CORRELATION MATRIX

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [4]:
```
# Load the dataset
df1=pd.read_csv(r"C:\Users\bhava\Downloads\cardio_train.csv")
df1
```

Out[4]:

	id;age;gender;height;weight;ap_hi;ap_lo;cholesterol;gluc;smoke;alco;active;cardio
0	0;18393;2;168;62.0;110;80;1;1;0;0;1;0
1	1;20228;1;156;85.0;140;90;3;1;0;0;1;1
2	2;18857;1;165;64.0;130;70;3;1;0;0;0;1
3	3;17623;2;169;82.0;150;100;1;1;0;0;1;1
4	4;17474;1;156;56.0;100;60;1;1;0;0;0;0
...	...
69995	99993;19240;2;168;76.0;120;80;1;1;1;0;1;0
69996	99995;22601;1;158;126.0;140;90;2;2;0;0;1;1
69997	99996;19066;2;183;105.0;180;90;3;1;0;1;0;1
69998	99998;22431;1;163;72.0;135;80;1;2;0;0;0;1
69999	99999;20540;1;170;72.0;120;80;2;1;0;0;1;0

70000 rows × 1 columns

In [8]:
```
df1=pd.read_csv(r"C:\Users\bhava\Downloads\cardio_train.csv",delimiter=';')
df1
```

Out[8]:

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
69995	99993	19240	2	168	76.0	120	80	1	1	1	0	1	0
69996	99995	22601	1	158	126.0	140	90	2	2	0	0	1	1
69997	99996	19066	2	183	105.0	180	90	3	1	0	1	0	1
69998	99998	22431	1	163	72.0	135	80	1	2	0	0	0	1
69999	99999	20540	1	170	72.0	120	80	2	1	0	0	1	0

70000 rows × 13 columns

In [16]:
```
# Select relevant features for the correlation matrix
features = ['age','gender', 'height', 'weight', 'ap_hi', 'ap_lo', 'cholesterol', 'gluc', 'smoke', 'alco', 'acti
features
```

Out[16]:
```
['age',
 'gender',
 'height',
 'weight',
 'ap_hi',
 'ap_lo',
 'cholesterol',
 'gluc',
 'smoke',
 'alco',
 'active',
 'cardio']
```

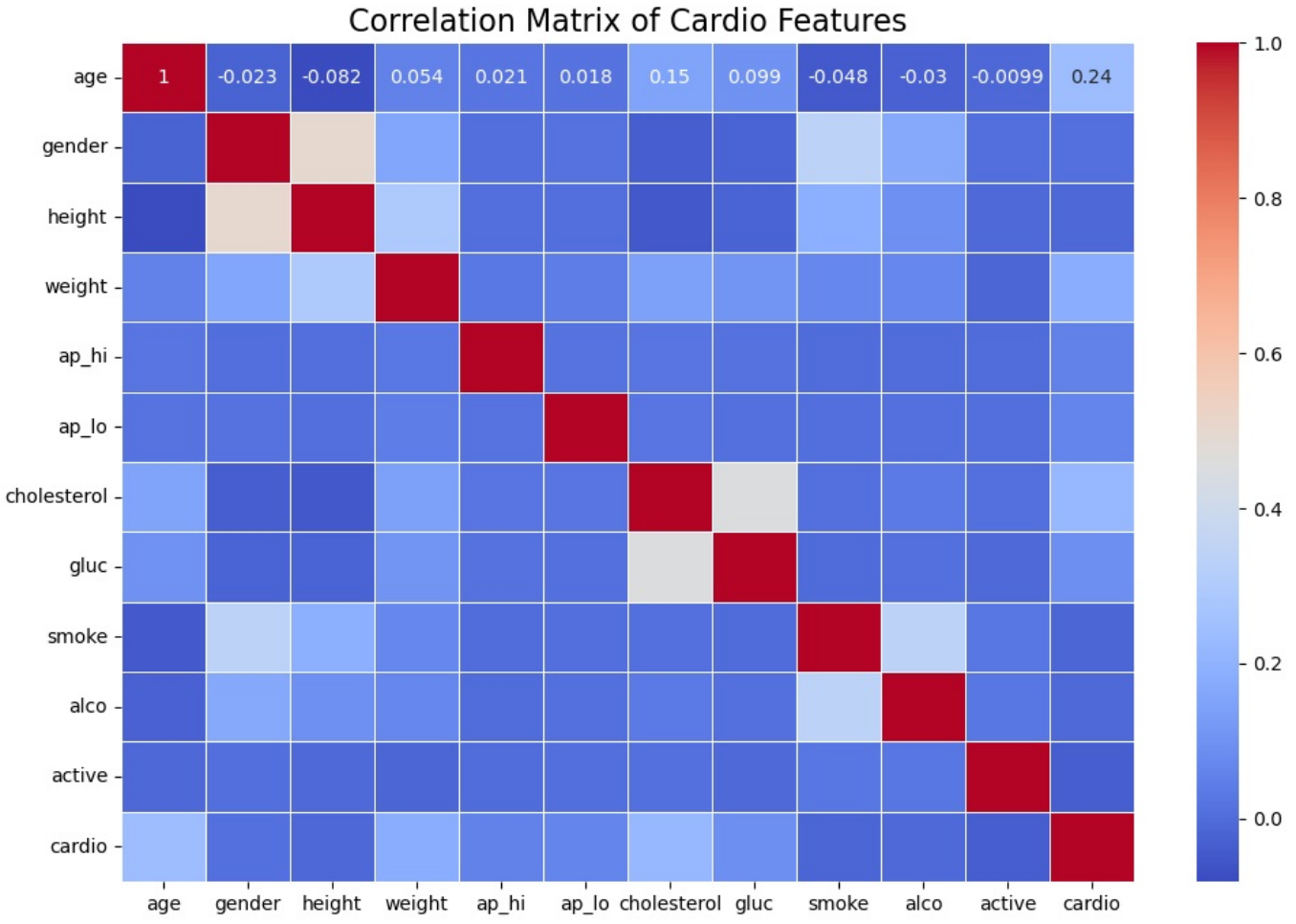
In [18]:
```
# Calculate the correlation matrix
correlation_matrix = df1[features].corr()
correlation_matrix
```

Out[18]:

	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active
age	1.000000	-0.022811	-0.081515	0.053684	0.020764	0.017647	0.154424	0.098703	-0.047633	-0.029723	-0.009927
gender	-0.022811	1.000000	0.499033	0.155406	0.006005	0.015254	-0.035821	-0.020491	0.338135	0.170966	0.005866
height	-0.081515	0.499033	1.000000	0.290968	0.005488	0.006150	-0.050226	-0.018595	0.187989	0.094419	-0.006570
weight	0.053684	0.155406	0.290968	1.000000	0.030702	0.043710	0.141768	0.106857	0.067780	0.067113	-0.016867
ap_hi	0.020764	0.006005	0.005488	0.030702	1.000000	0.016086	0.023778	0.011841	-0.000922	0.001408	-0.000033
ap_lo	0.017647	0.015254	0.006150	0.043710	0.016086	1.000000	0.024019	0.010806	0.005186	0.010601	0.004780
cholesterol	0.154424	-0.035821	-0.050226	0.141768	0.023778	0.024019	1.000000	0.451578	0.010354	0.035760	0.009911
gluc	0.098703	-0.020491	-0.018595	0.106857	0.011841	0.010806	0.451578	1.000000	-0.004756	0.011246	-0.006770
smoke	-0.047633	0.338135	0.187989	0.067780	-0.000922	0.005186	0.010354	-0.004756	1.000000	0.340094	0.025858
alco	-0.029723	0.170966	0.094419	0.067113	0.001408	0.010601	0.035760	0.011246	0.340094	1.000000	0.025476
active	-0.009927	0.005866	-0.006570	-0.016867	-0.000033	0.004780	0.009911	-0.006770	0.025858	0.025476	1.000000
cardio	0.238159	0.008109	-0.010821	0.181660	0.054475	0.065719	0.221147	0.089307	-0.015486	-0.007330	-0.035653

In [20]:

```
# Plotting the correlation heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=0.5)
plt.title('Correlation Matrix of Cardio Features', fontsize=16)
plt.show()
```



In [22]:

```
# Display the correlation matrix
print(correlation_matrix)
```

	age	gender	height	weight	ap_hi	ap_lo	\
age	1.000000	-0.022811	-0.081515	0.053684	0.020764	0.017647	
gender	-0.022811	1.000000	0.499033	0.155406	0.006005	0.015254	
height	-0.081515	0.499033	1.000000	0.290968	0.005488	0.006150	
weight	0.053684	0.155406	0.290968	1.000000	0.030702	0.043710	
ap_hi	0.020764	0.006005	0.005488	0.030702	1.000000	0.016086	
ap_lo	0.017647	0.015254	0.006150	0.043710	0.016086	1.000000	
cholesterol	0.154424	-0.035821	-0.050226	0.141768	0.023778	0.024019	
gluc	0.098703	-0.020491	-0.018595	0.106857	0.011841	0.010806	
smoke	-0.047633	0.338135	0.187989	0.067780	-0.000922	0.005186	
alco	-0.029723	0.170966	0.094419	0.067113	0.001408	0.010601	
active	-0.009927	0.005866	-0.006570	-0.016867	-0.000033	0.004780	
cardio	0.238159	0.008109	-0.010821	0.181660	0.054475	0.065719	

	cholesterol	gluc	smoke	alco	active	cardio
age	0.154424	0.098703	-0.047633	-0.029723	-0.009927	0.238159
gender	-0.035821	-0.020491	0.338135	0.170966	0.005866	0.008109
height	-0.050226	-0.018595	0.187989	0.094419	-0.006570	-0.010821
weight	0.141768	0.106857	0.067780	0.067113	-0.016867	0.181660
ap_hi	0.023778	0.011841	-0.000922	0.001408	-0.000033	0.054475
ap_lo	0.024019	0.010806	0.005186	0.010601	0.004780	0.065719
cholesterol	1.000000	0.451578	0.010354	0.035760	0.009911	0.221147
gluc	0.451578	1.000000	-0.004756	0.011246	-0.006770	0.089307
smoke	0.010354	-0.004756	1.000000	0.340094	0.025858	-0.015486
alco	0.035760	0.011246	0.340094	1.000000	0.025476	-0.007330
active	0.009911	-0.006770	0.025858	0.025476	1.000000	-0.035653
cardio	0.221147	0.089307	-0.015486	-0.007330	-0.035653	1.000000