IM2C 2025 Presentation

Rye Country Day School

George Zhan, Felipe Quintero Ochoa, Felix Mao, Aarush Dey

The Problem:
Creating A Global Sports League

Our Approach

1. **Identify** a sport.
2. **Identify** and **rank** the best teams in that sport.
3. Find a way to **reduce travel distance**.
4. Create a **schedule optimizer**.
5. Make the **schedule entertaining** and **maximize viewership**.
6. **Expand** the model.
7. **Explain** our **assumptions**.
8. **Submit** our solution.

Our Vision:
Global Women's Football League

But why Women’s Football?

Growing popularity

Opportunity for Global Attention

Competitive Balance

Assumptions

Teams fly to their “home” after away games.

A week is an adequate rest for travel

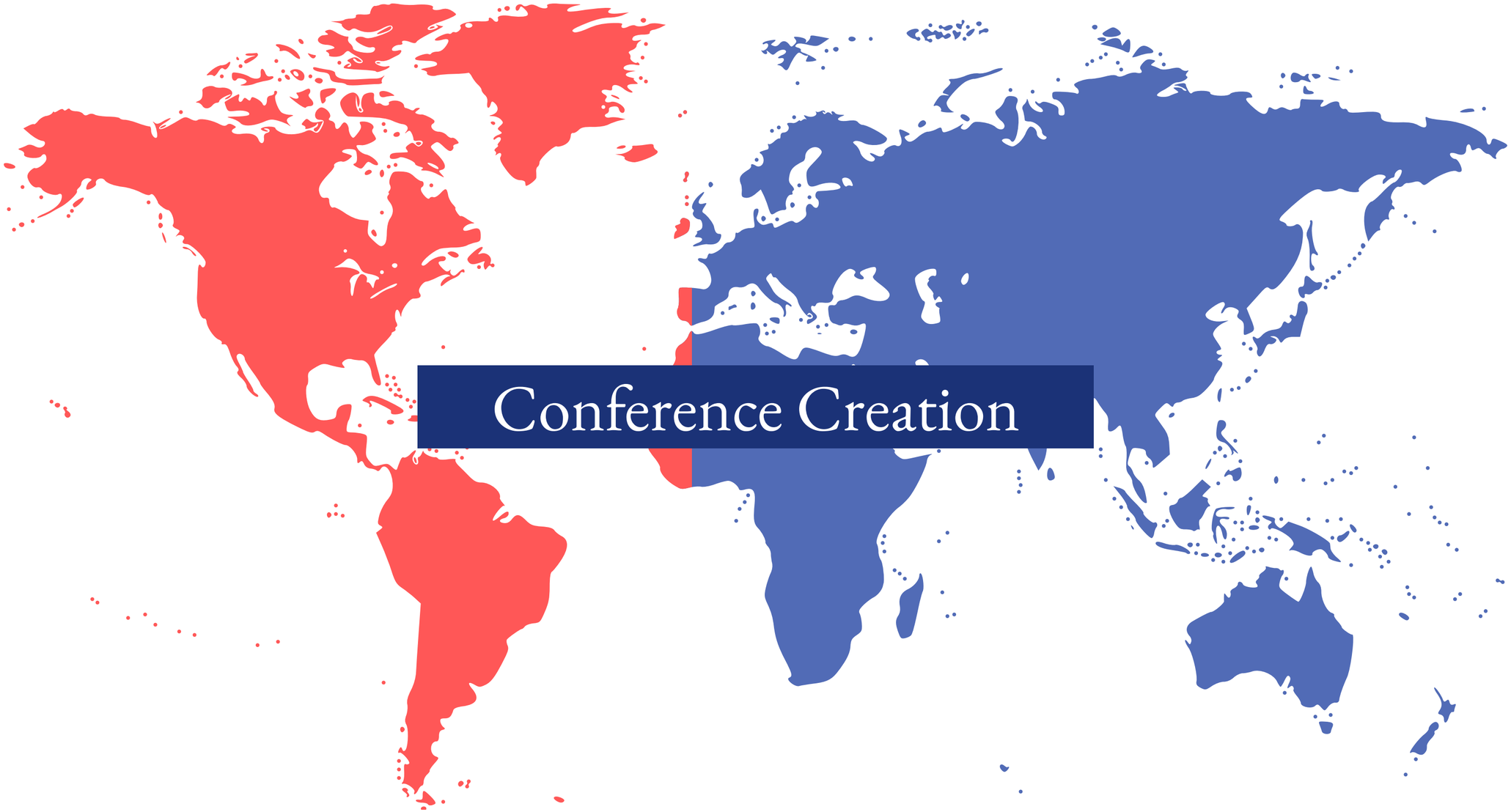
The schedule can be changed and/or adjusted for international club tournaments (i.e, the UEFA women’s Champions League and CAF Champions League)

Teams leave their original leagues to play

Flying time is exactly proportional to flight distance.

Our 3 Part Solution

Team Selection


Conference Creation


Scheduling

Our Approach to Team Selection

Identify the top five teams in the top women's soccer leagues in 2023-2024 .

Accquire a League Power Coefficient to Determine the International "Difficulty" or "Prestige of a League

Collect league data on the top five teams

Create a formula or set of formulas that incorporate the League Power Coefficient to determine the strength of a team

Select the top-ranking teams in each continent (with caveats)

Identify the Top Leagues

Continent	Leagues
Europe	Barclays WSL (England), Liga F (Spain), D1 Arkema (France), Frauen Bundesliga (Germany), Serie A Women's (Italy)
North America	NWSL (USA), Liga MX Femeníl (Mexico)
South America	Brasileiro Women (Brasil), Campeonato de Futbol Femenino (Argentina), Categoria Primera A - Womens (Colombia)
Africa	Moroccan Women's Championship (Morocco), SAFA women's league (South Africa)
Asia	WE League (Japan), WK League (South Korea), CWSL (China)
Australia	A-League Women (Australia)

RYE COUNTRY DAY SCHOOL
1869

Variables

Variable	Definition:
W	Wins
D	Draws
MP	Matches played by a single team in their respective league.
MPT	Matches played in the entire league
SLB	Skill level bonus.
TPRi	Initial Team Power Rating
NT	Number of Teams
UEC	The UEC stands for the universal error constant which is used to calculate potential error, decreasing a team's Team Power Rating (injuries, missing games due to international break, suspensions, etc.).
LPC	League Power Coefficient
LPR	League Power Rating

RYE COUNTRY DAY SCHOOL
1869

Formulas

Formula I - Initial Team Power Rating

$$TPRi = ((\frac{3W+D}{MP}) + (\frac{GD}{MP}) + (\frac{MP}{MPT})) * UEC$$

Formula II - Skill Level Bonus

$$SLB = (\frac{\sum TPRi\ of\ top\ 5\ teams\ in\ each\ league}{5}) * 0.02$$

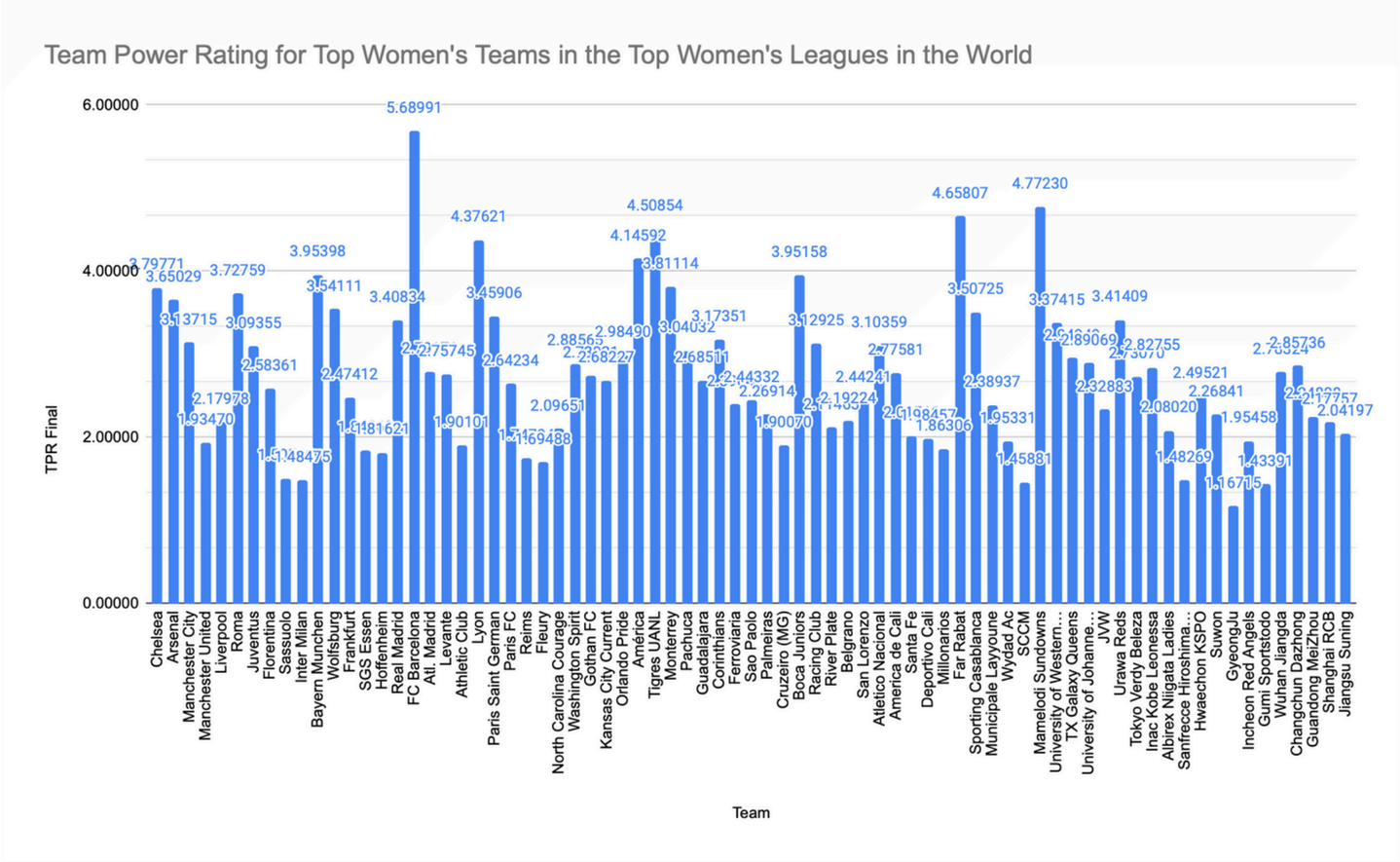
Formula III - League Power Rating

$$LPR = LPC * SLB$$

Formula IV - Final Team Power Rating

$$TPRf = TPRi * LPR$$

Results



RYE COUNTRY DAY SCHOOL 1869

Team Selection

Continent	Team	Location	League
North America	Tigres UANL	Monterrey, Mexico	Liga MX Femenil
North America	Club América	Mexico City, Mexico	Liga MX Femenil
North America	Orlando Pride	Orlando, USA	NWSL
North America	Washington Spirit	Washington DC, USA	NWSL
South America	BOCA Juniors	Buenos Aires, Argentina	Campeonato de Fútbol Femenino
South America	Racing Club	Buenos Aires, Argentina	Campeonato de Fútbol Femenino
South America	Corinthians	São Paolo, Brazil	Brasileiro Women
Africa	Far Rabat / AS Far	Rabat, Morocco	Moroccan Women's championship
Africa	Sporting Casablanca	Casablanca, Morocco	Moroccan Women's championship
Europe	Chelsea	London, England	Barclays WSL
Europe	FC Barcelona	Barcelona, Spain	Liga F
Europe	Roma	Rome, Italy	Serie A Womens
Europe	Lyon	Lyon, France	D1 Arkema
Europe	Bayern Munchen	Munich, Germany	Frauen Bundesliga
Africa	Mamelodi Sundowns	Pretoria, South Africa	SAFA Women's League
Asia	Urawa Reds	Saitama, Japan	WE League
Asia	Inac Kobe Leonessa	Kobe, Japan	WE League
Asia	ChangChun DaZhong	ChangChun, China	CWSL
Australia	Melbourne City	Melbourne, Australia	A-League Women
Australia	Sydney	Sydney, Australia	A-League Women

RYE COUNTRY DAY SCHOOL 1869

Conference Creation

Our Approach to Creating the Conferences

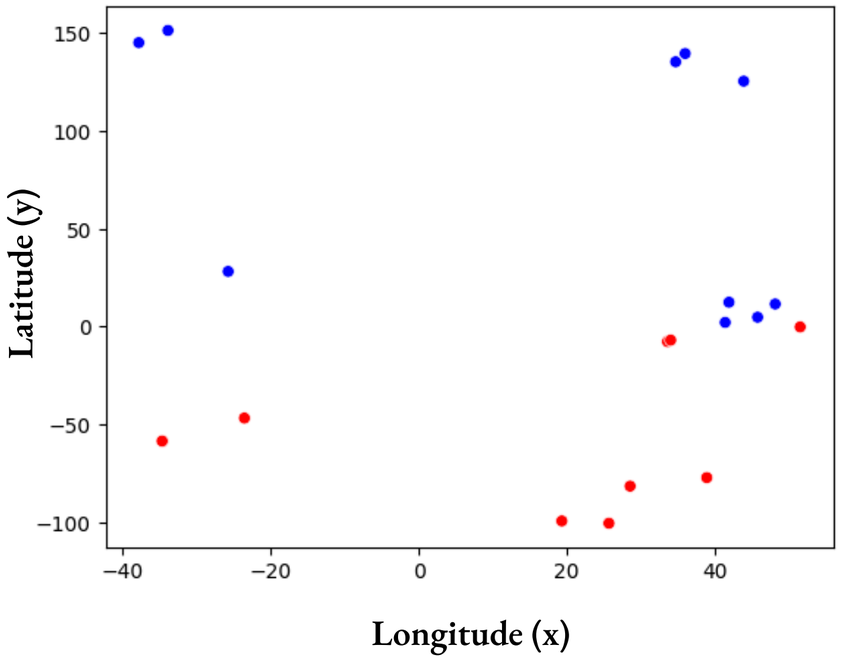
Haversine Formulas

$$d = 2r \cdot arcsin(\sqrt{sin^2(\frac{\Delta\phi}{2}) + cos(\phi 1) \cdot cos(\phi 2) \cdot sin^2(\frac{\Delta\lambda}{2})}$$

Having the distance between teams was essential to the creation of the conferences in the global sports league. The haversine formula enabled us to accurately calculate the distance.

K-Means Clustering

Results of K-Means Clustering



K-Means Clustering Formula

$$J = \sum_{i=1}^{n} \sum_{j=1}^{k} r_{ij} ||x_i - \mu_j||_2^2$$

Python Library Used

sklearn.cluster

RYE COUNTRY DAY SCHOOL
1869

Model Adaptability

RYE COUNTRY DAY SCHOOL
1869

Evaluate K-means Clustering

Calculate Average and Maximum Distances

$$Avg\ Distance\ =\ \frac{1}{n(n-1)}\sum_{i<j} distance(i,j)$$

$$Max\ Distance\ =\ max(i<j)\ distance(i,j)$$

Score To Compare

$$Score\ =\ 0.7\cdot Avg\ Distance\ +\ 0.3\cdot Max\ Distance$$

Visualization of the Conferences

RYE COUNTRY DAY SCHOOL
1869

Scheduling

The Approach

Minimize the Complexity of the Problem (Conference Creation)

↓

Identify the Constraints

↓

Develop the Code

↓

Test the Constraints

↓

Adjust the Code if Certain Constraints are Unachievable

↓

Explain the Code (Algorithims)

Identify the Constraints

Minimize the Distance.

Each team plays exactly 1 game per week.

Each team plays 9 home games.

Each team plays each divisional opponent twice.

Prevent inter-divisional games during intra-divisional game weeks.

No back-to-back games between the same team.

Limit of three home games and three away games in a row.

No more than 35,000 kilometers traveled per week.

Variables

Variables	Meaning
i,j	Both *i* and *j* are a set of teams (1-20)
$d_{i,j}$	Distance between teams *i* and *j*
$x_{i,j,w} \in \{0,1\}$	1 if team *i* hosts team *j* in week *w*, 0 otherwise
$h_{i,w} \in \{0,1\}$	1 if team *i* is home in week *w*, 0 otherwise
T	The collection of all 20 teams (e.g. *T* = 1,2,3…20)
w	A specific week
W	The set of all scheduled weeks (3-11,14-22)
w+k	A week relative to w, used to check consecutive weeks
$\forall i$	Repeats equation for every value of i

RYE COUNTRY DAY SCHOOL
1869

Mathematical Representaitons of Constraints (Part One)

1. Each team plays exactly one game in one week.

$$\sum_{j \neq i} (x_{i,j,w} + x_{j,i,w}) = 1 \qquad \forall i \in T,\ w \in W$$

2. Each team plays 9 home games.

$$\sum_{w \in W} \sum_{j \neq i} x_{i,j,w} = 9 \qquad \forall i \in T$$

3. Play each divisional opponent twice (1 Home, 1 Away)

$$\sum_{w \in W} x_{i,j,w} = 1 \qquad \forall i, j \in Western,\ i \neq 1$$

$$\sum_{w \in W} x_{j,i,w} = 1 \qquad \forall i, j \in Western,\ i \neq 1$$

RYE COUNTRY DAY SCHOOL 1869

Mathematical Representaitons of Constraints (Part Two)

4. Preventing inter-divisional games in intra-divisional gameweeks

$$x_{i,j,w} = 0 \qquad \forall i \in Western,\ j \in Eastern,\ w \in W$$
$$x_{j,i,w} = 0 \qquad \forall i \in Western,\ j \in Eastern,\ w \in W$$

5. No back-to-back (home) games between the same team

$$x_{i,j,w} + x_{i,j,w+1} \leq 1 \qquad \forall i, j \in T, i \neq j, w \in W$$
$$x_{j,i,w} + x_{j,i,w+1} \leq 1 \qquad \forall i, j \in T, i \neq j, w \in W$$

6. Track the number of home games (more of a variable)

$$\sum_{j \neq i} x_{i,j,w} = h_{i,w} \qquad \forall i \in T, w \in W$$

Mathematical Representaitons of Constraints (Part Three)

7. Consecutive home / away limit

$$\sum_{k=0}^{3} h_{i,w+k} \leq 3 \qquad \forall i \in T, w \in W \text{(up to week 19)}$$

$$\sum_{k=0}^{3} (1-h_{i,w+k}) \leq 3 \qquad \forall i \in T, w \in W \text{(up to week 19)}$$

8. Set the travel limit for consecutive game weeks.

$$\sum_{j \neq i} 2d_{i,j} \cdot x_{j,i,w} + \sum_{j \neq i} 2d_{i,j} \cdot x_{j,i,w+1} \leq 35{,}000 \qquad \forall i \in T, w \in W \text{ (except last week)}$$

Code it.

The Python Library

PuLP

RYE COUNTRY DAY SCHOOL
1869

How PuLP Works

PuLP is a Python Linear Programming library.

Linear Programming Minimization is a way of finding the optimal solution for a problem that can be expressed using linear relationships.

You define a problem variable which is equal to LpProblem(“title of problem”, minimize / maximize), and then add constraints to the problem.

We then use the .solve() function to solve the problem.

$$Minimize \sum_{i \in T} \sum_{j \in T, j \neq i} \sum_{w \in W} d_{i,j} \cdot x_{i,j,w}$$

RYE COUNTRY DAY SCHOOL 1869

Intra Division Schedule Results

18 Gameweeks

10 teams per Division

Minimal Distance

Games at 7pm of home team

Western Division games played on Saturday

Eastern Division games played on Saturday

Inter Division Matches

The Format

Regular Season

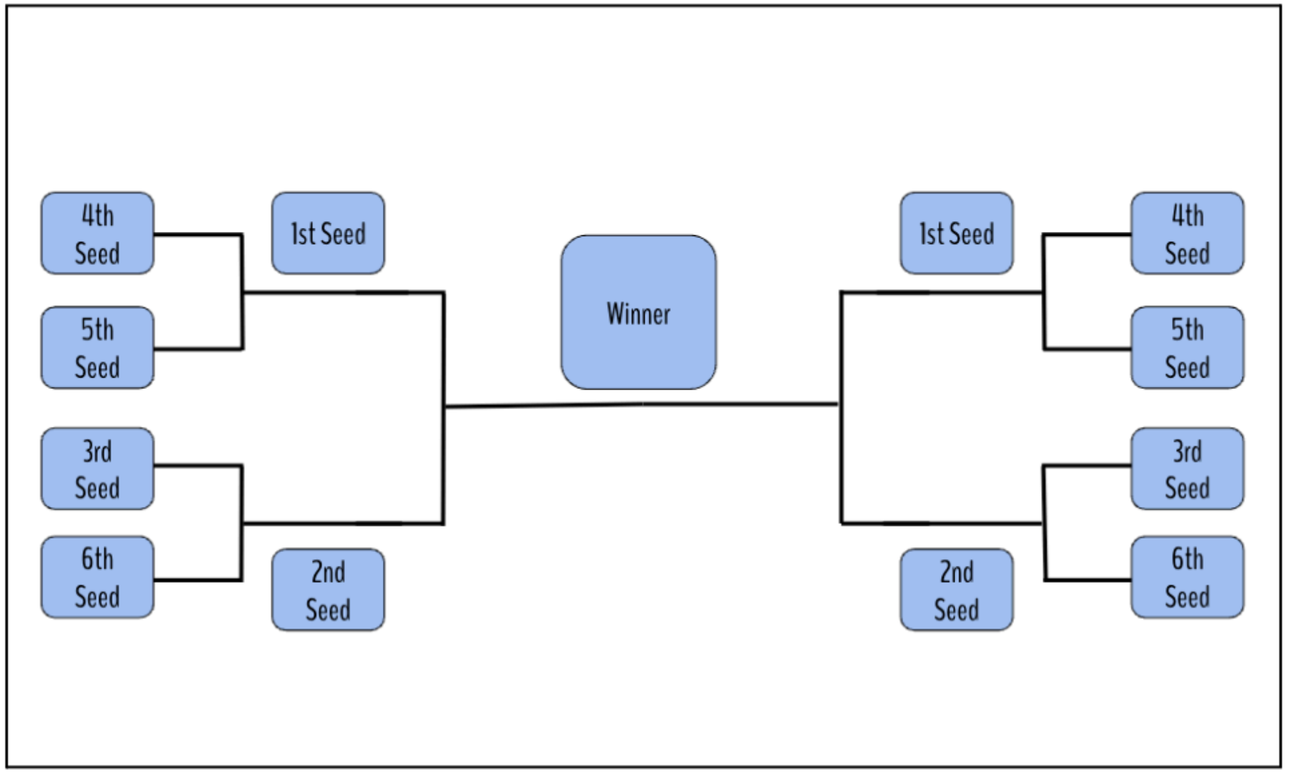
22 Matchweeks

18 intra-conference games

4 interconference games

The Format

Post Season



Expanding the Model

Expansion Teams

Goal: Add teams from continents with fewer amount of teams, that have a large potential for popularity, while maintaining high TPR

Scheduling Changes

Matchweeks: 22→28

18 intraconference games
(9 home, 9 away) :

$$\sum_{w \in W} \sum_{j \in T} x_{i,j,w} = 9 \qquad \forall i \in T$$

Travel Constraint Relaxation

35000km to 40,000km in consecutive weeks

$$\sum_{j \neq i} 2d_{i,j} \cdot x_{j,i,w} + \sum_{j \neq i} 2d_{i,j} \cdot x_{j,i,w+1} \leq 40{,}000 \quad \forall i \in T, w \in W \text{ (except last week)}$$

Conference Changes

2 Conferences of 12 teams

Result: East-West Split

Reflection

Thank you!