



Programming in Python

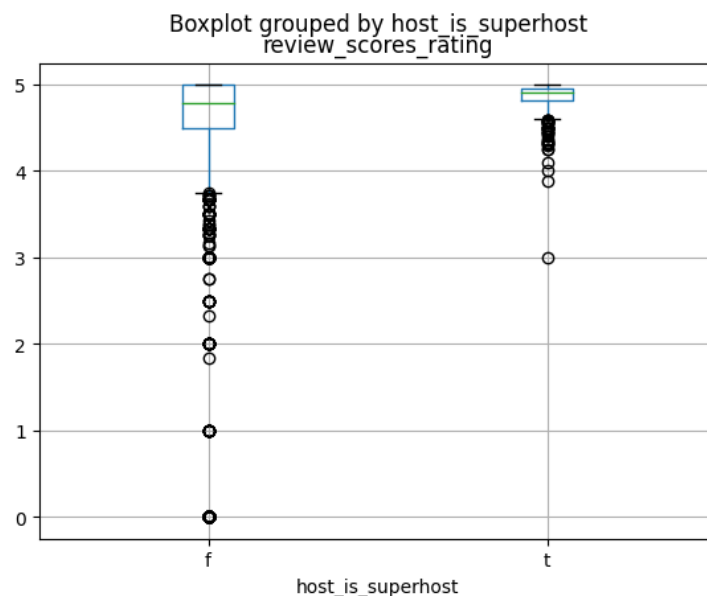
Final Project

`airbnb_listings.csv` is a dataset taken from Airbnb describing **7566** listings with a variety of quantitative and qualitative variables (**75** columns in total). Variables include listing amenities, review ratings, neighbourhoods, availability etc.

Part 1)

Highlight **5 trends** in the dataset using **data visualisation** techniques with Matplotlib, Pandas plotting or Seaborn.

One example might include the relationship between the variables `review_scores_rating` and `is_superhost`. We can see from this boxplot that listings whose host is a superhost generally have a higher rating score.



Comment on the trend and provide some statistical analysis (mean/mode/median, Interquartile ranges, standard deviation etc.)

Marks are awarded for:

- Appropriate labelling and styling of each plot.
- Clarity and comprehension of comment and statistical analysis.
- Complexity and depth of analysis. Some trends are less obvious and therefore more interesting. Perhaps, you could use a combination of variables or make new variables to highlight a trend. Furthermore, some columns require **data cleaning** before plotting and analysis

Part 2)

Create a **command-line interface** that enables a user to perform analysis and visualise data by entering queries through the terminal.

For example, a user could ask:

- Give me the top 10 rated Airbnb's where I can stay at least 8 nights and is instantly bookable
- Plot the host response rate against the host response time in a scatter plot
- Show me the amount of Airbnb's in Dublin, Ireland which have a cleanliness rating higher than 4.0
- Provide the cleanliness rating scores for listings 12-15

Marks are awarded for:

- **Creativity and intuition.** There are no strict rules for the command-line interface, the main aim is to provide the user with as much freedom as possible to explore the dataset
- **Complexity and depth.** Functions that are technically more difficult and provide the user to perform a deeper analysis will receive higher marks
- **Modularisation.** Appropriately split your code into functions and files so that each feature of your program is modular.
- **Maintainability and Readability.** Appropriately name your variables and functions. Leave comments where needed. Avoid overengineering by keeping your code as simple as it can be.
- **Error-handling.** Ensure your program doesn't crash when invalid input is given, or plotting/analysis fails with certain parameters.
- **Ease-of-use.** The command-line interface should be easy to use with appropriate input prompting. Imagine you are giving this program to someone who can't see or edit the code and can only control the program through the terminal.

When writing your program, you may edit the dataset in whatever way you please (remove columns, add columns, clean columns, change values etc.)

Upload your code to **GitHub** and provide the link to the repository in your submission.

Submit a **zip folder** containing your files on Moodle to complete the project. You can use either Jupyter Notebook or scripting mode (.py files) for Part 1 but you must use scripting mode for part 2. Cooperation for this assignment is permissible but copying is not tolerated. You must understand your code upon submission.