

# Abdirashid Omar

Data analyst and data science researcher focused on turning raw public, geospatial, text, and operational data into clean datasets, reproducible workflows, dashboards, and decision-support insights. Experienced with Python, SQL, data cleaning, visualization, NLP topic modeling, geospatial analysis, research reporting, and visual storytelling for clear data communication.

 [rashiedomar.github.io](https://github.com/rashiedomar)  [abdulrashid-o-matan](https://www.linkedin.com/in/abdulrashid-o-matan)  [rashiidmatan@gmail.com](mailto:rashiidmatan@gmail.com)

## EXPERIENCE

### GRADUATE STUDENT RESEARCHER | KOOKMIN UNIVERSITY

Dec 2024 - June 2026 | Seoul, South Korea

- Built reproducible data analysis workflows for remote sensing, geospatial, urban, and computer vision datasets using Python and structured experiment pipelines.
- Cleaned, organized, and analyzed large visual and geospatial datasets to identify patterns, evaluate model performance, and prepare research-ready outputs.
- Worked with public datasets, satellite imagery, GIS layers, annotation outputs, and model results to support evidence-based analysis and reporting.

### ICT OFFICER | T-HUB INNOVATION CENTER

Jul 2022 - Aug 2023 | Garoowe, Somalia

- Managed ICT operations and technology infrastructure supporting innovation programs, training sessions, workshops, and daily organizational activities.
- Maintained computer systems, software tools, local networks, data storage, and communication platforms for smooth internal operations.
- Provided technical support to staff, trainers, entrepreneurs, and program participants, resolving hardware, software, and connectivity issues.

### FIELD RESEARCHER / RESEARCH ASSISTANT | RAAGSAN CONSULTING

Mar 2023 - Aug 2023 | Garoowe, Somalia

- Supported field research projects focused on governance, community dialogue, and social development issues.
- Collected and analyzed qualitative data through stakeholder interviews, field notes, and community consultations.
- Organized transcripts, summarized findings, and extracted policy-oriented research themes for reporting.

## ADDITIONAL PROFESSIONAL BACKGROUND

**DATA LABELING & ANALYSIS** | Data organization, annotation, qualitative analysis, transcript review, and structured dataset preparation for reporting or model development.

**DATA VISUALIZATION & DESIGN** | Visual design, layout, and content organization for clearer dashboards, reports, presentations, project documentation, and public-facing data outputs.

**TECHNICAL TRAINING & WORKSHOP SUPPORT** | ICT tool preparation, participant support, and smooth delivery support for technology-focused training, workshops, and innovation programs.

## EDUCATION

### M.S. DATA SCIENCE / BIG DATA | KOOKMIN UNIVERSITY

Sep 2023 - June 2026 | Seoul

### B.S. COMPUTER SCIENCE |

UNIVERSITY OF BOSASO

Sep 2019 - Jun 2022 | Somalia

## TECHNICAL SKILLS

### CORE DATA

Python, SQL, Pandas, NumPy, CSV, JSON, JSONL, data cleaning, EDA, reporting

### VISUALIZATION & BI

Power BI, Tableau, matplotlib, dashboard layout, data storytelling, report design, chart interpretation, word clouds

### TEXT & NLP ANALYSIS

Text preprocessing, topic modeling, LDA, sentiment/text mining, corpus preparation, metadata structuring

### DASHBOARDS & GIS

MapLibre, interactive HTML dashboards, GitHub Pages, GeoJSON, GIS impact analysis, public spatial data

## LANGUAGES

English: Professional  
Korean: Conversational

## DATA FOCUS

- Public and open data analysis
- Geospatial data analytics
- Text analytics and NLP datasets
- Dashboard-driven reporting
- Research data preparation

## SELECTED DATA ANALYSIS PROJECTS

### TOURIST TASTE KOREAN FOOD | REDDIT TEXT ANALYTICS

[GitHub](#)

- Collected, cleaned, and analyzed Korean food discussion data from Reddit-style online conversations.
- Built preprocessing covering lowercase conversion, URL/punctuation/number removal, tokenization, POS filtering, lemmatization, and stopword removal.
- Produced processed CSV datasets, frequency tables, word clouds, summary plots, and an interactive LDA topic visualization.
- Used text analytics to surface dishes, ingredients, discussion themes, and tourist-facing food perceptions.

### SEOUL LIVE CITY LENS | PUBLIC URBAN DATA DASHBOARD

[GitHub](#) | [Live Dashboard](#)

- Processed live Seoul public-signal data across 121 hotspots with crowd, transit, traffic, weather, coordinates, and category fields.
- Normalized public realtime endpoint responses into a cached client-side JSON dataset for dashboard use.
- Built an interactive MapLibre dashboard for place-based urban exploration, signal comparison, and focus-radius summaries.
- Implemented a scheduled refresh workflow so deployed dashboard data can update from public Seoul realtime sources.

### LISTEN TO THE RIVER | FLOOD IMPACT ANALYSIS

[GitHub](#) | [Live Dashboard](#)

- Combined satellite-derived flood masks with GIS impact layers to summarize affected buildings, roads, and named places.
- Integrated Sentinel-2 optical water mapping, Sentinel-1 SAR pre/post-event change analysis, mask fusion, and geospatial impact reporting.
- Produced a Baladweyne 2023 flood case study with 196.03 km<sup>2</sup> final story mask, 6,460 affected buildings, 542 affected roads, and 19 affected named places.

### WHEN SEOUL WAS BUILT | BUILDING-AGE DATA MAPPING

[GitHub](#)

- Processed public Seoul building approval and geometry data into citywide building-age summaries and map outputs.
- Built reproducible Python scripts for source inventory, age-layer construction, data normalization, and static cartographic export.
- Produced generated summary artifacts and a building-age map from a deduped mapped layer of 747,095 buildings.

### SOMALI WIKIPEDIA CORPUS | NLP DATASET PREPARATION

[GitHub](#)

- Cleaned, merged, and structured approximately 9,500 Somali Wikipedia articles for NLP and low-resource language research.
- Combined Wikimedia dump extraction with targeted live scraping to preserve tables, infoboxes, lists, captions, categories, and coordinates.
- Produced article text files, JSONL records, and a manifest with title, URL, word count, source choice, table presence, and infobox presence.

### FINTECH | STRUCTURED LEGAL DATA PIPELINE

[GitHub](#)

- Built a structured legal-data preparation pipeline with parsed source data, annotation layers, finalized rule packs, runtime examples, and dashboard bundles.
- Created dashboard views for deterministic legal preparation flow and non-LLM runtime workflow.
- Used Python scripts to parse source materials, build bundles, validate outputs, and prepare reproducible dashboard artifacts.