BCV Data Science Consultant / Candidate Take-Home Assignment

Objective: The goal of this assignment is to assess your ability to gather, analyze, and communicate data-driven insights.

Dataset: You are provided with data on US startups that raised either Seed, Series A, or Series B funding rounds in the past three years.

BCV-ds-consultant-take-home.csv

Column names are self-explanatory, however, please refer to the organization object schema and funding rounds object schema to obtain more context on the columns if needed.

The data should include:

- Company identifiers - Name, URL, UUID
- Founding year
- Geographic location (City, Country)
- Funding information
- Company's category and Category group type
- Other columns that might be helpful in performing the analysis

Tasks:

1. **Exploratory Data Analysis (EDA):**
    - Identify and visualize the top sectors receiving the most investments in the three years.
    - Compare the geographic distribution of startups in H1 ‘23 vs. in H1 ‘22. Any emerging startup hubs? Are there any particular cities or countries that attract more funding?
    - Identify the top 5% of the investors
        - Which category is the best to invest in?
        - Can you identify a pattern here?
    - Display the feature correlation map of numeric variables in a format that is easy to understand and interpret.
        - Can you identify any noticeable patterns in this map?
2. **Statistical Analysis / Machine Learning:** (Use appropriate statistical tests.)
    - Test the hypothesis: "Startups in the Fintech sector tend to receive more funding than others.”
    - Can you identify seasonality or trends in the investment data, i.e., are there certain times of the year when startups tend to raise more?
3. **Communication:**
    - Summarize your key findings in a presentation deck. The deck should contain relevant visualizations, insights derived from EDA and statistical analysis, and recommendations for founders (if relevant).
    - Get ready for a live presentation in front of a mixed audience of technical and non-technical people!
    - The live presentation will be 1 hour: 30-40 min to present, 15 min for panel questions, and 5 min for you to ask us questions.
4. **Code Review:**
    - Share your code/scripts through a private GitHub repo. Make sure your code is well-documented and organized.

- We encourage you to store the API keys as local env variables. Please make sure you don’t declare the API keys in the code.

5. **Self-Evaluation:**
    - Provide a self-evaluation of your work. What went well? What challenges did you face? How did you overcome them?
6. **Extra Bonus Challenge:**
    - Use an ML model to predict the success of a startup (based on a success metric proposed by you) using features available in the data.

Deadline: Please submit your completed assignment by the date outlined via email.

Evaluation Criteria: We will be evaluating your assignment based on the accuracy of your results, the quality of your code, the clarity of your communication, and your ability to critically analyze your own work. Please note that we will place the highest emphasis on the accuracy of your results and the clarity of your communication.

Good luck! We're excited to see your insights.