

# Investing in Female Health

## How Clue Relied on Anaconda Distribution to Power their Menstrual Cycle Tracking App

### Executive Summary

Clue provides a mobile app that gives women a simple and secure way to track and predict their menstrual cycles. Clue's value proposition to users is that, by sharing individual health data through their app, Clue can provide personalized, predictive health recommendations. In order to make accurate predictions, the data science team at Clue needed tools that would enable them to gather and analyze data, build predictive models, and share these models within the organization.

Clue chose the Anaconda Distribution of Python because they enjoyed Python's friendly syntax and wanted the best open source packages for data analysis, machine learning, and data visualization.

While they were initially struck by Anaconda Distribution's ease of installation, they soon discovered myriad other benefits. Their data science team members found that, by using Jupyter Notebooks, scikit-learn, and seaborn, they could quickly gather and analyze data, prototype a number of predictive models, and share the results of their models via compelling visualizations. The Anaconda Distribution, which ships with all of these packages and more, makes this workflow seamless.

As stated by Data Science Team Lead Daniel Thomas, "[We] couldn't live without the Python data science ecosystem and libraries within Anaconda Distribution." Because of their success leveraging data to help users, Clue is expanding their health recommendations and collaborating with health centers and major universities to advance female health research.

### The Challenge

Clue's biggest challenge was finding the easiest way to build predictive models that would improve healthcare outcomes for their users. Their four-person data science team initially tried third-party analytics tools that offered drag-and-drop interfaces for analyzing data. However, they quickly realized they needed more powerful tools for predictive modeling and more flexible tools for data analysis and visualization.



### About Clue

Clue is working to help advance female health by giving users a way to track and establish patterns and predictions in their menstrual cycles. Based in Berlin, Clue strives to be the number one female health app in the world and is already being used across 190 countries.

Specifically, Clue's data science team wanted tools that enabled them to gather and analyze data, train and test a variety of predictive models, and share the results of those models with the engineering team. Their visual analytics tools could not meet these requirements.

Already familiar with Python, the data science team recognized that the Python ecosystem for data science was a great combination of power, flexibility, and ease of use in the open source community. They wanted a distribution that provided cutting-edge tools but did not require any effort to install or configure.

## The Solution

Clue's data science team turned to Anaconda Distribution to harness the power of the Python data science ecosystem. They found that downloading Anaconda Distribution was the easiest way to install the Python data science libraries they needed for their predictive modeling.

Over time, the libraries within Anaconda Distribution have become the core tools in their data science workflow. Specifically, they make heavy use of Jupyter Notebooks, scikit-learn, and seaborn, which all ship with Anaconda Distribution.

In a typical scenario, a data science team member performs exploratory analysis in Jupyter Notebooks. These notebooks are often shared internally with Clue team members, making it easy to share code, text, and visualizations in one place. They then transition to training and testing a variety of machine learning models within the scikit-learn library. Scikit-learn provides a simple interface to a number of complex machine learning models, thereby enabling Clue's data science team to prototype a number of different models rapidly. This, in turn, leads to better prediction performance.

After finishing their predictive model, Clue's data science team makes heavy use of matplotlib and seaborn to create compelling visualizations. These tools provide an interface that easily allows the team to build complex, interactive visualizations to share with the product team and business decision-makers.

The libraries within Anaconda Distribution are used at every stage of Clue's data science workflow. They rely on Anaconda Distribution to make installation simple and to ensure that all of their packages work together. Using open source tools like Jupyter Notebooks, scikit-learn, and seaborn, the data science team is able to nimbly and accurately respond to requests from their product team.

**Contact us at [ambassador@anaconda.com](mailto:ambassador@anaconda.com) or visit [anaconda.com](http://anaconda.com) to learn more.**

### About Anaconda, Inc.

With over 6 million users, Anaconda is the world's most popular Python data science platform. Anaconda, Inc. continues to lead open source projects like Anaconda, NumPy and SciPy that form the foundation of modern data science. Anaconda's flagship product, Anaconda Enterprise, allows organizations to secure, govern, scale and extend Anaconda to deliver actionable insights that drive businesses and industries forward. Visit [anaconda.com](http://anaconda.com).

---

*"[We] couldn't live without the Python data science ecosystem and libraries within Anaconda Distribution."*



**Daniel Thomas**  
Lead Data Scientist  
Clue