**EXPLORATORY DATA ANALYSIS (EDA) DOCUMENTATION FOR REAL ESTATE SALES\_2001-2020 GL**

1. ***Introduction***

This project is entails performing exploratory data analysis on Property Assessment and Sales data for Informed Decision-Making. The Exploratory Data Analysis (EDA) for real estate sales involves analyzing and visualizing key aspects of real estate data to gain insights into market trends, property values, and sales patterns. Main purpose is to analyzing the real estate sales data to extract valuable insights for stakeholders.

**Techniques taken includes:**

* Examining Data Distribution.
* Data Cleaning and Preprocessing (also involving identifying outliers).
* Exploring relationships between variables.
* Uncovering patterns that can inform decision-making in the real estate industry.

1. ***Background***

The Office of Policy and Management maintains a listing of all real estate sales with a sales price of $2,000 or greater between October 1 and September 30 of each year. For each sale record, the file includes town, property address, date of sale, property type (residential, apartment, commercial, industrial, or vacant land), sales price, and property assessment.

1. ***Data loading and description:***

* This dataset ‘Real\_Estate\_Sales\_2001-2020 GL’ includes information such as the assessed value, sale amount, sales ratio, property type, and more, offering a rich source of insights into the real estate landscape. The dataset consists of 997,213 rows and 14 columns.
* The columns include serial number, List year, Date recorded (for property information), Town, Address of the Property, assessed value, Sale amount, Sales Ratio, Property Type, Residential Type, Non-Use Code, Assessor Remarks and OPM remarks.

1. ***Problem Statement***

* Exploring Property Assessment and Sales data for Informed Decision-Making.

1. ***Project Objectives:***

* Assessment Accuracy: Evaluate the accuracy of property assessments by comparing assessed values with actual sale amounts, and identify any discrepancies.
* Market Trends: Analyze sales ratios to uncover trends in property market values, understanding how sale amounts relate to the assessed values across different property types and residential classifications.
* Geographical Analysis: Investigated variations in assessed values, sales amounts, and market ratios across different towns providing a localized understanding of real estate dynamics.
* Property Type Impact: Examine the impact of property type on assessment accuracy and market trends, exploring whether certain types (e.g. residential, commercial) exhibit distinct patterns.
* Non-Use Code Insights: Explore the significance of non-use codes in property assessment investigating how these codes influence assessed values and sales transaction.
* Assessor and OPM Remarks: Analyze remarks provided by assessors and the Office of Polic and Management (OPM) to identify factors influencing assessment decisions and potential areas for improvement.

1. ***CodeBase***

This jupyter notebook ‘REAL\_ESTATE\_SALES\_EXPLORATORY\_ANALYSIS.ipynb’ contains the code for the exploratory data analysis. It includes sections for data loading, data cleaning, data visualization, summary statistics and insights derived from the analysis.

1. ***Data Cleaning:***

* Handling missing values by Imputation *(“Simple Imputer”).*
* Checked for duplicates and removed if present.
* Corrected data types for accurate analysis.
* Handled Outliers using Interquartile Range Method (IQR)

1. ***Data Exploration (Univariate and Bivariate analysis) and Visualization:***

* Explored the distribution of sale amount and assessed values through histograms and summary statistics.
* Analyzed property types, residential types and towns with their frequencies using bar plots.
* Explored statistical summary on all numerical values in the dataset.
* Estimated the count of properties in each year from 2001 to 2020 using bar plot
* Investigated the relationship between sale conditions and property/residential characteristics.
* Explored the significance of non-use codes in property assessment.
* Evaluated the accuracy of property assessments by comparing assessed values with actual sale amounts, and identify any discrepancies.
* Analyzed sales ratios to uncover trends in property market value across different property types and residential classifications.
* Investigated variations in assessed values, sales amounts, and market ratios across different towns.
* Analyze remarks provided by assessors and the Office of Polic and Management (OPM).
* Analyzed the average sales ratio evolution across the years.

1. ***Feature Engineering:***

* Created new features such as Discrepancies, New Sales Ratio, Market Ratio to enhance analysis.

1. ***Outlier Detection:***

* Identified outliers in sale prices, assessed values, sales ratio and addressed them appropriately.

1. ***Summary Statistics:***

* Found that Property type has a positive correlation with Residential type as well as high positive correlation between Assessed value and Sale Amount.
* Identified common discrepancies and outliers.
* Identified a higher frequency of sales in certain property types like **Single-family homes**.
* Based on the Mean Sales ratio, a potential discrepancy between assessed values and actual sale amounts was indicated.
* **Single Family property type** have the highest sales ratio which means at average, the assessed value of Single-Family homes higher than their actual sale amounts, indicating strong demand or potentially overvalued properties in this category with their assessed values significantly exceeding their actual market sale prices.
* **CONDO residential type** had the lowest Sales Ratio indicating that the sale amounts are below the assessed values which means Condo residential types are selling below their assessed values, potentially a lower demand compared to other property types.
* The year with the highest average Sales Ratio appears to be **2011**.
* Found out that **Bridgeport** was the most frequently occurring town with count of 34201.
* **New Canaan Town** has the highest average Assessed value of 412,950 while **Darien Town** have the highest average Sale amount of 659,238.
* **Stafford Town** has the highest market ratio of 1.13 indicating assessed values are slightly higher than sale amounts on average.
* **Industrial Property type** have the highest average of Assessment Value and Sale Amount.
* **Public Utility Property type** have the lowest average of Assessment Value and Three-Family property type have the lowest average of Sales Amount.
* In terms of **Non-Use-Code***, 25-Other* has the highest value count followed by *14-Foreclosure* and *07-change in Property*.
* The **OPM Remarks** in the dataset are described as "GOOD SALE PER MLS" for most of all entries.

1. ***Dependencies:***

* Python 3
* Data Visualization (Matplotlib, Seaborn WordCloud)
* Jupyter notebook
* Data manipulation (NumPy, Pandas)
* Handling missing values (Scikit-Learn)
* Word preprocessing (Natural Language Toolkit(‘NLTK’))

1. ***Conclusion:***

* This EDA revealed insights into the factors influencing real estate sales, including property/residential characteristics and sale condition.

**RECOMMENDATIONS**

1. Improve the accuracy of property assessments and ensure consistency between assessed values and actual sale amount.
2. Investigate the cases where the discrepancy between assessed value and sale amount is significant (e.g., cases where the discrepancy exceeds the 75th percentile value of $158,900), determine the underlying reasons for these discrepancies, such as errors in assessment, market fluctuations, and property conditions.
3. Implement measures to improve the accuracy and reliability of property assessments. This may include enhancing assessment procedures, providing additional training to assessors, and conducting regular audits to identify and rectify discrepancies.
4. Evaluate the methods used for property assessment to ensure they accurately reflect the current market value of properties.
5. Establish a system for continuous monitoring and evaluation of assessment practices and discrepancies over time. Regularly review and analyze discrepancies to identify trends, patterns, and areas for improvement in the assessment process.
6. Foster collaboration between assessors, property owners, real estate professionals, and other stakeholders to exchange feedback and insights regarding assessment discrepancies.
7. Focus on maintaining competitive pricing strategies to capitalize on steady demand and explore investment opportunities in multi-family properties for rental income and long-term appreciation.
8. Consider negotiating favorable deals and value-added strategies like redevelopment to maximize returns
9. Conduct detailed market analysis in towns with lower average assessed values but higher sale amounts to understand the dynamics driving sales in those areas. There may be opportunities for growth or development.
10. Stakeholders should leverage data analytics and market intelligence tools to analyze sales trends, pricing dynamics, and demand-supply factors for various property types. This information can inform valuation decisions and help stakeholders understand how each property type is performing in the market.
11. Implement regular reviews and validations of property assessments involving non-use codes to ensure consistency and accuracy.
12. Recommend implementing a feedback mechanism to gather input from assessors and stakeholders on the effectiveness of utilizing Assessor Remarks in property assessments, enabling continuous improvement.

***By reviewing these recommendations, stakeholders can enhance the accuracy and reliability of property assessments, reduce discrepancies between assessed values and sale amounts, and ultimately improve trust and confidence in the assessment process among all stakeholders involved.***

***REFERENCES:***

Dataset Source: <https://drive.google.com/file/d/1xAUeLxh2n4Qmu43OzV24r-uXaeGBSqrY/view?usp=sharing>.