

**Canada Energy Regulator**

**Written Test – CANDIDATE Copy**

**Data Scientist – NEB 10**

**22-ERG-IA-0131**

**Selection Panel: Nancy Hedges, Jenny Ye, José Ribas Fernandes.**

**Candidate Name:**

**Date:**

**General Instructions:**

* You have received this test electronically and will be allowed ten days **(10)** to complete (due October 12, at 12 p.m. MST).
* **Please include your name and today’s date on your exam.**
* Return your answer as an email attachment to [cer.hr@cer-rec.gc.ca](mailto:cer.hr@cer-rec.gc.ca)
* You are expected to complete this written test on your own merit. Please do not save or distribute any copies of this exam.
* Failure to return the exam within the allotted time will result in your elimination from the assessment process.
* Should you be successful in responding to the question, you will be notified, and an interview will be scheduled.

**This test should be treated as CONFIDENTIAL by both the candidate and the Selection Board. Please do not save or distribute any copies of this exam.** **Failure to do so may jeopardize your own standing.**

Good luck!

The questions below will be used to assess the following competencies:

|  |
| --- |
| **OC4 Written Communication Apply/Guide**   * Communicates clearly, concisely and effectively with co-workers and clients verbally. * Accurately conveys complex technical information and concepts in terms that are appropriate for the audience. |
| **TC1 Data Analysis (Proficiency level – Guide)**   * Ability to analyze data, conduct research, create reports, and influence program direction, using appropriate techniques from statistics and machine learning, and data visualization, with an appropriate understanding of ethics and privacy, and incorporating best practices from software engineering. |

**ASSIGNMENT**

You work for the Consumer Financial Protection Bureau, the US agency that regulates financial institutions. The VP of Loan Regulation has an upcoming presentation to financial institutions in Illinois. The VP says that in Illinois it takes on average 30.5 days to get a mortgage (from loan application to decision), which carries a cost of $13 billion annually to consumers and financial institutions. The VP is interested in testing a regulatory sandbox, requiring financial institutions to provide a fast approval process for certain applicants, to reduce the administrative burden. Based on data submitted by financial institutions in 2020, in the state of Illinois, as part of the Home Mortgage Disclosure Act (state\_IL\_applicant.csv and Data fields.docx), you are asked to assess whether:

1. A model can be built, and what features should be included, and a discussion of performance (report performance using X\_test.csv);
2. A model should be built, and a discussion of privacy and bias.

Dataset (please download a local copy of these files):1

1. Data reported by financial institutions in 2020, regarding 777,843 loans requested and outcome, on applicant and loan characteristics [[state\_IL\_applicant.csv](https://074gc-my.sharepoint.com/:x:/g/personal/jose_ribasfernandes_cer-rec_gc_ca/EfKMDZai_jJMkh7GaZE4t2cBS-di8Gn5Hqt1V7ifw9I0fg?e=8HyCEJ)];
2. Description of variables [[Data fields.docx](https://074gc-my.sharepoint.com/:w:/g/personal/jose_ribasfernandes_cer-rec_gc_ca/ER97B0cHi19AjGg5_toRw4ABSQmJ43lAJHtnrhHUjUoCYw?e=gL2yfi)];
3. Dataset to be predicted [[X\_test.csv](https://074gc-my.sharepoint.com/:x:/g/personal/jose_ribasfernandes_cer-rec_gc_ca/EZGF2pGWeVlBmCtu5reYvjUBgcOD_blTe22jw4pIaeQJEQ?e=LmlhgR)].

Please submit (saving files with your initials):

1. A deck for your VP to present to a senior audience in financial institutions (maximum three slides), answering the questions posed above;
2. Code for the analysis;
3. Prediction results for X\_test in csv format (0 for loan not originated and 1 for loan originated).

Use of automated machine learning tools will not be accepted. Code will be reviewed. If you re-use code from other authors (Stack Exchange, Kaggle, etc.) please cite which sections were re-used, and their source.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1 Modified from Federal Financial Institutions Examination Council (n.d.). *Home Mortgage Disclosure Act Data Browser*. Retrieved from <https://ffiec.cfpb.gov/data-browser/>

END OF ASSIGNMENT