Sarah Langleben

| sarahlangleben@gmail.com | GitHub | LinkedIn

EDUCATION

Cornell University, Bachelor's of Science 2024. Double-major in Information Science and Economic Development.

TECHNICAL SKILLS

Languages: Python, SQL, Java, C++, OCaml, Ruby
Libraries/APIs: scikit-learn, pandas, spaCy, Selenium, BeautifulSoup, NLTK, PySpark
Tech Stack: Snowflake, Databricks, Palantir Foundry, Tableau, Airflow, PowerPoint, Excel, Kubernetes, Docker

EXPERIENCE

RBC Capital Markets, Data Engineer **Jul. 2024 – Present**

Lead for customer relationship management (CRM) projects. Collaborated with a cross-functional team of heads of business, cloud engineers, and project managers across five different offices in the USA, EU, UK and APAC to coordinate development and distribution of reports.

CRM Reporting

- Transformed individual instances of meeting data into a dashboard that shows month-by-month meetings, tracks client relationship degradation/growth over time, and compares meeting activity to previous years.
- Created Tableau dashboards for five industry-specific teams.
- Maintained 33 Tableau dashboards distributed weekly, reporting banker-client interactions.

Platform Decommission

- Saved team ~$450,000 / year on compute costs by migrating PySpark code from Palantir Foundry to Databricks.
- Built data processing pipeline using PySpark, Databricks, and Airflow.
- Authored documentation for extensive data pipeline on Palantir Foundry composed of legacy code written in PySpark.

RBC Capital Markets, Data Analyst **Jul. 2024 – Present**

Worked in a cross-functional team alongside data scientists and investment bankers. Provided bespoke data insights for banking teams and clients. Presented data analysis and findings to partnered banking team, explaining insights to a non-technical audience.

RBC Elements for Investment Banking, Uber Freight

- Consolidated geodata to report on shipping and freight activity across the United States and Canada.
- Developed Tableau dashboard to visualize freight activity, providing actionable insights for banking teams to assess client performance.

RBC Elements for Investment Banking, Maesa (Bain Capital Private Equity)

- Conducted SKU price analysis and web traffic assessment to understand customer demographics and preferences.
- Scraped and cleaned 200,000 online reviews and ratings for key products identified by SKU analysis using Selenium.
- Performed sentiment analysis on review data using spaCy. Calculated net sentiment for 5 brands and 15 unique products.

Banking, Cybersecurity Client:

- Reddit discussions collected with Selenium and analyzed using NLTK to gauge public perception and identify emerging trends.
- Used SQL queries on Snowflake to collect jobs posting data. Conducted time-series analyses on market demand for skill sets/roles within cybersecurity sector.
- Performed competitor analysis to benchmark client's offerings against industry peers.
- Analyzed user demographics to inform targeted marketing strategies and product development.

Cornell University Dept. of Information Science **Aug. 2022 – Dec. 2024**

Teaching Assistant. Led small-group instruction on fundamental data science concepts and data science libraries (NumPy, scikit-learn, pandas, and seaborn).

Cornell University Dept. of City and Regional Planning **Mar. 2023 – Dec 2024**

Research Assistant. Web scraped using BeautifulSoup4 to collect infrastructure and climate reports from government websites across the United States, China, and the European Union. Consolidated information to produce concise findings for future research publication.

PROJECTS

Data Science Final Project: Predicting Alcoholism in Portugal **December 2021**

Collaborated with a student group to develop a data science report. Conducted statistical analyses, authored documentation, summarized findings.

App Prototyping Final Project: Student Planner **May 2023**

Collaborated with a team of developers and designers to implement a mobile app prototype using Docker and Vue.js. Conducted user interviews, implemented state management using Vuex, and used Vue Router for seamless navigation.

PUBLICATIONS

Porciello, Jaron; Bourne, Thomas; Lipper, Leslie; Lin, Sammi; and Langleben, Sarah. 2021. *Mining the Gaps: Using Machine Learning to Map a Million Data Points on Agricultural Research from the Global South*. Colombo, Sri Lanka: Commission on Sustainable Agriculture Intensification.