



## Challenge for Data Scientist

Thanks for your interest in Natural Cycles! We have prepared a data science problem to test your abilities on our own real data.

The .csv file we are sharing with you contains data from a set of women who have been using **NC° Plan Pregnancy** to try to conceive. It includes the duration of the attempt to conceive (observation period, or exposure) expressed as a number of menstrual cycles, and the outcome of the attempt, i.e. whether the woman became pregnant or not. An unsuccessful attempt to conceive ends with the woman resuming contraception, or interrupting the usage of NC.

The dataset contains the following features:

- age when starting the attempt to conceive
- BMI
- country of residence
- previous pregnancies (answer to the question: “*Have you been pregnant before?*”)
- education
- sleeping pattern (answer to the question: “*What is your usual sleeping pattern? You wake up...*”)
- dedication towards using the NC app (fraction of days with logged temperature data)
- average length of the cycle in days
- standard deviation of the cycle length in days
- cycle regularity
- frequency of intercourse (fraction of days with logged sexual intercourse)
- number of cycles in the attempt to conceive (*n\_cycles\_trying*)
- outcome of the attempt to conceive: *pregnant* or *not\_pregnant* after the observed number of cycles *n\_cycle\_trying*

**We would like you to explore the dataset and provide answers to the following questions:**

1. What is the chance of getting pregnant within 13 cycles?
2. How long does it usually take to get pregnant?
3. What factors impact the time it takes to get pregnant?

Please prepare a short report (choose the format that you prefer) and send it back within **one week**. We are most interested in how you think, so please carefully explain the reasoning and any assumptions you make in approaching the problem.

Please also share with us your code, either sending us the files directly or the link to your github repository (make sure you don’t upload the dataset to the repository).

Your code and report will only be used for assessing how well suited you are for this role.

We hope you enjoy this exercise! We look forward to reading your report!