Basic Descriptive Statistics

**1.Mean**: 0.204783

* + The mean indicates that roughly 20.5% of the observations are smokers. This means that about 79.5% of the individuals are non-smokers, suggesting that smokers form a minority in this dataset.

1. **Median**: 0.0
   * The median value is 0, indicating that more than half of the observations are non-smokers.
2. **Mode**: 0
   * The mode is the most frequently occurring value, which is 0. This confirms that non-smokers are more prevalent in the dataset compared to smokers.
3. **Standard Deviation**: 0.403694
   * The standard deviation measures the variability of the data from the mean. Since this is a binary variable, a standard deviation of about 0.4 suggests some level of imbalance, with the majority being non-smokers.
4. **Variance**: 0.162969
   * Variance is the square of the standard deviation. The relatively low variance indicates that most values are close to the mean, meaning most individuals are non-smokers.
5. **Minimum**: 0
   * The minimum value is 0, indicating the presence of non-smokers.
6. **Maximum**: 1
   * The maximum value is 1, indicating the presence of smokers.
7. **Range**: 1
   * The range is 1, as expected for a binary variable (difference between 1 and 0).
8. **Percentiles (25th, 50th, 75th)**:
   * **25th Percentile**: 0.0 - At least 25% of the observations are non-smokers.
   * **50th Percentile (Median)**: 0.0 - At least half of the observations are non-smokers.
   * **75th Percentile**: 0.0 - Up to 75% of the observations are non-smokers, highlighting that smokers are in the minority.
9. **Scipy stats.describe**:
   * **Number of Observations (nobs)**: 1338 - The dataset contains 1338 observations.
   * **Min/Max (minmax)**: (0, 1) - The minimum and maximum values are 0 and 1, respectively.
   * **Mean**: 0.20478326 - Consistent with the earlier mean, showing that about 20.5% of the individuals are smokers.
   * **Variance**: 0.16296888 - Matches the variance calculated earlier.
   * **Skewness**: 1.46312353 - The positive skewness indicates that the distribution is right-skewed, meaning that there are relatively fewer smokers compared to non-smokers.
   * **Kurtosis**: 0.14073048 - A positive kurtosis close to 0 suggests the distribution is approximately normal in terms of tail thickness, but slightly peaked.

**Insights**:

* Only about 20.5% of the individuals in this dataset are smokers, indicating that smoking is relatively uncommon.
* The positive skewness suggests that most people do not smoke, and the minority that do creates a right-skewed distribution.
* The low variance further confirms that the data is not evenly distributed between smokers and non-smokers, with the majority being non-smokers.
* The percentiles and median values show that a significant proportion (at least 75%) of the individuals are non-smokers.