1. What kind of cleaning steps did you perform?

   After importing the data as a pandas DataFrame, I found all the missing values are represented by '?'. So I converted all the values with type of object to numeric values, so the missing values are represented by NA. Then I replaced NA with median or mode of that columns if needed.

2. How did you deal with missing values, if any?

   For the columns with numeric data, I filled missing values with median of the columns. The columns with numeric data are as follows,

   ['Number of sexual partners', 'First sexual intercourse', 'Num of pregnancies', 'Smokes (years)', 'Smokes (packs/year)', 'Hormonal Contraceptives (years)', 'IUD (years)', 'STDs (number)']

   For the columns with categorical data, I filled missing values with the mode of the columns. The columns with categorical data are as follows,

   cata_cols = ['Smokes', 'Hormonal Contraceptives', 'IUD', 'STD', 'STDs:condylomatosis', 'STDs:cervical condylomatosis', 'STDs:vaginal condylomatosis', 'STDs:vulvo-perineal condylomatosis', 'STDs:syphilis', 'STDs:pelvic inflammatory disease', 'STDs:genital herpes', 'STDs:molluscum contagiosum', 'STDs:AIDS', 'STDs:HIV', 'STDs:Hepatitis B', 'STDs:HPV']

3. Were there outliers, and how did you handle them?

   There were some outliers, but I would just leave them there for further analysis.