

P1 POWERLIFTING

Stefan Leif Brzost-Andersen

22/12/2021



AALBORG UNIVERSITY
STUDENT REPORT

Title:

OpenPowerlifting Datasæt

Project:

P1

Project Period:

08/12/2021 - 22/12/2021

Project Group:

B264a (SOLO)

Group Members:

Stefan Leif Brzost-Andersen

Supervisor:

Christophe Biscio

Number of Pages:

18

The Department of Computer Science

Data Science

Selma Lagerløfs Vej 300

9220 Aalborg Øst

<https://www.cs.aau.dk/>

Abstract:

Our goal is to find interesting correlations and appropriate data regarding the growing sport of modern Powerlifting. The data is collected through OpenPowerlifting, which are contributed to the Public Domain. We've chosen Linear Regression as the method to analyse the athletes and their traits to finalise our model for prediction.

The content of this report is freely available, but publication (with reference) may only be pursued due to agreement with the authors.

Indholdsfortegnelse

1	Introduktion	1
2	Exploratory Data Analysis	2
2.1	Data Typer	2
2.2	Data Cleaning	3
2.3	Numeriske Data	4
2.4	Kategorisk Data	9
2.5	Afrundning af databeskrivelse	11
3	Problemanalyse	12
3.1	Metodologi	12
3.2	Absolut Styrke og WILKS	14
3.3	Classic versus Equipment	14
3.4	Alders Effekt på Styrke	16
4	Problemformulering	17
5	Konklusion	17
6	Diskussion	17

Underskrift

Stefan Leif Brzost-Andersen

Forord

Projektet er designet for 1. Semester på Aalborg Universitet fra perioden d. 08-12-2021 til 21-12-2021. Der arbejdes med OpenPowerlifting Datasæt fra <https://www.openpowerlifting.org/>. Denne enkeltmands-gruppe siger også tak til Christophe Biscio for stress aflastning, forklaringer og hans tålmodighed. Projektet havde af psykiske årsager ikke været muligt at gennemføre og er derfor taknemmelig for hans hurtige hjælp.

1 Introduktion

Powerlifting er en styrkesport, hvor atleter kan stille op uden udstyr (Typisk refereret som 'RAW / Classic' dvs. uden udstyr) eller med udstyr. Udstyr defineres typisk i form af en mekanisk-støttende vest (Bench, deadlift, squat- suit) og 'wraps', hvis formål er at støtte ens banekurve igennem løftet og sørger for, at atleten kan klare tungere vægte, som i nogle tilfælde er over 100kg end hvad de ellers ville kunne have løftet i RAW. Alt dette er ulovligt i RAW / Classic, imens 'straps', 'bælte', 'squat shoes' og andet udstyr, hvis formål er til at minimere skaderisiko, er lovligt.

Atleten kan vælge at stille op i både squat, bænkpres og dødløft, men kan også stille op til solo-konkurrencer, hvor der kun udføres en enkelt disciplin bl.a. DM i Bænkpres. Ligesom i Olympisk Vægtløftning vil atleten modtage tre løft - uanset disciplin, hvor kun det stærkeste af teknisk-godkendte løft bliver registreret.

Konkurrencerne findes i de fleste steder i verden og er også en af de store paralympiske sportsgrene (bænkpres ONLY). Der findes internationale, interkontinental, nationale, lokale versioner, hvor præmien typisk er æren, fremfor en større økonomisk gevinst, som der ses i andre sportsgrene. Herefter følger der forskellige foreninger med til konkurrencerne, hvor nogle har bandlyst andre, da reglerne er forskellige fra forening til forening - både ift. doping-politik, men også teknik, krav for udstyr, valg af strømper, sko, trøje, etc.

Powerlifting har fået en del popularitet de seneste år og vi vil derfor kigge på, hvor hurtigt PL har vokset, hvorfor PL har vokset og om der er noget betydeligt grundlag for at større foreninger sanktionere mindre foreninger, såfremt deres reglement ikke støtter deres eget. En af de større problematikker vi vil fokusere på er om der ses en betydelig forskel mellem non-doping foreninger og doping foreninger. Til sidst vil vi se på nogle af de største faktorer til powerlifting, heraf WILKS og om en ny, mere retfærdig og non-bias løsning kan findes til bestemmelse af atleternes styrke-index.

2 Exploratory Data Analysis

OpenPowerlifting Datasæt indholder data fra 1962 til 2021 og samtlige konkurrencer, events og meets er registreret. Der er 41 variabler og ca. 2.5 millioner samples i datasættet. Udover dette vil en del variabler blive ekskluderet på grund af mangel på tid - dermed er der blevet valgt 16 ud af de 41 variabler. Udvalgte variabler vil blive analyseret og beskrevet i de kommende afsnit.

Ekskluderet			
Name	BirthYearClass	Division	Place
Dots	Glossbrenner	Goodlift	State
MeetState	MeetTown	Meetname	Event
Squat1Kg	Squat2Kg	Squat3Kg	Squat4Kg
Bench1Kg	Bench2Kg	Bench3Kg	bench4Kg
Deadlift1Kg	Deadlift2Kg	Deadlift3Kg	Deadlift4Kg
Dato	MeetCountry	Udstyr	Deadlift4Kg

2.1 Data Typer

Datasættet indholder forskellige datatyper og de vil derfor opdeles til to kategorier: *Numeriske data* og *Kategoriske data*.

Inkluderet	
<i>Numerisk Data</i>	<i>Kategorisk Data</i>
<i>Dtype = Float64</i>	<i>Dtype = Object</i>
Alder	Køn
Kropsvægt	Land
Maksimal Squat	Kategori
Maksimal Bænkpres	Vægtklasse
Maksimal Dødløft	Doping
Total Load	Parent Federation
Wilks	Federation

2.2 Data Cleaning

OPL-Datasæt er betydeligt større end det almindelige datasæt med sine 2.5 millioner samples. Der er blevet kortlagt en række betingelser til datacleaning, som minimere fejlagtig data og sørger for det er gennemskueligt til analyse. Da udgangspunktet i analysen er de inkluderede variabler vil de resterende variabler ikke renses og der opdeles to separate datasæt. Det rensede datasæt kører igennem fejlsøgning i form af loops og derefter er det klar til videre analyse.



Der blev fundet en betydelig mængde på mangel af værdier i Datasættet. Der blev også fundet en del N/A-værdier og besynderlige værdier, såsom kropsvægt på 0.5kg og en alder på 105 år. Dette bringer vores totalmængde ned på 462.965 samples og 16 features.

Fordelingen er dog ikke uniform blandt alle features og af denne grund vil der blive tilfældigt trukket 100.000(n) tilfældige samples. Da der kan være betydelig difference mellem data fra non-doping og doping federationer vil grafer bruge data fra federationer med godkendt doping-politik - uden andet er beskrevet.

Manglende Værdier (Procent)		
Data	Før	Efter
Sex	0.00	0.00
Event	0.00	0.00
Equipment	0.00	0.00
Age	37.91	0.00
BodyweightKg	1.30	0.00
WeightClassKg	1.15	0.20
Best3SquatKg	34.15	0.00
Best3BenchKg	12.09	0.00
Best3DeadliftKg	28.74	0.00
TotalKg	7.01	0.00
Wilks	7.88	0.00
Tested	26.22	0.00
Country	44.95	0.00
Federation	0.00	0.00
ParentFederation	35.47	11.11

2.3 Numeriske Data

2.3.1 Alder

Alder angiver hvor gammel den given atlet er på konkurrences tidspunkt. Denne variable bruges til at indstille dem til deres pågældende kategori (beskrives i 2.3.3 Kategori).

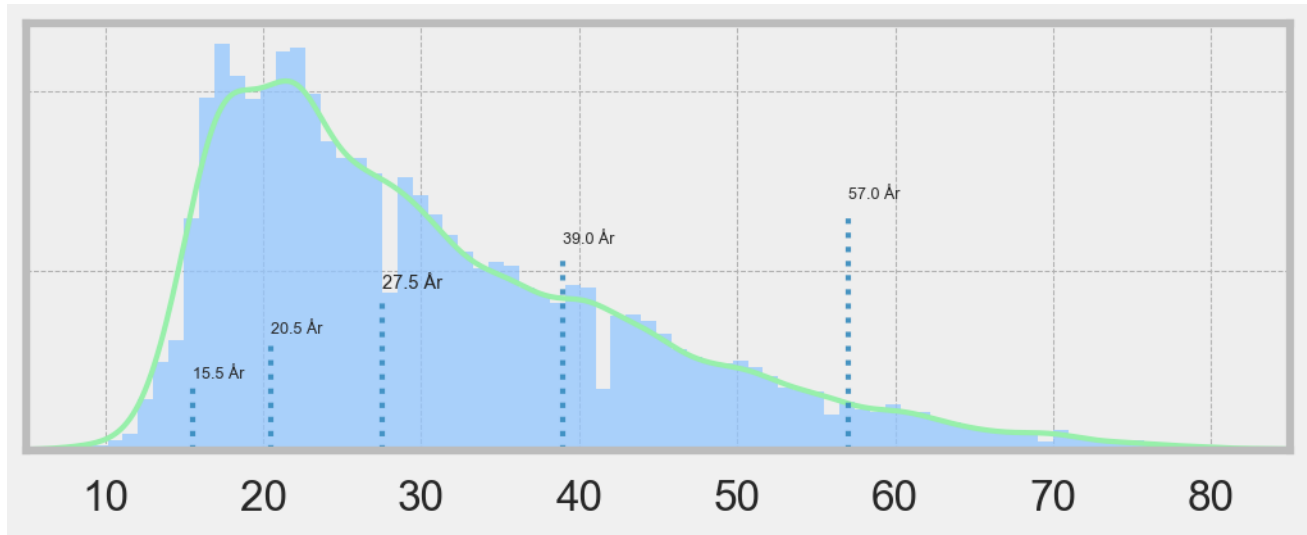


Figure 2.3.1: Graf viser fordeling pr. Alder og hhv. 5, 25, 50, 75, 95 kvantil.

Ved nærmere overview af OPL-Datasæt ser vi, at den ældste atlet er 95 år gammel, imens den yngste atlet er 7 år gammel. Fordelingen af alder hos atleterne kan ses i overstående histogram m. Kvartilsæt og KDE. Den gennemsnitlige alder ligger på 27.5 år og i de øvre kvartiller ses der hhv. 39 og 57 år. Sporten er derfor mest populær hos folk på 28 år og under, imens 57-årige og over udgør 5 procent. Dette tænkes som en tendens med PLs stigende popularitet i verden.

2.3.2 Kropsvægt

Atletens kropsvægt er en af de afgørende værdier indenfor PL, da ens absolut styrke stiger med ens total masse. Vi vil senere kigge på sammenhængen mellem kropsvægt og køn i kombination med absolut styrke.

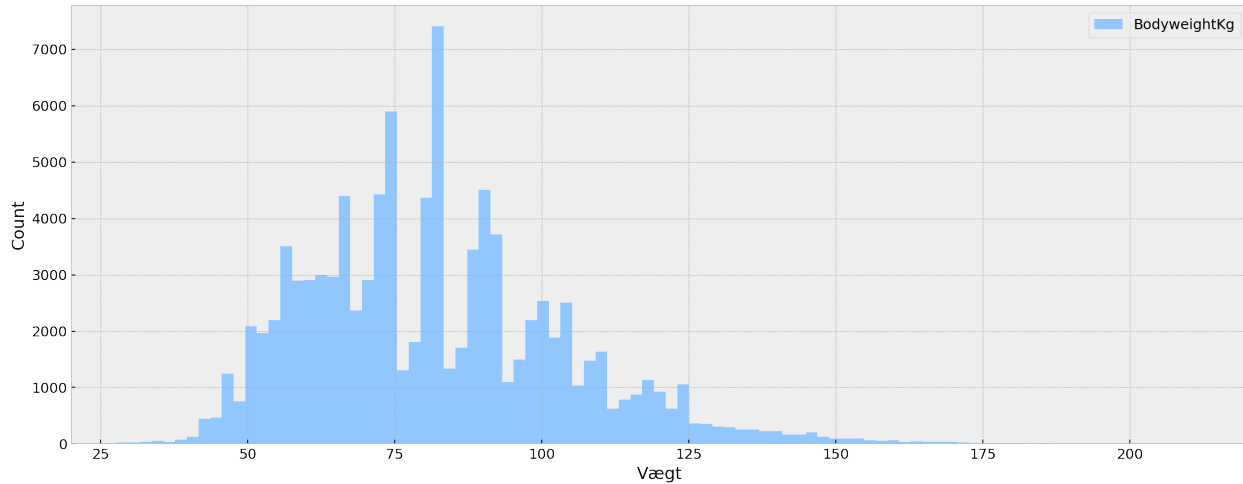


Figure 2.3.2: Graf viser fordeling i atleternes kropsvægt.

	Count	Mean	Std	Min	25%	50%	75%	Max
BodyweightKg	n = 100.000	81.87	22.17	21.7	65.5	80.5	94.4	220.3

Den gennemsnitlige atlet vejer 81.87 kilogram og afvigelsen fra mean er på 22.17 kilogram. Størstedelen af atleterne vejer mellem ca. 60kg til 104kg og det må derfor være de mest populære vægtklasser. Den mindste atlet vejer 21.7kg og den yngste atlet er 7 år gammel, dette hænger sammen med den forventede vægt for et barn på 7 år.

1. kvartil viser at 25 procent og under vejer 65.5kg, hvilket muligvis kan passe med kønsfordelingen mellem mænd og kvinder. Ved 2. Kvartil ser vi 80.5kg og ved 3. Kvartil ser vi 94.4 kg - det observeres derfor, at størstedelen af atleterne deltager i vægtklassen under 103kg. Den tungeste registreret atlet lå på 220.3 kilogram og betegnes som en outlier, som observeres på grafen.

2.3.3 Maksimal Squat, Dødløft, Bænkpress

Best3Squat, Best3Bench, Best3Deadlift er det stærkeste, teknisk-godkendt løft ud af alle atletens forsøg til en konkurrence. Hveri TotalKg er atletens samlede total løft, hvor man beregner summen af alle tre løft. Den stærkeste atlet er derfor den med det højeste total løft i alle discipliner.

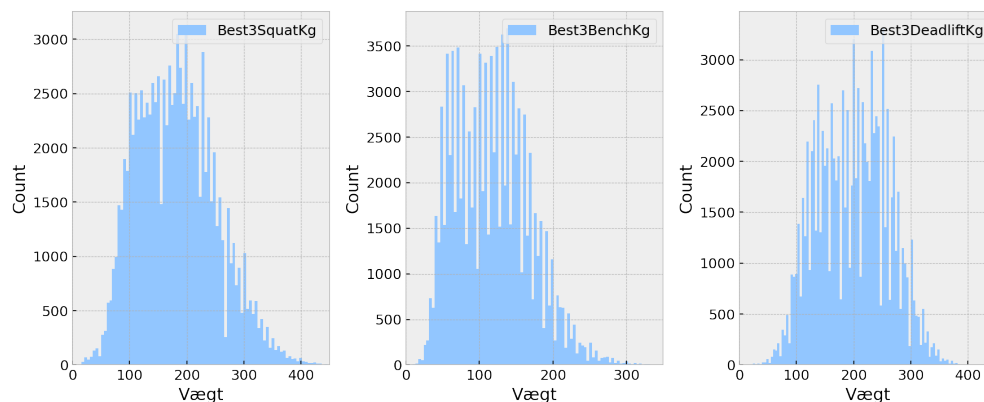


Figure 2.3.3: Graf viser fordeling i hhv. Squat og Bænkpress.

	Best3SquatKg	Best3BenchKg	Best3DeadliftKg
Count	n = 100.000	n = 100.000	n = 100.000
Mean	182.1	118.1	198.23
Std	68.82	50.5	61.5
Min	1	9	15
25%	128	75	147.5
50%	180	115	200
75%	230	150	245
Max	490	385	405

Vi kan observere ud fra deres mean, at den gennemsnitlige atlet generelt er stærkere i dødløft end i squat og bænkpres. De er generelt svageste i bænkpres, imens deres squat hænger tæt op ad dødløft. Derimod er der mindst afvigelse fra mean ved bænkpres, men højest i squat. Der ses også en mindste-værdi på 1 i squat, hvor det var besværligt at finde fejl i dette, da alle 0.5 værdier er fjernet efter datacleaning og 1kg ikke er et uværdigt løft for en 7-årig til en konkurrence. Derfor vil et teoretisk løft for en 7-årig med 1kg squat, 9kg bænkpres og 15kg dødløft ikke være usædvanligt. Som forventet er dødløft også den stærkeste løft hos de svageste atleter.

I hhv. 1. 2. og 3. kvartil ser vi, at 25 procent af atleterne har et maksimal løft på 128kg i squat, 75kg i bænkpres og 147.5kg i dødløft. Derefter er næste niveau på 180kg i squat, 115kg i bænkpres og 200kg i dødløft. Sidste niveau er på 230kg i squat, 150kg i bænkpres og 245kg i dødløft. Det skal dog noteres at der ikke er taget forhold for absolut styrke i disse værdier og de reelle tal er markant anderledes, når vægtklasser tages med i formlen.

2.3.4 Total Load

Total Load er summen af alle tre discipliner. Eksempelvis, hvis en atlet har et SBD-løft på hhv. 200kg, 150kg og 200kg vil det give et total load på 550kg. Total loads mean er på 498.55 kg med en afvigelse på 174.88kg. De fleste registrerede atleter vil derfor ligge mellem 323.67kg til 673.43kg i total løft. Dette forhold kan sammenlignes med atleternes kropsvægt, da det højeste total load er på 1205 kilogram og der skal være en tilsvarende, proportional kropsvægt for at opnå den total load.

TotalKg	Count	Mean	Std	Min	25%	50%	75%	Max
Total Load	n = 100.000	498.55	174.88	38.6	355	498.95	622.5	1205

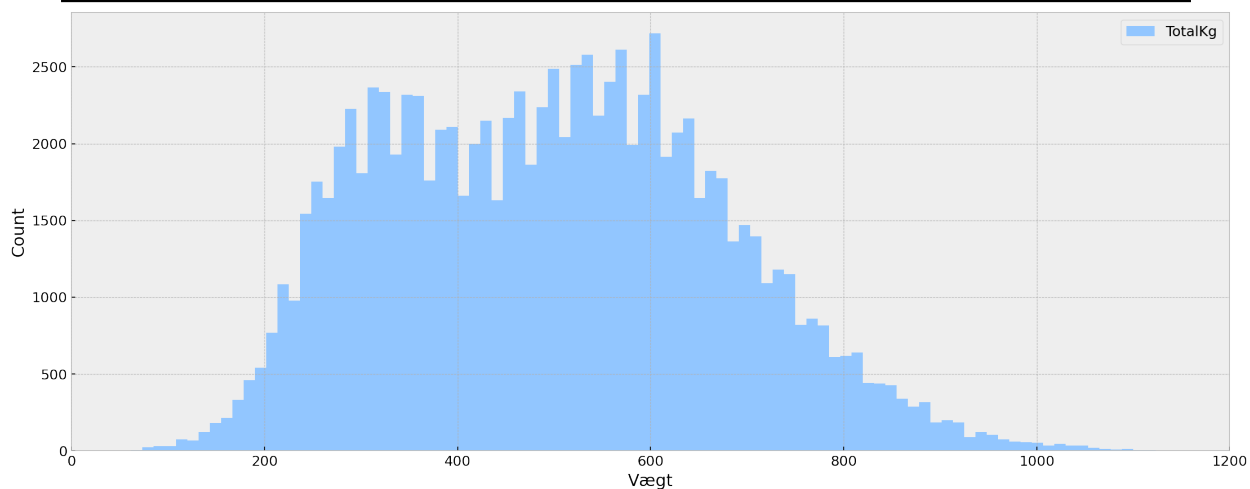
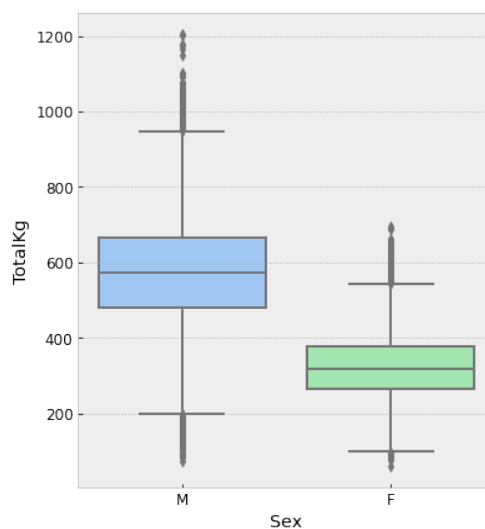


Figure 2.3.4: Graf viser fordeling i Total Load (KG)

I 2.3.5 har vi 'M' for mand og 'F' for kvinde. Grafen giver en nogenlunde visuel præsentation af styrkeforskellen mellem mænd og kvinder, hvor der ses højest afvigelse i de øvre outliers, der kan betegnes som de absolut stærkeste kvinder og mænd. Derimod er de svageste mænd og kvinder på nogenlunde samme niveau, dog er der talrige fejl i denne udtalelse og den er kun observeret fra denne graf.

Figure 2.3.5: x = Sex, y = TotalKg



2.3.5 Wilks

'*Mass Moves Mass*'-citateret kan bruges til forklaring af PLs vægtklasse system. Der har været gentagne matematiske modeller, som er designet til et enkelt formål; Målingen af relativ styrke mellem de forskellige vægtklasser - en model der skal neutralisere det naturlige hierarki, der ses ved måling af mekanisk styrke, hvor absolut styrke placeres i den absolutte top.

En af disse matematiske modeller er Wilks og er på nuværende tidspunkt en af de mest brugte metoder til at måle relativ styrke hos styrkeatleter. På denne måde vil man se hvem der er stærkest, når man fjerner '*Mass Moves Mass*' fra den naturlige formel. Problemet i modellen er, at den bestemt ikke er perfekt og der findes optimeret formler, som ville passe bedre til formålet. Wilks har i tidligere modeller haft en betydelig køns-bias og det ses typisk at mænd sættes højere op end kvinder. Imens der også ses en alders-bias i en enkelt vægtklasse (Open Class).

Der observeres en markant udligning af køns-bias mellem tidligere modeller og dette forhold kan ikke isoleres til reelt køns-bias ved kort omfang. Det kan diskuteres om en højere kønsdeltagelse ville neutralisere afvigelsen mellem mænd og kvinder ved sammenligning af WILKS.

	Mænd	Kvinder
a	47.46178854	-125.4255398
b	8.472061379	13.71219419
c	0.07369410346	-0.03307250631
d	-0.001395833811	-0.001050400051
e	$7.07665973070743 \times 10^6$	$9.38773881462799 \times 10^6$
f	$-1.20804336482315 \times 10^8$	$-2.3334613884954 \times 10^8$

Table 1: Oversigt over tal.

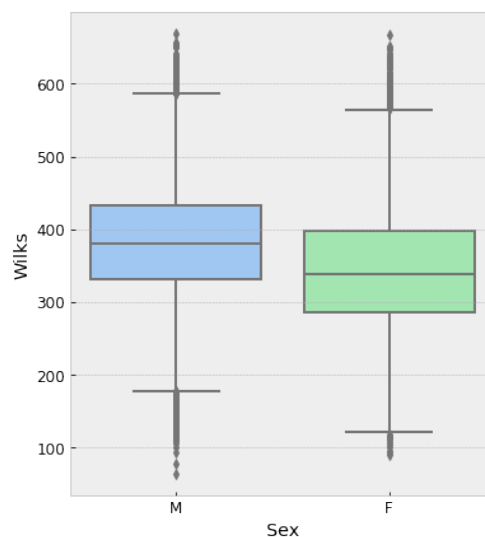


Table 2: Outlierboxplot

$$Coeff = \frac{500}{a + bx + cx^2 + dx^3 + ex^4 + fx^5}$$

Table 3: 2020-Version: Wilks Formula

2.4 Kategorisk Data

2.4.1 Land

Da Powerlifting er en international sport vil atleter fra forskellige lande tit kæmpe mod hinanden.

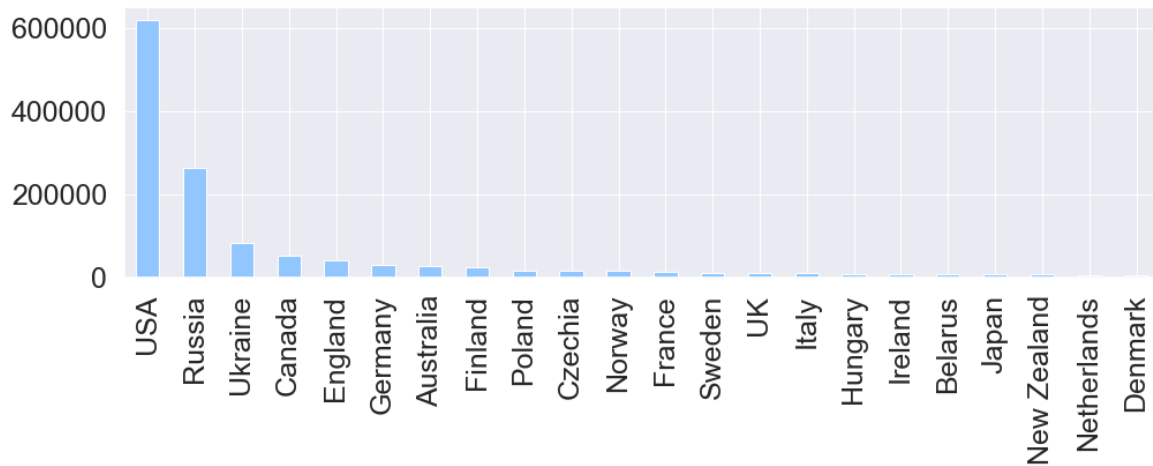


Figure 2.4.1: Deltagende Lande

Det ses at USA er det mest deltagende land, imens Rusland er lige efter. Ukraine er på tredje pladsen og deres version af sporten, samt træningsmetoder blev tit brugt af berømte trænere rundt i verden. Danmark ligger på en 23. plads og indirekte sidstepladsen, da der ikke er brugbar data for de resterende lande.

2.4.2 Tested (Non-Doping / Doping)

Variablen hentyder til om federationen håndterer kontrol af deres atleter. Typisk er det op til organisationen, dog er det en relativ stor økonomisk belastning, at kræve ens atleter bliver testet. I Danmark sørger staten for det igennem Anti-Doping Danmark (ADD). Sammen med Norge og Sverige konkurrerer disse tre lande betydeligt mod hinanden ift. andre lande, da kontrollen i andre lande ikke kan sammenlignes med den nordiske model. Denne Nordiske Alliance holder derfor flest konkurrencer mellem hinanden, da de ser det som mest fair for deres atleter.

2.4.3 Federation

Her ses de mest populære federationer i OPL-Datasæt. THSPA, USAPL og FPR tager hhv. 1. 2. og 3. pladsen. Dette ses som forventet, hvis man tager total mængde atleter pr. land i sammenligning. De mere strikse federationer, såsom IPF har i de seneste år indhentet en del, da atleter foretrækker klare regler og minimering af snyd.

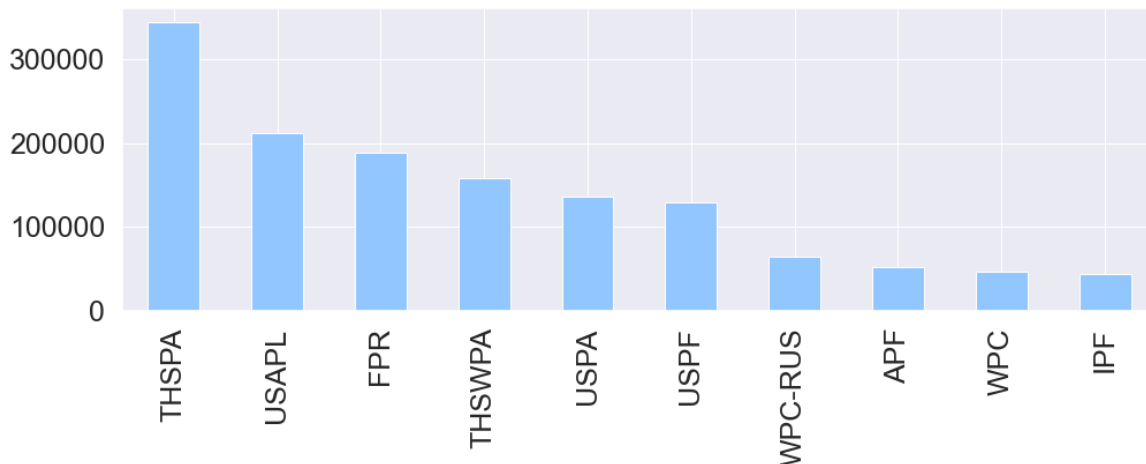


Figure 2.4.2: Mest populære federationer blandt OPL

2.4.4 Køn

'Sex' fastsætter hvilken kategori atleten skal deltage i. Der er tre kategorier i OPL-Datasæt: M, F og Mx.

'Mx' vil ikke bruges til videre analyse, da den primære interesse er for mand eller kvinde. En tidligere mandlig atlet, som der er skiftet til intetkøn vil have betydelig fordel ift. en kvinde, der skifter til intetkøn. Dette vil gøre en sammenligning besværlig, da man sammenligner miksede resultater.

Fordeling	
Data	Køn
M	69.27
F	30.72
Mx	0.000011

2.4.5 Vægtklasse

'WeightClassKg' er vægtklassen atleten har stillet op til. Den kan enten specificeres som et maksimum eller minimum - eksempelvis kan der stå '90', hvilket betyder atletens maksimale vægt er til og med 90 kilogram. Hvis der stod '90+' ville atleten stille op med en minimum kropsvægt på 91 kilogram, der skal betegnes som: 91...N Kilogram.

2.4.6 Dato

Dato henviser til en bestemt konkurrences tidspunkt. Da Powerlifting er en voksende sportsgren kan nedstående popularitetskurve give et overblik af dens vækst. Det ses at det først var ved 2008 til 2009 at sporten fik den eksplosive vækst vi ser i dag. Hvor at der var et stærkt fald i slutningen af 2019 og hen af starten af 2020, som der muligvis kan relatere til Corona og nedlukninger rundt i verden.

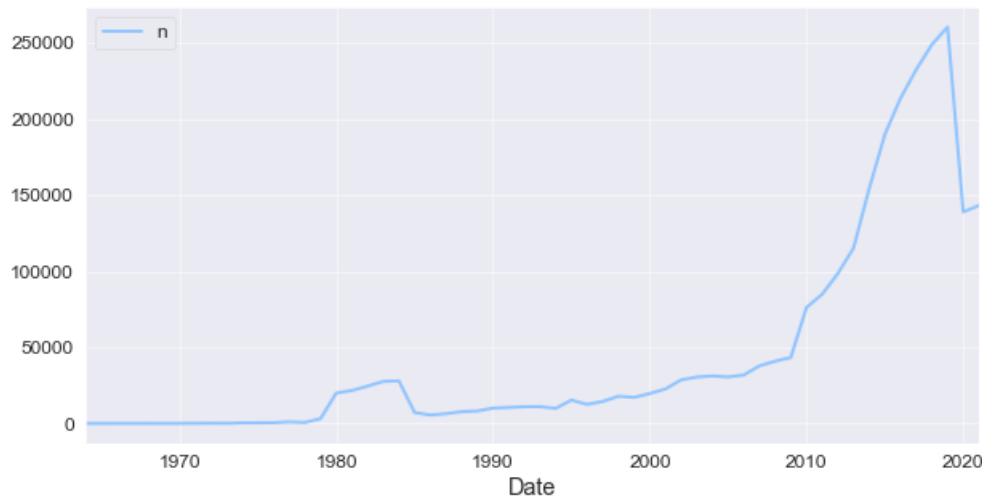


Figure 2.4.3: Popularitetskurve

2.5 Afrundning af databeskrivelse

I databeskrivelsen blev datasættet drejet rundt og der blev fundet relevant tendens til videre analyse. Det originale aspekt som opgaven skulle handle om, er dog ikke muligt indenfor den begrænsede tid, og en forsnævring af opgaven er uundgåeligt. Dette førte videre til to primære spørgsmål, som der er interessant at arbejde videre med og passer til solo-omfanget i opgaven.

- Kan man forudse atletens vægt ved brug af deres maksimale løft i hhv. squat, bænkpres og dødløft?
- Hvordan ser atleters progression ud over tid, og hvornår er de på deres stærkest / svageste?
- Hvilke variabler er vigtigst ift. absolut styrke?

Det vil derfor også være interessant med et korrelations heatmap for tjek af variablerne, samt en videre datanalyse af mænd og kvinders fysiske træk i Powerlifting.

3 Problemanalyse

Scikit-Learn er indbygget med en funktion kaldet LinearRegression, der vil blive brugt i det videre arbejde. Dens funktion omhandler minimeringen af afvigelse (RSS, Residual Sum of Squares) igennem den observeret data, hvorefter der vil blive modeleret en lineær linje, som passer bedst til den givne datas forhold. Denne model vil blive brugt til videreformulering af projektets problemanalyse. Modellen er ikke optimal til dette formål, dog er den valgt på grund af prioritering af tid fremfor prioritering af bedre, men ikke-realistiske resultater.

3.1 Metodologi

Tidligere metoder bliver brugt i dette tilfælde til videre arbejde med datasæt. En gennemgang af metoden vil minimere fejl, når der skal dannes en model. Herefter vil der udelukkende blive brugt data fra SBD-Federationen, da de har et højt antal atleter blandt alle løft. Ud af disse ender man med ca. 21.000 observationer efter datacleaning er færdigt og dette ses som et tilstrækkeligt antal til regressions modellen.

Derudover er det vigtigt vi primært bruger variabler som giver mening - de skal af den grund have en logisk relation til atletens styrke og vægt.

Overstående figur illustrere korrelationen mellem variablerne, hvilket er glimrende til søgen efter relevante variabler. Det ses at kropsvægt har nogenlunde samme korrelation til hhv. højeste løft i squat, bænkpres og dødløft. Den har en lav korrelation til Total Load og Alder.

Squat har derimod den højeste korrelation til Total Load, hvor bænkpres har den laveste. Hvilket betyder bænkpres har mindst indflydelse på ens stigning i Total Load og dens sum må derfor være domineret af Squat og Dødløft. Da vi helst vil have variabler med tættest og højest mulige korrelation til BodyweightKg, ligner det at Best3SquatKg, Best3BenchKg og Best3DeadliftKg bliver variabler til regression.

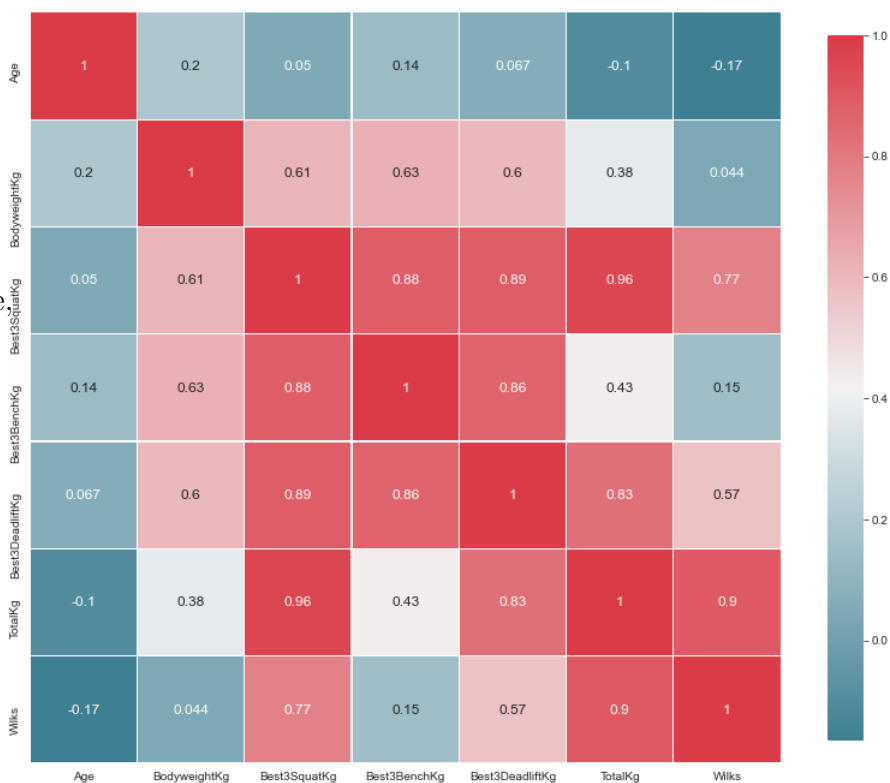


Figure 3.1.1: Korrelations Heatmap til visualisering af variablers sammenhæng.

Efter vi har udvalgt vores variabler kan der dannes en lineærgraf til visualiseringen. Dette vil give et bedre overblik over ens models håndtering af de enkelte variabler.

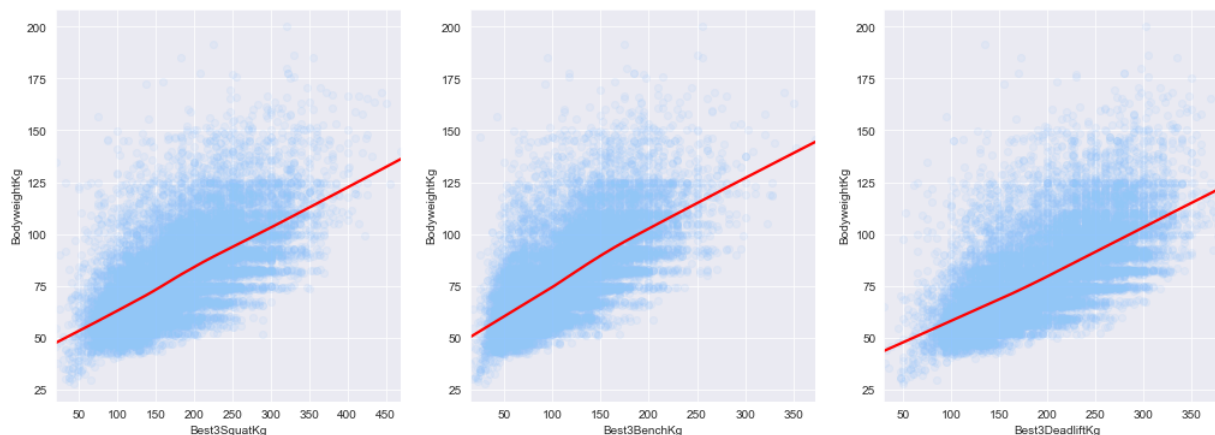


Figure 3.1.2: Visuel præsentation af den lineære sammenhæng mellem styrke og vægt.

Target		Features	Observations, n	R2
Data	BodyweightKg	Best3SquatKg	20.000	0.435
		Best3BenchKg		
		Best3DeadliftKg		
Cross-Validation				
1	2	3	4	5
0.44	0.455	0.40	0.43	0.44

Som forventet ser vi en stærk lineær sammenhæng mellem atletens styrke og vægt. Det ses atter igen at dødløft er den mest lineære, derefter squat og til allersidst bænkpres. Dog gælder naturens absolutte styrke i høj grad. Dens R2 værdi er forholdsvis lav og det kan skyldes den høje varians i alle dens samples, hvor det er mest tydeligt i de højere kropsvægte, hvor variansen er endnu højere.

Modellen vil passe bedre, hvis den blev specificeret til forskellige vægtklasser, samt ens samples kun inkluderet atleter, der har deltaget i en bestemt mængde konkurrencer for at øge deres effekt.

3.2 Absolut Styrke og WILKS

WILKS ide er at udligne absolut styrke og dens naturlige bias mod køn, vægt og deraf ens kropslige proportioner. Det ses dog at mænd stiger lineært i sammenhæng med deres kropsvægt, imens kvinder falder i WILKS. Udefra grafen kan man nævne af kvinder med en lavere kropsvægt, vil have nemmere ved en højere WILKS score end kvinder med en tungere kropsvægt. Det modsatte er gældende for mænd, hvor en højere kropsvægt vil give en højere WILKS.

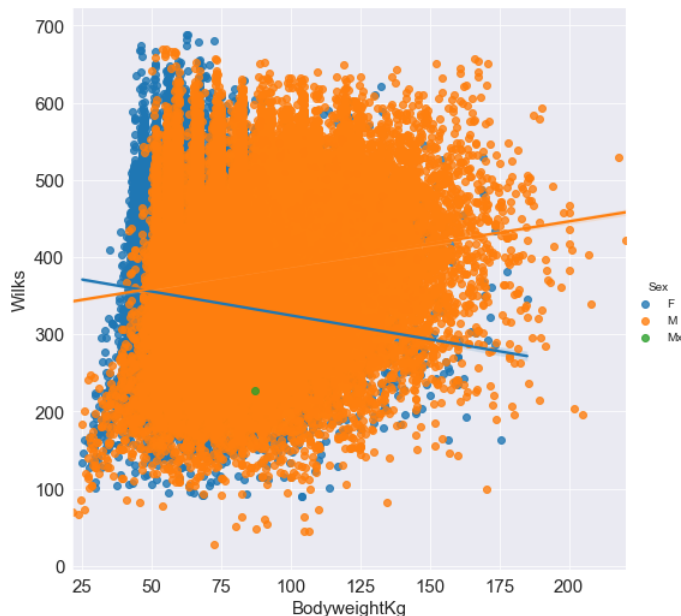


Figure 3.2.1: Sammenligning mellem forskellige køn's kropsvægt og WILKS

3.3 Classic versus Equipment

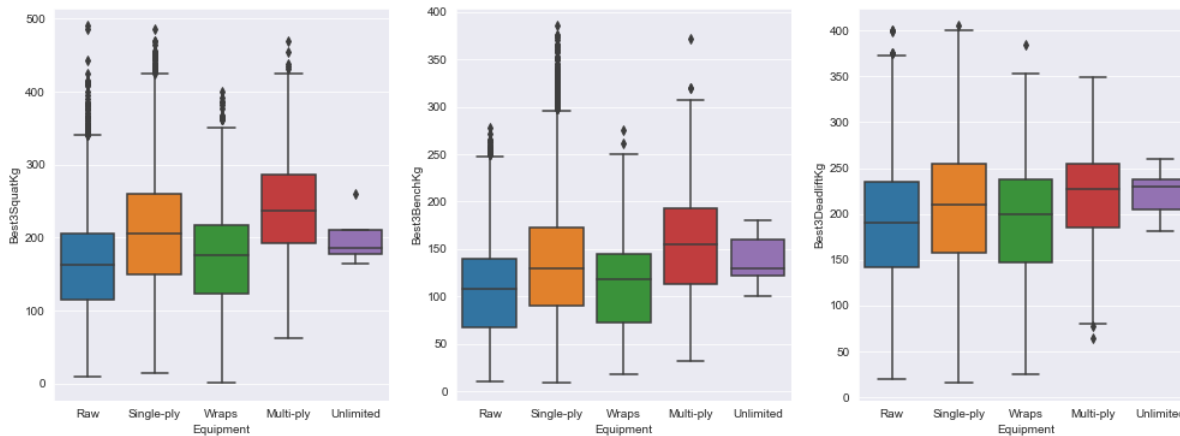


Figure 3.3.1: Styrkesammenligning

Ved nærmere analyse af styrkeforskellen mellem RAW / Classic og udstyr kan der ses en betydelig forskel. 'Single-Ply' og 'Multi-ply' hentyder til antallet af lag, som deres dragt må have - imens 'Wraps' er kraftige knæbind og 'Unlimited' er alt er tilladt ift. udstyr. Det er som forventet, at man er svagest uden udstyr og det er gældende indenfor alle tre discipliner.

Der ses en stor forbedring i løft indenfor alle typer udstyr, dog ikke besynderlig høj for Wraps. En dragt med et enkelt- eller flere lag har bedst effekt, da det maksimere støtten og kroppen skal derfor bruge mindre energi på balance for samme vægt. Det ses også at 'Unlimited' ikke er høj, som man ellers skulle forvente. Dette kan skyldes at det ikke er den mest populære kategori og der er derfor ikke nok atleter ift. andre kategorier.

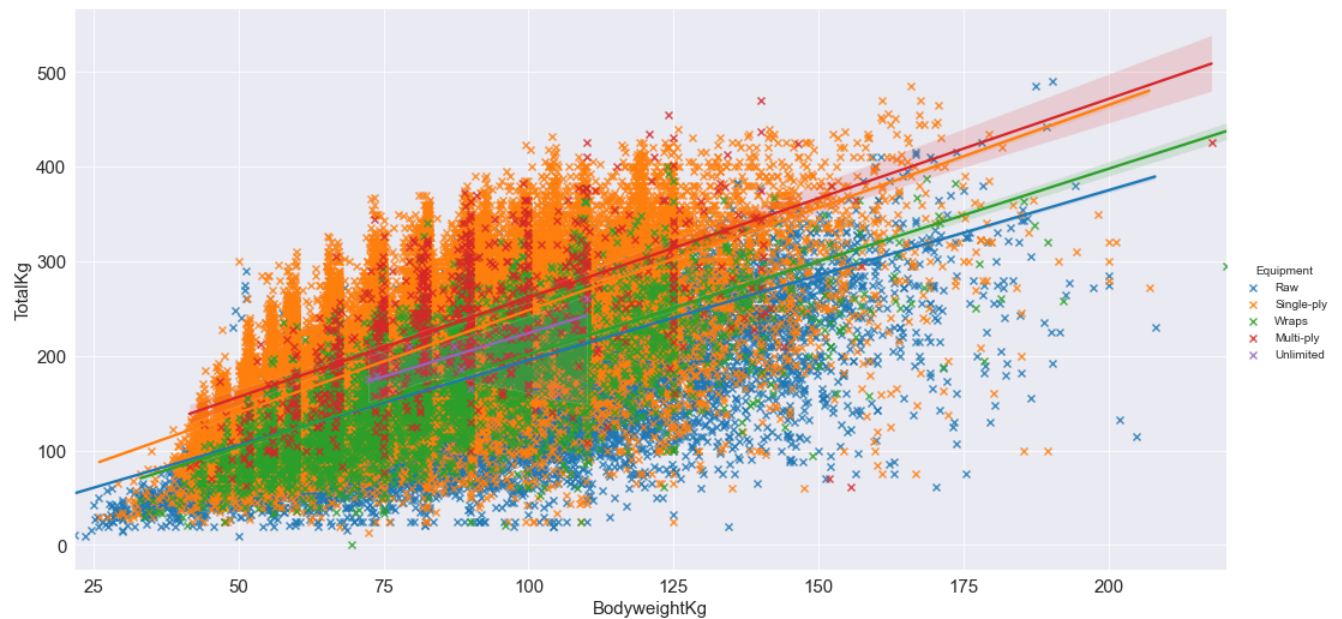


Figure 3.3.2: Sammenligning mellem Totalload og hver klasse af udstyr.

På overstående figur er der en yderligere konstatering af “Multi-ply” effekt på ens absolutte styrke. Grafen er sat med Total Load, fremfor enkelte discipliner, for at se dens totale effekt på summen. “Single-Ply” og “Multi-Ply” er dog ikke langt fra hinanden og det kan overvejes om atletens evne til at maksimere dens effekt er afgørende, da en uerfaren powerlifter ikke ville kunne gøre brug af dem i samme facon som en erfaren.

3.4 Alders Effekt på Styrke

Til undersøgelse af alderens effekt på ens styrke kan nedstående graf vises. Der sker en tydelig stigning fra den mindste observation indtil dens maksimum ved ca. 30 år. Tendensen går igen blandt alle tre discipliner og der noteres en tydelig aftagende lineær funktion. Datasættet har dog ikke en kategori for atleternes erfaring og der vil af den grund ikke kunne drages nogen konklusion udefra denne graf. Atleter der er kommet i gang senere end andre vil have lavere løft end dem med mest erfaring. Dette vil derfor være endnu et bias og grafen kan blive optimeret, hvis der blev dannet en kategori for atletens erfaring.

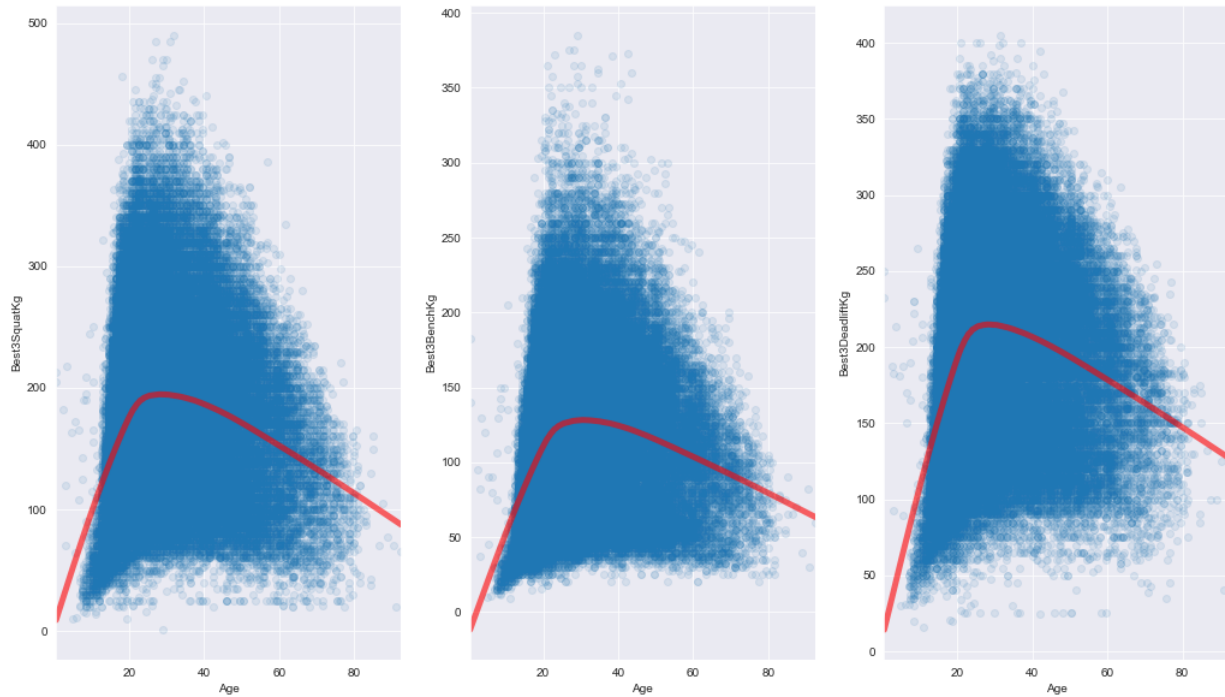


Figure 3.4.1: Alders effekt på styrke i alle tre discipliner.

4 Problemformulering

I tidligere kapitel blev der beskrevet, hvordan opgaven ville gå videre med dens initierende problem. Der er blevet undersøgt kilder, data og kigget på projektets indhold, hvor selve analysen forekom. Der vil derfor arbejdes videre med tidligere problem.

- Kan man forudse atletens vægt ved brug af deres maksimale løft i hhv. squat, bænkpres og dødløft?
- Hvordan ser atleters progression ud over tid, og hvornår er de på deres stærkest / svageste?
- Hvilke variabler er vigtigst ift. absolut styrke?

5 Konklusion

Dette projekts formål var at teste om absolut styrke er lineært, se på kønsforskelle og alderens indvirkning på ens mekanisk funktion, samt at finde de variabler som der er vigtigst for absolut styrke i datasættet.

Dette er der givet et forsøg på og det er bevist, at man i en vis omfang godt kan forudse en Powerlifters vægt ud fra hans styrke i de tre primære discipliner.

6 Diskussion

Der har været en del tidspres i projektet og det er ikke blevet lige så godt, som man ellers kunne ønske. Der kunne foretages en del mere datacleaning, samt optimering af ens features. Man kunne primært vælge atleter, som har stillet op til et minimum af bl.a. fem konkurrencer, for at sortere de mere erfarne frem.

Modellen kunne også blive optimeret pr. køn, vægtklasse, erfaring - der alt sammen er fine ideer, dog ikke realistiske i denne omfang.

Under selve Dataanalysen ville jeg ønske, at jeg kunne komme videre ind i kønsforskelle, vægtklasser og se mere ind på de fysiske karakteristikker, som danner en god powerlifter. Selve projektets model er desuden ikke ordentligt dannet og selvom den passer, samt giver en fin estimation. Så ville en anden model muligvis passe bedre, da man i sig selv ikke kan sætte en lineær funktion til absolut styrke, da det ville være ensbetydende med uendelig styrke.

OpenPowerlifting Dataset Author = "Sean Stangl" Addendum = "(Tilgået 12-18-2021)" Url = "<https://openpowerlifting.gitlab.io/opl-csv/>

World Health Organization Title = "WHO Vækstkurve", Author = "WHO", Addendum = "(Tilgået 12-12-2021)", Url = "<https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/weight-for-age-5to10-years>"