1. What is the probability of the word “data” occurring in each line ?

Answer : P(“data” in line) = 0.6364

1. What is the distribution of distinct word counts across all the lines ?

Chart, histogram

Description automatically generated

1. What is the probability of the word “analytics” occurring after the word “data” ?

Answer: P(“data analytics” | “data” ) = 0.3333