**Submitted By:**

**(FA24-BBD-069)**

**Noor Ul Ain Zahid**

**Submitted To:**

**Prof. Ahmad Qammar**

**Complete Analysis of the Melbourne Real estate Data File**

**1. *Data Loading***

* The dataset **"melb\_data.csv"** is loaded using pandas.read\_csv().
* The first few rows are displayed to inspect the structure.

**2. *Handling Missing Values***

* The percentage of missing values in each column is calculated.
* Columns with more than **20% missing values** are identified and dropped.
* Numeric columns' missing values are **filled with the median**.
* Categorical columns' missing values are **filled with the most frequent value**.

**3. *Data Filtering***

* Extracts properties located in **Richmond** with a **price greater than 1,000,000**.
* Filters properties where **land size is above 500 square meters**.
* Checks if the "BuildingArea" column exists before selecting specific columns.

***4. Data Transformation***

* **Sorting**: The dataset is sorted by **price in descending order** to identify the most expensive houses.
* **New Column Creation**:
  + "Price\_per\_Room": Computed as **Price / Number of Rooms**.
  + "Date" column is converted to **datetime format**.
  + Extracts the **year of sale** into a new column "Year\_sold".
* **Categorization**:
  + Adds a new column "Category":
    - "Expensive" if the price is **above the median**.
    - "Affordable" otherwise.

***5. Data Aggregation***

* Computes the **average property price per suburb**.
* Counts the **number of properties sold per suburb**.
* Identifies the **most expensive suburb** based on average price.

***6. Data Export***

* The transformed dataset is likely saved or displayed, though export functionality was not explicitly found in the extracted code.

**Conclusion**

* The notebook effectively performs **data cleaning, transformation, filtering, and aggregation**.
* It categorizes properties based on price and extracts **valuable insights** about expensive and affordable properties.