A Holistic Investigation of Powerlifting via Data Analytical Techniques

u2001820

Abstract

Powerlifting is a sport in which individuals compete to lift the most weight for their age/weight class in three barbell lifts: the squat, the bench press, and the deadlift. This report will be focussed on utilising techniques in data analysis to identify patterns in powerlifting between genders, drugs-tested athletes and the world's pound-for-pound best powerlifters.

I. Introduction

The fitness industry has increased in popularity and prevalence over the past decade moreso than any other. This is thanks in large part due its propagation through social media (through the likes of Gymshark, MyProtein, etc.) as well as the impressive feats of strength, endurance and agility on display from the world's best athletes in recent times. From Eddie Hall's 500kg deadlift in 2016 to Kelvin Kiptum's recent 2:00:35 marathon world record, the limits of human capabilities appear limitless, with previously impossible-seeming records smashed over and over again. In this report, we will focus our attention on the sport of powerlifting. In particular, we aim to explore factors affecting powerlifting performance, as well as investigating nuances in the available data with the help of techniques in data science and statistics.

Most of the data handling for this project has been completed in Python. The GitHub repository linked here contains all the relevant files. Where relevant, these files will be hyperlinked in each corresponding chapter of this report. Some analysis will also be conducted using Weka, an open-source software providing tools for data preprocessing, implementation and visualisation. The main Weka results discussed will be present in this report but not in the GitHub repository.

II. The Data

File name: Powerlifting_data_cleaning.ipynb

II.I. Origin

For this report, we are using data gathered from OpenPowerlifting.org, a public-domain archive of powerlifting history. This data is provided in a more convenient format on Kaggle, which is where we begin. The raw dataset provides an overview of the OpenPowerlifting archive up to April 2019. The dataset is comprised of 1,423,354 rows and 37 columns. Due to the sheer quantity of data, we decompose this report into multiple separate sections, each investigating an individual question related to the dataset. Although some initial general data cleaning is performed, each investigation may require bespoke cleaning methods: in any case, these will be mentioned in the respective section.

II.II. Technologies used

The data cleaning conducted for this report, unless stated otherwise, was entirely conducted using the Pandas package in Python. Plots were created using either Matplotlib or Seaborn. Other Python packages used include NumPy, for handling data arrays and vectorised data operations, and SciPy, for conducting statistical computations.

II.III. Overview of Features and Initial Cleaning

This section is dedicated to describing the main columns of the dataset and the initial means of data cleaning. The columns of the dataset are italicised for clarity.

1. *Event*: This column details the event the individual participated in. Some competitors only took part in one of the three the main lifts (of squat, bench and deadlift). 75% of individual records partipicated in all three lifts, indicated by the value 'SBD' in this column. These are the only rows we keep as part of our investigation; for the sake of this report, we are interested in performance across all three lifts. If a competitor only takes part in, say, the squat, they will have no dataset values for the bench press or deadlift, restricting our analysis considerably.
2. *Equipment*: The value 'Raw' means that the individual competed with no additional equipment. The value 'Single-Ply' refers to the individual's powerlifting suit. Powerlifting suits are designed with layers of high-tensile strength fabric. 'Single-Ply' refers to the suit consisting of just one of these layers, whereas 'Multi-ply' means two of these layers or more. The attribute 'Single-Ply' makes up 55% of the column's values, with 'Raw' as the second most common value at 33%: 'Multi-ply' shares the last 12% with the final unique value in the column, 'Wraps', which refers to the knee wraps worn to support lifters' knees, especially during the squat and deadlift.
3. The *Age* and *AgeClass* columns initially contained 573,802 and 555,719 missing values respectively. I first dropped the rows which contained neither value, leaving a dataset where at least one of the two values was present in every row. For each row containing an AgeClass but not an Age, the average of the age class was taken and assigned that value as the record's age. The converse case

was simpler, with the age simply inserted into the correct class.

4. There were only 12,389 missing *BodyweightKg* values. These represent very small proportions of the entire dataset and so, for simplicity, records with missing values in these columns were removed.
5. *WeightClassKg*: Different federations may break competitions down into different weight classes. As such, the values in this column are given with respesct to the corresponding competition in which the individual participated. This means that weight classes are inconsistent across the entire dataset. In order to resolve this, we have created a custom version of this column with increments of 10kg at a time, from 40-50kg all the way to 120-130kg and finally 130+kg, hence standardising the values and allowing for comparisons to be made between competitions as well as competitors.

II.IV. Rankings within categories

Rankings within categories: Each competitor is allowed three attempts of each of the three main lifts. The largest weight lifted for each main lift is taken and summed up to provide the value of *TotalKg*, which is the main value used to determine competitor placement. The dataset also includes values for fourth attempts, but these values were not included in the calculation for *TotalKg*. There were only 801 values missing in this column, and it was not difficult to compute and insert them manually. However, there were tens of thousands of records for which the values of *Best3SquatKg*, *Best3BenchKg* and *Best3DeadliftKg* were not present, and so these have been filled in in using the corresponding record values. For example, the value for *Best3SquatKg* for each record was assigned as the maximum of *Squat1Kg*, *Squat2Kg* and *Squat3Kg*. Any failed lift attempts were indicated in the dataset by the prefix of a minus sign. This means that, if a competitor failed all three lifts, my data imputation would leave a negative value in the *Best3_____Kg* column. I removed such rows from the dataset: in these cases, the competitor does not place well in the competition anyway and so is of little interest to us.

II.V. Rankings between categories

Rankings between categories: the bullet point above explains how rankings are calculated within age/gender/weight categories. However, it is also possible to compare powerlifters across separate categories as well. For example, powerlifting competitions can award trophies to the best junior athlete, encompassing all athletes between the ages 14 and 23, which spans multiple age categories and gender categories. This comparison is possible thanks to powerlifting formulas. The aim of powerlifting formulas is to evaluate the strongest lifters on a per-kilogram basis. The dataset is equipped with four different powerlifting formulas. The *Wilks* coefficient and the *IPFPoints* formula appear to be the most prevalent in scientific literature. The *McCulloch* and *Glossbrenner* points are also provided in the dataset, but they are very similar in formulation to the Wilks coefficient and do not seem to be as common. The Wilks coefficient of a lifter is given by

$$W = y \cdot \frac{500}{a + bx + cx^2 + dx^3 + ex^4 + fx^5},$$

for constants a, b, c, d, e, f and the IPF formula is given by

$$IPF = 500 + 100 \cdot \frac{y - (C_1 \cdot \log(x) - C_2)}{C_3 \cdot \log(x) - C_4},$$

for constants C_1, C_2, C_3, C_4 where x is the lifter's bodyweight in kg and y is the total weight lifted in kg (1, p. 573).

A study was conducted in (1) to investigate the hypothesis that the new IPF formula would be more efficient in determining the 'champion of champions' of the same sex and division. The study used the same dataset as the one we are using, except their study was restricted to results from January 1st 2012 onwards, focussing only on age categories between 24 and 38. It was concluded that Wilks was slightly more efficient than the IPF formula in this regard and so the hypothesis was rejected (1, pp 580-581). Hence, when evaluating inter-category strength in the way described, we will rely on the Wilks coefficient.

III. Battle of the Sexes

File name: Powerlifting_males_vs_females.ipynb

Scientific research suggests that men are stronger than women due to the difference in biology between the sexes. For example, a study described in (2) tested the one-rep maxes of 16 male and 14 female resistance-trained athletes in the squat, bench and deadlift. Female strength was reported to have been lower by 59.2%, 57.2% and 56.3% in the three lifts respectively (2, p. 8). That said, this was a small sample size and not specific to powerlifting. Hence, our aim in this section is to investigate powerlifting strength between men and woman for the dataset in question, aiming to quantify the distinction (if any) through empirical and probabilistic methods.

III.I. Comparing metrics

In order to simplify the comparisons made between the athletes, we filter the *Equipment* column to consider only contestants using 'Single-ply' suits. The 'Tested' column determines whether the lifter has been tested for drug use or not. There are only two values present in this column: 'Yes' or NaN (i.e. no value in the record). To guarantee that we are dealing with athletes who have not been using drugs, we filter the data to ensure that 'Tested = Yes'. Moreover, there are many competitors with multiple entries in the dataset. This could be due to said participants entering multiple competitions. As such, we group records by unique entrants, aggregating with the mean for their performance metrics in the three lifts. A sample of the resultant DataFrame after these modifications is shown in Figure 1.

With this done, we can construct a set of histograms depicting the total weight lifted (in kg) across males and females. Figure 2 includes both histograms across weights lifted and, as per the discussion on lifting formulae discussed in ....., also

	Name	Sex	Tested	Best3SquatKg	Best3BenchKg	Best3DeadliftKg	TotalKg	BodyweightKg	WeightClassKg
0	A Abduzhabarov	M	Yes	155.00	110.00	170.00	435.00	74.00	070 - 80 kg
1	A Akins	M	Yes	115.67	90.72	129.27	335.66	107.05	100 - 110 kg
2	A Allmehat	M	Yes	165.00	120.00	170.00	455.00	72.50	070 - 80 kg
3	A Almeida	F	Yes	45.00	25.00	75.00	145.00	44.00	040 - 50 kg
4	A Ashwin	M	Yes	180.00	95.00	210.00	485.00	81.70	080 - 90 kg

Figure 1: The first five rows of the dataset filtered for Question III.

includes kernel density plots according to the Wilks points accumulated. We observe that the claim certainly holds, at least empirically. The total weight lifted has a far larger variance in the case of the males, most likely because of the fact that there are more male competitors in the dataset than females. Moreover, the Wilks points accumulated by males is slightly offset from the curve in the female case. Although the distinction is not as evident as in the case of total amounts lifted, the whole point of the Wilks coefficient is to standardise the comparisons made: the fact that the distinction still remains, albeit in part, indicates the strength difference between the sexes.

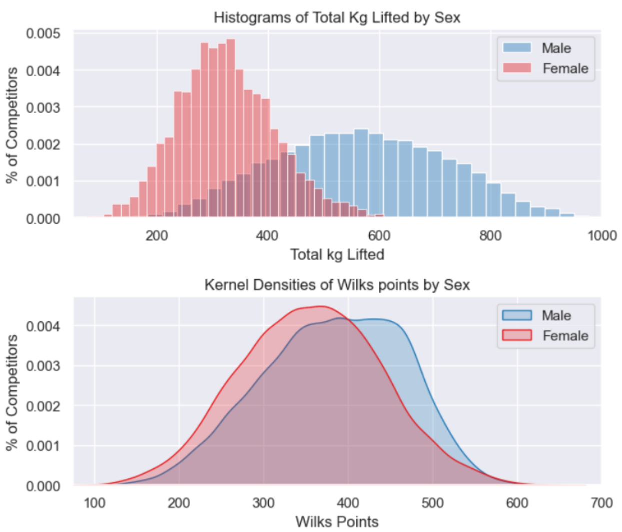


Figure 2: Performance distributions across male and female powerlifters.

III.II. Comparing a random male to a random female

Another question one may wish to answer is the following: if you select a random male and a random female from the dataset, what is the probability that the female is stronger than the male (i.e. larger total kg lifted in competition)? We can set up the problem probabilistically as follows: Suppose that M is a random variable following the distribution of total kg lifted by males, and F similar for females. Then the random variable representing the difference in total kg lifted between males and females is given by $D = M - F$. We can compute the mean and the variance of D. Assuming that D is normally distributed with these parameters, which is a reasonable assumption given the histogram distributions in 2 we can then plot the distribution of D and evaluate it either side of 0. For the mean, we have that, $\mu_D = \mu_M - \mu_F$. Assuming independence between the male and female data, we have that $\sigma_D^2 = \sigma_M^2 + \sigma_F^2$. If $D > 0$, then the male M is stronger than the female F; if $D < 0$, then F is stronger than M. This setup is illustrated in Figure 3.

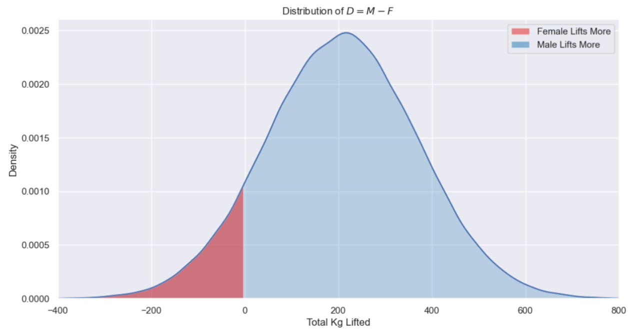


Figure 3: Distribution of D.

We can obtain our desired probabilities by appealing to the z-score formulation $z = \frac{X-\mu}{\sigma}$ and finding the area under the corresponding standard Gaussian. In conclusion, if selecting a natural male and a natural female powerlifter, both using single-ply powerlifting suits, there's a 9.07% chance of the female being stronger than the male and a 90.93% of the male being stronger than the female. This strongly supports the scientific literature.

III.III. Comparing individual lifts

It's one thing to compare the total kg lifted. But the dataset provides further granularity, namely the opportunity to compare squats, bench presses and deadlifts between males and females. One way to observe this numerically is to investigate the correlation coefficients between the gender of the lifter and their lift metrics. To do this, we employ the Pearson correlation coefficient. Concretely, give random variables X and Y with standard deviations σ_X and σ_Y respectively, the correlation coefficient is given by

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

For the data in question, we apply the empirical formula. Namely, given two feature vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ with means $\bar{x}$ and $\bar{y}$ respectively, the correlation coefficient between $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$\rho = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}.$$

Python's Seaborn package was leveraged to produce a correlation matrix between the subset of features *Sex*, *BodyweightKg*, *Best3SquatKg*, *Best3BenchKg*, *Best3DeadliftKg* and *Place*. The result is shown in Figure 4. Label encoding the *Sex* attribute with 0 for females and 1 for males, the gender of lifter is moderately positively correlated with all three main lifts, with PMCC values 0.56, 0.61 and 0.61 for the squat, bench press and deadlift respectively. There is also a positive, albeit weaker, correlation between *Sex* and *Wilks* point scored, with a PMCC value of 0.21.

It may be worth noting that the *Place* column appears to have no correlation with the other numerical values in this subset. The reason for this is most likely due to the fact that the dataset contains information related to many different powerlifting competitions. We comment further on this in Subsection VI.II.

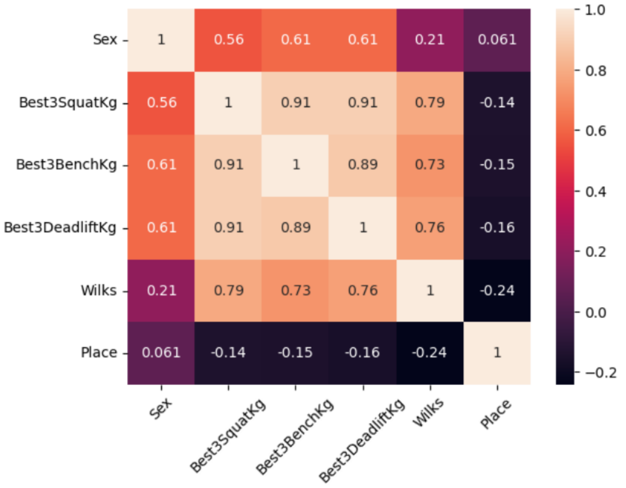


Figure 4: Correlation matrix with label-encoded *Sex* values

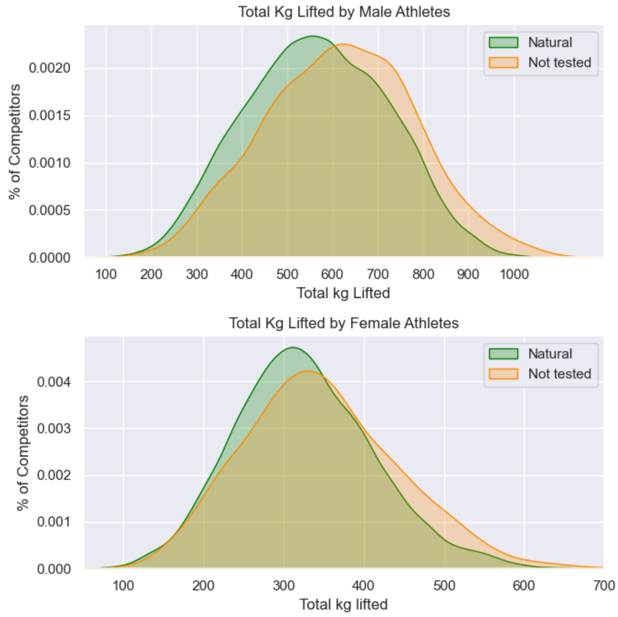


Figure 5: Kernel density estimates of *TotalKg* lifted between the tested and untested athletes.

IV. Natural or Not?

File name: Powerlifting_natural_or_not.ipynb

In sports, a 'natural' athlete is one who does not use Performance Enhancing Drugs, whereas those who do are referred to as 'enhanced' athletes. PED's typically involve hormonal injections which help to increase muscle mass, boosting athletic performances. However, the abuse of such substances comes with many health risks, both physical and mental (3). Studies have also been conducted investigating the negative impact on bodybuilders and powerlifters specifically (4). Given these facts, the question of whether PED's should be allowed in sport is a question of the health risks involved as well as the unfair advantage they provide to users compared to natural athletes. That said, there certainly exist advocates for their use, albeit in a medically controlled environment (5).

With regards to the data, any powerlifter with the value 'Yes' in the *Tested* column is most certainly a natural lifter, at least when tested in time for competition. That said, just because a lifter has not been drug tested (shown by column value NaN), does not necessarily mean that said lifter is abusing drugs. But can the dataset be used to draw a definite conclusion either way?

IV.I. Visual Comparisons

Our aim is to compare the total weight lifted betwen athletes who were drugs-tested against athletes who were not. Due to the conclusions drawn in Section III, we compare males and females separately. Moreover, we modify the 'Yes' and 'NaN' values in the *Tested* column to 'Natural' and 'Not tested' respectively, and, just like in Section III, we restrict our attention to athletes using single-ply powerlifting suits for a more fair comparison between lifts. The kernel density estimators for the male and female cases are shown in Figure 5.

It is difficult to determine the distinction in either case. The distributions in the case of female powerlifters is especially hard to see. Another way to compare the two data distributions is by comparing their quantiles. Figure 6 depicts Q-Q plots using the kde's from Figure 5. If the plots differ a sufficient amount from the line y=x, then one could conclude that the distributions are indeed different. In Q-Q plot for the male case, the curve fluctuates between staying below, and rising above, the baseline. For females, the plot begins to deviate from the baseline at around the 0.0015 mark.

IV.II. Building Classifiers

We continue our investigation by exploring whether classifiers built in Weka can pick out the untested athletes from the tested ones. If our classifiers can accomplish this relatively well, then we have evidence to suggest that untested atheletes are abusing PED's. In what follows, we utilise a train-test split of 66% for our classifiers. Also, we restrict the feature space to the subset *Sex, AgeClass, WeightClassKg, Best3SquatKg, Best3BenchKg, Best3DeadliftKg, Wilks, MeetCountry* and *Year*, which was extracted from the *Year* column. Removing rows with missing values culminated in a dataset of 49,517 records.

The J48 decision tree achieved an accuracy of 94.3%, with precisions of 95.5% and 71.0% for the 'Natural' and 'Not tested' classes respectively. The Naive Bayes classifier achieved an accuracy of 90.8%, with precision of 94.3%% on the 'Natural' class, but only achieved precision of 38.3% on the 'Not tested' class. Finally, k-th nearest neighbour algorithms were also implemented. The best result came from k=3 neighbours, yielding 92.9% accuracy, with precisions of 94.7% and 56.2%. All classifiers incurred relatively low TP rates on the 'Not tested' class, the best being that of the J48 decision tree

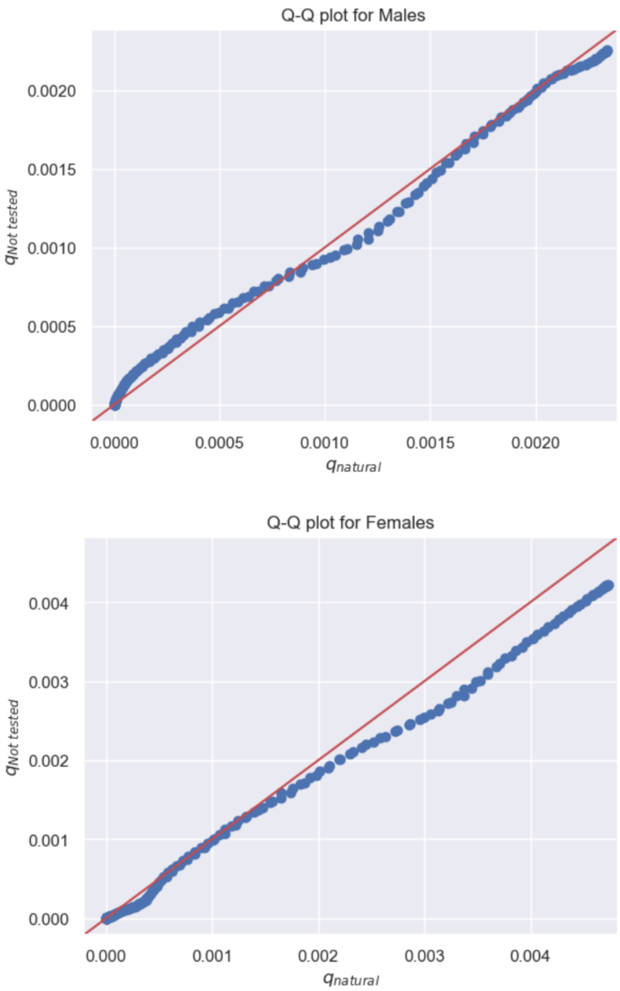


Figure 6: Q-Q plots between the tested and untested athletes.

at 44.2%. Figure 7 displays the full set of performance metrics of the J48 decision tree. Recall that we are interested in whether our classifiers can detect untested athletes and discern them from the tested ones. As such, the precision with respect to the 'Not tested' class is the most important metric for us to scrutinise, given by $\frac{TP}{TP+FP}$. Since the majority of our classifiers struggle to attain a high precision value, we cannot conclude through the use of classifiers alone whether or not untested athletes are abusing drugs to optimise competition performance.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.442	0.015	0.710	0.442	0.545	0.533	0.904	0.534	Not tested
	0.985	0.558	0.955	0.985	0.969	0.533	0.904	0.988	Natural
Weighted Avg.	0.943	0.516	0.936	0.943	0.937	0.533	0.904	0.953	

Figure 7: Performance metrics of the J45 decision tree classifier.

IV.III. Hypothesis Testing

We conclude this section by constructing a hypothesis test to rigorously evaluate whether or not the distribution of the values in the *TotalKg* column is statistically different between the tested and untested athletes. We can set up the hypothesis test as follows:

H_0: Lifters who have not been tested for PED use are natural athletes (i.e. are not using PEDs).

H_A: Lifters who have not been tested for PED use are enhanced athletes (i.e. are using PEDs).

In order to compare the distributions of *TotalKg*, we apply a 5% significance level to the 2-sided Kolmogorov-Smirnov test, a less powerful but more general version of the t-test. (More information regarding the K-S test can be found at (6, pp. 718-720)). Under this context, our null hypothesis H_0 is that the distribution of *TotalKg* for tested lifters is the same as that of non-tested lifters; our alternative hypothesis H_A is that the distributions are different. The test statistic generated by the K-S test compares the cumulative density functions between the two data series. However, the resultant p-values are computed as 0.394 and 0.270 for the male and female data respectively. In either case, the p-value is larger than our significance level and so we cannot reject the null hypotheses. This means that we cannot definitively conclude using this test whether untested athletes are abusing PED's for competitions.

V. Survivial of the Strongest

File name: Powerlifting_the_strongest.ipynb

It would be amiss to analyse a dataset on powerlifting without investigating the features which constitute the strongest athletes. There already exist studies comparing strength in terms of weight lifted (7). That said, given our brief analysis in Subsection II.V, we appeal to the Wilks value to determine the features which contribute to the strongest pound-for-pound powerlifters.

V.I. Clustering

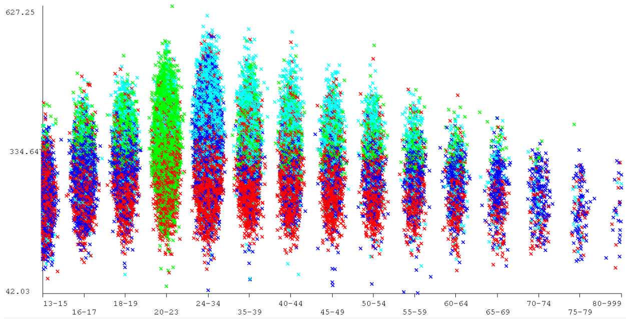
Our aim is to utilise clustering to observe whether there are any specific attributes contributing to the powerlifters with the largest Wilks values. We begin by filtering the dataset for lifters competing with no equipment. For our clustering, we use the following columns of data: *Sex*, *AgeClass*, *WeightClassKg*, *Best3SquatKg*, *Best3BenchKg*, *Best3Deadlift*, *TotalKg*, *Wilks* and we extract the year from the *Date* column to determine whether the year of lifting has any noticeable impact on Wilks score. This data pre-processing leaves us with 89,662 rows of data to investigate. The K-means clustering algorithm was implemented in Weka, with many different choices for K experimented with. K=4 yielded the most interpretable results. Jitter was applied to the plots in Figure 8 to help with data visualisation. We highlight the key feature distinctions in the following:

- Figure 8a illustrates the Wilks points distributed across the age classes of the no-equipment powerlifters. Almost all the age classes feature a similar distribution of data points across the 4 clusters. The main exception is the age class ‘20-23’, in which a very large proportion of data points belong to the green cluster, the cluster whose centroid has the second largest Wilks score. This suggests the importance of age in optimising one’s Wilks score: younger adults have the best pound-for-pound strength.

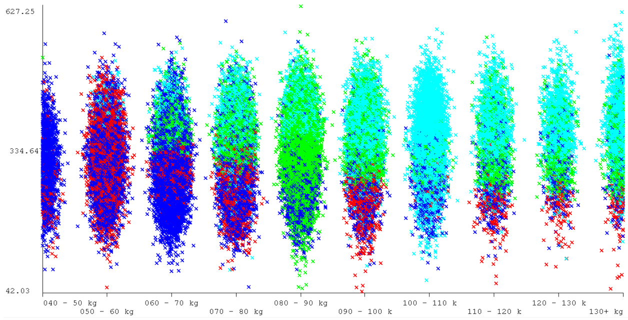
- Figure 8b portrays the Wilks points distributed across weight classes. Colour-wise, the ‘100-110kg’ weight class contains the largest proportion of data points belonging to the cluster with the largest centroid Wilks score. The second most notable weight class is ‘80-90kg’, in which we have data points belonging to the top two clusters in order of centroid Wilks score (the turquoise and green clusters). The ‘90-100kg’, ‘110-120kg’, ‘120-130kg’ and ‘130+kg’ classes share similar clustering patterns, while the classes on the lower end of the spectrum are populated moreso by data points belonging to the two clusters with the lowest centroid Wilks scores.

- Figure 8c showcases cluster arrangements across the years. Unfortunately, there is not enough data in the raw equipment category pre-2000s for any meaningful conclusions to be made. With the data we have, there does not appear to be an obvious relationship between year of competition and Wilks score achieved.

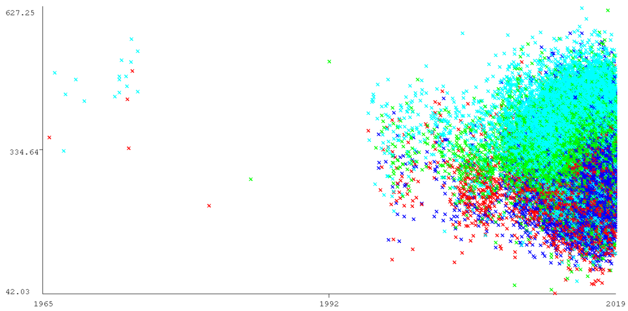
- Figure 8d highlights the distinction between males’ and females’ Wilks points. The datapoints for females are contained primarily within the low-Wilks cluster, whereas the males’ data points seem more evenly spread across the remaining three clusters. This was a rather surprising find, contrasting the low disparity between the kde plots from Figure 2.



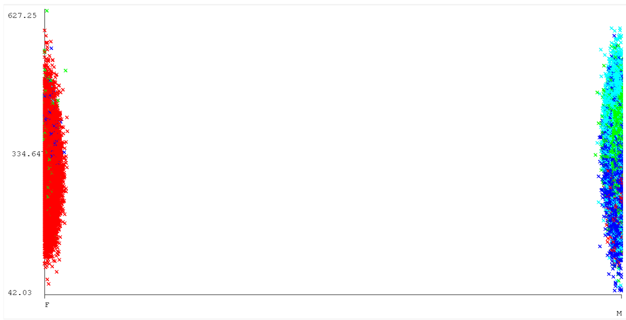
(a) Clustering across age classes



(b) Clustering across weight classes



(c) Clustering across years



(d) Clustering between the sexes

Figure 8: K-means implementations for Wilks scores in the ’Raw’ category, with K=4. The cluster colour order in ascending order of cluster centroid Wilks score is red, blue, green, turquoise.

VI. Conclusion

VI.I. Outcomes of Main Investigations

This report addressed three main questions surrounding the open powerlifting dataset. We outline the results from our investigations here:

1. **Males vs females**: In the context of drugs-tested athletes wearing single-ply powerlifting suits, men are able to lift more total weight than females. However, when compared using the Wilks coefficient, the contrast is not as evident, indicating that on a pound-to-pound basis the variation may not be as pronounced.
2. **Enhanced athletes**: It could not be concretely determined from the dataset as to whether untested athletes are abusing Performance Enhancing Drugs. Both visual and statistical techniques were applied to explore the question. While data plots portrayed differences in distributions between tested and untested athletes, neither Weka classifiers nor hypothesis tests could ratify the empirical observations. Hence, our discussion on this point remains inconclusive.
3. **Best pound-for-pound no-equipment lifters**: Although the Wilks coefficient promises the potential for inter-category comparisons, there were still some attributes which were favoured by the metric. In particular, the heavier weight classes provided more optimal Wilks scores, young adult categories featured more favourable score spreads, and male lifters enjoyed scores that were clustered in more optimal centroids compared to female lifters. A lack of temporal data in this area meant that a conclusion could not be drawn as to how Wilks scores have changed, if at all, over time.

VI.II. Limitations

Due to the sheer quantity of rows and columns in the dataset, a substantial amount of data cleaning was conducted, not only initially, but also before dealing with each smaller investigation. The decision was made early on to consider all the data on OpenPowerlifting.org, so as to accumulate a holistic summary of powerlifting through time and location. But if these investigations were to be conducted again, it is recommended that one focusses on a subset of the powerlifting competitions, such as considering only powerlifting meets in an individual year, or perhaps isolating the project scope to just one competition. This would help to restrict the variability of the results obtained from analysis- this is exemplified by the seeming lack of correlation between lifted weights and competition placement shown by the correlation matrix from Figure 4. The correct relationship ought to be that better lifts correspond to better competition placing, but this correlation is lost within the 5367 different competitions in the raw dataset. For example, one competition could rank a 200kg deadlift quite low, whereas a separate competition could rank the same lift very highly.

VI.III. Extensions

Due to the limitations on report page count, it was simply not possible for investigations to be conducted on all components of the dataset. However, this provides the perfect opportunity for further analysis. Namely, we did not discuss the data between different powerlifting meets, indicated by the columns *MeetCountry*, *MeetState* and *MeetName*. And althought the *Equipment* column was discussed in this report, no analysis was conducted between equipment categories. Possible questions to explore in relation to these include the following:

- Which countries/states hosted the best powerlifting performances? Are there any relationships between geographical locations and amounts of weight lifted?
- Are there any significant deviations between the frequency of athletes tested for drug use between different powerlifting meets? Could this lend insight into hotspots of PED prevalence geographically?
- Hypothetically, if one was to participate in a powerlifting competition given their own gender, age class, weight class, etc., in which location should they choose to participate in order to optimise their competition placement?

Acknowledgements

The author would like to thank the volunteers responsible for collating the data on OpenPowerlifting.org, as well as employees at Kaggle for providing the full amalgamated dataset in .csv format.

References

[1] P.-M. Ferland, M.-O. Allard, A. S. Comtois, Efficiency of the wilks and IPF formulas at comparing maximal strength regardless of bodyweight through analysis of the open powerlifting database, International journal of exercise science (2020) 568–582.
URL
`https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7523908/`

[2] S. Bartolomei, G. Grillone, R. Di Michele, M. Cortesi, A comparison between male and female athletes in relative strength and power performances, Journal of functional morphology and kinesiology (2021).
URL
`https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7930971/`

[3] Anonymous, Anabolic steroids and sports: winning at any cost, State of New York Department of Health (2008).
URL `https://www.health.ny.gov/publications/1210/`

[4] M. R. P. Pilar M. Sanjuan, J. L. Langenbucher, Conditional effects of appearance and performance enhancing drugs (apeds) use on mood in powerlifters and bodybuilders, Alcoholism Treatment Quarterly 33 (4) (2015) 444–457.
`arXiv:https://doi.org/10.1080/07347324.2015.1078155`,
`doi:10.1080/07347324.2015.1078155`.
URL `https://doi.org/10.1080/07347324.2015.1078155`

[5] U. Wiesing, Should performance-enhancing drugs in sport be legalized under medical supervision?, Sports Medicine 41 (4) (2011) 167–176.
`arXiv:https://doi.org/10.2165/11537530-000000000-00000`,
`doi:10.2165/11537530-000000000-00000`.
URL `https://doi.org/10.2165/11537530-000000000-00000`

[6] R. H. C. Lopes, Kolmogorov-Smirnov Test, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 718–720.
`doi:10.1007/978-3-642-04898-2_326`.
URL `https://doi.org/10.1007/978-3-642-04898-2_326`

[7] F. Pierre-Marc, A. Pollock, R. Swope, M. Ryan, M. Reeder, H. Kristin, A. S. Comtois, The relationship between physical characteristics and maximal strength in men practicing the back squat, the bench press and the deadlift., International journal of exercise science (2020) 281–297.
URL
`https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7039481/`