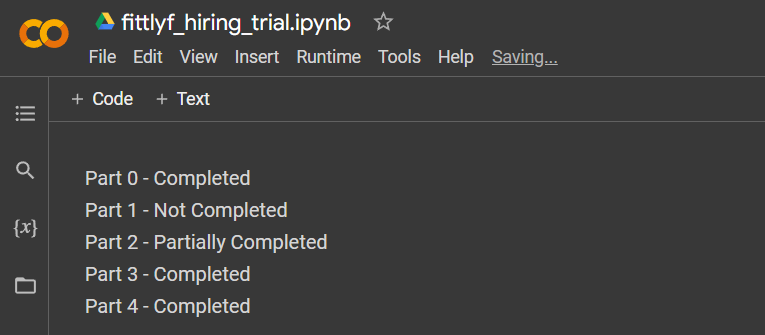
**Test**

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’** in the submission. The programming language that is acceptable is python.

**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.**
* **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.
* **The test has 5 parts,** **all of which are mandatory**. Failing to complete any one part would result in the rejection of the submission.
* 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. reference:**

**Part 0: Critical Thinking** ([link](https://www.analyticsvidhya.com/blog/2014/01/tips-crack-guess-estimate-case-study/) for reference)

* What do you think is the average revenue of your favourite youtuber? Explain how did you conclude this answer and what was your approach towards this question in brief

(OR)

* What is the per/day consumption of bread (grams) in Ukraine in the past 100 days? Explain how did you conclude this answer and what was your approach towards this question in brief

(OR)

* How many people would buy the premium face mask which is scientifically proven to be better than N95 of around 1000.rs in India? Explain how did you conclude this answer and what was your approach towards this question in brief

(OR)

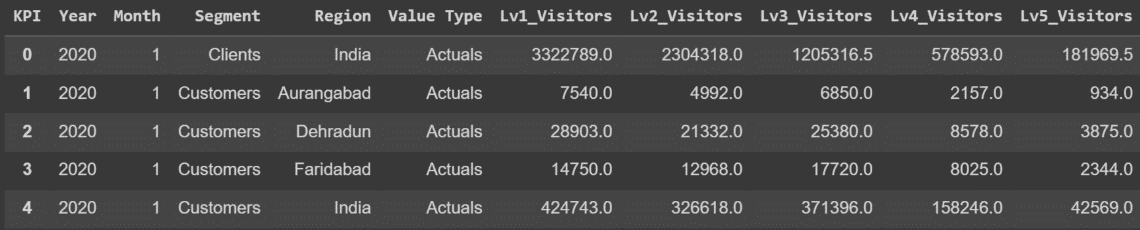
* What is the total sales of camping boots in the country of Switzerland? Explain how did you conclude this answer and what was your approach towards this question in brief.

**Part 1: Descriptive Analysis** (give clear comments for the codes)

* Please find the data (assignment.xlsx -> Funnel worksheet) and take it as the input (as data frame).
* Write a brief paragraph about what you think about this dataset along the lines of:
  + - What type of company this dataset belongs to?
    - Suppose that this dataset is for a website like Flipkart, what could be the possible definitions of the columns Level(visitors) 1, 2, 3, 4 and 5 in the given dataset? Do you observe any pattern?
* Give a pivot view summary and suitable visualization to answer the following question:
  + What was the total number of visitors segmented by each level, every month in each year?
  + What is the percentage difference in the number of visitors between different regions and years?
  + Look out for outliers and if there is any, use your preferred way to handle them. Use visualization to showcase outliers before and after treating them.

**Part 2: Prescriptive Analysis** (give clear comments for the codes)

* Transpose the data into a view as the reference given, ( numbers presented here are not precise )

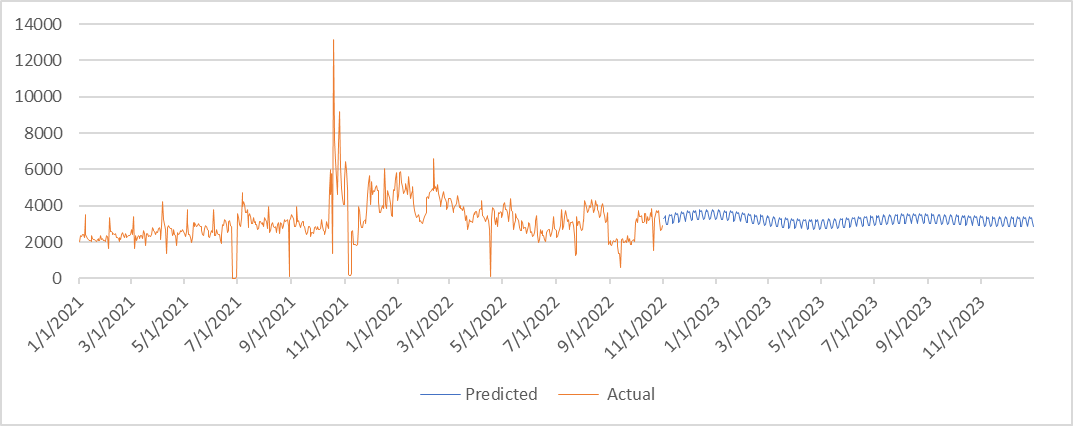


* Assume you are a data analyst at Fittlyf company, analyze the data for the region performing worst in all the years and prescribe what could be the reason and how to improve the number of visitors from that region.
* Based on the given data, identify which region is having a better YearOnYear growth.
* Guess what could be the meaning of Level1 visitors, level 2 visitors up to level 5 and then answer the following-
  + Create a new feature (Level 5 visitors/Level 1 visitors) and what are the top 3 states based on that created feature for all the available segments and each given year.
  + Create any other metric apart from (Level5 visitors/Level1 visitors) and perform the same task as above. Compare if the states are same in both the questions, if they are different try to create a hypothesis about the reason behind it.

**Part 3: Prediction** (give clear comments for the codes)

* Write a function called predict\_future(‘Region’,’Segment’) which, when called, would perform the following activity:
  + Predict “Level 5” future values for the next 6 months, given the parameters of the function. (Please make sure the parameters have default values in place) Also, plot it.
  + Generates the MAPE and RMSE of your prediction of the year 2022, 2021 & 2020 for the given parameters.

Plot a line graph of the level 5 actual numbers from 2020-2022 & in the same graph, there should be the predicted numbers for 2023. The x-axis should be the timeline from 2020 Jan to 2023 Jun and the y-axis should be the value of the level 5 column and predicted values. The below graph is just an example of how your plot should look like. You may use Rolling Average and ARIMA for forecasting. (link for reference:   
https://youtu.be/jiQM93dmUek)



**Part 4: A/B testing** (give clear comments for the codes)

* Using “AB\_TEST” sheet in the shared excel file, what is the possible metric you can create for A/B testing excluding no. of clicks and no. of visitors. (Any derived metric from the given no.of clicks and visitors ) – [link](https://segment.com/growth-center/a-b-testing-definition/metrics/) for refrence
* Perform an AB testing to find which variation whether control or treatment is better.