

Emmett Storts

Phone: (414) 218-9449 Email: emmett.storts@protonmail.com

LinkedIn: www.linkedin.com/in/emmett-storts/ GitHub: <https://github.com/erstorts>

Senior Data Scientist with eight years of experience spanning the full data stack, from Extract Transform Load (ETL) pipelines and deep-dive causal inference analytics to production Machine Learning and the tools that turn analytics questions into answers decision-makers can act on. Comfortable owning a data function end to end, having done exactly that on small teams at Product Hunt, across freelance engagements, and most recently at The Daily Wire. An applied-economics background keeps the focus on what is actually driving outcomes, always hunting for the specific lever that moves retention, not just predicting it. Paired with a builder's habit of shipping tools teams use rather than analyses that sit in a deck.

Experience

Data Scientist

October 2025 - May 2026

The Daily Wire | Nashville, TN

- Designed and deployed a complete overhaul of a core data model in dbt and Snowflake which was then viewed in a daily dashboard by the entire executive team.
- Built and deployed 4 production models (regressions, clustering, natural language processing (NLP), and random forests), and moved the team's modeling off local machines into a monitored CI/CD pipeline using dbt, Docker, and Github Actions so anyone could ship and retrain reliably.
- Built and maintained an end-to-end YouTube network analysis Streamlit app for the production team to better understand the YouTube algorithm and help with programming decisions.
- Designed and deployed a Lifetime Value (LTV) data model which was then used by marketing for balancing win-back spending and finance for financial projections.
- Developed a Streamlit application for on-platform comment insights, integrating AI summarization and vector embeddings which was used by the production team to find comments for audience engagement.

Founder & Data Scientist

January 2024 - October 2025

Knowledge Uncovered | Remote

- Medimap - Lead Data Scientist
 - Point person for all data requests; Constructed 10+ Power BI dashboards for self service analytics.
 - Built automated retention reports and dashboards for client subscription services, eliminating manual reporting entirely and improving visibility into churn trends.
 - Migrated from Fivetran to Hevo to save 50% on data engineering costs.
 - Generated 50k pieces of SEO content with LLMs (OpenAI API) leading to a 5% increase in sitewide traffic.
 - Applied clustering and PCA to build audience segmentation models for SEO and content personalization.
- Charter Communications - Insights Engineer
 - Built and Managed 5 Tableau Dashboards, including 1 specifically for upper management.
 - Handled 20+ Ad-hoc requests. Ensured data quality and accuracy through robust validation processes across all requests and dashboards.

Data Analyst

April 2022 - October 2023

Product Hunt | Remote

- Owned the full analytics function on a two-person team, from ad-hoc requests to experimentation, partnering directly with non-technical stakeholders.
- Applied causal inference and econometric methods (e.g., diff-in-diff, propensity scoring) in SQL and Python to evaluate product and marketing initiatives.
- Collaborated with Marketing on newsletter attribution resulting in a self-service dashboard for non-technical stakeholders.
- Led 23+ A/B experiments in collaboration with Leadership, Design, Product, and Engineering teams which led to product redesign that increased user engagement by 18%.
- Developed spam detection models which reduced Customer Experience time to action by 25%.
- Built ad impression forecasting models for the Sales team to optimize outreach strategy and accurately project pipeline revenue.

Senior Analytics Specialist

June 2020 - April 2022

Analytics Specialist

June 2018 - June 2020

WEC Energy Group | Milwaukee, WI

- Led 4 Machine Learning projects:
 - Anomaly Detection of Failed Meters using deep learning & neural networks, allowing for early intervention.
 - Predictive models for customer retention and payment agreements using gradient boosting and random forests which increased customer satisfaction by 10%.
 - Electric Outage Categorization using natural language processing and random forests to eliminate 5 hours of work across 3 people.
- Managed 3 Student Interns and 2 Data Science Outreach Projects.
- Provided internal support via 10+ data-centric projects, applying econometric and causal inference to evaluate program effectiveness and inform operational decisions.
- Optimized customer call center costs through a mix of full-time and seasonal hiring.
- Used SQL and Python to evaluate a company wide program, concluded it had no statistically significant impact on customer satisfaction (the primary metric for the program) which led to the end of the program.
- Reduced manual metric reporting by 90% for the largest IT project in company history (Customer Service 2022) using Python & Power BI.

Projects

Library Retrieval Augmented Generation (RAG) Search September 2025

MVP App: <https://library-ai-chat.streamlit.app/>

Github: <https://github.com/erstorts/destiny-rag>

- Built a retrieval-augmented generation system enabling middle school students to search a library catalog using natural language queries.
- Indexed 10,000+ books using OpenAI's text-embedding-3-large model and Pinecone vector database, with the OpenAI API handling query generation.
- Demoed to school staff with full build out in the works.

Local Catholic Business Directory Streamlit Dashboard

April 2025

Dashboard: <https://co-catholic-business-directory-report.streamlit.app/>

Github: <https://github.com/erstorts/co-catholic-business-directory-data>

- Designed and deployed a Streamlit dashboard serving a local church networking group.
- Featuring 100+ business listings with data ingestion, cleaning, and an interactive directory interface for end users.
- Stakeholders used this dashboard at monthly team meetings.

Skills

Languages & Tools: Python, R, SQL, dbt, Git/GitHub, Docker, scikit-learn, XGBoost
Machine Learning & Statistics: Causal Inference (diff-in-diff, propensity scoring), Gradient Boosting, Random Forests, Regression, Time Series, Clustering, Neural Networks, A/B Tests
GenAI: LLMs, RAG, Vector Embeddings & Search
Platforms: AWS, GCP, Snowflake, Streamlit
Visualization: Power BI, Tableau
Practices: MLOps, CI/CD, Non-technical Communication, Experimentation

Education

AI Bootcamp

April 2025

DeepAtlas | Remote

- Machine Learning theory, training, and deployment
- Relevant Activities: Lip Reading Model

Master of Arts in Applied Economics

December 2017

The University of North Carolina at Greensboro (UNCG) | Greensboro, NC

- Concentration: Policy Analysis and Quantitative Methods
- Relevant Activities: Graduate Research Expo Participant, Graduate Student Association Member

Bachelor of Business Administration in Economics and Entrepreneurship

May 2016

The University of Wisconsin at Whitewater (UWW) | Whitewater, WI

- Double Major: Economics and Entrepreneurship
- Relevant Activities: Economics Society President, Fed Challenge Member, Economic and Entrepreneurial Independent Studies, Entrepreneurial travel study in Scotland