

[Building Atlas](#) is a climate-tech startup that focuses on decarbonising the built environment. We work with companies that own or manage large portfolios of commercial buildings to help them find the best energy efficiency projects across their entire portfolios.

Our platform needs just the addresses of buildings: we then find and ingest data from a variety of public and private data sources and use advanced AI modelling to replicate physics-based building energy simulations.

We are looking for **a team of 2-3 Informatics students** who can support our product development team with data engineering and machine learning tasks. We are a small, mission-driven organisation. The students will get to work directly with our head of technology, full-stack engineers, the data science team, and the ex-Google founders.

Proposed start date: February 17, 2025

Responsibilities / tasks

- Explore and evaluate a variety of datasets related to commercial buildings in the UK and the EU; extract data from external APIs and integrate it into our platform
- Ingest, clean and transform data from relevant datasets to make it trustworthy and easy to use
- Design and implement appropriate data storage and processing solutions, ideally cloud-based solutions (we use GCP)

Milestones

- Procure and evaluate data on commercial leases from the Land Registry.
- Create a web scraper to download Building Air Conditioning reports and work with our data scientists to train a natural language/LLM model that extracts relevant insights from the reports.
- Identify sources of electrical grid capacity data in the UK, then ingest and normalise data to make it available for our optimisation algorithms.
- Explore Planning Applications data to evaluate its usefulness for our customers and complexity of ingestion; make a recommendation on how we should proceed.
- Gather examples of past retrofit projects from our partners, extract data on costs and use it to improve our costing calculations.

Qualifications and skills

- Experience working with externally generated datasets and a variety of structured, semi-structured, and unstructured data
- Familiarity with data storage technologies (e.g., relational DB, object stores, NoSQL)
- Experience with cloud platforms and cloud-native data services
- Experience in Python and SQL
- Strong analytical and problem-solving skills and desire to learn and experiment
- Bonus: experience with geographic, energy, or building datasets and interest in climate tech