Introduction

The junior data analyst assignment will assess candidates in three main areas: critical thinking, research and analytical skills; understanding of statistical and mathematical concepts; and knowledge of emerging technologies in the data field.

Section 1: Critical Thinking, Research and Analytical Skills Testing

As a data analyst supporting the research team in curating data and insights to back up their research on civic technology, please perform the following task;

1. Highlight the steps you will take to curate credible secondary datasets that support the research work.
2. Extract the data from the identified sources
3. Clean and prepare the data for analysis
4. Analyse the data, drawing actionable insights
5. Document all the insights in a shareable Word document, slides or PDF

Share the code used to extract, clean and analyse data in a reproducible manner.

Section 2: Statistical Skills Testing

Using a sample dataset for talent_survey, perform the following tasks:

1. Generate descriptive statistics for the demographic characteristics of the respondents.
2. Identify the most common programming languages that students were exposed to during their coursework.
3. Use an appropriate statistical method to test whether the average number of required industrial training sessions is equal to 2.
4. Apply a suitable inferential statistics method to draw conclusions about student satisfaction with the quality of lectures received.
   Where 1 = Very dissatisfied, 2 = Dissatisfied, 3 = Indifferent, 4 = Satisfied, 5 = Very satisfied
5. Use the chi-square method to analyze students' perceptions of whether the course design meets the current needs of the tech industry.
   where 1=strongly disagree, 2=disagree, 3=indifferent, 4=agree, and 5=strongly agree

Please share the code used to conduct the analysis above.

Section 3: Emerging Technologies Skills Testing

Using the text_data attached, perform the following tasks:

1. Define entity recognition
2. Based on your understanding of the concept of entity recognition, extract the following information from the text: geographical locations, organizations, people, dates, and amounts mentioned in the text. Present the results graphically.
3. Perform topic modelling on the text to identify key thematic areas that the authors discuss in the article.
4. Using transformer models, summarize the text into one precise paragraph.

5. Apply unsupervised machine learning techniques to classify the texts into the following categories, generating ratios for each class.
   Categories: *Advanced Gene Editing, mRNA Technology, CAR-T Cell Therapy, Organoids and Tissue Engineering, Single-Cell Genomics, Synthetic Biology, Biological Computing, Wearable Biosensors, Microbiome Therapeutics, Nanomedicine*
6. Extract and explain emerging trends within the industry
7. Draw conclusions.

Please share the code used to conduct the analysis above.

Take note of the data sources links below;

1. talent_survey
2. text_data