John Salt

Intro to Data Science

Project 3

March 1, 2022

https://github.com/J-Salt/IDS/tree/main/HW3

Life Expectancy

Purpose

The purpose of this project is to use our knowledge of MySQL and Python to analyze the Life Expectancy dataset.

Methodology

We will use MySQL to cleanse the data and perform basic sorting on the set. Then using Python, we will perform correlation analysis.

Conclusion and Summary

1. **Do various predicting factors which has been chosen initially really affect Life expectancy? What are the predicting variables actually affecting life expectancy?**

   Of the various predicting factors I plotted, the only ones to show strong correlations were Schooling vs Life Expectancy and Adult Mortality Rate vs Life Expectancy. I believe that schooling shows a strong relationship because education can play a major role in someones health.

2. **Should a country having a lower life expectancy value(<65) increase its healthcare expenditure in order to improve its average lifespan?**

I believe based on what I have seen, that they should increase their healthcare expenditure. But, I also believe that they should simultaneously increasing their schooling.

3. **How does the Adult mortality rate affect life expectancy?**

   As adult mortality rate increases, life expectancy decreases. This is simply becauseif more people die before the age of sixty the life expectancy is going to be lower to reflect that.

4. **Does Life Expectancy have a positive or negative correlation with eating habits, social factors, drinking alcohol, etc.?**

   I noticed that with the eating and drinking habits a positive coorelation was shown despite what I would expect. For social factors I used schooling as the other factors I tested didn't seem to show any strong correlation. This was the one that made the most sense as education typically leads to more healthy individuals and more trust in healthcare.

5. **What is the impact of schooling on the lifespan of humans?**

   Schooling has a strong positive affect on the lifespan of humans as shown in the graphs at the end of this report.

6. **Does Life Expectancy have a positive or negative relationship with drinking alcohol?**

   From my testing it had a slight positive correlation, but this could be because typically richer countries have more money to spend on alcohol, but this also means they have better healthcare. So this could explain why the plot shows the opposite of expected.

7. **Do densely populated countries tend to have a lower life expectancy?**

From my analysis this was one that was hard to conclude. I included a plot at the end showing a linear regression and it shows a slight negative correlation between population and life expectancy. Since this was such a low r2 score, I would not conclude that there is a relationship in this dataset between population and life expectancy.

MySQL:

To start, I used SQL to remove the countries with a zero population. I also replaced missing values of life expectancy, adult mortality, alcohol, BMI, and GDP to their respective averages. I did not do this for total expenditure or percentage expenditure since it is possible for these values to be zero. Next, I found the countries with the highest and lowest averages for mortality rate, population, gdp, schooling, and alcohol consumption between 2010-2015. Below are the pictures of each.

country	Avg Adult Mortality
Lesotho	436.0000
Central African Republic	435.5000
Zimbabwe	421.0000
Swaziland	409.5000
Nigeria	366.0000
Chad	363.3333
Sierra Leone	362.6667
Angola	353.6667
Malawi	346.8333
South Sudan	346.8333
South Africa	322.5000
Cameroon	310.6667

country	Avg Adult Mortality
Tunisia	10.6667
Italy	31.5000
Iceland	35.3333
Israel	42.1667
Montenegro	48.5000
Kiribati	49.8333
Spain	51.3333
Netherlands	51.6667
Australia	52.3333
Switzerland	53.0000
Myanmar	53.8333
Cyprus	55.1667

country	Avg population
India	281099848.5000
Indonesia	175751726.3333
Pakistan	123341557.5000
Nigeria	90382108.3333
Brazil	70799653.8333
Bangladesh	59971506.3333
Russian Federation	55182376.1667
Turkey	51513111.5000
Ethiopia	49418339.5000
Mexico	42514971.5000
Algeria	37971074.1667
Philippines	36966300.0000

country	Avg population
Palau	292.0000
Tuvalu	1819.0000
Sri Lanka	2522.8333
Maldives	8454.3333
Georgia	9383.3333
Kiribati	14193.5000
Tonga	15189.8333
Israel	43490.3333
Seychelles	63329.6667
Samoa	104704.8333
Japan	106435.1667
Sao Tome...	127502.8333

country	Avg gdp
Switzerland	70817.14495
Luxembourg	63600.08135
Australia	61393.17152833334
Austria	40276.874234999996
Netherlands	34698.751229999994
Denmark	33730.662534999996
Canada	33583.381264999996
Israel	29932.439365
Ireland	27964.952261666665
Finland	26069.414426666666
Iceland	24840.910004999998
Italy	24713.54498333333
Sweden	23298.503781666663

country	Avg gdp
Senegal	94.40303758333334
Burundi	192.02427953333336
Sierra Leone	298.3913750666667
Niger	322.34020160000006
Guinea	353.5075783666666
Madagascar	377.95140244999993
Malawi	399.30295415
Ethiopia	405.5988584499999
Haiti	413.83100093333337
Liberia	414.5753463333333
Central Afri...	415.0316701333333
Guinea-Bissau	434.2454787666666
Rwanda	449.4002187333333

country	Avg schooling
Australia	20.083333333333332
Iceland	18.783333333333335
Ireland	18.5
Denmark	18.2
Netherlands	17.766666666666666
Norway	17.583333333333332
Spain	17.216666666666665
Argentina	17.166666666666668
Slovenia	17.066666666666666
Finland	16.95
Germany	16.900000000000002
Greece	16.900000000000002
Lithuania	16.566666666666666

country	Avg schooling
Tuvalu	0
South Sudan	4.083333333333333
Eritrea	5.05
Niger	5.066666666666666
Djibouti	6.1000000000000005
Central African Republic	6.966666666666668
Sudan	7.033333333333334
Chad	7.099999999999999
Burkina Faso	7.183333333333334
Pakistan	7.7
Mali	7.8500000000000005
Mauritania	8.049999999999999
Ethiopia	8.316666666666665

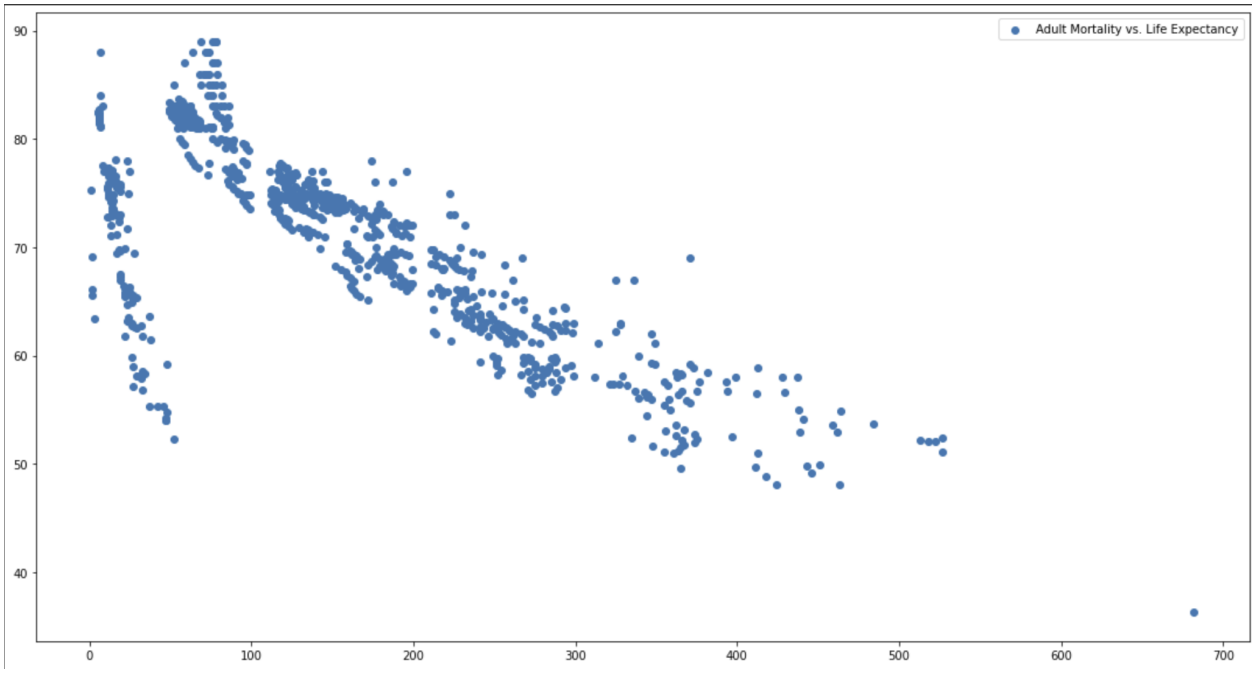
country	Avg alc consumption
Belarus	13.500424285044543
Lithuania	12.538757618377877
Austria	10.807090951711208
Croatia	10.768757618377876
France	10.31709095171121
Bulgaria	10.142090951711209
Luxembourg	10.107090951711209
Ireland	10.105424285044544
Portugal	10.05209095171121
Germany	9.975424285044541
Belgium	9.847090951711207
Poland	9.828757618377876

country	Avg alc consumption
Afghanistan	0.01
Tuvalu	0.01
Eritrea	0.615
Bangladesh	0.7254242850445417
Mauritania	0.7254242850445417
Pakistan	0.7454242850445417
Comoros	0.7670909517112084
Niger	0.7670909517112084
Iraq	0.7820909517112083
Indonesia	0.7870909517112085
Guinea	0.7920909517112085
Bhutan	0.8070909517112085

Do densely populated countries tend to have lower life expectancy?

From the data I cannot conclude that countries with a higher population density have lower life expectancy. By looking at the data, it seems that GDP has more of a correlation with life expectancy. But even this doesn't seem to have a strong positive relationship.

Python:



1. **How does Adult mortality rates affect life expectancy?**

Adult mortality rate has a negative correlation with life expectancy. This means that as adult mortality rate increases, the countries life expectancy decreases. This makes sense since a higher adult mortality rate means more people are going to pass away before the age of 60, thus lowering the life expectancy.

2. **Does life expectancy have positive or negative correlation with eating habits, drinking alcohol, social factors, and economic factors?**

   After doing various plots with linear regression models to see the relationship between the different factors. From this I can conclude that BMI has a positive correlation with life expectancy. This is the opposite of what I would expect, but typical countries with higher BMIs tend to have more GDP and thus more to spend on healthcare. Next, I compared alcohol consumption to life expectancy and noticed that this also showed a positive correlation, possibly for the same reason as BMI. For social factors I chose to compare life expectancy and schooling and saw a strong positive correlation. This also could be explained by the increased funding. Then, I did GDP and saw a slight positive correlation, which is not what I expected considering the outcome of the past few. In order to confirm my results with GDP I did the same for percent and total expenditure and saw similar correlations.

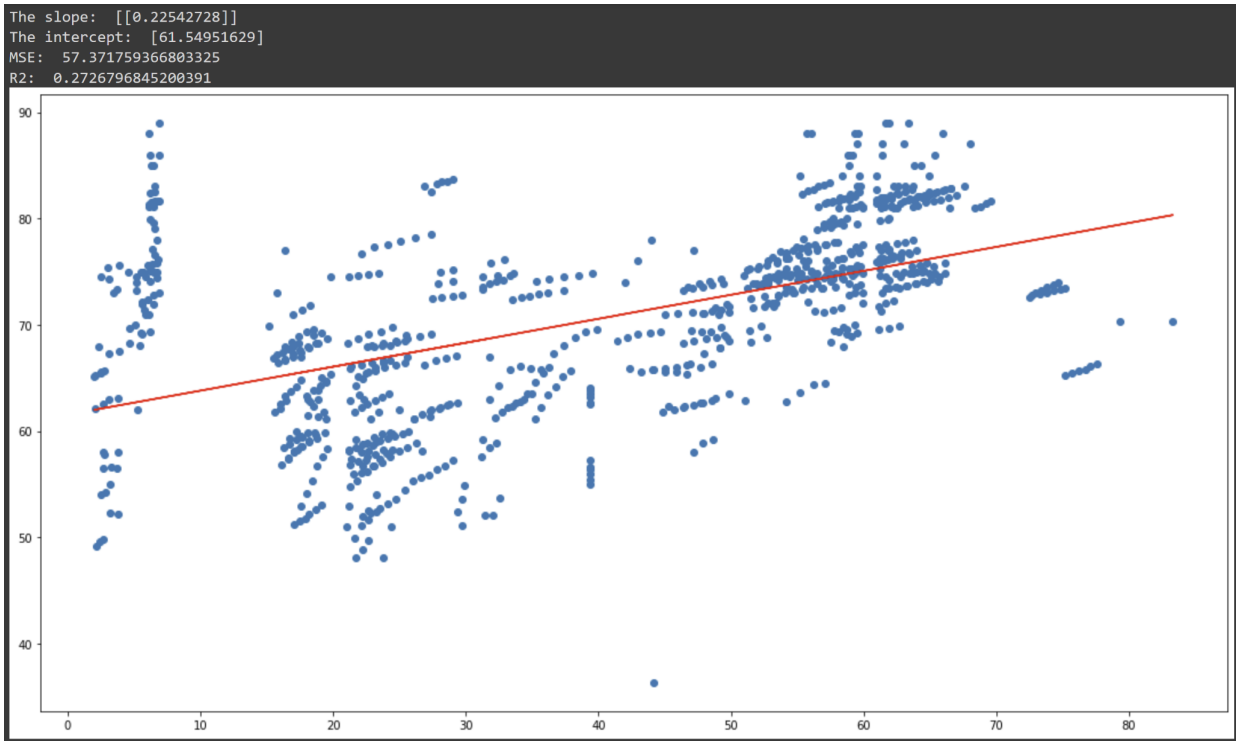
3. **What is the impact of schooling on the lifespan of humans?**

Schooling seemed to have the largest positive impact on life expectancy. I believe that this is do to not only an increase in healthcare funding in these countries, but also better education leading to people knowing better health habits than those in poorer countries.

Images:

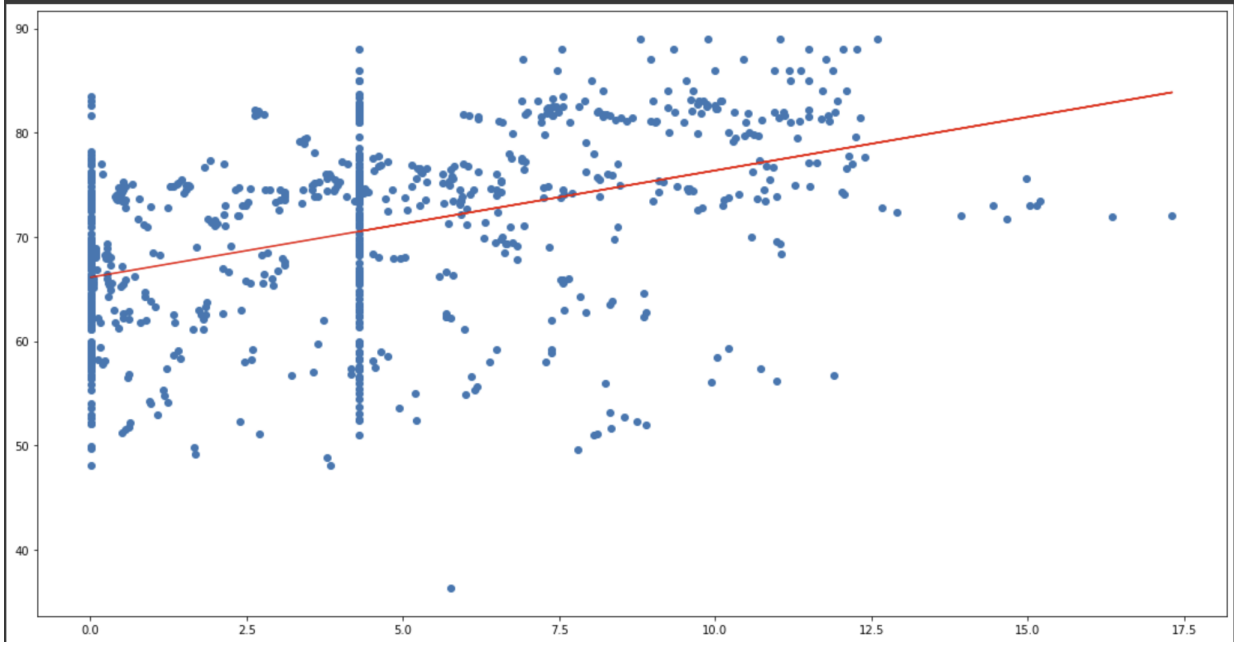
Below is all of the various plots with their linear regression models. All of these are vs. Life Expectancy

BMI:



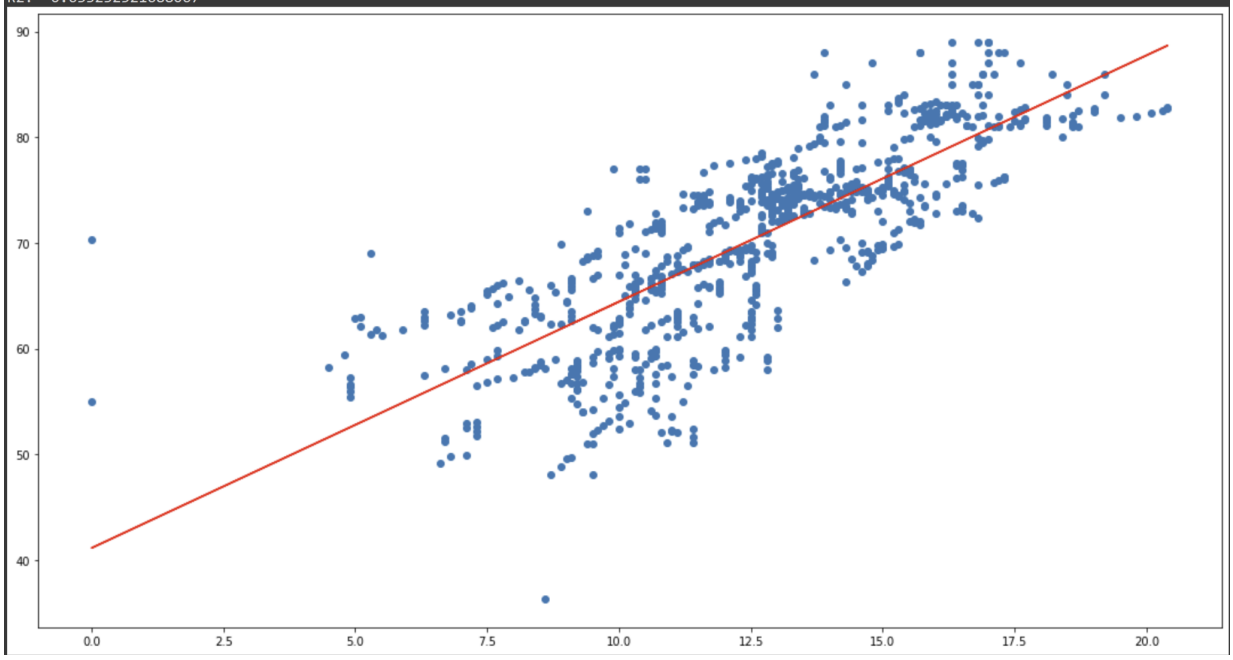
Alcohol Consumption:

```
The slope:  [[1.02459767]]
The intercept:  [66.12905862]
MSE:  63.85472783756602
R2:  0.19049300024467264
```



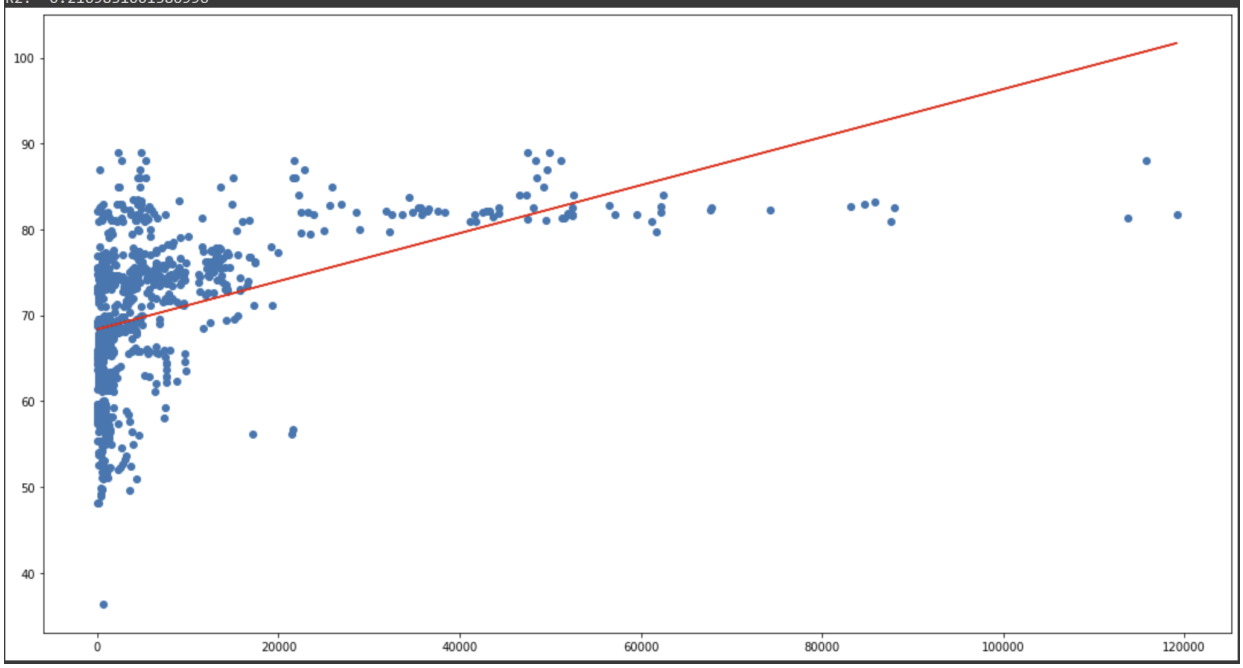
Schooling:

```
The slope:  [[2.32932063]]
The intercept:  [41.15398235]
MSE:  28.930988261197868
R2:  0.633232521688067
```



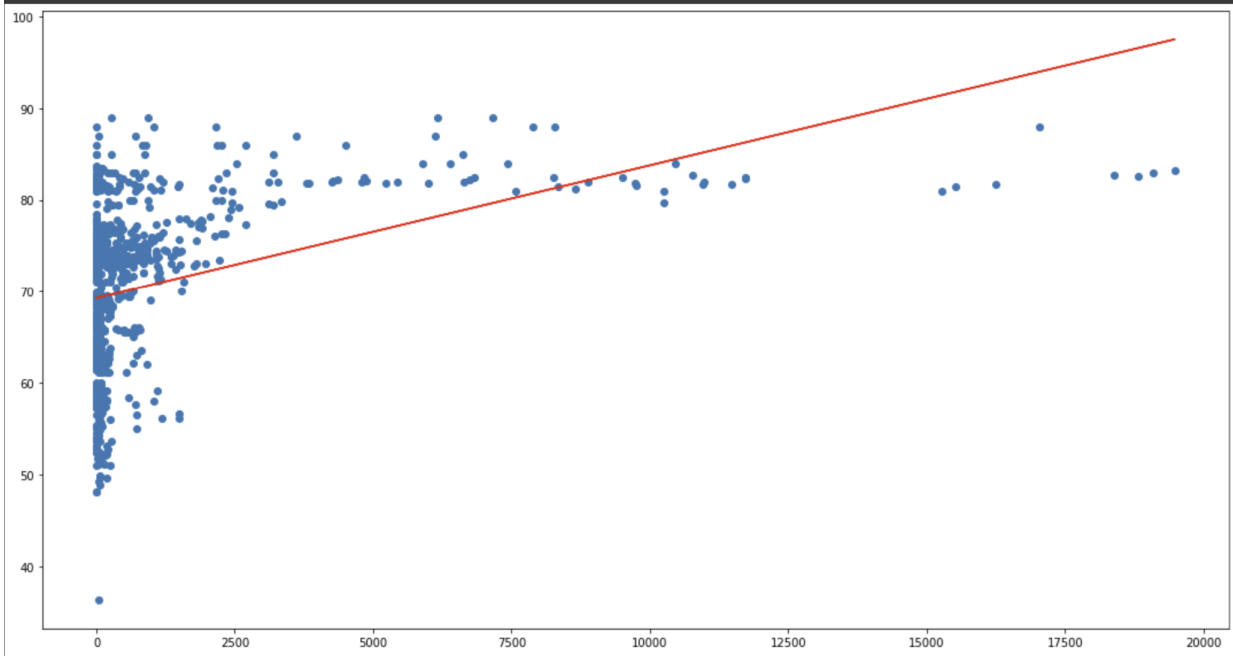
GDP:

```
The slope:  [[0.00027976]]
The intercept:  [68.37469563]
MSE:  61.76516206526969
R2:  0.2169831001386996
```



% Expenditure:

```
The slope:  [[0.00144934]]
The intercept:  [69.28682153]
MSE:  67.19986168764358
R2:  0.14808565847908461
```



Total Expenditure:

```
The slope:  [[0.41348749]]
The intercept:  [68.3974701]
MSE:  77.00300429699331
R2:  0.02380805475872705
```