Data Scientist Intern Position Home Task

Part 1 – Data Gathering & Analysis

Context:

Many European cities are facing housing affordability issues. In this exercise, we want you to explore the relationship between short-term rentals (e.g., Airbnb) and rental market prices in two cities of your choice (you can choose any two cities but they need to have good housing data).

1. You will have to download a housing dataset coming from a public source (such as https://data.london.gov.uk/)
2. Access a short-term rental price dataset either with a platform like https://insideairbnb.com/ or by scrapping a short-term rental website yourself (using BeautifulSoup or similar package)
3. Explore if there is any observable relationship between the concentration of Airbnb listings and rising rental prices in different neighbourhoods
4. Prepare a short summary of your findings including 2-3 data visualisation
5. Bonus: Try mapping the data using folium, geopandas, or another geographic visualisation library.

Part 2 – Analysis

6. Using the combined datasets, can you uncover any interesting trends, correlation or insights (use basic ML or clustering if relevant)

Part 3 – Open Ended Challenge

7. Assuming one of the two cities you picked is experiencing a higher increase in rental prices than the other. How would you try to explain that incremental difference? *(just explain how you would solve that question)*
   a. What data would you use to explain that difference?
   b. How would you structure your investigation and analysis?

Submission Guidelines

- Please submit your work as a Jupyter Notebook with comments and explanations.
- Make sure the notebook runs from top to bottom.
- We will be evaluating data handling, insight generation and clarity.
- Optional: include a short 2–3 slide (max) summary to present your findings. Don't be too technical. We are more interested in your ability to present your work.