

**Problem statement:** At Fasal we provide data driven intelligence to our growers/customers. One of the major features we have is to provide data driven disease prediction to the growers. We use sensor data captured via our IoT device at a customer's farm to predict certain disease risk and spray schedule. These risks and spray schedule are developed using research work done by our team.

In the following problem you have to extract data, build a disease risk model and spray schedule based on conditions provided and SampleData, and finally design an experiment to validate the developed model in the field. For each question the data and problem is explained.

**Preferred language for the assignment: Python**

**Please create a git repository** and commit your codes, output data, answers to the repository and share the same with us. For any questions reach out to Madhumita Dash (madhumita.dash@wolkus.com)

**Q1.** You have a NoSQL database with 3 collections. The *customer* collections contain information for all your customers, the *farm* collection has information about the farm details of each customer whereas the *Device* collection has real time data from the IoT device installed on each farm. **Write a function (get\_data) to extract hourly data for a given crop, input date range and set of sensor values.**

For the function `get_data(crop_name, start_date, end_date, sensor_list)` where *crop\_name* will be a string like "tomato", *start\_date* and *end\_date* will be in the format "YYYY-MM-DD" and *sensor\_list* will be array of parameters from device collection whose value is required. E.g. if we need temperature in celsius (TC), humidity in % (HUM), leaf wetness in % (LW) and rainfall data in mm (PLV2) the array will be ["TC", "HUM", "LW", "PLV2"].

You can find sample documents for each collection below and details of each element.

#### **customer**

```
{
  "_id" : "600c1806289947de938c68ea",
  "name" : "Ram Mohan",
  "age" : 32,
  "gender" : "male",
  "dateJoined": 2019-09-30T09:18:29.044Z,
  "state" : "MH",
  "country": "IN"
}
```

\_id: unique ID for each customer; name: customer name; age: customer age in years; gender: customer gender; dateJoined: datetime the customer joined in ISO format; state: State to which the customer belong; country: Country of origin of the customer

#### **farm**

```
{
  "_id" : "uTMEtfc77M8eLdX7E",
  "customer" : "600c1806289947de938c68ea",
  "name" : "farm_tomato",
  "crop" : "tomato",
  "deviceId": "7HGRXY",
  "lat" :20.632189,
  "long": 74.8961
}
```

\_id: unique ID for each farm; customer: unique ID of the customer to whom the farm belongs (use this to join this collection with customer collection); name: farm name; crop: crop grown in the farm; deviceId: unique ID of the device installed in the farm (use to join with Device collection); lat: latitude of the farm; long: longitude of the farm

### Device

```
{
  "_id" : "uTMEtfc77M8eLdX7E",
  "deviceId" : "7HGRXY",
  "TC" : 24,
  "HUM" : 84,
  "datetime": 2021-07-30T09:18:29.044Z,
  "LW" : 0,
  "PLV2": 0
}
```

\_id: unique ID for each document; deviceId: unique ID of the device installed in the farm; TC: temperature recorded in celsius; HUM: Relative humidity recorded in %; datetime: exact datetime in ISO format when the data is captured; LW: Leaf wetness in %; PLV2: rainfall in mm.

**Note:** The datetime frequency could be different for each device (ranging from every 15 minutes to 1 hour). You would have to resample the device data to get hourly data.

**Q2.** Please see the 'Disease\_risk\_conditions.docx' for the conditions to calculate risk and spray advisories. In SampleData daily sensor data for a farm growing pomegranate is provided (similar to what you will get from the function from Q1). Based on the conditions provided, develop a daily risk condition and spray schedule for the farm provided in the sample data. Hint: Your final data will have farm data, date and 2 columns with risk (H, M, L) and spray (Yes/No).

**Q3.** If you have to validate the above model in the field how would you design an experiment to do so. What factors or variables will you collect and how many samples would you need? What all resources would be required to carry out the experiment?

**Scoring Criteria:**

- Code quality and documentation
- Analytical approach used
- Problem understanding
- Final solution presentation

In short, the less time it takes me to review your assignment the happier I will be :).