

**MELBOURNE HOUSING MARKET DATASET ANALYSIS**

|  |  |  |
| --- | --- | --- |
| **NAME** | **ID** | **COURSE** |
| **HASHIR AHMED** | **26454939** | **DATA MINING AND VISUALIZATION** |

# **Task 4: Dataset interpretation**

Based on the interactive plots, this dataset reveals key trends in property market using pricing and distribution.

* Properties that are closer to the CBD are generally command prices higher, although these local variations exists within suburbs.
* Houses ‘h’ that tend to be more expensive than units ‘u’ or the townhouses ‘t’, reflecting differences in size and desirability.
* Regional disparities are evident, with certain regions exhibiting higher property values and price volatility.
* Outliers in this dataset represent premium properties, which are scattered throughout the dataset.
* Price distribution is skewed, indicating a higher frequency of lower-priced properties compared to luxury homes.
* And the interaction visualization enhance this analysis by enabling detailed examination of individual data points and localized trends.
* In 2016 Melbourne housing market shows a wide range of property prices influenced by CBD, property type and region.

# **Task 5: Learning Experience**

This interactive visualization exercise has significantly improved my understanding of Exploratory Data Analysis (EDA), with its superpower of creating high static plots using matplotlib and seaborn, providing a foundation for understanding basic visualization principles. However, the transition provided by interactive plots of Plotly revealed a new dimension of data interaction with its zooming and hovering feature. This ability to zoom, hover, and filter data points within the plots transformed the way I could explore the Melbourne Housing Dataset. With the help of this, I could quickly identify outliers, compare distribution across different regions, and examine the impact of features like property type and distance to the CBD on housing prices. These interactive box plots were invaluable for comparing price distributions across regions, uncovering insights that would have been difficult from static plots alone. The histogram allowed for a deeper understanding of price distribution, with a rug plot that provided additional context on data density.

Furthermore, the way of choosing appropriate interactive features, such as hover data and color coding, forced me to think critically about the message I wanted to convey, and this is the best representation of data. I learned that effective interactive visualizations are not just about making pretty pictures, but they are about empowering the user to explore data and discover insights for themselves. This hands-on experience has equipped me with valuable skills in data visualization, making me more confident in my ability to analyze and interpret complex datasets.

From the dataset visualizations, we observe key insights about property prices. The scatter plot shows a negative correlation between distance and price properties closer to the city tend to be more expensive. The static and interactive box plots indicate that houses ('h') have a wider price distribution, with more outliers, compared to units ('u') and townhouses ('t'). Regional variations also affect property prices, as seen in the interactive box plot. Some metropolitan areas exhibit significantly higher property values. These insights suggest that location and property type play crucial roles in determining real estate prices, with urban areas demanding a premium. And the interaction visualization enhance this analysis by enabling detailed examination of individual data points and localized trends.