Kevin Kiding

Housing Price Prediction

Background:

In the volatile world of real estate, the ability to forecast property prices is as essential as a six-gun on a cowboy's hip. It is what distinguishes the shrewd from the foolish, the prudent from the rash. One can make informed decisions, plan their finances, and analyze the ever-changing market with this formidable ability. Partner, this skill can make or ruin a transaction. Accurate forecasts are like a trusty six-shooter on this wild frontier of commerce, where both consumers and sellers travel. They enable these courageous individuals to navigate the perilous terrain of complex markets with confidence, enabling them to make shrewd decisions and devise devious strategies. With such insight, they confidently ride, knowing that success is within reach.

Objective:

The purpose of the research is to conduct an exploratory data analysis (EDA) to investigate and evaluate the provided dataset in order to identify and clarify underlying patterns, relationships, and valuable insights. This study aims to enhance the decision-making process in relation to property prices and their associated characteristics. Using data visualization and analysis techniques, it is possible to gain valuable insights regarding categorical and numeric variables. This contributes to the advancement of subsequent analysis stages.

Introduction

I'm excited to share my precise data processing methods with you. Explore this amazing Google Colab link to see how information can change your life. Be astonished! Click this intriguing link to explore data processing. This fascinating procedure will enchant you. Explained code bits are intriguing. Prepare for an exciting adventure that will teach you data processing. Take advantage! Select now for this amazing voyage. Data processing techniques are difficult, yet they uncover a wealth of knowledge. Be intrigued as you learn the workflow's secrets and how it operates. Prepare for an unforgettable discovery. Access the intriguing links immediately to discover many exciting options.

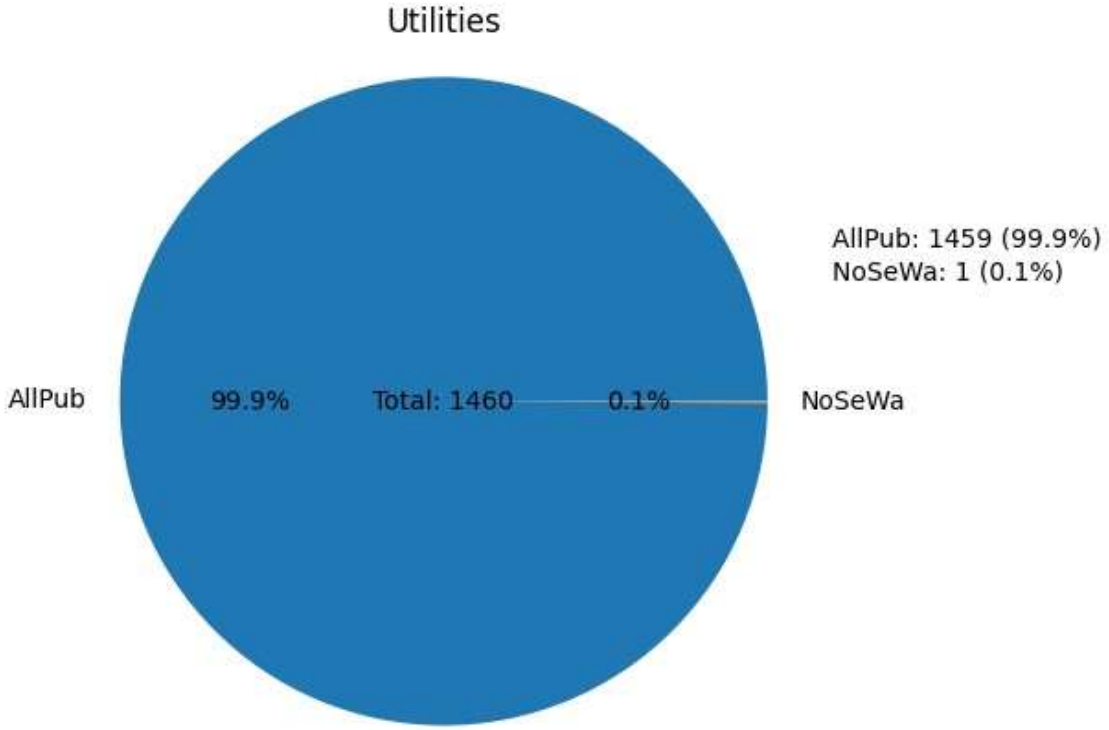
Click Here ☺

Thank you for your attention

Utilites

This document provides information about the amenities and services available at the location.

- The term "AllPub" indicates that the property has unrestricted access to all publicly available utilities, including Electricity (E), Gas (G), Water (W), and Sanitation (S).
- The term "NoSeWa" indicates that the residence has limited access to electricity (E) and gas (G). It is possible that properties with this value lack access to public water and sanitation systems.

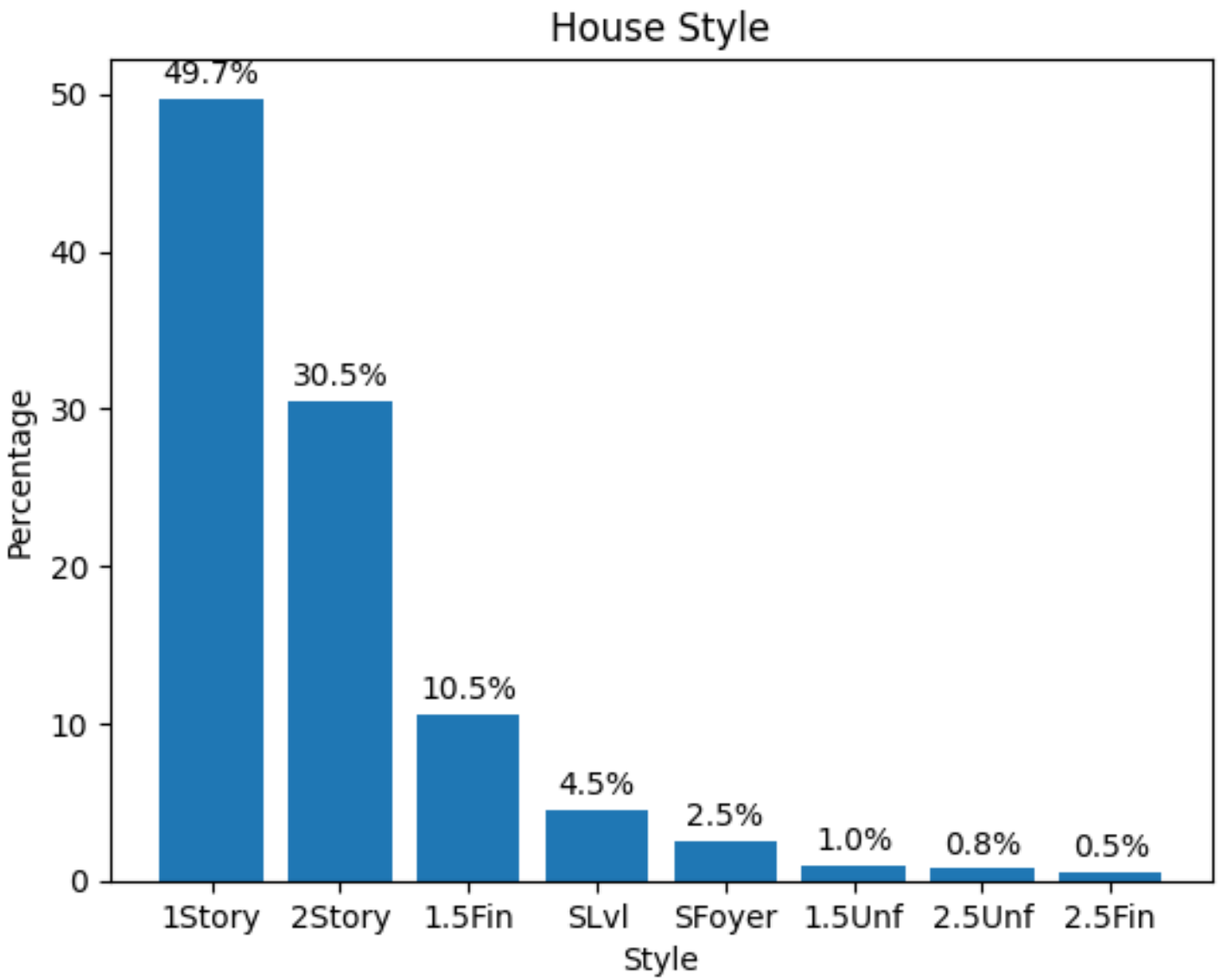


- The utility classification 'AllPub' is assigned to the vast majority of properties (99,99%), indicating that they have access to all public utilities.
- One property, or 0.01% of the total, lies under the utility type 'NoSeWa', signifying limited utility access.

This analysis provides purchasers, property developers, and policymakers with valuable information regarding the availability and accessibility of essential services. This is accomplished by displaying the distribution of utility categories within a particular area.

House Style

The HouseStyle box tells you about the design or style of each home listed. The given data shows how single-story, two-story, half-story, split-level, and other types of houses are spread out based on their architectural styles. It shows both the total number of homes in each type and the percentage of those houses. Getting to know the different types of homes can help you figure out what the latest building trends are in the housing market. With this information, people who work in the real estate business can make better investing choices.


House Style
49.7%
30.5%
10.5%
4.5%
2.5%
1.0%
0.8%
0.5%
50
40
30
20
10
0
Percentage
1Story
2Story
1.5Fin
SLvl
SFoyer
1.5Unf
2.5Unf
2.5Fin
Style

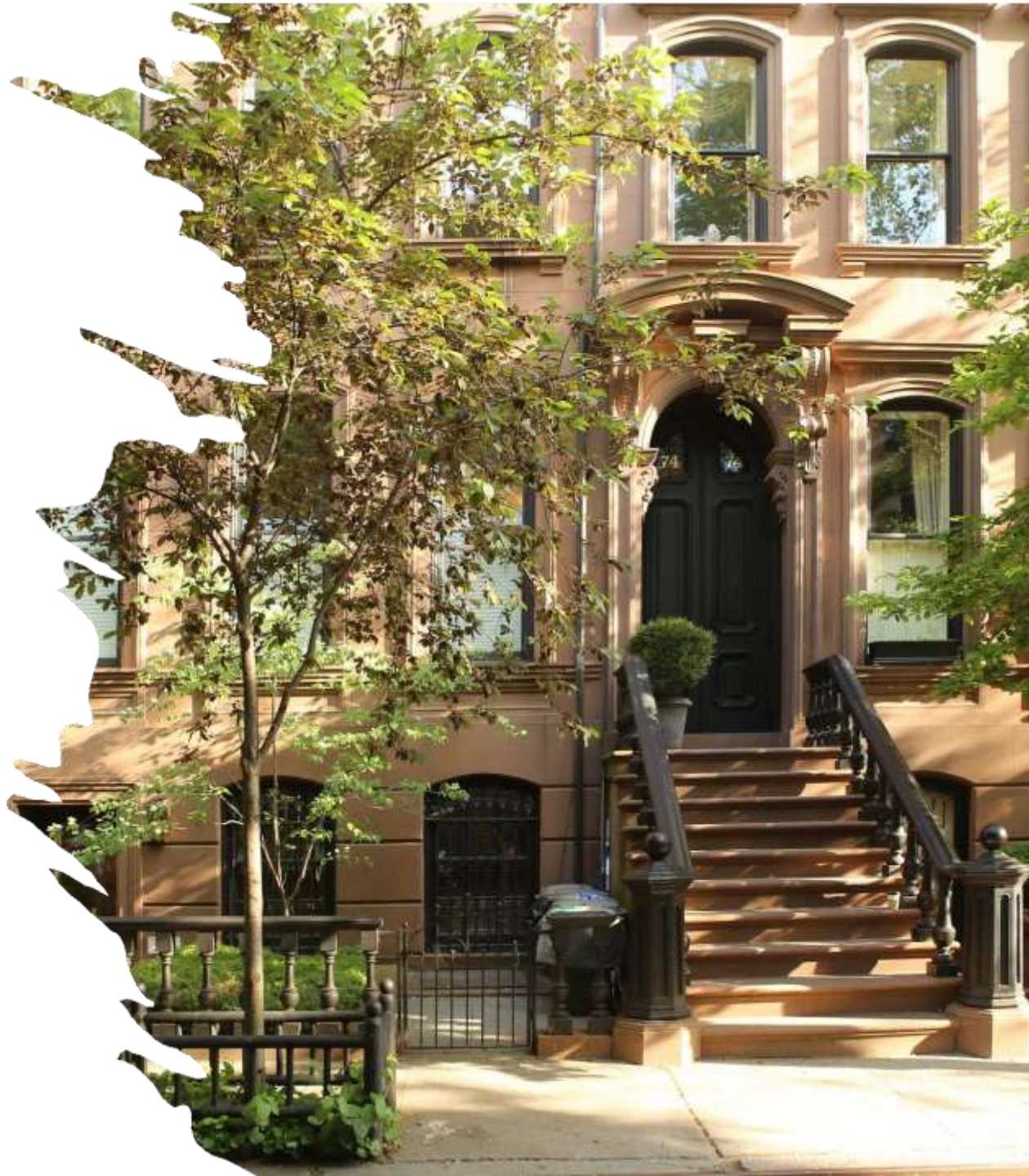
Explanation

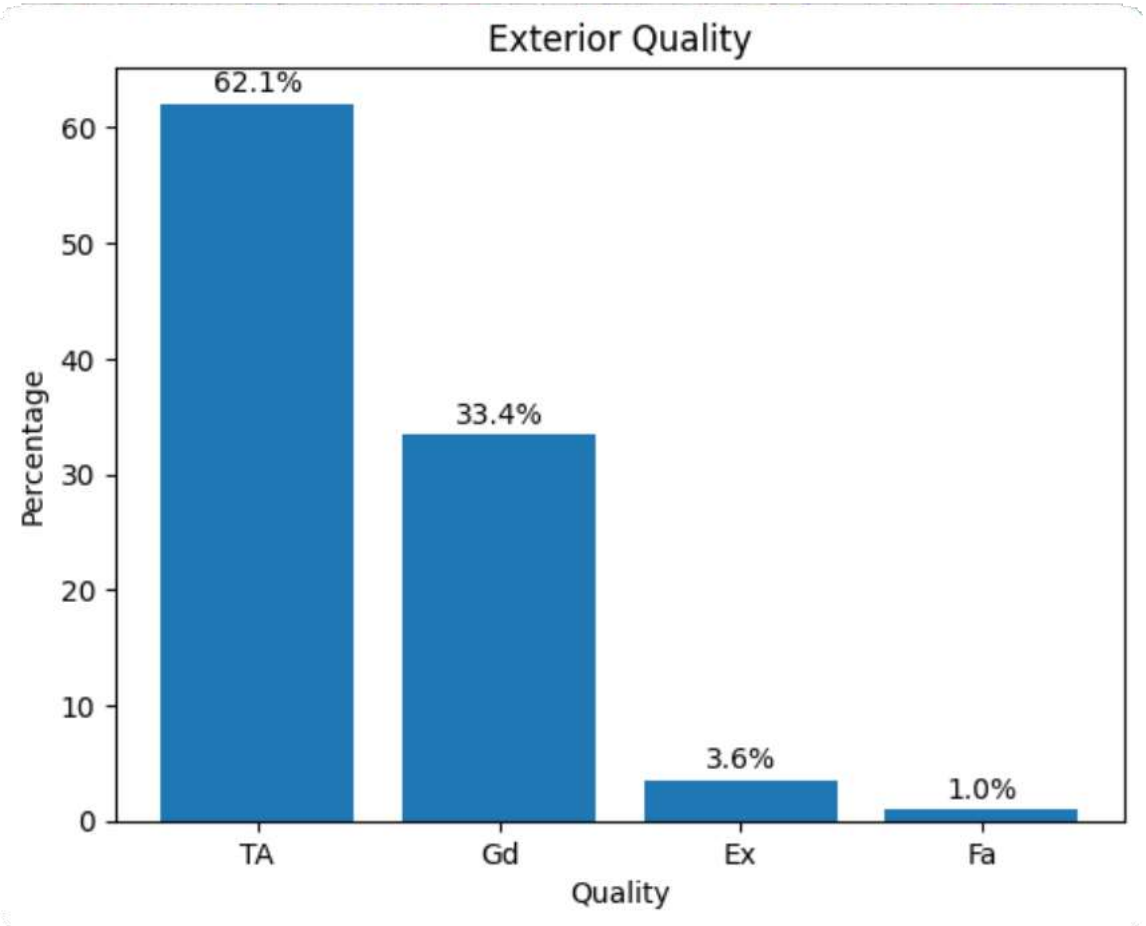
- This data shows people's housing preferences and market performance. 49.7% of homes are single-story. Accessing this one-story design is easy. 30.5% of homes are two-story, family-friendly designs.
- The 1.5-Fin house design (10.5%) comprises one and a half storeys and a finished second level.
- 4.5% of homes are S-Lvl split-levels with distinct living zones.
- 2.5% of homes have S-Foyer entrances to higher and lower floors.
- A 1.5-Unf home (1%) has an incomplete second storey. Customizing the dwelling is easy.
- 2.5-Unf (0.8%) is a two-story house with an unfinished second storey.
- 2.5-Fin houses (0.5%) have two and a half storeys and a finished second story.

Understanding the distribution of various dwelling types might inform housing sector choices regarding design preferences and market trends. Each type of property has unique features that attract buyers and affect its price.

Exterior Quality (ExterQual)

- In the dataset, the field "ExterQual" shows the grade of the material used on the outside of the building. This piece looks at how the outside of the house looks and how well it is kept. This field has numerical numbers that show different levels of quality, from "outstanding" to "not good enough."

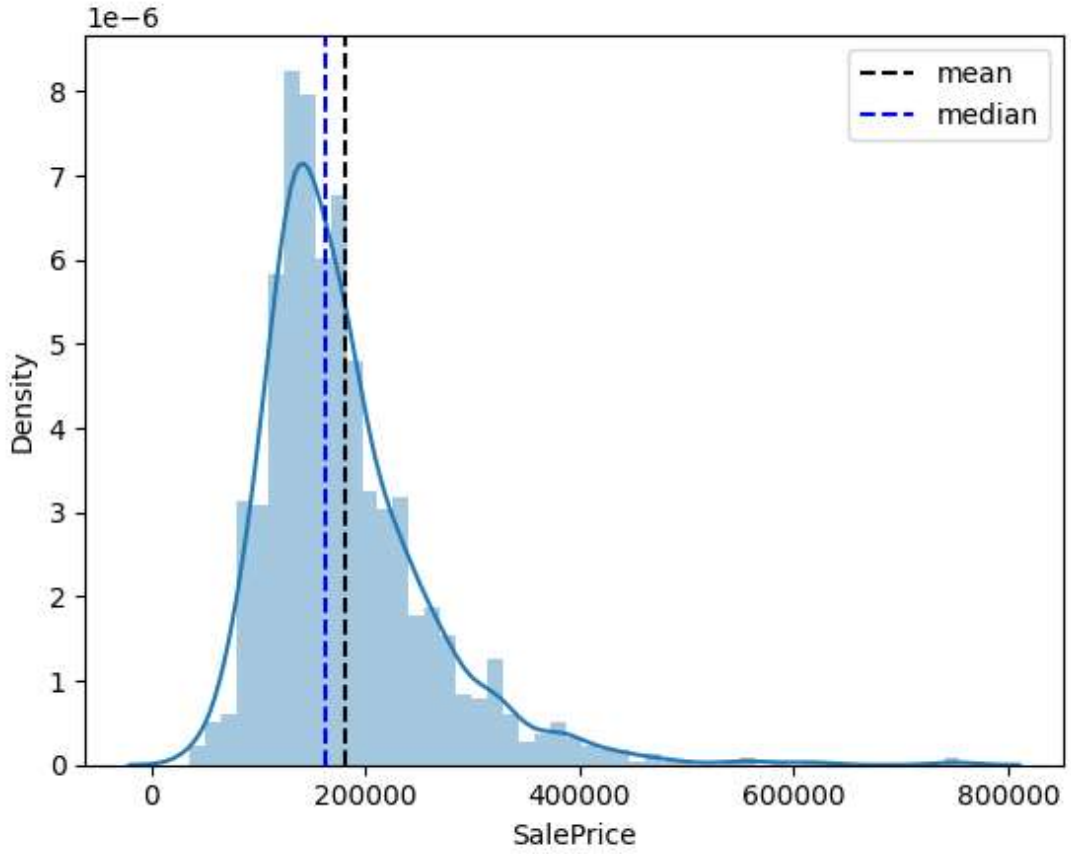




- A large number of properties, 62.1%, have a medium or average level of quality on the outside, while 33.4% are of excellent quality.
- ExterQual data is critical to understand in real estate deals because it can tell you important things about the outside of a building.
- How a house looks significantly affects its market value and how desirable it is to potential buyers. Properties with better curb appeal are more likely to sell for a higher price and attract more buyers.

SalePrice

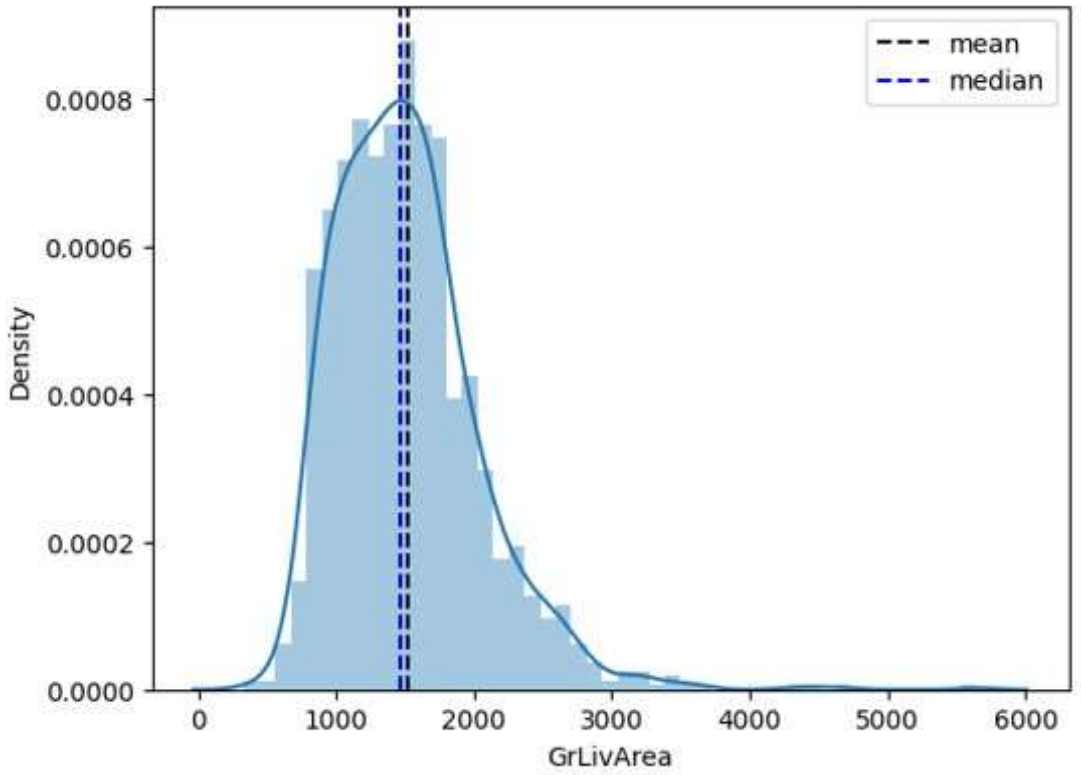
- The "SalePrice" column provides insightful information and facilitates our understanding of the numerous factors that affect property prices. The column's analysis facilitates well-informed judgments concerning the purchase or sale of real estate. It contributes to maximizing the property's value and profits. Individuals benefit from the information in the "SalePrice" column because it enables them to identify market trends, make necessary budget adjustments, and seize opportunities in the competitive real estate industry.
- When purchasing or selling real estate, individuals can make more informed and strategic decisions if they comprehend the factors that influence home prices.



- Histograms are used in this study to show how the prices of homes sold vary. The middle price, which is $163,000, and the average price, which is $180,921,196, get the most attention. The prices in the list are very different from one another.
- Even though there are more expensive homes, most of them cost between $100,000 and $300,000. This makes them a popular and attractive choice for many buyers.
- To make smart choices about real estate, you need to know how home prices are spread out and what the most usual price ranges are. With this knowledge, people can make good choices based on their own wants, tastes, and the factors that affect prices within these levels.

GrLivArea

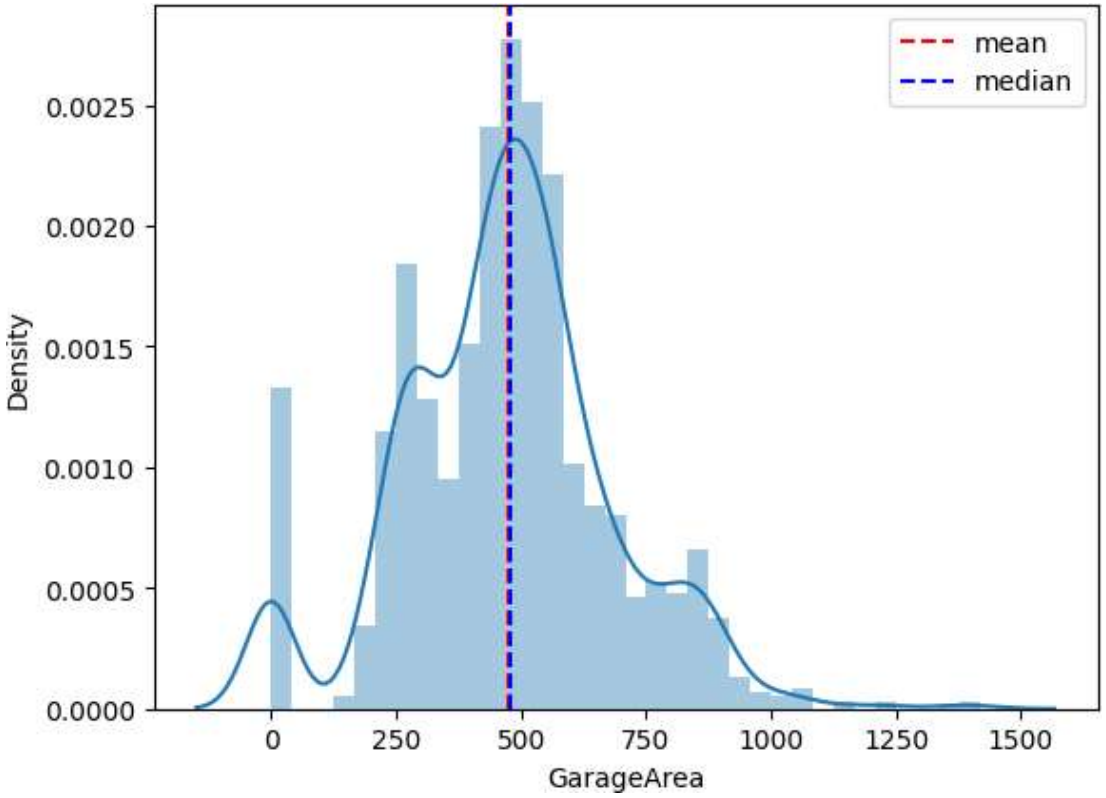
Within the dataset, there is a column labeled "GrLivArea" that details the total amount of livable space offered by each individual property. This information includes details concerning the size or extent of the house's livable space. The living space is significant because of the direct impact it has on the ease and efficiency with which the people living in the house go about their daily lives.



- The dataset includes a column called GrLivArea, which is the living area of each property. The distribution of this variable is right-skewed, so it suggests that the majority of homes in United States have with a living area ranging from 1000 to 2000 square feet.
- Understanding the dimensions of the property's living space has importance for purchasers as well as vendors to be able to make pertinent decisions.
- In terms of my opinion, the square footage representation represented by its size is significant in relation to both value and desirability. Generally speaking, a larger square footage points toward higher price point as well as greater interest from prospect buyers. Yet it is important to consider additional variables while assessing cost along with decision-making process associated with acquiring a property.

GarageArea

- The dataset includes a column called 'GarageArea' that represents the size of each property's garage. The garage area can provide useful information regarding a garage's ability to accommodate vehicles and provide additional storage space.



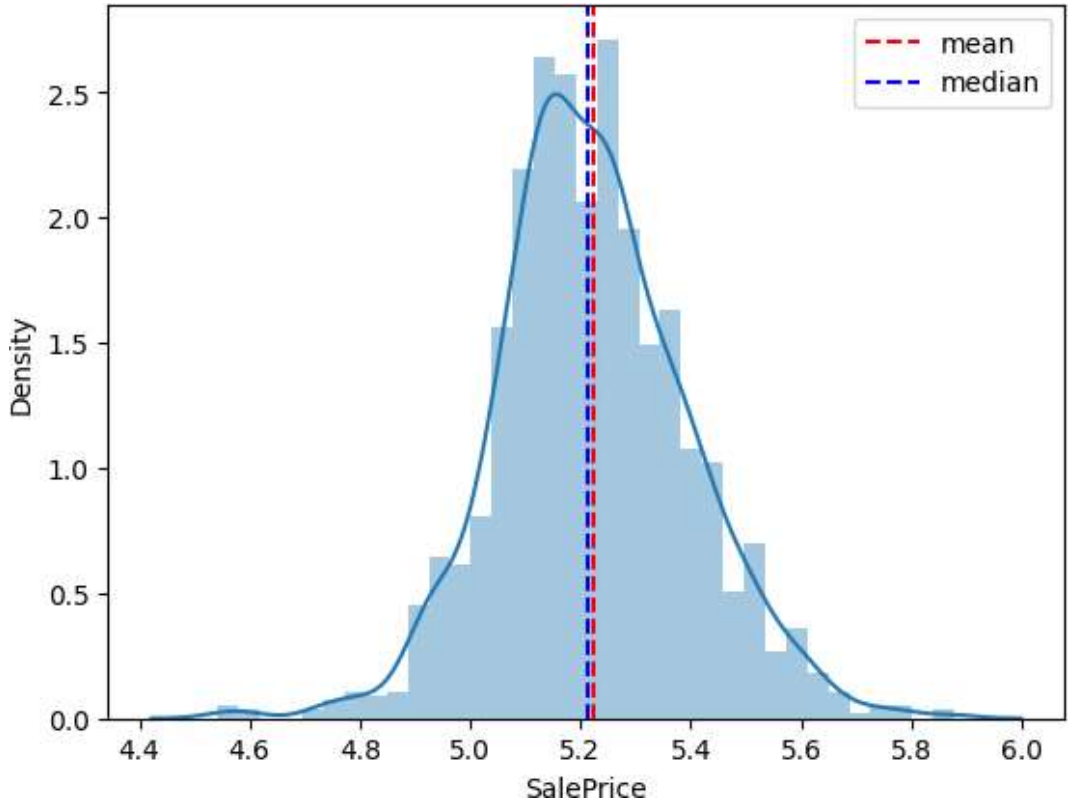
- The garage area for each property is indicated in the data set's GarageArea column. The distribution of garage data leans towards the left.
- Garages range in size from 250 to 860 square feet, according to the histogram representation. The majority of garages fall within this range, but there are also properties with garage areas outside of it, showing that garage sizes can differ.
- Determining the desires and requirements of both buyers and sellers of a property is crucial in evaluating the functionality of a garage and understanding its square footage. The convenience and value of a property can be influenced by the square footage of its garage. But it should be noted that the ideal size for a garage can differ based on individual preferences and needs.

Processing Data Numerical varaibel

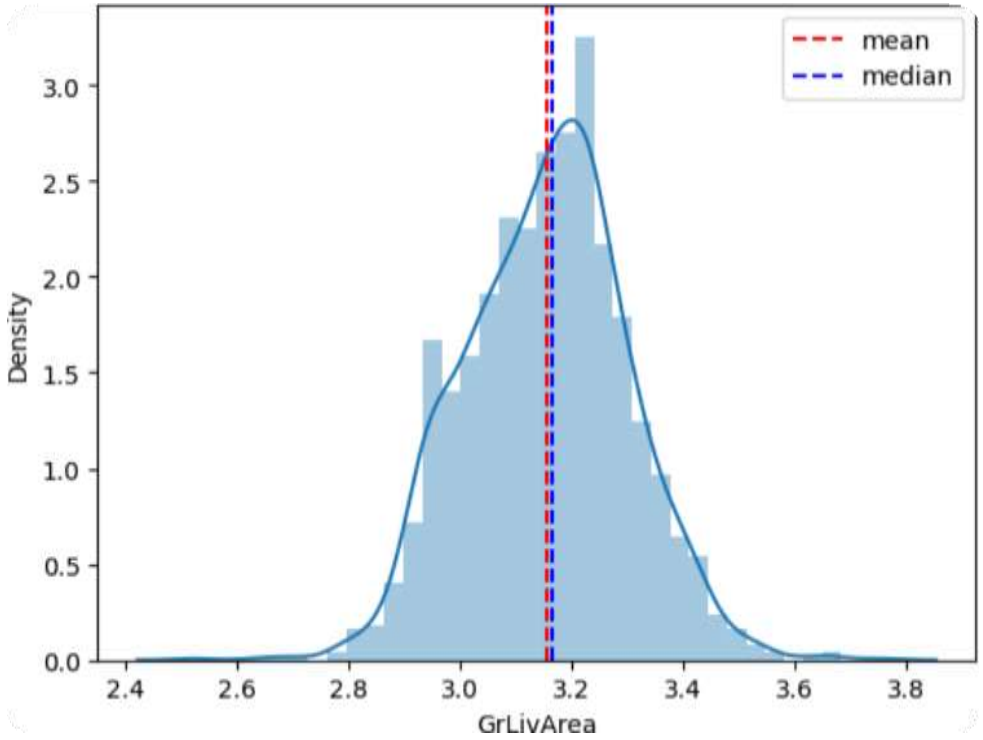
- In this analysis, the data in the SalePrice and GarageArea columns are transformed using the log-10 method. The purpose of this study is to improve understanding of the various factors that influence residential property prices. This will enable individuals to engage in real estate transactions with greater knowledge and discernment.
- The use of the log-10 technique for data processing in the SalePrice column normalizes the magnitude of home price values, thereby facilitating the comparison and analysis of variables such as location and property size.
- This methodology is extremely important because it facilitates a deeper comprehension and enables individuals to make well-informed property transactions decisions. By understanding the various factors that affect housing prices, individuals can confidently engage in property transactions and maximize the efficiency of their transactions.

SalePrice

- The 'SalePrice' column in the dataset represents the monetary value at which the house was sold. A log-10 transformation was applied to the column in order to enhance the comprehensibility of the variation in house prices.
- Following the transformation, the median value is observed to be 5,212, while the mean value is 5,222, accompanied by a standard deviation of 0.173. The log-10 transformation is used to proportionally rescale house price values, enabling more effective comparisons between houses of different sizes.
- The application of a log-10 transformation to the ‘SalePrice’ column is beneficial for addressing asymmetrical distributions and enhancing the understanding of patterns in house price fluctuations. The utilization of a standardized scale allows individuals to effectively compare house prices across various properties and obtain a more precise comprehension of the variations in house prices within this dataset.



GarageArea



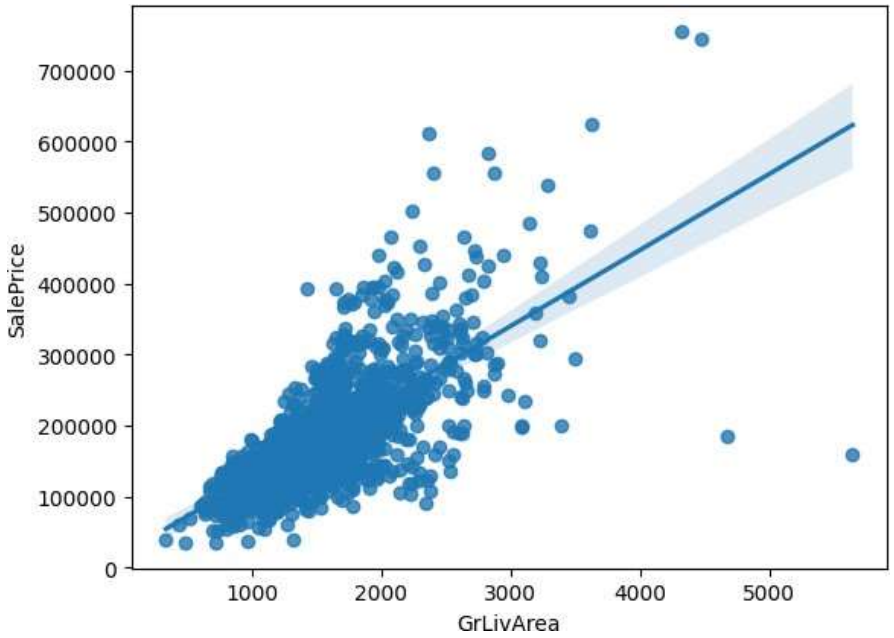
- The garage area of each property in the dataset is reflected in the 'GarageArea' column. A log-10 transformation was applied to this column to acquire a better understanding of the variance in garage area.
- The results of the transformation reveal a median of 3.166, a mean of 3.156, and a standard deviation of 0.145.
- The 'GarageArea' column's log-10 transformation delivers considerable benefits by leveling the data distribution and providing a deeper understanding of the variance in garage area. This variable is critical in establishing a property's value and attractiveness. The transition improves garage area data analysis, allowing for more informed property-related decision-making.

Delete zero values in the 'GarageArea' column

```
[ ] # see how many zero values there are in GarageArea
    print("Number of non-zero values: ", np.sum(house_numeric["GarageArea"] != 0))
    print("Number of zero values: ", np.sum(house_numeric["GarageArea"] == 0))

    Number of non-zero values:  1379
    Number of zero values:  81

    # Removing zero values from GarageArea
    x = house_numeric["GarageArea"][house_numeric["GarageArea"] != 0]
    sns.distplot(x, axlabel=x.name)
    line1 = plt.axvline(x.mean(), color='r', linestyle='--', label='mean')
    line2 = plt.axvline(x.median(), color='b', linestyle='--', label='median')
    first_legend = plt.legend(handles=[line1, line2], loc=1)

    print('Median value after removing zero values from GarageArea: {:.2f}'.format(x.median()))
    print('Mean value after removing zero values from GarageArea: {:.2f}'.format(x.mean()))
    print('Standard deviation value after removing zero values from GarageArea: {:.2f}'.format(x.std()))

    plt.show()

    <ipython-input-33-9d2ceeeb3fb5>:3: UserWarning:

    `distplot` is a deprecated function and will be removed in seaborn v0.14.0.

    Please adapt your code to use either `displot` (a figure-level function with
    similar flexibility) or `histplot` (an axes-level function for histograms).

    For a guide to updating your code to use the new functions, please see
    https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

      sns.distplot(x, axlabel=x.name)
    Median value after removing zero values from GarageArea: 484.00
    Mean value after removing zero values from GarageArea: 500.76
    Standard deviation value after removing zero values from GarageArea: 185.68
```

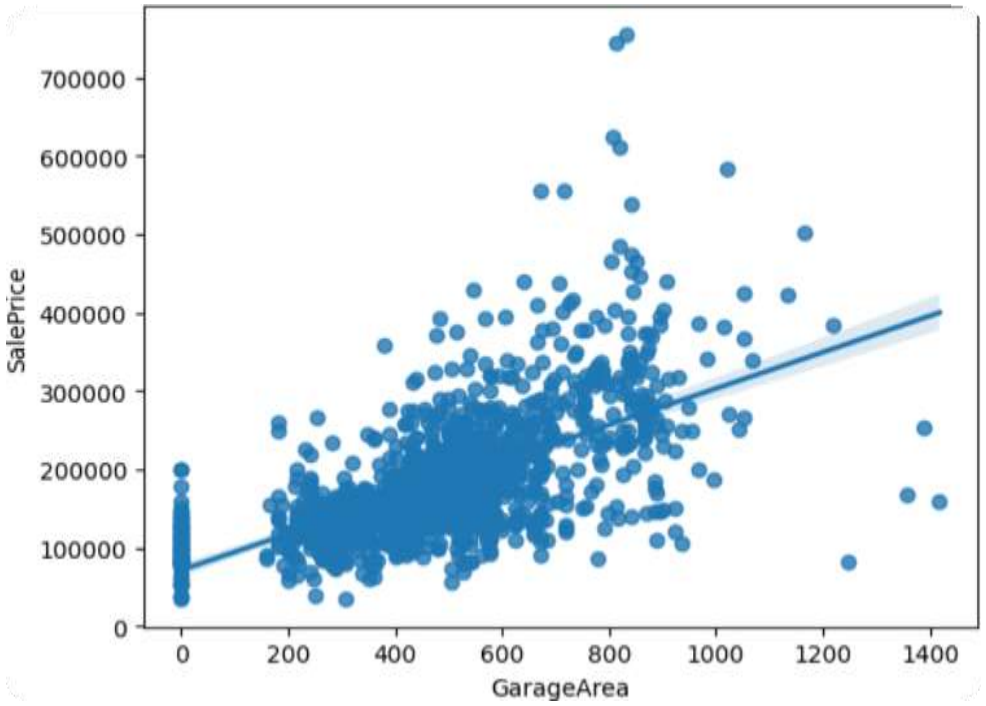
- To get a more accurate evaluation of the garage area in properties with garages, the analysis removed all instances where the 'GarageArea' column had a value of zero, indicating there was no garage on the property.
- After removing blank values, the results of this analysis show that garage area has a median value of 484.00, a mean value of 500.76, and a standard deviation of 185.68. Analyzing the variation in garage area between properties that have garages is important for property valuation and attractiveness.
- By removing properties with a value of zero, the analysis can be narrowed down to properties that specifically have garages, allowing for a more targeted examination of their characteristics. The ability to make more informed decisions when buying or selling a property is facilitated by this feature. Analysis of a property's garage area can provide valuable insights for optimizing property transactions and understanding the factors that influence home prices.

Relationship between ‘GrLivArea’ and ‘SalePrice’



- The correlation between the variables 'GrLivArea' and 'SalePrice' in the columns is depicted by the graph. The interaction between these two components increases their overall efficacy. In other words, the price of a property correlates positively with the extent of the residential area it encompasses. This observation illustrates the relationship between property values and the type of neighborhood in which a home is located.
- Due to its significant impact on the real estate industry, the extent of the residential area must be considered when determining the price of a property. This information provides vendors with insight into the increased value of a larger living space, which can indicate the level of comfort and desirability to potential buyers.
- By comprehending the positive correlation between square footage and selling price, buyers can make more informed purchase decisions and accurately assess property values. By recognizing the relationship between these two factors, real estate organizations can boost customer satisfaction and improve pricing strategies.

Relationship between ‘GarageArea’ and ‘SalePrice’



- The regression plot shows that there is a positive relationship between the 'GarageArea' and 'SalePrice' columns. This means that as the garage area goes up, the sale price tends to go up, too. There is a good relationship between how big the garage is and how much the house sells for.
- The information given could help figure out the selling price or worth of the property that is being considered for buy. When planning and building new homes, developers can take into account what buyers want in a garage area.
- One thing to think about is how the garage affects the selling price of a home. Location, the state of the property, and any extra features also play a big role in figuring out the total price.

Conclusion

- Accurately predicting prices is a key part of making smart decisions in the ever-changing real estate market.
- Architectural Preferences: The research found that people like different types of architecture for their homes, which gives property producers important information.
- ExterQual, or how a house looks from the outside, is a big part of how much it's worth and how appealing it is.
- Positive associations were found between the size of the living area, the size of the closet, and the sale price, which suggests that these factors have a big effect on the value of a property.

Thank You