Data Science Career Chat

Cogs 108

Brendan Tomoschuk

Career Timeline

- Penn State University (2010 - 2014)
  - Majors: Neuroscience, Linguistics
- UC San Diego (2014 - 2019)
  - PhD Cognitive Psychology
- Insight Data Science (2019)
  - Data Science Fellow
- Cambly (2019-2021)
  - Data Scientist
- Cruise (2021 - present)
  - Senior Data Scientist

Typical week

10-12 hours **meeting** time

- Team meetings (Data Science Brown Bag, Work planning)
- Regular One-on-Ones (Meeting with Manager, teammates)
- Misc project planning, co-working
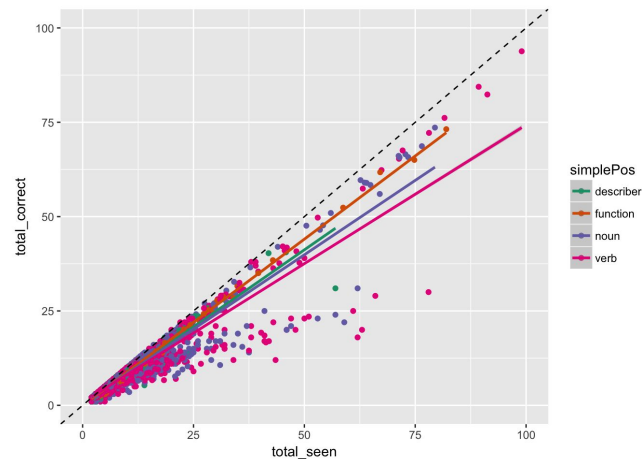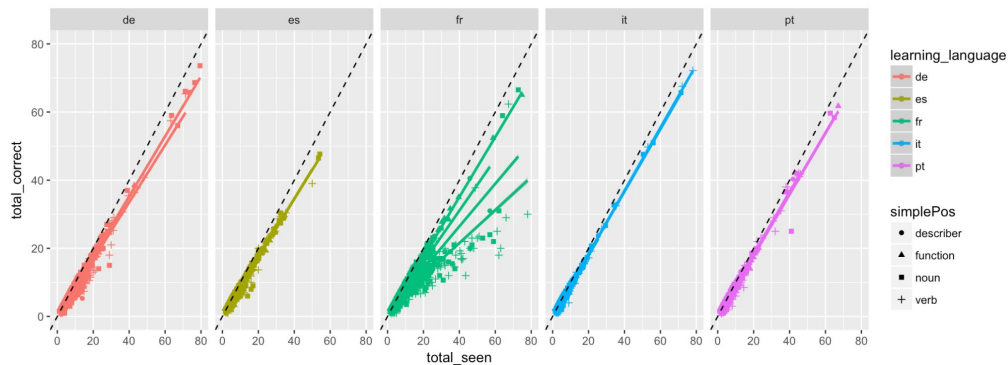
5-10 hours **document** time

- Writing project plans
- Editing/Commenting teammate docs
- Documenting code or repos

~20 hours "**coding**" time

- Includes data cleaning
- Pipeline development
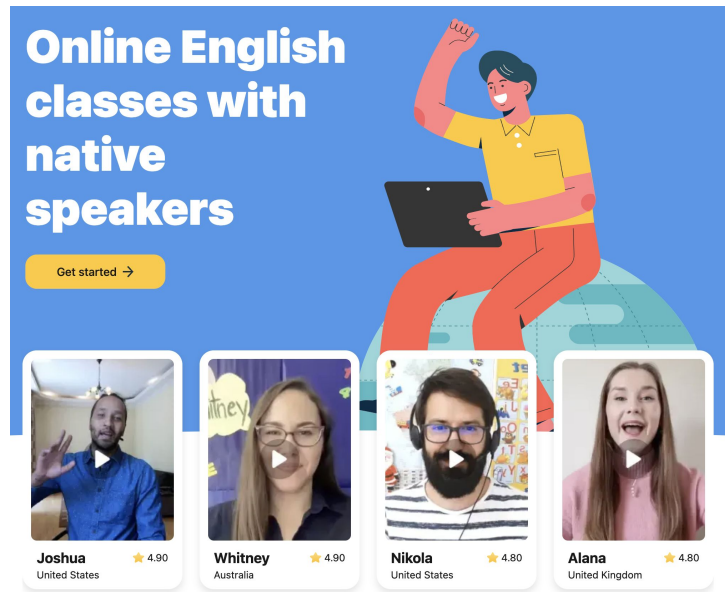- Data Analysis
- Modeling

Example Project - **Portfolio**

- Which languages on Duolingo are hardest to learn?
  - Downloaded a public dataset
  - Analyzed, graphed
  - Wrote up and linked online

Example Project - Cambly
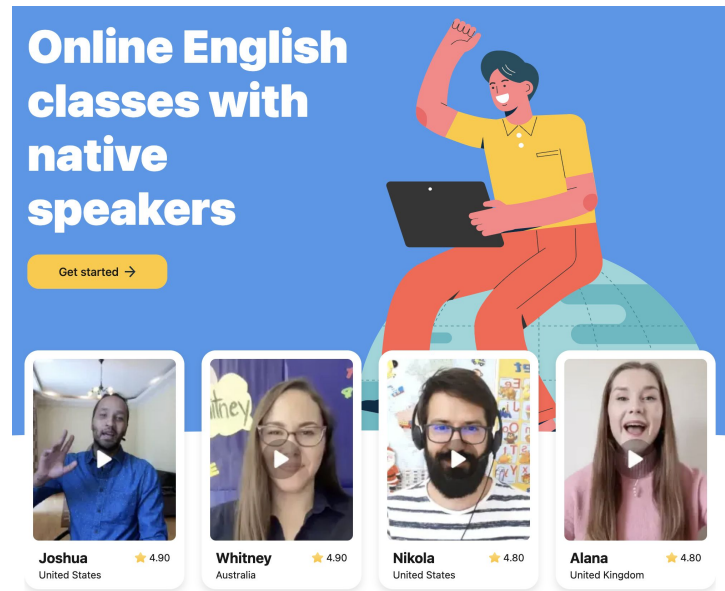
Does training tutors help book
and retain students?

Example Project - Cambly

- Experimentation
  - Tutors split into three groups
    - no instructional material
    - optional instructional material
    - mandatory instructional material
  - A/A testing
    - Make sure tutors in each group don't differ drastically
  - Metrics
    - Trial conversion
    - Student 1 month retention

Example Project - Cruise
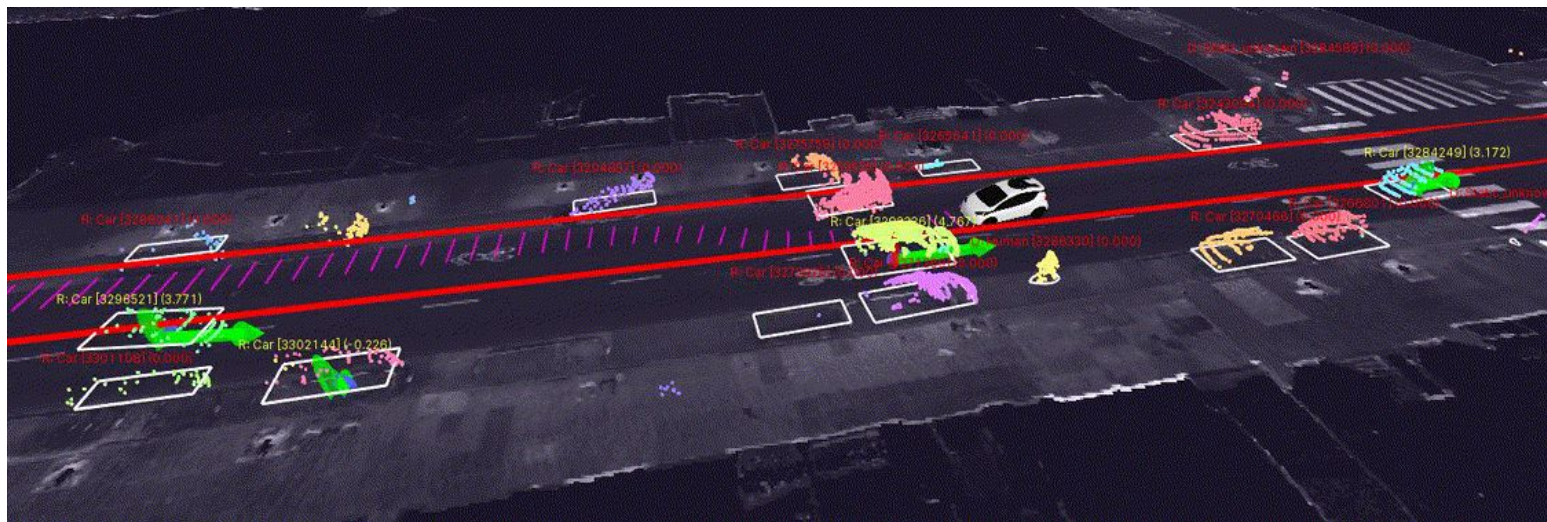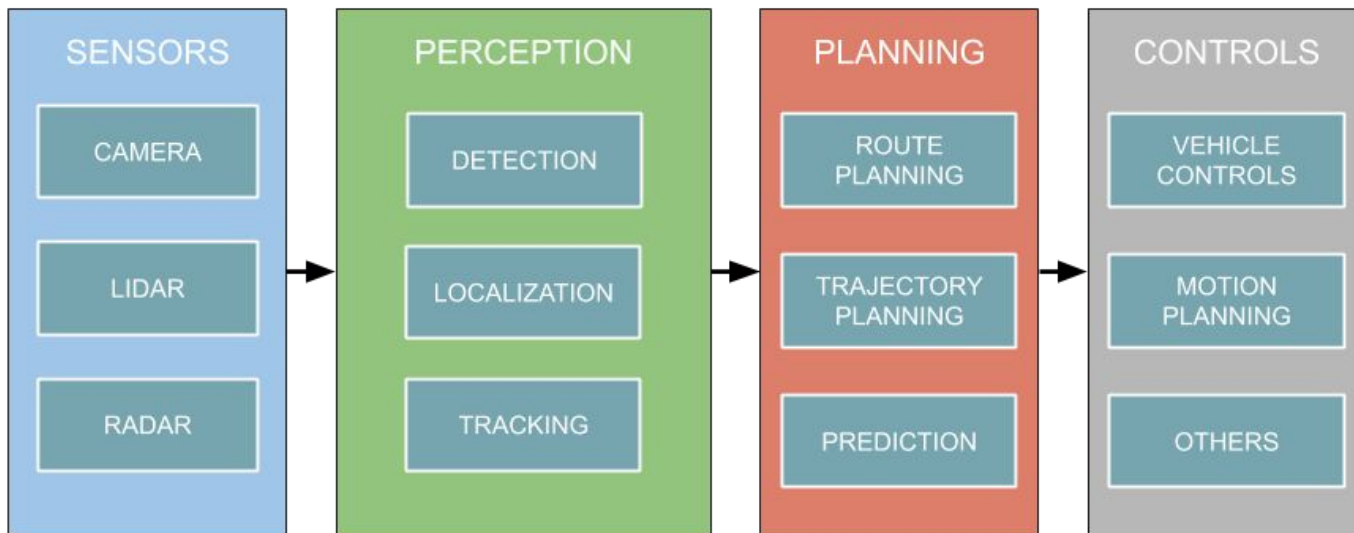
Example Project - Cruise

Example Project - Cruise

- Trip Monitoring Metrics
  - Gathering, cleaning, scaling input data
    - Structured data
      - Trip Rating 1-5
      - Thumbs up/down of driving categories like Pickup/Dropoff
    - Unstructured data
      - Written feedback
    - Vehicle data
      - Acceleration, jerk etc

Example Project - Cruise

- Trip Monitoring Metrics
  - Deciding how to unify and present data
    - Model?
      - Binary output (trip good or bad?)
      - Trip score (out of 100?)
      - Regression? Random forest? Heuristics?
    - How do we balance?
      - Should a bad pullover affect the score more or less than a hard brake?
    - Is the model comprehensible to non-technical audiences?
    - How do we know the metric remains correct or relevant?

Job searching

- During grad school:
  - Applied to ~15-20 internships, none accepted
- During Insight program
  - Applied to ~50 companies (Over 6 months)
    - ~8 interviews
      - 2 offers
- When leaving Cambly
  - Applied to ~30 companies (Over 4 months)
    - ~6 interviews
      - 2 offers

Landing a job

- Portfolio
- Tech Interviewing
    - "Leetcode" coding questions
    - Statistics
    - Problem solving
    - Effective communication around projects
- Referrals / "Networking"

Questions?

https://bit.ly/brendan-wi24