Patrick O'Boyle
CSC 346 Introduction to Data Science
Project 3
3/1/2022
Repo:

1.
```
SET SQL_SAFE_UPDATES = 0;
/*select population from LifeExpectancy*/
DELETE FROM LifeExpectancy where population = 0;
DELETE FROM LifeExpectancy where Total_Expenditure=0;
select * from LifeExpectancy;
```

Here I deleted the rows with population 0 and also the rows with Total Expenditure 0.

1.
```
select count(*) from LifeExpectancy;
```

Result Grid | Filter Rows: | Expor

| count(*) |
|---|
| 714 |

Here is the total count of the countries after I cleansed the data.

2. Here is the highest mortality rate in its country.

```
SELECT Country, MAX(Adult_Mortality) FROM LifeExpectancy GROUP BY Adult_Mortality desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Country | MAX(Adult_Mortality) |
|---|---|
| Haiti | 682 |
| Lesotho | 527 |
| Lesotho | 522 |

Here is the lowest mortality rate in its country.

```
SELECT Country, MIN(Adult_Mortality) FROM LifeExpectancy GROUP BY Adult_Mortality;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Country | MIN(Adult_Mortality) |
|---|---|
| Palau | 0 |
| Fiji | 2 |
| Botswana | 3 |

3. Here is the country with the highest average population.

```
SELECT Country, MAX(population) FROM LifeExpectancy GROUP BY population desc;
/*SELECT Country, min(population) FROM LifeExpectancy GROUP BY population desc;*/
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Country | MAX(population) |
|---|---|
| India | 1293859294 |

Here is the country with the lowest average population.

```
SELECT Country, min(population) FROM LifeExpectancy GROUP BY population ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Country | min(population) |
|---|---|
| Maldives | 41 |
| Hungary | 123 |
| Palau | 292 |

4. Here is how I got the countries with the highest and lowest GDP's.

```
SELECT Country, min(GDP) FROM LifeExpectancy GROUP BY GDP desc ;
SELECT Country, max(GDP) FROM LifeExpectancy GROUP BY GDP desc ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Country | max(GDP) |
|---|---|
| Luxembourg | 119172.7418 |

5. Here is the list of countries with the highest and lowest average schooling years.

```
SELECT Country, max(Schooling) FROM LifeExpectancy GROUP BY Schooling desc ;
SELECT Country, min(Schooling) FROM LifeExpectancy GROUP BY Schooling ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Country | min(Schooling) |
|---|---|
| Tuvalu | 0 |
| Niger | 4.5 |
| Niger | 4.8 |

6. Here is how I determine which countries have the highest and lowest average alcohol consumptions. Lowest: Palau 0: , Highest: Belarus 17.31.

```
SELECT Country, min(Alcohol) FROM LifeExpectancy GROUP BY Alcohol;
SELECT Country, max(Alcohol) FROM LifeExpectancy GROUP BY Alcohol desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Country | max(Alcohol) |
|---|---|
| Belarus | 17.31 |
| Belarus | 16.35 |
| Lithuania | 15.19 |
| Lithuania | 15.14 |
| Lithuania | 15.04 |

7. The trend I am seeing is that yes, densely populated countries tend to have lower life expectancies.

```
select country, population, Adult_Mortality
from LifeExpectancy
group by Adult_Mortality,population desc
```

Result Grid | Filter Rows: | Export:

| country | population | Adult_Mortality |
|---|---|---|
| Central ... | 4499653 | 451 |
| Swaziland | 122843 | 459 |
| Malawi | 1516795 | 462 |
| Sierra Le... | 779162 | 463 |
| Zimbabwe | 14386649 | 464 |
| Lesotho | 289928 | 513 |
| Lesotho | 2117361 | 518 |
| Lesotho | 2145785 | 522 |
| Zimbabwe | 1486317 | 527 |
| Lesotho | 24551 | 527 |
| Haiti | 9999617 | 682 |

By using these instructions in MYSQL I was able to see the mortality rates side by side with the populations of that specific country. I could see that usually, countries with high populations tend to have higher mortality rates.

Part 2:

This is how i connected to the database in python:

dataScienceProject3

File Edit View Insert Runtime Tools Help All changes saved

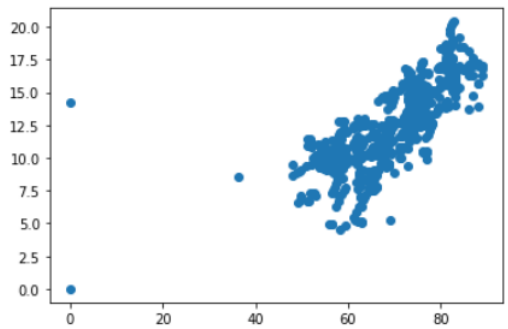
+ Code + Text

```
[1] import pandas as pd

[2] from tabulate import tabulate

pip install mysql-connector-python

Collecting mysql-connector-python
  Downloading mysql_connector_python-8.0.28-cp37-cp37m-manylinux1_x86_64.whl (37.6 MB)
     |████████████████████████████████| 37.6 MB 24.8 MB/s
Requirement already satisfied: protobuf>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from mysql-connector-python) (3.17.3)
Requirement already satisfied: six>=1.9 in /usr/local/lib/python3.7/dist-packages (from protobuf>=3.0.0->mysql-connector-python) (1.15.0)
Installing collected packages: mysql-connector-python
Successfully installed mysql-connector-python-8.0.28

[4] import mysql.connector as sql

[5] # use the same credentials you use to connect to MySQL via Workbench

    db_connection = sql.connect(host='208.109.18.154', database='ids16db', user='ids16', password='Wky3401')
    db_cursor = db_connection.cursor()
    db_cursor.execute('SELECT * FROM LifeExpectancy')

    table_rows = db_cursor.fetchall()

    lifeExpectDF = pd.DataFrame(table_rows)

    print(tabulate(lifeExpectDF, headers='keys', tablefmt='fancy_grid'))

    685 | Ukraine | 2011 | 75 | 24 | 8.48 | 415.163 | 59 | 6.98 | 3569.76 | 45761 | 14.9
```

1s completed at 10:15 AM

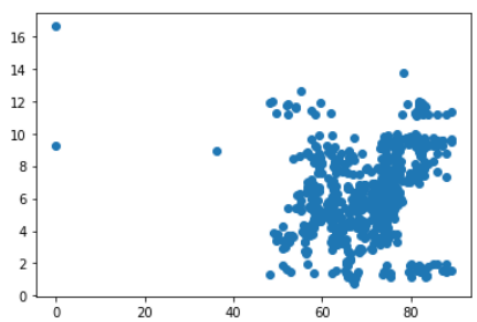
This is the correlation between Life expectancy and Body mass index.

```python
from sklearn.linear_model import LinearRegression
import numpy as np
import matplotlib.pyplot as plt
X=lifedf[2]
Y=lifedf[10]
# Visualize the dataset and the regression line:
plt.scatter(X, Y)

plt.show()
#seems to be a large negative covaraince
#this shows the correlation between 2 and 5 which means life expectancy and BMI
```



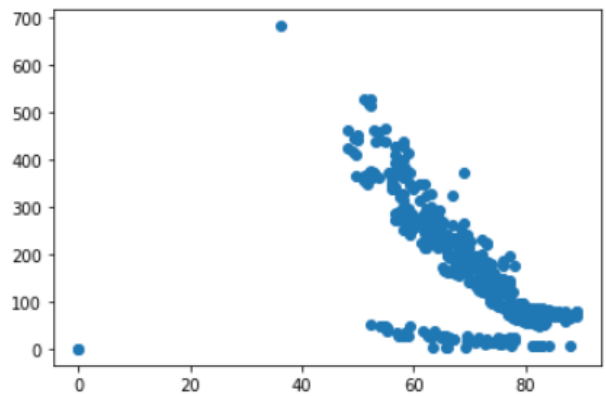

This is the correlation between Life expectancy and Alcohol.

```python
X=lifedf[2]
Y=lifedf[7]
# Visualize the dataset and the regression line:
plt.scatter(X, Y)

plt.show()
#seems to be a large negative covaraince
#this shows the correlation between 2 and 7 which means life expectancy and Alcohol
```




This is the correlation between Life expectancy and Adult morality.

```
X=lifedf[2]
Y=lifedf[3]
# Visualize the dataset and the regression line:
plt.scatter(X, Y)

plt.show()
#seems to be a large negative covaraince
#this shows the correlation between 2 and 7 which means life expectancy and adult morality
```




Findings:
I decided to examine the correlation between life expectancy and three other factors. First I did BMI which I think looks like a positive correlation. Next I examined alcol. This was really scattered and I couldn't gather any information from it. Lastly I examined adult morality which showed me a negative correlation.

**How do Adult mortality rates affect life expectancy?**

This was a negative correlation.

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

Through doing this project I would definitely say that there is a correlation between doing things that are bad for your health as well as where the person is when talking about life expectancy.

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

I found a positive correlation between them.

<u>Report</u>

1. The purpose of this assignment was to get comfortable with python and mysql while using big data sets. The data set was very big and a little intimidating but easy to examine using python and mysql. I was able to make

accurate assumptions about general trends just by plotting the columns in python.

2. In each problem we were tasked to examine the data set and clean it up. In mysql I did most of the cleaning. I took out a lot of rows just because it had no data would mess up my future findings.
3. The technologies I used were mysql workbench and python in google colab.
4. My conclusion is that it would have been so hard to examine this data if I could not use mysql and Python. By using them I was easily able to make inferences from huge data sets.
5. I used the slides from class to learn how to connect the database to work with python.
   https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html
   https://pythontic.com/pandas/serialization/mysql
   https://dev.mysql.com/doc/connector-python/en/connector-python-example-connecting.html
   https://www.w3schools.com/python/python_mysql_getstarted.asp