# JP Morgan MLCOE TSRL 2026 Internship Question 1:

# Application for lending department of a bank

In the banking sector, strategic lending decisions are fundamental to financial success and growth. The Strategic Lending Division of a bank plays a crucial role in driving profitability by identifying and capitalizing on lending opportunities. This division is integral to the bank's operations, contributing significantly to its revenue stream.

Lending and credit activities are major revenue drivers for banks, often accounting for 40-60% of their total income. This highlights the importance of effective credit risk assessment and strategic decision-making in optimizing returns and managing risks. Developing a solution that enhances these capabilities is essential for supporting the bank's revenue generation efforts.

Your task is to create an application that analyzes the financial health of businesses seeking loans, enabling the bank to make informed, data-driven lending decisions. This challenge requires a combination of analytical skills and innovative problem-solving. By accurately assessing risk and potential returns, your solution will contribute to the bank's ability to allocate capital efficiently and profitably.

Part 1. Modeling of a balance sheet

1. We would like to forecast the balance sheet of a company.  Unfortunately, the different fields of a balance sheet are not independent.  Hence we have to construct a model that respect these identities.  For a short introduction of the problem, please consider the papers Pareja(07), Pareja(09) and Pelaez(11).  For a much more detail exposition of the problem, please consult Shahnazarian(04) and the textbook “financial forecasting, analysis and modelling” by Samonas, as well as other standard accounting textbooks.
2. Construct a very simple model of the balance sheet based on the tools of Velez-Pareja(09) and Velez-Pareja(10).  Please write down the mathematical equations government the evolution of the fields of balance sheet.  Is it possible to model this problem as a time series?  How do we handle the accounting identities?
3. Implement the model in Tensorflow and python
4. You can get income statement and balance sheet data from yahoo finance.  This blog post may help you.  <https://rfachrizal.medium.com/how-to-obtain-financial-statements-from-stocks-using-yfinance-87c432b803b8>
5. Choose some companies to apply your model to.  How are you going to train your model?  How can you test if your model is good at forecasting the balance sheet of the company?  How can you ensure that your forecast at least respect the accounting identities, and at least satisfying the asset = liability + equity identity as other relationship stated in the papers quoted here?
6. Can you use your model to forecast earnings?
7. What are the ML techniques we can use to your model to make it better?
8. Hint: simulation is highly related to prediction.  Suppose that you can simulate y(t+1) given y(t).  The prediction problem is very simple to implement numerically.  A general form of the model can be written as y(t+1) = f( x(t), y(t) ) + n(t),  where n(t) is some noise term to be specified, and x(t) are additional sets of variables that are relevant for the simulation.  What should x(t) be?

Part 2. Consider the problem of paper Kim(24) applying large language model to financial statement analysis.

1. Choose your favourite LLM to apply the problem of financial statement analysis.
2. Let’s try the task of balance sheet forecast using the same set of data as collected in part 1, does the LLM you picked perform better or worse than your model?
3. Is it possible to combine your model in part 1 and LLM to create an ensemble model that performs better than the individual model in balance sheet forecast?
4. Given the results of your analysis, pick a company you have analysed, what would you recommend to the CFO or CEO of this particular company given your results?
5. Consider the annual report of General Motors here. <https://investor.gm.com/static-files/1fff6f59-551f-4fe0-bca9-74bfc9a56aeb> on page 56 you can the income statement, and on page 57, the balance sheet.
6. Given this PDF file, use any tool you can think of, including any LLM, or any other tools such as chatPDF, pypdf, IronPDF, build an automatic tool to extract the income statement and balance sheet to answer the following questions:
   * 1. What’s the net income of the current year?
     2. What’s the cost-to-income ratio?
     3. What are the quick ratio, debt-to-equity ratio, debt to assets ratio, debt-to-capital ratio, debt-to-EBITDA ratio?
     4. What’s the interest coverage ratio?
7. if you are using LLM, is the model output robust to different runs? And please provide the model API version you use, such as 'gpt-4o-2024-08-06' for our testing. If you use other tools, please state clearly what you do and the version of the tools.
8. Does your program work for other annual reports? For example consider the one from LVMH, <https://lvmh-com.cdn.prismic.io/lvmh-com/Z5kVBpbqstJ999KR_Financialdocuments-December31%2C2024.pdf>
9. Is it possible to get your program to work on the annual report of Tencent, Alibaba, JP Morgan Chase, Exxon Mobil, and Volkswagan? How about Microsoft and Google?

Additional 50% for completing all the bonus questions.

Part B. Bonus question 1 : Suppose we would like to build a model to assign credit ratings given the annual report.

1. How does the mathematical form of such a model should look like?
2. Where can you find the data to train your model? Please indicate clearly your source with a copy of the data used uploaded to your repository.
3. Consider the annual reports of the Chinese real estate company Evergrande. What credit rating does your model assign to the company given the annual report? <https://www.evergrande.com/ir/mobile/en/reports.asp?year=2022>
4. How can you know that if the information of the annual report is correct? Can you build some tools to detect such issues? The book “Financial Shenanigans” by Howard Schilit may be helpful to you. Please test this with any of the bankrupted companies in the past to see if your program can at least generate warning signals.

Part C. Bonus question 2 :

Risk warnings:

1. How can we extract important points from the annual report that can be potential warnings. The bulk of the annual report are probably not important. However, some parts such as “qualified opinions” are very important. Examples of qualified opinions can be found here, <https://www.hkicpa.org.hk/professionaltechnical/assurance/example_auditors/SME-FRS-Jul2009.pdf>
2. What are the other important points to notice when reading an annual report?
3. Can we build an automatic extraction engine for extracting all the relevant paragraphs that show sign of risks?
4. Can you find a set of annual reports of bankrupted companies to test your program? What are the results?

Part D. Bonus Question 3

Suppose we would like to build a model to suggest the interest to be charged for the loan. Let’s only consider term loans. Borrowers of a term loan has to pay a fixed interest every period and pay the principle back at maturity. Term loans are usually priced as a spread to the US treasury bond of the same maturity. If the spread is 2%, and the US treasury bond of the same maturity has a yield of 5%, the interest of the term loan is 7%.

1. What kind of model will you use? Can you give a brief survey of the literature on the pricing of loans, or in general any defaultable bond or loan?
2. Where can you get the data to train your model? Please upload the data you use to train your model to the Github repository together with your code.
3. Suppose that a new client comes to the bank and would like to borrow money. The client has no publicly tradable loans or bonds. They are not listed and their shares are not traded anywhere. Does your model work for this case? How can you build a model to handle this case?
4. We would like to forecast the reselling price of the loans to other financial institutions after holding the loan for a month. Can your model handle this forecasting problem?
5. Can your model compute the 95% confidence interval of the price of the loan in a month?

# References

[Kim(24)] A. Kim, M. Muhn, and V. V. Nikolaev, “Financial statement analysis with large language models,” Chicago Booth Research Paper Forthcoming, Fama-Miller Working Paper, 2024, https://arxiv.org/abs/2407.17866

[Pelaez(11)] Mejia-Pelaez, F., and I. Vélez-Pareja, 2011, “Analytical Solution to the Circularity Problem in the Discounted Cash Flow Valuation Framework”, *Innovar*, 21 (42), , pp. 55-68. Available at: http://ssrn.com/abstract=1596426.

[Pareja(07)] Vélez-Pareja, Ignacio, (2007), “Forecasting Financial Statements with No Plugs and No Circularity”, Available at SSRN: http://ssrn.com/

abstract=1031735.

[Pareja(09)] Vélez Pareja, Ignacio, (2009), Constructing Consistent Financial Planning

Models for Valuation (August 15). Available at SSRN:

http://ssrn.com/abstract=1455304.

[Shahnazarian(04)] Shahnazarian, H 2004, 'A Dynamic Microeconometric Simulation Model for Incorporated Business', Sveriges Riskbank Occasional Paper Series, vol. 11. <https://www.riksbank.se/contentassets/09fcffd6cace4be8a915f3de3c010852/occp_11.pdf>