Malachi Dunn
Data Engineer • Data Scientist

malachi@mindwyre.org | (+1) 208-625-1747 | Oldtown, ID 83822 United States
linkedin.com/in/malachiomally | github.com/malachi-mindwyre | mindwyre.org

Profile Summary

- Accomplished Data Engineer proficient in Python, SQL, Airflow, Spark, and Scala, including building ETL pipelines predominantly in a GCP cloud environment.
- Solid proficiency in data modeling, database architecture, and utilization of cloud-based data solutions and back-end systems for distributed data processing.
- Expert in streaming services, advertising, mobile apps, and web-scale architecture, with extensive experience working with REST APIs in batch and real-time processing of data.
- Skilled in designing and executing scalable, data-driven solutions for Martech and Adtech.
- Specialized in predictive analytics, time-series data, and utilization of ML models in MMM, MTA, and other marketing and advertising strategies, leveraging these skills to drive data-driven decision-making and optimization strategies.
- Effective collaborator within agile environments and SDLC, utilizing interpersonal skills to reach alignment among various teams with differing priorities, to ensure steady progress even within tight deadlines.

Education

Georgia Institute of Technology | Masters, Analytics — Atlanta, GA (In Progress)

Lewis-Clark State College | Bachelors, Mathematics — Lewiston, ID | GPA: 3.700/4.0

Professional Experience

CopperWyre | Senior Data Engineer, Martech

Contract, Remote — Aug. 2024 - Current

- Performed ETL on user data to send to Experian and Liveramp that allowed advertising partners to target specific users based on different attributes in targeted MMM and MTA ad-campaigns.
- Increased user match-rates to advertising partners for user targeting by restructuring legacy SQL and Python code that optimized internal data collection techniques.
- Developed a comprehensive media mix model that identified potential savings by optimizing channel allocation, maintaining overall sales performance while reducing marketing spend.
- Automated A/B test experiments by creating executable VertexAI / Google Colab notebooks that automated test and control Martech campaign KNN matching for Google PLA feed experiments designed to increase revenue on targeted advertisements.

Disney | Lead Data Engineer, Adtech

6mo. Contract, Remote — Jan. 2024 - Jul. 2024

- Built revenue-impacting, direct-to-client data pipelines for the entirety of Hulu, Disney+, and GAM ad-logs for vendor reporting, ad-targeting, and maintaining CCPA and GDPR data compliance within internal tables using GCP, Snowflake, Airflow, and Python, enabling real-time campaign performance visibility for enterprise advertisers.
- Optimized data processing architecture through multi-stage Snowflake pipelines (raw → staging → complete) ensuring data integrity while accommodating delayed ingestion from streaming sources.
- Engineered a self-service reporting solution by integrating Habu clean room as a request management system, reducing redundant Data Engineer workload, eliminating duplicate client charges, and increasing advertiser confidence through transparent campaign analytics.

Pinterest | Data Analyst, Martech

12mo. Contract, Remote — Jan. 2023 - Jan. 2024

- Built an automated insights generation system that transformed manual reporting processes into programmatic workflows, reducing analysis time from days to hours for high-value advertisers by leveraging Presto, Spark SQL, and Pandas.
- Conducted sophisticated cohort analysis identifying statistically significant behavior patterns among targeted user segments, enabling advertisers to optimize campaign targeting and increase conversion rates.
- Engineered data pipelines that automatically populated Google Sheets with visualizations and used internal ML tools to generate contextual narratives, dramatically scaling the insights delivery process.

- Led technical investigation identifying causal factors affecting content engagement rates based on upload methodologies, uncovering critical platform optimizations for publisher growth.

Kochava | **Senior Data Analyst, Adtech**

Full-Time, Sandpoint, ID   Jan. 2019 - Jan. 2023

- Developed a proprietary attribution modeling system processing data from $3.5B in annual ad spend, using time-delta analysis to distinguish between organic and ad-driven conversions.
- Created advanced fraud detection algorithms analyzing temporal patterns in user behavior, identifying non-human traffic patterns that would have otherwise inflated campaign performance metrics.
- Engineered a lifetime value calculation framework connecting ad impressions to in-app purchase events, enabling advertisers to optimize campaigns based on true incremental ROI rather than click-through rates.
- Directed incremental lift studies and control group analytics for flagship clients, including TikTok, utilizing advanced data analysis to reveal strategic insights for ad optimization and communicating complex findings to C-suite and non-technical stakeholders, influencing data-driven decision-making.

Projects Experience

Polling subreddit posts with Airflow, Google Cloud Function, and BigQuery   **Jan. 2023**

- Built an automation system that polls a subreddit via REST API, dedupes the posts in a database, and streams the data to a pub/sub topic. I also wrote a consumer client that consumes from the topic and relays updates to its downstream clients.

Predicting Lichess chess winners with RandomForest and XGBoost   **Jan. 2020**

- Predicted chess winner based on a series of first opening moves and associated features, and then determined the best openings to play - which matches current chess literature.

Interaction factoring and age prediction of drug users in Colorado   **Dec. 2019**

- Developed R plots that showed interactions of various narcotics and narcotics found within overdose victims and determined likelihood of death based on demographic group via clustering algorithm.

Publications

App Developer Magazine, "New Lookback Attribution Windows for SAN Networks."   **Feb. 2022**

Certifications

- **Data Engineering, Big Data, and Machine Learning on GCP Specialization**
- **AWS Cloud DevOps Certification**
- **AWS Certified Machine Learning**

Technical Skills

Languages: Python, R, Scala, SAS, Matlab
Databases: Snowflake, Redshift, PostgreSQL, MySQL, NoSQL, DynamoDB, Aurora, RDS
Data Engineering Tools: Airflow, dbt, Git, GitHub, Snowpipe, Docker, Kubernetes, Databricks
Data Visualization: Tableau, Looker, Power BI, Matplotlib, Seaborn
Cloud and DevOps: GCP (Google Cloud Platform), AWS (Amazon Web Services), Microsoft Azure, Databricks, Docker, Kubernetes, SageMaker
Frameworks & Libraries: Pandas, NumPy, Sci-kit learn (SciPy), TensorFlow, PyTorch, XGBoost, LightGBM, CatBoost, Matplotlib, Seaborn
Big Data Technologies: Apache Spark, PySpark, Hadoop, Hive, Pig
Cloud Services: Pub/Sub, Datastream, Vertex AI, Google Kubernetes Engine (GKE), Snowpipe, AWS Batch, AWS Data Pipeline, EC2, Kinesis, Elasticache, SageMaker, Docker

Keywords

BigQuery, Apache, Business Intelligence, BI, ETL, Pipeline, Statistics, Clustering, Backend, Back End, DevOps, Information Security, API, Project Management, Performance Optimization, Problem Solving, Attention to Detail, Advanced Analytics,

Data Architecture, Cloud Computing, Machine Learning, Predictive Analytics, Data Science, Campaign Optimization, AWS, Azure, Power BI, PowerBI, DynamoDB, Dynamo DB, Decision Making, Business Objectives, Cloud Infrastructure, Infrastructure