Trusted Housesitters

Principal Data Scientist Task

Problem Definition:

For this task, we are providing you with a dataset* with anonymised and realistic records for homeowners, housesitters, pets, listings. And attributes such as location, pet preferences, availability, experience, ratings, and interaction history.

The attached dataset contains individual CSV files representing portions of our internal tables, which you can use as a basis for your clustering work. Documentation describing each table and its fields is also included for your reference.

We would like you to propose and execute a clustering strategy to:

- Improve matching (e.g., recommending the best housesitters to homeowners or vice versa).
- Identify potential underserved segments (e.g., housesitters for exotic pets or in rural areas).
- Understand user personas (e.g., casual vs. professional housesitters).

In your submitted work, we would like you to demonstrate ability in data processing, clustering methodology, data interpretation through analysis and visualisation, and finally suggesting actionable insights.

Aspects of these would include:

- Handling missing or noisy data
- Engineering meaningful features from complex data

- Experimenting with a clustering algorithm of your choice and justifying your selection
- Presenting clear visualisations and a narrative explaining your clusters
- Presenting how clustering results would be interpreted in actionable next steps

Evaluation Criteria

Data Preprocessing, clustering Methodology, interpretation Actionable Insights and Stretch Goals.

Expectation

For this task, please submit the following:

- Your code, saved in a private GitHub repository (please share the link and grant us access).
- Your slides/presentation of your clustering findings.
- We recommend 2-4 hours to complete the task.

Links to Dataset*:

- https://drive.google.com/file/d/1qQkDWIpa47tOVlQf6VHcuhTkx4RaYhfA/view?usp=sharing
- https://drive.google.com/file/d/1yHuV53uh8W_jWqT98PL8D2bxRZ4hCBJA/view?usp=drive_link