Артем, привет  
Выложил ноутбук с кратким анализом и описанием.  
  
В датасете POSTS есть столбец Body (описание вопроса) и Title(в основном NaN).  
Как я вижу, в идеале должно быть так - на вход подается вопрос, этот вопрос сравнивается с уже существующими и выдаются, допустим, топ-5 похожих по убыванию. К этим топ-5 вопросам прикрепляются ответы (тоже, например, 5 в порядке рейтинга) и комментарии к ответам (в порядке рейтинга)

p.s. я взял отпуск, попробую успеть