Test

We are pleased to invite you to the interview process for our Data Science Team! This is a practical exercise that will test your programming and analytical skills, please **submit your codes as a ‘Google Colab link’** & ‘**Power point presentation’** in the submission.

**Instructions: Please read carefully**

* **Submit 1 colab link with all the answers. The submitted colab notebook’s name should be in ‘<your\_full\_name>\_<date>’ format and the same goes for the PPT.**
* **Your code, comments & output should be present in the colab notebook. Please make sure that all the output code and text are organized and readable in the submitted colab notebook.**
* You may not consult with any other person regarding the test.
* You may use internet searches, books, or notes you have on hand.
* Read the questions carefully and answer accordingly. Failing to complete any one part would result in the rejection of the submission. **Code should be commented properly**.
* In case of doubts please make thoughtful assumptions.

**Start your colab notebook with a checklist mentioning the parts you were able to complete / was not able to complete.**

**Funnel Sheet questions**

The goal of the business is to increase the LV5/LV1 ratio. The price of the service is 100 INR/Person. Once the visitor has paid, we can say that we have acquired that visitor & we may now call them a customer.

|  |  |
| --- | --- |
| Lv1\_Visitors | Visited the home page |
| Lv2\_Visitors | Visited the sign up page |
| Lv3\_Visitors | Visited the thank you page after signing up |
| Lv4\_Visitors | Visited the pricing page after signup |
| Lv5\_Visitors | Paid for services after sign up |

1. Look at the data and replace the blank values in the ‘value’ column with logical values.
2. Create a function which takes region, Customer segment, Start Year & Month, End year & month as an input parameter and gives the following as an output.
   1. A line graph & a bar graph to show the trend of KPIs column for the given date range (i.e., between the Start Year & Month and End year & month). The graph should have appropriate labels, titles and other factors which would make it readable.
   2. Based on the input parameters, forecasts the Value column using the following algorithm and spit out the actual values and the forecasted values appended in a single csv file for each algorithm:
      1. SARIMA
      2. Long Short-Term Memory (LSTM) Networks
      3. Moving average
   3. From the csv created above find out which algorithm best predicted the values for the input parameters. The function should also create a summary table for the forecast where we could see the accuracy of the time series forecasting model. You may add more evaluating criteria, but the following are a must have:
      1. Mean Absolute Percentage Error (MAPE)
      2. Mean Squared Error (MSE)
3. You are the star data scientist of the company & are part of the strategy team which is tasked with growing the business. The team is relying upon you to find data backed strategies from the give data. Please create a power point presentation where you would present the data to answer the following questions.
   1. If we always spend 1 lakh INR/month on all the regions and segments, can you perform an EDA and find out for which Region, Customer segment, Year & month the CAC (Cost of acquiring Customer) was highest. (CAC = (number of visitors paid \* price of the service)/marketing spend). We would like to see visually how you arrived at the answer so some graph should be in order.
   2. If we have always allocated 1 lakh INR/month to Uddeppy, Dehradun, Ujjain & Indore region for marketing can you tell us, based on your best forecasting model values, which region should we allocate less money for the upcoming quarters? (The goal of the business is to increase the LV5/LV1 ratio and optimize the Cost of acquiring Customer)

**AB\_Test Sheet questions**

* [**https://abtestguide.com/calc/**](https://abtestguide.com/calc/)
* [**https://www.evanmiller.org/ab-testing/sample-size.html**](https://www.evanmiller.org/ab-testing/sample-size.html)

Fittlyf wanted to test which color of the sign-up button would yield higher Click through rate (Click/Visitors). We designed an AB test where control had the default color on the sign-up button and the treatment had a new color in the sign-up button. We ran the test, and the data is collected in the ‘AB test’ sheet. You are from the Global Data science team and are tasked with analyzing the AB test.

1. Assuming MDE = 3%, Significance level α: = 95% & Statistical power 1−β = 80%. Can you tell us?
   1. Have we reached the required amount of sample size to conclude the test?
2. At 95% confidence can you tell us
   1. Has the test reached statistical significance?
3. Is there a difference in your inference of the test with respect to the statistical significance when you compare the overall results against the results of
   1. Only Desktop data
   2. Only Mobile data
4. Write a function which automate the task you performed above. You can write 3 functions, one for each question, with appropriate input parameters.

**Writing skills**

In less than 250 words can you write how will you make sure you will make the most out of this Full-Time internship & how this opportunity would benefit your career.