Self-Reflection on **Apartment for Rent Classified** Dataset Project

This self-reflection outlines the insights, skills, challenges, and new learnings from my work on the Apartment for Rent Classified dataset. Throughout the project, I gained hands-on experience in data analysis, clustering techniques, and problem-solving. The project helped me refine my data science skills, but it also highlighted the complexities that arise when working with large, unstructured datasets.

# Insights

This project provided valuable insights into the relationships between apartment features and rental prices. The exploratory analysis revealed significant patterns, such as the correlation between price and apartment size, and how rental prices vary by region. The clustering models helped uncover hidden structures within the data, allowing me to group apartments based on similarities in their features.

# Skills

The project allowed me to develop a wide range of skills. Key skills I strengthened include data preprocessing, exploratory data analysis (EDA), and clustering algorithms like K-Means and Hierarchical Clustering. Additionally, I gained hands-on experience in feature encoding, handling missing values, and scaling data for machine learning. I also gained familiarity with Scipy's Pdist method, which was helpful when working with non-squared data for Hierarchical Clustering.

# Challenges

One of the main challenges I faced during this project was working with hierarchical clustering on a large, non-squared dataset. Initially, the model was not running, and I was unsure how to proceed. After consulting with my professor, I was introduced to the Pdist method from the Scipy package, which allowed me to handle non-squared datasets. Despite this, I still faced performance issues, and the session would frequently crash. In response, I reduced the dataset to the first 1000 rows to make the model more manageable. Moreover, I encountered challenges when visualizing the dendrograms, so my professor suggested using the truncate feature for clearer visualization. I also experimented with different p-levels to compare the dendrograms.

# New Learnings

Throughout this project, I learned valuable lessons about managing large datasets and applying clustering techniques effectively. I gained a deeper understanding of hierarchical clustering, especially how to handle non-squared data using Pdist. I also learned how to troubleshoot performance issues, optimize models, and visualize clustering results in a more meaningful way. These experiences have expanded my skill set and prepared me for more complex data analysis tasks in the future.