```
import pandas as pd
import json
import csv
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns
```

```
    /usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWa
      import pandas.util.testing as tm
```

```
#json to csv
data = json.load(open("data.json"))

names = data["Health"]
data_file = open("data.csv", "w")
csv_writer = csv.writer(data_file)
print
csv_writer.writerow(names[0].keys())
for name in names:
    csv_writer.writerow(name.values())
data_file.close()

df = pd.read_csv("data.csv")
df
```

3	48	214	108	138	F
4	54	195	122	150	M
5	39	339	170	120	M
6	45	237	170	130	F
7	54	208	142	110	M
8	37	207	130	140	M
9	48	284	120	120	F
10	37	211	142	130	F
11	58	164	99	136	M
12	39	204	145	120	M
13	49	234	140	140	M
14	42	211	137	115	F
15	54	273	150	120	F
16	38	196	166	110	M
17	43	201	165	120	F
18	60	248	125	100	M
19	36	267	160	120	M
20	43	223	142	100	F
21	44	184	142	120	M
22	49	201	164	124	F
23	44	288	150	150	M
24	40	215	138	130	M

```
#cleaning
df.dropna()
#outliers
df.describe()
df = df.loc[df['Cholesterol'] < 270+((270-202.5)*1.5)]
df.describe()
```

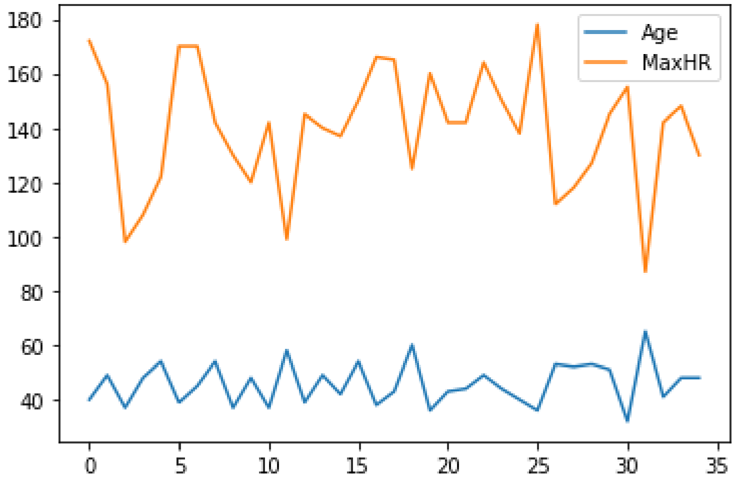
	Age	Cholesterol	MaxHR	RestingBP
count	34.000000	34.000000	34.000000	34.000000
mean	45.676471	232.676471	140.235294	126.970588
std	7.752695	42.419779	22.741044	14.072292
min	32.000000	164.000000	87.000000	100.000000

```
#correlation
df.corr()
```

	Age	Cholesterol	MaxHR	RestingBP
Age	1.000000	0.095111	-0.547935	0.054135
Cholesterol	0.095111	1.000000	-0.131795	-0.146915
MaxHR	-0.547935	-0.131795	1.000000	-0.121600
RestingBP	0.054135	-0.146915	-0.121600	1.000000

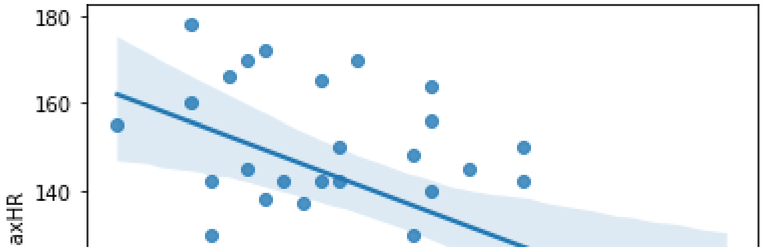
```
df1 = df[['Age', 'MaxHR']]
df1.plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fc48a6f4450>
```

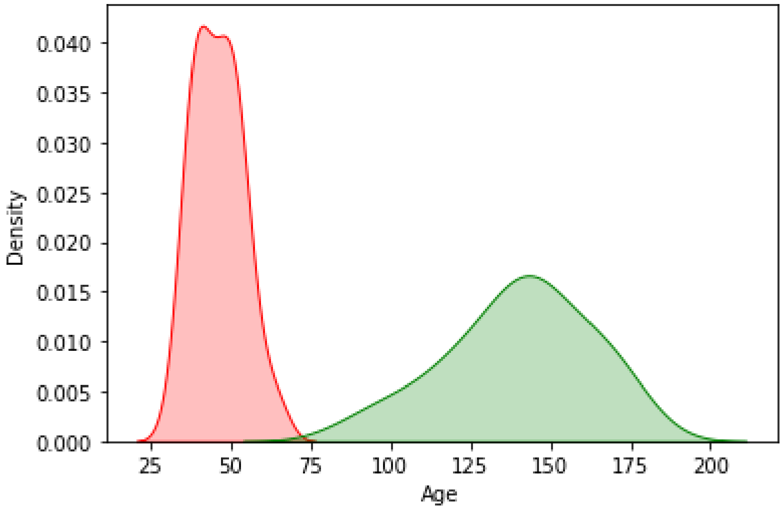


```
#scatterplot
sns.regplot(x=df["Age"], y=df["MaxHR"])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f300acaab90>
```



```
#density
fig = sns.kdeplot(df['Age'], shade=True, color="r")
fig = sns.kdeplot(df['MaxHR'], shade=True, color="g")
plt.show()
```



```
#OLS regression
XVar = df['Age']
YVar = df['MaxHR']
linearModel = sm.OLS(YVar, XVar)
results = linearModel.fit()
print(results.summary())
```

```
                                 OLS Regression Results
=================================================================================
Dep. Variable:                  MaxHR   R-squared (uncentered):
Model:                            OLS   Adj. R-squared (uncentered):
Method:                 Least Squares   F-statistic:
Date:                Tue, 07 Dec 2021   Prob (F-statistic):                     8
Time:                        22:09:08   Log-Likelihood:                         .
No. Observations:                  34   AIC:
Df Residuals:                      33   BIC:
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
Age            2.9435      0.150     19.659      0.000       2.639       3.248
==============================================================================
Omnibus:                        3.385   Durbin-Watson:                   2.489
Prob(Omnibus):                  0.184   Jarque-Bera (JB):                2.377
Skew:                          -0.639   Prob(JB):                        0.305
Kurtosis:                       3.216   Cond. No.                         1.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
```

```
#conclusions
# Our null hypothesis is that x and y have no relation. Since the p-value is 0,
# this means that we can reject the null hypothesis, so there is a relationship
# between the x and y variables, age and max heart rate respectively.

# The R-squared value of 0.921 also indicates that there is a strong correlation
# between the two variables age and max heart rate
```