

[LINK TO THE NOTEBOOK](#)

? **Word Cloud of Host Names:**

- **User Response:** The user concatenates all host names into a single string to feed into a word cloud generator. A custom stopwords list is created to remove non-persons such as company names and generic terms.
- **Why:** The user's approach ensures that the word cloud focuses on real hosts by cleaning the data beforehand. This improves the accuracy of the visualization by eliminating irrelevant words that could skew the results.

? **Geospatial Analysis of Listings:**

- **User Response:** Longitude and latitude values are extracted and plotted using a scatter plot to visualize listing locations. Later, an interactive map is generated using folium to allow zooming and exploration of geographical distribution.
- **Why:** The initial scatter plot gives a quick overview of how listings are distributed, while the folium map allows for more detailed inspection, which is valuable for spatial analysis.

? **Neighborhood Pricing Boxplots:**

- **User Response:** Boxplots are created to show price distributions across neighborhoods. The user merges neighborhoods with similar names to ensure consistency and reduces unnecessary variation in the x-axis labels.
- **Why:** Cleaning and merging neighborhood names ensures that the boxplots present clearer comparisons between areas, preventing misleading results caused by inconsistent naming.

? **Descriptive Analysis of Neighborhoods:**

- **User Response:** The user normalizes the neighborhood names by converting them to lowercase and stripping whitespace, ensuring consistent grouping. They then calculate descriptive statistics, including mean prices, and include a breakdown of room types.
- **Why:** Consistent neighborhood names avoid errors during aggregation and ensure that descriptive statistics are accurate. This helps to highlight trends such as which neighborhoods or room types are more common or expensive.

? **Regression Modeling to Predict Prices:**

- **User Response:** The user builds three regression models using features like neighborhood, room type, and number of bedrooms. They split the dataset into training and test sets and then fine-tune model parameters. Additional, seemingly irrelevant, features (e.g., "number of reviews") are tested to see how they affect model performance.
- **Why:** The user experiments with different models and features to identify which combination yields the highest accuracy. Testing irrelevant features shows the impact of feature selection on model performance, helping to refine the final model.

? **Model Evaluation:**

- **User Response:** The user evaluates the models, adjusting hyperparameters like the number of neighbors and distance metrics in k-nearest neighbors (KNN). They test several configurations and document how different models perform, noting when certain features degrade accuracy.
- **Why:** Tuning the model parameters and including/excluding features help optimize the accuracy. Documenting these changes allows the user to understand which variables most influence the predictions, leading to better model selection.