Samuel Harris

Washington, DC | harrissamuel130@gmail.com | 412-527-9194
https://www.linkedin.com/in/samuel-harris-sh/ | https://github.com/harrissamuel/projects

EDUCATION

The George Washington University **Washington, DC**
B.S. in Data Science and Mathematics | Minor in Statistics **May 2025**

- GPA: 3.94 / 4.0 | Dean's List Every Semester
- Relevant Coursework: Data Mining, Data Visualization, Data Warehousing, Multivariable Calculus, Linear Algebra, Mathematical Statistics, Real Analysis, Engineering Computations, Bayesian Statistics and Modern Physics

TECHNICAL SKILLS

Languages: Python | R/RStudio | SQL and NoSQL (MongoDB and Neo4j) | Power BI | SAS
Packages/Libraries: Pandas, NumPy, Plotly, Scikit-Learn, Matplotlib, Seaborn, Selenium, JAGS, ggplot2, tidyr, dplyr
Skills: Supervised Learning (Regression Models/Random Forests), Unsupervised Learning (Clustering), Data Mining, Data Visualization, Data Cleaning, Data Wrangling, Bayesian Modeling (MCMC Methods), Web Scraping

EXPERIENCE

Amtrak **Pittsburgh, PA**
Voluntary Safety Reporting Intern **May 2024 - August 2024**

- Analyzed safety incident data from frontline rail workers leveraging Power BI to create visualizations for the monthly Chief Safety Officer report and briefing. Produced safety metrics by department and type of incident
- Performed two data pulls isolating safety incidents of interest for legal and marketing purposes. Conducted an additional data pull to formulate an infographic design promoting the diversity of Amtrak's safety reporting culture

GW Data Science for Sustainable Development Hub **Washington DC**
Data Scientist: Research Team Head **January 2023 - Present**

- Research team head of GW's Data Science for Sustainable Development (DSSD) chapter. DSSD provides pro bono data science services to mission-driven organizations operating in the development and humanitarian sphere
- Led the development of a dashboard in Python (utilizing packages like Pandas, NumPy, and Plotly) for energy and water trends in hundreds of buildings in Washington DC for a project reviewing sustainability metrics by university

TECHNICAL PROJECTS

World Disasters in Context **Washington DC**
Author **January 2024 - May 2024**

- Placed over 15,000 world disaster events since 2000 in the context of development indicators by country for Capstone Project in Python. Developed a risk score methodology to map the most severe disaster types by country

Underrated Fantasy Football Players 2022 **Washington DC**
Creator **June 2023 - August 2023**

- Deployed linear regression with K-fold cross-validation to predict fantasy football points per game for wide receivers and running backs for a personal supervised learning project in Python using Pandas, Seaborn, and Scikit-Learn
- Determined the difference between predicted point values and real point values to rank fantasy players from overrated to underrated, thereby enhancing insights into making optimal player selections

Baseball Over the Last 40 Years **Washington DC**
Author **March 2023 - May 2023**

- Examined trends among baseball statistics like strikeouts and home runs over the last 40 years applying regression in R for a Data Visualization project. Project included interactive and animated visualizations using Plotly and ggplot2

AWARDS

- 2023 Outstanding Academic Achievement Award - awarded to the top 2% of students by GPA within each school with over 60 credits earned