

Language Learning on Duolingo

By Kate Christensen



What is Duolingo?

Home page
for a course a
user is
enrolled in:

The screenshot shows the Duolingo home page for a German course. The top navigation bar includes links for LEARN, STORIES, DISCUSS, SHOP, and MORE. The user's profile is shown with a German flag, a crown icon, and a score of 449. The main content area displays a grid of lesson icons, each with a yellow circular icon and a small yellow tag indicating the number of lessons completed. The lessons are organized into rows: Basics 2 (5 lessons), Phrases (5 lessons), Idioms (1 lesson), Flirting (1 lesson), Christmas (1 lesson), Acc. Case (5 lessons), Intro (5 lessons), and Food 1 (5 lessons). A blue button labeled 'TRY PLUS' is visible next to the Phrases lesson. On the right side, there is a 'Daily Goal' section with a large orange circular progress indicator showing 558 DAYS. To the right of the progress indicator, it says '13/10 xp goal met' and '10 hours left'. Below the progress indicator is a line graph showing the user's progress over the last seven days (Th, F, Sa, Su, M, Tu, W). The graph shows a peak on Thursday (Th) at approximately 28, followed by a drop on Friday (F) to approximately 15, a rise on Saturday (Sa) to approximately 20, a drop on Sunday (Su) to approximately 12, and then a relatively flat trend on Monday (M) at approximately 14, Tuesday (Tu) at approximately 14, and Wednesday (W) at approximately 13.

LEARN STORIES DISCUSS SHOP MORE

German 449 558 2049

TRY PLUS

Basics 2 Phrases

Idioms Flirting Christmas

Acc. Case Intro

Food 1

Daily Goal EDIT GOAL

558 DAYS

13/10 xp goal met

10 hours left

40 30 20 10 0

Th F Sa Su M Tu W

About the Data

- Each observation is an individual word a user encounters in a single lesson
- Data was taken from users over 8 different courses:
 - English speakers learning German, French, Spanish, Italian, and Portuguese
 - Portuguese, Spanish, and Italian speakers learning English
- Overall dataset had over 12 million observations

Target Variable and Features Analyzed

Target variable:

Overall Accuracy (overall times correct/overall times encountered)

Word metadata:

- Part of speech:
 - Noun, Verb, Adjective (adj), Adverb(adv), Determiner, Preposition
- Length
 - word_len
- Number
 - Singular and plural

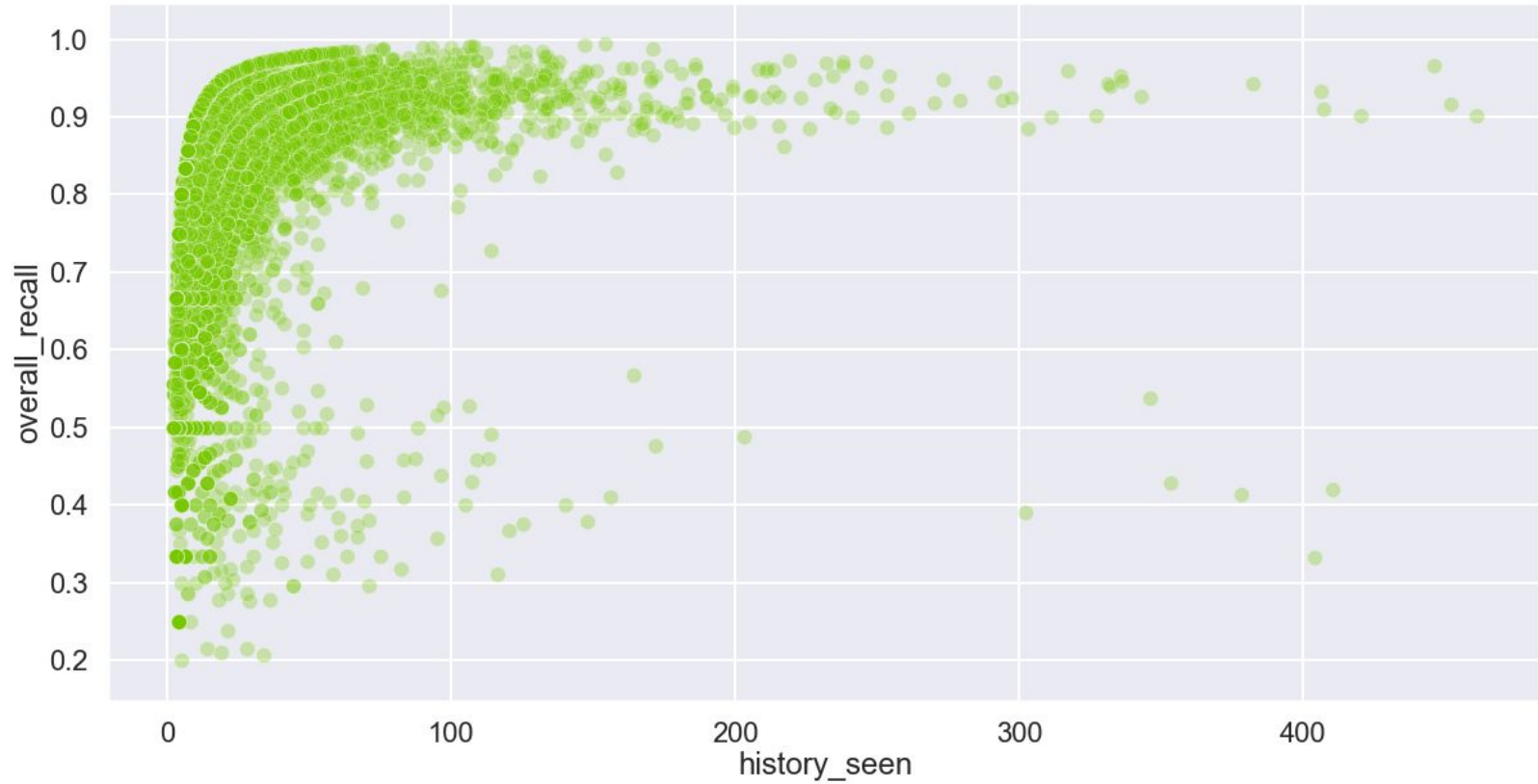
User data:

- Delta- time since they last saw that word
- Times the user has encountered the word:
 - Session_seen: encounter in a lesson
 - History_seen: over all lessons
- Session_comp- Number of lessons the user has completed in that course
- P_recall- Accuracy within individual lessons

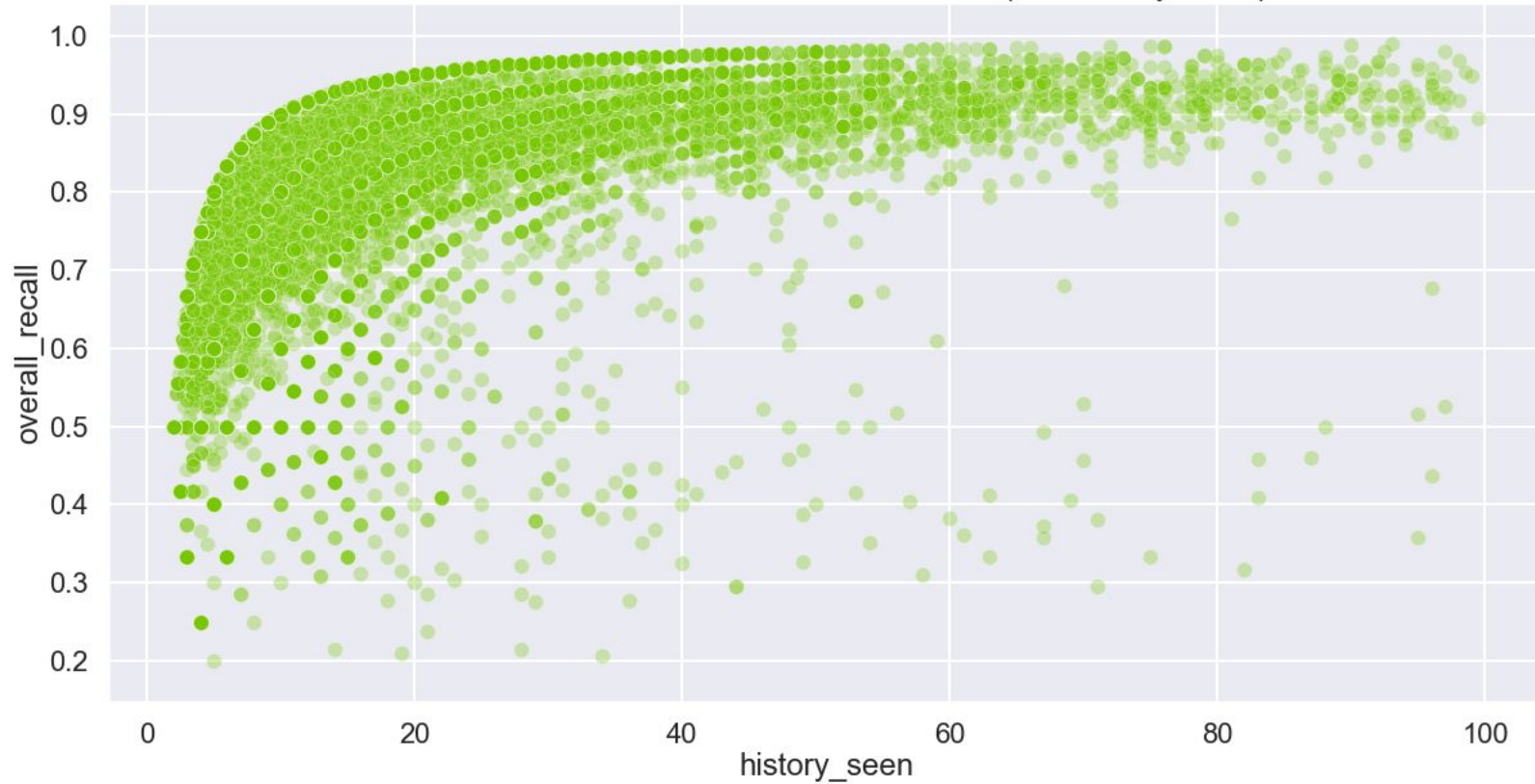
Notes on the Data

- Points where the overall accuracy was 1 or 0 were dropped, since we're interested in learning
- Filtered the data so each word was the last time a given user saw that word
- Data was aggregated by taking the means for each word encountered in a certain session
 - Example: the averages for the word "have" when it popped up in users' first session, "have" in their second session, etc.
 - This dataframe had 24,276 total observations

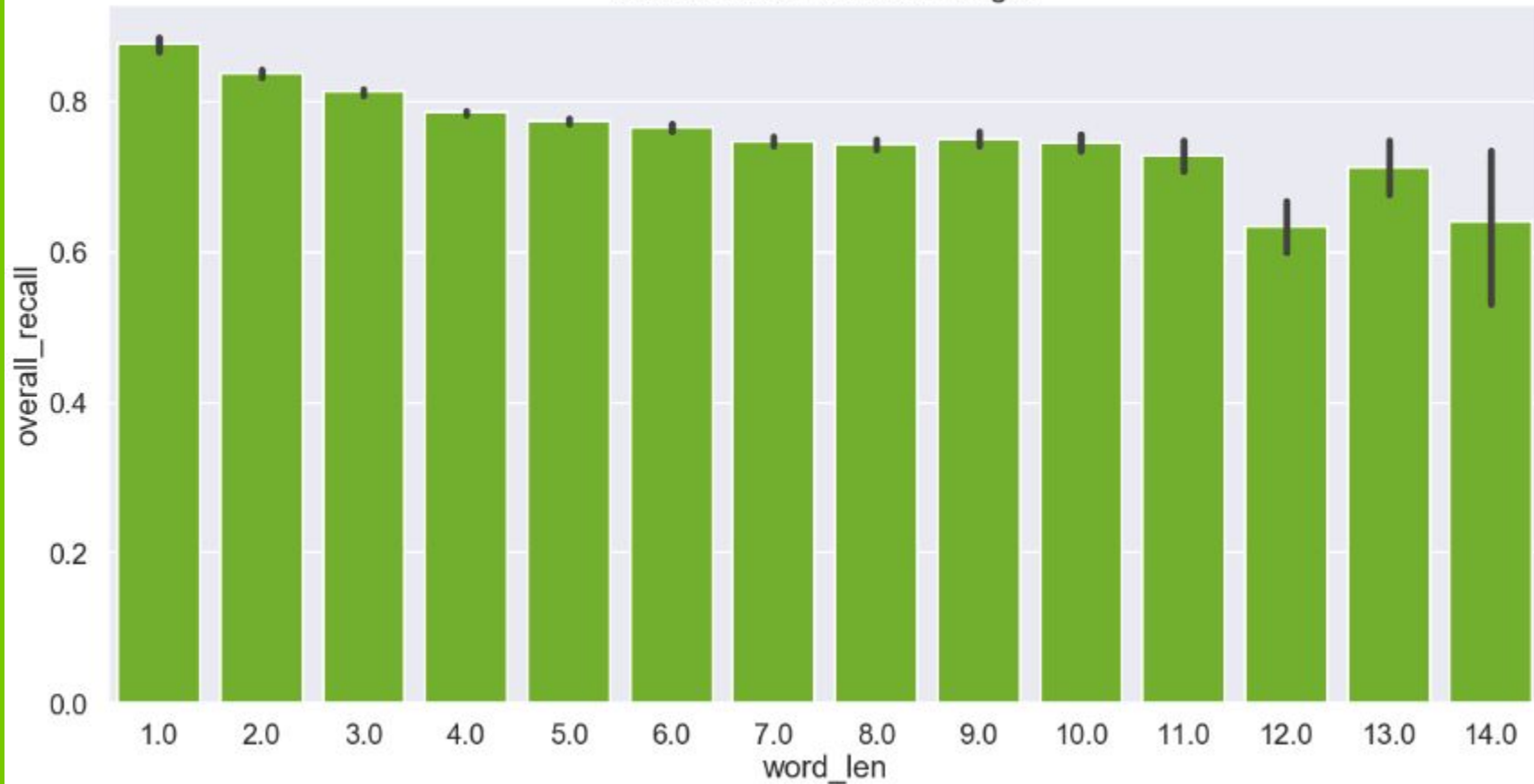
Overall Recall vs Times a User Sees a Word



Overall Recall vs Times a User Sees a Word (Max history = 100)



Overall Recall vs Word Length



Ordinary Least Squares Regression Results

R squared: 0.952

Mean Squared Error: 0.03

Critical Points from the Results

- Overall:
 - P_recall (0.39)
- Within word metadata:
 - Prepositions (0.24)
 - Determiners (0.22)
 - Adverb (0.21)
- Session seen (0.04) greater than history seen (0.00021)
- All p-values except for past tense (0.40) are below 0.005

Conclusions

- Doing well in a lesson means you're likely to be doing well overall
- Prepositions, Determiners, and Adverbs are the easiest types of words to learn
 - Prepositions being higher than determiners is unexpected
- Seeing a word more in an individual session will help you with overall accuracy more than just seeing it a lot overall