1. What kind of cleaning steps did you perform?

Step 1: Generate a pandas dataframe that includes the image directory paths and the type of cervix image for the training dataset

Step 2: Visualize some sample images of each cervix type of the training dataset to get a sense how different the types are

Step 3: Create a new dataframe that include the following columns for the training dataset:

+ Image directory
+ Cervix type
+ Image height
+ Image width
+ Image channels

Step 4: Write the dataframe generated in step 3 to a csv file (training_images_size.csv)

Step 5: Create a dataframe that includes the following columns for the testing dataset:

+ Image directory
+ Image height
+ Image width
+ Image channels

Step 5: Convert the generated dataframe in step 5 to a csv file for later use (testing_images_size.csv)

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

There are no missing values since the training and the testing dataset for this project only involve images.

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

The outliers will be explored in the exploratory data analysis portion of this project.