

DAVID CUI

cuidavid02@gmail.com | [Github](#) | [LinkedIn](#) | davidcui.net | +1(617) 866-9522

EDUCATION

Northeastern University, Khoury College of Computer Sciences

Boston, MA

Candidate for BS in Computer Science and Economics

Expected December 2024

- **GPA:** 3.72/4.0, Dean's List
- **Selected Coursework:** Algorithms & Data Structures, Computer Systems, Object Oriented Design, Database Design, Logic and Computation, Programming in C++, Applied Econometrics, Linear Algebra

SKILLS & TECHNICAL TOOLS

- **Programming Languages:** Python, Java, SQL, R, C++, C, LaTeX, Racket, HTML, Javascript, ACL2, Lean
- **Technologies:** AWS, Linux, Snowflake, MySQL, Docker, Octoparse
- **Frameworks:** Flask, React.js, PyTorch, JUnit, Pytest
- **Development Tools:** IntelliJ, Visual Studio, RStudio, DataGrip, Eclipse, PyCharm, Overleaf, DrRacket
- **Speaking Languages:** Conversational Mandarin Chinese and Spanish

PROFESSIONAL EXPERIENCE

Starday Foods (CPG Startup)

Junior Data Science Software Developer (Part-time)

September 2024 - Present

Junior Data Science Software Developer

July 2024 - August 2024

- Improved access to Meta Llama 3.1 and Anthropic Claude 3.5 models using AWS Bedrock and SageMaker, enabling efficient processing of over 100,000 queries through chatbot integrations and the development of chatbot wrappers for seamless user interactions
- Produced custom APIs and Lambda functions to analyze over 100,000 product images and descriptions, extracting key attributes and reducing manual analysis time by over 500%
- Designed and implemented an end-to-end process in AWS that extracts attributes from over 100,000 food and beverage products, automating their analysis for deeper insights and efficiency
- Engineered an automated data pipeline on AWS S3 with Lambda functions, connecting over 10,000 consumer purchase records and survey insights to CDP profiles, driving seamless data integration and boosting analytics precision

Junior Data Science Software Developer (Co-op)

Feb 2023 - December 2023

Fully Automated TikTok Post Ingestion | AWS (Athena, Glue, S3, Lake Formation), Python (Pandas, JSON)

- Enhanced an automated data retrieval system that collected over 1,000,000 TikTok videos and their details daily, efficiently transforming JSON data into organized tables for seamless analysis
- Built and managed a scalable 12-table data platform on AWS to ingest and store TikTok video data, leveraging tools such as S3, IAM, Lake Formation, Glue Notebooks, Glue Crawlers, and Athena

Similar Hashtag Generator | AWS (Lambda, API Gateway), Python (Pandas, JSON), Dataiku

- Created an applet that ingests a set of terms and generates a list of up to 250 similar hashtags per term by developing an AWS Lambda function that uses an external API to gather data
- Expanded Starday's food and beverage ontology keyword database by 25% in 1 month by deploying the applet via AWS API Gateway and Dataiku DSS

Product Database & GUI | Octoparse, Appsmith, Snowflake, SQL, Python (Pandas)

- Utilized Octoparse to web scrape data from over 100,000 products across major retailers like Target and Sprouts, organizing and structuring the collected data into distinct database tables
- Integrated data tables into Snowflake and developed a user-friendly GUI, streamlining interaction and query processes for stakeholders

Fake Survey Response Detector | Python (Pandas, Numpy, Scikit-learn)

- Designed a data preprocessing module that converts categorical survey responses into numerical values, enabling the utilization of machine learning algorithms for classification
- Trained and deployed several models on the transformed numerical data to predict the authenticity of over 3000 survey responses, saving stakeholders hours of manual inspection