FMC Data Challenge

Thank you for your interest in FMC. As part of your interview process, we would like you to complete a data challenge that will allow you to showcase your technical and analytical aptitude as well as your ability to communicate to a variety of business and technical stakeholders.

Background

The provided data set entails a subset of field trap data. The field trap data capture pest count from pheromone traps placed in a grower’s field. As the growing season progresses, the traps in the field are checked and the counts are recorded. These data can lead to insights on field dynamics and pest pressure presence. This insight can be used to help growers plan the best time to apply pest management solutions to their crops.

A brief overview of the metadata for field\_trap\_data.csv ​is given below:

* trap\_id: the unique identifier of the trap
* date: the date the trap was checked
* crop\_name: the name of the crop
* region: the name of the region
* pest: the name of the pest
* trap\_count: the count of bugs from scouting
* is\_sticky\_sheet\_replaced: Boolean flag indicating whether or not the sticky sheet was replaced
* Lat: Latitude of trap
* Lon: Longitude of trap

Deliverables

Your task for this data challenge is to prepare a presentation intended for multiple stakeholders (product management, commercial and regional leadership, engineering, and data science). The body of the presentation should be for the general audience, but also include a technical appendix for discussion with the data science team.

Tasks and questions to consider for this data challenge:

* Do an exploratory data analysis (EDA) and provide any general insights about the provided data set.
* Consider constructing a “forecast or predictive” model. Given the sequence of trap checks from the field, how would you produce a prediction/forecast for the next trap check?
* You can use Python 3 and any available library/package you prefer, but provide any written code used in this process in the form of source files or notebooks. Please only show relevant items in the notebook/code – not super long please.
* As you analyze and code up this portion, here are some questions to consider:

1. What features do you think would be useful for this model?
2. What additional features could provide information for the model?

* What potential data sources would you like to have access to in developing the model?
* If this were a project undertaken by you and your team:
  + Which algorithm architectures would you consider exploring (either classical or recently developed) and why?
  + How far into the future would you build the model to forecast for? And why?
  + What error or performance metrics would you use for model scoring and comparison and why?
* A natural desire is to provide advance insight to growers and customers (dealers) about pest presence in their field by providing accurate and timely prediction/forecast. With the “predictive” model explored, how would you deploy it in a production environment?
  + Considering that these predictions/forecast could happen in in real time, does that change any considerations about the algorithmic structure of the model?