



Data Scientist

PesaKit is on a mission to empower every mobile money agent in Africa to achieve and deliver more digital financial services to the underbanked and underserved.

Our objective is to tackle the inherent supply-side challenges of mobile money and agency banking through:

- **Easing** an agent's daily liquidity and operations pressures
- **Consolidating the distribution model** by providing last-mile distribution of financial services on behalf of Banks, SACCOs, MFIs, MNOs, Remittance and Insurance companies.

The company is in search of a Data Scientist who will support our products and services, sales, leadership, and marketing teams with insights gained from analyzing company data with a priority focus on effectively building and managing an agent network in our target markets.

The company is rapidly scaling and as the Data Scientist, you will need to be adept at using large data sets to find opportunities for product and process optimization and using models to test the effectiveness of different courses of action. The right candidate will have a passion for discovering solutions hidden in large data sets and working with stakeholders to improve business outcomes. If this sounds like you, please share your resume to andrew@pesakit.co.ke

Some of the Roles you will be in charge of include:

- Work with stakeholders throughout the organization to identify opportunities for leveraging company data to drive business solutions.



- Mine and analyze data from company databases to drive optimization and improvement of product development, marketing techniques and business strategies.
- Assess the effectiveness and accuracy of new data sources and data gathering techniques.
- Develop custom data models and algorithms to apply to data sets.
- Use predictive modeling to increase and optimize customer experiences, revenue generation, ad targeting and other business outcomes.
- Develop company A/B testing framework and test model quality.
- Coordinate with different functional teams to implement models and monitor outcomes.
- Develop processes and tools to monitor and analyze model performance and data accuracy.

Requirements:

- Strong problem-solving skills with an emphasis on product development.
- Experience using statistical computer languages (R, Python, SLQ, etc.) to manipulate data and draw insights from large data sets.
- Experience working with and creating data architectures.
- Knowledge of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.
- Knowledge of advanced statistical techniques and concepts (regression, properties of distributions, statistical tests and proper usage, etc.) and experience with applications.
- Excellent written and verbal communication skills for coordinating across teams.
- A drive to learn and master new technologies and techniques.
- We're looking for someone with experience manipulating data sets and building statistical models, and is familiar with the following software/tools:
 - ❖ Coding knowledge and experience with several languages: C, C++, Java, JavaScript, etc.
 - ❖ Knowledge and experience in statistical and data mining techniques: GLM/Regression, Random Forest, Boosting, Trees, text mining, social network analysis, etc.
 - ❖ Experience querying databases and using statistical computer languages: R, Python, SQL, etc.
 - ❖ Experience using web services: Redshift, S3, Spark, DigitalOcean, etc.



- ❖ Experience creating and using advanced machine learning algorithms and statistics: regression, simulation, scenario analysis, modeling, clustering, decision trees, neural networks, etc.
- ❖ Experience analyzing data from 3rd party providers: Google Analytics, Site Catalyst, Coremetrics, Adwords, Crimson Hexagon, Facebook Insights, etc.
- ❖ Experience with distributed data/computing tools: Map/Reduce, Hadoop, Hive, Spark, Gurobi, MySQL, etc.
- ❖ Experience visualizing/presenting data for stakeholders using: Periscope, Business Objects, D3, ggplot, etc.