**Aim of project:**

1. Which location in Kuala Lumpur’s area has the lowest monthly rental?

2. Which location in Kuala Lumpur’s area has the lowest house price?

**Type of data:**

|  |  |  |
| --- | --- | --- |
| Variable | Data type | Description |
| property\_id | integer | Unique identifier for the properties, serving as primary key |
| transaction\_type | string - categorical | Indicates whether the property is listed for sale or rent |
| Property\_category | string - categorical | The general category to which the property belongs, such as "House" or "Apartment/Condominium." |
| property\_type | string - categorical | The specific type of property, providing more detailed information about its structure or style, e.g., "2-storey Terraced House" or "Condominium." |
| location\_city | string – categorical | The city where the property is situated, in this case, Kuala Lumpur. |
| Location\_subarea | string - categorical | A more specific area inside the city, for example, both Ampang and Cheras which are in KL city. |
| bedrooms | Integer | The number of bedrooms in the property |
| bathrooms | Integer | The number of bathrooms in the property |
| property\_size\_sqft | Integer | The size of the property in square feet |
| property\_price | float | The price at which the property is listed for sale. This column may be empty for properties listed for rental. |
| monthly\_rent | float | The monthly rental cost if the property is listed for rent. This column may be empty for properties listed for sale. |

**Method of data collection:**

To collect primary data, I was collecting data using Web scraping. Using python requests, I am able to scrape house price and monthly rental cost from mudah.my website. Please feel free to refer my code at Kaggle: <https://www.kaggle.com/code/tanyongsheng/learn-scrape-mudah-my>

The reason I choose primary data is because it’s more quality, reliable, and understanding. With getting up-to-date data of housing price and monthly rental cost, it will be easier for its users to take fast action (e.g., finding best affordable place to rent or buy).

Meanwhile, I choose python script to collect data is to save time to collect large rows of data (around 10k rows) compared to manual copy and paste. Also, we could run the script again in future whenever we need the latest data in future.