PREDICTION CHALLENGE

Due date: March 26th, 2019 (10 PM) – *any delay will be penalized.*

Hello dear students,

As mentioned during the 8th session class, you'll have a prediction challenge to realize as a final exercise for the data camp.

The purpose of the exercise is to use what you learnt in NLP in order to predict the rating of a review. To do this, we scrapped data from Novotel resorts all over the world, which as a hotel chain is quite related to Center Parcs. You will have to make a prediction for each review (one review = one rating) of the evaluation dataset. You are free to use any data modelisation on the training dataset.

If you have any question, do not hesitate to contact us by slack or by mail,

Thank you and have fun,

The Capgemini Team.

About data

In attachment of this pdf file you will find three files:
- A training dataset: with the following columns: hotel_name, trip_date, title, review, & rating
- An evaluation dataset: with the same columns excepted the rating one that you will have to predict
- A prediction dataset: example of what your submission file must look like

About score

We will evaluate your predictions using a F1-Score metric.

How to get a better score? When you think you have good predictions, you can send us your predictions. We'll then compute your score and tell you the result.
You can send us your results at the following mail addresses:
- Ariel.ibaba@capgemini.com
- Thibaud.lamothe@capgemini.com
- Thomas.clavier@capgemini.com

About results

At the end of the challenge, we'll ask you to send us your code. Both results and code will be taken into account in the final mark of the exercise.

NB: this is a sort of competition between the 12 groups, and the 3 winners will earn a price.