JOHN CHEGE MAINA

Entry Level Data Analyst

☎ 0726 058 103

✉ jcmaina400@gmail.com

in linkedin.com/in/john-chege-maina

PROFILE OBJECTIVE

Motivated junior data analyst aiming to leverage a strong understanding of data analytics techniques, statistical analysis, and data visualization tools to extract meaningful insights from large datasets. Proficient at interpreting complex data, collaborating with teams, and delivering solutions to align with business objectives. Ready to learn from seniors and contribute to the organization's data-driven decision-making process.

EDUCATION

08/ 2018 to 11/2022 Bachelors in Mathematics and Computer Science

Murang’a University of Technology

WORK EXPERIENCE

09/2021 to 12/2021 **Attachee**

Kenya National Bureau of Statistics

- Executed field data collection and management processes, ensuring a 98% data accuracy rate and timely delivery of analysis reports.
- Efficiently accomplished and analyzed data, extracting valuable insights from trends, resulting in an 85% accuracy rate in identifying key patterns.
- Engineered research surveys and accurately managed data collection process, achieving an impressive 95% response rate; provided a comprehensive dataset that served as a valuable resource for in-depth analysis.
- Effectively monitored and evaluated tasks, achieving a 95% completion rate and providing actionable feedback for process improvement.
- Demonstrated proficiency in data collection methods, achieving a 90% response rate and accurate representation of Kenya

PROJECTS

Used Car Price prediction app

- Downloaded data from Kaggle, and analyzed it to come up with insights. Trained a model with an accuracy of 87% for predicting the value of a used car. Deployed the model through Streamlit.

Laptop Price prediction using machine learning

- Downloaded, and analyzed it to unearth valuable insights. Trained a model that got an accuracy of 85% for predicting the value of a laptop based on the specifications that the buyer chooses. Deployed the model through Streamlit.

Customer Segmentation

- Using KMeans clustering, I have trained a model to predict which cluster a customer belongs to based on salary and spending capability, the model is performing quite well.

Obesity prediction system

- This system prompts the client to enter the Age, height, and weight. It calculates the body mass index for machine learning to predict the obesity status. The model attained an accuracy of 100%.

Data Analysis for Socio-Economic Decision Making

- Collected and analyzed data on actual agricultural outputs and their impact on various socio-economic outcomes.
- Performed regression analysis to identify significant correlations between agricultural performance and socioeconomic indicators.
- Prepared data visualizations and reports for policymakers through concrete data-backed recommendations for optimizing agricultural policies and improving socioeconomic outcomes.

Kenya Household Master Sampling Frame Analysis

- Collected data through survey experiments, interviews, and questionnaires and did data cleaning processes for the sampled households.
- Steered statistical analysis to generate demographic insights.
- Outcome: Provided accurate and comprehensive data on the household demographics, helping policymakers make informed decisions.

United States Accidents Analysis

- Performed data cleaning processes on the dataset and analyzed it to generate demographic insights.
- Established charts and provided concrete data-based commendations on strategies to curb road accidents.

Amazon food customer reviews analysis

- Conducted sentiment analysis on the Amazon food customer reviews dataset.
- Generated insights into the opinions of the customers and hence made visualizations and reports on the strategies to be put in place to improve sales.

Titanic data analysis and classification

- Cleaned the Titanic dataset from Kaggle and Analysed it to uncover some insights about the dataset.

- Through the analysis, we managed to know which cabin had the highest number of occupants, and which passenger class had the most people, between male and female, who survived and perished the most.
- Through the use of machine learning models, classified the gender, deaths, cabins, and age of the Titanic travelers.

Programming languages popularity analysis

- Performed data cleaning for the dataset and conducted statistical analysis to generate insights.
- Came up with a visualization for easy understanding.

SKILLS

- Microsoft Excel
- Python
- SQL
- Power Bi
- Tableau
- Machine Learning
- Web Scrapping
- STATA
- SPSS
- Hypothesis Testing
- Sentiment Analysis
- Data Science
- Statistical Analysis

CERTIFICATION

Data Entry
Ajira Digital (2021)

Quantitative Data Analysis
Sonek Data School October – December (2022)

Business Intelligence Analyst
Sonek Data School January – May 2023.

Standard Bank Data Science Job Simulation
Forage – November 2023

British Airways Data Science Job Simulation
Forage – November 2023

REFEREES

Available on request