Ben Bubnick

Advisory Data Scientist · Data Integration Specialist

3419 Superior Park Drive, Cleveland Heights, Ohio, 44118, U.S.

(+1) 216-556-1612 | ben.bubnick@gmail.com | benbubnick

"I want to use my 15+ years experience to reimagine healthcare by creating, assembling, and improving our business relationships through data science. I believe that when I inspire people to do the things that inspire them, we can improve the relationship between caregiver and patient."

Summary

SME/Utility Roles	AI & Machine Learning, Data Curation/Wrangling & Integration, Training & Mentoring
Big Data Platform	Hadoop & Azure, transformation with Spark/Hive/Impala/Pig/kafka/Ruby/Python/R
Health Care Analytics	HEDIS Measures & Tailored Analytics from EMR, HL7(C-CDA, CCD, FHIR), 837, CMS, etc.
Relational and NoSQL DB	ETL & EDW design with SSIS/SQL/T-SQL/PostgreSQL, NoSQL, DB2, IMS in CI/CD pipeline
Demonstrated Past Success	Band Promotion, Lab Services Award, Eminence and Excellence Award, Honey Badger

Experience

Data Scientist

Merative

Clinical Operations

Jul. 2022 - Present

Technical lead on the delivery of large & complex AI and Cognitive led transformation programs and works with Merative & Partners to optimize tools and delivery methods. Identifies, analyzes, and extracts key data from heterogeneous systems; and leads technical requirements gathering, data discovery, documentation, and implementation of data projects. Works closely with senior leaders and colleagues to identify opportunities, set objectives, and formulate analytic strategies that result in measurable impact.

Advisory Data Scientist

IBM

Provider Analytics

Nov. 2021 - Jul. 2022

Tools: Python (NumPy, Pandas, SciPy, scikit-learn, matplotlib), GoLang, Azure, Docker, K8s, C#, SQL/PostgreSQL, Tableau

Technical lead on delivery of Cognitive, Automation & AI led programs using diverse data sources, such as electronic health record data, insurance claims data, scheduling and financial data, social determinants of health data, and any other relevant sources. Integration of disparate and centralized Provider Analytics data into data marts, data mining, statistical model development, advanced analytic methods. Manages client resources and drives technical solutions to timely completion. Engages with key clients & stakeholders to effectively communicate value proposition on analytics led insights and embed AI solution in client processes.

- Created machine learning Covid-19 tool that reduced need for hospitalization identification time by up to two weeks
- Led HIPAA patient deidentification project from 200+ million individual patients for content analytics and measures
- Led delivery on tailored analytics and reporting project for client administrative data
- Created testing framework for ETL & implementation processes on new data platform
- Developed data migration validation analysis on 150+ billion patient records
- Developed surgical smart case card using combined CNN/RNN model
- Awarded Data Science Profession Certificate

Senior Data Curation Architect

IBM

Provider Analytics

Dec. 2020 - Nov. 2021

Tools: Gradle, Python/R, Hadoop/HDFS, Hive/Impala, Spark, kafka, Flink, Docker, K8s, Elasticsearch, Kibana, SQL/SSIS/Oracle

Senior member of Data Onboarding & Governance for integration of healthcare data and enrichment of metadata. Configuration of data models, analytics, reports, and visualizations. SME to project data architects, delivery, analytics, integration, etc. Directed the configuration of the data onboarding platform and associated infrastructure, and developed improvements to advance stakeholder's business priorities using tailored analytics.

- Developed large clinical data integration of 100+ billion structured and semi-structured records for Life Sciences analytics projects
- Led HIPAA patient deidentification project from more than 100 data sources for content analytics and measures
- Led cross-function technical team as clinical data analyst for 2020/2021 measure updates project
- Developed topic modelling project on nearly 6 million semi-structured provider notes data
- Curated 25 pages of AI workflow documentation and mentored new hires in AI tuning process
- Created AI Surgical Scheduling Optimization tool currently evolving into new offering
- Closed documentation gaps, organizing more than 100 past projects
- Advised on development project for new data platform

Data Curation Architect

IBM

Watson Health

Nov. 2017 - Nov. 2020

Tools: Ruby, Pig, Java, Gradle, Perl, Python/R, Hadoop/HDFS, SQL/SSIS/Oracle, NoSQL, Hive/Impala, Spark, kafka, Elasticsearch, Kibana

Lead member of Data Onboarding & Governance, member of the Watson Foundation for Health Solutions Architecture team. Built the data onboarding platform and associated infrastructure, and developed improvements to reducing customer implementation cycle time and data triage activities. Reported on the "health of the data" during data onboarding. Remediation of data issues during UAT. Tuned machine learning classificiation algorithms for patient matching, and led the internal review of similar work by my peers.

- Managed and developed ambitious multi-org enterprise with over 70 interconnected, project-sized components
- Spearheaded new-hire mentoring program that led to ≈4,500 increase in total billable hours
- Developed new data validation and review process that reduced total rework by 25%
- Reduced mapping rework by 45% with through new knowledge-driven initiatives
- Awarded Lab Services Award in 2018

Software Developer

IBM

EPM Application Suite

Sep. 2016 - Sep. 2017

Tools: Ruby, Pig, Java, Perl, Jenkins, Hadoop/HDFS/CDH, SQL/mySQL/Oracle, DB2, IMS, PL/SQL

Technical delivery lead for client integrations which has included the ordering, installation, and configuration of required hardware, including VPN devices to support client integrations. Configuration and development of software necessary to support client integrations, including ETL and Data Validation.

- Reduced time spent on knowledge transfer by ≈80% by reorganizing project documentation
- Eliminated need for a $30,000 annual contract by developing data curation tool
- Reduced cost-per-hire by $5,210 by re-developing education program
- Eliminated average of 3 weeks from employee onboarding process
- Awarded Eminence and Excellence Award in 2017

Data Scientist

IBM

EPM Application Suite

Nov. 2015 - Sep. 2016

Tools: Ruby, Pig, Hadoop/HDFS/CDH, Shell, SQL/mySQL/Oracle, Jenkins, Java

Data Ingestions, pre-analytic enrichments, analytic augmentation, and data governance. Identified and extracted key data from multiple heterogeneous systems; and led technical requirements gathering, data discovery, documentation, and implementation of assigned client integrations. Configuration and development of software necessary to support client integrations.

- Acted as technical expert point of contact for customers in order to identify & resolve issues
- Led team of 12 remote data science contractors for 36 integration projects throughout 2015 - 2017
- Created over 125 new knowledge base articles, making up 30% of the team's online content
- Developed automated data integration tool to cut out 28% unnecessary development time
- Principal integration specialist on over 60 projects for 20 different clients

Data Scientist

Explorys

METL

Jun. 2015 - Nov. 2015

Tools: Ruby, Pig, Hadoop/HDFS, React, NodeJS, JavaScript, Shell, SQL, Oracle, Jenkins

Identified clinical, financial, and operational data elements within different health provider systems to develop data transformations and providing meaning to the data. Performed data ingestions, pre-analytic enrichments, analytic augmentation, and data governance.

- Developed cron-based widget to define blackout periods for client data pulls
- Performed troubleshooting on production issues across multiple environments
- Principal integration specialist for 10+ client database warehouses

Software Engineer

AmTrust Financial Services

Commercial Package Policies

Mar. 2014 - Jun. 2015

Tools: VB.NET, ASP.NET, C#, CSS, Javascript, HTML5, SQL, PL/SQL

Developed web forms for insurance using ISO standards and created multi-tier MVC components, Web APIs, and UIs; Maintained and improved existing codebase, and intermediary between business analysts and development team when gathering requirements and reporting progress.

- Development on two software architecture projects in 2015 and principal developer on offering product
- Analyzed page timings on 65+ ISO rating based web forms with R for optimization efforts
- Steered development group into a unit testing project on multi-tier MVC components

Color Scientist — *PPG*

Automotive Refinish — *Apr. 2012 - Jun. 2013*

Tools: VB.NET, C#, SQL, C++, FORTRAN

Developed software and radiative transfer models on complex paint mixtures to determine formulation methods and for spectroscopic profile analysis on color match samples.

- Improved color matching classification algorithm performance by 40% through new numerical methods
- Developed cross-instrument correlation software for risk analysis project
- Awarded Color Lab Automation Award in 2012

Adjunct Professor — *various colleges*

Lorain County Community College, Miami University, Cincinnati State Technical and Community College — *Aug. 2010 - May. 2019*

Developed curricula in introductory physics & astronomy courses, and technical presentations of complex physical concepts for major, non-major, and remedial students. Wrote concise laboratory instruction manuals, trained in precise measurement techniques & data analysis using DataStudio, and coordinated analysis documentation.

- Tracked a 28% reduction in student confusion by reorganizing online education materials
- Exceeded feedback evaluation goals by an average of 16% in 2017 and by 17% in 2018
- Recorded over 700 hours of lecture material for convenient consumption

Research Assistant — *University of Cincinnati*

Department of Physics — *Sep. 2006 - Sep. 2010*

Gathered data using SpeX, a 0.8–5.4 μm spectrograph, at the NASA Infrared Telescope Facility. Observations in the short-wavelength, crossdispersed mode SXD. Multiple analysis projects comparing 2MASS with measured data, statistical classification without independent photometry for cluster properties, and photometric analysis bootstrapped to independent measurements.

- Volunteered 120 hours of Quality Control for the Cambridge Astronomical Survey Unit Variable Stars database
- Published 3 scholarly articles on near-infrared astrophysics over a 24-month period
- Awarded Mary J. Hanna Research Fellowship in 2010

Awards

2018 - 2022	**Gold Champion**, For 1,477 classroom hours and earning 43 digital credentials (4 consecutive years)	*IBM*
2018	**Lab Services Award**, For Coordinating Internal Teams to deliver client needs on time	*IBM*
2017	**Eminence and Excellence Award**, For developing and implementing a new data ingestion & curation tool	*IBM*
2012	**Color Lab Automation Award**, For expanding data collection and analysis of pigments & paints	*PPG*

Talks

AI Hackathon - Surgery Scheduling Optimization — *IBM Provider Analytics*

Data Scientist — *Dec. 2020*

Estimating operative case-time duration accurately is crucial for operating room scheduling optimization. Our goal was to create statistical models on a large retrospective dataset to improve case-time duration estimation in comparison to current standards. A unique algorithm was designed to accurately anticipate surgery duration using a three-stage process that first uses previous utilization data and current waiting list information to manage case mix distribution.

IBM Watson Care Insights Reach Initiative — *Watson Health*

Trainer — *Apr. - Sep. 2019*

Watson Care Insights aims to analyze longitudinal patient records and real-time data points through HL7 feeds to provide practitioners insights into the patient's data, and address these challenges by integrating within the physician's workflow in the electronic health record system.

Content & Data Analytics Learning Exchange — *Watson Health*

Peer Presenter — *Apr. - May. 2018*

Introduction of implementation of machine learning algorithms and their applications using real-world health data analytics examples.

2018 IRI Fall Networks Conference — *Innovation Research Interchange*

Presenter for the Information Services/Information Technology Network Talks — *Sep. 2018*

Cognitive computing can use NLP to extract insights from the unstructured data that holds most of the relevant information, but also it can also perform the necessary transformation accurately at the earliest stages of the data integration pipeline.

Cross Team Training Series — *Explorys*

Lead Presenter for the METL Group — *Sep. 2016*

Seven part series in data ingestion approach using the Hadoop software framework. Data can be aggregated much more quickly and cost-effectively using a data lake framework than in traditional data warehouses, because of the speed and low-cost of massively parallel computing.

Writing

Surgical scheduling optimization and procedure duration prediction[a] — *IBM*

Technical Development Team — *Dec. 2020*

Improvement on traditional fixed ratio methods for total procedure time estimation, using linear regression models based on relevant variables.

Process and Requirements for CCD Data Integration — *IBM*

Lead Integration Developer — *Oct. 2017*

Technical white paper on processes and requirements, overcoming the many difficulties with flat file data curation of CCDs.

AI Paint Formula Prediction from Spectroscopic Measurements[b] — *shinyapps.io*

Student Developer — *May. 2015*

Naive bayesian color and effects matching algorithm for the formula of complex automotive paint formula.

Characterization of Surface Coatings using Radiative Transfer Models for Color Matching — *PPG*

Radiative Transfer Physicist — *Dec. 2010*

Approximation of Chandrasekhar radiative transfer model for atmospheric scattering is applied to color matching. (U.S Patent Application)

Massive Stellar Clusters in the Disk of the Milky Way Galaxy[c] — *OhioLINK ETD*

Graduate Research Assistant — *Dec. 2010*

Using spectroscopic and photometric analysis of two open clusters to determine the structure of the plane of the Milky Way Galaxy.

Near-Infrared Spectroscopy of Candidates Members of the Galactic Cluster [BDS2003] 107 — *PASP*

Undergraduate Research Assistant — *Feb. 2008*

New near-infrared classification spectra for nine candidate members and comparative 2MASS photometry for several cluster members.

A Novel Method for Low-Toxic Photo-Etch Printing[d] — *MAPC*

Provost Undergraduate Researcher — *May. 2003*

Aquatint etching techniques, which have long been employed in intaglio printmaking, may be applied to a novel method for photo-etch procedures.

Volunteer

After School Computer Science Supplemental Education — *CSforAll/CMSD*

Instructor — *Oct. 2018 - May. 2019*

Our goal was to provide computer science instruction in all of the District's high schools, with a special emphasis on students who do not have access to a computer or a stable Internet connection at home.

Data Science High School Competition — *HIT in the CLE*

Coach — *Mar. 2017 - Mar. 2019*

Teams of five students work together to solve a challenge using a large data collection. The purpose is to educate students understand the potential of a career in health information technology and to assist IT companies in Northeast Ohio in developing a talent pipeline.

RTAHeatMap Civic Hacking Project — *Code for America*

ETL Guild — *Jan. 2015 - Jun. 2015*

As IT technology evolves, the number and quality of data available to planners and managers changes. The initiative was brought about to apply data science tools to improve and expand public record access.

After School Education Program — *T.C.P. World Academy*

Tutor — *Sep. 2008 - May. 2009*

Our goal was to teach students how to understand the value of their learning experience and how to engage in a global society. Students at T.C.P. World Academy become academically involved and independent learners by participating in society simulation activities for higher learning.

Education

2022 **Certification**, Data Science[a] — *IBM*

2015 **Certification**, Data Science[b] — *Bloomberg School of Public Health*

2010 **Master of Science**, Astophysics[c] — *University of Cincinnati*

2004 **Bachelor of Fine Arts**, Printmaking[d] — *Ohio University*