

# Data Science Task

## Exercise:

The dataset `airbnb_dataset.csv` presents data about some US airbnb properties. Work with this dataset to answer the following questions:

1. How many features are represented in the dataset? In terms of statistics, what type of variables are "host\_since", "host\_is\_superhost", "room\_type", "bedrooms" and "price"?
2. What are the two most correlated variables?
3. What are the two least correlated variables?
4. Create a plot for the correlation matrix. According to this matrix, what are the most relevant attributes? How can you use these findings for feature selection?
5. How does the price distribution vary by state? And by room type? Create a few slides summarizing your findings about this.

## Take into account:

- Use python or R.
- Submit the code used to solve all the questions.
- Make sure that your code is readable and reproducible.
- Make sure that you name the deliverables and add comments to the code to identify who the files belong to

## Expected deliverables:

- Code file
- Slides