

Hamza Musse

Education

Master of Public Policy, Quantitative Analysis - *University of Minnesota*
Bachelor of Arts, Political Science - *University of Minnesota*

Sept. 2015—Sept. 2017
Sept. 2012 – Aug. 2016

Skill Profile

Python - 2 years; Data Structures and Algorithms, ETL, Pandas, Numpy, Plotly, Big Data Analytics, PySpark, Machine Learning
R Programming - 1 year; ETL, statistical analysis, Data.Table, DPLYR, Ggplot2, data visualization, Markdown
SQL - 3 years; MySQL, PostGreSQL, NoSQL (Cassandra, MongoDB), Data Warehousing, Data Modeling, , Spark, Airflow
Cloud Engineering - 2 years; AWS (Glue, Lambda, S3, RDS, Redshift), GCP (BigQuery), Snowflake, MongoDB Atlas, Databricks

Software Projects

COVID-19 Data Pipeline and Visualization 07.2020 - 07.2020

- Created ETL pipeline from Johns Hopkins MongoDB Cloud database to extract COVID-19 deaths for each US state.
- Applied principles of functional programming to clean, transform, and visualize heat map of deaths over time.
- Tools included Python, MongoDB, Shell scripting, Git, and Heroku. View at <https://hamza-musse-portfolio.herokuapp.com/>

AWS Health Data Analytics Pipeline 09.2018 - 11.2018

- Configured cloud data warehouse with AWS RDS to ingest millions of health records from disparate sources (CVS, APIs, etc.)
- Transformed data with advanced SQL and Python methods to create a unified dataset available to internal data users.
- Uncovered \$900M of excess costs for heart disease treatment in Maryland by combining clinical data with ACS variables.
- Utilized Python ETL, AWS, SQL, SFTP, Tableau, and client management to identify business objectives and data needs.

AWS Serverless ETL Pipeline and Data Lake 09.2019 - 10.2019

- Scraped housing price data from Craigslist and stored it in S3 using Python, AWS Lambda, and AWS Glue.
- Query json data stored in S3 using standard SQL via AWS Athena and connected Athena to Power BI.
- Tools included Python, AWS Lambda, Linux Command Line, Git, AWS Glue, AWS S3, AWS Athena, and Microsoft Power BI.

Work Experience

ASR Analytics - Washington, DC Data Engineer 10.2019 - Present

- Created R data pipeline that automatically pulls data from diverse IRS data sets and merges into a single source of truth.
- Organized meetings with frontend developers to translate dashboard design into upstream data wrangling tasks.
- Reduced coding time by creating R functions that auto-generate and run common SQL queries.
- Created detailed documentation of the data pipeline using Gitlab issue tracking and Git for version control.

Socially Determined Inc - Washington, DC Business Intelligence Engineer 05.2018 - 10.2019

- Organized regular meetings with external clients to identify financial opportunity in their customer usage patterns.
- Provided rapid turnaround Power BI dashboards with customer demographics and revenue.
- Provided the business intelligence and machine learning team with data best-practices to maintain quality and consistency.
- Led conversations with CTO and Product team on automation opportunities using tools such as AWS Glue and Lambda.

MN Department of Health - Saint Paul, MN Data Analyst 09.2017 - 05.2018

- Transformed data using R and Excel, and created over 40 visuals for PowerPoint presentation.

Skills & Competencies

Communication - Adapt communication style for diverse stakeholders to translate business objectives into an engineering strategy.
Analytical Rigor - I seek critique and question my assumptions, leaving no rock unturned; trust but verify, especially with data!
Curiosity - Data technology is rapidly changing and learning (and teaching) new Big Data tools is an exciting responsibility.