



Data Science Engineer

Location: Mumbai, Maharashtra

Information Links

www.educategirls.ngo

[Founder Safeena's TED Talk , April 2019](#)

[2023 WISE Prize in Education](#)

[Worlds First Development Impact Bond in Education](#)

[First TED Audacious Project to be chosen from Asia](#)

About Educate Girls:

- Established in 2007, Educate Girls' is a non-profit that focuses on mobilizing communities for girls' education in India's rural and educationally backward areas. Educate Girls' scalable, replicable and sustainable program model resides on top of India's 'Right to Education Act'.
- From a 500-schools pilot in 2007, Educate Girls has metamorphosed into a 24,000+ schools, over 18,000 villages across more than 33 districts in the states Rajasthan, Madhya Pradesh, Uttar Pradesh & Bihar.
- Since inception, Educate Girls has enrolled 1.4+ million Out-of-school girls (OOSG); 1.9+ million children have benefitted from the remedial learning curriculum. Over 18.6 million total beneficiaries have been impacted from Educate Girls' program model. For more information log on to www.educategirls.ngo
- In the next 10 years, Educate Girls aims to build one of the largest second-chance program - Pragati - for adolescent girls and young women. The objective of Pragati is to create an ecosystem to enable 10th grade credentialing and access to better life chances for adolescent girls and women who have dropped out of the formal school-system. The goal is to transform 10 million lives in the next 10 years. For more information about the project, visit <https://www.educategirls.ngo/project-pragati/>

About the role:

The Data Science Engineer is expected to lead the data science and engineering team help EG maximise its impact by leveraging data and technology. This role entails both individual contribution as well as team management. This role is instrumental in improving the predictive analytics capabilities of the team and enabling rapid courses. This document reflects the job content at the time of designing the job description and will be subject to periodic change in the light of changing operational and environmental requirements. Such changes will be incorporated in the job description on annual basis correction in program delivery by building models and automated processes. Efficient management and mining of primary and secondary data sources to enable timely and error-free reporting to our stakeholders will also be a key responsibility of this role. This role will help grow data-driven decision making in the organization.

In this role, you will be responsible for:

Areas of Responsibility:

- **Data Modelling & Analysis**
 - Interpret data and analyze results and trends, using statistical techniques where appropriate.
 - Develop comprehensive dashboards to monitor performance & quality of program delivery.
 - Leverage predictive modeling to optimize targeting, improve program delivery, increase the pace of course correction, and more.

- Capture operational requirements, formulate the ML questions and liaise with IT vendor to implement the ML models.
 - Prototype ideas, research and develop custom data/ML models and algorithms as required; test model quality through A/B testing.
- **Data Engineering**
 - Build efficient and well documented methods for easy, repeatable extraction, cleaning and validation of data from primary and secondary data sources.
 - Build automated data pipelines & add data sources as required.
 - Automate manual processes; build the infrastructure necessary for optimal extraction, transformation, and loading of data (from a variety of sources leveraging Azure/AWS/Python/SQL technologies)
 - Optimize data delivery and data management practices of EG.
 - **People Management**
 - Lead the data science and engineering team.
 - Strive for excellence in performance of both self and team.
 - Prepare questionnaire/assessment and conduct interviews to hire consultants and/or employees in data related roles for the QAMIS team or other teams.
 - **Liaise with:**
 - Internal:** Program & Operations
 - External:** None

Education Background:

Minimum is a Bachelor's Degree in a relevant data science field or a related field of study. A Masters Degree in Science/Engineering/Management is desirable.

Work Experience:

- Minimum 5 years of relevant work experience data analytics/data visualization/ data management
- Minimum 3-5 Years of **SQL** experience and demonstrated ability to design and analyze data in relational databases. Experience with **PostgreSQL desirable**.
- Minimum 3-5 years' experience in **data manipulation/cleaning using Python**. Experience in **Pandas** and **machine learning** using **Python** is desirable.
- Experience in designing and developing **dashboards**.

○ Skills Set:

- **Soft skills:**
 - Excellent communication skills
 - Critical & analytical thinking
 - Target driven

- Interpersonal skills

- **Technical skills:**

- Intermediate to advanced skills in SQL.
- Basic to intermediate skills in Python and Linux.
- Demonstrated experience in dashboard design & development.
- Demonstrated analytical and assessment skills of quantitative and qualitative data.
- Strong skills in data modelling and analysis.
- Experience of machine learning.
- Knowledge of statistical analysis software (such as Stata, SPSS) would be a bonus.
- Experience in version control using github.
- Proficient in Linux.
- Good to have experience in AWS EC2, S3, Lambda, API Gateway etc.

- ✓ **Travel requirement:**

- Willingness to travel up to 25% of the time at the State/district level in program geographies

Apply : <https://forms.gle/9NH3uZnCfnk1sf2t5>