Jeff Brennan

jeffbrennan10@gmail.com | jeffbrennan.dev

SUMMARY

Careful preparation of public health data improves the lives of underserved communities. I have five years of experience working at every level of the healthcare data lifecycle, and would like to apply what I have learned in a collaborative team that uses modern approaches to tackle difficult problems.

EXPERIENCE

Data Engineer – Medisolv   **Jul 2023 – Present**
- Oversee the ingestion and transformation of patient data (10B+ records/week, 150TB+ data lake) for hundreds of hospital clients
- Created a Python CLI wrapper around Databricks Asset Bundles to programmatically generate Databricks Workflows, enabling a 20x runtime improvement and $100k+ yearly savings through task-specific cluster tuning
- Designed our central git repository to store Databricks jobs, tests, and utilities. Trained 12 analysts and engineers on software development best practices enabling faster iteration, regression testing, and improved process visibility for management.
- Led the migration of 1,500+ Azure Data Factory (ADF) pipelines to a source-controlled IaC implementation. Developed tests to validate our pipelines before they were published to production, reducing our error rate by 34%
- Developed a Spark metrics parser using Plotly Dash to guide data-driven transformation rewrites, saving thousands in wasted compute time and storage access costs.

Data Analyst – NewYork-Presbyterian   **Dec 2020 – Jul 2023**
- Managed the calculation, tracking, and reporting of quality metrics, leading to $20M+ in savings
- Built and maintained weekly analytics ELT pipeline (100M+ records for 35k Medicare patients)
- Led the adoption of geographic analysis by creating custom address cleaning and geolocation workflow
- Developed data cleaning helper functions in Python and R used by 9 other analysts
- Reduced Tableau loading times from minutes to seconds by designing composable data models

Data Engineer (contract) – UTHealth   **May 2020 – Sep 2023**
- Created and maintained daily Texas COVID-19 data pipeline from state and third-party sources
- Web scraped data with Python (REST apis, beautiful soup, selenium)
- Developed a monitoring Slack bot and unit tests to ensure consistent data quality

Research Coordinator I – Baylor College of Medicine   **Jan 2019 – Dec 2020**
- Authored thesis (100+ citations) on biomarkers of traumatic brain injury (TBI) and provided data support for other research efforts
- Applied variable selection on hundreds of biomarker combinations to identify TBI predictors
- Built analysis pipelines and created publication-ready data visualization in R

EDUCATION

UTHealth Houston
Master of Science - Major in Epidemiology, Minor in Biostatistics   **May 2020**
Certificate: Data Science

The University of Texas at Austin
Bachelor of Science - Major in Public Health   **May 2018**

SKILLS

- **Languages**: Python (advanced), SQL (advanced), R (advanced), Scala (intermediate), Golang (learning)
- **Data Stack**: Transformation (pyspark, dbt, polars); Orchestration (azure data factory, airflow, dagster); Validation (pytest, pydantic); Distributed Computing (spark, docker, kubernetes) Databases (sql server, postgres, mongodb); Cloud Platforms (azure, aws, databricks)
- **Certificates**: Epic Data Model, AZ-900, Databricks Data Engineer Professional (in progress)