

Screening Case Study – Synergy Sense

Deadline: 19th August 2023

Please note: Late submission will not be considered under review. This is because you will have deadline for any paper and submission within time is mandatory

You are a data scientist of a Analytics company **Rinnegan Inc.** You have been given a task to study the data of 5000 youtubers (part-time, full-time) to understand their demographics and their earning analysis. You will be given with two datasets

- Revenue Data
- Calendar Data

Your task to do the following:

- Clean the data
- Answer the following questions:

SQL based analysis

- Trends of number of quarterly contracts beginning and end date (tutor count and generated revenue)
 - Trend of contract value based on their start date and end date QoQ
 - Trend of contract counts based on their start date and end date QoQ
- Number of customers whose revenue is greater than the average value in their respective departments.
- The share of revenue each region is generating w.r.t their department

Python based analysis

- Create two models
 - Regression – To predict the revenue which the tutor is capable of generating
 - Classification – To determine what is the geo of the tutor (cannot use region as independent variable)
- For each model, provide the following details
 - Important variables for the model
 - How much is the variable contributing (Convert coeff to elasticity)
 - Elasticity is defined as the share of the coefficient of total sum of coefficient
 - Model Health
 - Visualization (based on analysis)

The reviewer is not as much interested in the documentation, but in the coding skill. You will need to comment the code so that to clarify the purpose of the code.