Sam Deery-Schmitt

Data Scientist based in Seattle, WA

samdeeryschmitt@gmail.com **|** linkedin.com/in/samds **|** sampds.github.io

EDUCATION

University of Notre Dame **|** BS, Environmental Sciences, 2014
Syracuse University **|** MS, Applied Data Science, 2023

SKILLS

Programming: **SQL**, **Python, Spark** (via PySpark), **Apache Airflow** (via Python), **Git, R |** IDEs: **Databricks**, **MS SQL Server Management Studio**, **Anaconda**, **RStudio, Google Colab**, **Azure Data Studio|**
Software: **Azure Cloud Services**, **MS 365 Power Platform**, **Squarespace**, **Trello, Excel, SharePoint**

PERTINENT EXPERIENCE

Data Scientist, Axé Engineering (Contract). Albuquerque, New Mexico (Remote). Aug–Nov 2023

- Axé Engineering contracts with the US Department of Defense to manage critical construction and services projects at military bases, such as building physical and digital cybersecurity infrastructure
- Collaborated with stakeholders to gain working knowledge of domain in order to design, implement, and monitor technical solutions to business challenges, including automating core business processes
- Constructed ERD and logical model of database schema based on business rules, wrote SQL to build the schema in Azure SQL Server, and employed SQLAlchemy to migrate data from SharePoint
- Utilized SQL, Python, Azure, PowerPlatform, and Apache Airflow to automate data validation and ETL
- Built dashboards and PowerApps linked to queries, views and stored procedures providing descriptive analytics, while allowing user to utilize GUI to execute queries against the database

Data Science Trainee, RSG Media. New York, New York (Remote). Jun–Aug 2022

- Working in Spark via PySpark within the Databricks IDE, utilized predictive text mining to perform fuzzy matching on large (millions of rows, dozens of columns), disparate datasets in order to improve accuracy of metadata used in production by data science team
- Collaborated with management and colleagues to strategically improve efficacy of methodology

PROJECTS

Network TV Movie Trend Analysis

- Used R to aggregate and visualize Nielsen ratings from Q4 2016–Q4 2017 by age, sex, daypart, genre, network and network group to understand viewership demographics with high granularity
- Analysis parallels analytics performed in industry to hypertarget scheduling and advertising

Minimizing Labor Costs for a Real Estate Development and Home Building Firm

- Applied DMAIC process to quantify previously unmeasured inefficiency in supply chain—identifying $50,000 annual loss—and create granular view of issue using 7 quarters of internal and external data
- Worked with stakeholders to develop and implement process improvements to increase efficiency

Understanding Cryptocurrency Volatility: Exploring Correlations and Building Predictive Models

- Wrote R script to analyze popular cryptocurrency exchange data sets, discerning trends and building machine-learning models to forecast prices

Building Data Warehouse, ETL Pipeline, and Business Intelligence Dashboard
- Performed dimensional modeling to merge two databases
- Used Microsoft Visual Studio to create ETL pipeline for staging, transforming, and loading data into dimension and fact tables
- Created dashboard in PowerBI to provide visualizations and front-end queries of data warehouse

Database Development for Educational Consulting Company
- Mapped business rules to ERD and logical model, then wrote SQL to build tables
- Used R to implement ETL, and used MS Access to give users a GUI to perform CRUD operations
- Structured views and stored procedures to provide descriptive analytics for use in administration, sales, and marketing

USGS Earthquake Data Aggregation and Analysis
- Accessed USGS API through Python to pull .geojason data then parse and store it in MongoDB
- Created visualizations and dashboards to provide summary and analysis of earthquakes greater than magnitude 4.5 in the past 30 days

Sentiment Analysis and Communication of Insights: Twitter Sentiment Classification
- Built on script created for USGS project to perform ETL on tweets using the Twitter API, cleaning JSON with regular expressions and storing in MongoDB
- Performed sentiment analysis using VADER and TextBlob to discern public sentiment towards specific hashtags

Matching and Analyzing Records from Disparate Sources to Analyze College Football Coaches
- Scraped websites to collect records on team performance and coaches salaries
- Wrote fuzzy matching script to identify and correct mismatches, then combine data from multiple datasets
- Conducted in-depth exploratory analysis and made model to predict salaries of coaches