Real Estate Sales Analysis

Step by Step Process of Analysis using Python on Jupyter Notebook:

- Loaded the data on Jupyter Notebook and looked at the raw data.
- In the columns named Non Use Code, Assessor Remarks, OPM Remarks and Location we found more than 70% data loss. The values in these columns were redundant and is of no use, so we dropped the column.
- We handled the null values in the column, Residential Type, Property Type and Address.
- Converted the Date Recorded column from object to datetime format.
- The columns Assessed Value and Sale Amount had 0 and 1 as the value. As the data is related to real estate, the Assessed Value and Sale Amount of any listed property can never be such low amount and so we removed these values from the data.
- Using the boxplot for Assessed Value, Sale Amount and Sales Ratio, we understood that there are outliers in the data and data is skewed too much and will disrupt the analysis. So, we used the quantile function to get the IQR (Inter Quartile Range) and set the data as per the upper bound and lower bound data of these columns.
- Later we saved the file in excel for further visualization in Tableau.

Visualization on Tableau:

- We created Scatter plot graph using Assessed Value and Sale Amount we understood that there is a clear positive correlation as Assessed Value increase, the Sale Amount also increase.
- We crated the bar chart and pie chart to see the data distribution in Residential Type column and Stacked bar chart for distribution of Residential Type across top 20 Town in the Connecticut, USA.
- We also understood that the Single Family Residential Type is highly overpowered in the data. But there was one trend we found after 2020, the Non-Residential rate heavily increased.
- We created these line graph and bar graph to understand the relation between the Assessed Value, Sale Amount and properties listed over the years.
- We also found that after 2007, there was increase in the listing of properties till 2010 along with that the Assessed Value and Sale Amount increased, but then after 2010 the listing dropped and along with that the Assessed Value and Sale Amount dropped on a higher rate. By 2023 the Assessed Value and Sale Amount depreciated more than what it was back in 2007.