Congratulations on making it to our take-home assignment round.

Within this zip folder, there is a file named ‘dataset’. This is a garbled dataset created by our team to enhance the process of hiring. Data in the file doesn’t correspond to any actual places and is randomly generated. Please follow the instructions and submit your response by the time given in the mail.

1. Read the file in Python
2. It has data of certain imaginary hotels
3. Perform EDA using visual analysis and mention any interesting insights in your code file. Think about what data could be interesting for Travee team and find and explain any insights around them. Lean on graphical representation if you think they are necessary to explain your point of view.
4. Test the file for inconsistencies and mention the same. Also mention your thoughts on how will you handle the same in a pipeline.
5. Build the following models on this data. Focus on model components and not on performance metrics. You will not be scored for performance metrics. Using ChatGPT will get you immediately **disqualified** from the hiring process.
   1. Build a model to predict the star rating of a hotel given the reviews related to that hotel
   2. Build a model to extract key features of the hotel from their reviews. Put these key features in a comma separated file and send with your submission.
6. For each model, you need to send separate
   1. Properly commented Ipython notebook after all code has been run
   2. HTML download of the same after all your code has run
   3. The model in pickle file which can be reloaded
7. Include a small description about possible edge cases and how you will handle them in the pipeline code.