

Fyttlyf Data Science Team Test

We are pleased to invite you to the interview process for our Decision Science Team! This is a practical exercise that will test your programming and analytical skills, please **include your codes as a PDF** in the submission. The programming language that is acceptable: Python or R

Instructions: Please read carefully

- Submit 1 pdf file with all the answers. Your code, comments & output should be present in the pdf.
- You may not consult with any other person regarding the test.
- You may use internet searches, books, or notes you have on hand.
- The test has 7 parts, all of which are mandatory. Failing to complete any one part would result in the rejection of the submission.
- A thoughtful, clean & commented code is expected as a submission.

Part 0: Reading the data

- Please find the data (Fyttlyff_DS_Interview.csv) and read it as a Dataframe
- Observe the column names and the data types in each column.

Part 1: Data cleaning

- Write a function called `data_cleaning()` which, when called, would perform the following activity:
 1. Replaces the NA values with 0s in the data
 2. In column 'B' replace Jan with 1, feb with 2, march with 3 and so on...
 3. In column 'E' Replace "Came_From_Google" with "Google" and "Landed_on_the_page_Directly" with "Direct_traffic"

Part 2: Descriptive statistics

- Write a function called `descriptive_stats()` which, when called, would perform the following activity:
 1. Generates the summary statistics (Mean, Median, Quartile, standard deviation) of all the numerical columns
 2. Produce a list of all the unique values & data types present in the non-numeric columns

Part 3: Prescriptive statistics

- Can you write code and present the data which would help us answer (Text in "" are column names) :
 1. "Which_Place_in_India?" has the highest "How_many_Landed_on_the_our_Page?"
 2. "How_many_Landed_on_the_our_Page_and_clicked_on_a_button_and_started_filling_the_Form_and_Completed_and_submited_the_form?" divided by "How_many_Landed_on_our_Page?" is highest for "Which_Place_in_India?"

Part 4: Simple Machine learning questions

- Write a function called `pred_future()` which, when called, would perform the following activity:

1. Predict
“How_many_Landed_on_the_our_Page_and_clicked_on_a_button_and_started_filling_the_Form_and_Completed_and_submited_the_form?” for the complete year of 2022
2. Generate the overall MAPE of your prediction for the year 2021.

You may use linear regression, Huber regression, ARIMA or prophet.

Part 5: Visualization

- Please write a code to display :
 1. A line graph for “How_many_Landed_on_the_our_Page_and_clicked_on_a_button?” for the different “Which_Place_in_India?” over the months of the year 2019 & 2020. (Hint : On x axis there should be months for 2019 & 2020 and Y axis should be the “How_many_Landed_on_the_our_Page_and_clicked_on_a_button?” and there should different lines depicting different regions of “Which_Place_in_India?”)
 2. A line graph of the actual and projected number of “How_many_Landed_on_the_our_Page_and_clicked_on_a_button_and_started_filling_the_Form_and_Completed_and_submited_the_form?” for the months of the year 2021(Actuals values) & 2022 (Predicted values). (Hint : It should be a line graph)

Part 6: About the Previous projects

- Please describe any interesting project you did in the Data Science domain in more than 250 words. Attach Github links if possible.

Part 7 : Time management

- Can you please share your thoughts, in less than 120 words, on “If you get selected, how will you manage your time for this full-time internship opportunity”

Best of luck!