# Melbourne\_Housing\_Snapshot

## **About Dataset:**

The dataset available at the given Kaggle link provides information about the Melbourne Housing Snapshot. It includes various attributes related to 'Rooms', 'Price', 'Distance', 'Postcode', 'Bedroom2','Bathroom',' Land size', 'Latitude', 'Longitude', 'Property count' and other relevant statistics.

## **Libraries:**

### NumPy:

The NumPy library, which stands for Numerical Python, is a fundamental package in Python for scientific computing. It provides powerful data structures, multidimensional array objects, and a collection of mathematical functions to efficiently perform numerical computations.

### Pandas:

Python has a robust data analysis and manipulation module called Pandas. For working with structured data, such as tabular data in the form of Data Frames, effectively, it offers data structures and functions. We read and loaded the dataset using Pandas, cleaned, filtered, and transformed the data, computed many summary statistics, and then explored the data.

### Sklearn:

Sklearn, often known as the scikit-learn library, is a well-known open-source machine-learning toolkit for Python. For numerous machine-learning tasks, including classification, regression, clustering, dimensionality reduction, model selection, and data preprocessing it offers a complete collection of tools and functionalities. Sklearn is compatible with the larger Python ecosystem because it is built on top of other well-known scientific computing libraries like NumPy, SciPy, and matplotlib. It provides a dependable and clear API, making it reasonably simple to use and comprehend, especially for those new to machine learning.

### Matplotlib:

A popular Python data visualization package is Matplotlib. It offers a versatile and extensive range of features for building different kinds of plots and charts. A variety of visualizations, such as bar graphs, histograms, scatter plots, box plots, heatmaps, and pie charts, were produced using Matplotlib. We were able to gather insights, spot patterns, and effectively explain the data's conclusions thanks to these visualizations.

### Seaborn:

Seaborn is a statistical data visualization library built on top of Matplotlib. It provides a higher-level interface for creating visually appealing and informative statistical graphics. We used Seaborn to enhance the visual aesthetics of our plots, apply color palettes, and create more advanced visualizations, such as scatter plot matrices and correlation heatmaps.

## **Overall Working:**

I mainly tracked and analyzed Melbourne city home prices, giving information on the number of rooms, bathrooms, and areas covered in relation to different locations. Additionally, I ran additional data sanity checks, looking for NA, NULLS, and missing values in the columns for "Building Area," "Year Built," and "Council Area," but I was unable to locate any in the dataset. Then I performed exploratory data analysis where I plotted histograms & Box plot on continuous variable (Housing price, Housing land size, bedrooms, bathrooms, Rooms) & finding the relationship between the columns in the data collection by applying correlation. After applying correlation, we are identifying the important columns for our analysis are (Rooms, Price, Distance, Bedroom2, bathroom, Longitude). Now apply scatter plot matrix on the identifying columns. After doing all of these analyses also find the average price of houses by region and property type. And lastly apply some machine learning models which are Linear regression, SVM & XGboost and evaluate the performance of the model by calculating mean squared error, root mean square error and R-square score.

## **Results**

A picture containing screenshot, design

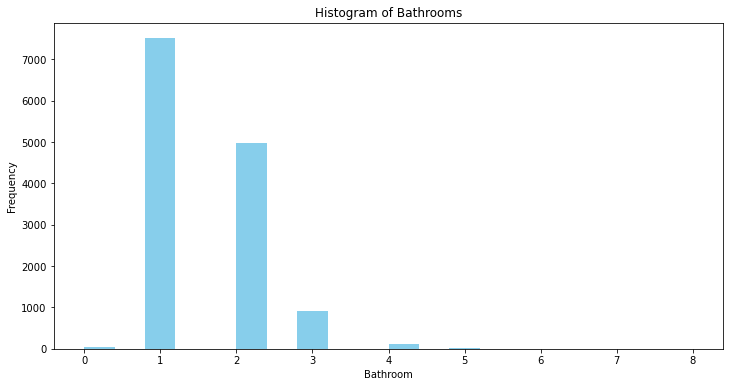
Description automatically generated

By analyzing the histogram of the housing prices, we can visually understand the distribution and range of prices in the Melbourne market, which is crucial for assessing affordability and market trends.

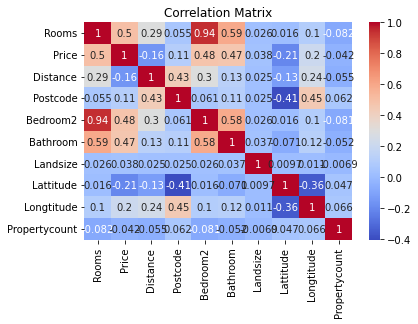
A picture containing screenshot, rectangle, design

Description automatically generated

The Melbourne Housing Snapshot dataset also includes information about the number of bedrooms in residential properties. Analyzing the histogram of the number of bedrooms can provide valuable insights into the distribution and frequency of different bedroom configurations in the Melbourne housing market. This information can help in understanding the preferences and needs of homebuyers and renters and can also provide insights into the housing supply and demand dynamics in different bedroom categories.



The Melbourne Housing Snapshot dataset contains data on the number of bathrooms in residential properties. Analyzing the histogram of bathroom counts allows us to visually examine the distribution and frequency of different bathroom configurations in the Melbourne housing market. This analysis provides insights into the prevalent bathroom standards and preferences among homebuyers and renters. It can also be indicative of property value and desirability, as the number of bathrooms is often a significant factor in determining the convenience and functionality of a property.



After applying correlation analysis on the Melbourne Housing Snapshot dataset, several important columns have been identified for further analysis. Rooms, Price, Distance, Bedroom2, bathroom, Longitude these are the important columns for the data set because these columns values are near to the 1 which shows that these columns are correlated with each other.

A picture containing colorfulness, screenshot, rectangle, graphics

Description automatically generated

The graph showing the average house price by region provides insights into the variation in housing prices across different regions in Melbourne. Each region is represented on the x-axis, while the average house price for that region is shown on the y-axis.

By examining this graph, we can observe the relative affordability or premium associated with different regions. Regions with higher average house prices indicate areas where properties tend to be more expensive, while regions with lower average house prices suggest relatively more affordable housing options.

A picture containing screenshot, rectangle, colorfulness, square

Description automatically generated

The graph depicting the average house price by property type provides valuable insights into the variation in housing prices based on different property types in the Melbourne Housing Snapshot dataset.

Each property type, such as houses, townhouses, or units, is represented on the x-axis, while the average house price for each property type is shown on the y-axis.

By analyzing this graph, we can observe the relative price levels associated with different property types. Property types with higher average house prices indicate that they are generally more expensive, while property types with lower average house prices suggest relatively more affordable options.