STATISTICAL ANALYSIS

In [9]:
```
import pandas as pd
```

In [10]:
```
df=pd.read_csv("heart.csv")
```

In [20]:
```
df.head()
```

Out[20]:

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

In [21]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   age       303 non-null    int64  
 1   sex       303 non-null    int64  
 2   cp        303 non-null    int64  
 3   trtbps    303 non-null    int64  
 4   chol      303 non-null    int64  
 5   fbs       303 non-null    int64  
 6   restecg   303 non-null    int64  
 7   thalachh  303 non-null    int64  
 8   exng      303 non-null    int64  
 9   oldpeak   303 non-null    float64
 10  slp       303 non-null    int64  
 11  caa       303 non-null    int64  
 12  thall     303 non-null    int64  
 13  output    303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

DESCRIPTIVE STATISTICS

In [13]:
```
print("Mean:\n", df.mean())
print("\nMedian:\n", df.median())
print("\nMode:\n", df.mode().iloc[0])
print("\nStandard Deviation:\n", df.std())
print("\nVariance:\n", df.var())
print("\nRange:\n", df.max() - df.min())
print("\nSkewness:\n", df.skew())
print("\nKurtosis:\n", df.kurt())
```

```
Mean:
 age          54.366337
sex           0.683168
cp            0.966997
trtbps      131.623762
chol        246.264026
fbs           0.148515
restecg       0.528053
thalachh    149.646865
exng          0.326733
oldpeak       1.039604
slp           1.399340
caa           0.729373
thall         2.313531
output        0.544554
dtype: float64

Median:
 age          55.0
sex           1.0
cp            1.0
trtbps      130.0
chol        240.0
fbs           0.0
restecg       1.0
thalachh    153.0
exng          0.0
oldpeak       0.8
slp           1.0
caa           0.0
thall         2.0
output        1.0
dtype: float64

Mode:
 age          58.0
sex           1.0
cp            0.0
trtbps      120.0
chol        197.0
fbs           0.0
restecg       1.0
thalachh    162.0
exng          0.0
oldpeak       0.0
slp           2.0
caa           0.0
thall         2.0
output        1.0
Name: 0, dtype: float64

Standard Deviation:
 age          9.082101
sex          0.466011
cp           1.032052
trtbps      17.538143
chol        51.830751
fbs          0.356198
restecg      0.525860
thalachh    22.905161
exng         0.469794
oldpeak      1.161075
slp          0.616226
caa          1.022606
thall        0.612277
output       0.498835
dtype: float64

Variance:
 age          82.484558
sex           0.217166
cp            1.065132
trtbps      307.586453
chol       2686.426748
fbs           0.126877
restecg       0.276528
thalachh    524.646406
exng          0.220707
oldpeak       1.348095
slp           0.379735
caa           1.045724
thall         0.374883
output        0.248836
dtype: float64

Range:
 age          48.0
sex           1.0
cp            3.0
trtbps      106.0
chol        438.0
fbs           1.0
restecg       2.0
thalachh    131.0
exng          1.0
oldpeak       6.2
slp           2.0
caa           4.0
thall         3.0
output        1.0
dtype: float64

Skewness:
 age         -0.202463
sex         -0.791335
cp           0.484732
trtbps       0.713768
chol         1.143401
fbs          1.984652
restecg      0.162522
thalachh    -0.537410
exng         0.742532
oldpeak      1.269720
slp         -0.508316
caa          1.310422
thall       -0.476722
output      -0.179821
dtype: float64

Kurtosis:
 age         -0.542167
sex         -1.382961
cp          -1.193071
trtbps       0.929054
chol         4.505423
fbs          1.959678
restecg     -1.362673
thalachh    -0.061970
exng        -1.458317
oldpeak      1.575813
slp         -0.627521
caa          0.839253
thall        0.297915
output      -1.960783
dtype: float64
```

INFERENTIAL STATISTICS

In [24]:
```
from scipy import stats
trtbps_values = df['trtbps']


population_mean = 130


t_stat, p_value = stats.ttest_1samp(trtbps_values, population_mean)

print(f"T-Statistic: {t_stat}")
print(f"P-Value: {p_value}")
```

```
T-Statistic: 1.6116108638133604
P-Value: 0.10809121014912249
```

INTERPRETATION

- t-statistic represents that sample mean is 1.61 standard errors below the population mean
- p-value is larger than 0.05 so,we fail to reject the null hypothesis

CONFIDENCE INTERVAL

In [25]:
```
import numpy as np
from scipy import stats

# Sample mean and standard error for TRTBPS
sample_mean = np.mean(trtbps_values)
standard_error = stats.sem(trtbps_values)

# Compute 95% confidence interval for TRTBPS
confidence_interval = stats.norm.interval(0.95, loc=sample_mean, scale=standard_error)

print(f"95% Confidence Interval for TRTBPS: {confidence_interval}")
```

```
95% Confidence Interval for TRTBPS: (129.64902030398173, 133.59850444849354)
```

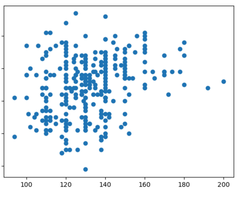
REGRESSION ANALYSIS

In [7]:
```
import statsmodels.api as sm

# Define independent variable (add constant for intercept)
X = sm.add_constant(df['trtbps'])

# Define dependent variable
y = df['age']

# Fit linear regression model
model = sm.OLS(y, X).fit()

# Print model summary
print(model.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                    age   R-squared:                       0.078
Model:                            OLS   Adj. R-squared:                  0.075
Method:                 Least Squares   F-statistic:                     25.48
Date:                Mon, 16 Sep 2024   Prob (F-statistic):           7.76e-07
Time:                        17:23:26   Log-Likelihood:                -1085.6
No. Observations:                 303   AIC:                             2175.
Df Residuals:                     301   BIC:                             2183.
Df Model:                           1                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         35.3255      3.806      9.283      0.000      27.837      42.814
trtbps         0.1447      0.029      5.048      0.000       0.088       0.201
==============================================================================
Omnibus:                        1.665   Durbin-Watson:                   2.015
Prob(Omnibus):                  0.435   Jarque-Bera (JB):                1.782
Skew:                          -0.108   Prob(JB):                        0.410
Kurtosis:                       2.693   Cond. No.                     1.01e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.01e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [10]:
```
import matplotlib.pyplot as plt
```

In [22]:
```
plt.scatter(df['trtbps'],y)
plt.show()
```



CONCLUSION

we have R-SQUARED = 0.078

it means that only 0.78% of the variance in age is defined by the TRTBPS

In [ ]: