HOUSING DATASET

GOWTHAM S

DA&DS FEB 2025

**OBJECTIVES :**

The housing dataset provides comprehensive information on various attributes associated with residential properties, including price, number of bedrooms and bathrooms, square footage, location details, and other relevant features. The objective of this project is to conduct an in depth analysis of the dataset to derive valuable insights for stakeholders in the real estate industry.

**DATASET OVERVIEW :**

The dataset contains following information:

1. Date: The date when the property information was recorded.

2. Price: The price of the residential property.

3. Bedrooms: The number of bedrooms in the property.

4. Bathrooms: The number of bathrooms in the property.

5. Sqft\_living: The total square footage of living space in the property. 6. Sqft\_lot: The total square footage of the lot or land area associated with the property.

7. Floors: The number of floors in the property.

8. Waterfront: Indicates whether the property has a waterfront view (binary: 0 for no, 1 for yes).

9. View: An index from 0 to 4 representing the quality of the view from the property.

10. Condition: An index from 1 to 5 representing the overall condition of the property.

11. Sqft\_above: The square footage of the interior space above the ground level.

12. Sqft\_basement: The square footage of the basement space in the property.

13. Yr\_built: The year when the property was built.

14. Yr\_renovated: The year when the property was last renovated. 15. Street: The street address of the property.

16. City: The city where the property is located.

17. Statezip: The state and zip code of the property.

18. Country: The country where the property is located.

**TOOLS USED :**

* Python

**STEPS INVOLVED :**

* DATA CLEANING
* Handled Missing Values
* Removed Duplicates
* Replaced Null values
* Datatype Changed
* **Read the Housing dataset**

Df = pd.read\_csv('housing.csv')

* **Rename the table header names**

df.rename (columns={'date' : ' Registration\_Date',

'price' : 'Price',

'bedrooms' : 'Bedrooms',

'bathrooms' : 'Bathrooms',

'sqft\_living' : 'SQFT\_Living',

'sqft\_lot' : 'Overall\_Lot\_SQFT',

'floors' : 'Floors',

'waterfront' : 'Waterfront',

'view' : 'View',

'condition' : 'Condition',

'sqft\_above' : 'Above\_SQFT',

'sqft\_basement' : 'Basement\_SQFT',

'yr\_built' : 'Built\_Year',

'yr\_renovated' : 'Renovated\_Year',

'street' : 'Street',

'city' : 'City',

'statezip' : 'State\_Zip',

'country' : 'Country'},inplace=True)

* **Change the O and 1 into 'Yes' and 'No'**

df['Changed\_waterfront'] = df['Waterfront'].apply(lambda x: 'Yes' if x==1 else 'No')

* **Change the condition values**

condition\_mapping = {

1: "Poor",

2: "Fair",

3: "Average",

4: "Good",

5: "Excellent"}

df['Condition'] = df['Condition'].map(condition\_mapping)

* **Change the view values**

view\_mapping = {

0: "No view",

1: "Poor view",

2: "Average view",

3: "Good view",

4: "Excellent view"}

df['View'] = df['View'].map(view\_mapping).astype(str)

**OUTLIERS:**

checking\_cloumns=['Price','SQFT\_Living','Overall\_Lot\_SQFT','Above\_SQFT','Basement\_SQFT']

def outliers\_removing\_function(df,columns,inplace=True):

for i in columns:

Q1 = df[i].quantile(0.25)

Q3 = df[i].quantile(0.75)

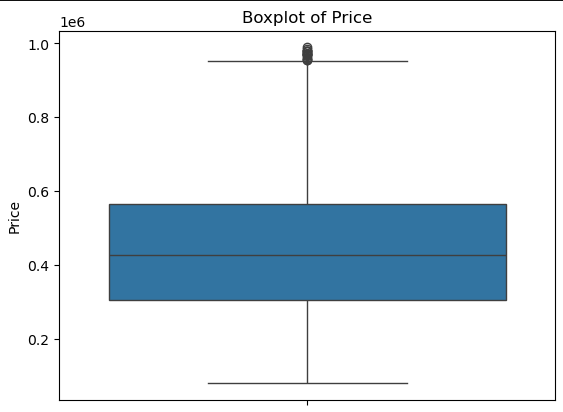
IQR = Q3 - Q1

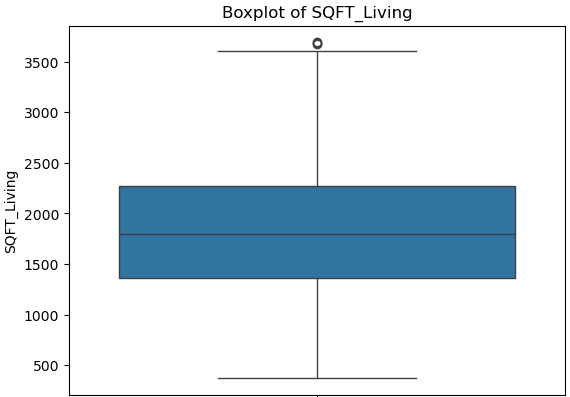
lower\_bound = Q1 - 1.5 \* IQR

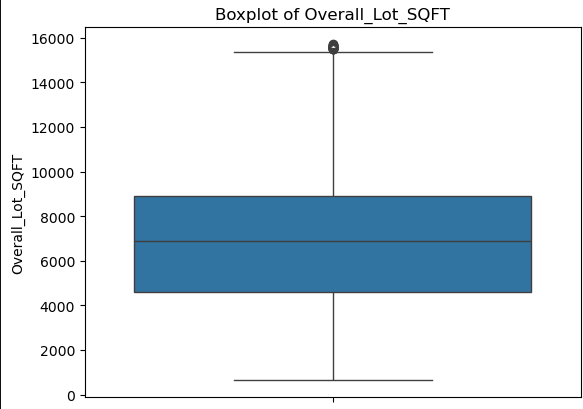
upper\_bound = Q3 + 1.5 \* IQR

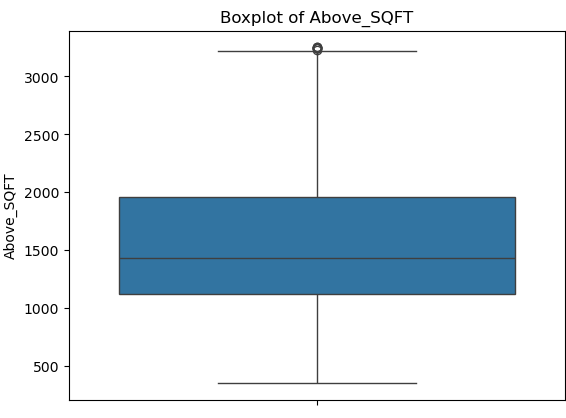
df = df[(df[i] >= lower\_bound) & (df[i] <= upper\_bound)]

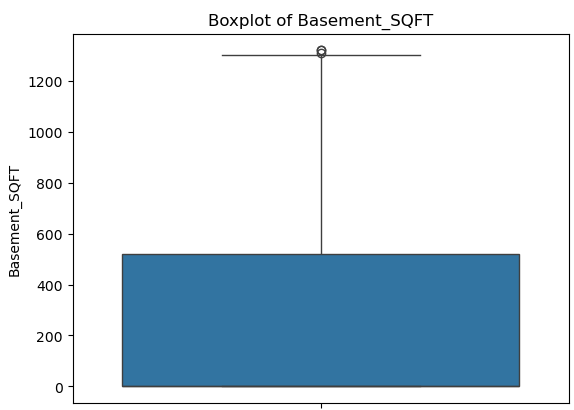
return df









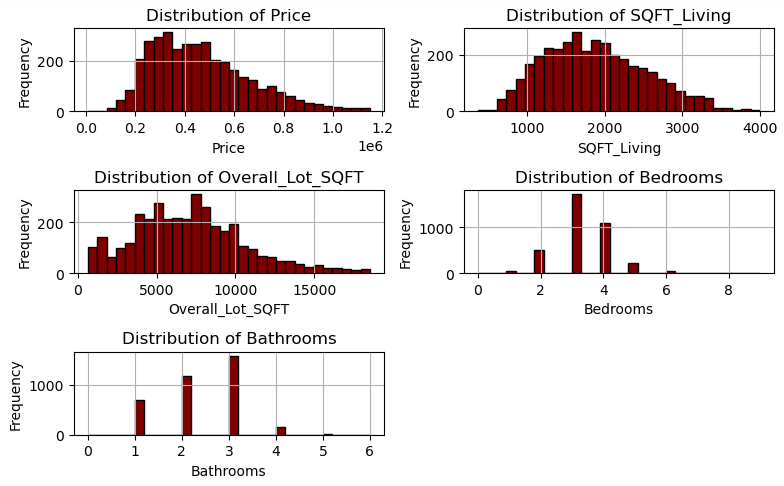




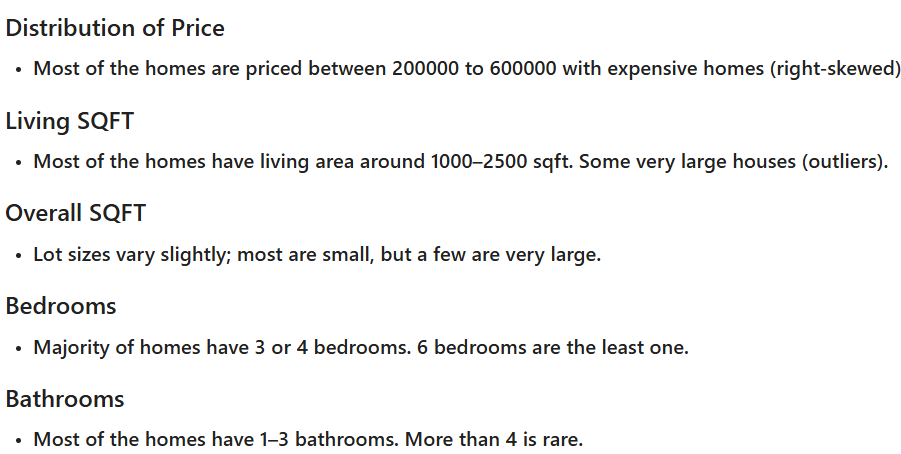
**Exploratory Data Analysis (EDA):**

1. **Univariate Analysis:**

Explore distributions and summary statistics of individual variables such as price, square footage, and number of bedrooms and bathrooms.



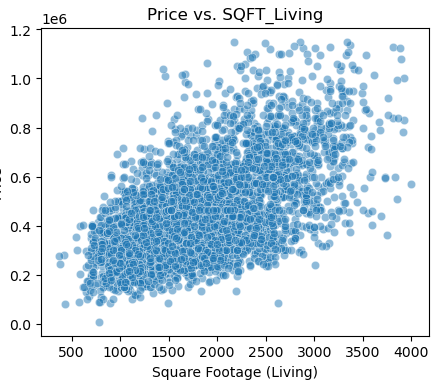
* **INSIGHTS**



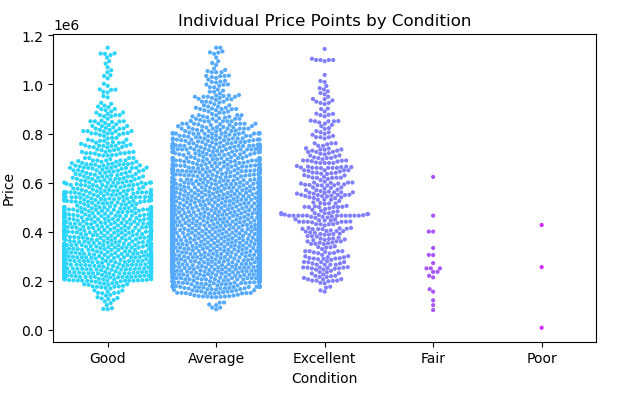
1. **Bivariate Analysis**

Bivariate Analysis: Investigate relationships between pairs of variables, such as price vs. square footage, bedrooms vs. bathrooms, etc., using visualizations and statistical methods.

* **Relationships between price vs sqft\_living**



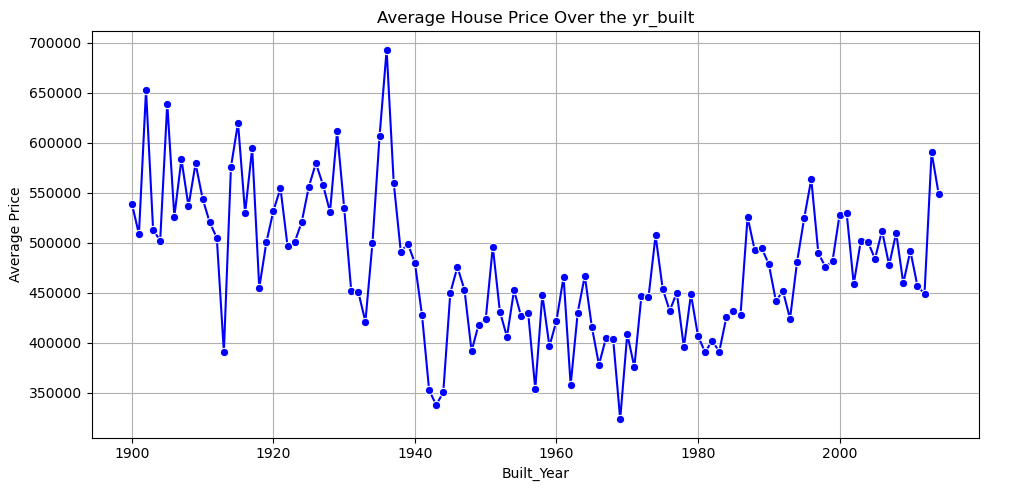
* **Price by Condition**



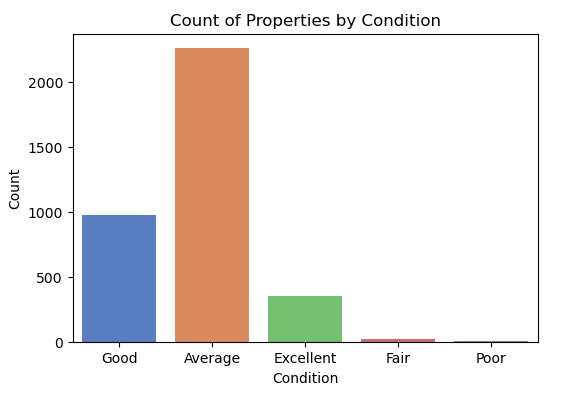
* **Relationship between price and view**



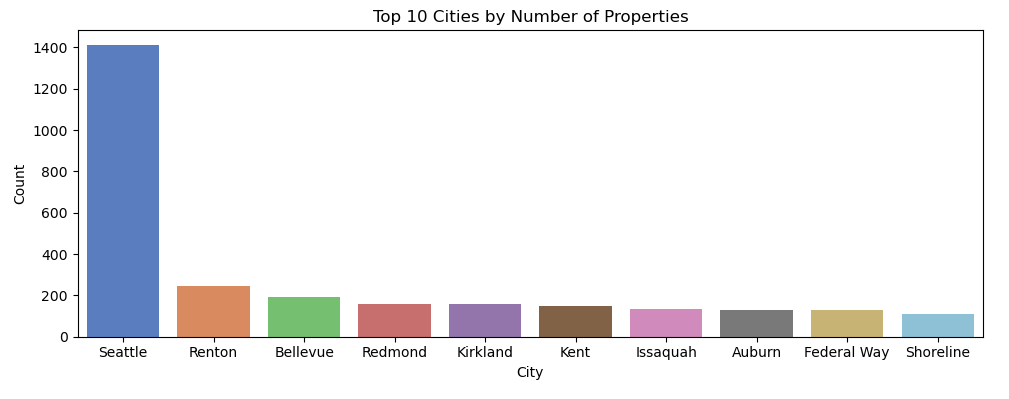
* Average House Price Over the yr\_built



* Count of Properties by Condition

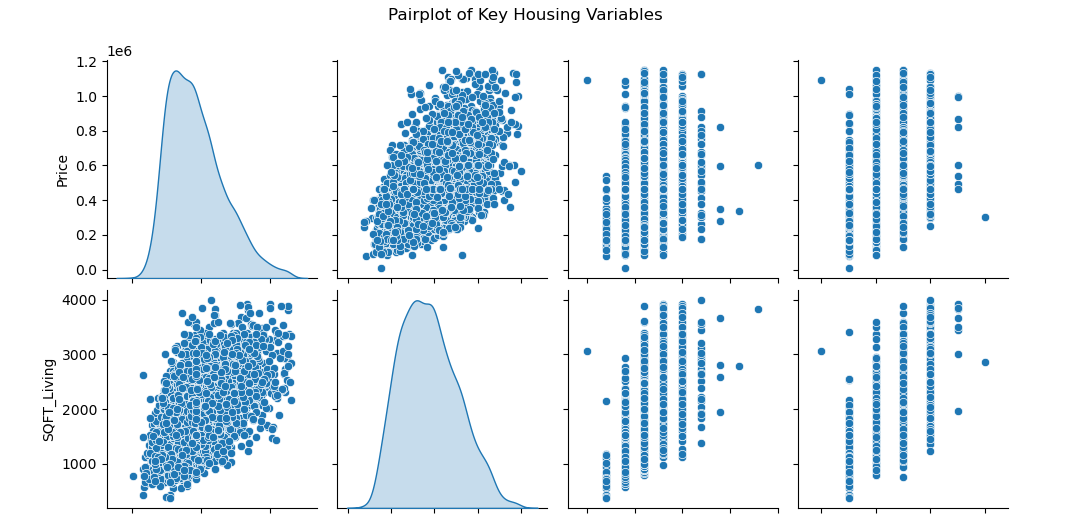


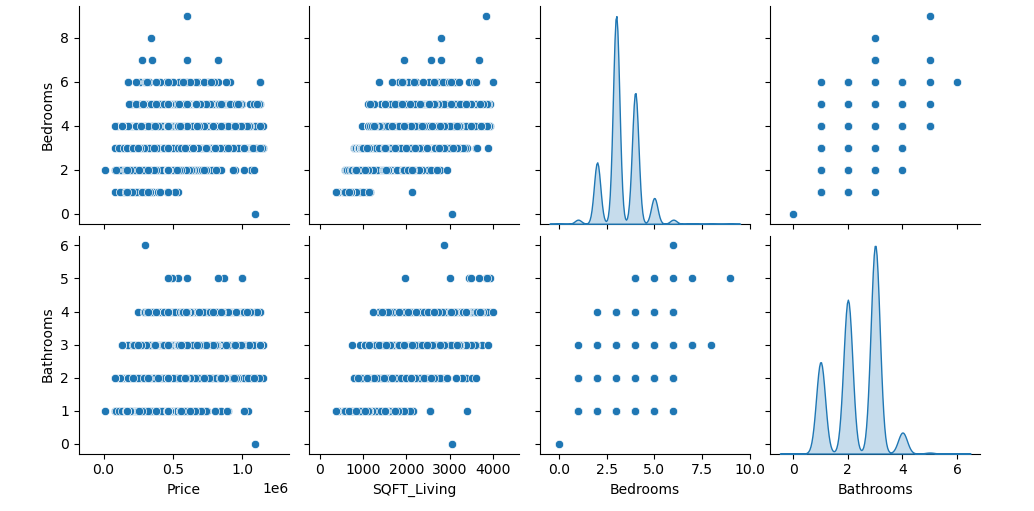
* Number of Properties per City



1. **Multivariate Analysis**

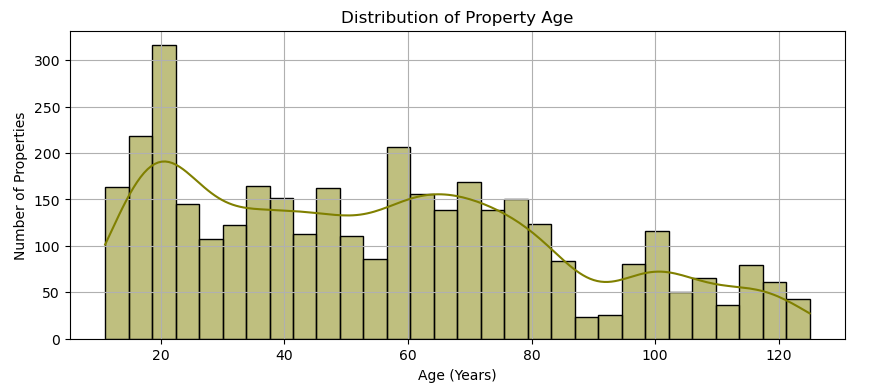
Multivariate Analysis: Examine interactions and dependencies among multiple variables, identifying correlations and patterns that may influence housing prices.

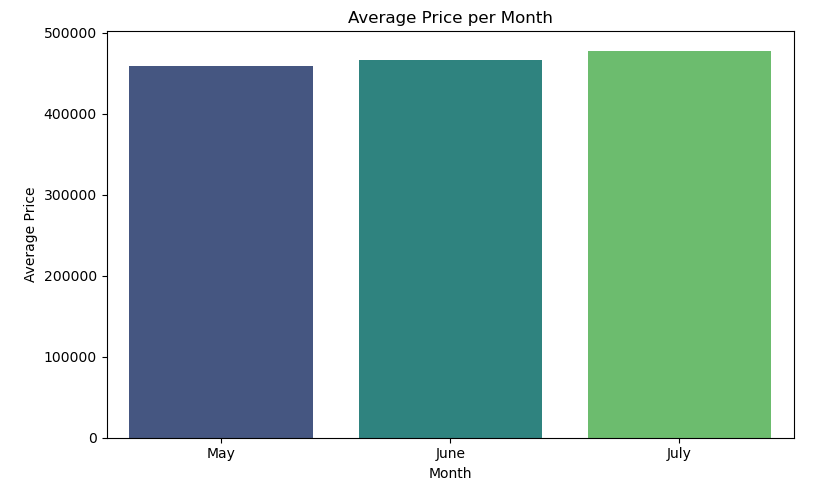




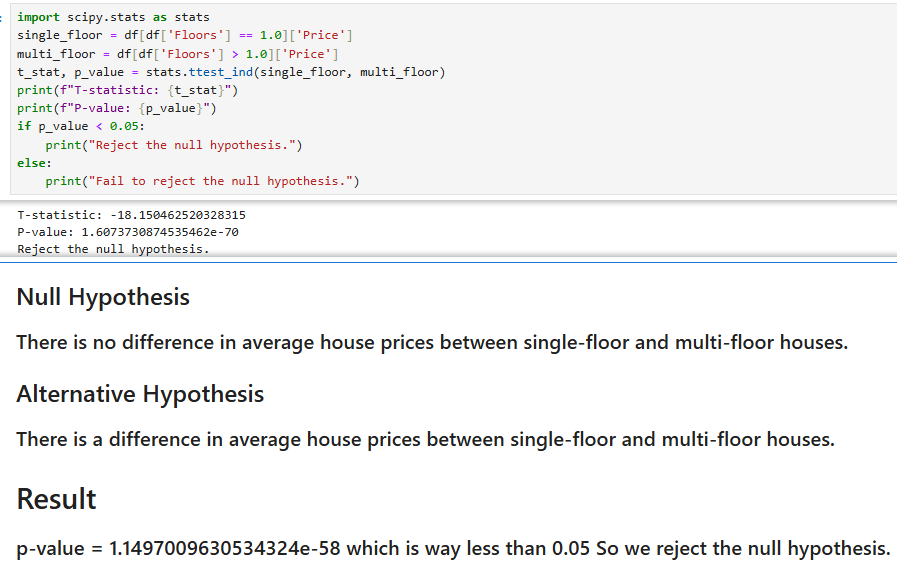
Feature Engineering

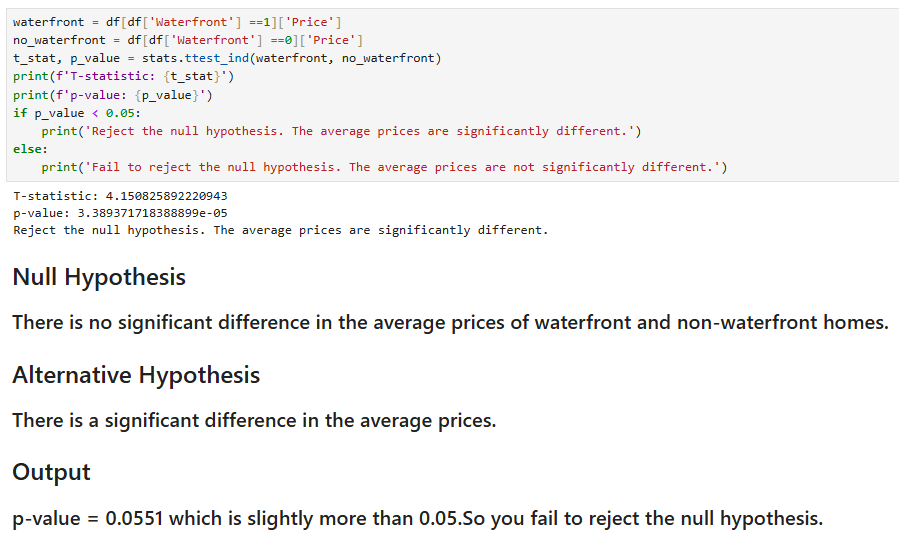
Extract or create new features that may enhance predictive models or provide additional insights into housing prices, such as age of the property, price per square foot, etc. • Perform dimensionality reduction if necessary to simplify the dataset while preserving important information.





* **Hypothesis Test**
* Performing a hypothesis test in two-sample t-test to compare house prices between
* Houses with single floor
* Houses with multi floor





OVERALL INSIGHTS

* **Price Insights**
* July has the highest average housing price compare to other months.
* Most housing prices between 250,000 to 600,000.
* Some properties are priced above 900,000, indicating high-value outliers.
* **City-Level Insights**
* Beaux Arts Village is the most expensive city with average prices over 750,000 this is likely luxury housing area .
* Seattle has 1300+ properties making it the most common city in major housing hub .
* **Multivariate & Trend Insights**
* SQFT\_Living has a positive correlation with Price larger homes cost is more.
* Most homes have 3 or 4 bedrooms and 1–3 bathrooms.
* Homes in excellent condition are few but are priced the highest.
* Most homes are in good or average condition.
* **Investment Recommendations**
* Focus on properties in Seattle for volume-based opportunities.
* Luxury areas like Beaux Arts Village can target rich buyers.
* Prioritise houses with good views and newer construction for better resale value.

**RECOMMENDATIONS**

* Invest in renovations for older homes to increase property value and appeal.
* Focus on high-view or waterfront properties for premium pricing potential.
* Use sqft based pricing strategies combined with condition and location.
* For sellers: list properties in peak months (May, June) for better visibility and pricing.