Ben Deatsman

Classwork 3b

In [125…
```
import pandas as pd
df = pd.read_csv('heart.csv')
df
```

Out[125…

	age	sex	chest_pain	rest_bp	chol	fbs	rest_ecg	max_hr	exang	old_peak	slope	ca	thal	disease
0	63	Male	typical	145	233	1	left ventricular hypertrophy	150	0	2.3	3	0.0	fixed	0
1	67	Male	asymptomatic	160	286	0	left ventricular hypertrophy	108	1	1.5	2	3.0	normal	1
2	67	Male	asymptomatic	120	229	0	left ventricular hypertrophy	129	1	2.6	2	2.0	reversable	1
3	37	Male	nonanginal	130	250	0	normal	187	0	3.5	3	0.0	normal	0
4	41	Female	nontypical	130	204	0	left ventricular hypertrophy	172	0	1.4	1	0.0	normal	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	45	Male	typical	110	264	0	normal	132	0	1.2	2	0.0	reversable	1
299	68	Male	asymptomatic	144	193	1	normal	141	0	3.4	2	2.0	reversable	1
300	57	Male	asymptomatic	130	131	0	normal	115	1	1.2	2	1.0	reversable	1
301	57	Female	nontypical	130	236	0	left ventricular hypertrophy	174	0	0.0	2	1.0	normal	1
302	38	Male	nonanginal	138	175	0	normal	173	0	0.0	1	NaN	normal	0

303 rows × 14 columns

In [126…
```
#There are 6 missing values. 4 in 'ca' and 2 in 'thal'
df.isnull().sum()
```

Out[126…
```
age          0
sex          0
chest_pain   0
rest_bp      0
chol         0
fbs          0
rest_ecg     0
max_hr       0
exang        0
old_peak     0
slope        0
ca           4
thal         2
disease      0
dtype: int64
```

In [127…
```
df['ca'].fillna(df['ca'].mode()[0], inplace=True)
df['thal'].fillna(df['thal'].mode()[0], inplace=True)
```

In [128…
```
df.isnull().sum()
```

Out[128…
```
age          0
sex          0
chest_pain   0
rest_bp      0
chol         0
fbs          0
rest_ecg     0
max_hr       0
exang        0
old_peak     0
slope        0
ca           0
thal         0
disease      0
dtype: int64
```

In [129…
```
risk = df['age']/(df['rest_bp'] + df['chol'] + df['max_hr'])
```

In [130…
```
df['risk'] = risk
```

In [131…
```
df.head()
```

Out[131…

	age	sex	chest_pain	rest_bp	chol	fbs	rest_ecg	max_hr	exang	old_peak	slope	ca	thal	disease	risk
0	63	Male	typical	145	233	1	left ventricular hypertrophy	150	0	2.3	3	0.0	fixed	0	0.119318
1	67	Male	asymptomatic	160	286	0	left ventricular hypertrophy	108	1	1.5	2	3.0	normal	1	0.120939
2	67	Male	asymptomatic	120	229	0	left ventricular hypertrophy	129	1	2.6	2	2.0	reversable	1	0.140167
3	37	Male	nonanginal	130	250	0	normal	187	0	3.5	3	0.0	normal	0	0.065256
4	41	Female	nontypical	130	204	0	left ventricular hypertrophy	172	0	1.4	1	0.0	normal	0	0.081028

In [132…
```
df['rest_ecg'].unique()
```

Out[132…
```
array(['left ventricular hypertrophy ', 'normal', 'ST-T wave abnormality'],
      dtype=object)
```

In [160…
```
vmap = {'left ventricular hypertrophy ': 'lvh', 'ST-T wave abnormality': 'stt_wav_abn'}
df['rest_ecg'] = df['rest_ecg'].replace(vmap)
```

In [162…
```
df.head()
```

Out[162…

	age	sex	chest_pain	rest_bp	chol	fbs	rest_ecg	max_hr	exang	old_peak	slope	ca	thal	disease	risk
0	63	Male	typical	145	233	1	lvh	150	0	2.3	3	0.0	fixed	0	0.119318
1	67	Male	asymptomatic	160	286	0	lvh	108	1	1.5	2	3.0	normal	1	0.120939
2	67	Male	asymptomatic	120	229	0	lvh	129	1	2.6	2	2.0	reversable	1	0.140167
3	37	Male	nonanginal	130	250	0	normal	187	0	3.5	3	0.0	normal	0	0.065256
4	41	Female	nontypical	130	204	0	lvh	172	0	1.4	1	0.0	normal	0	0.081028

In [169…
```
df.describe(include = ['int', 'float', 'object'])
```

Out[169…

	age	sex	chest_pain	rest_bp	chol	fbs	rest_ecg	max_hr	exang	old_peak	slope	ca	thal	disea
count	303.000000	303	303	303.000000	303.000000	303.000000	5	303.000000	303.000000	303.000000	303.000000	303.000000	303	303.0000
unique	NaN	2	4	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	3	N
top	NaN	Male	asymptomatic	NaN	NaN	NaN	lvh	NaN	NaN	NaN	NaN	NaN	normal	N
freq	NaN	206	144	NaN	NaN	NaN	4	NaN	NaN	NaN	NaN	NaN	168	N
mean	54.438944	NaN	NaN	131.689769	246.693069	0.148515	NaN	149.607261	0.326733	1.039604	1.600660	0.663366	NaN	0.4587
std	9.038662	NaN	NaN	17.599748	51.776918	0.356198	NaN	22.875003	0.469794	1.161075	0.616226	0.934375	NaN	0.4991
min	29.000000	NaN	NaN	94.000000	126.000000	0.000000	NaN	71.000000	0.000000	0.000000	1.000000	0.000000	NaN	0.0000
25%	48.000000	NaN	NaN	120.000000	211.000000	0.000000	NaN	133.500000	0.000000	0.000000	1.000000	0.000000	NaN	0.0000
50%	56.000000	NaN	NaN	130.000000	241.000000	0.000000	NaN	153.000000	0.000000	0.800000	2.000000	0.000000	NaN	0.0000
75%	61.000000	NaN	NaN	140.000000	275.000000	0.000000	NaN	166.000000	1.000000	1.600000	2.000000	1.000000	NaN	1.0000
max	77.000000	NaN	NaN	200.000000	564.000000	1.000000	NaN	202.000000	1.000000	6.200000	3.000000	3.000000	NaN	1.0000

In [ ]:
```
```