

Data Scientist for the Development Data Lab

We founded the Development Data Lab with the idea that harnessing new streams of data can dramatically improve the way that policymaking and research are conducted in developing countries. Across the world, there is a severe shortage of data that policymakers and researchers can use to design, implement, and evaluate effective policies. We need capable, motivated individuals to help us build an open data universe for poverty research, and to deploy these data to build better policy and contribute to making a better society.

If you work with our team, you will learn cutting edge methods in both computer science and anti-poverty research, and will deploy your data science toolkit across the stack and the data pipeline. Your ambition to improve your data science skills and take initiative will be rewarded with ample opportunities to learn, take ownership, and deploy your abilities in a creative and societally meaningful manner.

The role:

The [Development Data Lab](#) is seeking to hire a full-time Data Scientist to work with our team on the determinants of local economic development in India. This position will offer the opportunity to have a significant impact in the development and dissemination of new research, and the deployment of unprecedented data resources in India for public benefit. Qualified candidates will have strong quantitative and computer skills, with experience in data analysis, management, and visualization using R or Python and willingness to learn Stata. A commitment to coding best practices and fluency with version control software is required. Other programming skills, especially GIS software and front end visualization (e.g. D3.js), are highly desirable. Tasks will include scraping, managing, and analyzing very large datasets on poverty and growth in India, working with new methods in computer science and statistics, building interactive visualizations for media and the public, and supporting research projects as needed.

The work will focus on three areas. The first is contributing to the advancement of the recently released Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) - an open access repository currently comprising dozens of datasets covering India's 500,000 villages and 8000 towns using a set of a common geographic identifiers that span 25 years. The data scientist will contribute to the automation of testing and verification of the data, will expand the database (including the deployment of machine learning tools), and will have extensive opportunities to improve the stack and develop socially meaningful interactive visualizations. The second is to contribute to a new project on data harmonization for the public good, focusing on large, underutilized administrative datasets in India. The third is to support the advancement of several additional research projects including examining judicial bias against marginalized groups and the ability of politicians to influence judicial outcomes, patterns and impacts of urban segregation, and the impacts of mining on neighboring rural communities. These projects include partnerships with the World Bank, the Government of India, and research partners across multiple top-tier universities.

Grow with us. Compensation will be competitive and based on experience.

Location: Washington, DC, Hanover, NH, or Delhi preferred, but also the possibility of working remotely

Deadline to apply: Applicants are reviewed on a rolling basis

Desired start date: As soon as possible

Length of Commitment: 1 year commitment is required for this position, extension is possible

Compensation: competitive and commensurate with experience/qualifications

Please Note: US work authorization required.

Qualifications:

- A Bachelor's or Master's degree in a quantitatively rigorous field such as a natural science, computer science, statistics, or economics
- Advanced knowledge of Python and/or R
- Knowledge of Stata (desirable but not required)
- Ability to write and maintain clean code
- Knowledge of GIS software, interactive visualization techniques (e.g. D3.js) (desirable but not required)
- Familiarity with machine learning classification techniques (desirable but not required)
- Excellent management and organizational skills along with strong quantitative skills
- Fluency and excellent communication skills in English
- Flexible, self-motivating, able to manage multiple tasks efficiently, and team player
- Experience living in a developing country preferred (India ideally) but not required

If you are interested, please send a CV, transcript, single-authored writing sample and code sample to info@devdatalab.org. In the message, please describe your reasons for applying for this position and any other skills that you think will be valuable in this work that are not listed in your CV.

The Development Data Lab is strongly committed to diversity and inclusion. Applications by members of all underrepresented groups are encouraged.