HR191	**POSITION DESCRIPTION**	UNIVERSITY OF CAPE TOWN IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

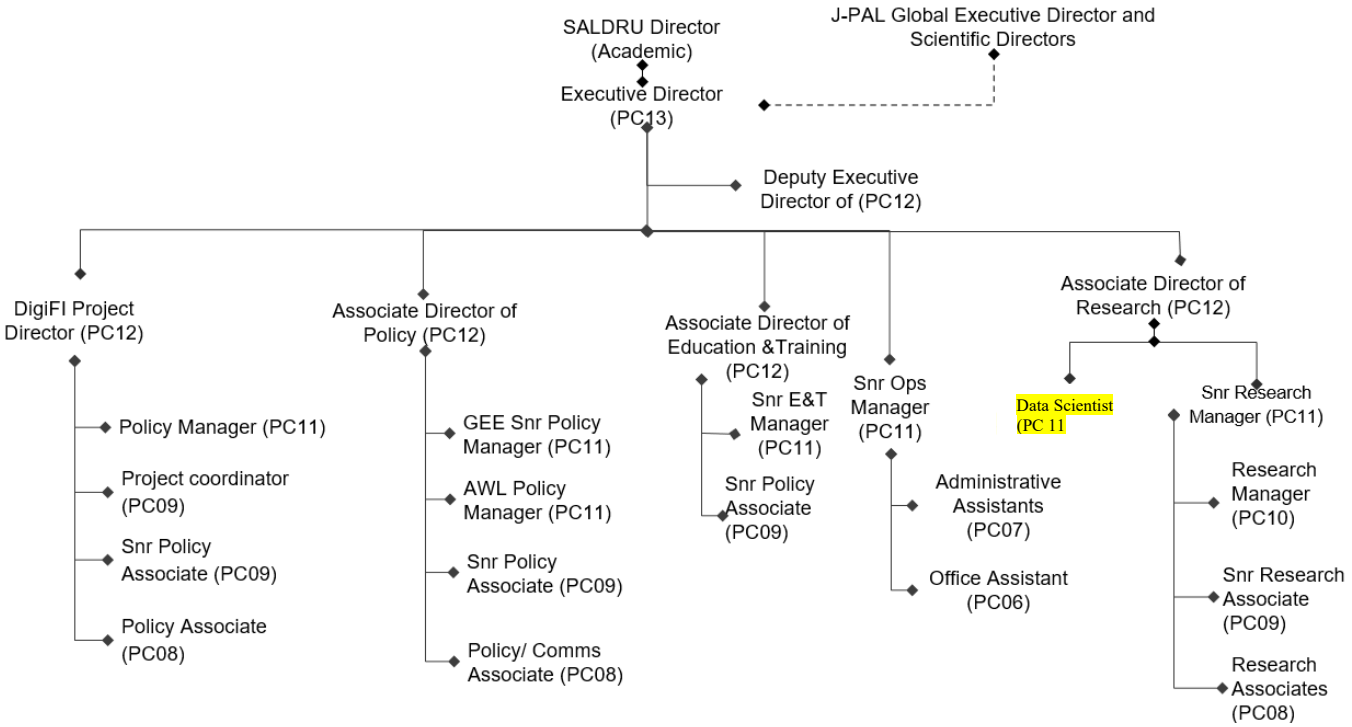
NOTES
- Forms must be downloaded from the UCT website: http://forms.uct.ac.za/forms.htm
- This form serves as a template for the writing of position descriptions.
- A copy of this form is kept by the line manager and the position holder.

POSITION DETAILS

Position title	Data Scientist		
Job title (HR Business Partner to provide)			
Position grade (if known)		Date last graded (if known)	N/A - new position
Academic faculty / PASS department	Commerce / economics		
Academic department / PASS unit	SALDRU		
Division / section	J-PAL Africa		
Date of compilation	10.06.2024		

ORGANOGRAM

(Adjust as necessary. Include line manager, line manager's manager, all subordinates and colleagues. Include position grades)



PURPOSE

The Data Scientist will be hired under J-PAL Africa's Water, Air, and Energy (WAE) Lab, which is a collaboration with the City of Cape Town (CCT) to co-generate research and inform the scaling of evidence-based programs and policies that effectively improve access to clean air, water, and reliable energy for all residents of Cape Town. The data scientist will be seconded to the Data Science team at the CCT and focus on two areas: 1) to design, develop and implement data infrastructures, pipelines, platforms and products to enable data functions to facilitate greater evidence-based and data-driven policy making within the City of Cape town; and 2) to support WAE Lab research projects to access and utilize relevant administrative data by liaising with line departments across the CCT and building relevant data sets the research teams can access. The WAE Lab project data sets will serve as case studies to enable the data scientist to improve the broader CCT data infrastructure, quality checks, governance, autonomous data pipelines, and more, helping to create a blueprint to further scale the infrastructure. Opportunities to contribute to research design, analysis and production of visualization may be available, and would come with the potential for coauthorship of academic papers.

CONTENT

Key performance areas		% of time spent	Inputs (Responsibilities / activities / processes/ methods used)	Outputs (Expected results)
1	Professional data technical functions	60%	1. Data analysis and synthesis ● Assist in conducting data analytics and synthesis; ● Keeps abreast of advances in digital analytics tools and data manipulation products. ● Undertakes data profiling and source system analysis, including different mechanisms for interfacing with source systems ● Presents clear insights to support the end-use of the data, including collect, collate, cleanse, synthesize and interpret data to derive meaningful and actionable insights; ● Use analytical tools and techniques. 2. Data product development process ● Design, build, and test complex or large-scale data products; ● Complete data integration services. ● Assist designing, building and testing data products based on feeds from multiple systems using a range of different storage technologies and/or access methods; ● Creates repeatable and reusable data products. ● Design build and maintain data pipelines so as to ensure continuous data availability to downstream consumers 3. Data Integration design ● Deliver data solutions in accordance with agreed organizational standards that ensure services are resilient, scalable and secure ● Expose data from systems (e.g. through application programming interfaces (APIs)), link data from multiple systems and deliver streaming services. 4. Data modeling ● Understands industry-recognised data modeling patterns and standards and when to apply them ● Compares and aligns different data models for multi-use. ● Understands the concepts and principles of data modeling and various data modeling tools; ● Produce, maintain, and update data models for specific business needs;	1. Data translated into valuable insights that inform decisions across City of Cape Town line departments and WAE Lab research projects 2. Data feeds integrated and separated in order to map, produce, transform, and test new data products. 3. Fit-for-purpose, resilient, scaleable, and future-proof data serviced and developed to meet user needs of stakeholders internal and external to the CCT 4. Data models produced, with an understanding of when and where to use different types of data models

			● Reverse-engineer data models from a live system 5. Metadata management ● Assist in designing and maintaining appropriate metadata for managed datasets ● Understanding of a range of tools for storing and working with metadata. ● Works with metadata to complete complex tasks such as data and systems integration impact analysis; 6. Programing and build ● Use agreed programming standards and tools to design, code, test, correct, and document moderate-to-complex programs and scripts ● Collaborate with others to review requirements and specifications where appropriate. ● Follow modern, iterative development approaches such as the use of version control, continuous integration, service architectures, etc ● Understands security, accessibility, and version control in the context of programming. 7. System administration ● Implement new technologies and solutions, physical and virtual, based on business needs and consistent with industry best practices ● Implement, and maintain procedures to ensure consistent provisioning, uptime, and performance. ● Provide automation and auto provisioning of infrastructure (Virtual and Physical) ● Install, configure, maintain, troubleshoot and performance-tune software, and peripheral devices. Anticipate issues and employ mitigation strategies. ● Perform standard configuration, management, and maintenance tasks of all systems both physical and virtual ● Monitor, test, and tune system performance; preserve and provide system log files and performance metric data as needed. ● Ensure secure user access and role validation processes ● Document and perform regular system backups and restore and ensure that systems/assets can be recovered ● Recommend and execute modifications to systems to improve efficiency, reliability, and performance in accordance with established Change control protocol.	5. Appropriate metadata repositories designed and maintained, which enable the organization and external researchers to understand CCT data assets, which will enable additional policy-relevant research and data-informed policymaking. 6. Code designed, written, and iterated from prototype to production-ready 7. Shared data science infrastructure designed and proactively managed

			● Create/provide required reports in response to business user needs ● Develop systems documentation for educating end users 8. Testing ● Review requirements, specifications, and define data-focused test conditions; and ● Analyze and report test activities and results. ● Automate data-focused testing where appropriate 9. Data engineering capability development ● Able to promote continuous professional development with regard to data engineering within the organization; ● Able and willing to develop and share knowledge of cutting-edge techniques; ● Able to develop and support data engineering capability across the organization.	8. Tests planned, designed, managed, executed, and reported, using appropriate tools and techniques, and working within testing regulations. 9. Data engineering knowledge continuously developed and improved, using multiple sources. Data engineering practices shared across departments and in industry.
2	Data innovation	10%	● Understands emerging trends on the organization in data tools, analysis techniques, and data usage; ● Introduces new techniques and platforms within the organization when appropriate and to the benefit of the City's residents	Data products and the data engineering domain consistently and proactively improved through various avenues of innovation.
3	Data-related problem resolution	10%	● Respond to problems with data processing related infrastructure, e.g. databases, ETL products and services as they occur; ● Initiates actions, monitors service and identifies trends to resolve problems in data-related infrastructure ● Determines the appropriate remedy and assists with the implementation of solutions and preventative measures.	Data-related problems logged, analyzed, and managed; appropriate solutions in turn identified and implemented.
4	Project collaboration	10%	● Work with professionals, technologists and analysts in the organization to produce, transform and test new scalable data products that meet user needs. ● Work with third parties, e.g. academia, industries, NGO's, etc on data analytics problems	Technical and non-technical CCT officials and external academics meaningfully engaged, co-identifying potential data use cases to inform policy priorities of officials and the research of academics.
5	Stakeholder relationship management	10%	● Consider opposing views to reach consensus; ● Uses and presents evidence to explain decisions made to stakeholders. ● Understands who the data engineering stakeholders are, what their priorities are, and the importance of managing relationships with them	Relationships with various internal CCT and external stakeholders (including from J-PAL Africa, academia, and others) developed, managed, and maintained; priorities, scope and timelines of requested products identified and aligned.

MINIMUM REQUIREMENTS

Minimum qualifications	A relevant tertiary qualification (honours NQF 8) preferably in Computer Science, Engineering, Statistics, Applied Mathematics, Economics, Information Technology or Information Systems or relevant field.
Minimum experience (type and years)	Two or more years working with large and complex data sets in a high performing environment, including experience extracting data from complex systems.
Skills	- A passion for working with data, including experience creating large data sets and analyzing administrative data and a belief in creating resources for public good and knowledge-sharing - Advanced written and oral communication skills in English, particularly an ability to communicate complex topics effectively to both technical and non-technical audiences - Proven ability to handle several different projects/tasks at one time, successfully plan, delegate and complete tasks, collaborate with a range of stakeholders, and meet deadlines while ensuring high quality work - Demonstrated ability to create and cultivate relationships with a variety of stakeholders, such as researchers or policymakers - Strong attention to detail and precision - Entrepreneurial spirit and ability to self-motivate and learn - Proficiency in programming languages such as Python /R, and SQL. - Demonstrable experience with data modeling, data warehousing, and cloud computing. - Experience with data visualization tools and dashboarding (Tableau, Power BI, etc.). - Familiarity with data security and governance best practices
Knowledge	- Comfort working with version control (e.g. GIT) is a plus, but not required - Completed coursework in microeconomics, econometrics, and development economics would be advantageous, but not required
Professional registration or license requirements	Non applicable
Other requirements (If the position requires the handling of cash or finances, other requirements must include 'Ability to handle cash or finances'.)	Ability and willingness to work from the Cape Town Civic Centre as project needs dictate.

Competencies (Refer to UCT Competency Framework)

Competence	Level	Competence	Level
Analytical thinking / problem solving	3	Individual Leadership	2
Building interpersonal relationships	2	Professional knowledge and skill	3
Communication	3	Resource Management	3
Creativity and innovation	3	Teamwork / collaboration	3

SCOPE OF RESPONSIBILITY

Functions responsible for	- Design, develop and implement data infrastructures, pipelines, platforms at the City of Cape Town - Relationship building across CCT line departments, J-PAL Africa staff, and academics working on WAE Lab projects - Extracting data from CCT systems for WAE Lab projects - Identifying opportunities to enhance data systems for improved evidence- and data-driven research and policymaking
Amount and kind of supervision received	Managed by the Associate Director of Research at J-PAL Africa and the head of the Data Science team at the City of Cape Town, though conducts much independent work.
Amount and kind of supervision exercised	Will supervise research associates on a project-by-project basis.
Decisions which can be made	- Process, approach, and strategy for designing and developing data infrastructures - Day-to-day extracting of data from CCT systems for WAE Lab projects
Decisions which must be referred	- Final approval to implement newly proposed data infrastructure - Must follow data sharing agreement and proper approval processes on which data can be shared with whom, and when

CONTACTS AND RELATIONSHIPS

Internal to UCT	J-PAL Africa Executive Director and WAE Lab Senior Research Manager, Associate Director Research
External to UCT	City of Cape Town Data Science Team, Kelsey Jack (University of California, Berkeley)