# Subpage 1: Cleaning agricultural data

## Sustainable farming at Mars

Crop diversification can lift yields, suppress pests, minimize risk of single crop failure, and improve income streams for farmers. Our client, Mars, is at the forefront of sustainable cocoa production.

[Feel-good pic of sustainable farmers at work]

*“We’re working with our supply chain partners to boost agricultural production without extending our overall land footprint. This means focusing on efficient, sustainable land use, as well as rehabilitating degraded land… Increasing yields will help us hold flat our land footprint, even while our business continues to grow.”*

## The question: economic outcomes

Done right, crop diversification offers many environmental and economic benefits. Done wrong, it can be a costly change to tried and tested farming methods.

So how do we know if we’re doing crop diversification right? We need good data and good analyses.

[Farm layout map]

In an ongoing partnership, Mars has been measuring the economic costs and benefits of crop diversification on experimental plots in Sulawesi, Indonesia. Using this data, EcoData performs the necessary analyses and forecasts to help our client to do sustainable farming right.

## The problem: messy data

Good analysis requires good data, which is challenging to collect directly from living, breathing agricultural systems. It can be hard to capture all the factors that might impact key metrics – climate, wildlife, soils, different farming products and methods, to name a few.

[Picture of farm complexity, like pollinators at a crop with human workers in the bg and a rainy sky]

Even if we collect everything we need, all data is prone to human error at the point of collection. Agricultural workers are no exception – they need to get on with the job of growing crops.

[farm workers collecting data in the field]

The more humans involved, the more errors that can creep in. Clean data become even trickier at the scale of a client like Mars, let alone across cultural-linguistic boundaries.

## Our solution: ETL pipeline

EcoData was well-equipped to work with the messy realities of our client’s agricultural data, without the need for specialized training at the point of collection.

*“All of their data scientists have PhDs. They’ve spent as much time getting their hands dirty on their own field collections as they’ve spent at a computer, and it shows. They had no trouble adapting to the data we provided.”*

[detailed flowchart of the ETL process]

Our Python data pipeline handles common issues like inconsistent formatting, incorrect data types, and entry errors. We also validated key metrics such as revenues against third party market data. Here’s a sample:

[code block from python ETL]

# Subpage 2: Empowering Mars with monitoring tools

## Why good metrics & visualisation matter

How are we going, and where are we headed? These are two critical questions for successful businesses.

Managers need easy access to key performance measures that can cut through the crunchy details and tell them to correct or stay the course.

*“EcoData distilled our data into something we could monitor at a glance. Their metrics and visualisations are helping us to optimise expenses and revenues and improve our farming methods.”*

## Bringing the data to our client’s fingertips

Using the high-quality data from our ETL pipeline, we developed a feature-rich dashboard for the crop managers at Mars. Our consultants worked closely with their team to ensure they had access to the hard numbers they needed.

[consultants discussing a brief]

Raw data collected from different timescales were standardised to meet monthly reporting requirements. We broke down expenses, revenues, and return on investment per crop.

Don’t take our word for it – check out our demo:

[**Big centrepiece** of this subpage: embedded interactive Tableau dashboard view with dummy data, or a nice video/gif if that’s not allowed. Ideally let it speak for itself.]

## So how was Mars going, and where were they headed?

With a few years of data under the hood, our visualisations immediately highlighted a concern. Though our client is in the business of sustainable cocoa, the return on investment was surprisingly poor compared to other crops.

[fully grown annual crops in the field beside tree saplings]

But cocoa trees fruit for 30 years and may provide a better ROI over time. So how might future yields lift cocoa performance against other crops?

This field data doesn’t exist. But instead of waiting another 25 years to collect it, we used simulations to forecast an answer.

# Subpage 3: Forecasting crop yields

## How we use forecasts to help

Growers want to know how management decisions will impact their bottom line. The sooner they have the information they need to make informed decisions, the better.

But how can get total return on investment for a crop that will live another 25 years *sooner* than that? How can we try hundreds of different cropping layouts when we only have the land and manpower to grow ten?

[Figure forecasting future yields]

To answer these questions for Mars, we developed simulations. Leveraging our curated data, we can forecast a range of potential economic outcomes for their agricultural systems.

## How we build useful biological simulations

*“All models are wrong. But some are useful.” --George EP Box*

A simulation is a kind of model. Useful models don’t capture every single part of a system. They zero in on only the most important bits, so we don’t lose the forest for the trees so to speak.

[schematic of an elegant model that captures the essentially of an imaginary system]

A useful model of a real system is also built around the data that’s available, or feasible. A model that can perfectly predict pest damage isn’t very useful if you have to tell it exactly how many aphid nymphs were on your whole farm in the fourth week of Spring.

[schematic of a perfectly useless model, the kind of thing that might give a farmer ptsd from a past encounter with a clueless wonk]

So how do we know what the most important factors are in a system, or what data we might need or be able to collect? We need people who understand how the system works.

What truly sets Ecodata apart from other consultancies is our domain knowledge. Our data scientists come from agricultural and ecological backgrounds. We bring the necessary expertise to accurately capture the real-world complexities of living systems, and to ask the right questions of growers to factor in the idiosyncrasies of their farms.

[consulting with farmers to close knowledge gaps]

### How the Mars simulation works

We built a simulation for Mars around their core business question, within their data collection capabilities.

“*Okay, so based on these data: what are the costs, revenues, and yields for plots of different crops? We want to know the answer projected over the 25-year productive lifespan of our key crop: cocoa.”*

Our client provides us with the starting conditions: how many of each type of crop was planted in the plot, and breakdowns of various costs and yields at different points in the lifecycle of each crop. Here’s a snippet of what the cleaned inputs look like:

[table showing sample from the crosswalks]

Our simulations then forecast materials and labor expenses, yields accounting for crop loss, and revenues. We do this in Python by modeling crop growth and cultivation each month over a 25-year time horizon. Here’s a snippet of the code:

[chunk of code showing off e.g. part of a timestep in the simulation]

We forecast under both optimistic and pessimistic assumptions, so our client can assess the full range of opportunities and risks associated with their cropping systems.

### Summarizing the results

Building on our dashboard for static financial performance, we designed a new Tableau dashboard to summarize our simulation results.

[screenshot or gif from the dashboard]

Once again, we empowered the team at Mars to slice through the results. They can, for example, assess how yields, expenses, and revenues evolve each month on month or year on year over the course of the simulation, looking for pain points in each crop’s lifecycle.

## New business questions

Our simulations allow Mars to ask how different agricultural plot designs might perform. They provide the specifications and we feed these ‘what-if’ scenarios into our model.

[example new plot design]

With the tools we’ve developed our client can explore the potential opportunities and risks when adopting new crop varieties, rotations, or intercropping strategies before ever planting a new seed in the ground.

[some kind of computerized plant/field render – a simulated farm]

With our help, Mars has evaluated hundreds of agroforestry plots for 30 different crops to discover designs that maximize total and average monthly returns and crop yields for their farmers.