IEGMA analyst online test and recorded interview

Context: Imagine that you are an analyst working on an IEG evaluation for which you need to identify a preliminary portfolio of projects and provide some basic information to the team leader on the composition and size of this portfolio. To accomplish your task you will need to use two datasets. You have 48 hours to accomplish the task but we don't expect the work to take more than 5 hours.

Outputs expected: You will need to send back the following files to us by email to Zorayda Buco Opana <zopana@worldbankgroup.org> **by 8:00AM (US Eastern time) on Monday October 28th**

1) Code: Scripts (.r or .py or .do) and / or notebook for topic modelling task and the optional bonus task. We should be able to run this code with minimal intervention.
2) Answers: Word Document or PDF file with your answers to the tasks, including the visualizations pasted in the document
3) If you decide to submit the 'bonus exercise', you will need to submit your code for this additional task as well as a term frequency table. For code, see 1) above. For term frequency table, please share .csv file.
4) Short recorded interview to explain your analytical approach: You will receive a separate email from interviews@hirevue.com, please login to the platform and follow the steps.

Part 1. Portfolio description

Instructions: Based on the two datasets provided, please do the following. For items 1. through 5., you will only need to use the Workbook "all.xlsx". For item 6., you will use both "all.xlsx" and "IEG_ICRR-PPAR_Lessons_2024-06-20.xlsx".

For all projects with status ‘active’ or ‘closed’ with board approval date July 1st, 2009, onwards, identify those mapped to the level 3 theme code ‘access to energy’ or sector codes related directly to renewable energy. The theme percentage and sector percentage must be strictly greater than zero. This is the ‘evaluation portfolio’. Based on this ‘evaluation portfolio’, please do the following.

1. Calculate the total number of projects in the ‘evaluation portfolio’. Hint: the correct number is between 500 and 600.
2. Calculate the total amount of IDA commitments for the ‘evaluation portfolio’.
3. Calculate the total amount of IDA commitments for the ‘evaluation portfolio’ going towards renewable energy-related sectors. Hint: the sector percentages indicate the proportion of a project’s commitments going towards various sectors.
4. Create the following tables or graphs based on the ‘evaluation portfolio’. For graphs, make sure that each graph has clear labels, a title that briefly explains what the graph is showing. For tables and graphs also add a sentence to interpret each.
    i. A table showing total amount of annual (based on calendar years) IDA commitments in ‘evaluation portfolio’ over time.
    ii. A table showing proportion of total amount of annual (based on calendar years) IDA commitments in ‘evaluation portfolio’ to total amount of overall annual (based on calendar years) IDA commitments over time. Please only include active and closed projects in the denominator as well.
    iii. A table and a graph showing the use of the following lending instruments for IDA commitments in the ‘evaluation portfolio’ over time: development policy lending, investment project financing, and program-for-results financing.
    iv. A table and a choropleth map showing total amount of IDA commitments in ‘evaluation portfolio’ by country.

v. A table showing the top 5 sectors based on volume of total amount of IDA commitments in the ‘evaluation portfolio’, excluding the sectors directly related to renewable energy which were used to identify projects. Hint: the sector percentages indicate the proportion of a project’s commitments going towards various sectors.

vi. A table showing the top 5 level 3 theme codes in the ‘evaluation portfolio’ based on average theme %, excluding the theme code ‘access to energy’.

vii. A table showing the top 5 co-financiers in the ‘evaluation portfolio’ based on their commitments, excluding IDA.

5. A heat map showing IDA commitments by project geolocations. Hint: assign the full total commitment for a project to each of the project’s sites.

6. For all projects in the evaluation portfolio, write Python/R/Stata code to do the following. Extract the ICRR and PPAR lessons from the lessons file. If a project has multiple lessons, group them. For these lessons, build a topic model (such as LDA) to identify 10 salient topics in the text. Based on the results of this, produce the following outputs:

   i. Table which shows the top 10 terms per topic.

   ii. Table which shows the probability of each lessons section (i.e. all the lessons text for a project) belonging to each of the 10 topics.

   iii. Table which shows the most frequent topic by country. Hint: use the most probable topic per document.

   iv. In 250 words or less describe your approach for task 6, and provide a summary of your results, including their interpretation.

Part 2: Bonus task: API Use

In addition to the above work, you can receive bonus points for the following task if completed successfully:

- For each project in the evaluation portfolio, use the World Bank documents and reports API (available here: https://documents.worldbank.org/en/publication/documents-reports/api) to bulk download their Project Appraisal Documents (PAD). PAD is the document type. Hint: use the project ids from the 'evaluation portfolio'.
- Then, search for the exact phrases "energy access" and "access to energy" in each document and generate a document term frequency table. The rows should correspond to the project ids associated with each document.
- Your output for this bonus will include both the R/ Python/ Stata code as well as the document term frequency table.

Part 3: Short recorded interview

Please follow the link sent to you by email by **interviews@hirevue.com**, check your junk mail folder if you have not received the email. You will be asked three questions related to the position and your analytical process to complete the tasks above.