

Group Members: Viera Garcia, Ting Jennings, Erin Zheng, Ricardo Benitez, Tina Huang

Outline: We are planning on bringing 3 csv files into the postgres database and cleaning out the columns so we can join it with the data sources from years 2019-2020 for comparison purposes.

Research Question: Which country has the best chess players? Does age play a role in chess rankings? Compare results between the 2 years.

Data: <https://www.kaggle.com/rohanrao/chess-fide-ratings>

Description:

- Players.csv has a list of all the players including name, FIDE_id, gender, YoB, and country
- Ratings_2019 & ratings_2020 include FIDE_id and rating for every month for the selected year

Work: Chess Ratings 2019-2020

Done by Saturday: Viera, Erin

- Merge 2-3 data sets, remove duplicates
- Show only columns we need (rating_standard, federation/country, name, chess title, YOB)

Saturday-Monday: Ricardo, Tina, Ting, Viera, Erin

- Clean data (python)
 - Categorize by country & pull highest rated player
 - Section by age group then find the rating average in each age group
 - <20|20-30|30-40|40-50|50-60|60-70|>70
 - <2001 | 2001-1991 | 1991-1981 | 1981-1971 | 1971-1961|>1951

2019 country	Tina
2019 age	Ting, Ricardo
2020 country	Erin
2020 age	Viera

- ETL diagram
- Combine data in SQL

Monday-Tuesday (Together)

- Github push
- Final writeup