The use of the datasets named “User.csv, Period.csv, Symptom.csv” is this folder are for the purpose of participating in the School of AI Global Health Hackathon 2019 only, and should not be used for commercial purposes or shared to other parties.

**Regarding the data:**

1. The data is taken from a subset of real users who were using an iOS app to record their feminine health data, for real (please pardon us if the data set is not entirely clean).
2. **User.csv**
   1. Personally identifiable information has already been removed
   2. The column named “dob” refers to a user’s birthday, and empty values mean that the user did not provide their birthday to the app
   3. The column named “cycle\_length\_initial” refers to the value (in days) that a user provides to the app during their onboarding to indicate how many days they think their menstrual cycle is
   4. The column named “period\_length\_initial” refers to the value (in days) that a user provides to the app during their onboarding to indicate how many days they think their periods last for
   5. The column named “id” is the User ID and can be used as a foreign key to link a given user to data rows with the same “user\_id” in Period.csv and Symptom.csv data tables
3. **Symptom.csv**
   1. Each symptom has a default & minimum value of 0, and a maximum value of 100, with 100 being the most severe
4. **Period.csv**
   1. The column named “start\_date” refers to the first day of bleeding of a period
   2. The column named “end\_date” refers to the last day of bleeding of a period
5. **Period trivia**
   1. Example of “What is period length?”
      * 1st Jan 2019: 1st day of bleeding (a.k.a period start date)
      * 5th Jan 2019: Last day of bleeding (a.k.a period end date)
      * Period Length = 5 days
   2. Example of “What is cycle length?”
      * 1st Jan 2019: 1st day of bleeding (a.k.a period start date)
      * 1st Feb 2019: 1st day of bleeding (a.k.a next period start date)
      * Cycle Length = 31 days