

KGATHOLA PUKA

Data Scientist, Data Engineer, Data Analyst

📞 067 355-7808

@ kgatholapuka@gmail.com

🌐 <https://www.linkedin.com/in/kgathola-puka-994295226/>

📍 South Africa, Johannesburg



SUMMARY

Data Scientist familiar with gathering, cleaning and organizing data for use by technical and nontechnical personnel. Advanced understanding of statistical, algebraic and other analytical techniques. Highly organized, motivated and diligent. Well-qualified Data Scientist experienced working with vast data sets to break down information, gather relevant points and solve advanced business problems. Skilled in predictive modeling, data mining and hypothetical testing.

EXPERIENCE

Data Engineer Intern

Nedgroup Investments- Johannesburg

📅 02/2022 - 06/2023 📍 Location

- Extracted data from NGI website, focused on FAQs and general information. Transformed data into txt and CSV formats and loaded it into an AWS S3 bucket.
- Deployed Haystack's InMemoryDocumentStore and achieved {95%} reduction in response time for the FAQ chatbot.
 - Utilized ODQA (Open-domain Question Answering) to extract information from diverse documents, resulting in a {30%} increase in accuracy and effectiveness.
 - Employed Haystack Retriever for efficient document search and scoring, contributing to a {40%} improvement in search performance.
 - Addressed challenges of Extractive QA, overcoming slower speed with efficient computational resources and reducing processing time by {50%} for large text data analysis.
 - Developed expertise in ETL pipeline implementation, resulting in a {50%} improvement in data extraction, transformation, and loading processes.
 - Developed expertise in ETL pipeline implementation and gained experience in data extraction, transformation, and loading processes
 - Acquired knowledge of ODQA systems and document retrieval, familiarity with utilizing ODQA for extracting information from diverse sources and employing Haystack Retriever for effective document search and scoring

Data Engineer Intern

Explore AI Academy - Johannesburg

📅 07/2022 - 01/2023 📍 Location

- Explored common data terminology and learned problem definition techniques
- Managed databases and performed data operations using SQL, ensuring efficient data management processes.
 - Acquired fundamental Python programming skills, including working in notebooks and manipulating data structures for efficient data analysis.
 - Gained hands-on experience with various AWS services for data and application management in the cloud, resulting in {80%} reduction in infrastructure costs.
 - Learned to connect and integrate large data sources to create data lakes in the cloud, leading to improved data accessibility and scalability.
 - Delivered government-mandated data wiretapping project in 40 markets by deadline, thereby avoiding penalties
 - Analyzed and update the data of 20 distributors and prepared Excel sheet weekly
 - Developed skills in data pipeline management using Apache Hadoop and Apache Spark, resulting in a {30%} increase in data processing speed.
 - Utilized Apache Spark to write functions and test quality metrics for big data processing, ensuring accurate and efficient data analysis.
 - Documented data lineage and gained practical experience in data analytics for big data, improving data governance and compliance practices.
 - Participated in team projects and collaborated with experienced data scientists, contributing to the successful completion of project deliverables.
 - Enhanced problem-solving abilities and communicated findings to stakeholders resulting in informed decision-making and improved business outcomes.

LANGUAGES

English

Proficient



Afrikaans

Intermediate



Isizulu

Proficient



Sepedi

Native



SKILLS

AWS

apache

data analysis

Data Extraction

EXCEL

Microsoft Office

PowerPoint

Problem Solving

SQL

Teamwork

Video

Video Conferencing

Data Visualization

PowerBI

Statistical analysis

python

Machine Learning

Data Wrangling

Web Scrapping

Cloud Computing

Model Deployment

Data Warehousing

Distributed Computing

Python Scripting

ETL/ELT Development

Cloud Automation Scripting

CERTIFICATION

Big Data Foundations - Level 1

IBM

AWS Academy Graduate - AWS Academy Cloud Foundations

Amazon Web Services (AWS)

EXPERIENCE

Business Data Analyst

Openserve South Africa - Johannesburg

📅 09/2021 - 12/2021 📍 Location

Utilized advanced data analysis tools and techniques to analyze complex data sets

- Collaborated with stakeholders to develop quarterly roadmaps based on data-driven insights, leading to effective strategic planning and goal achievement.
- Established a cloud-based SQL database to efficiently store and retrieve client data for analysis, improving data accessibility and reducing storage costs.
- Developed and implemented innovative forecasting models to optimize business productivity, resulting in a {15%} increase in revenue.
- Conducted rigorous evaluations of the accuracy and effectiveness of various data sources and analytical techniques, ensuring reliable and high-quality data analysis.
- Designed and maintained an automated dashboard using PowerBI to provide regular updates on business performance, improving data visualization and reporting capabilities.

Data Science Intern

Explore AI Academy - Johannesburg

📅 01/2021 - 07/2021 📍 Location

Cleaned and validated data using SQL and Python

- Performed SQL data manipulation and normalization using JOIN statements, Sub-Queries, and table relationships, improving data accuracy and efficiency.
- Conducted descriptive and inferential data analysis, including natural language processing on Twitter data, uncovering valuable insights for business decision-making.
- Built a movie recommender system using advanced techniques, resulting in a {20%} increase in user engagement and satisfaction.
- Utilized PowerBI for data interpretation and report building, producing comprehensive and insightful reports for stakeholders.

EDUCATION

Information Technology

Explore AI Academy

📅 01/2021 - 12/2021

Information Technology

Explore AI Academy

📅 07/2022 - 07/2023

Matric (Grade 12)

Barnato Park High School

📅 12/2019

CERTIFICATION

Python Intermediate Course

Dataquest.io

Python for Data Engineering: Fundamentals Part II Course

Dataquest.io

Python for Data Engineering: Fundamentals Part I Course

Dataquest.io

Programming Concepts with Python Course

Dataquest.io

FIND ME ONLINE



GitHub

<https://github.com/kgatholapuka>



LinkedIn

<https://www.linkedin.com/in/kgathola-puka-994295226/>



Personal website

<https://kgatholapuka.github.io/>