KTP Data Science Technical Assessment

This test is designed to test the technical and reasoning skills required for the Data Science role. Please use this as an opportunity to demonstrate your reasoning, attention to detail and coding style.

Feel free to reach out to the hiring manager if you have any blocking questions as you go. If you have minor questions or get stuck, please simply make a reasonable assumption and call that out in your work so we can follow your logic.

1. Data Details

The dataset "Technical Task Dataset.csv" contains company level data at yearly granularity. Each row corresponds to data from a company. There are five rows per company, each corresponding to a financial account from a different year.

The "Trading Status" field indicates the case where a company has defaulted. Active = non-default, Dissolved or Liquidated/Receivership = default.

2. Task Details

Using the data provided, please create a model that returns the probability that a company defaults given a set of inputs that you define.

Please restrict your analysis to data in the last financial account (i.e., "Account Year" = year in "Latest Accounts Date").

You may rely on any available literature and material to assist, but please use Python to code the model.

Once you've built your model, use an open-source app framework of your choice (e.g., Streamlit, Hex, etc) to turn your algorithm into a data-app that can be used to illustrate your model (e.g., inputs, outputs, visualizations, etc). You are free to illustrate your model via the data-app in whatever way you think is suitable. Think of business users as your audience.

3. Task Presentation

3.1.Code

Please create a Github repository with all the files you think necessary to demonstrate your solution. Please share this with:

- https://github.com/mattarderne
- https://github.com/MBazzi

- https://github.com/lotzma

Please include in your solution answers to the following questions:

1. In two or three paragraphs, please describe your methodological approach to the problem (e.g., how you framed the problem, any assumptions you made, why you chose certain techniques, etc). If applicable, please include any references to the literature that you used.
2. In 3 bullet points, please explain what feedback you'd give to the engineering team responsible for the data API to help them improve any aspect of the tool that you think would benefit.
3. In 3 bullet points, please give guidance to the business regarding any suggestions you'd give them for using this model in production.
4. What two things would you do to improve this test? One line for each.

3.2.Presentation

Please be prepared to run through your solution with a presentation (up to 10 minutes). This presentation should be pitched to non-technical business stakeholders, in order to convince them of the potential value of the model you have created.