

Assignment - Data Analyst

There are 3 questions in the assignment. You can make any assumptions if needed, but make sure that you mention them clearly in the document. For the submission, you need to share the running code, and pdf document where it is required.

Question 1:

As a part of the implementation of the Agroforestry program, we collect multiple farmer details. Like their name, location details, Total area they have, how much area they are giving for the plantation, document uploading status, dates of different activities. This data looks similar to dummy data [created here](#).

1. Many times we have observed that there are multiple data errors in the entered data. So, can you develop the data quality test cases, based on below program details.
 - a. We identify the farmer and fill in the initial details.
 - b. We check the details like, water availability, electricity availability and decide to onboard the farmer and collect the payment.
 - c. Then we do the carbon credit training and contract with the farmer.
 - d. Distribution of the plant and drip is done after contract.
 - e. Plantation of around 350 to 450 trees is done after the sampling distribution.
2. It is important for our Field Manager to get the field updates in real time, e.g. Amount of area identified, amount of plantation done, amount of data collected etc. So, create real time end to end data visualization/ dashboard connecting to above data. Instead of connecting directly to the above link, initially download the google sheet, upload it in your drive and then connect it to the dashboard. Make sure that, Data Quality checks you prepared above are part of this pipeline.

Question 2:

Farmers are getting to know about the F4F program from social media and some other means. These farmers call back to our call center. [This attached data](#) consists of dummy data similar to a small portion from call center data. How can we obtain some numerical insights from the data including comments which will help us to better understand the data.

Question 3:

As a part of the estimation of the tree biomass, we need the tree trunk diameter. Which is the tree diameter at breast height(DBH). But, we don't get the Diameter at breast height data from the drone orthomosaic. So, we decided to get it from the 3 variables given below -

- a. Tree species
- b. Tree height

c. Tree crown size

We have some ground data collected, which is uploaded in [this link](#).

So, please develop and train the Machine Learning Model to get the Tree DBH from the Tree Species, Tree Height and Tree Crown Size.

Share the code you have used to train the model and some description on the methodology and results in the document.