Basic Descriptive Analysis

1. **Mean**: 0.505232
   * The mean represents the average of all values. Here, it indicates that roughly 50.5% of the observations are male (value 1). This suggests a fairly even split between male and non-male categories.
2. **Median**: 1.0
   * The median is the middle value when the data is sorted. A median of 1.0 means that more than half of the observations are 1 (male), indicating a slight skew towards males in this dataset.
3. **Mode**: 1
   * The mode is the most frequently occurring value, which is 1. Therefore, there are more male observations than non-male ones.
4. **Standard Deviation**: 0.50016
   * This measures the average spread of values from the mean. Given that this is a binary variable, the standard deviation being close to 0.5 indicates a roughly even distribution between the two categories.
5. **Variance**: 0.25016
   * Variance is the square of the standard deviation. For a binary variable, the variance is maximized when there is an even split between the categories (as is approximately the case here).
6. **Minimum**: 0
   * The minimum value is 0, indicating the presence of non-male observations.
7. **Maximum**: 1
   * The maximum value is 1, indicating the presence of male observations.
8. **Range**: 1
   * The range is the difference between the maximum and minimum values, which is 1 here, as expected for a binary variable.
9. **Percentiles (25th, 50th, 75th)**:
   * **25th Percentile**: 0.0 - This means that at least 25% of observations are non-male.
   * **50th Percentile (Median)**: 1.0 - Indicates that at least half of the observations are male.
   * **75th Percentile**: 1.0 - This indicates that 75% of the values are either 1 (male) or 0, with most being 1.
10. **Scipy stats.describe**:
    * **Number of Observations (nobs)**: 1338 - There are 1338 total observations in the dataset.
    * **Min/Max (minmax)**: (0, 1) - The minimum value is 0 and the maximum value is 1.
    * **Mean**: 0.50523169 - Consistent with the earlier calculated mean.
    * **Variance**: 0.25015959 - Consistent with the earlier calculated variance.
    * **Skewness**: -0.0209279 - The skewness is very close to 0, indicating that the data is nearly symmetric.
    * **Kurtosis**: -1.99956202 - A negative kurtosis indicates that the distribution is flatter than a normal distribution, meaning it has lighter tails.

**Insights**:

* The dataset is nearly evenly split between males and non-males, with males making up slightly more than half.
* The metrics indicate a fairly balanced distribution, with no significant skewness.
* The negative kurtosis suggests fewer extreme values compared to a normal distribution, consistent with a binary variable where most values are either 0 or 1.

PLOTTING