

Market Need

A common problem faced by companies working in analytics is the difficulty in creating "random" datasets that resemble actual company data.



Companies need purpose-built data for a variety of reasons, including the demonstration and testing of systems and analytical models.

Companies are often reluctant to use their own company or client data due to privacy issues, but they need data with the same characteristics as their real data.

Technology solution

DataGen - a tool that allows users to generate synthetic data. DataGen offers two modes of operation:

- ▣ **Manual generation:** The user specifies the features, rules and inter-feature relationships.
- ▣ **Automatic generation:** The data generation is seeded with an existing dataset. Data is then generated to match the features and inter-feature relationships of the existing dataset.

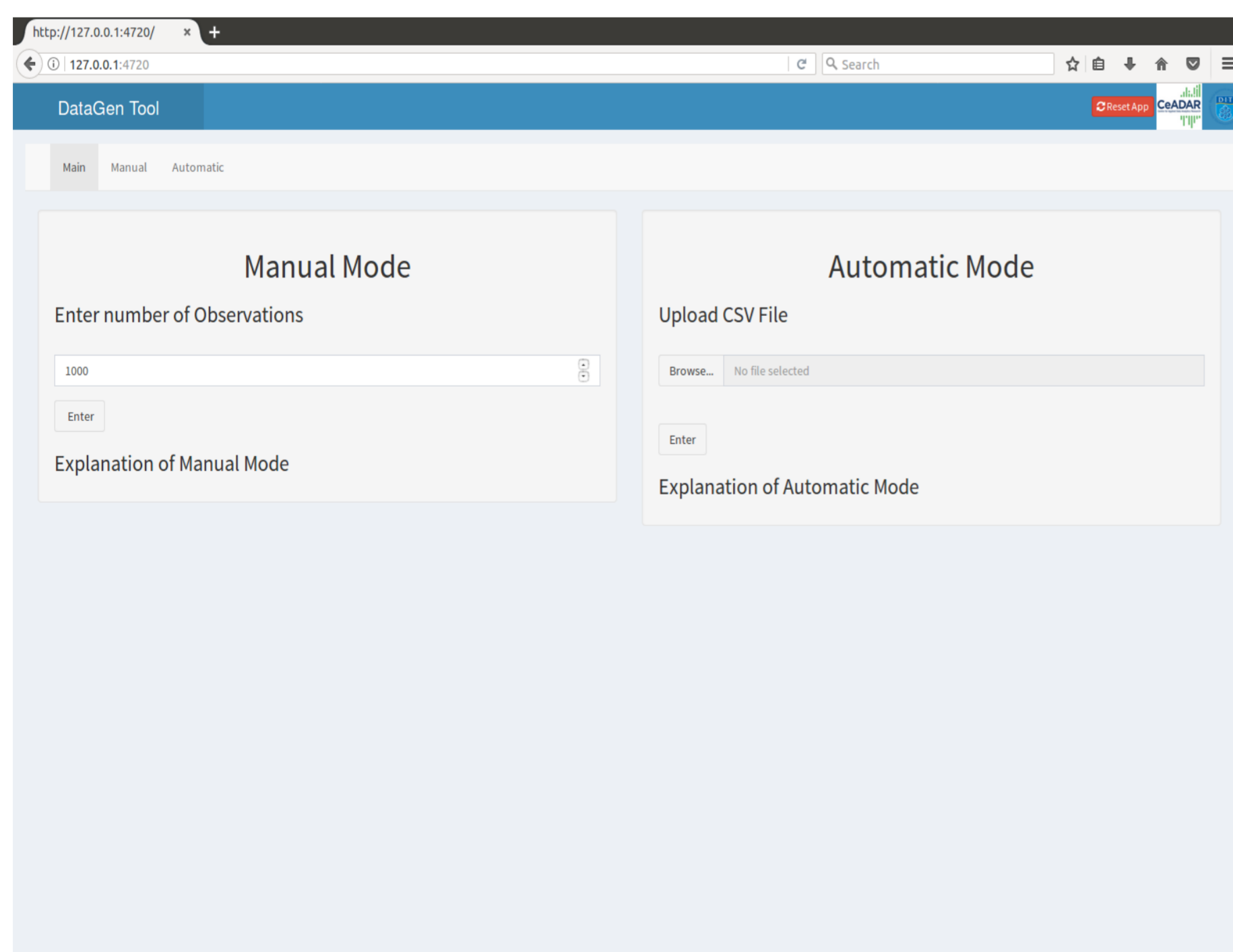


Figure 1: Manual and Automatic generation

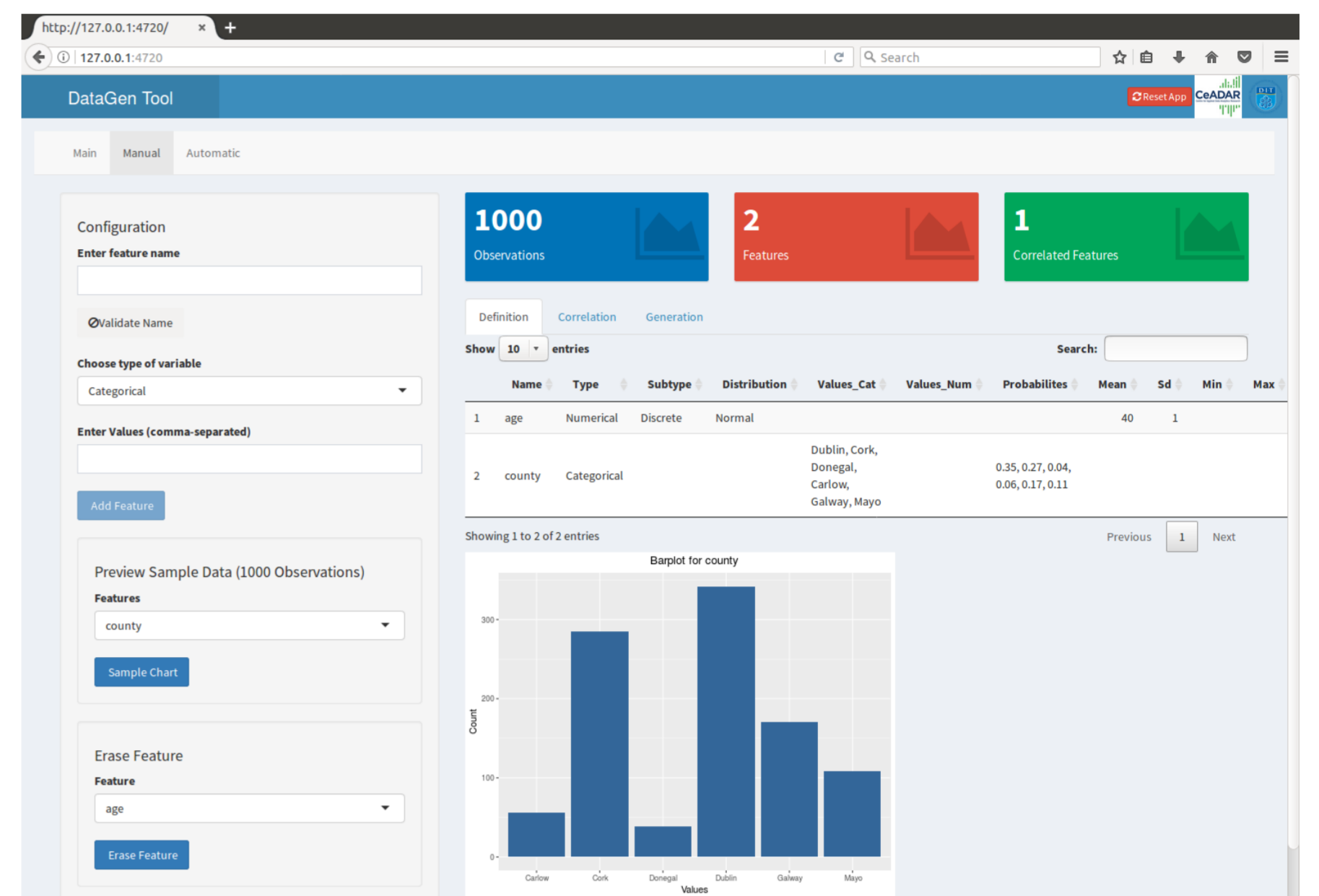


Figure 2: Manual generation

Key features

- ▣ The ability to generate datasets based on user-defined features and inter-feature correlations.
- ▣ The ability to use existing datasets to automatically define the characteristics of generated data.
- ▣ Flexible, easy to use functionality.
- ▣ Easy-to-interpret visualisations.

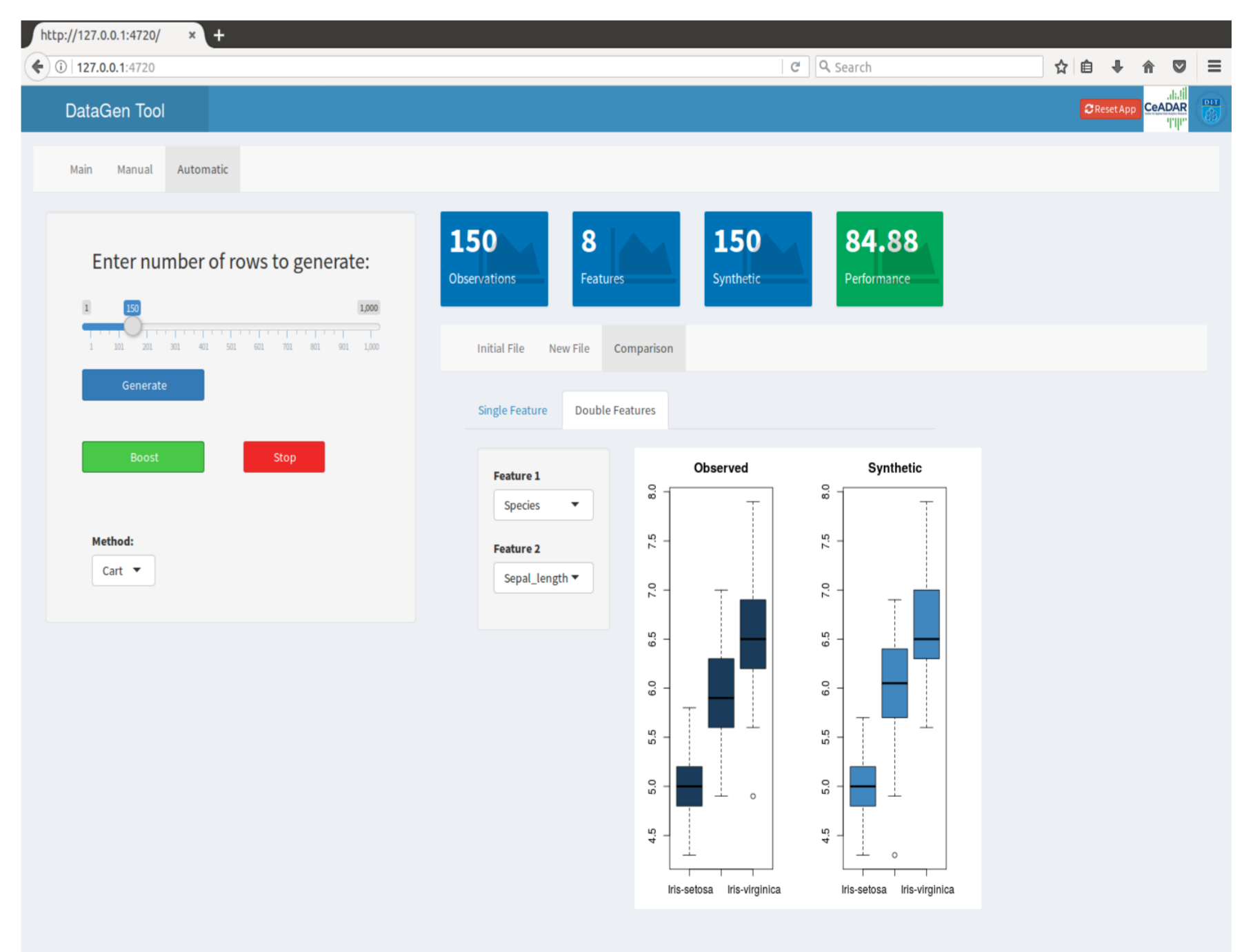


Figure 3: Automatic Generation

Research Team

Msc. André Rios Dublin Institute of Technology
Dr. Thibaut Lust Dublin Institute of Technology
Dr. Susan McKeever Dublin Institute of Technology

Follow **CeADAR Ireland:**

