Surf Scores

Predicting WSL competition results...

Yew!

Hello! My name is Andres Perez.
I am a Data Scientist
aperezperret.com
andresignacioperez@gmail.com

Outline

1

Problem Statement

Can we predict average heat scores in a Championship Tour competition, using only buoy data?

2

Data

Insights from gathering 10+ years of data

3

Results

Data-modeling results, conclusions and recommendations.

First a little context:

Competitive surfing:

- Highest level **was** the Championship Tour (CT)
- Organized by the World Surf League (WSL)
- Tournament style contests at each major ‘break’
- Sum of **two best waves** is the final score
- Judging criteria

1

Problem Statement

Can we predict average heat scores in a Championship Tour competition, using only buoy data?

Status Quo

Surf Forecasts

Surfline supplies very sophisticated surf forecasts

Human ‘Call’

WSL officials decide to go based on forecasts and qualitative info

Scores?

Scoring potential is considered based on forecasted waves only

Opportunity!

Predictive models that incorporate scores can be leveraged to optimize for scoring potential, increasing broadcast values.

2

Data

What we found in 14 years of scores around the world

Data Sources

WSL

Average Scores per Athlete per Heat, going back to 2008

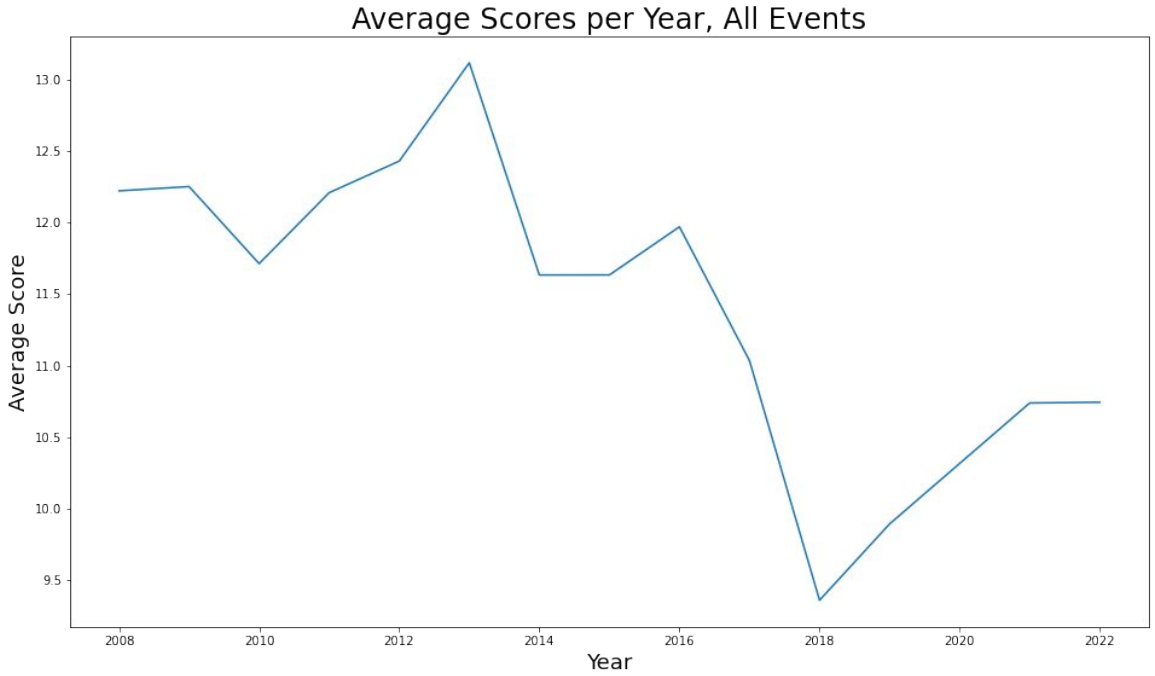
NOAA

National Buoy Data Center historical data for wind and waves.

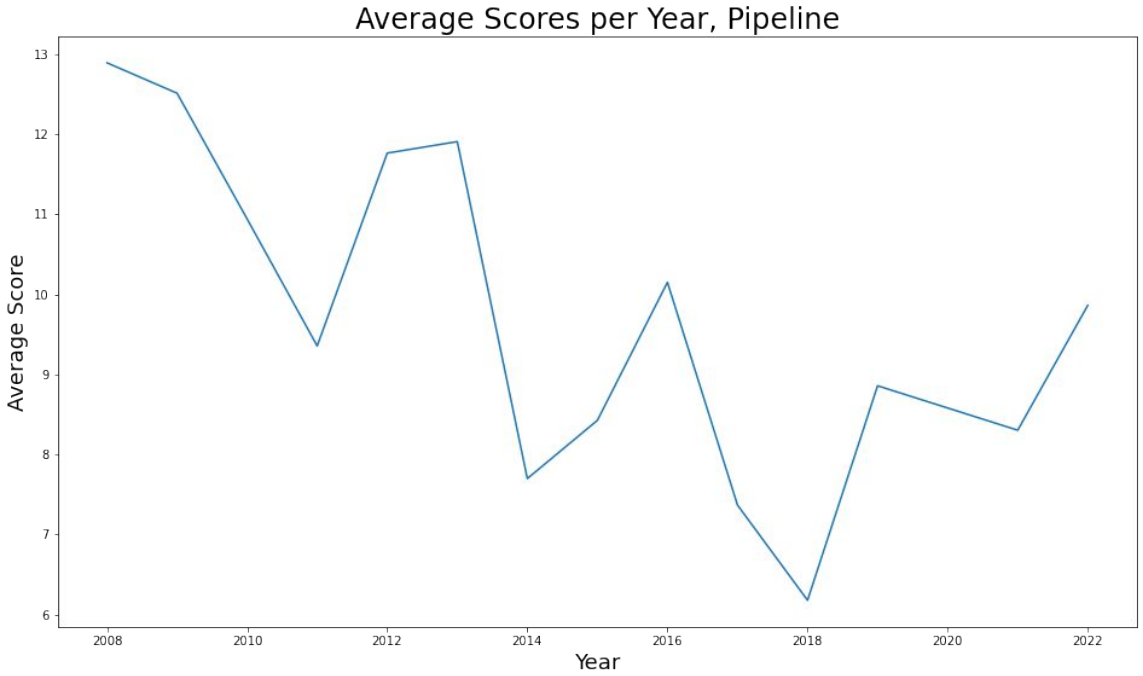
Most common 'venues'

1. **Pipeline**
   Over 800 heats surfed since 2008
2. **Gold Coast**
   ~ 700 heats surfed
3. **France**
   ~ 700
4. **Trestles**
   ~ 500, **with buoy data**

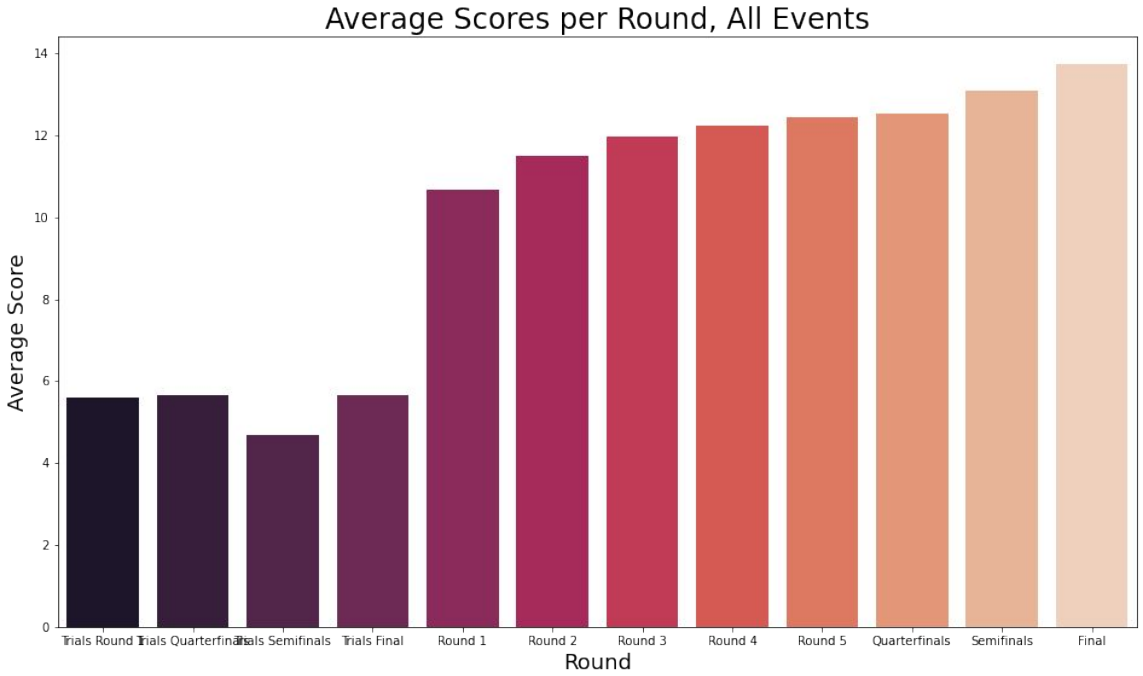
All-time average scores



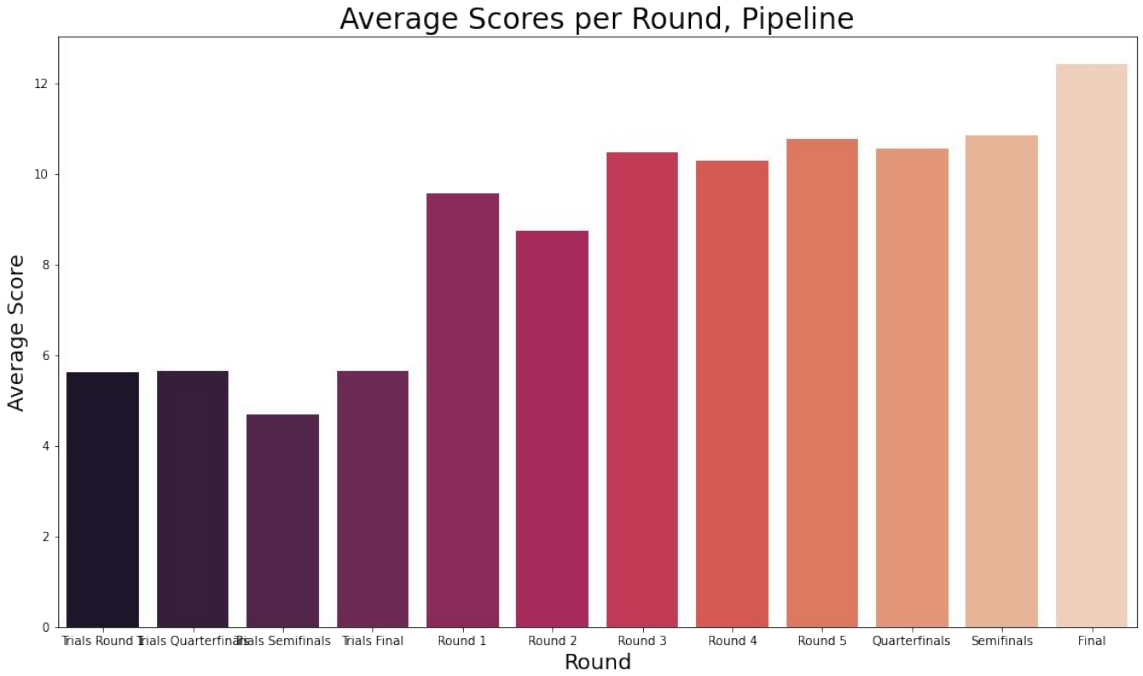
All-time average scores – Pipeline



All-time average scores – per Round

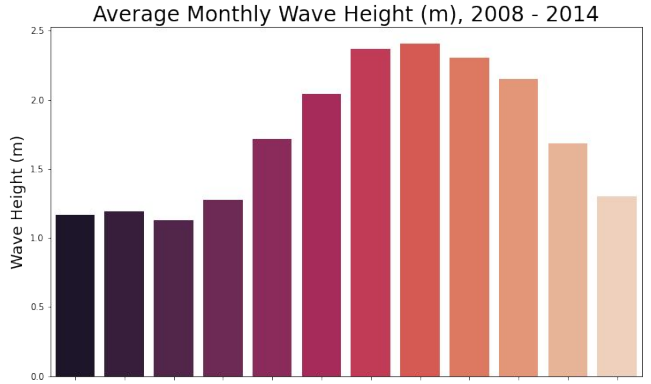


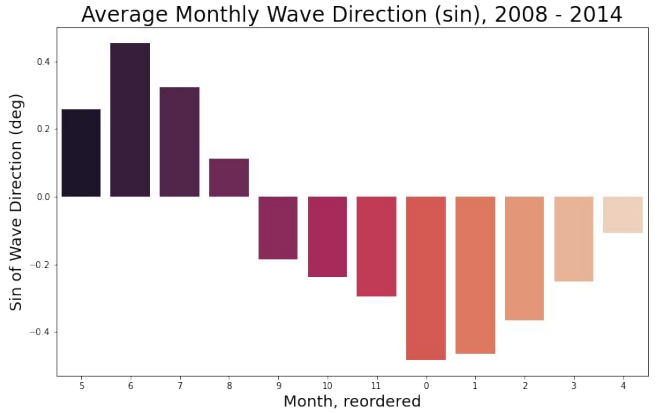
All–time average scores – per Round, Pipeline



Winter waves

Wave seasonality at Pipeline



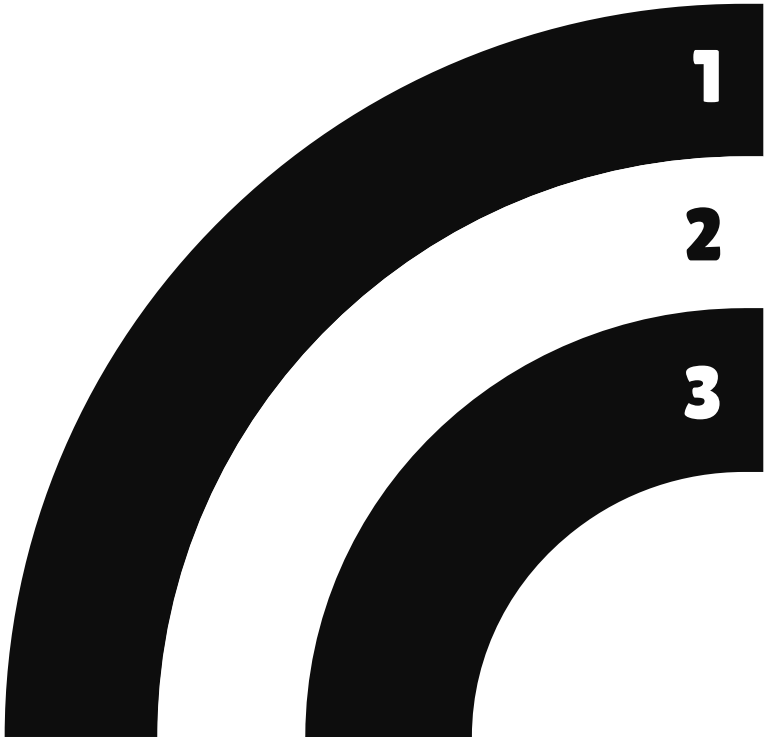


3

Results

It’s about to get *slightly* technical

Three types of models were used...



1. Linear Regression
2. eXtreme Gradient Boosting
3. Neural Networks

... each with its trade-offs...

1. LinReg is interpretable and easy to use
2. XGBoost is less interpretable
3. NN, difficult to tune and data hungry

... and in the end they promised hope!

Baseline RMSE	3.167511

	RMSE train	RMSE test
LinReg (Lasso)	2.923231	2.820082
XGBoost	1.719407	2.871119
Neural Net	2.419217	2.856377

A vision of the future

Live buoy data

Updated every 30 minutes at NOAA

More locations

Trestles is the next target

Computer vision

Can Neural Nets *see* waves?

Wearables

Track athlete's movements

Thank you!

Do you have any questions?

andresignacioperez@gmail.com
+1 415 565 9040
aperezperret.com