Fitness Challenge

Data analytics approach

Agenda

1. Introduction
2. Data Presentation
3. Data Description
4. Data Exploration
5. Model Development
6. Interpretation

Introduction

FIT 143 Challenge

The Fit 143 challenge involves 143 days of tracking workout routine data which would be used to create data dashboards. Workout routines tracked include Running, Push-ups, dumb-bell, sit-ups, bench-press, and jumping-jacks.
The data is stored on a Google Sheet that can be found here: https://docs.google.com/spreadsheets/d/133xo8gUubOxyeB9vTJabCg0aY8DysDvwrxLgWliqBgw/edit?usp=sharing

Description of Columns in Dataset

- Push-ups: Number of push-ups done each day
- Bench-press: Number of bench-presses done each day
- Squats: Number of squats done each day
- Distance run (miles): Distance covered running each day measured in miles
- Duration (minutes): Time taken to complete run measured in minutes
- Average pace (min/mile): Average time taken in minutes to complete one mile while running
- Calories: Amount of calories burnt while running measured in calories.
- Sit-ups: Number of sit-ups done each day
- Jumping jacks: Number of jumping jacks done each day
- Date: Date for each tracked activity

SOURCES OF DATA

This data used in research study are secondary data and were gathered using the Runkeeper. The Runkeeper app is used to track running and generate data.

Data Presentation

push_ups	bench_press	dumb_bell	squats	distance_run (miles)	duration (minutes)	average_pace (min/mile)	calories	sit_ups	jumping_jacks	date	days	week
40	110	20	25	1.1	15.4	14.1	120	200	90	8/11/2020	Tuesday	1
45	112	20	25	1.7	20.2	11.59	194	215	100	8/12/2020	Wednesday	1
30	110	20	25	1.11	21.22	19.18	121	165	50	8/13/2020	Thursday	1
35	110	20	25	1.4	17.57	12.47	164	175	90	8/14/2020	Friday	1
35	60	25	20	1.12	14.09	12.36	127	100	50	8/15/2020	Saturday	1
20	60	30	20	1.12	17.39	14.36	134	50	55.9	8/16/2020	Sunday	2
35	60	20	20	1.42	17.39	12.27	158	100	50	8/17/2020	Monday	2
35	60	20	20	1.23	20.14	16.28	130	100	50	8/18/2020	Tuesday	2
35	100	20	20	1.23	16.5	13.41	138	100	50	8/19/2020	Wednesday	2
35	110	20	25	1.42	18.11	12.47	159	100	50	8/20/2020	Thursday	2
35	60	30	20	1.36	20.04	14.42	150	100	50	8/21/2020	Friday	2
35	60	20	20	1.57	19.57	12.44	179	100	50	8/22/2020	Saturday	2
35	60	20	20	1.3	15	11.35	150	100	50	8/23/2020	Sunday	3
35	110	20	25	1.24	15.35	12.37	140	100	50	8/24/2020	Monday	3
40	60	20	20	1.3	14.23	11.04	151	100	50	8/25/2020	Tuesday	3
35	60	20	20	1.24	13.05	10.31	145	100	50	8/26/2020	Wednesday	3
35	80.1	21.1	30	1.32	16.08	12.17	151.7	123	50	8/30/2020	Sunday	4
35	80.1	21.1	23.3	1.32	16.08	12.17	151.7	123	55.9	8/31/2020	Monday	4
35	110	20	25	1.34	14.01	10.27	160	200	50	9/3/2020	Thursday	4
35	60	20	25	1.32	13.57	10.34	157	100	50	9/4/2020	Friday	4
35	60	20	20	1.13	9.4	8.34	134	100	50	9/7/2020	Monday	5
35	60	20	25	1.23	12.22	10.02	147	100	50	9/8/2020	Tuesday	5
35	110	20	40	1.73	16.08	9.19	206	200	50	9/9/2020	Wednesday	5
35	60	20	20	1.46	13.31	9.17	174	100	50	9/10/2020	Thursday	5

Data Description

Descriptive Statistics						
	N	Minimum	Maximum	Sum	Mean	Std. Deviation
push_ups	24	20.0	45.0	840.0	35.000	4.1703
bench press	24	60.0	112.0	1922.2	80.091	23.7045
dumb bell	24	20.0	30.0	507.3	21.136	2.9236
squats	24	20.0	40.0	558.3	23.261	4.5704
distance run / mile	24	1.10	1.73	31.71	1.3214	0.17061
duration(min)	24	9.40	21.22	386.01	16.0836	2.90572
average pace (min/mile)	24	8.34	19.18	292.09	12.1705	2.38991
calories	24	120.0	206.0	3641.5	151.727	21.2382
sit ups	24	50.0	215.0	2950.9	122.955	43.7299
jumping jacks	24	50.0	100.0	1341.8	55.909	14.6433
Valid N (listwise)	24					

Data Description

(1) PUSH UPS
Observations
- The challenge was done in the space of 24 days
- The minimum push ups per day during the 24 days was 20 push ups
- The maximum push ups per day during the 24 days was 45 push ups
- The average push ups per day was 35 push ups
- While the deviation is of 4.17 push ups
- 840 push ups was done during the 24 days fitness challenge

(2) BENCH PRESS
Observations
- The challenge was done in the space of 24 days
- The minimum bench press per day during the 24 days was 60 bench press
- The maximum bench press per day during the 24 days was 112 bench press
- The average bench press per day was 80 bench press
- While the deviation is of 23.7 bench press
- 1922 bench press was done during the 24 days fitness challenge

(3) SQUATS
Observations
- The challenge was done in the space of 24 days
- The minimum squats per day during the 24 days was 20 squats
- The maximum squats per day during the 24 days was 40 squats
- The average squats per day was 23.261 squats
- While the deviation is of 4.57 squats
- 558 squats was done during the 24 days fitness challenge

(4) DISTANCE RUN / MILES
Observations
- The challenge was done in the space of 24 days
- The minimum distance run (miles) per day during the 24 days was 1.1 miles
- The maximum distance run (miles) per day during the 24 days was 1.73 miles
- The average distance run (miles) per day was 1.3214 miles
- While the deviation is of 0.171 miles
- 31.71 miles was covered during the 24 days fitness challenge

Data Description

(5) AVERAGE PACE MIN / MILE
Observations
- The challenge was done in the space of 24 days
- The minimum duration (minutes) per day during the 24 days was 9.4 minutes
- The maximum duration (minutes) per day during the 24 days was 21.22 minutes
- The average duration (minutes) per day was 16.1 minutes
- While the deviation is of 2.91 minutes
- 292.1 minutes per mile was ran during the 24 days fitness challenge

((6) DURATION / MINUTES
Observations
- The challenge was done in the space of 24 days
- The minimum average pace (min/mile) per day during the 24 days was 8.34 min/mile
- The maximum average pace (min/mile) per day during the 24 days was 19.18 min/mile
- The average average pace (min/mile) per day was 12.171 min/mile
- While the deviation is of 2.39 min/mile
- 386.01 minutes was ran during the 24 days fitness challenge

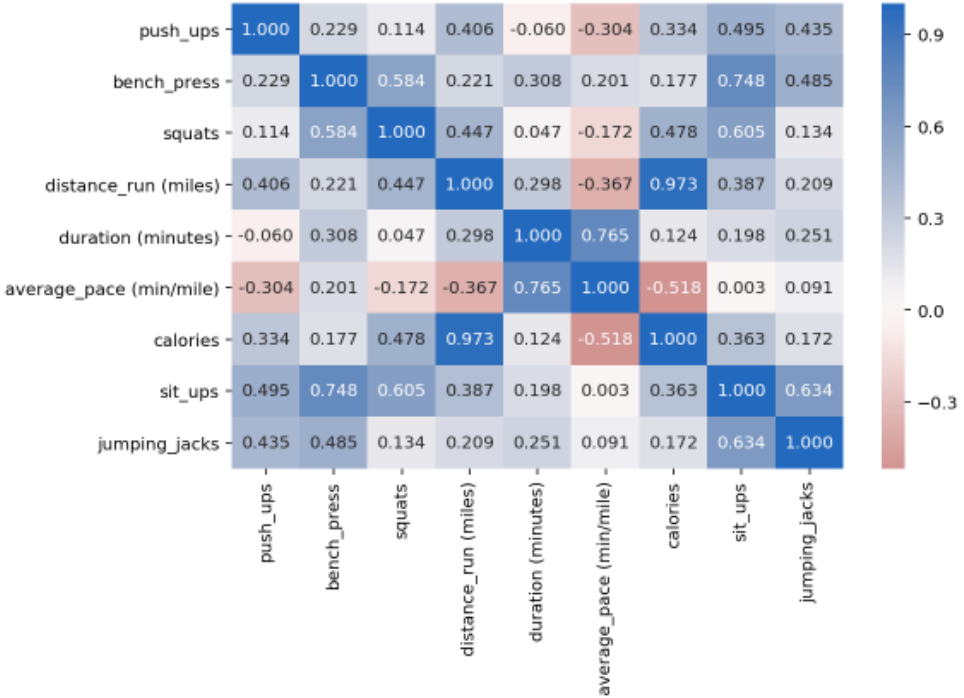
(7) CALORIES BURNT WHILE RUNNING
Observations
- The challenge was done in the space of 24 days
- The minimum calories burnt per day during the 24 days was 120 calories
- The maximum calories per day during the 24 days was 206 calories
- The average calories burnt was 151.73 calories
- While the deviation is of 21.24 calories
- 3641.5 calories was burnt during the 24 days fitness challenge

(8) SIT UPS
Observations
- The challenge was done in the space of 24 days
- The minimum sit ups per day during the 24 days was 50 sit ups
- The maximum sit ups per day during the 24 days was 215 sit ups
- The average sit ups per day was 123 sit ups
- While the deviation is of 44 sit ups
- A total of 2951 sit ups was done during the 24 days challenge

(9) JUMPING JACKS
Observations
- The challenge was done in the space of 24 days
- The minimum jumping jacks per day during the 24 days was 50 jump
- The maximum jumping jacks per day during the 24 days was 100 jump
- The average jumping jacks per day was 56 jump
- While the deviation is of 44 jump
- A total of 1342 jumping jacks was done during the 24 days challenge

Data Exploration

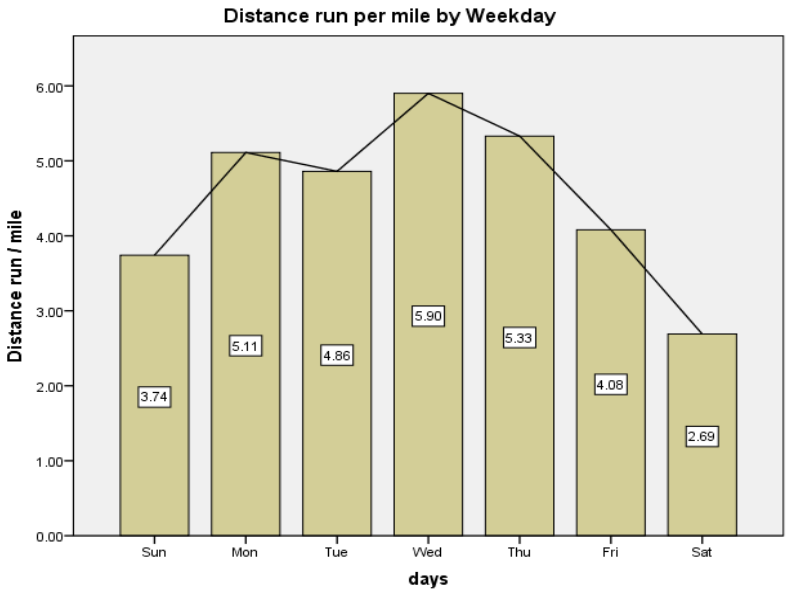
CORRELATION MATRIX



- In these results, the Pearson correlation between calories and distance run per mile is about 0.973. The Pearson correlation between average pace (min/mile) and duration (min) is about 0.765, and between sit ups and bench press is about 0.748; which indicates that there is a strong positive relationship between the variables. The relationship between these variables is positive, which indicates that, as a variable increases, the other variable also increases.
- Some variables are also moderately correlated with each other; (squats and sit ups (r=0.605), jumping jacks and sit ups (r=0.634), (squats and bench press (r=0.584)), which also indicate a positive relationship between the variables.
- Some variables also showed a weak positive correlation between each other.
- While some variables were also negatively correlated with each other, which indicates as a variable increases, the other variable tends to decrease and vice versa.

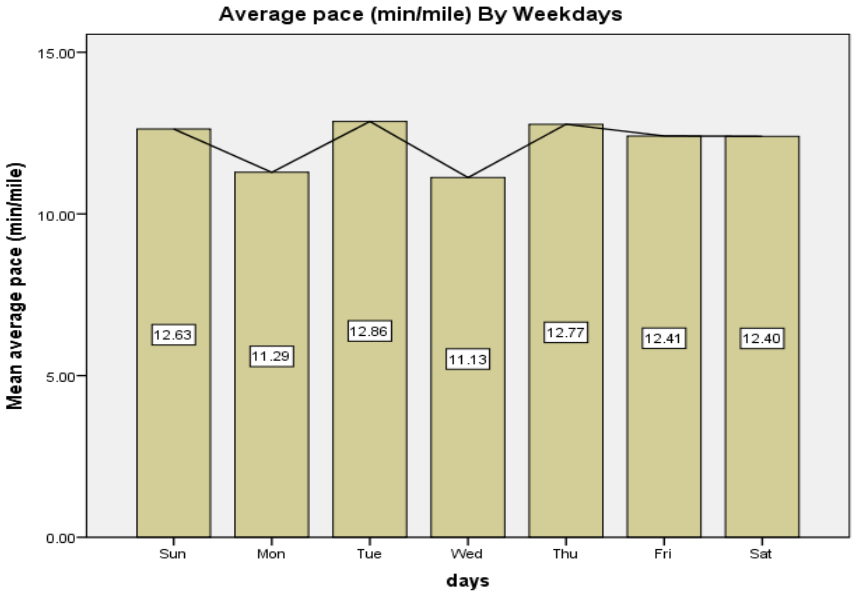
Data Exploration

(1) Distance run per mile by Days



- More distance was covered on Wednesday than the other days
- The least distance covered during the fitness challenge was on Saturday

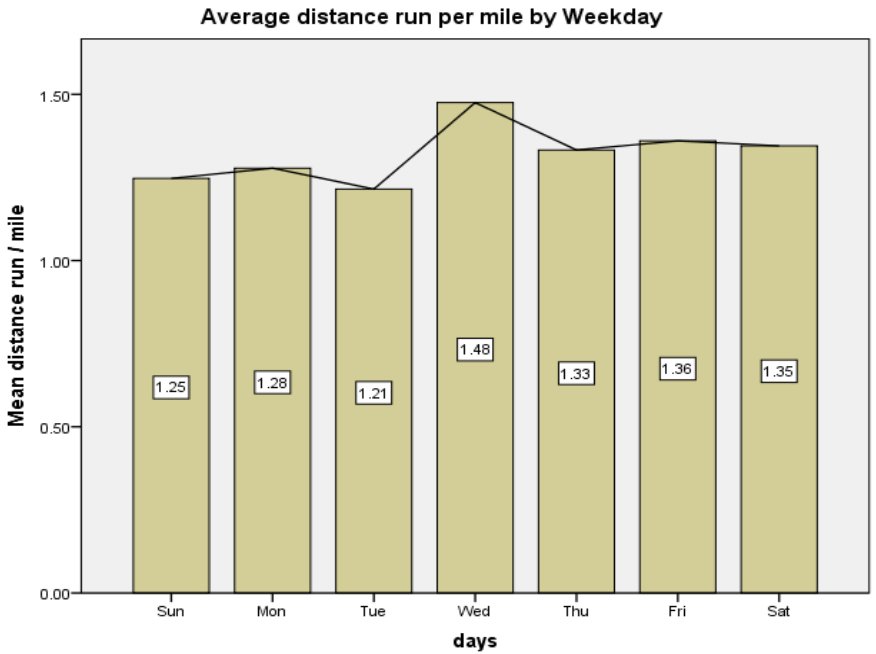
(2) Average pace (min/mile) by Days



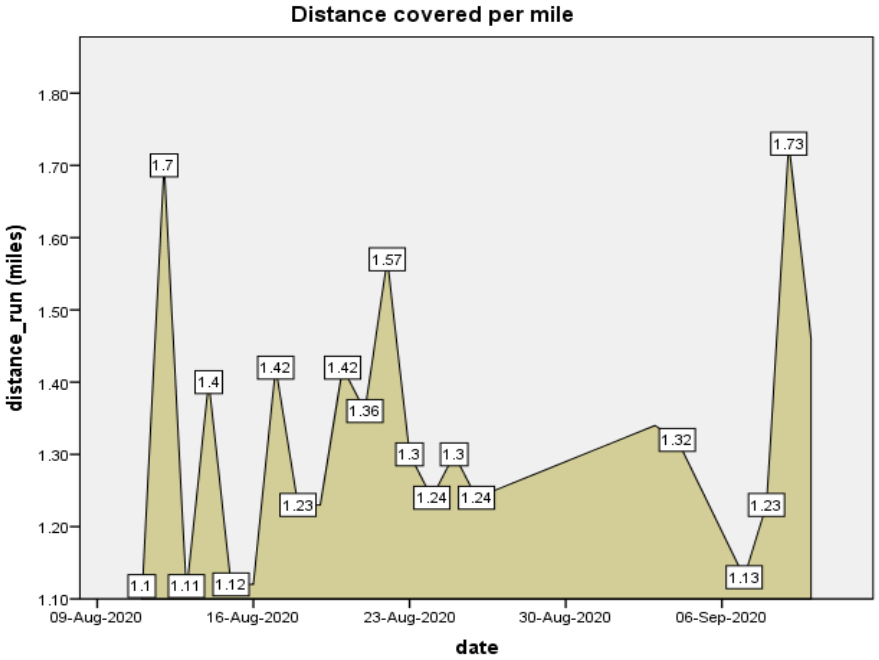
- The fastest pace was ran mostly on Tuesday
- But the pace was somewhat constant during the fitness challenge

Data Exploration

(3) Average distance run per mile per day

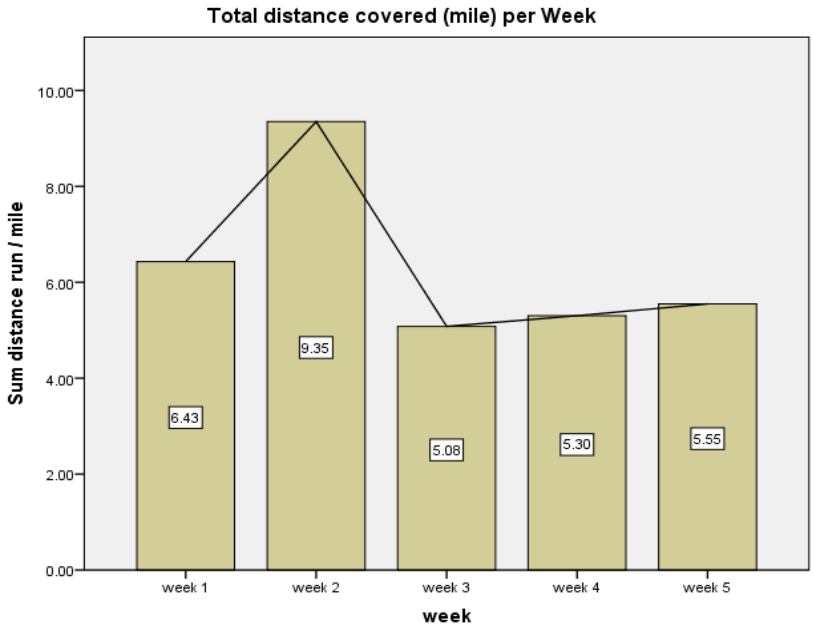


(4) Distance covered per mile



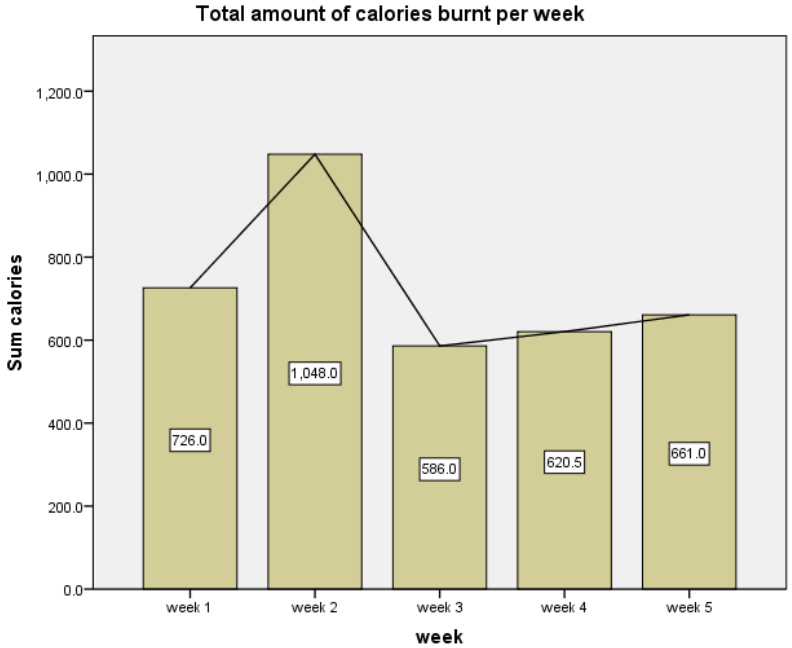
Data Exploration

(5) Distance covered(mile) per Week



- More distance was covered on Week 2 than the other weeks

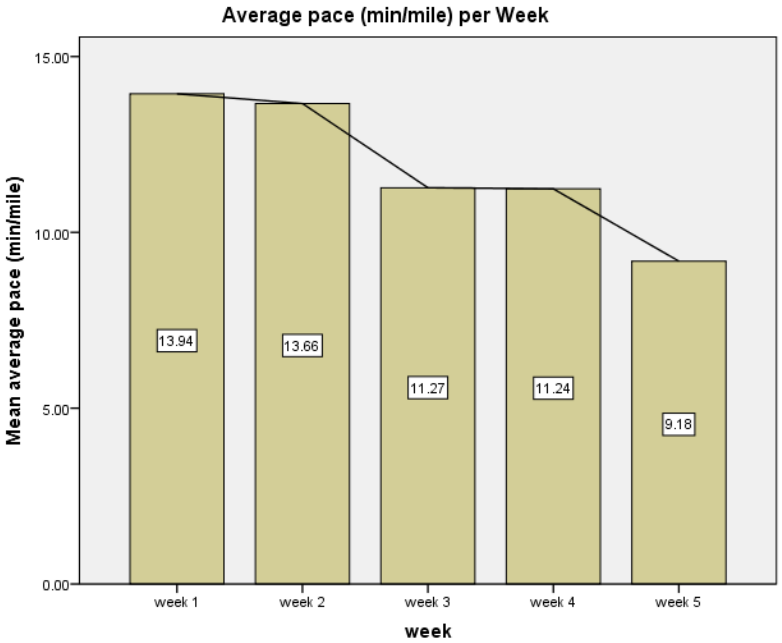
(6) Calories burnt per week



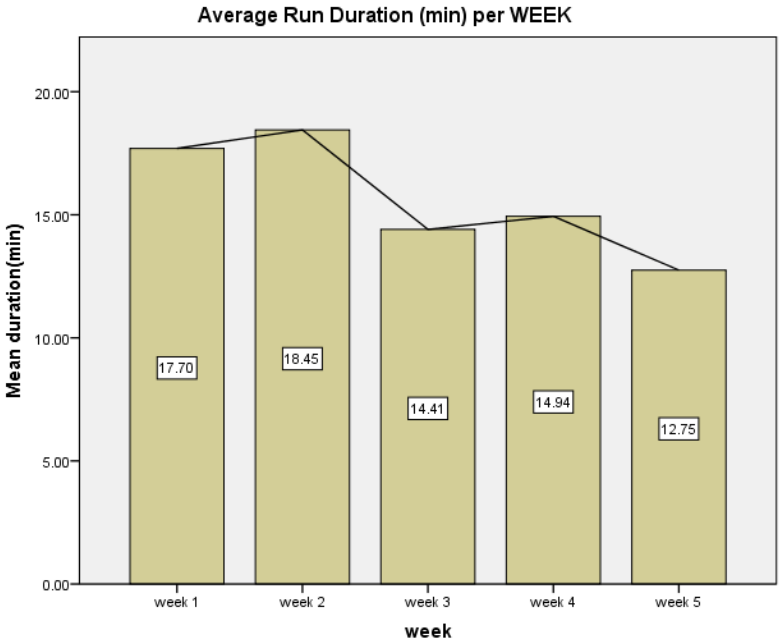
- More calories was burn on Week 2 relative to other weeks
- Seeing that there is a strong positive correlation between distance covered and calories burnt, which brings about the total number of calories burnt is highly dependent on distance covered.

Data Exploration

(7) Average pace (min/mile) per Week



(6) Run Duration(min) per week



Model Development

REGRESSION ANALYSIS

Analysis one

Using multiple linear regression, I want to fit a model to help predict calories burnt based on these variables: distance run/mile, duration(min), push ups, and jumping jacks

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	10293.527	4	2573.382	604.850	.000[e]
	Residual	80.837	19	4.255		
	Total	10374.364	23			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
	(Constant)	20.834	4.676		4.456	.000	11.048	30.621		
	distance run / mile	135.468	2.959	1.088	45.784	.000	129.275	141.661	0.726	1.378
	duration(min)	-1.653	.167	-.226	-9.896	.000	-2.003	-1.303	0.785	1.274
	push_ups	-.769	.129	-.151	-5.944	.000	-1.040	-.499	0.635	1.575
	jumping jacks	.097	.034	.067	2.812	.011	0.025	.169	0.729	1.372

Model Development

MODEL COEFFICIENTS

- calories = (Intercept) + $\beta_1 X1$ + $\beta_2 X2$ + $\beta_3 X3$ + $\beta_4 X4$

- calories = (20.8342) + (-0.7695) * (push_ups) + (135.4682) * (distance run/mile) + (-1.6531) * (duration(min)) + (0.0967) * (jumping_jacks)

Interpreting the Output

- We can see here that this model has a very high R-squared value of 99.2%, meaning that this model explains 99.2% of the variance in our dependent variable. Whenever we add variables to a regression model, R^2 will be higher, but this is a pretty high R^2. Overall the model is a good fit of the data.
- The value of calories when all variables are kept constant is 20.8342. We can see that push ups, distance run/mile, duration(min), and jumping jacks are all statistically significant in predicting (or estimating) the median calories burnt; we see that as push ups increases by 1 unit, calories will decrease by 0.7695 and vice versa keeping the remaining variables constant; also when distance run/mile increases by 1 unit, calories will increase by 135.47 and vice versa keeping the remaining variables constant; also when duration(min) increases by 1 unit, calories will decrease by 1.6531 and vice versa keeping the remaining variables constant; and lastly when jumping jacks increases by 1 unit, calories will increase by 0.0967 and vice versa keeping the remaining variables constant

CONCLUSION

A multiple regression was put to predict calories burnt from push ups, distance run/mile, duration(min), and jumping jacks. These variables statistically significantly predicted calories burnt $F(4,24) = 604.850$, $p(0.001) < 0.05$, Adjusted R-squared= 0.991
All four variables statistically significantly predicted calories burnt

Predicted calories is equal to (-13.9265) + (11.3925) * (distance_run (miles)) + (1.2289) * average pace (min/mile)

Model Development

REGRESSION ANALYSIS

Analysis two

Using multiple linear regression, I want to fit a model to help predict duration (min) based on these variables: distance run (miles) and average pace (min/mile)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	188.966	2	94.483	379.562	.000[b]
	Residual	5.227	21	.249		
	Total	194.194	23			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-13.926	1.203		-11.579	.000	-16.428	-11.425		
	distance run / mile	11.392	.655	.669	17.381	.000	10.029	12.756	.865	1.155
	average pace (min/mile)	1.229	.047	1.011	26.264	.000	1.132	1.326	.865	1.155

Model Development

MODEL COEFFICIENTS

- distance covered (min) = (Intercept) + $\beta_1 X1$ + $\beta_2 X2$

- distance covered (min) = (-13.9265) + (11.3925) * (distance_run (miles)) + (1.2289) * average pace (min/mile)

Interpreting the Output

- We can see here that this model has a very high R-squared value of 97.3%, meaning that this model explains 97.3% of the variance in our dependent variable. Whenever we add variables to a regression model, R^2 will be higher, but this is a pretty high R^2. Overall the model is a good fit of the data.
- The value of distance covered(min) when all variables are kept constant is -13.9265.
- We can see that distance run/mile, and average pace (min/mile) was statistically significant in predicting (or estimating) the median distance covered(min).
- A unit increase in distance run/mile when average pace (min/mile) is held constant will cause a 11.4 increase in distance covered (min), and also a unit increase in average pace (min/mile) when distance run/mile is held constant will cause a 1.2289 increase in distance covered (min)

CONCLUSION

A multiple regression was put to predict distance covered in minutes from distance run per mile and average pace (min/mile). These variables statistically significantly predicted distance covered (min): F(2,24) = 379.562, p(0.001) < 0.05, Adjusted R-squared= 0.971
The two variables statistically significantly predicted distance covered (min).

- Predicted distance covered (min) is equal to (-13.9265) + (11.3925) * (distance_run (miles)) + (1.2289) * average pace (min/mile)