moySurf

A comparative analysis on WSL surfing data. Beat your mates at fantasy surf with statistics Kai Moy 3/14/22
Senior Project

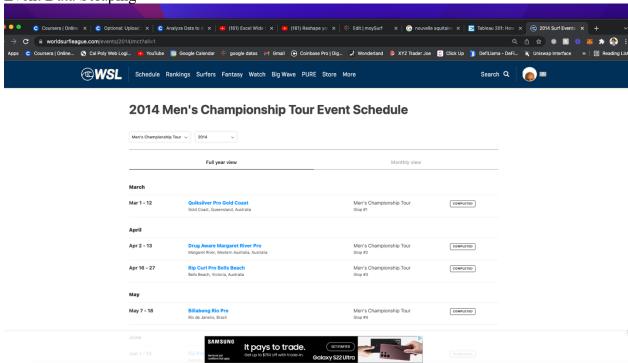
Event Log

- Gather
- Clean
 - o Remove unnecessary variables
 - o Unmerge cells
 - o Remove empty fields
 - o remove hyperlinks
 - Substring values and delimit
 - Start_date=DATE(B2,VLOOKUP(LEFT(C2,3),Sheet4!\$A\$2:\$B\$13,2,FALS E),MID(C2,5,2))
 - End_date=DATE(B2,VLOOKUP(LEFT(C2,3),Sheet4!\$A\$2:\$B\$13,2,FALSE),RIGHT(C2,2))
 - o Split String based on lowercase to capital letter delimiter.
 - Quiksilver Pro Gold CoastGold Coast, Queensland, Australia
 =LEFT(A1,MATCH(2,MMULT(N(ABS({109.5,77.5}-CODE(MID(MID(A1,ROW(INDIRECT("1:"&LEN(A1)-1)),2),{1,2},1)))<13),{1;1}),0))
 - Quiksilver Pro Gold Coast
 - o Keep one summary of all events with specific heat id
 - o Remove duplicates for an encompassing list of all events

Data Details:

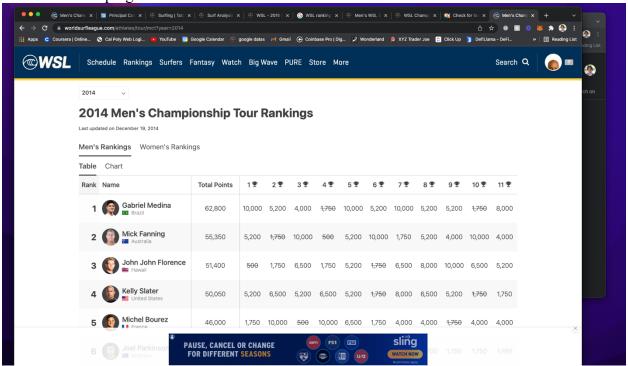
- Dates included in study 2014-2019
- Only Male surfers
- WSL judge scores per event
- 11 events per season
- Event location and order cycle each year

Event Data Scraping



https://www.worldsurfleague.com/events/2014/mct?all=1

Score Data Scraping

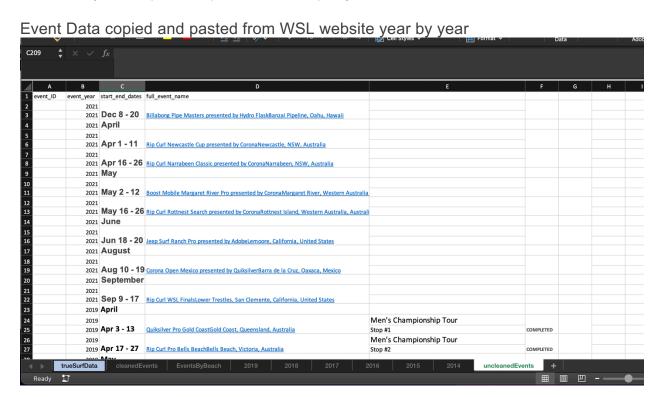


https://www.worldsurfleague.com/athletes/tour/mct?year=2014

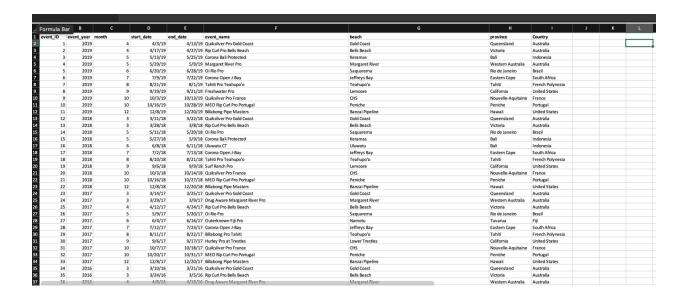
- Points in **Red** indicate athlete withdrew or was injured from the event.
- Points with Strikethrough are events that have been dropped from the athlete's points total.
- Points with *Asterisk indicate athlete was a replacement for the event.

Assumptions:

- Beach: CHS = Capbreton / Hossegor / Seignosse
- Throwaways and injured points will stay within the data set
- Only use top 40-50 placed surfers per year



Event Data after thorough cleaning



Formatting Score Data was the same process of copying and pasting from WSL site

										Cell Styles		1 11	ilat •		Data
$\stackrel{\blacktriangle}{_{\blacktriangledown}}$ × \checkmark f_X 4745															
A	В	С	D	E	F	G	н	- 1	J	К	L	М	N	0	Р
ne	1	2	3	4	5	6	7	8	9	10	11	year			
o Ferreira	10,000	4,745	1,330	4,745	1,330	7,800	1,330	3,320	7,800	10,000	10,000	2019			
riel Medina	4,745	4,745	1,330	1,330	4,745	10,000	7,800	10,000	3,320	3,320	7,800	2019			
ly Smith	6,085	6,085	1,330	4,745	7,800	3,320	6,085	4,745	3,320	7,800	1,330	2019			
e Toledo	3,320	7,800	4,745	1,330	10,000	6,085	3,320	7,800	1,330	4,745	1,330	2019			
he Andino	7,800	1,330	4,745	7,800	6,085	6,085	1,330	1,330	4,745	4,745	3,320	2019			
oa Igarashi	3,320	3,320	10,000	3,320	4,745	4,745	1,330	3,320	1,330	6,085	1,330	2019			
n John Florence	6,085	10,000	1,330	10,000	4,745	265	265	265	265	265	4,745	2019			
y Slater	265	4,745	6,085	3,320	3,320	3,320	1,330	3,320	1,330	3,320	6,085	2019			
en Wright	3,320	3,320	1,330	3,320	1,330	4,745	10,000	6,085	1,330	1,330	1,330	2019			
my Flores	1,330	3,320	7,800	1,330	265	1,330	4,745	1,330	10,000	1,330	265	2019			
n Wilson	1,330	1,330	3,320	6,085	4,745	1,330	3,320	4,745	3,320	1,330	3,320	2019			
n Moniz	4,745	3,320	1,330	4,745	1,330	1,330	6,085	1,330	3,320	265	3,320	2019			
hel Bourez	3,320	1,330	3,320	3,320	3,320	3,320	3,320	265	3,320	1,330	4,745	2019			
n Callinan	1,330	6,085	3,320	4,745	1,330	3,320	1,330	1,330	4,745	265	1,330	2019			
Freestone	1,330	1,330	3,320	1,330	1,330	1,330	3,320	1,330	6,085	4,745	4,745	2019			
fin Colapinto	1,330	265	1,330	1,330		1,330	3,320	6,085	1,330	3,320	6,085	2019			
Ibelli	265*	1,330*	1,330*	6,085*	1,330*	1,330*	4,745*	1,330	1,330*	6,085*	3,320*	2019			
de Carmichael	4,745	1,330	4,745	265	3,320	1,330	1,330	3,320	3,320	3,320	1,330	2019			
an Buchan	1,330	1,330	4,745	265	265	4,745	1,330	4,745	4,745	1,330	1,330	2019			
ner Coffin	4,745	3,320	3,320	3,320	1,330	1,330	1,330	1,330	1,330	3,320	1,330	2019			
erson Crisanto	1,330	3,320	1,330	3,320	265	3,320	1,330	265	1,330	4,745	3,320	2019			
o Dora	3,320	1,330	1,330	1,330	1,330	1,330	1,330	4,745	3,320	1,330	4,745	2019			
vid Silva	1,330	3,320	1,330	1,330	3,320	3,320	3,320	3,320	265	1,330	1,330	2019			
ian Cardoso	3,320	3,320	1,330	1,330	1,330	3,320	1,330	3,320	1,330	1,330	1,330	2019			
e Mendes	1,330	265	3,320	1,330	3,320	265	1,330	1,330	1,330	3,320	3,320	2019			
hael Rodrigues	1,330	1,330	6,085	1,330	3,320	1,330	265	265	1,330	3,320	265	2019			
astian Zietz	1,330	265	265	3,320	1,330	4,745	1,330	3,320	265	1,330	1,330	2019			
kiel Lau	1,330	1,330	265	1,330	1,330	3,320	1,330	3,320	3,320	265	1,330	2019			

- Dashed out values turned to 0
- Slashes removed from values

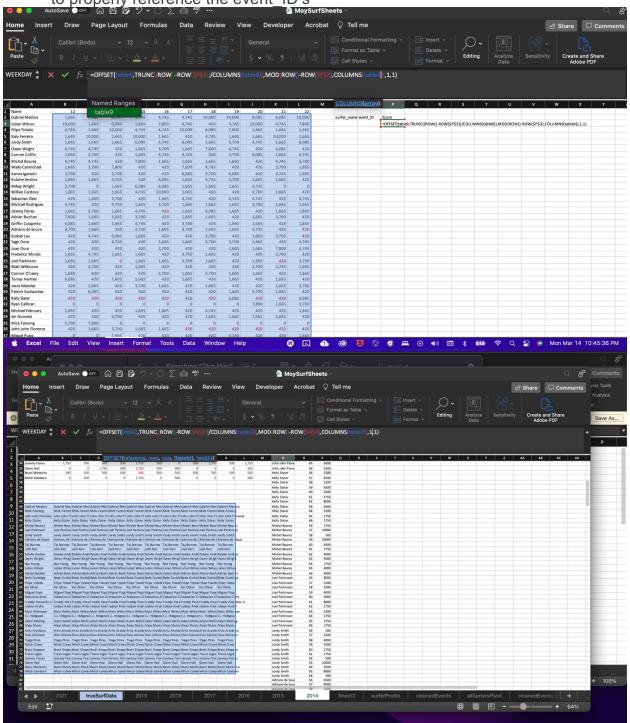
=TRANSPOSE(A57:A67)

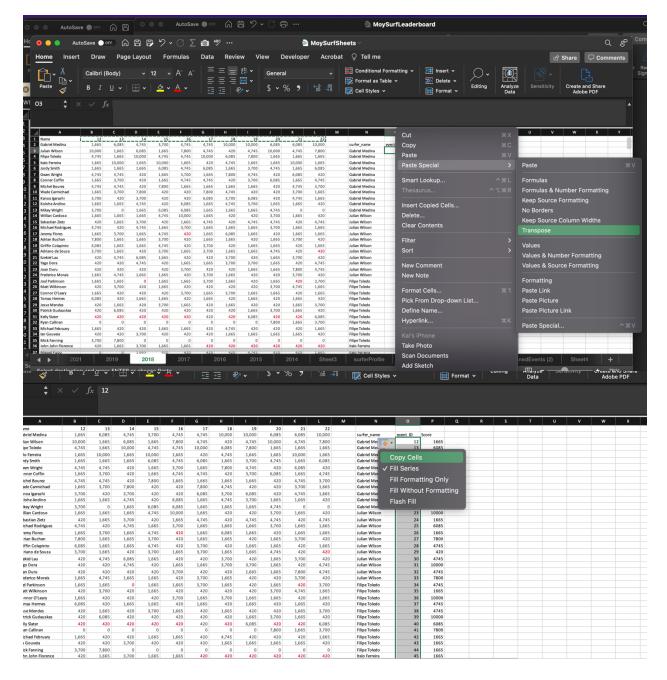
Used for placing vertical event_ID's horizontal to match scoring data.

= OFFSET(table 1, TRUNC((ROW()-ROW(\$P\$3))/COLUMNS(table 1)), MOD(ROW()-ROW(\$P\$3), COLUMNS(table 1)), 1, 1)

Used OFFSET function and table matrixes to transpose wide data to long data

*This was by far, the most difficult and headaching part of the data wrangling process. This became such an issue because the data which I scraped manually from the WSL website was in a wide fashion and I needed long data to properly reference the event_ID's

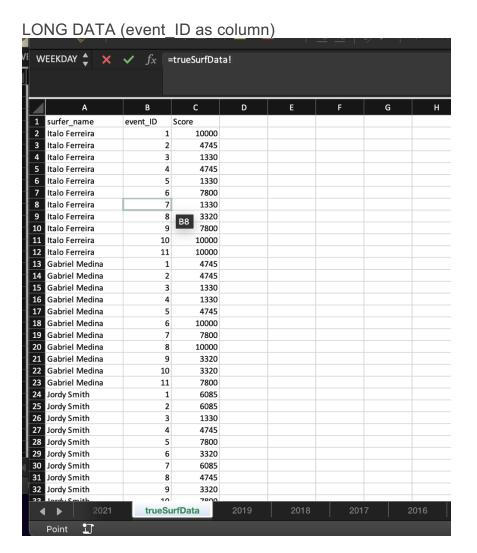




- I repeated this transposing process on the 6 data sets that I had for scoring.
 2014-2019
- Combined them all into one single csv file.

WIDE DATA before transposing (event ID on first row)

A	В	С	D	E	F	G	н	1	J	К	L	М
Name	45	46	47	48	49	50	51	52	53	54	55	
Adriano de Souza	6,500	8,000	10,000	1,750	1,750	5,200	1,750	8,000	6,500	1,750	10,000	
Mick Fanning	5,200	10,000	1,750	4,000	4,000	8,000	1,750	10,000	5,200	1,750	6,500	
Gabriel Medina	1,750	5,200	500	1,750	1,750	5,200	8,000	6,500	10,000	5,200	8,000	
Filipe Toledo	10,000	5,200	C3 ₅₀₀	10,000	1,750	1,750	4,000	6,500	500	10,000	1,750	
Owen Wright	1,750	5,200	4,000	5,200	10,000	1,750	6,500	4,000	5,200	500	500	
Julian Wilson	8,000	1,750	5,200	500	8,000	8,000	1,750	1,750	6,500	500	1,750	
Italo Ferreira	4,000	500	1,750	6,500	5,200	1,750	5,200	4,000	5,200	8,000	1,750	
Jeremy Flores	500	4,000	4,000	1,750	6,500	500	10,000	1,750	4,000	5,200	4,000	
Kelly Slater	1,750	4,000	5,200	1,750	4,000	6,500	5,200	4,000	1,750	1,750	5,200	
Nat Young	1,750	6,500	6,500	1,750	1,750	4,000	500	5,200	1,750	4,000	500	
Josh Kerr	1,750	6,500	4,000	5,200	500	1,750	5,200	1,750	1,750	1,750	5,200	
Bede Durbidge	5,200	500	500	8,000	1,750	1,750	1,750	1,750	8,000	1,750	1,750	
Joel Parkinson	1,750	4,000	1,750	500	5,200	1,750	1,750	5,200	500	5,200	4,000	
John John Florence	1,750	1.750	8,000	4,000	500	500	1,750	500	5,200	1.750	4,000	
Wiggolly Dantas	5,200	500	500	1,750	5,200	4,000	4,000	5,200	500	500	500	
Tai Burrow	5,200	1.750	6,500	1,750	6,500	500	500	1,750	0	0	1.750	
Kai Otton	500	500	1.750	1.750	5,200	5.200	5,200	1,750	1.750	500	1.750	
Matt Wilkinson	4,000	1.750	500	6,500	500	1,750	1,750	1,750	4,000	1.750	500	
Adrian Buchan	500	500	1.750	1,750	1.750	6,500	1,750	5,200	1,750	500	1,750	
Keanu Asing	500	1.750	500	4,000	1,750	4,000	500	500	1,750	4.000	4,000	
Michel Bourez	500	500	5,200	500	500	4,000	500	1,750	1,750	4,000	1,750	
Jadson Andre	500	4,000	1,750	5,200	1,750	500	1,750	500	4,000	500	500	
Adam Melling	500	1,750	1,750	1,750	4,000	1,750	500	1,750	500	500	5,200	
C.J. Hobgood	500	500	500	500	500	1,750	6,500	500	1,750	1.750	5,200	
Kolohe Andino	1,750	1,750	500	500	1,750	1,750	500	1,750	4,000	4,000	5,200	
Sebastian Zietz	1,750	1,750	4,000	1.750		500	1,750	500	500	1,750	1.750	
	,	500	,	500	1,750 500	500	500	4.000	500	500	500	
Miguel Pupo	6,500		1,750	500	500	500	500		500	500		
Jordy Smith	4,000 500	5,200	1,750		500			500			1,750	
Brett Simpson		1,750	500	500		500	1,750	500	1,750	6,500	500	
Glenn Hall	4,000	500	1,750	500	500	500	500	1,750	500	500	1,750	
Ricardo Christie	500	500	1,750	5,200	500	500	500	500	500	1,750	500	
Freddy Patacchia Jr.	1,750	1,750	1,750	500	1,750	1,750	500	1,750	0	0	0	
Matt Banting	1,750	1,750	500	4,000	500	500	500	500	500	500	500	
Mason Ho	0	1,750	0	0	0	0	0	0	0	1,750	6,500	
Alejo Muniz	0	0	500	500	1,750	5,200	0	0	500	0	0	
Dusty Payne	1,750	500	500	1,750	500	500	1,750	500	500	500	500	
Dane Reynolds	500	0	0	0	4,000	1,750	0	0	1,750	0	0	
Vasco Ribeiro	0	0	0	0	0	0	0	0	0	6,500	0	
Aritz Aranburu	0	0	0	0	500	0	4,000	500	500	500	0	
Jay Davies	0	0	5,200	0	500	0	0	0	0	0	0	
Frederico Morais	0	0	0	0	0	0	0	0	0	5,200	0	
Bruno Santos	0	0	0	0	0	0	4,000	0	0	0	0	
Tomas Hermes	0	0	0	0	0	500	0	500	1,750	500	0	
Caio Ibelli	0	0	0	0	0	0	0	0	500	1,750	0	



In Tableau

Imported cleanedEventDetails.csv and trueSurfData.csv Join data frames where event_ID = event_ID

- Various Worksheets with Surfer scoring, (AVG, Total)
- Filtering for years and surfers
- From total score sheet, the champion of the season is the highest total score.
- Data insights based on geographic variables

Check out my Viz!

https://public.tableau.com/views/moySurfComplete/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link