MorningStar Ventures: Investment Trends in SEA and Oceania : An analyst report

You are a Senior Analyst at MorningStar Ventures Pte. Ltd, a top Venture Capital firm based out of Singapore investing in various sectors with special focus in Technology, Mobility, Energy, Safety. Your focus markets are SE Asia(without China and Japan) and Oceania (NZ and Aus).

Your boss has decided to understand trends in investing over last 5 years and has tasked you to lead the study. This time, you have to decide the methods and problem statement for the team. Welcome to the real world fellas!

You decide the problem statement and is as follows:

1. You decide to look at Linkedin, news articles and Crunchbase to look for startups with Headquarters ONLY in above mentioned geographical areas, which have wither Seed, or Series A, or Series B or Series C funding, or have been funded to less than 100mn US$. Each one of you will look for 20 such startups, and everyone has to find 20 unique startups. (We will provide a form to submit entries and an excel sheet to check if your startup has been named before by someone else. This means, START EARLY).
   a. This part is mandatory and you will receive 0 marks if you move ahead without doing this part.
   b. This part will constitute 40% of the weightage.
2. You fill the excel sheet with the relevant details under headings mentioned there and also identify a corpus of text relevant to the startup. This can be but is not limited to "relevant news articles about the startup", the "About us introduction on their webpage" etc. In a real analytics job rarely you get a PS or relevant data, but you are smart enough that you not only now has the access to the data (Once the part 1 deadline is over, we will compile all data and share with everyone).
   a. You will clean this NLP corpus by using standard pre processing techniques to remove stop words, punctuations etc. and provide the corpus and the link from which it was generated in the form above. The example will be provided beforehand.
   b. This will constitute 20% of the assignment.
3. Your boss is surprised at your strategy and look forwards to your results. Now that you have all the details of funding and a text corpus, you can use various analytical techniques to present trends and results.
   a. You can receive 20% credits for the algorithm part and 20% credits for the results and insights.
   b. Depending on the sophistication of the technique you use and results and insights you provide, you can additionally receive upto 25% bonus for the algorithms part and upto 25% bonus for the results and insights part.
   c. An example of a simple model with limited insights could be (But not necessarily) a Bayesian model, whereas a little sophisticated model with stellar results could be use of text vectorisation or even better, using word embeddings and utilising them to develop a topic model and utilising a topic model to generate funding insights.
   d. You do have the capability now to make claims like : "since 2016 the investments have focussed on blockchain technologies in Singapore markets whereas in Indonesia the major funding has drifted towards E-commerce" yada, yada, yada…

This assignment can earn you upto 150% of the weightage, with easily scoring upto 95% of the scores with minimal effort.