

# Jake Haines

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

## Education

### North Carolina State University, BS Statistics

Raleigh, NC

**Major:** Statistics | **Minors:** Computer Science, Music Technology

August 2021 – May 2025

**Coursework:** Linear Algebra, Probability, Natural Language Processing, Python for Data Science, Statistical Computation & Data Management, Applied Data Analysis, Database Management Systems, Data Structures, Research, Recording Studio Practicum, Audio Engineering

**Activities:** Data Analytics Club member, Automotive Club member, collegiate division e-sports team member

## Experience

### Tesla

Palo Alto, CA

Data Engineering Intern

August 2022 – February 2023

- ETL using SQL queries to analyze big data on 10mil rows from 100k vehicles (PostgreSQL, Pyspark, Spark, pandas, plotly)
- Managed big data pipelines handling 8 months of granular sensor data from 100k vehicles in Airflow, reduced query time by 52%
- Built dashboard w/ sensor and firmware data visualizations from Airflow pipeline (plotly, Tableau API, PostgreSQL, Spark), increasing firmware deployment frequency

### Tesla

Fremont, CA

Research Intern

May 2022 – August 2022

- Data analysis and testing, improved supplier vitals sensing accuracy by 76% (pandas, numpy, plotly, scikit-learn)
- Collection and ETL of vitals sensing, CAN, geospatial, Raspberry Pi sensors in Python (pandas, numpy, plotly, Mapbox),
- Research design and execution of user experience studies for early-stage projects (pandas, scipy, scikit-learn)
- Built 6 proof-of-concept projects with potential implementation to over 250k production vehicles
- Authored technical documentation, visuals, and Gantt chart for studies, code, and databases (Clickup, Sharepoint, Visio)

### Eagle Parts & Products

Augusta, GA

Data Engineering Intern

May 2021 – August 2021

- Data cleaning, parsing, and ETL on 15k rows of unstructured inventory data in Python (pandas, regex)
- Restructured data for migration by automating data entry, 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

Collect user's live listening data via Spotify REST API and live sensor data from Subaru ECU. Analyze driving and listening behavior, add tracks to queue based on driving habits using Spotify API.

### Data storage and warehousing infrastructure for mobile productivity app

London, UK

Startup

April 2023 – Present

Built AWS and Firebase infrastructure (PostgreSQL, IAM, RDS, EC2, Lambda, Firestore) with 0 expense. Implemented Firebase Apple authentication for user signup and login (Node.js, XCode, React Native). Generated over 800 waitlist signups, featured on *There's An AI For That*, available on Testflight. Authored 10 research papers for features from research in clinical executive dysfunction management strategies

### Traffic analysis using object recognition model

Asheville, NC

Academic

January 2021 – May 2021

Collected traffic flow data via TensorFlow object classification model containerized in Docker on Raspberry Pi. Performed ETL in SSIS and Microsoft SQL Server reducing 7mil rows of data to 5k rows. Data mining, analysis, and visualization in Tableau and presented at 2 universities.

## 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, Firebase, Git, Github, Jira, Jupyter, Node.js, REST, Tableau, Visio, XCode