ARIEL ASPER
DATA SCIENTIST

Cupertino, CA
408.422.6479
arielmasper@gmail.com
https://www.linkedin.com/in/ariel-asper
https://arielmas.github.io/

Results-driven Data Scientist with extensive experience in statistical analysis, machine learning, and data engineering. Passionate about automating workflows, optimizing data pipelines, and delivering actionable insights that drive business value. Adept at collaborating with cross-functional teams and effectively communicating complex analytical findings.

SKILLS

Python, R, SQL, Snowflake, PostgreSQL, Git/Github, DuckDB, Linux/MacOS, AWS SageMaker, Docker, NLP, ML Supervised/Unsupervised models, Hugging face, CI/CD, Github actions, sklearn, TensorFlow, Spacy, Statistical modeling, Exploratory Data Analysis, Streamlit, Data mining, Web scraping.

EXPERIENCE

Data Scientist | Graticule, Inc. *2024 – 2025*

- Led evaluation and selection of NLP tools, optimizing patient privacy workflows using Python, Docker, and DuckDB, leading to a ~60% reduction in manual effort, enhanced scalability, and increased revenue through new customer acquisition.
- Designed and implemented an ETL pipeline to automate raw data extraction, preprocessing (cleaning, QC, linking), and cohort generation, improving data consistency and reducing processing time by 50%, providing scalability for future projects, and increasing revenue.
- Developed Python functions and packages to automate tasks, saving ~30 hours per project and improving efficiency by 25%. This reduced operational costs, boosting profitability and scalability across projects.
- Developed data pipelines and analyzed 3M+ patient records using Snowflake, AWS Sagemaker, Python, and R. Delivered insights on high-risk patient populations and drug effectiveness, supporting FDA regulatory approval. This helped drive revenue growth by securing additional projects in the U.S. and Europe.
- Developed a demo using BERTopic to categorize clinical notes, with the potential to improve data quality, and accelerate QC processes which could be sold as services by the company - driving business value.

- Developed remote development workflows using virtual machines, Docker, ngrok, and SSH, improving team collaboration by 40% enabling efficient scaling of services for new clients, and ensuring cost-effective development.

Data Scientist | KSM Research and Innovation   *2023 – 2024*

- Led data analysis of 2M+ patients to assess healthcare resource utilization in clinical cohorts, in order to justify new drug development, improve patient outcomes, reduce healthcare system burden, and offer cost-saving opportunities for insurers while driving revenue growth for the pharma company.
- Created and tested federated learning algorithms to perform classic statistical models (e.g., Chi-square, Linear regression, Logistic regression, Kaplan-Meier Survival analysis) on multi-source healthcare data, ensuring privacy compliance and enhancing the scalability, accuracy, and representativeness of research findings.

Data analyst | Sheba Medical Center   *2021 – 2023*

- Analyzed 4+ million individuals using statistical models such as Cox and Logistic regressions, to assess medication-related outcomes, strengthening client trust and influencing pharma decisions on new drug development to improve patient well-being.
- Conducted cohort-based income analysis using statistical models such as Poisson regression Social Security data, informing policy decisions and optimizing financial support for underserved clinical groups.

EDUCATION

- **Data Science and Machine Learning Certificate |** Bar-Ilan University   *2022 – 2023*
- **M.A., Rehabilitation Psychology |** Ariel University   *2019 – 2021*
- **B.A., Behavioral Sciences** | Ariel University   *2016 – 2019*