Task for Candidates:

Junior Data Scientist, Product

Attached is a file containing long-term care claims data.

Before your next interview, please answer the following questions. For each answer, add your code (python / R), including documentation.

1. Explain the structure of the data file: what does each row represent?
2. Can you explain why there are missing values in the fields "claim_end_date" and "eb_reassessment_date"?
3. Describe the distribution of claimants age upon first approved claim. Add plot(s) as needed.
4. What is the distribution of genders among first time claimants?
5. Is there a difference between males and females in the prevalence of claimants who were ever denied a claim? Is this difference statistically significant?

Please answer the above questions in English.

For any questions regarding this task, please contact hila@assuredallies.com