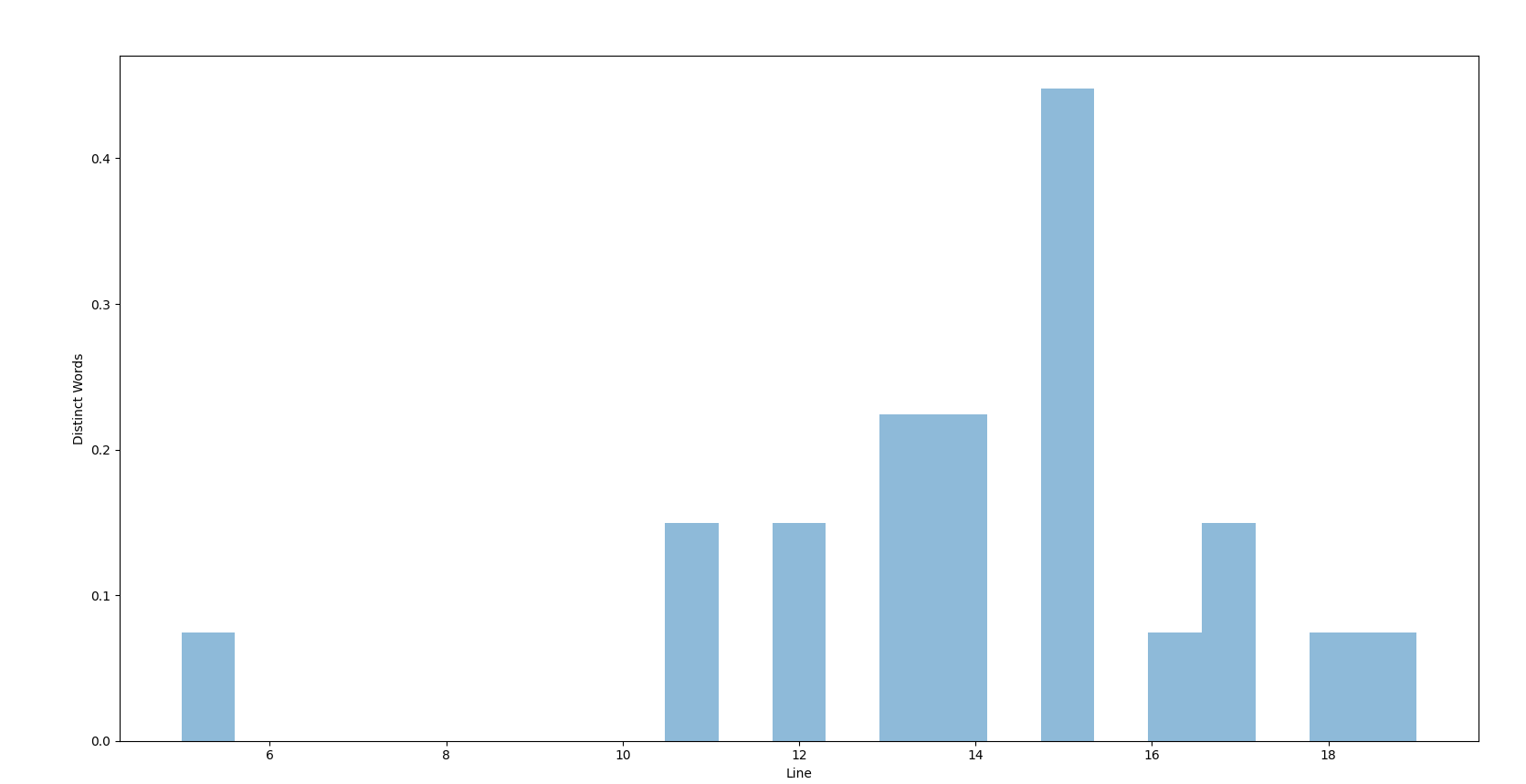
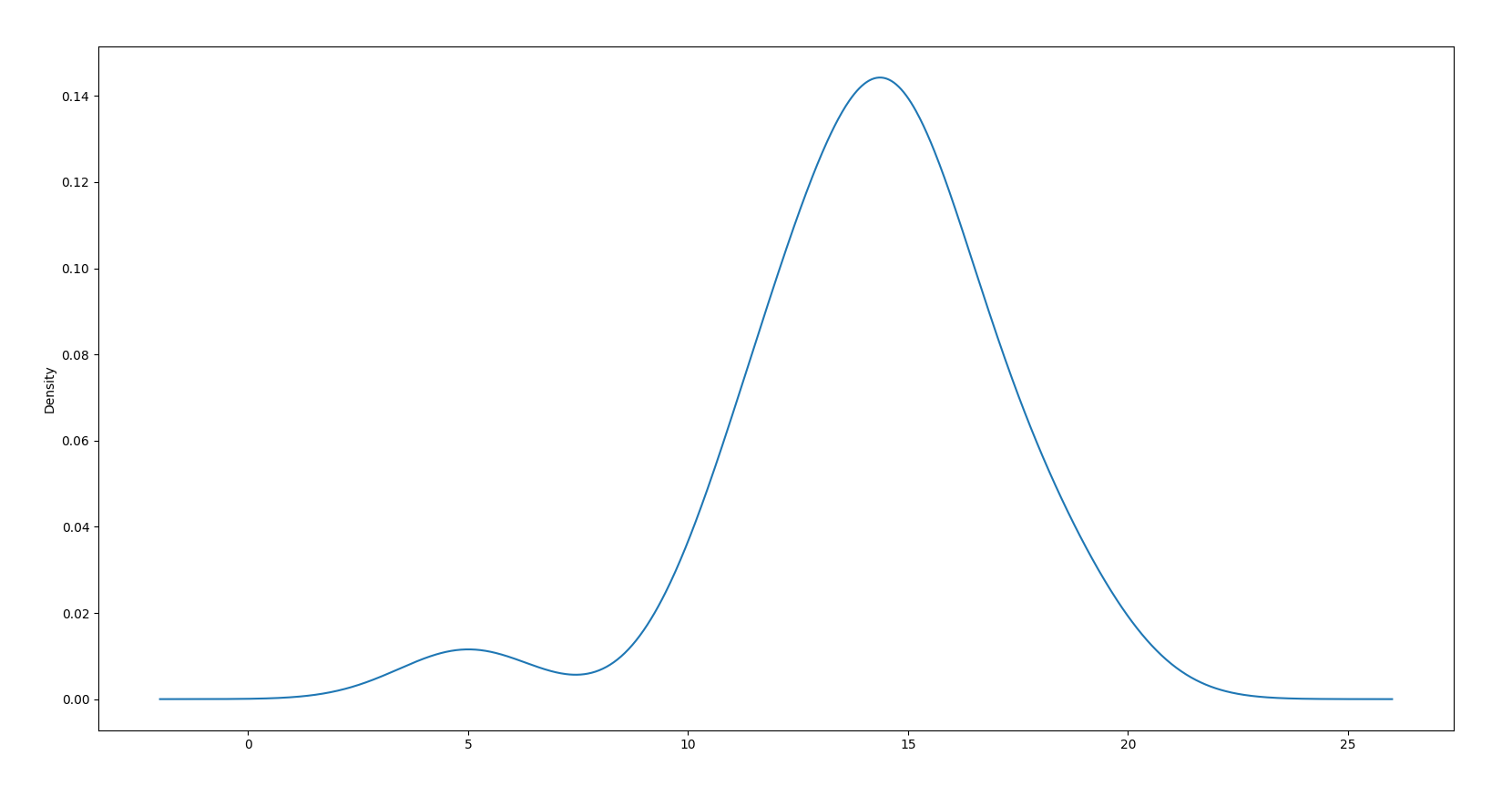
3.

a) The probability of the word “data” occurring in each line is 0.818

b)





From the histogram and the KDE plot, we can see that the distribution is largely normal.

c) The probability of the word “analytics” occurring after the word “data” is the probability of the word “analytics” given the word “data” is **0.333**. Note that this includes “analytics” appearing on the **next line**, since the question didn’t specify that it had to appear within the same line.