Yalemzewod Gelaw

13 Astoria Blvd, Hilbert, WA 6112 yalassefa@gmail.com +61469738918 https://yalemzewodgelaw.com

SUMMARY

I am a Senior Data Analyst at HBF. I use Microsoft Fabric for data deployment and data visualization, SQL Server and Snowflake for data extraction and transformation, and R and Python for data management and machine learning. I am proficient in a range of computing skills over eight years of experience as a data analyst, researcher, and statistical modeler, working with different data in academia and public services. I am skilled in project management, stakeholder engagement, and effective communication.

EXPERIENCE

Senior Data Analyst | HBF | March 2025 – Present | Perth, WA, AU

- Managed and analysed complex episode-level data from Snowflake to identify the conditions associated with high claims.
- Developed a Streamlit app for quotes and claims related to a medical relational database.
- Created a diagnostic database.
- Collaborated with the broader health program team to understand and forecast diseases associated with high claims for intervention design.

Senior Data Scientist | HQIU, WA Department of Health |January 2024 – March 2025, Perth, AU

- Managed and analysed complex, large-scale, and longitudinal cloud-based datasets (in terabytes) to provide a better understanding of cancer patient journey, with a focus on patient safety and clinical quality.
- Performed patient-level data extraction and transformation large hospital datasets from Snowflakes and SQL server.
- Designed data flow, data models and deployment pipelines in Power BI Desktop and Power BI Service
- Developed reports and dashboards using Power BI.
- Engaged with dashboard user groups to monitor and evaluate the quality of care provided and to identify areas for improvement.
- Collaborated on machine learning models to identify diabetes-related discharges from discharge summary clinical synopses.

Postdoctoral Research Officer | Telethon Kids Institute |May 2021 - Present, Perth, AU

- Developed four data pipelines for annual production of global malaria burden estimates.
- Managed climate and health service datasets (in terabytes) to spatiotemporally predict malaria transmission risk using the Bayesian Framework.
- Designed, built, and maintained automated systems for four surveys and five surveillance data collection projects, streamlining data gathering, transformation and integration to reduce time and effort.
- Presented and published findings in scientific publications, reports, and educational materials.
- Collaborated with international stakeholders and provided support in cleaning and utilizing local data gathered from 3,385 health centres for decision-making.
- Led data extraction, cleaning, manipulation, and cross-validation for 42 data points.
- Contributed to spatial analysis in R and QGIS, prepared a training manual, and conducted a session on data management and spatial analysis for over 50 data analysts and program officers.
- Provided R-programming technical support to four research assistants, improving team productivity by preparing guide scripts.

Data analyst and researcher | University of New South Wales |March 2020 - May 2021, Sydney, AU

- Developed, documented, and implemented data cleaning, analytics, and reporting for five projects with four years of longitudinal data points.
- Collaborated on and led data cleaning, transformation, integration and analysis of four datasets, resulting in successful publications.
- Applied decision tree analysis using a machine learning algorithm to identify the influential risk factor for socio-emotional trajectories of children in NSW out-of-home care services using the *R caret* package.
- Provided statistical support to eight clinicians and three PhD students on utilizing R-programming for hospital record data management, analysis, and reporting.

Epidemiologist | Queensland Department of Health | September 2019 - February 2020, Brisbane, AU

- Analyzed routine surveillance data to provide spatial and temporal patterns of weekly reportable disease for program recommendations.
- Provided analytical support for Communicable Disease Unit project planning and program evaluation.
- Conducted monitoring and surveillance activities (three outbreaks), including data collection, management, and analysis.
- Utilized ETL process to extract reportable diseases and examined and visualized data to identify patterns on maps for monitoring.

Lecturer | The University of Gondar | September 2013 - July 2016, Gondar, Ethiopia

- Delivered face-to-face biostatistics courses to undergraduate Nursing, Environmental Health, and Medicine (400+) and Master of Public Health Students (200+)
- Mentored 15 Master of Public Health students and 20 Bachelor of Science (Nursing and Medicine) students.
- Led and coordinated 100+ health science students in 4 round field community professional practices.

EDUCATION

PhD in Epidemiology and Biostatistics, The University of Queensland, Brisbane, Australia (2016 – 2020)
Thesis: Towards sustainable TB control in Ethiopia–profiling high-risk geographical areas using spatial modelling
Key skills developed: Mathematical modelling, data visualization, statistics, machine learning, spatial modelling, meta-regression, predictive analysis, and scientific writing.

MPH in Epidemiology and Biostatistics, The University of Gondar, Gondar, Ethiopia (2011 – 2013)
Thesis: Effect of residence on mothers' health care seeking behaviour for common childhood illness in Northwest Ethiopia: a community-based comparative cross-sectional study.
Key skills developed: Field data collection, Data entry, EpiInfo, REDCap, Data analysis, survey data management, and scientific writing.

BSc in Environmental Health Science, Debub University, Hawassa, Ethiopia (2002 – 2006)
Key skills developed: risk assessment, surveillance, data collection and community engagement.

SKILLS

- Programming: R and Python
- Statistics, machine learning, predictive analysis, spatiotemporal analysis, data ingestion and modelling
- Data management & analysis: RStudio, Panda, SQL
- Data warehousing (snowflake)
- Data visualization: Power BI, MS Excel, Ggplot2, R-shiny, seaborn, matplotlib, Streamlit
- Communication & Negotiation
- Structural Thinking & Problem-Solving
- Scientific writing
- Leadership

PUBLICATIONS

Over 45 peer-reviewed publications, with 15 first and senior-author papers. Full list available on Google Scholar. For conference presentations, grants and awards please review my webpage.

PROJECT PORTFOLIO

MS Excel dashboard
R and Python programming for Data Science (GitHub)
https://yalemzewodgelaw.com/

REFERENCE

Up on request