# Real Estates in Connecticut Final Report Introduction

The project was tasked to find what a hypothetical property might be assessed at, and what a property might sell for in the Real Estate market in Connecticut. Using the dataset found on Kaggle ([link](https://www.kaggle.com/datasets/utkarshx27/real-estate-sales-2001-2020-gl)), the analysis carried forward is limited insight pertaining to assessment value vs property value to verify the popular claim that property value is increasing.

The method used in this report is simple, aimed for Investors, Property Managers, and Home Buyers to understand without the usage of complex analytical models, terminologies, or exaggerated long reports.

Its purpose is to provide a snapshot of the overall Real Estate market condition for the past 20 years to give a realistic picture without the marketing hype.

# 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.

# Data

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)

# Data Analysis & Computation

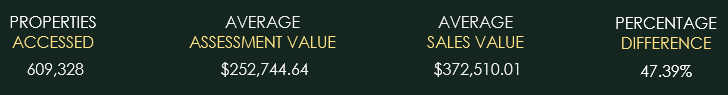
## Data Cleaning

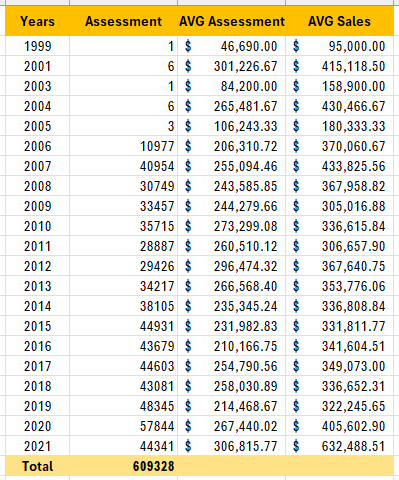
* Filter “Address” by Smallest to Largest to isolate irregular fields.
* Remove all irregular “Address” entries to a new Worksheet - Blank Address, Numerical Address, Date Address
* Filter Property Type by - Residential, Condo, Single Family, Two Family, Three Family, Four Family, Apartment
* Copy all filtered record into a new Cleaned Data Worksheet
* Saved New copy of excel with only the cleaned data

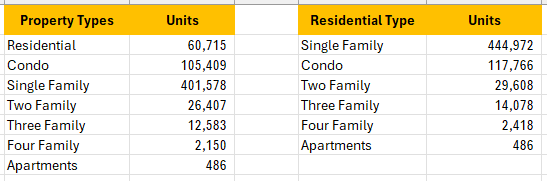
## Analysis Methods

Analysis conducted and used for the datafolio.

* Analysis #1 - Total Records  
  To provide a general idea of the dataset as a whole.

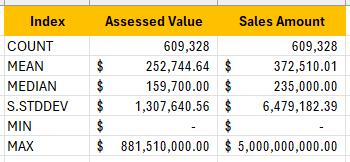


* Analysis #2 - Records by Year  
  Define trends among the cleaned data.  
  
* Analysis #3 - Property Type Distribution

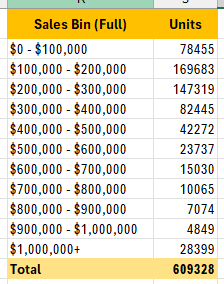
For further breakdown of the analysis by the types of property.  


* Analysis #4 - Basic Statistics

Using statistics to verify numerical accuracy of previous analysis.



* Analysis #5 - Property Value Distributions

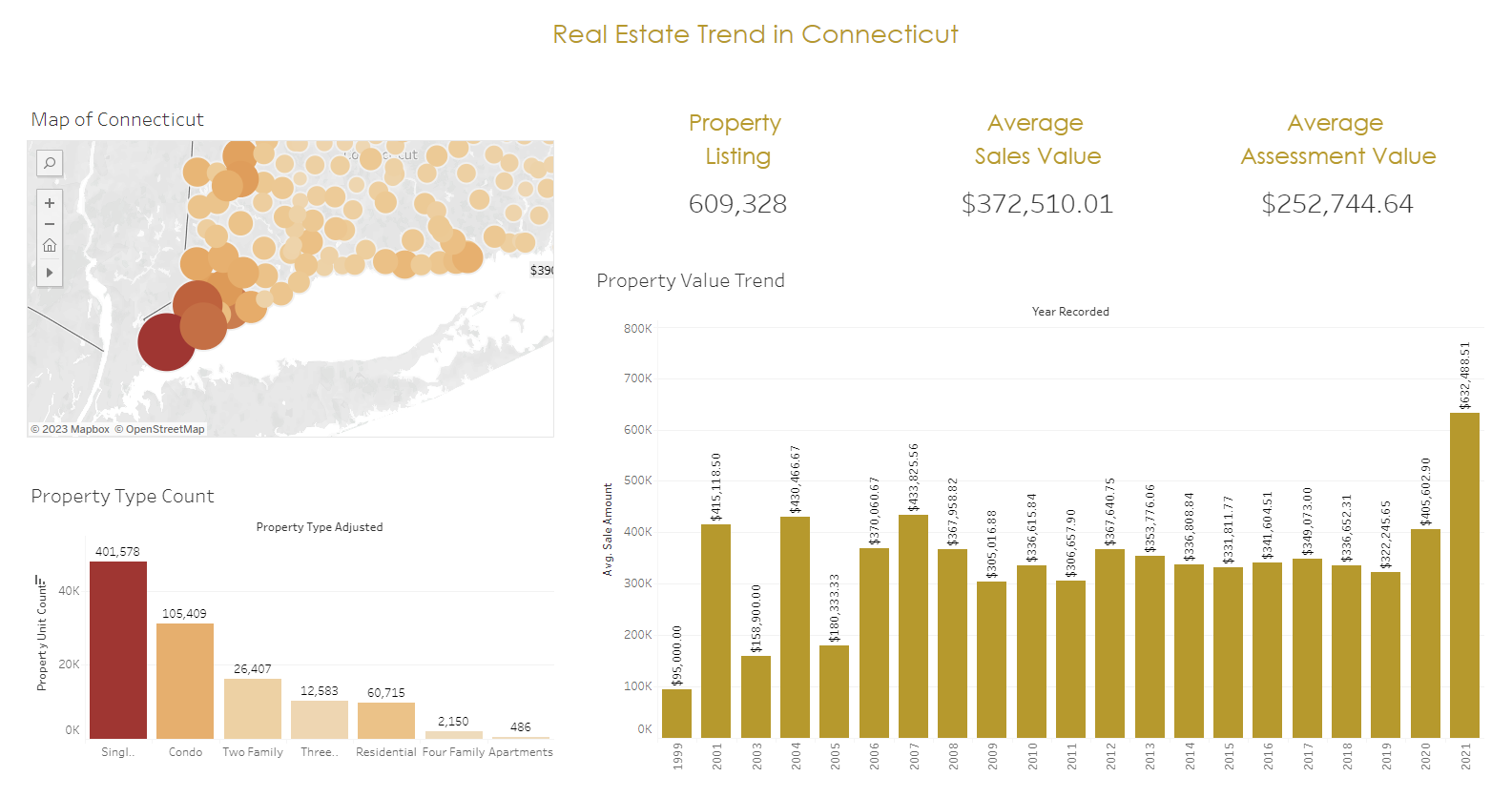


Analysis that was discarded due to skew data or introduce unnecessary complexity to the final report.

* Reduced Sample Size
* Assuming high value property as outlier
* Isolate trend by property type
* Histogram of Property Sales Value distribution

# Dashboard

The Dashboard’s ( [Link](https://public.tableau.com/app/profile/frankie.law3674/viz/Real_Estate_In_Connecticut_2000-2021/Main) ) intended purpose is to provide an interactive filtering system to allow readers to further explore Real Estate market trends by local region. The dashboard breakdown will inform the reader of the evolution of property prices. It also shows the breakdown of the property by their residential type.



# Challenges

The dataset used can only determine the numerical trend of the Real Estate market for the past 20 years. It does not hold information for the causation of those trends.

Filtering the dataset can be an issue. There are records of assessment for every level of Real Estate. For example, there are single family homes that range from $50,000 to $15 Million that can cause skewness of an analysis. The records can be removed although it is not necessary.

# 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. It is only until 2020 during the COVID pandemic where we begin to see the raise of property sales value.

Over the past 20 years, property sales have consistently sold for more on an average of 47.39% than the assessed value, which can perhaps feel that the Real Estate Market has rapidly increased.

Study of Property Sales Value distribution suggests that the majority of the transaction occurred in the bracket of $100,000 to $200,000 range and as property value increases, home sales decrease. This can be caused by the local standard of living.

# Future Works

The dataset provides a rich dataset for the past 20 years. In combination with datasets such as demographic, crime rate, or civil planning data, future analysis can be conducted to assess the overall quality of Connecticut.

This dataset can be used alongside with global events to find its effect against the Real Estate market.

In this study, it only included Residential property. This study can further expand into Commercial and Industrial properties.