Detailed Project Description

1. Project Title

Scoring Bias of Judges in International Figure Skating

2. Project Summary

The purpose of our project is to shed light on the scoring bias of the judges in international figure skating competitions, mainly based on nationality. We will analyze the data on points given to skaters by the judges collected from such competitions.

Initial analysis will be on judges having biases for skaters from their own nationalities. Secondary analysis will be on the positive or negative bias of certain countries for other countries. We will make a website where the users can view and filter the data and analysis on tables and graphs.

3. Description

While we are not figure skating officials, we hope to bring awareness to bias in figure skating judging through a statistics-based approach. Biases can be seen through judges consistently giving skaters lower or higher scores relative to other judges on the judging panel.

We will compare the points given to similar moves made by the skaters by different judges, scores that increase the standard deviation and scores that stand out as vastly different from other judges' scores to the same skaters. Using this analysis, we will design a website and showcase our results using various graphs and charts. We plan the website to be view-only, where people can filter the data and use different methods to display it.

4. Usefulness

Subjective sports scoring and officiating are often biased because they are human. An example is the 2022 Salt Lake City Olympics figure skating scandal, where the scoring investigation resulted in judges being dismissed. Most sports scoring/officiating is generally fine (like who is the fastest in swimming/track, who wins tennis matches, etc.), but more so for sports with an artistic component, it is often biased because of the subjectivity of the human factor. These biases can include factors such as politics, gender, faith, etc. Our database and analysis of the data are designed to bring awareness to such bias and encourage discussions regarding ways to improve scoring. Specifically, we analyze whether a particular judge will always score a player of a particular country lower than other judges. If this is the case, is it possible to reduce the weight of the judge's score in the overall scoring to achieve a fairer score?
This was pointed out during the 2022 Winter Olympics in Beijing (source: https://twitter.com/SkatingScores/status/1573349500810567680?cxt=HHwWgIDR9Z7C1NUrAAAA) , but his analysis and visualization of the data only showed the results for one skater's data analysis. What we want to do is to combine all the scores of each judge to determine if the judge may have bias.

5. Realness

Our data sources are currently decided as three data sites, the first one is https://skatingscores.com/, the second one is https://github.com/BuzzFeedNews/figure-skating-scores/tree/master/data/json, and the last is https://github.com/BuzzFeedNews/2018-02-figure-skating-analysis/tree/master/data . This data includes information about the athletes, the judges and detailed scoring.
All we need to do is read the JSON file and insert it into the corresponding table, e.g. grab the judges' information and insert it into table Judges. Skating scores' data is highly visualized, so we can better understand how to analyze each part.

6. Functionality

a) The website stores data regarding international figure skating competitions that took place in 2016-2017. These competitions were held in various countries, such as Russia, France, Japan and Canada. This data comes from BuzzFeed News's figure-skating-scores public repository on GitHub. It will be split into 4 entities: competitions, skaters, judges and scores

Each Competition table has info regarding the competition name, location it was hosted in, and it was held. The Skater table will have tuples of every skater who competed in the presented competitions, and attributes will include their name and nationality. Skaters will be tied to competitions via a Performance relationship table that has a performanceID, the name of the skater tied to the ID, the name of the competition tied to the ID, and the specific program the skater competed in (i.e., Free Skating). The Judge table will hold every judge that scored the above competitions, with attributes being their own name and nationality. There is also an entity table for Scores, with attributes including total scores and subscores, like segment score, element score and component score, as well as the number of deductions the skater received from a particular judge. The score will also include the name of the judge that gave the score, the role of the judge in that scoring, and the performanceId of the score to match it to a player and competition.

The fifth and final entity we will have does not come from the BuzzFeed News dataset. It is the Admin entity, which will hold the admin's ID, name and date of account creation. This entity represents the approved and trusted users that can make edits to the website. It will be auto-generated for testing purposes, but it would be user-populated if the website were actually going live.

b) Users will be able to navigate different links on the website to view the raw data of certain entities in table form. These entities are Competition, Skaters and Judges. For each entity, the tuples will be displayed in table format. Users can then change how the tuples are ordered (i.e., sort competitions by date, performers by nationality in alphabetical order). They can also use a search bar above each table to search for entries either by primary key (aka name) or other attributes like nationality for performers. For the competition table, they will be able to click on a given competition to see more details about it (i.e., specific performances of the players who competed in that competition) To see bias in judges, the user will be able to access a page that represents the average scores of judges to skaters of various nationalities graphically, allowing for a visual representation of the discrepancies

(described in more detail below.) In terms of user-manipulation of data, there will be an option to log-in to the page as an administrator. Administrators will have the power to add/update to existing databases, as well as delete information. We reserve this as an administrator privilege, because we don't want unverified users to potentially sabotage the data to make their own nationality look unbiased.

c) Our creative component is the graphical representation of judges' biases. Aside from being able to view the discrepancies of a selected judge's scores, the user will be able to manipulate the further chart by selecting multiple judges of the same country to see if any seemingly favored scoring is a coincidence, or a part of a greater trend. The user can also filter the chart by country overall, to see which countries have the greatest biases in scoring.

7. Low Fidelity UI Mockup

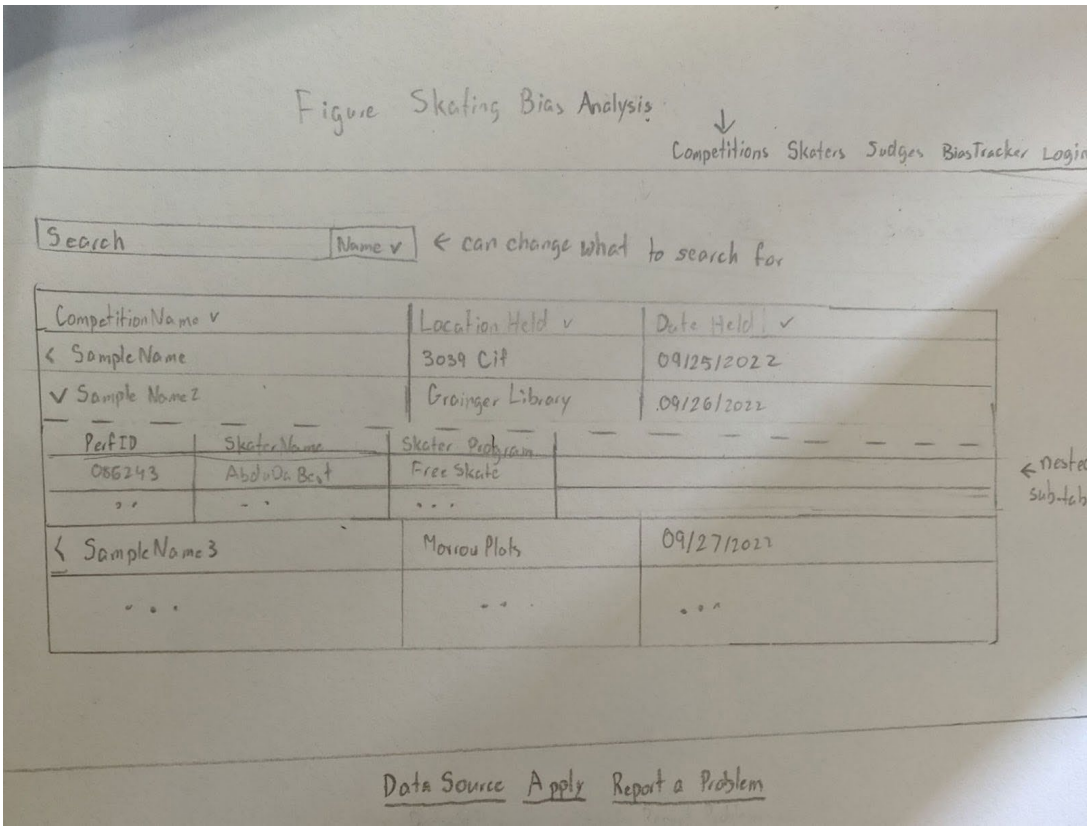
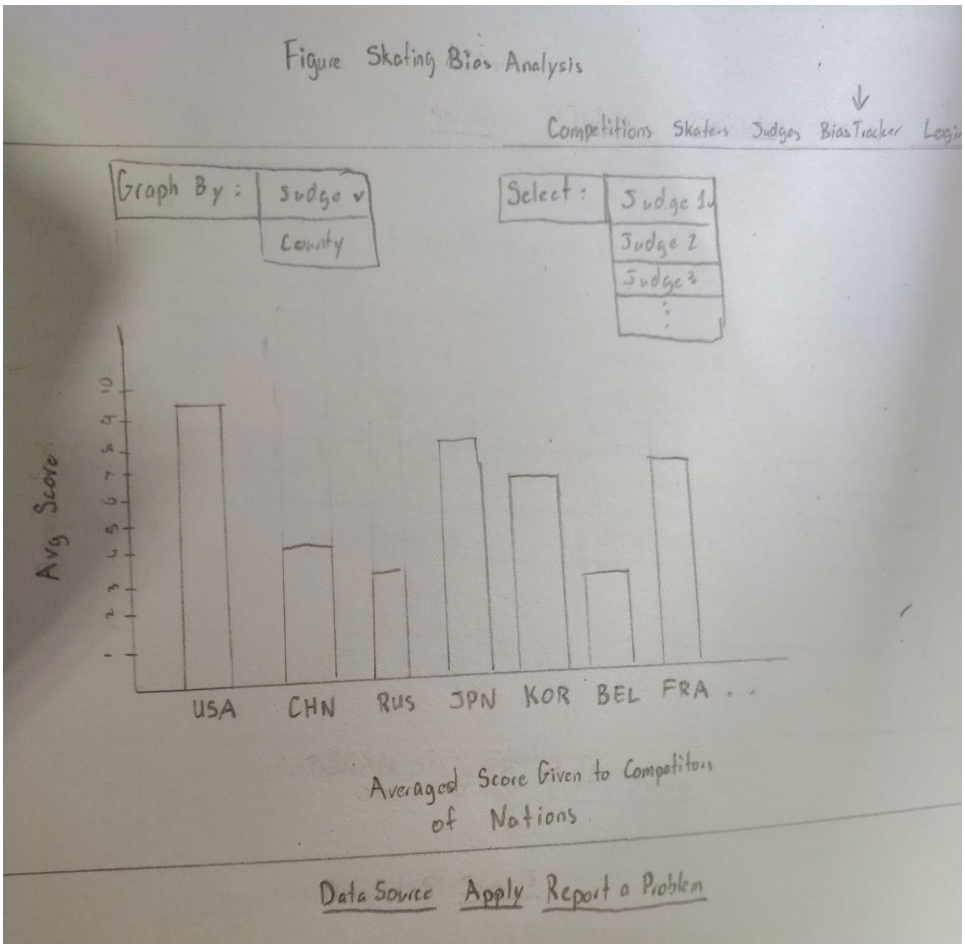


Figure Skating Bias Analysis
Competitions Skaters Judges BiasTracker Login
Search
Name
← can change what to search for
CompetitionName
Location Held
Date Held
SampleName
3039 CIT
09/25/2022
Sample Name2
Grainger Library
09/26/2022
PerfID
SkaterName
Skater Program
086243
AbdulDa Beat
Free Skate
SampleName3
Morrou Plots
09/27/2022
Data Source
Apply
Report a Problem



8. Project Work Distribution

1st step - parsing data

- Parsing JSON - Athena, Murat
- Parsing CSV data - Victoria, Zepei
- Compare and choose React vs Flask - Athena, Murat, Victoria, Zepei

2nd step - DB creation - Athena, Murat, Victoria, Zepei

3rd step - inputting in Data -> DB - Athena, Murat, Victoria, Zepei

4th step - features (req both frontend and backend for querying)

- Search bar/filtering for tables - Zepei
- Admin (login, manipulation of tables) – Victoria
- (Creative component) Visual/graphical representation of judge biases - Athena, Victoria