Aim of project: Which location in Kuala Lumpur's area has the lowest monthly rental?

Type of data:

Variable	Data type	Description
property_id	integer	Unique identifier for the properties
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
property_type	string - categorical	The specific type of property, providing more detailed information about its structure or style
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
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
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 this secondary 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 secondary data is because it's easier to collect the data without spending too much time or resources into it, as it is collected by others already.

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.

Project Timeline:

I spend around 4 hours to complete this scraping project on 04 September.

Contact Information:

- Project Lead: Tan Yong Sheng

- Phone: +6011-3920 0209