**< INTRODUCTION TO DATA MANAGEMENT>**

**(Project Semester January-April 2025)**

**(TITLE OF THE PROJECT)**

# Real-estate Data Analysis


**Submitted by**

**(Y.V. Sai Yaswanth Reddy)**

**Registration No: 12307855**

**Programme and Section: Computer Science and Engineering(K23GW)**

**Course Code: INT217**

**Under the Guidance of**

**(Dr. Baljinder Kaur)**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

# DECLARATION

I hereby declare that this project titled *"Real-Estate Data Analysis"* submitted for the subject of Microsoft Excel is my original work and has not been submitted earlier for the award of any degree or diploma. All the data used in this project has been taken from reliable sources and has been used only for academic purposes.

**Signature:**
**Date:12-04-2025**
**Name: Y.V. Sai Yaswanth Reddy**

# CERTIFICATE

This is to certify that the project titled *"Real-Estate Data Analysis"* is a bonafide work completed by **Y.V. Sai Yaswanth Reddy** under my guidance for the course requirements of Introduction to Data Management. The work is found to be satisfactory.

**Signature of the Guide**
Name: Dr. Baljinder Kaur
Department of CSE
Institution Name: Lovely Professional University

# ACKNOWLEDGEMENT

I extend my heartfelt thanks to the BALJEENDAR KAUR MAM. With data on sale amounts, assessed values, property types, residential classifications, and sales ratios across various towns, this dataset has been an essential resource for my project. I deeply appreciate the hard work and dedication that went into compiling and making this data accessible, as it has allowed me to explore and analyze real estate market trends effectively. My gratitude goes out to everyone involved in this effort, which has significantly contributed to the success of my work.

Y.V. Sai Yaswanth Reddy
12307855
Computer Science and Engineering

# Introduction

## Introduction

The "Real-Estate Data Analysis" dataset is a comprehensive collection of real estate transaction records, designed to facilitate in-depth analysis of property market trends and dynamics. It includes detailed information on properties sold across various towns, capturing key metrics such as sale amounts, assessed values, sales ratios, property types, and residential classifications. The dataset also provides specifics like serial numbers, list years, dates recorded, addresses, and towns, offering a granular view of real estate activities. With data points covering residential properties, vacant land, commercial properties, and more, this dataset serves as a valuable resource for studying market behaviours, property valuation patterns, and regional differences in real estate transactions. It is particularly useful for researchers, analysts, and enthusiasts seeking to derive insights into the factors influencing property sales and market performance.

## Objectives

The key objectives of this project are as follows:

- Analyse Average Sale Amounts by Town
- Assess Property Type Distribution
- Compare Assessed Values and Sale Amounts
- Evaluate Sales Ratios by Property Type
- Explore the count and distribution of property

## Scope

The scope of the Real-Estate Data Analysis project is defined by the geographical coverage of various towns, as highlighted in the "Town vs sale amount" bar chart, which includes locations such as Bridgewater, Darien, Greenwich, Kent, Lyme, Middlebury, New Canaan, Old Saybrook, Washington, and Wilton. This analysis will be limited to these specific towns unless additional data is incorporated, ensuring a focused regional perspective on real estate trends.

The project will also encompass an evaluation of different property types, as depicted in the "Property Type" pie chart, which shows a dominant 91.8% Residential category alongside 18.2% Vacant Land, with potential inclusion of other types like Commercial or Apartments. This includes a detailed breakdown of residential subtypes as illustrated in the "Count of Property Type" chart, allowing for a comprehensive understanding of market composition.

Financial metrics form a critical part of the scope, involving the analysis of average sale amounts the comparison of average assessed values versus sale amounts and the assessment of sales ratios .These metrics will be compared across towns and property types to evaluate market valuation accuracy and identify pricing trends.

Data segmentation and visualization are integral to the scope, leveraging interactive filters to segment the data effectively. The project will utilize bar charts, pie charts, and plot areas to represent counts, averages, and ratios, providing a clear and interactive visual representation of the dataset's insights.

**Methodology**

The project methodology is structured into the following phases:

1. **Data Collection:** The relevant dataset was sourced, containing information on country-wise manufacturing figures. The data was assessed for accuracy, relevance, and completeness.
2. **Data Cleaning and Preparation:** Raw data was cleaned by handling missing values, standardizing country names, and ensuring consistency for seamless integration with Excel tools.

**Source of Dataset**

This data is from the Data.gov

And here is the link of the dataset: https://catalog.data.gov/dataset/real-estate-sales-2001-2018

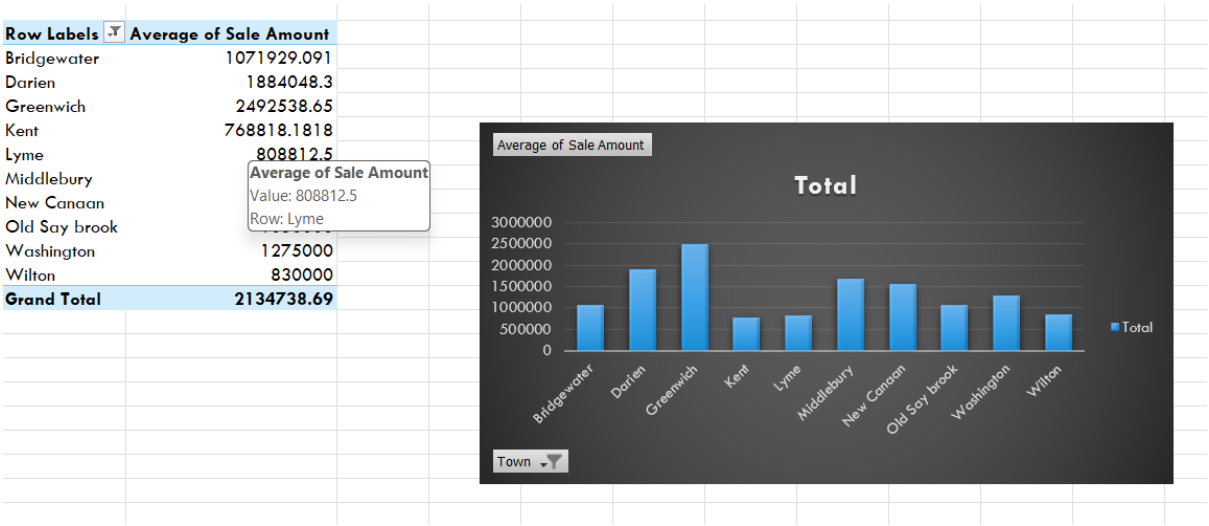
**Key Columns:**

- Town
- Address
- Assessed Value
- Sale Amount
- Sales Ratio
- Property Type
- Residential Type

## 3. Analysis on Dataset

### Objective 1: Analyse Average Sale Amounts by Town

This objective focuses on evaluating the average sale amounts across different towns as derived from the dataset's "Sale Amount" and "Town" columns. The "Town vs sale amount" bar chart in the dashboard visually represents these averages, allowing for a clear comparison of high-value markets versus lower-value ones The goal is to identify regional disparities, highlight lucrative markets, and understand economic or demand-driven factors influencing sale prices across towns.



| Row Labels | Average of Sale Amount |
|---|---|
| Bridgewater | 1071929.091 |
| Darien | 1884048.3 |
| Greenwich | 2492538.65 |
| Kent | 768818.1818 |
| Lyme | 808812.5 |
| Middlebury | |
| New Canaan | |
| Old Say brook | |
| Washington | 1275000 |
| Wilton | 830000 |
| **Grand Total** | **2134738.69** |

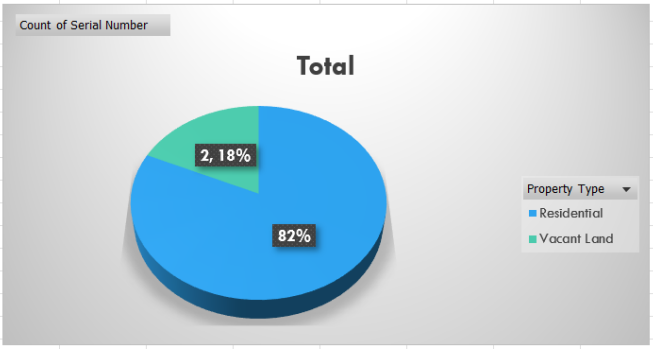


### Objective 2: Assess Property Type Distribution

This objective involves analyzing the distribution of property types within the dataset, categorized under the "Property Type" column. The dashboard's "Property Type" pie chart provides a visual breakdown, emphasizing the dominance of residential properties. By exploring the count and proportion of types like Commercial or Apartments ,the aim is

to understand market composition, assess the prevalence of different property categories, and identify potential investment opportunities or market saturation in specific types.

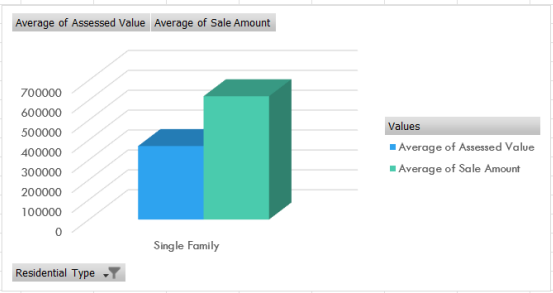
| Row Labels | Count of Serial Number |
|---|---|
| Residential | 9 |
| Vacant Land | 2 |
| **Grand Total** | **11** |




**Objective 3: Compare Assessed Values and Sale Amounts**

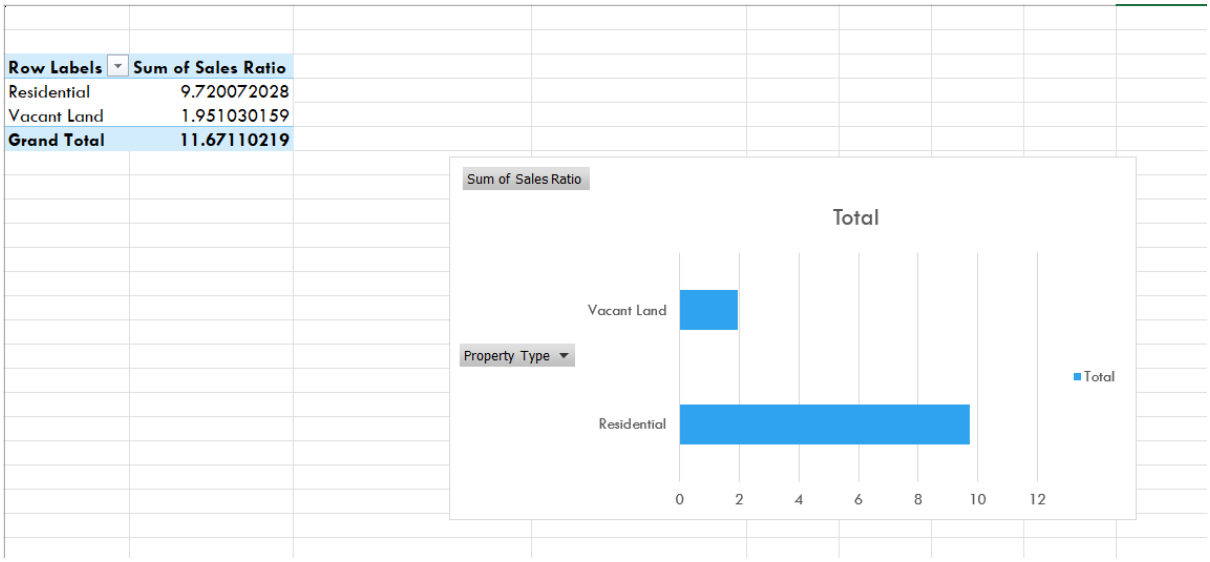
This objective entails comparing the "Assessed Value" and "Sale Amount" columns to evaluate how closely property valuations align with actual sales. The dashboard's "Assessed value vs sale amount" plot, focusing on Single Family properties, shows a significant gap indicating potential under- or over-valuation. The analysis will explore these differences across property types and towns to assess valuation accuracy, market trends, and the reliability of assessed values as predictors of sale prices.

| Row Labels | Average of Assessed Value | Average of Sale Amount |
|---|---|---|
| Single Family | 369037.7317 | 619296.6452 |
| **Grand Total** | **369037.7317** | **619296.6452** |



**Objective 4: Evaluate Sales Ratios by Property Type**

This objective centres on analysing the "Sales Ratio" column, which measures the ratio of assessed value to sale amount. The dashboard's "Sales Ratio" plot highlights these variations, with Residential showing a higher aggregated ratio. The goal is to evaluate how well assessed values reflect market prices across property types, identify anomalies and provide insights into assessment fairness or market conditions affecting different property categories.



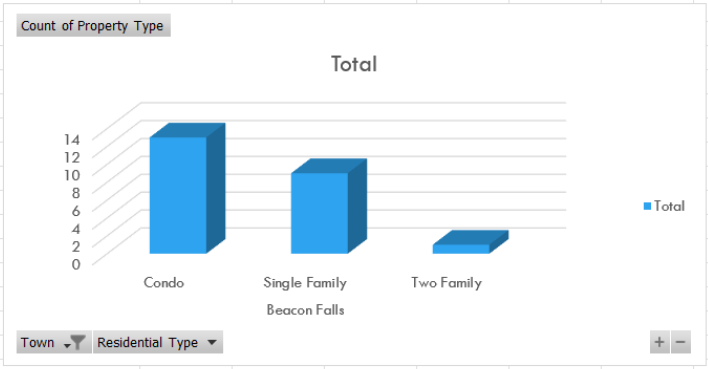
| Row Labels | Sum of Sales Ratio |
|---|---|
| Residential | 9.720072028 |
| Vacant Land | 1.951030159 |
| **Grand Total** | **11.67110219** |



**Objective 5: Explore the Count and Distribution of Property**

This objective involves examining the "Count of Serial Number" and "Count of Property Type" from the dataset, visualized in the dashboard's "Count of Property Type" bar chart. The analysis will assess the distribution of properties across towns and residential types using the interactive filters to segment data. The aim is to understand market activity levels, identify regions with high transaction volumes, and explore preferences for specific property types or residential classifications within the dataset.

| Row Labels | Count of Property Type |
|---|---|
| Beacon Falls | 23 |
| Condo | 13 |
| Single Family | 9 |
| Two Family | 1 |
| Grand Total | 23 |




## Conclusion

The analysis of the "Real-Estate Data Analysis" dataset offers valuable insights into the real estate market across various towns, highlighting significant regional variations in property values. The dominance of residential properties over other types underscores a market heavily focused on housing, with a noticeable gap between assessed values and sale amounts suggesting potential inconsistencies in valuations. The distribution of property types and transaction counts reveals active markets and clear preferences in residential subtypes. Overall, this analysis provides stakeholders with a strong foundation to make informed decisions, identify investment opportunities, and address valuation challenges, while also indicating the need for further exploration into regional and temporal trends to enhance future real estate strategies.

## Future Scope

The "Real-Estate Data Analysis" dataset holds significant potential for future exploration and expansion, offering opportunities to deepen insights and broaden its applicability. Future analyses could incorporate additional years of data to identify long-term market trends and seasonal patterns, enhancing the understanding of economic cycles. Integrating external factors such as interest rates, demographic shifts, or economic indicators could provide a more holistic view of market influences. Predictive modelling, using machine learning techniques, could be developed to forecast property values and sales ratios, aiding proactive decision-making. Expanding the geographical scope to include

more towns or regions could reveal broader market dynamics, while adding data on property could refine valuation analyses. Furthermore, developing interactive tools or dashboards with real-time updates could make the dataset more accessible to stakeholders, supporting ongoing real estate planning and investment strategies.

**References**


1. Microsoft Excel Documentation
    - Official documentation and user guides provided by Microsoft for understanding the features and functionalities of Excel, including advanced analysis tools like Power Query, Power Pivot, and data visualizations.
    - URL: https://support.microsoft.com/excel
2. TutorialsPoint / W3Schools Excel Guide
    - Extensive guides and tutorials on Excel's functionality, formulas, pivot tables, data analysis, and visualization techniques. Useful for beginners to advanced users seeking to leverage Excel for analytics.
    - URL: https://www.tutorialspoint.com/excel
    - URL: https://www.w3schools.com/excel
3. Stack Overflow Discussions on Excel Analytics
    - A collection of relevant discussions and solutions regarding Excel's capabilities in data analysis, including formula assistance, pivot tables, and dynamic data visualizations. This forum is an essential resource for resolving common issues encountered during Excel projects.
    - URL: https://stackoverflow.com/questions/tagged/excel