Data Description for Ancestry Data

Let's say that a ficticous deceased man, Joseph Wallace, has childen who have sought the help of an experienced genealogist to help with the family history. The genealogist then recorded the names of their ancestors from the legitimate sources they gave him, and he then created the csv file 'ancestors-raw-data.csv', to give to a data analyst. The genealaogist says that each row listed in the table describes one person, each related to Joseph Wallace. However, the genealogist admits there are errors in the file, since he had the help of a novice genealogist for data entry, so it may need some minor edits. He also asked that our analysis only include ancestors who were born after the year 1600 and died before the year 2000.

Note: While the ancestor names are ficticious, the cities, states and countries listed are all real places, and the amount of time between births marriages and deaths are believable.

Column Descriptions: Each of the columnns is described below.

- person_id: A unique 5 character id (with dash) for each person (ancestor) listed. (Example: 'A-052')
- first_name: The person's first name, with the first letter capitalized
- last_name: The person's last name, with the first letter capitalized. In this column we only list a woman's maidenn name, not her married name. For example, Charlotte Reed's last name would be "Church," because Charlotte Church is married, her name is now Charlotte Maria Reed. But in this column we only want a woman's maidenn name, not her married name.
- sex: The person's gender. Either "M" or "F". "M" stands for male. "F" stands for female.
- living: Values should be either 'Living', or 'Deceased'. Tells whether the person is living or deceased.
- birth_date: The person's date of birth, in the YYYY-MM-DD format. For example, October 16, 1875 would be "1875-10-16".
- birth_place: City, State, Country. For example, Shreveport, Louisiana, United States of America
- mother_id: The id of the person's mother. Ids that are not null have rows present in the file.
- father_id: The id of the person's father. Again, if an id is listed for father, the father's information is listed in another row.
- spouse_id: The id of the person's spouse. If the id of the person's spouse is listed, then the spouse's information is listed in another row of data.
- marriage_date: The date the person and their spouse were married. Again, the format used was YYYY-MM-DD.
- marraige_place: The place where the person and spouse were were married.
- death_date: The date of the person's death. gain, use the format YYYY-MM-DD.
- death_place: The place the person died. If the death place is not given, list the burial place. Use the format "City, State, Country".