

## **PROJECT NAME: WEB SCRAPING REAL ESTATE DATA FROM NOBROKER.COM**

### **Project Description:**

In my project, I focus on extracting valuable real estate data from NoBroker.com, a disruptive platform revolutionizing property transaction by eliminating brokerage fees. Leveraging Python and essential libraries like pandas, BeautifulSoup, and Selenium, I've crafted a comprehensive solution to gather critical information for property seekers and analysts.

### **Key Data Extracted:**

- APRTMENT\_LOCATION: Location of the apartment.
- APRTMENT\_RENT: Rent amount for the apartment.
- APRTMENT\_DEPOSIT: Deposit required for renting the apartment.
- APRTMENT\_MAINTENANCE: Maintenance charges associated with the apartment.
- APARTMENT\_SQFT: Area of the apartment in square feet.
- RENT\_TYPE: Type of rental agreement.
- APARTMENT\_FURNISHING: Furnishing status of the apartment.
- APRTMENT\_TYPE: Type of apartment. (e.g., 1BHK, 2BHK, etc.)
- PREFERRED\_TENANTS: Preferred tenants specified for the apartment.
- APRTMENT\_AVAILABLE\_FROM: Availability date for the apartment.

### **Methodology:**

I begin by importing necessary libraries including pandas, BeautifulSoup, and Selenium. Utilizing Selenium, I navigate the dynamic interface of NoBroker.com to access property listings in Thane (India). To handle the infinite scroll feature of the website, I implemented a while loop that continuously scrolls down the webpage until reaching the end. With BeautifulSoup, I parse the HTML structure of the webpage to pinpoint and extract specific data elements. The find and find\_all functions in BeautifulSoup are utilized to locate all instances of specific HTML elements that contain the desired data. Core Python programming language is employed to execute data extraction tasks, utilizing features like loops, conditionals, and string manipulation to extract and process data efficiently. Data extraction is performed through targeted selection and iteration

over relevant elements, ensuring accuracy and completeness. Once the data is extracted and organized into a structured dataframe, I utilize the pandas library's `to_csv` function to convert the dataframe into a CSV (Comma Separated Values) file. This conversion process ensures portability and compatibility of the dataset across various platforms and software applications, facilitating further analysis and utilization.

**Dataset Information:**

The extracted dataset comprises approximately 2478 rows and 10 columns, providing a comprehensive repository of real estate information in the city of Thane.

**Potential Utilization:**

My project not only streamlines the process of gathering real estate data but also unlocks opportunities for in-depth analysis and insights. By leveraging further analysis techniques, users can uncover crucial insights such as identifying high-priced areas within Thane, understanding apartment owner expectations, and exploring the influence of location and apartment size on rental prices. Through meticulous examination and interpretation of the extracted data, users can derive actionable conclusions and make informed choices in their real estate endeavors.

My project empowers users with the tools to navigate and understand the dynamic real estate landscape, facilitating informed decision-making for tenants, landlords, and analysts alike.