A photograph of a Crossfit competition. In the foreground, a shirtless man with a bun hairstyle and tattoos on his arms is running towards the right. Behind him, several other athletes are running, including a man with a beard and tattoos, a man in an orange tank top, and a woman in a white tank top with 'LEADER' printed on it. The background shows a crowd of spectators and some equipment like dumbbells.

Crossfit

Data - science project

MADE BY : BEN TURGEMAN, ORI RONEN

30-01-22

Research question

- Is it possible to predict which athletes will reach the final -“games” based on date that collects during the “open” - the first stage.

background and motivation

- CrossFit is a branded fitness regimen that involves constantly varied functional movements performed at high intensity
- incorporating elements from high-intensity interval training, Olympic weightlifting, plyometrics, powerlifting, gymnastics, kettlebell lifting, calisthenics, strongman
- The CrossFit Games are styled as a venue for determining the "Fittest on Earth," where competitors should be "ready for anything."
- The first stage the Open has been described as the largest participatory sporting event in the world over 200,000 competitors
- Why CrossFit?

Big -date:

- volume
- velocity
- variety
- volatility
- Veracity

thousand Competitors
each and everyone has
is own profile

Cross-fit - classic case of big date

OPEN	START	LEADERBOARDS	WORKOUT	COMPETITION	STORIES	THE SPORT OF FITNESS
RANK	NAME	POINTS	21.1	21.2	21.3	21.4
1	- JEFFREY ADLER	101	20th (11:55) 605 reps Judged by Caroline Lambray at CrossFit Wonderland.	8th (9:14) 225 reps Judged by Caroline Lambray at CrossFit Wonderland.	27th (8:15) 180 reps Tiebreak: 8:15 Judged by Caroline Lambray at CrossFit Wonderland.	46th (317 lbs) Weight lifted: 317 lbs. Judged by Caroline Lambray at CrossFit Wonderland.
2	- SCOTT PANCHIK	141	33rd (12:25) 605 reps Judged by Christin Panchik at CrossFit Mentality.	47th (9:52) 225 reps Judged by Christin Panchik at CrossFit Mentality.	5th (7:48) 180 reps Tiebreak: 7:48 Judged by Christin Panchik at CrossFit Mentality.	56th (316 lbs) Weight lifted: 316 lbs. Judged by Christin Panchik at CrossFit Mentality.
3	+ TRAVIS MEAD	165	87th (13:02)	24th (9:38)	48th (8:26)	6th (345 lbs)
4	+ SAXON PANCHIK	217	5th (11:25)	68th (9:59)	87th (8:41)	57th (316 lbs)
5	+ RICHARD FRONING JR.	254	58th (12:45)	91st (10:04)	5th (7:48)	100th (312 lbs)
6	+ NOAH OHLSEN	272	11th (11:41)	21st (9:33)	2nd (7:44)	238th (301 lbs)
7	+ SAMUEL COURNOYER	276	158th (13:27)	79th (10:02)	20th (8:09)	19th (328 lbs)
8	+ COLE GREASHABER	403	82nd (13:01)	68th (9:59)	46th (8:25)	207th (303 lbs)
9	+ AGUSTIN RICHELME	417	30th (12:19)	137th (10:15)	1st (7:40)	249th (300 lbs)
10	+ OLEG LASCENKO	428	53rd (12:43)	22nd (9:34)	102nd (8:46)	251st (300 lbs)
11	+ CÉDRIC LAPOINTE	454	62nd (12:49)	9th (9:16)	27th (8:15)	356th (296 lbs)
12	- GIORGOS	510	207th (12:40)	42nd (9:51)	129th (8:54)	122nd (200 lbs)

Four events, all the events are variety in content and score.

Stages

1

crawling

2

Data Exploration

3

Data Engineering

4

Eda-visualization

4

Machine Learning

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X		
1																										
2	0	Jeffrey	Adler	101	20th (11:55)	8th (9:14)	7th (9:31)	42nd (317 lbs)	16th (7:41)	24th (9:31)	8th (8:41)	57th (316 lbs)	25	161 cm	175 lb	177 lb	90 lb	567 lb	511	54	2:02	1:16	7:18	15:17	0:59	19:15
3	1	Scott	Panchik	141	33rd (12:25)	47th (9:52)	47th (7:41)	10th (11:46)	11th (8:46)	24th (9:31)	8th (8:41)	57th (316 lbs)	25	161 cm	175 lb	177 lb	--	--	--	--	--	--	--	--	--	--
4	2	Travis	Mead	165	87th (13:02)	24th (9:31)	8th (8:41)	100th (312 lbs)	34	5'9"	180 lb	470 lb	360	195 lb	285 lb	560 lb	454	60	1:58	1:11	6:20	15:26	1:10	20:12		
5	3	Saxon	Panchik	217	5th (11:25)	68th (9:59)	68th (8:41)	100th (312 lbs)	34	5'9"	180 lb	470 lb	360	195 lb	285 lb	520 lb	--	62	2:07	1:09	7:18	--	0:52	19:12		
6	4	Richard	Froning Jr.	254	58th (12:45)	91st (10:04)	5th (7:48)	100th (312 lbs)	34	5'9"	194 lb	475 lb	370	195 lb	305 lb	570 lb	508	75	2:13	1:11	--	--	--	--		
7	5	Noah	Chilsen	272	11th (11:41)	21st (9:33)	2nd (7:44)	238th (301 lbs)	30	5'7"	190 lb	465 lb	355	195 lb	285 lb	540 lb	421	72	1:58	1:16	--	19:05	0:59	18:50		
8	6	Samuel	Cournoyer	276	158th (13:27)	79th (10:02)	20th (8:09)	19th (328 lbs)	26	180 cm	205 lb	435 lb	370	195 lb	305 lb	530 lb	--	--	2:05	1:30	--	--	--	--		
9	7	Cole	Greashaber	403	82nd (13:01)	68th (9:59)	46th (8:25)	207th (303 lbs)	22	6'0"	206 lb	--	350	195 lb	275 lb	--	--	--	--	--	--	--	--	--		
10	8	Agustin	Richelme	417	30th (12:19)	137th (10:15)	1st (7:40)	249th (300 lbs)	24	173 cm	83 kg	195 kg	160	125 kg	220 kg	--	79	1:58	--	--	--	--	--	--		
11	9	Oleg	Lascenko	428	53rd (12:43)	22nd (9:34)	102nd (8:46)	251st (300 lbs)	24	175 cm	92 kg	--	--	--	--	--	--	--	--	--	--	--	--	--		
12	10	David	Opdyke	457	11th (11:46)	11th (9:59)	13th (8:41)	15th (316 lbs)	25	171 cm	175 lb	455 lb	355	125 kg	275 lb	--	70	1:47	--	--	--	--	--	--		
13	11	Giovanni	Kralis	519	17th (12:51)	12th (10:51)	13th (8:41)	15th (316 lbs)	27	170 cm	170 lb	420	355	120 kg	140 kg	--	50	2:18	1:40	--	--	--	--	--		
14	12	Stas	Solodov	548	111th (13:14)	2nd (8:52)	67th (8:35)	368th (295 lbs)	37	175 cm	85 kg	160 kg	115	95 kg	200 kg	--	50	2:30	2:15	9:00	--	1:05	21:05	--		
15	13	Jonne	Koski	564	1st (11:08)	109th (10:09)	25th (8:14)	429th (293 lbs)	27	171 cm	185 lb	--	--	--	--	--	--	--	--	--	--	--	--	--		
16	14	Elliot	Simmonds	582	308th (13:57)	36th (9:47)	68th (8:36)	170th (305 lbs)	28	185 cm	87 kg	405 lb	315	125 kg	265 lb	485 lb	--	--	--	--	--	--	--	--		
17	15	Logan	Collins	589	53rd (12:43)	358th (10:42)	15th (8:06)	163rd (306 lbs)	30	5'7"	182 lb	455 lb	340	125 kg	275 lb	535 lb	--	90	2:02	1:05	--	--	--	18:00		
18	16	Colin	Gough	611	9th (11:40)	50th (9:53)	18th (8:07)	534th (289 lbs)	22	175 cm	82 kg	--	--	--	--	--	--	--	--	--	--	--	--	--		
19	17	John	McLennan	621	248th (13:47)	3rd (9:04)	15th (8:06)	355th (296 lbs)	23	5'4"	185 lb	--	--	--	--	--	--	--	--	--	--	--	--	--		
20	18	Jayson	Hopper	630	308th (13:57)	165th (10:20)	132nd (8:53)	25th (326 lbs)	24	6'1"	215 lb	475 lb	355	125 kg	295 lb	500 lb	408	--	2:15	2:00	--	0:52	20:05	--		
21	19	Patrick	Vellner	638	47th (12:39)	13th (9:27)	4th (7:46)	574th (287 lbs)	31	5'11"	195 lb	455 lb	355	125 kg	285 lb	595 lb	389	54	2:14	1:17	9:46	20:16	1:06	18:50		
22	20	Alex	Kotoulas	642	23rd (12:05)	12th (9:22)	30th (8:17)	577th (287 lbs)	28	170 cm	78 kg	180 kg	135	110 kg	210 kg	--	--	--	--	--	--	--	--	--		
23	21	Brent	Fikowski	644	15th (11:48)	41st (9:50)	231st (9:10)	357th (295 lbs)	30	6'2"	220 lb	--	--	300	120 kg	--	--	--	--	--	7:08	--	--	--		
24	22	Hunter	Hallfield	708	230th (13:41)	142nd (10:31)	12th (8:22)	162nd (312 lbs)	24	5'12"	195 lb	485 lb	345	120 kg	300 lb	500 lb	--	--	2:05	1:20	--	--	--	--		
25	23	Colby	Alvarez	708	17th (13:03)	146th (10:03)	25th (8:22)	255th (302 lbs)	29	5'4"	180 lb	405 lb	320	120 kg	255 lb	420 lb	489	73	2:06	1:40	7:47	16:18	0:59	22:33		
26	24	Mitch	Wagner	708	134th (13:23)	111th (10:10)	94th (8:44)	369th (295 lbs)	32	5'8"	185 lb	--	--	--	--	--	--	--	1:57	--	--	--	--	--	--	
27	25	Phil	Toon	741	572nd (14:28)	87th (10:03)	58th (8:29)	24th (326 lbs)	23	5'10"	195 lb	575 lb	375	125 kg	275 lb	660 lb	--	--	1:57	1:30	7:30	--	--	--	--	
28	26	Aleksandar	Ilin	748	649th (14:35)	18th (9:30)	44th (8:24)	37th (320 lbs)	31	175 cm	87 kg	185 kg	--	120	230 kg	--	--	2:15	1:30	--	--	--	0:58	21:40	--	
29	27	Joshua	Al-chamaa	749	62nd (12:49)	79th (10:02)	563rd (9:48)	45th (318 lbs)	30	182 cm	206 lb	215 kg	168	130 kg	240 kg	--	50	2:24	--	--	--	--	--	21:00	--	
30	30	Fabian	Beneito	769	238th (13:45)	181st (10:23)	100th (8:45)	250th (300 lbs)	25	178 cm	90 kg	--	--	--	--	--	--	--	--	--	--	--	--	--		
31	31	Sean	Gordon	807	154th (13:26)	55th (9:56)	48th (8:26)	550th (288 lbs)	23	5'8"	185 lb	--	--	--	--	--	--	--	--	--	--	--	--	--		
32	32	Khan	Porter	812	182nd (13:35)	63rd (9:58)	313th (9:23)	254th (300 lbs)	32	183 cm	205 lb	200 kg	160	133 kg	250 kg	468	50	2:01	1:10	--	16:40	--	19:31	--		
33	33	Lazar	Dukić	817	413th (14:10)	15th (9:29)	23rd (8:12)	366th (295 lbs)	26	182 cm	90 kg	--	160	125 kg	--	--	--	--	--	--	--	--	--	--		
34	34	Michael	McKenna	818	124th (12:22)	219th (10:30)	290th (9:31)	174th (310 lbs)	24	5'10"	191 lb	485 lb	350	120 kg	290 kg	464	50	2:27	1:40	8:00	17:50	1:07	22:20	--		
35	35	Lukas	Gulab	842	17nd (13:11)	104th (10:03)	40th (8:22)	549th (288 lbs)	25	5'8"	180 lb	465 lb	365	120 kg	300 lb	500 lb	--	--	--	--	--	--	--	--	--	
36	36	Nick	Mathew	850	19th (13:18)	274th (10:35)	284th (9:19)	173rd (305 lbs)	28	5'9"	200 lb	545 lb	365</td													

Stage 1 - Crawling

- At first we try to scrape the date using the method from the Course
 - But we failed.
 - Then we found out about selenium - headless browser.
 - The methods was - first:
 - scraping/crawling at the main page, at every competitor
 - divine into the his unique profile and scraping his date.
 - At the end we had more then 5.000 competitors (rows), each has 23 colum



File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3

Data Exploration

```
In [ ]: 1 pip install -r requirements.txt
```

```
In [1]: 1 from selenium.webdriver.common.by import By
2 from selenium.webdriver.chrome.options import Options
3 from webdriver_manager.chrome import ChromeDriverManager
4 from selenium.webdriver.support.ui import WebDriverWait
5 from selenium.webdriver.chrome.service import Service
6 from selenium.webdriver.support import expected_conditions as EC
7 from selenium import webdriver
8
9 import numpy as np
10
11 import requests
12 from bs4 import BeautifulSoup
13
14 import pandas as pd
```

```
In [20]: 1 # כניסה מערך עם כל המינים לטבלה אולין הראותית - לנוברים
2 mans_open_main_table_page_links = []
3 temporary_index_number = 2
4 mans_open_main_table_page_links.append("https://games.crossfit.com/leaderboard/open/2021?view=0&division=1&region=0")
5 for _ in range(99): #Enter here how many pages the program will take (+1 -> 3 its 4 pages)
6     mans_open_main_table_page_links.append("https://games.crossfit.com/leaderboard/open/2021?view=0&division=1&region=0")
7     temporary_index_number += 1
8
```

```
In [49]: 1
2 # from the main table
3 first_names = []
4 last_names = []
5 all_scores = []
6 event_21_1 = []
7 event_21_2 = []
8 event_21_3 = []
9 event_21_4 = []
10
11 # from the profile
12 profile_age = []
13 profile_country = []
14 profile_hight = []
15 profile_weight = []
16 year = []
17 division = []
18 rank_worldwide = []
19 rank_by_region = []
20 rank_by_country = []
21 profile_Back_Squat = []
22 profile_Clean_and_Jerk = []
23 profile_Snatch = []
```

```
In [2]: 1 df = pd.read_excel(r'/Users/wryrwnn/Downloads/all_data 1-100-pages.xlsx')
```

```
In [3]: 1 df
Out[3]:
```

	Unnamed: 0	First Name	Last Name	Points	Event 21.1	Event 21.2	Event 21.3	Event 21.4	Age	Country	Snatch	Deadlift	Fight Gone	Max Pull	Fran	Grace	Helen	Filthy Bad Ups
0	0	Jeffrey	Adler	101	20th (11:55)	8th (9:14)	27th (8:15)	46th (317 lbs)	27.0	Nan	290 lb	567 lb	511	54	2:02	1:16	7:18	15:17
1	1	Scott	Panchik	141	33rd (12:25)	47th (9:52)	5th (7:48)	56th (316 lbs)	34.0	Nan	--	--	--	--	--	--	--	
2	2	Travis	Mead	165	87th (13:02)	24th (9:38)	48th (8:26)	6th (345 lbs)	35.0	Nan	285 lb	560 lb	454	60	1:58	1:11	6:20	15:26
3	3	Saxon	Panchik	217	5th (11:25)	68th (9:59)	87th (8:41)	57th (316 lbs)	25.0	Nan	295 lb	520 lb	--	62	2:07	1:09	7:18	--
4	4	Richard	Froning Jr.	254	58th (12:45)	91st (10:04)	5th (7:48)	100th (312 lbs)	34.0	Nan	305 lb	570 lb	508	75	2:13	1:11	--	--
4995	4995	Alejandro	De la Guardia Santana	35067	23406th (378 reps)	6575th (13:15)	1544th (10:52)	3542nd (258 lbs)	23.0	Nan	--	--	--	--	--	--	--	
4996	4996	Christophe	Graux	35069	9192nd (389 reps)	5594th (13:00)	7856th (14:04)	12427th (231 lbs)	27.0	France	100 kg	250 kg	--	--	--	--	--	
4997	4997	Tomasz	Żukowski	35070	5287th (395 reps)	2073rd (11:55)	21867th (155 reps)	5843rd (249 lbs)	32.0	Nan	--	--	--	--	--	--	--	
4998	4998	Dan	Kohutanycz	35080	4108th (433 reps)	6514th (13:14)	3512th (12:12)	20946th (219 lbs)	28.0	Nan	--	--	--	--	--	--	--	
4999	4999	Danny	Kenny	35088	7770th (391 reps)	6243rd (13:10)	6079th (13:24)	14994th (226 lbs)	35.0	Nan	103 kg	220 kg	--	48	2:45	1:48	8:14	--

5000 rows x 24 columns

5000 rows * 24 columns = 120,000

Data Engineering

- Missing values
- The amount off the date
- Most of the contents was mixed - numeric and categorical alike
- The measures was mixed - kg/lbs, cm/inches/

```

1 # drop col:
2
3 cols = ['Fight Gone Bad', 'Max Pull Ups', 'Fran', 'Grace', 'Helen', 'Filthy', 'Sprint', 'Run 5K']
4 for x in cols:
5     df.drop([x], axis=1, inplace=True)
6

```

Most of the date missing

	First Name	Last Name	Points	Event 21.1	Event 21.2	Event 21.3	Event 21.4	Age	Country	Hight	Weight	Back Squat	Clean&Jerk	Snatch	Deadlift	ght Gone Bad	Max Pull Up	Fran	Grace	Helen	Filthy	Sprint	Run 5K
0	Jeffrey	Adler	101	20th (11:55)	8th (9:14)	27th (8:15)	46th (317 lbs)	27	5'9"	197 lb	475 lb	377 lb	290 lb	567 lb	511	54	2:02	1:16	7:18	15:17	0:59	19:15	
1	Scott	Panchik	141	33rd (12:25)	47th (9:52)	5th (7:48)	56th (316 lbs)	34	5'9"	187 lb	--	--	--	--	--	--	--	--	--	--	--	--	--
2	Travis	Mead	165	87th (13:02)	24th (9:38)	48th (8:26)	6th (345 lbs)	35	6'1"	205 lb	465 lb	350 lb	285 lb	560 lb	454	60	1:58	1:11	6:20	15:26	1:10	20:12	
3	Saxon	Panchik	217	5th (11:25)	68th (9:59)	87th (8:41)	57th (316 lbs)	25	5'9"	180 lb	470 lb	360 lb	295 lb	520 lb	--	62	2:07	1:09	7:18	--	0:52	19:12	
4	Richard	Froning Jr.	254	58th (12:45)	91st (10:04)	5th (7:48)	100th (312 lbs)	34	5'9"	194 lb	475 lb	370 lb	305 lb	570 lb	508	75	2:13	1:11	--	--	--	--	--
5	Noah	Ohsen	272	11th (11:41)	21st (9:33)	2nd (7:44)	238th (301 lbs)	30	5'7"	190 lb	465 lb	355 lb	285 lb	540 lb	421	72	1:58	1:16	--	19:05	0:59	18:50	
6	Samuel	Cournoyer	276	158th (13:27)	79th (10:02)	20th (8:09)	19th (328 lbs)	26	180 cm	205 lb	435 lb	370 lb	305 lb	530 lb	--	--	2:05	1:30	--	--	--	--	--
7	Cole	Greashaber	403	82nd (13:01)	68th (9:59)	46th (8:25)	207th (303 lbs)	22	6'0"	206 lb	--	350 lb	275 lb	--	--	--	--	--	--	--	--	--	--
8	Agustin	Richelme	417	30th (12:19)	137th (10:15)	1st (7:40)	249th (300 lbs)	24	173 cm	83 kg	195 kg	160 kg	125 kg	220 kg	--	79	1:58	--	--	--	--	--	--
9	Oleg	Lascenko	428	53rd (12:43)	22nd (9:34)	102nd (8:46)	251st (300 lbs)	24	175 cm	92 kg	--	--	--	--	--	--	--	--	--	--	--	--	--
10	Cédric	Lapointe	454	62nd (12:49)	9th (9:16)	27th (8:15)	356th (296 lbs)	25	5'9"	187 lb	465 lb	355 lb	275 lb	485 lb	--	70	1:47	--	--	--	--	--	--
11	Giorgos	Karavis	510	207th (13:40)	42nd (9:51)	138th (8:54)	123rd (309 lbs)	27	177 cm	90 kg	220 kg	153 kg	120 kg	240 kg	--	50	2:18	1:40	--	--	--	--	--
12	Stas	Solodov	548	111th (13:14)	2nd (8:52)	67th (8:35)	368th (295 lbs)	37	175 cm	85 kg	160 kg	115 kg	95 kg	200 kg	--	50	2:30	2:15	9:00	--	1:05	21:05	

some of the date missing

```

1 df['Clean&Jerk'].str.split(' ', expand=True)
2 df[['Clean&Jerk', 'kg-lb']] = df['Clean&Jerk'].str.split(' ', n=1, expand=True)
3
4 df['Clean&Jerk'] = pd.to_numeric(df['Clean&Jerk'], errors='coerce')
5 df = df.replace(np.nan, 0, regex=True)
6 df['Clean&Jerk'] = df['Clean&Jerk'].astype(int)
7
8 df.loc[df['kg-lb']=='lb', ['Clean&Jerk']] *= 0.45

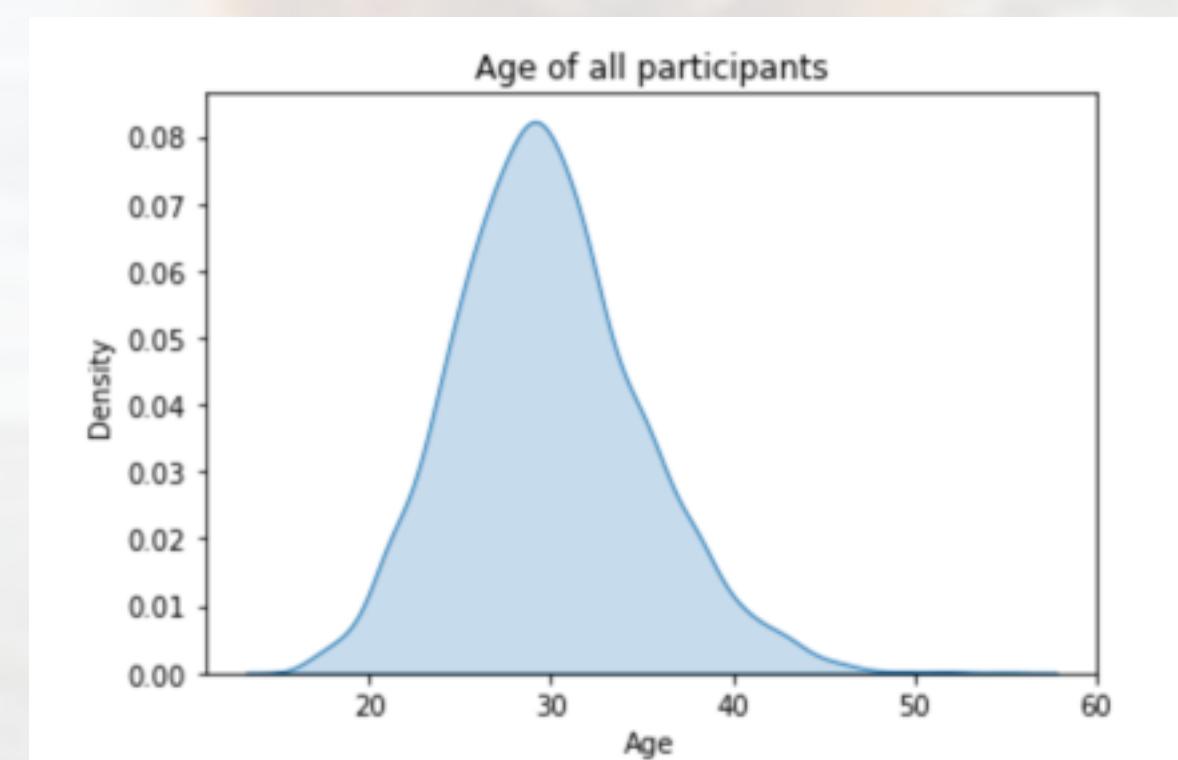
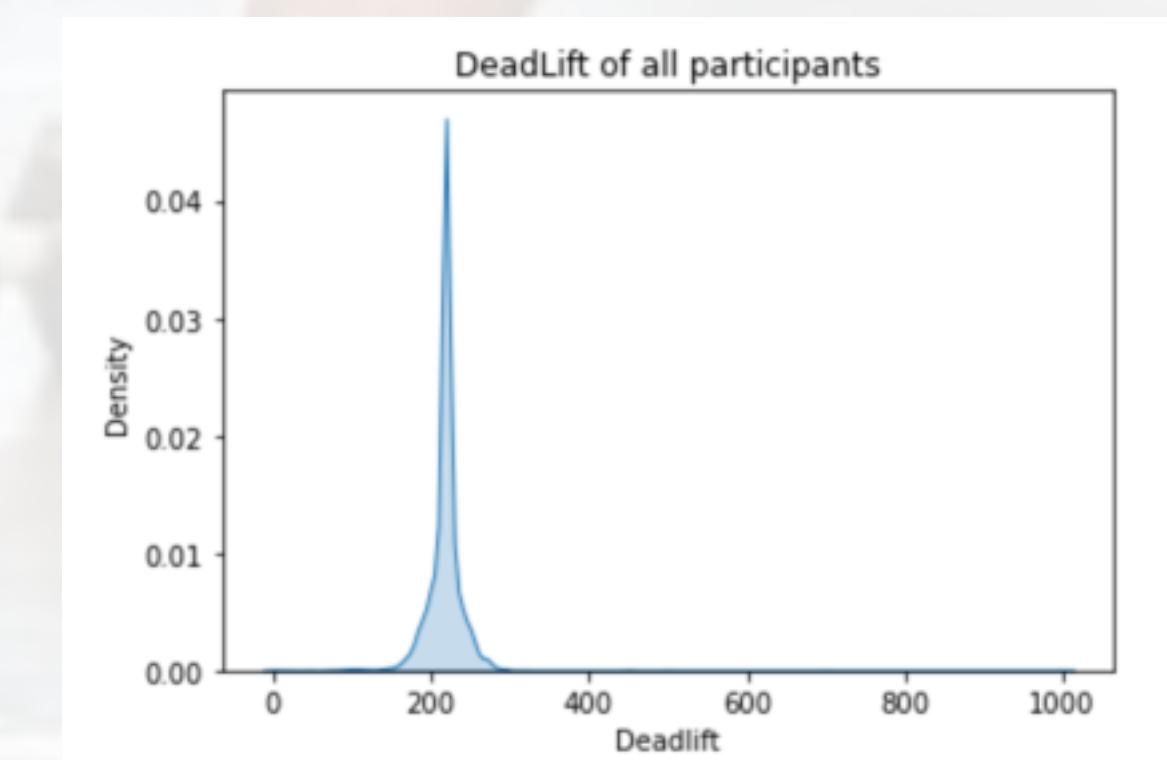
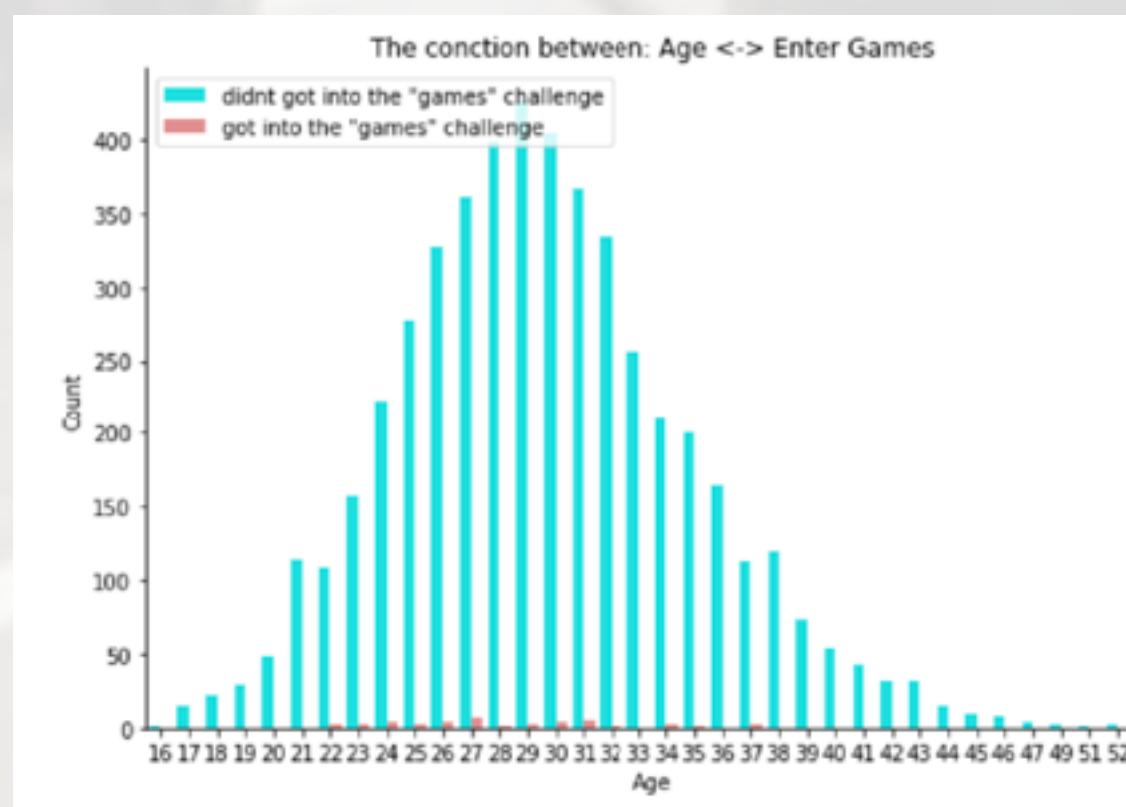
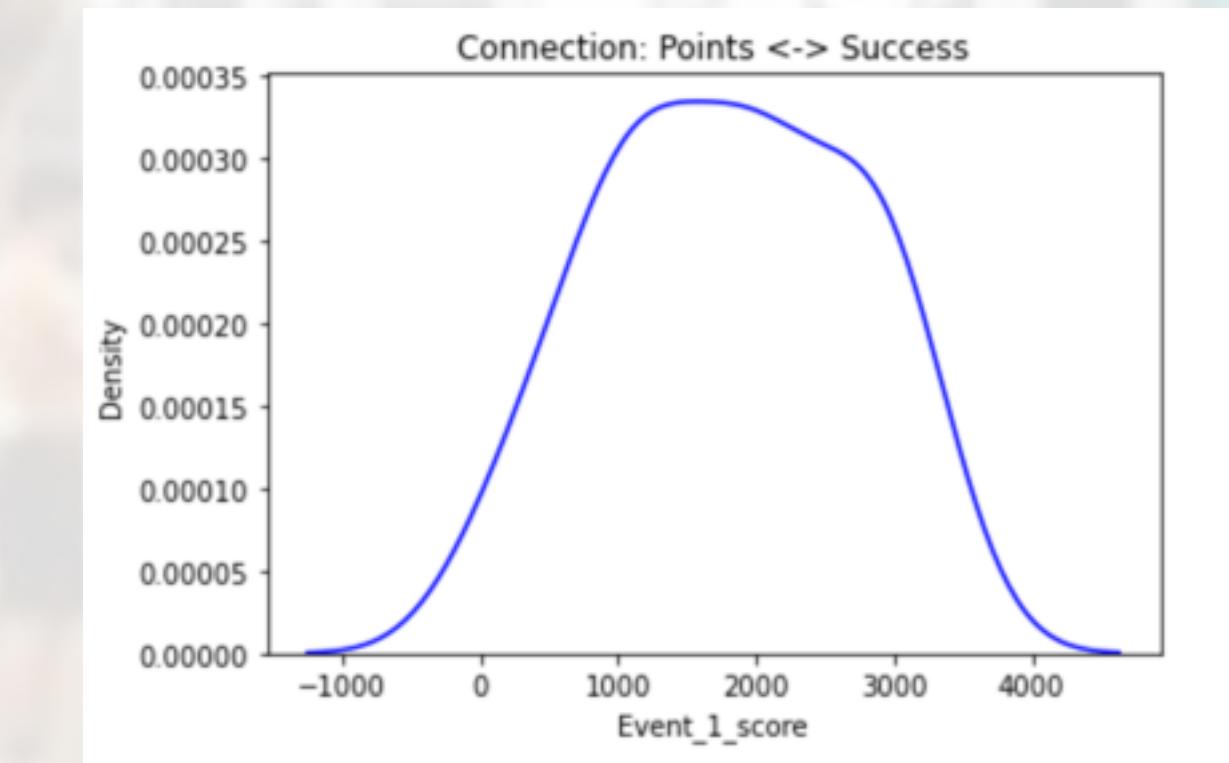
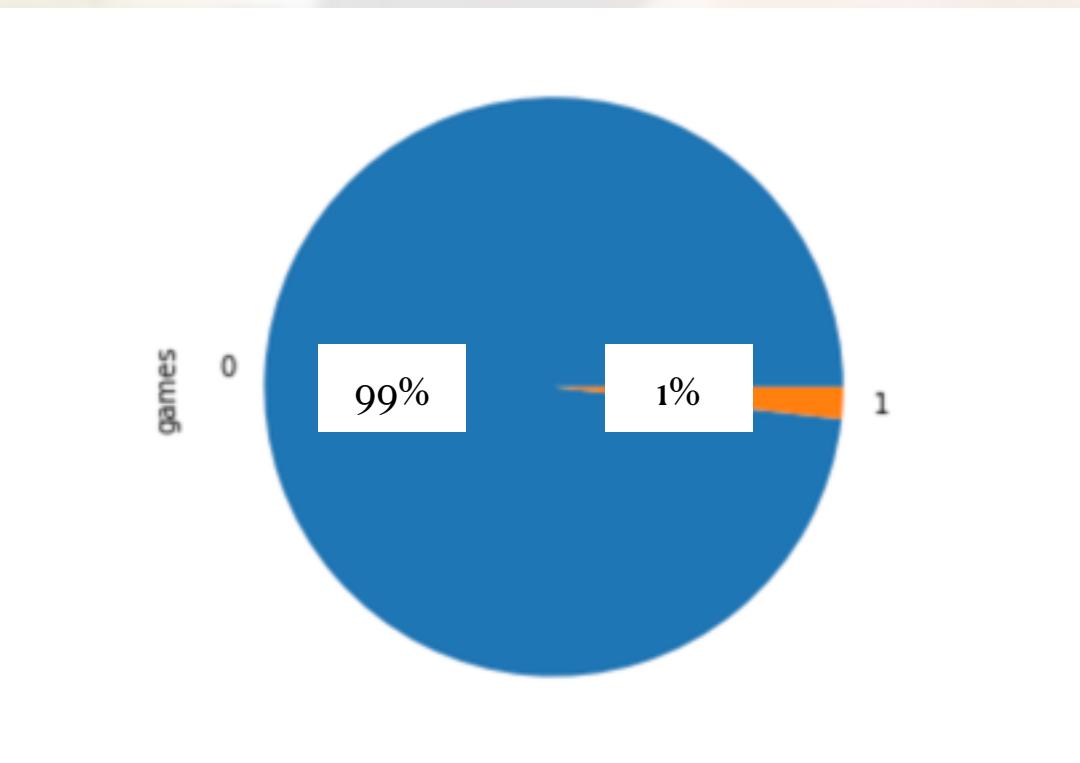
```

```

1 df = df.replace(0, np.nan, regex=True)
2 df.fillna(value=df['Clean&Jerk'].mean(), inplace =True)

```

Eda-visualization

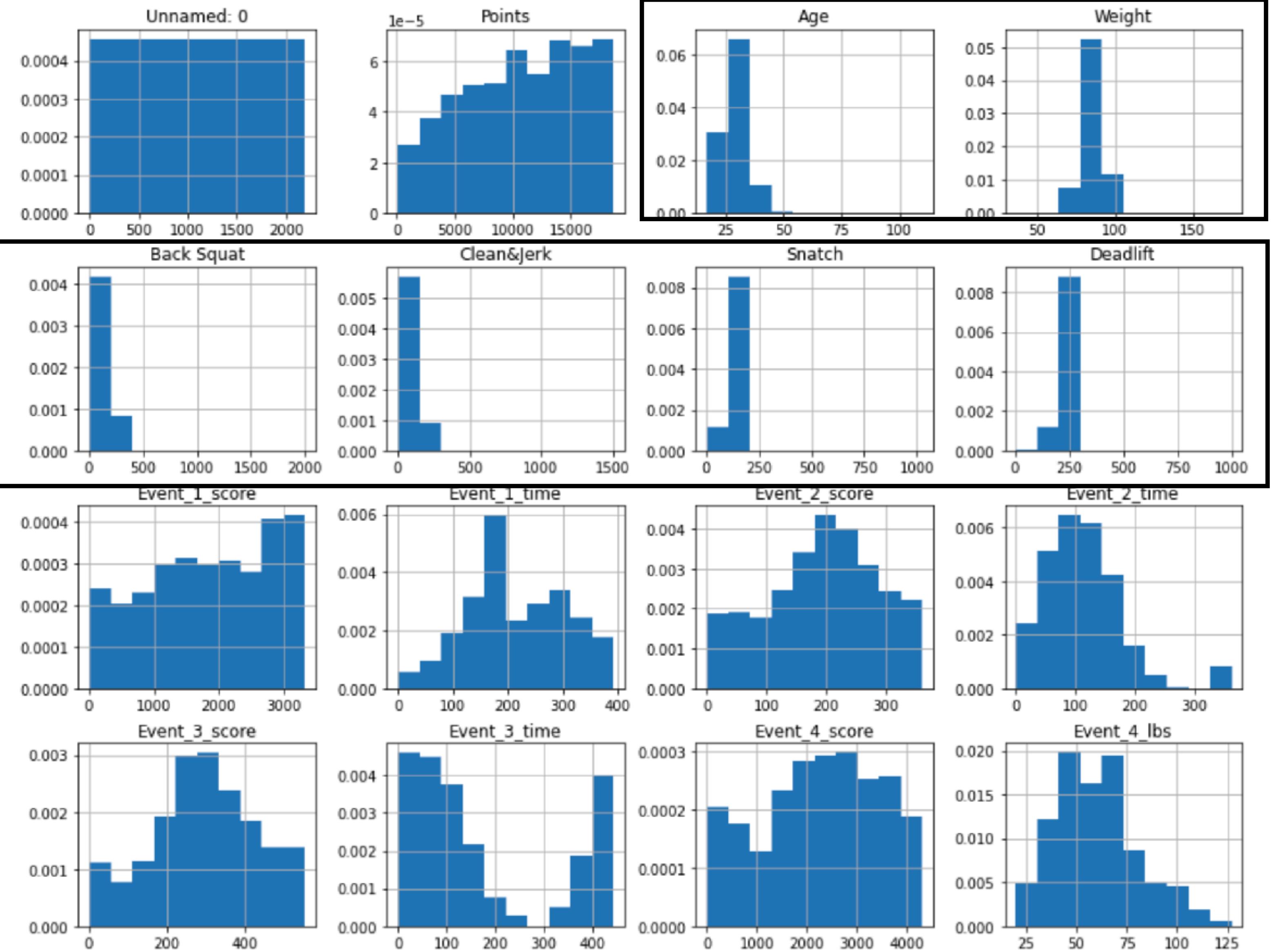


Eda-visualization

[6] :

	Points	Age	Weight	Back Squat	Clean&Jerk	Snatch	Deadlift	Event_1_score	Event_1_time	Event_2_score	Event_2_time	Event_3_score	Event_3_time	Event_4_score	Event_4_lbs	games
Points	1.000000	0.102140	-0.084363	-0.093299	-0.082375	-0.094790	-0.122106	0.088270	0.205969	0.048452	0.263857	-0.004161	-0.372301	0.119707	-0.546140	-0.244254
Age	0.102140	1.000000	0.056970	0.059870	0.023924	0.016948	0.075466	0.017151	0.050957	0.005916	0.019241	-0.013239	-0.065184	0.025709	-0.015996	-0.040593
Weight	-0.084363	0.056970	1.000000	0.221123	0.254982	0.291154	0.301315	0.067036	0.037305	0.007377	-0.042560	0.014803	-0.027887	-0.070009	0.355747	0.066235
Back Squat	-0.093299	0.059870	0.221123	1.000000	0.272786	0.316713	0.428861	0.020872	-0.018695	0.002304	-0.041676	-0.012780	0.028098	0.003766	0.144756	0.033160
Clean&Jerk	-0.082375	0.023924	0.254982	0.272786	1.000000	0.954602	0.433092	0.029545	-0.041857	0.005610	-0.010229	0.037087	0.040926	-0.043908	0.180718	0.039744
Snatch	-0.094790	0.016948	0.291154	0.316713	0.954602	1.000000	0.526408	0.028691	-0.048758	0.017847	-0.014657	0.032751	0.049691	-0.049249	0.184909	0.055060
Deadlift	-0.122106	0.075466	0.301315	0.428861	0.433092	0.526408	1.000000	0.040144	-0.051875	-0.013231	-0.050135	0.025802	0.045653	-0.054815	0.195328	0.060930
Event_1_score	0.088270	0.017151	0.067036	0.020872	0.029545	0.028691	0.040144	1.000000	-0.552315	-0.021203	0.001720	0.004702	-0.041231	-0.020795	0.085470	-0.005035
Event_1_time	0.205969	0.050957	0.037305	-0.018695	-0.041857	-0.048758	-0.051875	-0.552315	1.000000	0.029496	0.004134	-0.019309	-0.142452	0.050316	-0.108627	-0.167738
Event_2_score	0.048452	0.005916	0.007377	0.002304	0.005610	0.017847	-0.013231	-0.021203	0.029496	1.000000	0.179453	0.000075	-0.012811	-0.000465	0.054141	0.030958
Event_2_time	0.263857	0.019241	-0.042560	-0.041676	-0.010229	-0.014657	-0.050135	0.001720	0.004134	0.179453	1.000000	0.005751	-0.083994	0.038727	0.038479	0.158844
Event_3_score	-0.004161	-0.013239	0.014803	-0.012780	0.037087	0.032751	0.025802	0.004702	-0.019309	0.000075	0.005751	1.000000	0.281470	-0.008112	0.052635	0.007750
Event_3_time	-0.372301	-0.065184	-0.027887	0.028098	0.040926	0.049691	0.045653	-0.041231	-0.142452	-0.012811	-0.083994	0.281470	1.000000	-0.057244	0.152078	0.160993
Event_4_score	0.119707	0.025709	-0.070009	0.003766	-0.043908	-0.049249	-0.054815	-0.020795	0.050316	-0.000465	0.038727	-0.008112	-0.057244	1.000000	-0.178867	-0.023258
Event_4_lbs	-0.546140	-0.015996	0.355747	0.144756	0.180718	0.184909	0.195328	0.085470	-0.108627	0.054141	0.038479	0.052635	0.152078	-0.178867	1.000000	0.206522
games	-0.244254	-0.040593	0.066235	0.033160	0.039744	0.055060	0.060930	-0.005035	-0.167738	0.030958	0.158844	0.007750	0.160993	-0.023258	0.206522	1.000000

Even though that in some aspects there is correlation
In the big picture it is difficult to find something beyond doubts



It look like in the athlete's parameters the changes are minority

Machine Learning

```
In [32]: 1 y = df['games']
2 X = df.drop('games', 1)

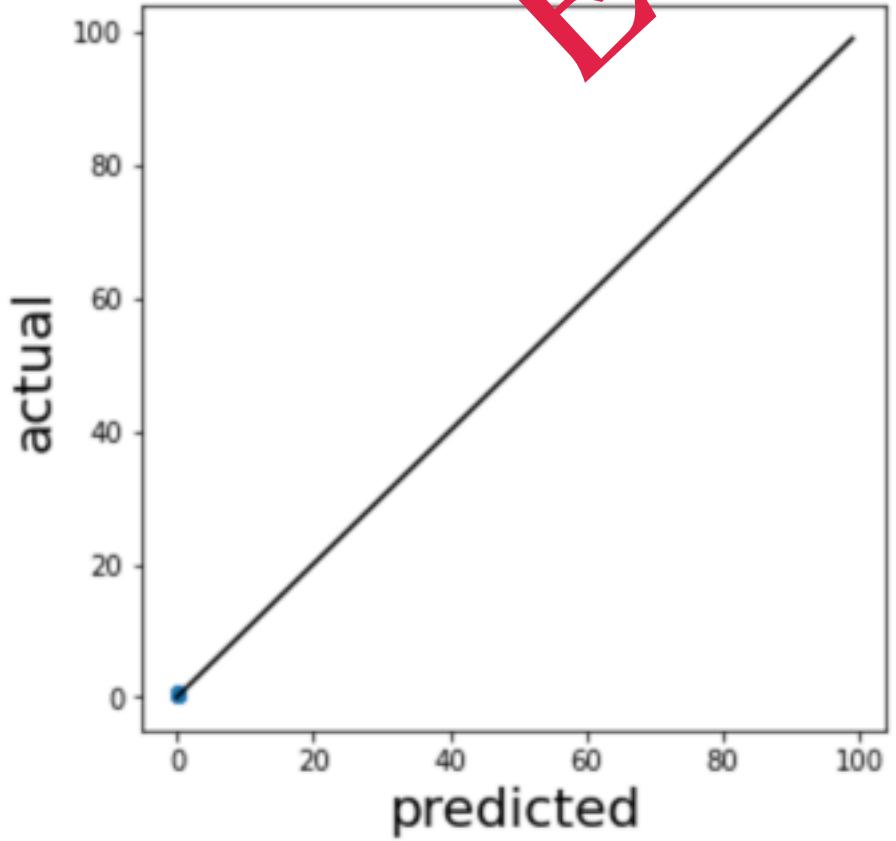
In [33]: 1 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=None, random_state=42)

In [34]: 1 model = ElasticNet().fit(X_train, y_train)

In [35]: 1 y_pred = model.predict(X_test)

In [36]: 1 plt.figure(figsize=(5,5))
2 plt.plot(y_pred, y_test, '.')
3 plt.xlabel("predicted", fontsize=20)
4 plt.ylabel("actual", fontsize=20)
5 plt.plot(range(100), range(100), "-k")

Out[36]: <Figure size 360x360 with 0 Axes>
Out[36]: [<matplotlib.lines.Line2D at 0x7fcec05cb120>]
Out[36]: Text(0.5, 0, 'predicted')
Out[36]: Text(0, 0.5, 'actual')
Out[36]: [<matplotlib.lines.Line2D at 0x7fcec05cb250>]
```



```
In [37]: 1 r2_score(y_test, y_pred)

Out[37]: 0.09270041764605785
```

```
In [26]: 1 y = df['games']
2 X = df.drop('games', 1)

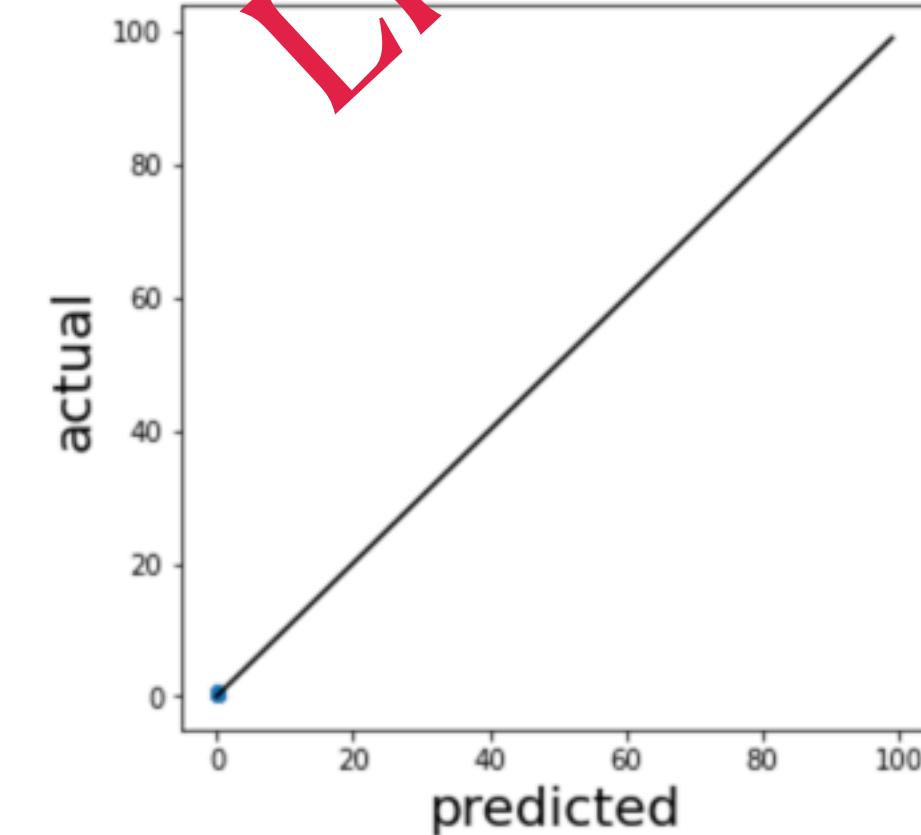
In [27]: 1 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=None, random_state=42)

In [28]: 1 model = LinearRegression().fit(X_train, y_train)

In [29]: 1 y_pred = model.predict(X_test)

In [30]: 1 plt.figure(figsize=(5,5))
2 plt.plot(y_pred, y_test, '.')
3 plt.xlabel("predicted", fontsize=20)
4 plt.ylabel("actual", fontsize=20)
5 plt.plot(range(100), range(100), "-k")

Out[30]: <Figure size 360x360 with 0 Axes>
Out[30]: [<matplotlib.lines.Line2D at 0x7fcec1d531f0>]
Out[30]: Text(0.5, 0, 'predicted')
Out[30]: Text(0, 0.5, 'actual')
Out[30]: [<matplotlib.lines.Line2D at 0x7fcec1d53d00>]
```



```
In [31]: 1 r2_score(y_test, y_pred)

Out[31]: 0.07492093894519458
```

LogisticRegression

```
In [219]: 1 X_train, X_test, y_train, y_test = train_test_split(df_2.drop('games',axis=1), df_2['games'], random_state =0)

In [222]: 1 clf=LogisticRegression()
2 clf.fit(X_train,y_train)
3 acc=clf.score( X_test,y_test)

/Applications/anaconda3/lib/python3.8/site-packages/sklearn/linear_model/_logistic.py:763: ConvergenceWarning: lbfgs
failed to converge (status=1):
STOP: TOTAL NO. OF ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
n_iter_i = _check_optimize_result()

Out[222]: LogisticRegression()
```

```
In [223]: 1 print (acc)

0.9708029197080292
```

```
In [68]: 1 X_train, X_test, y_train, y_test = train_test_split(df_3.drop('games',axis=1), df_3['games'], random_state =0)
```

```
In [69]: 1 clf=LogisticRegression()
2 clf.fit(X_train,y_train)
3 acc=clf.score( X_test,y_test)
```

```
/Applications/anaconda3/lib/python3.8/site-packages/sklearn/linear_model/_logistic.py:763: ConvergenceWarning: lbfgs
failed to converge (status=1):
STOP: TOTAL NO. OF ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression
n_iter_i = _check_optimize_result()
```

```
Out[69]: LogisticRegression()
```

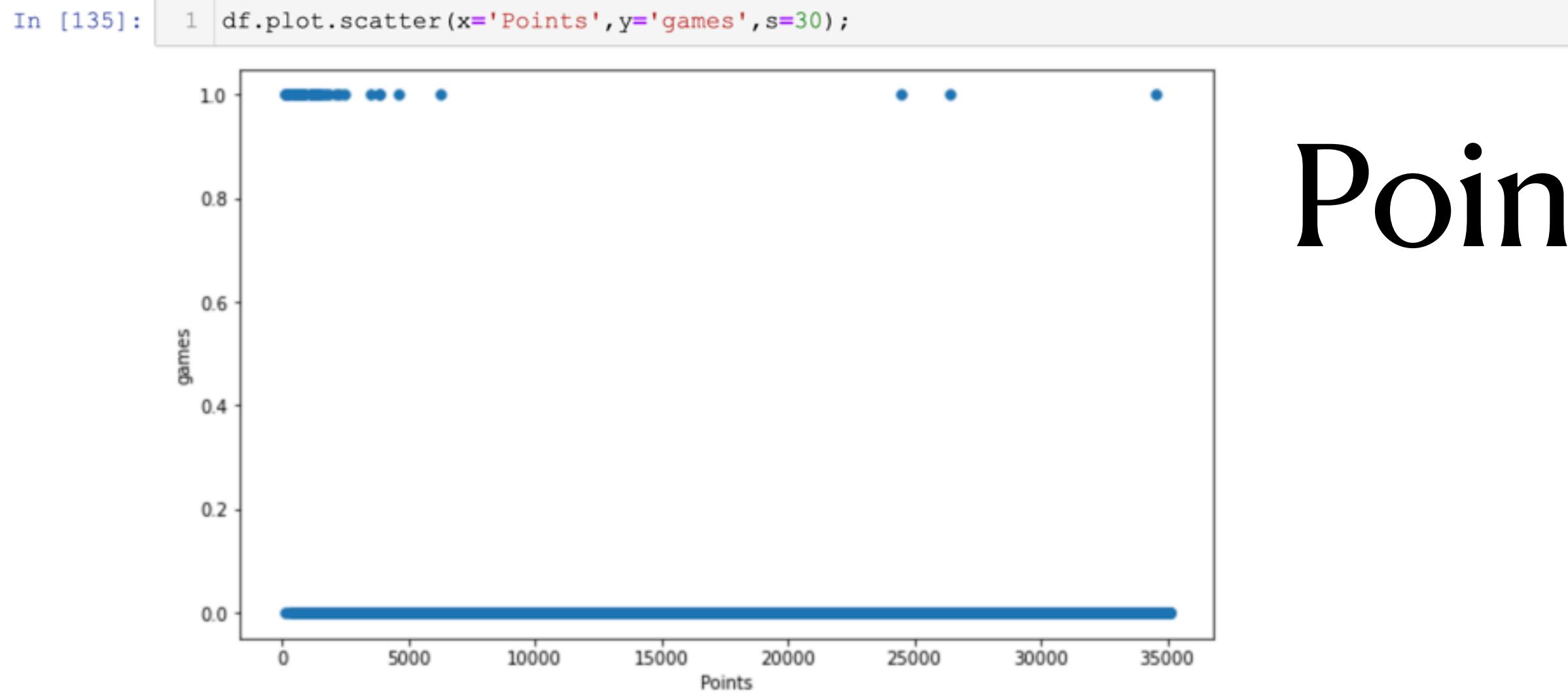
```
In [70]: 1 print (acc)

0.9896
```

A photograph of three male CrossFit athletes standing on a podium. The athlete on the left is wearing a black t-shirt with 'FIT ON' and '2' on it, and is holding a Canadian flag. The central athlete is wearing a white t-shirt with 'FIT ON' and '1' on it, and is holding an American flag. The athlete on the right is wearing a grey t-shirt with 'FIT ON' and '3' on it, and is holding a Canadian flag. They are all smiling and have medals around their necks. In the background, there is a large banner with the words 'CROSSFIT NO BULL'.

Even though there is thousands of competitors
It is look like the differences that we can see are insignificant thus
unfortunately we were unable to find good correlation and can't predict

Personal record



Points