In [27]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

In [28]:
```
df = pd.read_csv('medical_examination.csv')
df
```

Out[28]:

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

70000 rows × 13 columns

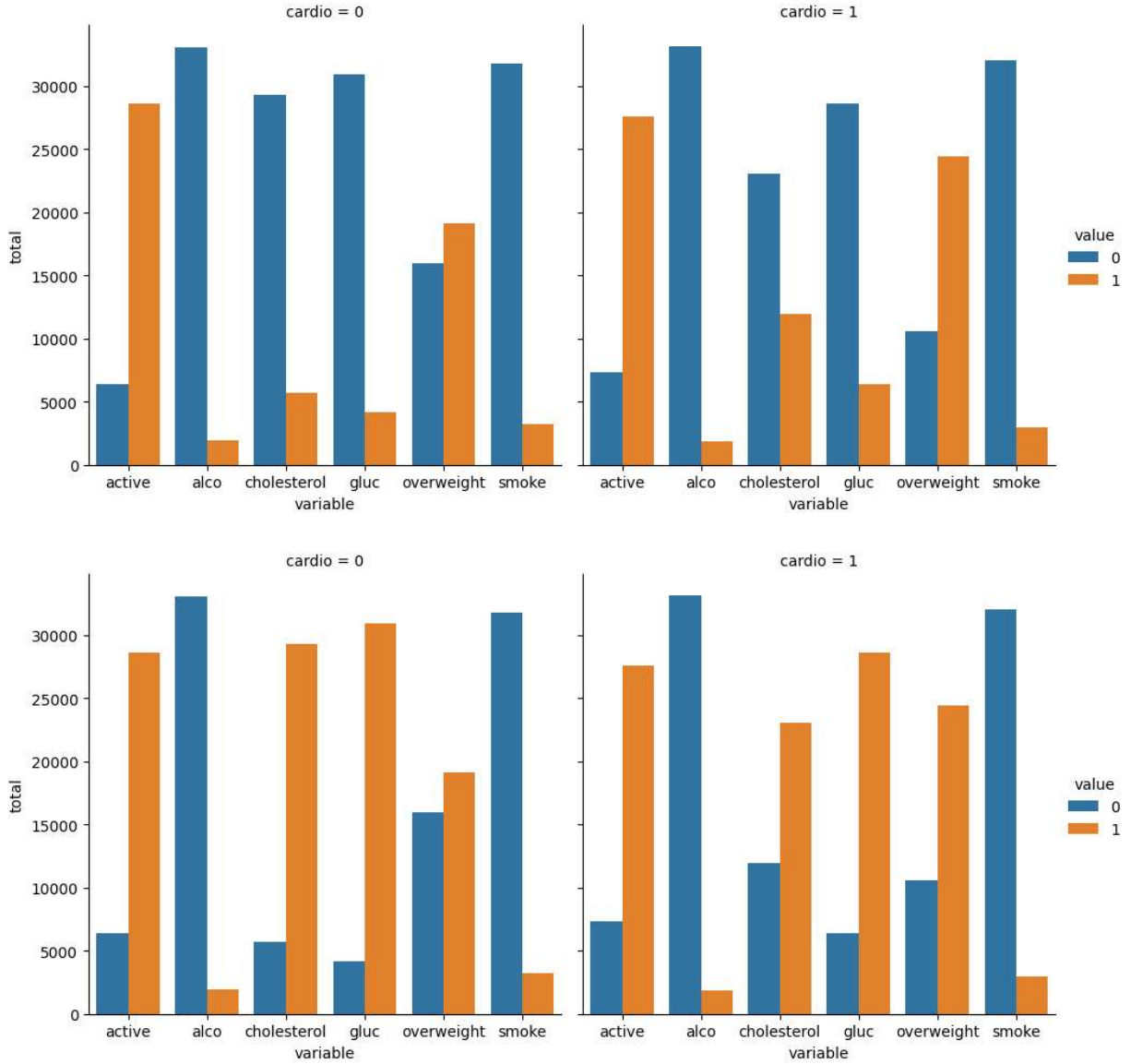
In [29]:
```
df.info()
```

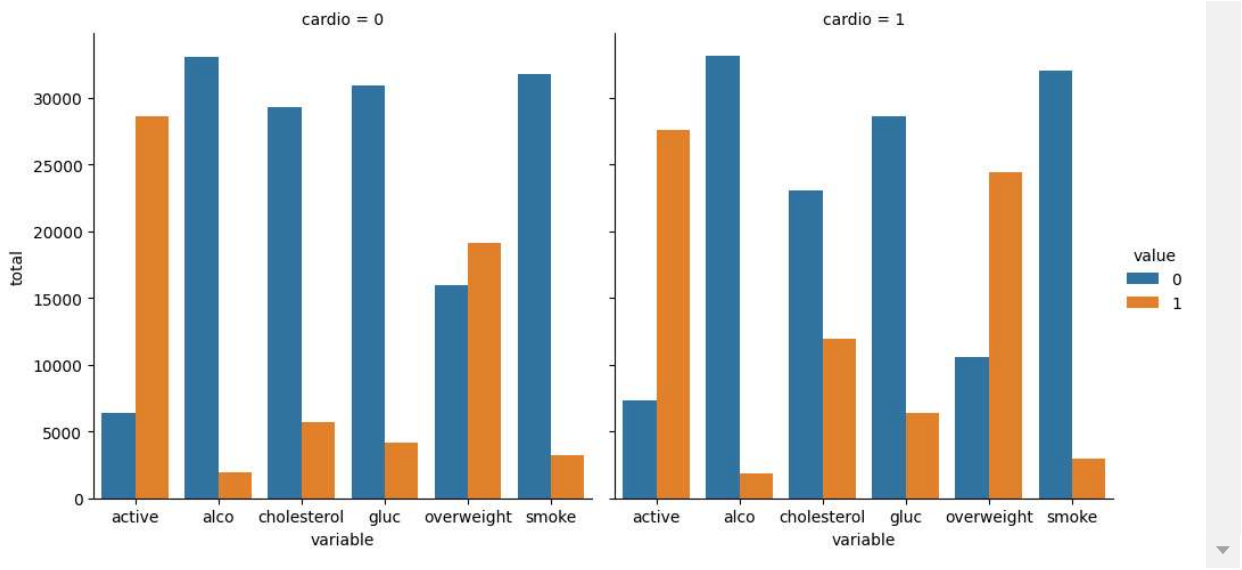
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 70000 entries, 0 to 69999
Data columns (total 13 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   id           70000 non-null  int64
 1   age          70000 non-null  int64
 2   sex          70000 non-null  int64
 3   height       70000 non-null  int64
 4   weight       70000 non-null  float64
 5   ap_hi        70000 non-null  int64
 6   ap_lo        70000 non-null  int64
 7   cholesterol  70000 non-null  int64
 8   gluc         70000 non-null  int64
 9   smoke        70000 non-null  int64
 10  alco         70000 non-null  int64
 11  active       70000 non-null  int64
 12  cardio       70000 non-null  int64
dtypes: float64(1), int64(12)
memory usage: 6.9 MB
```

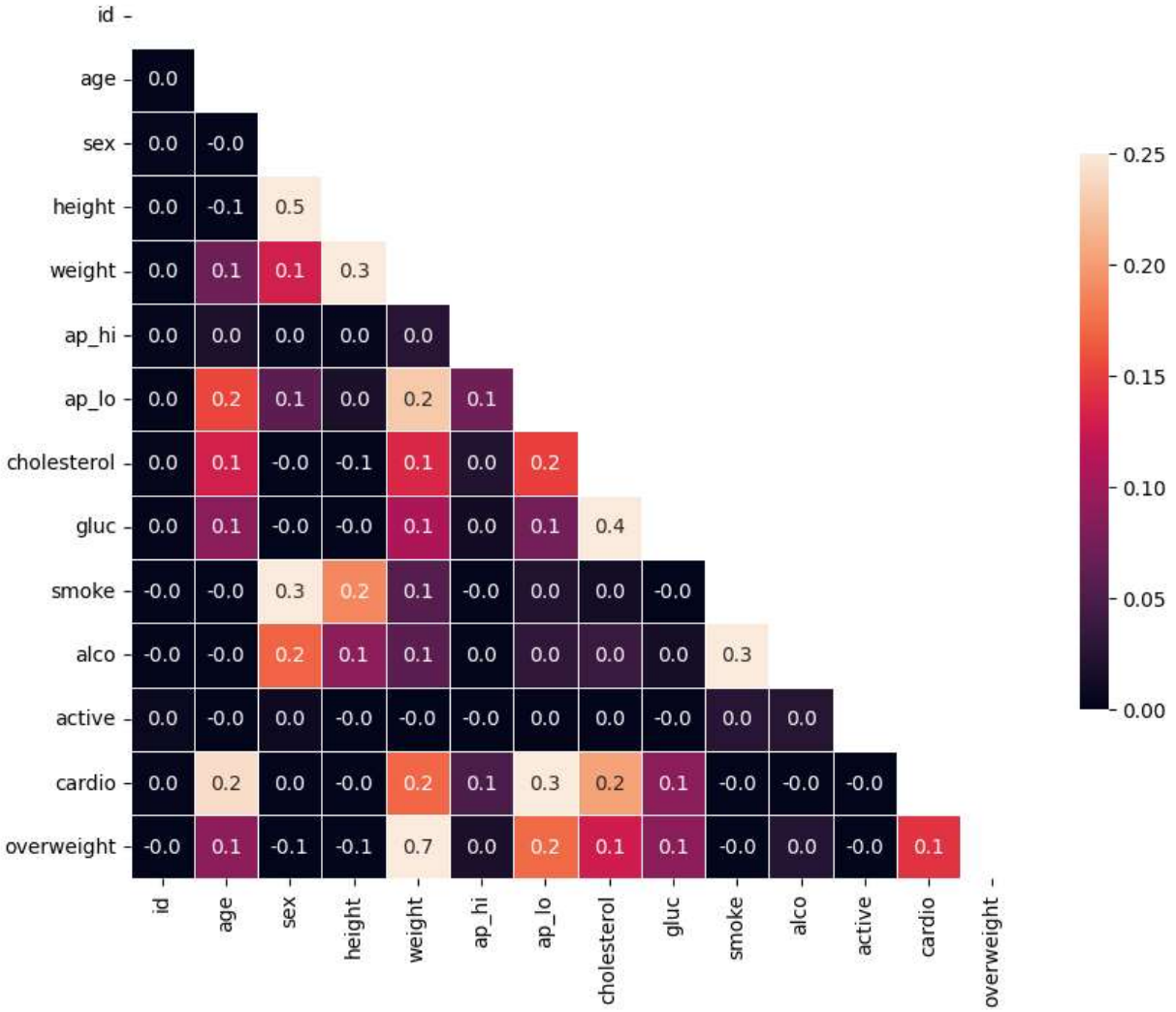
In [30]:
```
df['overweight'] = np.where(df['weight'] / ((df['height'] * 0.01) ** 2) > 25, 1,
df['cholesterol'] = np.where(df['cholesterol'] == 1, 0, 1)
df['gluc'] = np.where(df['gluc'] == 1, 0, 1)
```

In [31]:
```
df_cat = sorted(['cholesterol', 'gluc', 'smoke', 'alco', 'active', 'overweight'])
df_cat = pd.melt(df, id_vars = 'cardio', value_vars = df_cat)
graph = sns.catplot(x='variable', col='cardio', hue='value', kind='count', data=d
fig = graph.fig
```

In [32]:
```
#Drawing a heat map
df_heat = df.loc[(df['ap_lo'] <= df['ap_hi']) & (df['height'] >= df['height'].qua
corr = df_heat.corr()
mask = np.zeros_like(corr)
mask[np.triu_indices_from(mask)] = True
fig, ax = plt.subplots(figsize=(10, 10))
ax = sns.heatmap(corr, vmin=0, vmax=.25, square=True, cbar_kws={"shrink": .50}, a
plt.show()
```




cardio = 0
cardio = 1
total
variable
value
0
1
active
alco
cholesterol
gluc
overweight
smoke


id
age
sex
height
weight
ap_hi
ap_lo
cholesterol
gluc
smoke
alco
active
cardio
overweight

In [12]:

In [ ]: