Jake Haines

Princeton, NJ | +1 908 323 3690 | hello@jakehaines.net | linkedin.com/in/jhainesnc | jakehaines.net

Education

North Carolina State University   **Raleigh, NC**
Bachelor of Science in Statistics | **Minors:** Computer Science | **Activities:** Data Analytics Club   August 2021 – May 2025
Coursework: Python for Data Science, Applied Data Analysis, Probability, Data Structures, Statistical Computation & Data Management, Database Management Systems, Computer Vision, Natural Language Processing

Experience

Tesla   **Palo Alto, CA**
Data Engineering Intern   August 2022 – December 2022
- Optimized Airflow big data pipeline using, Python, Spark, SQL, Hadoop for data from 100k vehicles, reduced bottlenecks
- Built ETL pipeline for 10mil rows from 100k vehicles using PostgreSQL, Python, Pyspark, Spark, pandas, plotly to analyze sensor data
- Built dashboard with Python, plotly, Airflow, Tableau API, Postgres to visualize sensor data, increased firmware deployment rate

Tesla   **Fremont, CA**
Research Intern   May 2022 – August 2022
- Analyzed data using Python, pandas, numpy, plotly, scikit-learn to test vitals sensors, improved sensor accuracy
- Performed ETL on vitals sensor, CAN, geospatial, climate sensor data using Python, pandas, numpy, plotly, Mapbox to analyze data
- Designed and executed research studies using Python, pandas, scipy, scikit-learn, Sharepoint, Visio for 6 early-stage products' user experience with future implementation to over 250k production vehicles

Eagle Parts & Products   **Augusta, GA**
Data Engineering Intern   May 2021 – August 2021
- Built ETL pipeline using Python, pandas, regex on 15k rows of unstructured inventory data for parsing data
- Cleaned, parsed, and restructured data by automating data entry for data migration, reducing labor time required by 50%

Projects

Spotify API and Subaru ECU data streaming pipeline using Airflow and Postgres   **Princeton, NJ**
Personal   February 2024 – Present
- Built data pipeline using Spotify REST API, Python, Linux, and vehicle ECU data to aggregate real-time user streaming data and sensor data for analysis
- Analyzed data for driving and listening behavior using Spotify API, Python, plotly to add songs to queue based on driving data

Mobile productivity app   **London, UK**
Startup   April 2023 – June 2024
- Built cloud infrastructure using AWS, Google Firebase, PostgreSQL, IAM, RDS, EC2, Lambda, Firestore with 0 expense.
- Implemented GCP Firebase API using Node.js, XCode, React Native for Apple iOS user authentication
- Authored 10 research papers using clinical research and networking with international psychologists for building innovative features.
- Received over 1000 waitlist signups, featured on *There's an AI for That*, available on Testflight, received seed investments.

Traffic analysis using object classification model   **Asheville, NC**
Academic   January 2021 – May 2021
- Collected traffic flow data using TensorFlow object classification model, Docker, Raspberry Pi , Python for data analysis
- Performed ETL and reduced 7mil rows of data to 5k rows using SSIS and Microsoft SQL Server for data analysis
- Analyzed and visualized data using Tableau and presented at 2 universities to show traffic trends for facilitating infrastructure improvements

Skills

Programming Languages: Python, SQL, Bash, Linux

Python Libraries: Beautiful Soup, Keras, Matplotlib, Numpy, OpenCV, Pandas, Plotly, Pyspark. Scikit-learn, Tensorflow

Database Tools: Airflow, Firestore, Hadoop, Microsoft SQL Server, Postgres, RDS, Spark

Technologies: AWS (EC2, IAM, Lambda, RDS), Docker, GCP Firebase, Git, Github, Jira, Jupyter, Node.js, REST, Tableau, Visio, XCode