

Candidate instructions

As the next step in the selection process, you are being asked to complete a Job Demonstration. A Job Demonstration asks you to complete a task that is similar to the actual work completed on the job. Review these instructions carefully and in full.

Job Demonstration preparation:

- Review the **Detailed task description** section below.
- You will have **70 hours** to complete your Job Demonstration.
- If you require an accommodation to complete this simulation exercise, please contact your recruiter immediately.
- *Your work will NOT be evaluated for its compliance with company-specific policy.*

Follow-up discussion:

- A follow-up discussion will occur to discuss your Job Demonstration with the assessor(s).
- The follow-up discussion will last **1 hour**. During the follow-up discussion, you will have **20 minutes** to deliver your presentation.
- The follow-up discussion will be in-person. Refer to the meeting invitation for details.

Submission requirements:

- Submit a copy of your Job Demonstration to the Administrator by the specified deadline. If your Job Demonstration is not received by the deadline, you will be removed from consideration. See the body of the email you received for detailed submission instructions.

Detailed task description

- **Goal of Job Simulation**
This job simulation gives us an opportunity to assess your skills related to data analysis, model development, interpreting the results, documentation, and presentation. You have to analyze the data in the attached datasets, develop multiple models using different techniques, interpret the results, and document the analysis as well as findings. The data used in this study are simulated and do not correspond to any real accounts.

This exercise provides you with a simplified example of what our quants do on a day-to-day basis: developing and using models to solve business problems while clearly and concisely communicating the results to different stakeholders.
- **Task Description**
Financial institutions that lend to consumers rely on models to help decide on who to approve or decline for credit (for lending products such as credit cards, automobile loans, or home loans). In this job simulation, your task is to develop models that review credit card applications to determine which ones should be approved. You are given historical data containing one response (binary) and 20 predictor variables from credit card accounts for a hypothetical bank XYZ.

There are two datasets: training data with 20,000 accounts and test data with 5,000 accounts. Information about the variables is given in the Appendix.

As part of the job simulation, you are asked to:

- Conduct an exploratory analysis of the data, provide a summary, do any necessary data pre-processing in preparation for modeling;
- Develop and fit a logistic regression (LR) model, assess its performance, and interpret the results;
- Develop an additional model based on a machine learning (ML) algorithm selected from one of the following: Random Forest, Gradient Boosting (XGBoost or another implementation), or Feedforward Neural Network; assess its performance, and interpret the results. Make sure to explain why you chose this particular algorithm.
- Compare the results from the ML algorithm with those from logistic regression model; discuss them; select one of these models for credit approval; and describe the reasons for your selection;
- Describe how you would use it to make decisions on future credit card applications.

The analysis and model development can also cover any other considerations about the data or models that you deem important. In addition, you should address the following practical questions in your report:

- Do customers who already have an account with the financial institution receive any favorable treatment in your model? Explain your answer.

- **Deliverables and Guidelines**

- Please submit a report (pdf file) that describes all important steps in your data analysis, model development, model interpretation (for example, which predictors are important, what are the input-output relationships, are there any other interesting structure in the model, etc.), comparison of the models, and justification for your final model selection. The body of the report should be no more than 12 pages in length (font size 11 and spacing 1.2). In addition, you can include an appendix of no more than five pages that contains additional tables and figures. Include the important figures, tables, and discussion in the body of the report.
- You should also submit the code you used for the analysis with brief but adequate annotation so that we can review it. Indicate clearly the software packages and versions (if appropriate) that you used for the analysis. You should use Python, R, or SAS for your analysis.
- Please also submit a presentation deck (pdf file) with no more than 12 slides that summarizes your results and conclusions. Note that you will be presenting these results during superday to an interview panel. Make sure that the presentation is accessible to a general audience with the credit card application but not necessarily with the technical details in your models.
- You are allowed to review textbooks, published papers, websites, and other open literature in preparing for this case study. Note, however, that the material you submit in your report must be based on your own analysis and writing. If you relied on published scholarly work and open-source software for your analysis and findings (beyond what is generally known), you should provide references at the end of the report.

- **Appendix: Description of Dataset**

- Two datasets are attached:
 - Training dataset A with 20,000 accounts; and

- Test dataset B with 5,000 accounts. Use dataset A for developing your model and dataset B for predictive performance assessment.
- Description of variables:

VARIABLE NAMES USED IN THE DATASET	DESCRIPTION OF VARIABLES
Response: Def_Ind	Indicator of Default: Binary: 1 = account defaulted after an account was approved and opened with bank XYZ in the past 18 months; 0 = not defaulted; (Default means no payments for 3 consecutive months)
Predictors: Applicant's attributes derived from information obtained by bank XYZ at the time of application	
tot_balance	Total available balance (amount owed by applicant) for all credit products (credit cards, auto-loans, mortgages, etc.)
avg_bal_cards	Average balance (amount owed by applicant) for all credit cards
credit_age	Age in months of first credit product obtained by the applicant
credit_age_good_account	Age in months of oldest credit product obtained by the applicant with no past due payments
credit_card_age	Age in months of applicant's oldest credit card
num_acc_30d_past_due_12_months	Number of accounts that are 30 or more days delinquent within last 12 months (Delinquent means payment not made)
num_acc_30d_past_due_6_months	Number of accounts that are 30 or more days delinquent within last 62 months
num_mortgage_currently_past_due	Number of mortgages delinquent in last 6 months
tot_amount_currently_past_due	Total amount past due currently for all credit accounts
num_inq_12_month	Number of inquiries in last 12 months (An inquiry occurs when a lender requests the applicant's credit history from a credit bureau. This occurs when a consumer applies for credit.)
num_card_inq_24_month	Number of credit card inquiries in last 24 months
num_card_12_month	Number of credit cards opened in last 12 months
num_auto_36_month	Number of auto loans opened in last 36 months
uti_open_card	Utilization on open credit card accounts (Utilization is ratio of balance divided by credit limit)
pct_over_50_uti	Percentage of open accounts with over 50% utilization
uti_max_credit_line	Utilization on credit account with highest credit limit
pct_card_over_50_uti	Percentage of open credit cards with over 50% utilization
ind_XYZ	Indicator: 1 if applicant already has some account (checking/savings, etc.) with the bank XYZ; 0 otherwise

Simulation Exercise – Job Demonstration

Virtual Prep/In-Person Delivery

Selection & Assessment

Artificial Intelligence & Quantitative Analytics_Quantitative Analytics_Quantitative Analytics Associate

Job Code - 100621

10/28/2022



VARIABLE NAMES USED IN THE DATASET	DESCRIPTION OF VARIABLES
rep_income	annual income (self-reported by applicant and not verified)
rep_education	education level (self-reported by applicant and not verified) Four levels: high-school or below; college degree; graduate degree; other