Non Llm String Similarity Statistics by Model