Office of Sustainability

Sustainability Data Engineer

Department: Office of Sustainability
Reports to: Director of Sustainability, Julie Newman
Hours: Fulltime, 40 hours
Schedule: Hybrid (minimum 2 days per week in-person, on campus)
Salary Grade: 10

The Office of Sustainability at MIT is seeking a Sustainability Data Engineer to lead projects that produce actionable data insights and develop tools to help MIT advance campus wide climate and sustainability commitments. Our mission is to make MIT a leading example for generating scalable solutions that respond to the pressing challenges of a rapidly changing planet. We strive to achieve this by fostering collaboration that elevates diverse voices and promotes operational excellence, education, research, and innovation on our campus.

In this role, you will work collaboratively with the members of the Office of Sustainability team, as well as stakeholders from a mix of administrative and academic units across the Institute, and report to the Director of the Office of Sustainability. We share a common desire to solve problems, address environmental issues and injustices. If you are looking for an opportunity to work with a caring, innovative, and impactful team, we encourage you to apply for this position. Once established, the culture of the office encourages additional training and growth to continue to advance the position and overall responsibilities.

Position Overview:

As our Sustainability Data Engineer, you will be responsible for updating, building and maintaining the data architecture to track and report on all campus sustainability activities, including but not limited to climate mitigation and resiliency, food, waste, water and transportation. You will be responsible for formalizing the data acquisition, analytical structure and procedures to curate and organize data from multiple sources for inclusion in the MIT Sustainability Data Pool, our centralized data repository that is available to all members of the MIT community. This position will seek to enable data pipeline automation in its cloud infrastructure where feasible; optimize, monitor, and scale applications and associated infrastructure. Once these initial data sources are up to date, there is additional data that warrants incorporation and room for creativity to grow the platform. We envision an individual that can identify and build new tools and apps to inform future business intelligence.

Principal Duties and Responsibilities:

1. Refine and develop protocols to efficiently and robustly source, ingest and manage data, building on existing MIT processes and systems and in alignment with campus-wide data management practices.
2. Partner with Office of Sustainability team and student researchers to identify critical data gaps and needs.Coordinate a data curation and management process in alignment with and leveraging IS&T campus data management practices, strategies, and software.
3. Collaborate with and provide support to Office of Sustainability project management team in obtaining data from other MIT offices and external vendors.

4. Design and implement robust data ingestion and integration process, leveraging a range of software and technologies for data cleaning, validation, integration and governance, to populate, maintain and update the MIT Sustainability Data Pool with diverse and complex data sets,
5. Collaborate with data providers to maintain data quality, control, and integrity during data ingestion and processing, to align with their expectations and troubleshoot any anomalies such as duplicated lines.
6. Determine when statistical learning techniques (machine learning) can be applied and where they would add value. Execute these techniques and clearly communicate limitations to stakeholders.
7. Automate reporting for city compliance, stakeholder transparency, operational performance, and decision-making. Reports include, but are not limited to supporting MIT's greenhouse gas inventory, transportation trends, and waste management.
8. Support the preparation of data via the Data Pool to communicate findings
9. Other duties as required

Required Qualifications:

- Bachelor's degree from 4-year College or University in Computer Science/Engineering/ /Business/Math/Policy or related technical field is required. Master's degree preferred
- A minimum of 7 years of related work experience
- Demonstrated knowledge of traditional relational databases (SQL), big data technologies (Hadoop, Spark), computer programming experience (Python, R), and experience with APIs.
- At least 5 years of experience with an open-source data science programming language (i.e. *R, python)*
- Strong data visualization skills with at least 3 years of experience with data visualization software such as Tableau.
- Demonstrated experience collaborating with others to obtain, ingest, organize and clean data sets to enable analysis
- Demonstrated ability to work collaboratively and effectively as part of a team.
- Commitment to promoting justice, equity, diversity, and inclusion, demonstrated through previous work and interest in initiatives integrating racial, economic, and climate justice.
- Demonstrated self-awareness, cultural competency and inclusivity, and ability to work with colleagues and stakeholders across diverse cultures and backgrounds and serving the needs of diverse populations.
- Experience with cloud technologies, preferably AWS

Additional Preferred Qualifications:

- Big data experience (Spark, hive, Hadoop)
- Application of machine learning approaches
- Experience working with energy or other sustainability metrics
- Demonstrated ability to collaborate and work effectively with others and function well as part of a team
- Experience working in higher education