

# Christopher Lum

[lum@ucsd.edu](mailto:lum@ucsd.edu) • (425) 295-8028 • San Diego, CA • [github.com/ch-lum](https://github.com/ch-lum) • [linkedin.com/in/ch-lum](https://linkedin.com/in/ch-lum)

## EDUCATION

---

### University of California, San Diego

*Bachelor of Science in Data Science, Minor in Business*

**La Jolla, CA**

*Grad: March 2025*

- **Cumulative GPA:** 3.86/4.00
- Relevant Coursework: Probabilistic Modeling and Machine Learning, Data Viz, Recommender Systems

## PROJECTS

---

### First-Authored Peer Reviewed Publication

Pre-Print: [https://bit.ly/ch-experts\\_novices](https://bit.ly/ch-experts_novices)

Christopher Lum, Guoxuan Xu, Sam Lau. “I’m not sure, but...”: Expert Practices that Enable Effective Code Comprehension in Data Science. ACM Technical Symposium on Computer Science Education (SIGCSE), 2025.

- First author on forthcoming research paper comparing novices and experts on difficult data analysis tasks

### Facebook Comment Sentiment Analysis

Live website: <https://ch-lum.github.io/comment-vis/>

- Used the DistilBERT model from Hugging Face to classify sentiment on 1,300-comments from a Biden Facebook post
- Implemented custom interactive visualization using D3.js and JavaScript
- Highlighted as **top-1 best project in class of 150 students** by instructor

### Rocket League Possession Model

Exemplar code: <https://bit.ly/ch-rl-parser>

- Rocket League is a physics-based video game centered around the premise of cars playing soccer
- Scraped over 10 GB of match replays, consisting of 30 frames-per-second physics data that recreates five-minute matches in the game’s client and parsed the resulting JSONs into a cleaned CSV dataset for analysis
- Trained a predictor to classify possession to serve as the first frame-by-frame possession analysis model in the Rocket League professional scene as a personal project

### AI Grading Assistant

Exemplar code: <https://bit.ly/ch-ai-partial-credit>

- Implemented OpenAI API in a full-stack web application in Next.JS and Python to give partial credit on previously all-or-nothing questions
- To be deployed in DSC 20 for winter quarter 2025 for use by tutors for class of 400 students

## PROFESSIONAL EXPERIENCE

---

### DSTL Lab @ UC San Diego

*Research Assistant – Visualizing Pandas Tables*

**La Jolla, CA**

*Apr 2024 – Present*

- Lead project group in developing ClassBuzz, an in-lecture web application (Next.js) for student engagement
- Pandas Tutor automatically visualizes Pandas code, used by over 30,000 people per year across 150 countries
- Enhanced project capabilities by increasing the dataset size limit from 1 MB to 100 MB, allowing for more complex data visualizations in Pandas Tutor
- Conducting user study on how people reason about data tables and how they get displayed in notebooks to understand data comprehension and presentation (see “Peer Reviewed Publication”)

### San Diego County Taxpayers Association

*Data Science Intern*

**San Diego, CA**

*May 2023 – Sep 2023*

- Before I joined, the data gathering was done by hand over the course of several months
- Finding place for improvement, I developed a ProPublica scraper for 990 forms with BeautifulSoup which scraped documents with 99% accuracy with error flagging for inaccuracies, doubling the size of our dataset with fewer errors than when done by hand
- Digitized SD County’s city budgets from PDF to XML preparing groundwork for financial and meta-analysis