GROUP CODE: A099

STUDENT IDS: GITHUB USERNAMES:

GIT URL: https://github.com/7COM1079-A99/A99.git

DATASET NO: DS003

REASON FOR CHOICE OF DATASET: The reason for choosing this dataset was because we believed that this dataset could be analysed to explore the factors influencing property prices. Factors such as number of bedrooms and

bathrooms, location preferences, furnished status and so on. We believe analysing these will provide insight into which of

these features cause increased prices of property.

COLUMN HEADERS: WHAT DOES EACH COLUMN REPRESENT

- ✓ **price**:The price of the property (in a specific currency, dollar was assumed here).
- ✓ **area**: The total dimension area of the property in square feet or square meters.
- ✓ **bedrooms**: The number of bedrooms in the property.
- ✓ **bathrooms**: The number of bathrooms in the property.
- ✓ **stories**: The number of floors or stories in the property.
- ✓ **mainroad**: Whether the property is accessible from a main road (e.g., yes or no).
- ✓ **guestroom**: Whether the property has a guest room (e.g., yes or no).
- ✓ **basement**: Whether the property has a basement (e.g., yes or no).
- ✓ **hotwaterheating**: Indicates if the property has a hot water heating system (e.g., yes or no).
- ✓ **airconditioning**: Whether the property is equipped with air conditioning (e.g., yes or no).
- ✓ **parking**: The number of parking spaces available with the property.
- ✓ **prefarea**:Whether the property is in a preferred area (e.g., yes or no).
- ✓ **furnishingstatus**: The furnishing status of the property, it is categorical with values such as:
  1. furnished: Fully furnished.
  2. semi-furnished: Partially furnished.
  3. unfurnished: No furniture included.
  4.

OUR DEPENDENT VARIABLE IS: price OUR INDEPENDENT VARIABLE IS: area

DATA TYPE OF DEPENDENT VARIABLE: Interval

RESEARCH QUESTION: Is there a correlation between the price (in dollars) of a house and its dimension area (in

square feet)

NULL HYPOTHESIS: There is no correlation between the price and area of the house

ALTERNATIVE HYPOTHESIS: There is a correlation between the prices of the house and its areas in square feet.

IS IT A NORMAL DISTRIBUTION: non-normal distribution

DO YOU HAVE A HISTOGRAM: yes

THE REASON FOR CHOOSING THIS DATASET:

The reason for choosing this dataset was because we believed that this dataset could be analysed to explore the factors influencing property prices. Factors such as the number of bedrooms and bathrooms, location preferences, furnished status and so on. We believe analysing these will provide insight into which of these features cause increased prices of property

MAX OF 400 WORDS EXPLAINING WHAT THE DATA WILL SHOW AT THE END:

By the end of this project, we expect the dataset to uncover valuable insights into the factors that influence property prices and how they interact. Here's what we anticipate the data will reveal:

The analysis will likely confirm that larger properties and those with more amenities (such as additional bathrooms, multiple stories, or parking spaces) tend to have higher prices. Features like the property's furnishing status, its proximity to main roads, or being located in a preferred area may also play a significant role in driving up property values. Furthermore, premium features like air conditioning or hot water heating are expected to contribute to higher price tags.

In terms of relationships between features, we can expect to see moderate to strong positive correlations between the price and the area of the property. Larger homes are typically more expensive. Similarly, features such as the number of bathrooms or stories may also show a positive correlation with price, reflecting their impact on a property's overall value.

These findings will have practical implications for various stakeholders, such as buyers, investors, and real estate agents. Buyers will be able to identify the best value for their budget, while agents can use these insights to market properties more effectively.

GIT LOG CODE: git log --pretty=format:'"%h","%ad","%an","%s"' --date=iso-strict > git_log.xlsx

R SCRIPT