Sydney Casey

614-506-0610 / sydneycasey124@gmail.com / Linkedin / Github / Tableau / Portfolio / Columbus, OH

I bring a strong foundation in data science, mathematical modeling, and statistical analysis, with experience in designing predictive models and analyzing large datasets. My background has honed my ability to perform rigorous statistical analyses and communicate results effectively to diverse audiences. With a commitment to continuous learning and a passion for applying data-driven insights to real-world challenges.

EDUCATION

Master of Science in Health Data Science, University College London (UCL) Sept. 2023

GPA: 3.75 / 4.00 (UK Distinction)

Relevant coursework: machine learning for healthcare & biomedicine, advance machine learning for healthcare, regression modeling, advanced statistics for records research, data methods for health research, programming with python for health research

Bachelor of Arts in Biochemistry, Mount Holyoke College May 2016

Cum laude, Mount Holyoke Leadership Award

KEY COMPETENCIES

- **Programming languages**: Python, SQL, STATA, R
- **Data Science & Machine Learning**: Mathematical modeling, Model Evaluation, Hypothesis testing, Data science pipeline (cleansing, wrangling, visualization, modeling, interpretation), Python (eg. scikit-learn, numpy, pandas, matplotlib)
- **Data Visualization & Analysis Tools**: Tableau, Power BI, Matplotlib, Seaborn, JupyterLab, Streamlit, R Shiny
- **Additional Tools & Technologies:** Microsoft Office, Excel, MySQL, PostgreSQL, Docker, Survival Analysis, Statistics, Experimental Design, Apache Airflow, API integration
- **Technical Reporting:** Strong skills in scientific reporting and presenting data to diverse audiences

SELECTED PROJECTS

End-to-End Unsupervised ML Pipeline for Pharmaceutical Adverse Events Analysis | Personal April 2025 -

- Developed a scalable, containerized data analytics pipeline for FDA drug adverse event data using unsupervised machine learning:
  - Built ETL workflow with Docker, Apache Airflow, and OpenFDA API
  - Designed object-oriented framework to process structured and unstructured data
  - Applied scikit-learn clustering to structured adverse event data, leveraging spaCy to preprocess coded text fields for pattern discovery
  - Implemented PostgreSQL persistence layer; designed for AWS/GCP deployment

PubMed Insight Explorer – NLP-Enhanced Literature Analysis Tool | Personal 2024 -

- Developed an interactive web application that interfaces with the PubMed API to retrieve and organize biomedical literature based on user-defined keywords.
- Engineered a data pipeline to extract, clean, and structure metadata (e.g., publication date, journal, authors) into downloadable CSV files.
- Designed a keyword co-occurrence analysis module using NetworkX and Plotly to visualize semantic relationships across retrieved abstracts.
- Planning and prototyping NLP-based features, including text summarization, topic modeling, sentiment analysis, and bias detection to enrich article insights

Predicting Length of Critical Care Days in Welsh Intensive Care | UCL June 2023 - Sept. 2023

- Developed and optimized a two-stage model by integrating a Random Forest classifier with an XGBoost regressor for continuous prediction of critical care days from using a novel dataset of electronic health records.
- Achieved a mean absolute error (MAE) of 1.732 with a sensitivity of 0.892 for values less than nine days, utilizing Shapley Additive Explanations (SHAP) to identify the most influential features and enhance model transparency.

PROFESSIONAL EXPERIENCE

Research Assistant

Nationwide Children's Hospital, Columbus, OH Sept. 2020 - Aug. 2021

- Collected and interpreted experimental data, providing insights that influenced key decisions in ongoing research projects and clinical studies.
- Delivered weekly research presentations to peers and senior researchers, effectively communicating complex data analyses to guide experimental adjustments and future research directions.

Postbaccalaureate Award Fellow

National Institutes of Health, Bethesda, MD April 2017 - June 2019

- Compiled and categorized a published comprehensive review of neuronal ceroid lipofuscinoses (NCLs), establishing it as a valuable resource for researchers and clinicians, with over 90 citations.
- Spearheaded individual research projects targeting diverse molecular pathways in infantile NCL, amplifying the knowledge base within the field and catalyzing advancements in treatment strategies.

PROFESSIONAL DEVELOPMENT

- Datacamp Machine Learning Engineer Career Path
- Datacamp Deep Learning Skill Path
- Datacamp Associate AI Engineer for Data Scientists Career Path
- DataExpert-io Data Engineering Bootcamp
- Imperial College - Mathematics for Machine Learning (Coursera)