ML Engineer - Technical Assessment

Thanks for applying to AltaML and congratulations on being selected for a technical assessment! Upon successfully passing this assessment you will be invited to a final step of our interview process - a team interview.

For the technical assessment, we have provided you with a sample dataset to explore. This dataset, although artificial, is a small representation of the kinds of problems we work on day to day. In this case, a farming company has come to you with a dataset detailing some of the information they gather about the farms they own.

It is your job to look through this data and find patterns and make predictions that possibly will help the customer make more money.

Dataset
Information about yield from 1000 farms across a country.
- Id - Identifier
- Water - the average amount of water received by hectare
- UV - the average amount of light received by hectare
- Area - the size of the farm in hectares
- Fertilizer_usage - the level of fertilization
- Yield - total crop yield by farm
- Pesticides - the amount of pesticides used per hectare
- Region - region code
- Pesticides Used - comma-separated list of pesticides used

Requirements
Ideally, you will spend about 3-4 hours over the next week exploring the data in an attempt to find something interesting. As one of the tasks please identify the target variable, build a prediction model for it and evaluate its performance. Additionally, please provide insights into the data. As the vast majority of our projects are done in python, we ask that you use python 3 with pandas and submit a Jupyter notebook with your code along with report that highlights insights and describes your thought process (we expect your Jupyter notebook to serve the purpose of such report, please don't send additional files).

Additionally, please provide us with what your next steps would be if you were to proceed further with this project.