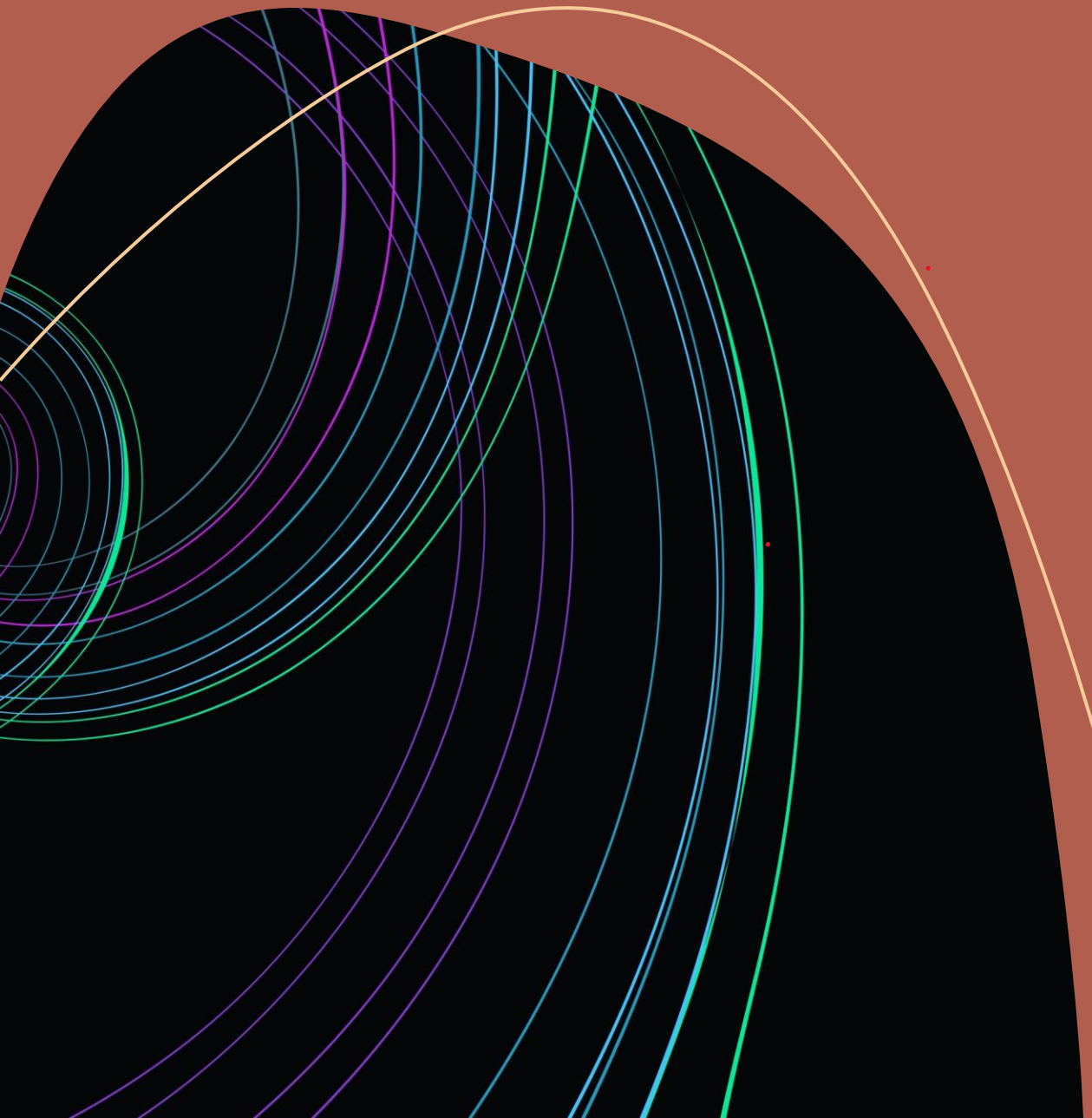


How Are Individual Lifts & Attributes Driving Powerlifting Totals?

Charlie Evert

Background

- How can we predict lifting totals using a combo of variables?
  - Which models are best for quantitative vs qualitative variables?
- Powerlifting competitions measure 3 lifts & score based on totals
  - Bench Press
  - Squat
  - Deadlift
- Competitions vary with PED* tests, equipment**, genders & dates

*Performance Enhancing Drugs such as Anabolic Steroids, Human Growth Hormone, etc.

**Lifting Straps, Bench Press Shirts, wraps, etc. that increase performance potential on lifts

Data Summary

- Source: Open Powerlifting
  - https://openpowerlifting.gitlab.io/opl-csv/
- Nominal Variables → Binary Columns
- N = 2,536,437 rows of data
  - n = Randomly sampled 1000 rows of data from N*
    - 70% Training (700), 30% Validation (300)
  - 16 Variables | 6 Interval | 10 Binary | 1 Target (Total Weight Lifted in kg)

*Population data would crash SAS Enterprise Miner, and 1000 rows didn't crash the application

Attribute Descriptions

Attribute	Description
Male	Gender where Male = 1 and Female = 0
Multi-ply	Two layers of equipment or more are permitted
Raw	No equipment is permitted
Straps	Lifting straps on wrists to increase grip are permitted
Unlimited	All equipment is permitted
Wraps	Tight cloth wraps to support joints are permitted
Bodyweight Kg	Bodyweight of the lifter, in kilograms
Best of 3 Squat Kg	Highest amount squatted of 3 attempts, in kilograms
Best of 3 Bench Kg	Highest amount bench-pressed of 3 attempts, in kilograms
Best of 3 Deadlift Kg	Highest amount deadlifted of 3 attempts, in kilograms
Total Kg	Aggregate of the Best of 3 Bench, Squat and Deadlift
Not Tested for PED	Not tested for Performance Enhancing Drugs (AKA PEDs are permissable) = 1, and tested for PED = 0
Year	The year that the lift took place

2 Sets of Models for Different Results

- There are 2 models because of the direct influence each individual lift has on predicting totals
- We want to also examine attributes, so the model must be split

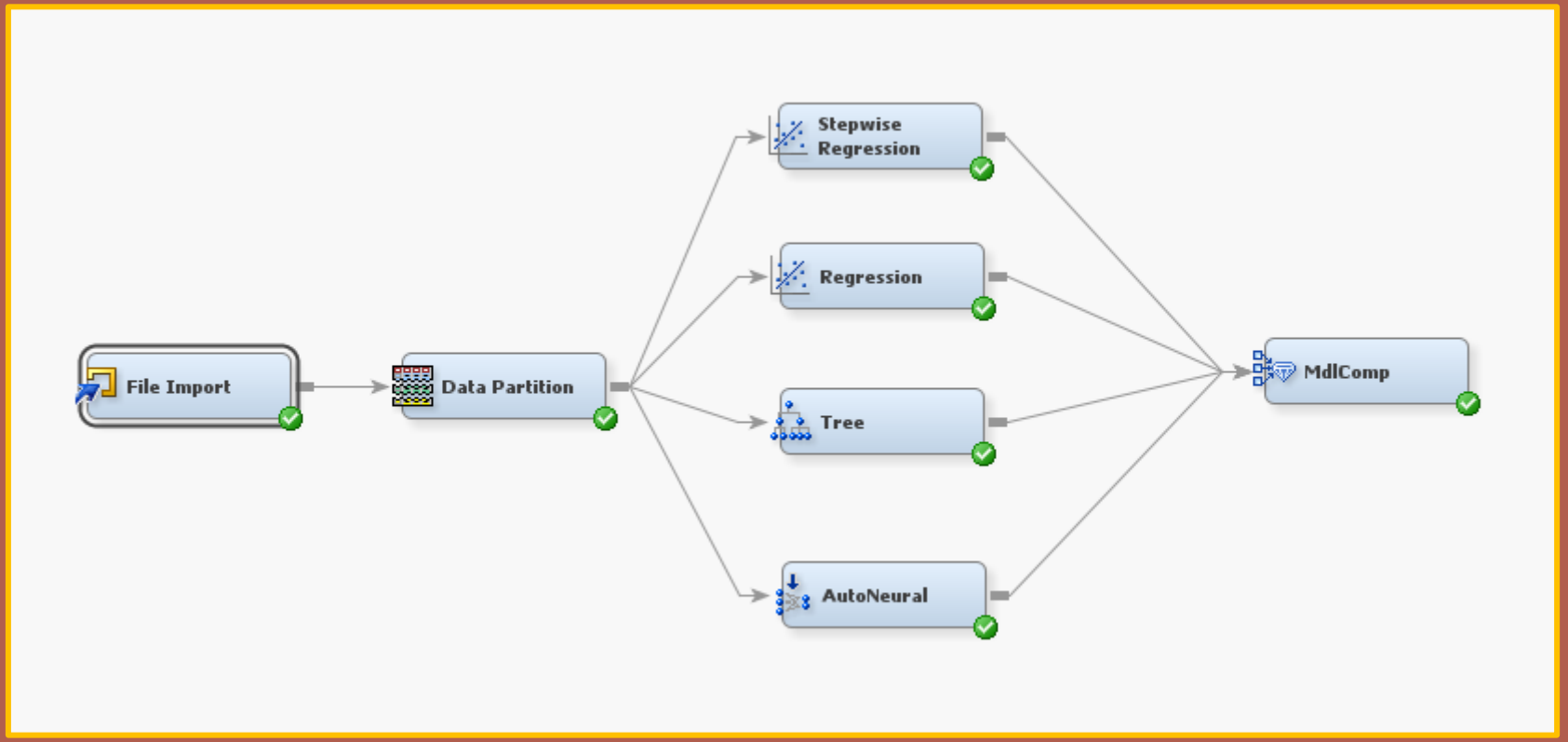
Model 1: Includes 3 Lifts & All Variables

Name	Role	Level	Drop
Best_of_3_Benc	Input	Interval	No
Best_of_3_Dead	Input	Interval	No
Best_of_3_Squa	Input	Interval	No
Bodyweight_Kg	Input	Interval	No
Female	Input	Binary	Yes
Male	Input	Binary	No
Multi_ply	Input	Binary	No
Not_Tested_for_	Input	Binary	No
Raw	Input	Binary	No
Single_ply	Input	Binary	Yes
Straps	Input	Binary	No
Tested_for_PED	Input	Binary	No
Total_Kg	Target	Interval	No
Unlimited	Input	Binary	No
Wraps	Input	Binary	No
Year	Input	Interval	No

Model 2: Excludes 3 Lifts

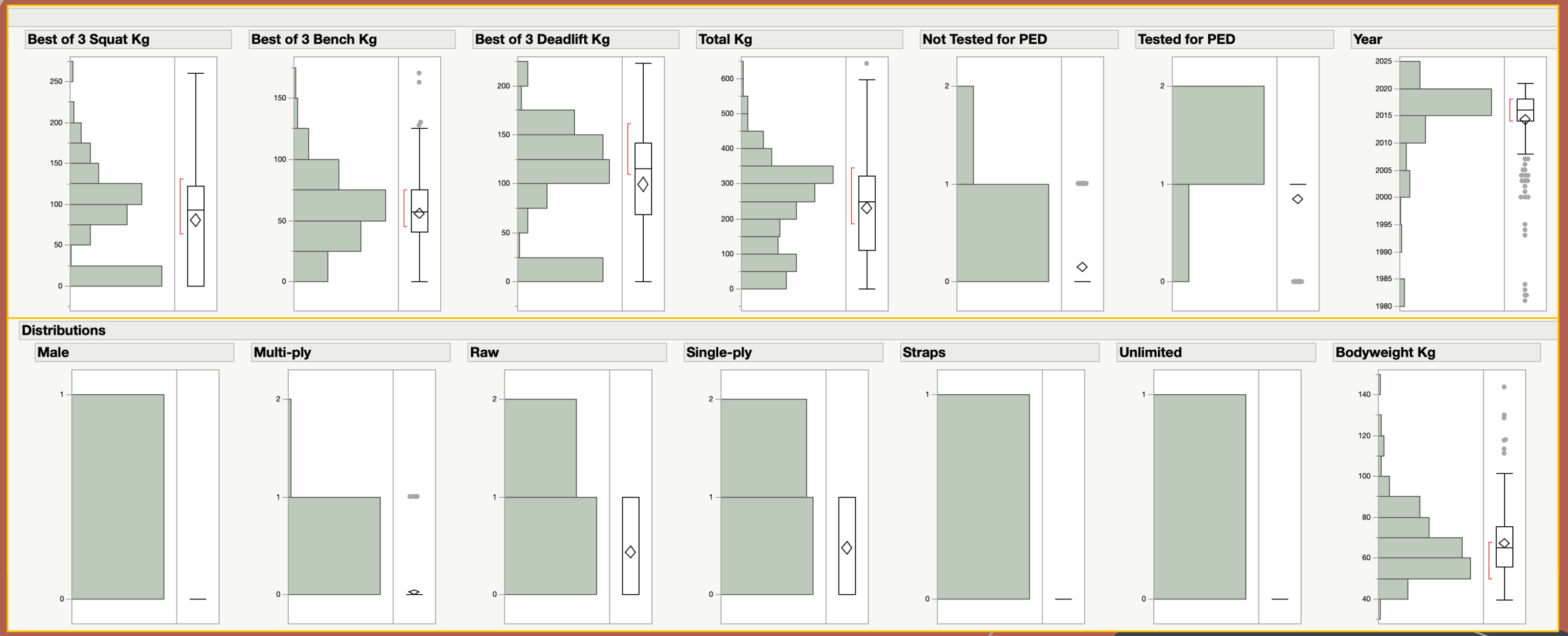
Name	Role	Level	Drop
Best_of_3_Benc	Input	Interval	Yes
Best_of_3_Dead	Input	Interval	Yes
Best_of_3_Squa	Input	Interval	Yes
Bodyweight_Kg	Input	Interval	No
Female	Input	Binary	Yes
Male	Input	Binary	No
Multi_ply	Input	Binary	No
Not_Tested_for_	Input	Binary	No
Raw	Input	Binary	No
Single_ply	Input	Binary	Yes
Straps	Input	Binary	No
Tested_for_PED	Input	Binary	No
Total_Kg	Target	Interval	No
Unlimited	Input	Binary	No
Wraps	Input	Binary	No
Year	Input	Interval	No

SAS Enterprise Miner Diagram*



*Diagrams are the same, despite the inputs & outputs differing

Variable Distributions



Model Comparison

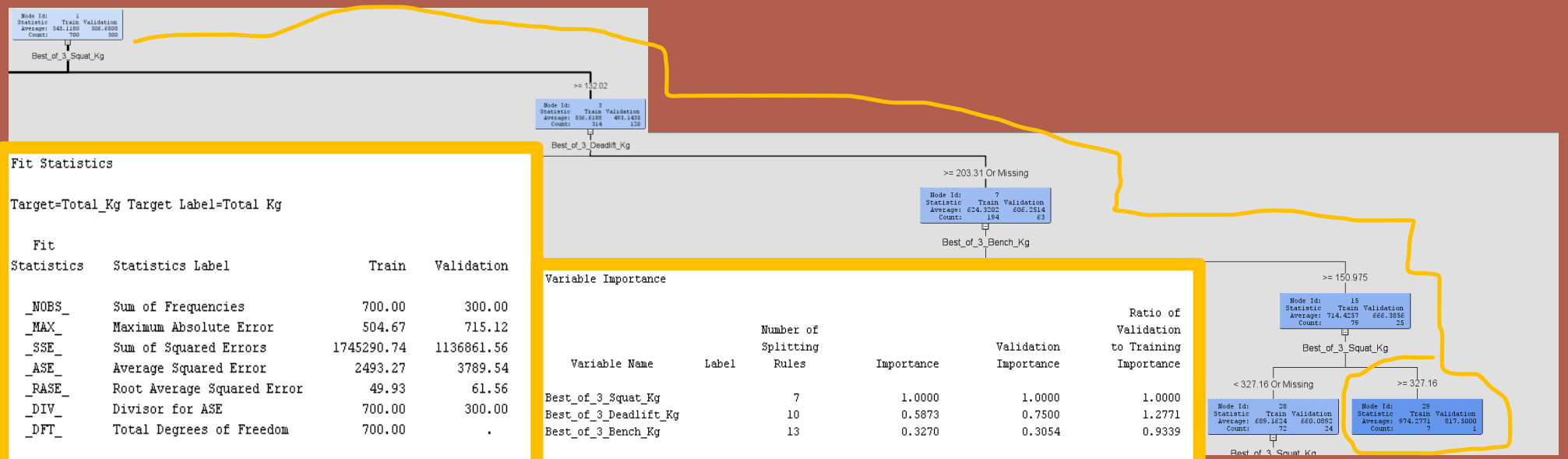
Model 1: Decision Tree has lowest MSE of any model

Selected Model	Predecessor Node	Model Node	Model Description	Target	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Total Degrees of Freedom	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Number of Estimated Weights	Train: Akaike's Information Criterion	Train: Schwarz's Bayesian Criterion	Train: Average Squared Error	Train: Maximum Absolute Error	Train: Divisor for ASE	Train: Sum of Frequencies
Y	Tree	Tree	Tree	Total_Kg	Total Kg	3789.539	700	.	.	.	.	.	2493.272	504.6691	700	700
	AutoNeural	AutoNeural	AutoNeural	Total_Kg	Total Kg	4462.11	700	615	85	85	5296.176	5683.018	1514.905	447.5514	700	700
	Reg2	Reg2	Stepwise...	Total_Kg	Total Kg	5016.612	700	695	5	5	5485.788	5508.544	2496.274	551.4567	700	700
	Reg	Reg	Regressi...	Total_Kg	Total Kg	5062.489	700	688	12	12	5493.971	5548.584	2475.615	553.7066	700	700

Model 2: Neural Network has lowest MSE of model 2

Selected Model	Predecessor Node	Model Node	Model Description	Target	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Total Degrees of Freedom	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Number of Estimated Weights	Train: Akaike's Information Criterion	Train: Schwarz's Bayesian Criterion	Train: Average Squared Error	Train: Maximum Absolute Error	Train: Divisor for ASE	Train: Sum of Frequencies
Y	AutoNeural	AutoNeural	AutoNeural	Total_Kg	Total Kg	38512.87	700	655	45	45	7456.882	7661.68	37201.95	802.6793	700	700
	Reg2	Reg2	Stepwise...	Total_Kg	Total Kg	39769.52	700	693	7	7	7428.199	7460.056	39803.6	683.0605	700	700
	Reg	Reg	Regressi...	Total_Kg	Total Kg	40279.2	700	691	9	9	7430.291	7471.251	39695.28	655.2707	700	700
	Tree	Tree	Tree	Total_Kg	Total Kg	42333.23	700	.	.	.	.	.	39108.98	669.7326	700	700

Model 1’s Decision Tree: Optimal Path



Fit Statistics

Target=Total_Kg Target Label=Total Kg

Fit Statistics	Statistics Label	Train	Validation
NOBS	Sum of Frequencies	700.00	300.00
MAX	Maximum Absolute Error	504.67	715.12
SSE	Sum of Squared Errors	1745290.74	1136861.56
ASE	Average Squared Error	2493.27	3789.54
RASE	Root Average Squared Error	49.93	61.56
DIV	Divisor for ASE	700.00	300.00
DFT	Total Degrees of Freedom	700.00	.

Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
Best_of_3_Squat_Kg		7	1.0000	1.0000	1.0000
Best_of_3_Deadlift_Kg		10	0.5873	0.7500	1.2771
Best_of_3_Bench_Kg		13	0.3270	0.3054	0.9339

Model 1's Decision Tree: Optimal Path

The Highest-Lifting Branch

- Squats at/over 132kg
- Deadlifts at/over 203kg
- Bench-presses at/over 150kg
- Squats at/over 327kg

- This branch makes up only 7 of the 1000 total observations
- Thus, clearly squat numbers drive the highest lifting totals and should be trained the most for a competitive advantage

Model 1's Decision Tree Results & Output

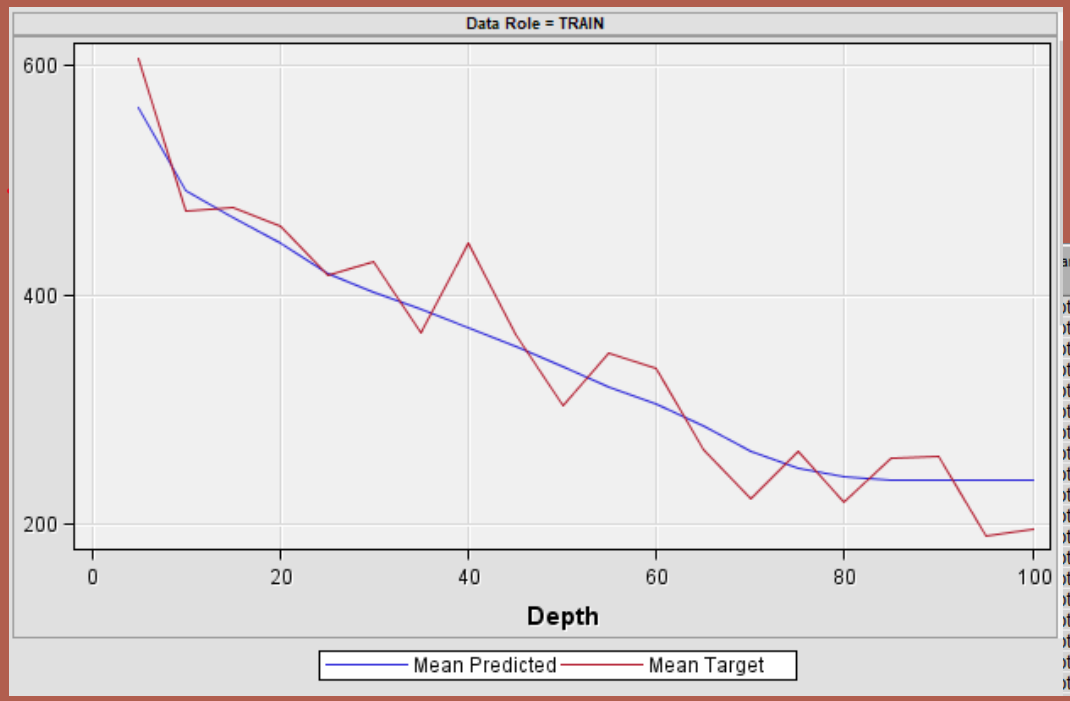
- Squat is the most important predictor of a heavy total, followed by deadlifts and then bench-press
- This model omits every other variable due to the importance of the 3 (despite them being inputs)

Assessment Score Rankings

Data Role=TRAIN Target Variable=Total_Kg Target Label=Total Kg

Depth	Number of Observations	Mean Target	Mean Predicted
5	37	785.351	785.351
10	44	658.806	658.806
15	50	601.136	601.136
20	20	562.583	562.583
25	28	524.031	524.031
30	53	483.443	483.443
35	28	436.448	436.448
40	50	373.657	373.657
45	58	318.207	318.207
55	66	269.572	269.572
65	27	234.584	234.584
70	62	182.375	182.375
75	33	162.346	162.346
80	5	157.100	157.100
85	66	118.109	118.109
90	11	64.091	64.091
95	30	56.248	56.248
100	32	25.331	25.331

Model 2's Neural Network



The Neural Network is great for basing predictions on attributes other than lifts

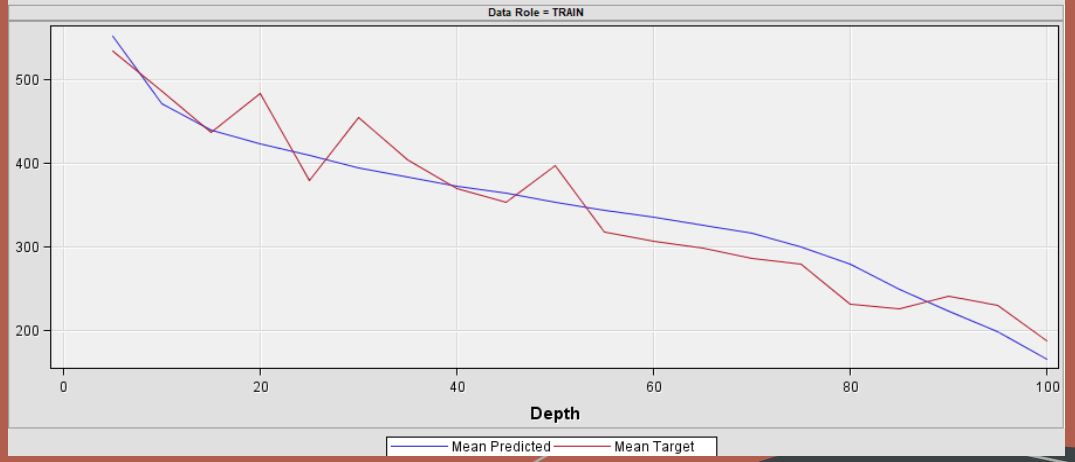
arget	Target Label	Fit Statistics	Statistics Label	Train	Validation
tal Kg	Total Kg	DFT	Total Degrees of Freedom	700	.
tal Kg	Total Kg	DFE	Degrees of Freedom for Error	655	.
tal Kg	Total Kg	DFM	Model Degrees of Freedom	45	.
tal Kg	Total Kg	NW	Number of Estimated Weights	45	.
tal Kg	Total Kg	AIC	Akaike's Information Criterion	7456.882	.
tal Kg	Total Kg	SBC	Schwarz's Bayesian Criterion	7661.68	.
tal Kg	Total Kg	ASE	Average Squared Error	37201.95	38512.87
tal Kg	Total Kg	MAX	Maximum Absolute Error	802.6793	530.7106
tal Kg	Total Kg	DIV	Divisor for ASE	700	300
tal Kg	Total Kg	NOBS	Sum of Frequencies	700	300
tal Kg	Total Kg	RASE	Root Average Squared Error	192.8781	196.247
tal Kg	Total Kg	SSE	Sum of Squared Errors	26041367	11553862
tal Kg	Total Kg	SUMW	Sum of Case Weights Times Freq	700	300
tal Kg	Total Kg	FPE	Final Prediction Error	42313.67	.
tal Kg	Total Kg	MSE	Mean Squared Error	39757.81	38512.87
tal Kg	Total Kg	RFPE	Root Final Prediction Error	205.7029	.
tal Kg	Total Kg	RMSE	Root Mean Squared Error	199.3936	196.247
tal Kg	Total Kg	AVERR	Average Error Function	37201.95	38512.87
tal Kg	Total Kg	ERR	Error Function	26041367	11553862

Model 2's Stepwise Regression

- Stepwise Regression for model 2 shows how each attribute affects totals
- Not being tested for PEDs increase totals more than any other attribute
- Wraps are the most beneficial equipment for higher totals

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept		1	4852.6	1855.8	2.61	0.0091
Bodyweight_Kg		1	2.0482	0.3535	5.79	<.0001
Male	0	1	-51.4957	9.9733	-5.16	<.0001
Not_Tested_for_PED	0	1	21.5564	9.4620	2.28	0.0230
Raw	0	1	16.8751	8.4585	2.00	0.0464
Wraps	0	1	-73.8934	17.8189	-4.15	<.0001
Year		1	-2.3140	0.9211	-2.51	0.0122



Limitations & Future Research

- SAS Enterprise Miner crashes with more than 1k rows
  - More data would be ideal for analysis to get better insights
- It would be interesting to have country of origin as a variable
  - Some countries produce better lifters; this could aid in predictions
  - This isn’t possible due to SAS Enterprise Miner crashing
    - Countries coded binarily would create 200+ columns, which it can’t handle

Conclusion

- Using PEDs & equipment undoubtedly increases lifting totals
- Focusing on squats will increase totals the most
- Neural Networks are ideal for prediction from attributes
- Decision Trees are ideal for predicting from 3 lifts
- Stepwise Regression is ideal for understanding attribute effects