Dataset Overview

```
1  person_id,age,weight_kg,height_cm,steps_per_day,calories_burned
2  301,18,52.99,173.99,3000,2075
3  302,33,70.89,192.65,12840,2796
4  303,19,57.21,186.68,9229,2102
5  304,46,71.4,192.25,14381,1764
6  305,20,92.28,180.18,6638,2788
7  306,56,88.11,177.46,11447,3081
8  307,41,85.46,167.24,6079,2041
9  308,29,99.63,163.65,12525,2624
10 309,44,57.14,170.02,8059,1759
11 310,53,97.05,173.09,3228,2478
12 311,61,89.95,155.98,12794,1712
13 312,31,65.53,152.13,14839,2298
14 313,27,92.78,172.23,7790,3126
15 314,27,71.18,188.56,10545,2585
16 315,28,79.19,156.67,7305,2193
17 316,53,67.99,162.63,13572,2578
18 317,47,70.13,198.59,7516,1881
19 318,38,75.71,177.9,4181,2949
20 319,62,69.69,174.33,6769,1756
21 320,50,73.0,199.54,10929,1670
22 321,37,84.02,178.56,11474,2950
23 322,51,53.79,167.83,12597,2969
24 323,54,60.1,157.16,3753,2389
25 324,62,96.33,168.5,11207,2061
26 325,37,55.65,190.51,4760,3123
27 326,61,61.86,156.46,9628,2779
28 327,18,87.57,157.72,4759,1757
29 328,21,76.78,163.12,13612,2635
30 329,28,79.67,167.04,5271,1832
31 330,22,75.39,192.32,8492,2643
```

	person_id	age	weight_kg	height_cm	steps_per_day
count	30.000000	30.000000	30.000000	30.000000	30.000000
mean	315.500000	39.133333	75.282333	173.833000	8973.966667
std	8.803408	14.778207	13.803452	13.772684	3619.636203
min	301.000000	18.000000	52.990000	152.130000	3000.000000
25%	308.250000	27.250000	66.145000	163.252500	6218.750000
50%	315.500000	37.500000	74.195000	172.660000	8860.500000
75%	322.750000	52.500000	87.042500	185.055000	12262.250000
max	330.000000	62.000000	99.630000	199.540000	14839.000000

	calories_burned
count	30.000000
mean	2379.800000
std	486.005066
min	1670.000000
25%	1921.000000
50%	2433.500000
75%	2785.750000
max	3126.000000

Univariate Visualizations Boxplots

Age_std :
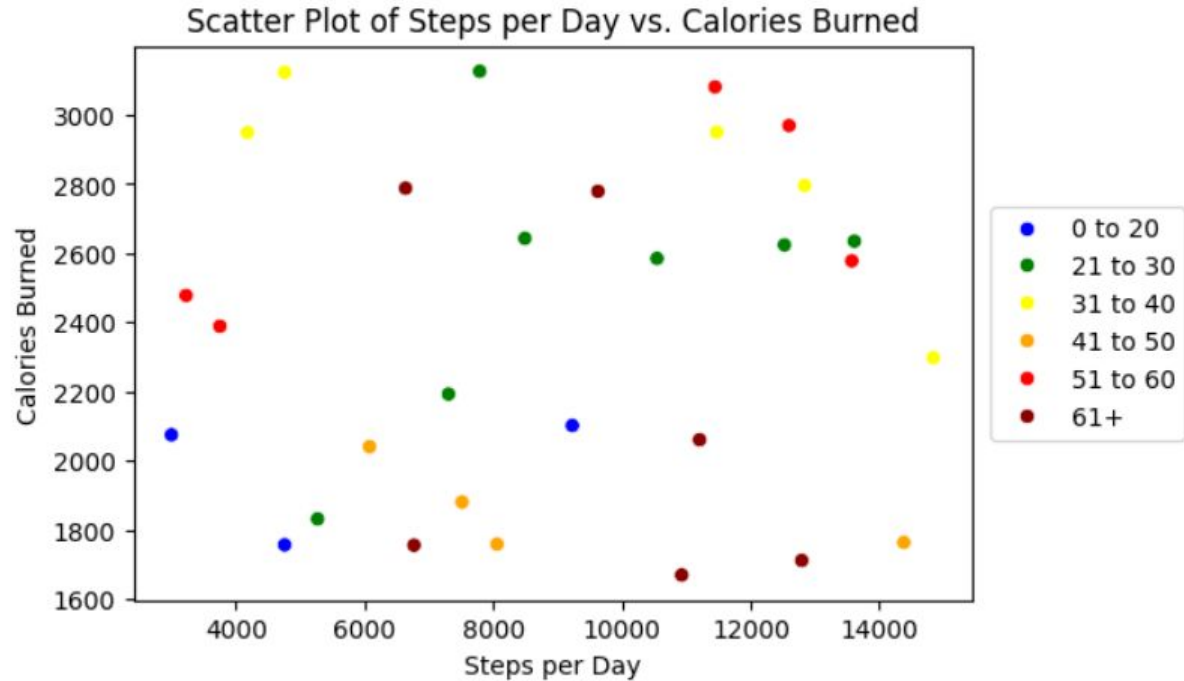14.778207

Weight_std :
13.803452
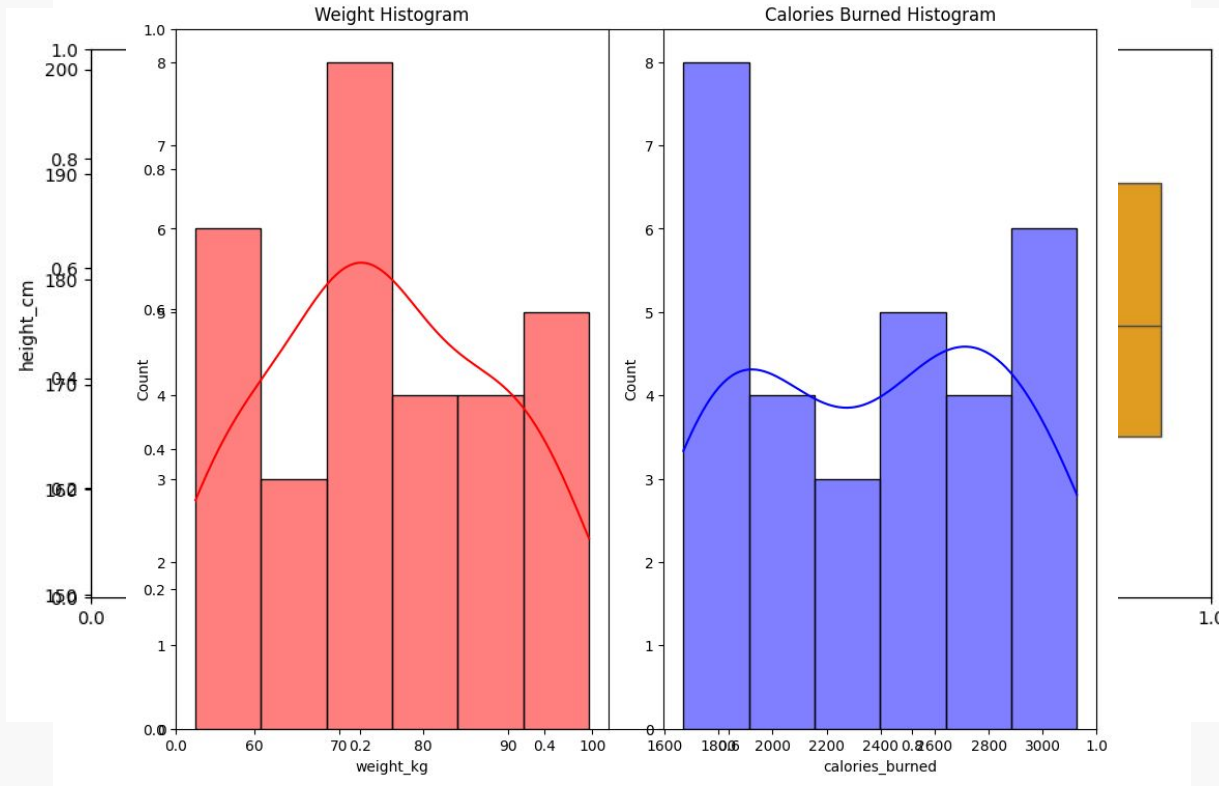
Height_std :
13.772684

StepsPerDay_
std :
3619.636203

CalBurned_std
: 3619.636203

Multivariate Variables

- No strong association between calories burnt and steps walked per day
- Different age distribution shows no pattern either

Scatter Plot of Steps per Day vs. Calories Burned

Legend:
- 0 to 20
- 21 to 30
- 31 to 40
- 41 to 50
- 51 to 60
- 61+

Multivariate Variables

- 21 to 30 y/o's weighed the most
- 21 to 30 and 61+ y/o's took the most steps per day
- 41 to 50 y/o's burnt the most calories
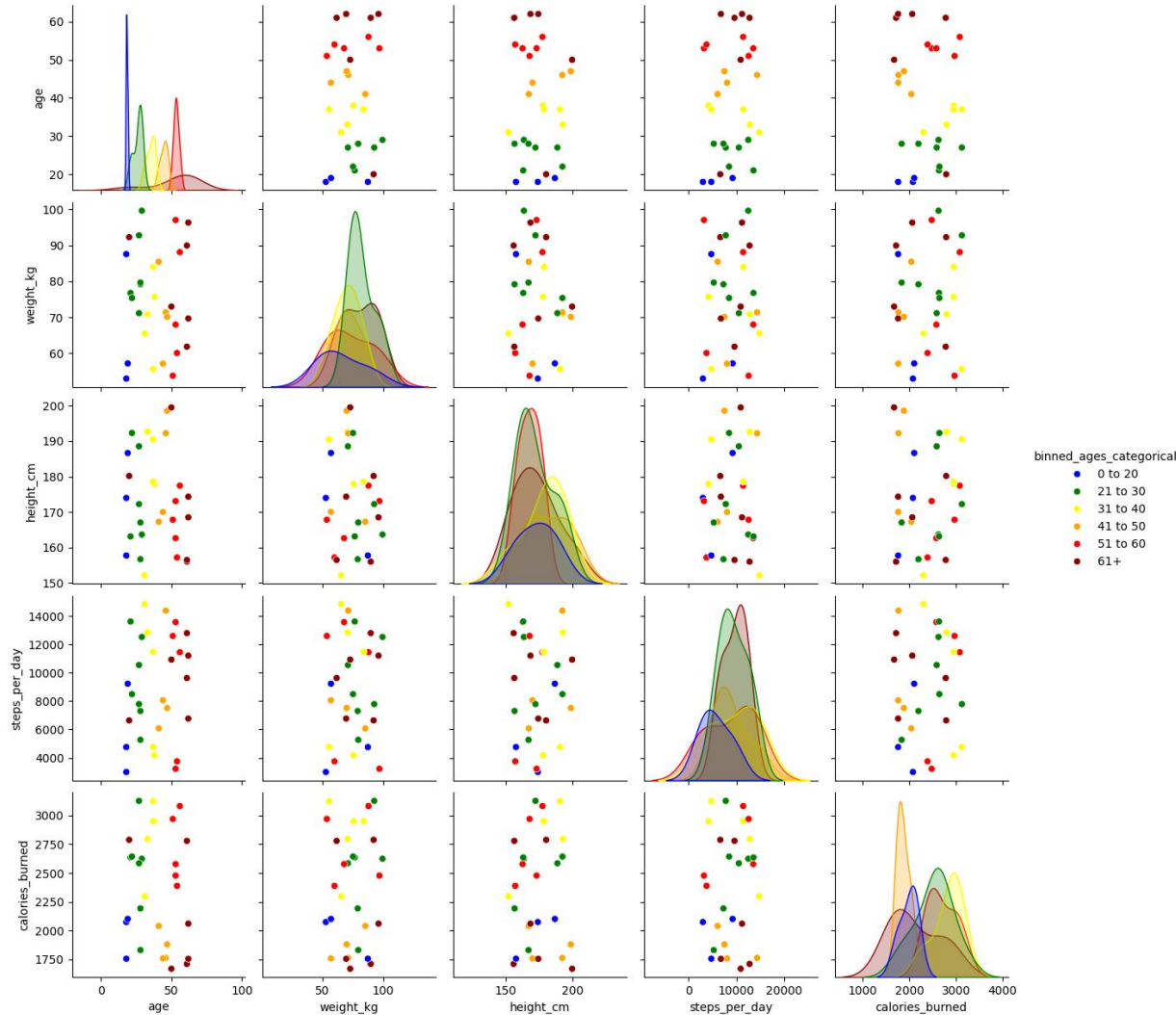
Correlation Matrix Heatmap

- Exclude 'person_id' since this is an identifier, not a meaningful value for correlations.

- Mix of positive and negative correlations, none very strong in absolute terms.

- But what about relative to one another? Which values are the most/least correlated?

Correlation Matrix Heatmap (excluding person_id)
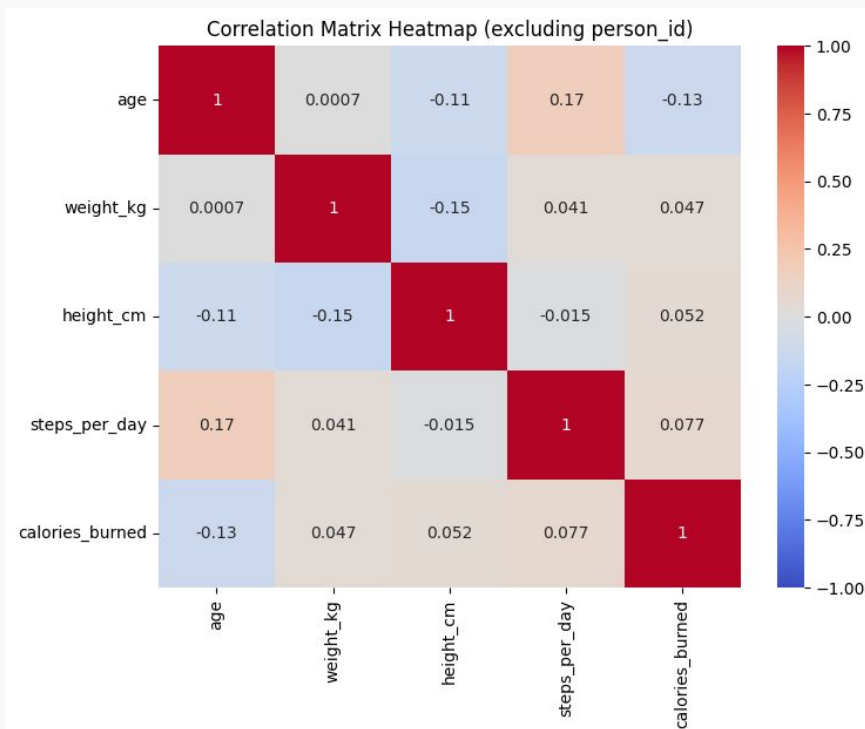
Correlations by Rank

- Rank by absolute value

- Hard to understand at a glance

- Visualize by plotting values using barh (horizontal bar plot

Rank	Variable 1	Variable 2	Correlation
1	age	steps_per_day	0.169239
2	weight_kg	height_cm	-0.151204
3	age	calories_burned	-0.128699
4	age	height_cm	-0.114970
5	steps_per_day	calories_burned	0.076832
6	height_cm	calories_burned	0.051617
7	weight_kg	calories_burned	0.046506
8	weight_kg	steps_per_day	0.040808
9	height_cm	steps_per_day	-0.015362
10	age	weight_kg	0.000703

Correlations by Rank, continued

Correlation Strength by Rank
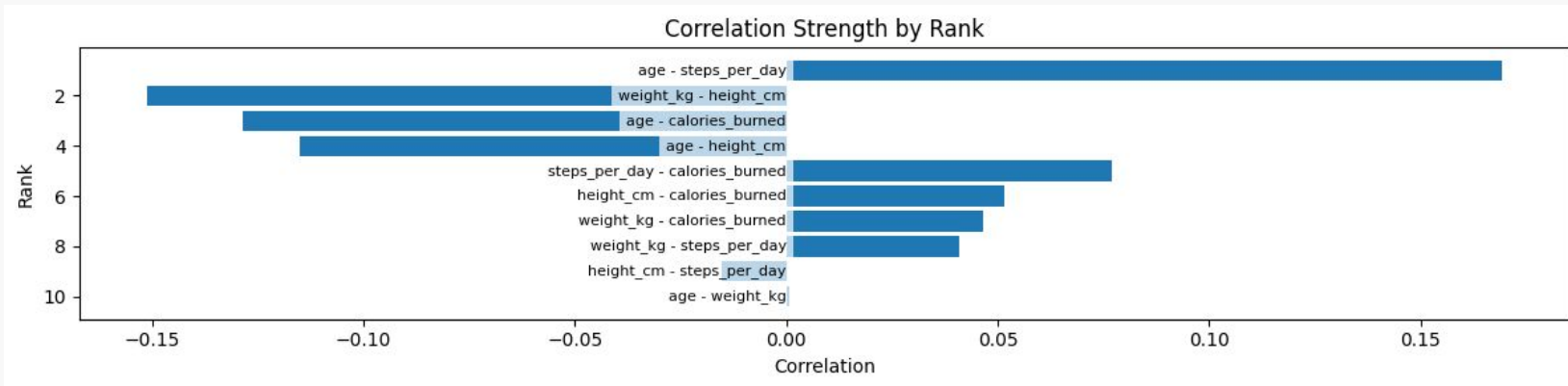
- Strongest: age-steps per day
- Weakest: age-weight
- Positive correlation of steps per day-calories burned makes sense
- Height- and Weight-calories burned also positively correlated—larger people burn more calories
- Surprising negative correlation in weight-height