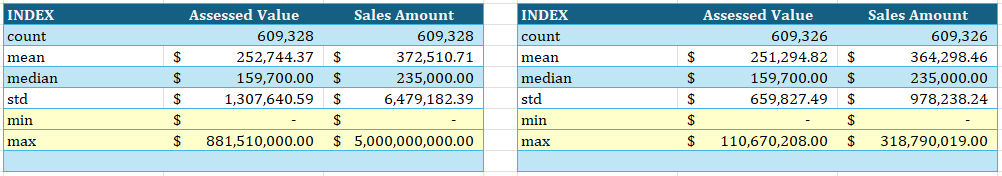
Two charts can be created to demonstrate the distribution of Property Types and the conclusion that the majority of sales are derived from Single Family homes is evident.

|  |  |
| --- | --- |

## **Analysis #4 - Basic Statistics**

To further expand on the analysis. Distribution of Property Value can potentially highlight the overall state of the Real Estate Market in Connecticut. Using basic statistics, it reveals the dispersion of property values are leaning towards high end properties with standard deviation of $6,479,182.39, meanwhile the average is $372,510.71. A plausible cause is due to an overly exaggerated maximum value of $5 Billion property that may appear to have been an outlier. However, in the Real Estate market, it is plausible to have a $5 Billion property.

After careful consideration, it would be wise to test the effects of those records. After eliminating the highest property records, the standard deviation significantly dropped, however the average was not significantly impacted.

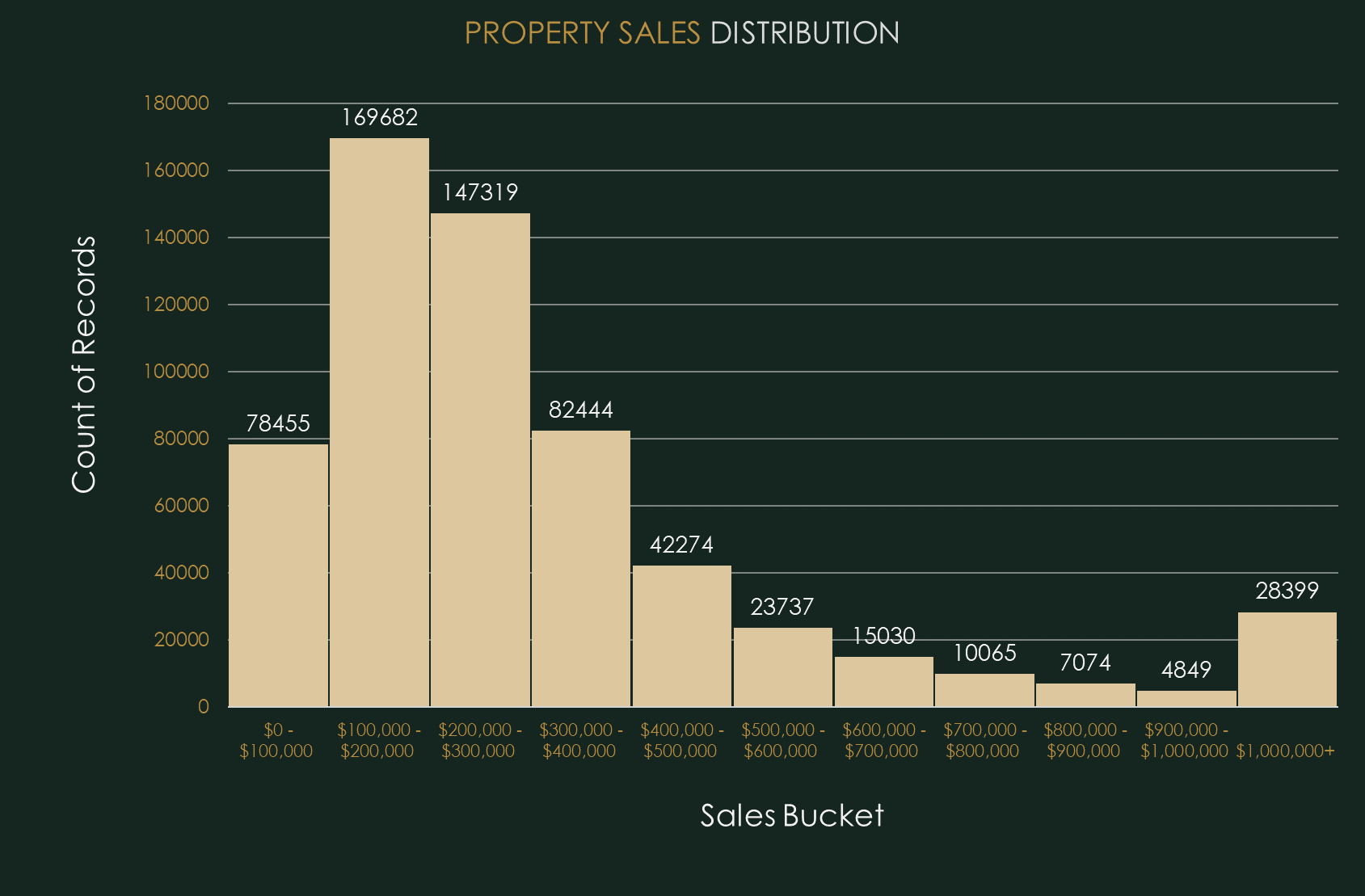


In conclusion, the records were kept due to the nature of the Real Estate business and that its existence has minimal influence over the trends of the market.

## **Analysis #5 - Property Value Distribution**

“Property Values” are grouped together using predefined customized Bins to illustrate the distribution. The bins are set to $100,000 increment to a maximum of $1,000,000. Anything beyond $1,000,000 is grouped together as one.

73% of Residential Property consist of Single Family homes. The distribution chart illustrates the plausible price range of Single Family Homes anywhere between $0 - $100,000 to $300,000 - $400,000, which means that the majority of the homes are sold for below average price.



Visually, the distribution is right skewed in favor of low property value homes.

The distribution of property is an important indicator for investors who wish to understand more deeply about the Real Estate market in Connecticut.

In the case of the Problem Statement. Showing the Property distribution might not have much influence over the trends of the Real Estate Market.

# Conclusion

Out of the 609,328 residential records. The dominant Residential Type are Single Family Homes that consist of 73% of the records, followed by 19% Condominium. The trend over the past 20 years indicates a steady trend of property value despite global concern for the rise of house prices, which the dataset does not conform to the original hypothesis. Even during the 2008 Real Estate Market crash, the effects had bare minimum influence on the market in Connecticut. It was only until COVID pandemic in 2020 that began the escalation of property sales value. A plausible explanation is due to lockdowns with citizens unable to work that leads to an influx of foreclosed homes with many investors capitalizing on the opportunity that drives property values up. In addition, the US Federal Reserve had also increased printing of fiat currency during this period causing massive inflation. If the cause of pandemic and excessive fiat currency in circulation. The overall Real Estate Market in Connecticut would have maintained its stability.