Jake Haines

+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: Applied Data Analysis, Regression Analysis, Applied Spatial Statistics, Bayesian Analysis, Probability, Data Structures, Statistical Computation & Data Management, Database Management Systems, Computer Vision, Natural Language Processing

Experience

WEX **Raleigh, NC**

Data Scientist Intern September 2024 – Present

- Analyze behavioral data using linear and logarithmic regression, random forest, Python, Scikit-learn, Pandas, and Pytorch to develop user experience for generative telematic data insights application used by 5 corporations
- Build backend infrastructure using Flask, JavaScript, ThoughtSpot API for embedding dashboards into data visualization application

Myndmap **Princeton, NJ**

Cofounder 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.

Tesla **Palo Alto, CA**

Data Engineer Intern August 2022 – December 2022

- Optimized Airflow big data pipeline using, Python, Spark, SQL, Hadoop for data from 100k vehicles, reduced query time 76%
- 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 Engineer Intern May 2022 – August 2022

- Analyzed data using Python, pandas, numpy, plotly, scikit-learn to test vitals sensors, improved sensor reliability by 25%
- 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

Projects

Automation for calendar organization **Princeton, NJ**

Personal July 2024 – August 2024

- Implemented productivity app API using Python script, TickTick API, pandas, and scipy to automate calendar scheduling and task prioritizing based on task categories
- Built automation using Python, TickTick API, and pandas to reschedule events in calendar if event timing is changed

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 to show traffic trends facilitating infrastructure improvements, presented at 2 universities

Skills

Programming Languages: Python, SQL, Bash, Linux, JavaScript

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

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

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