

Naresh Kalluri (10755799)

Intro to Data Science (CSC 346 D01 Spring 2022).

Assignment – 3 (Project 3).

Date: 3-1-2022.

GitHub Repository: <https://github.com/nk755799/IDS>

Project 3

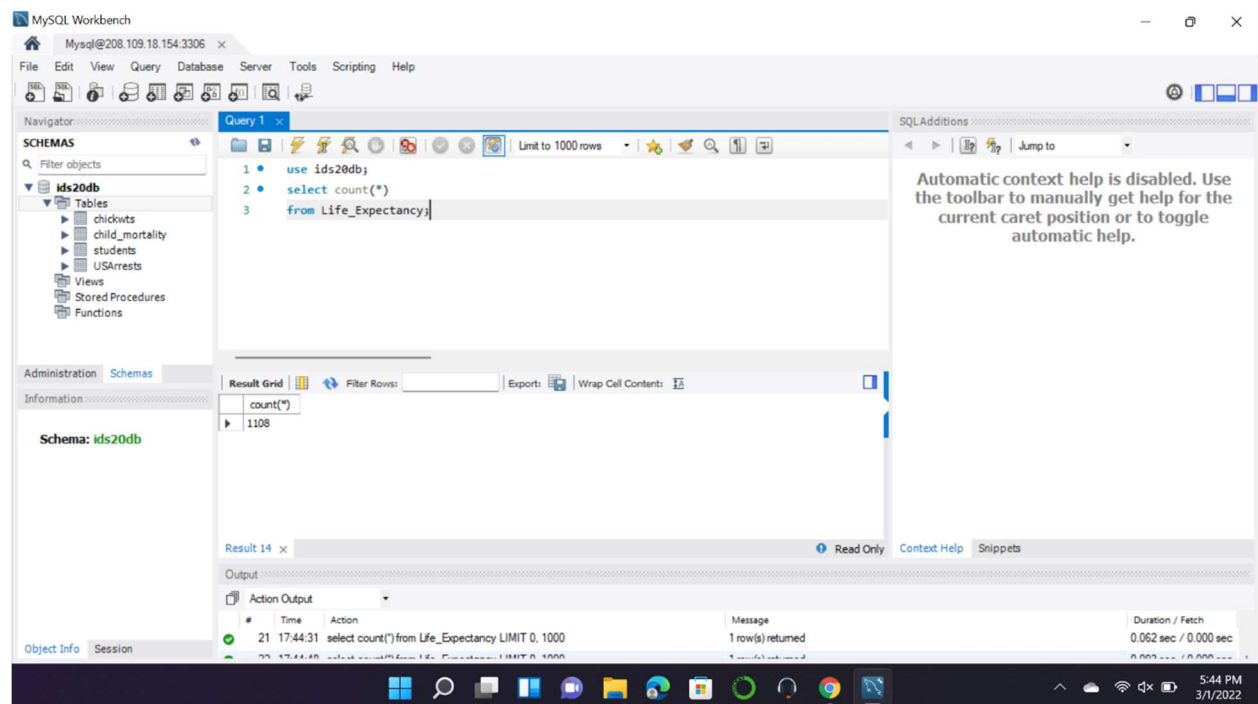
MySQL:

Purpose of the project:

Life Expectancy: Statistical Analysis on Factors Influencing Life Expectancy

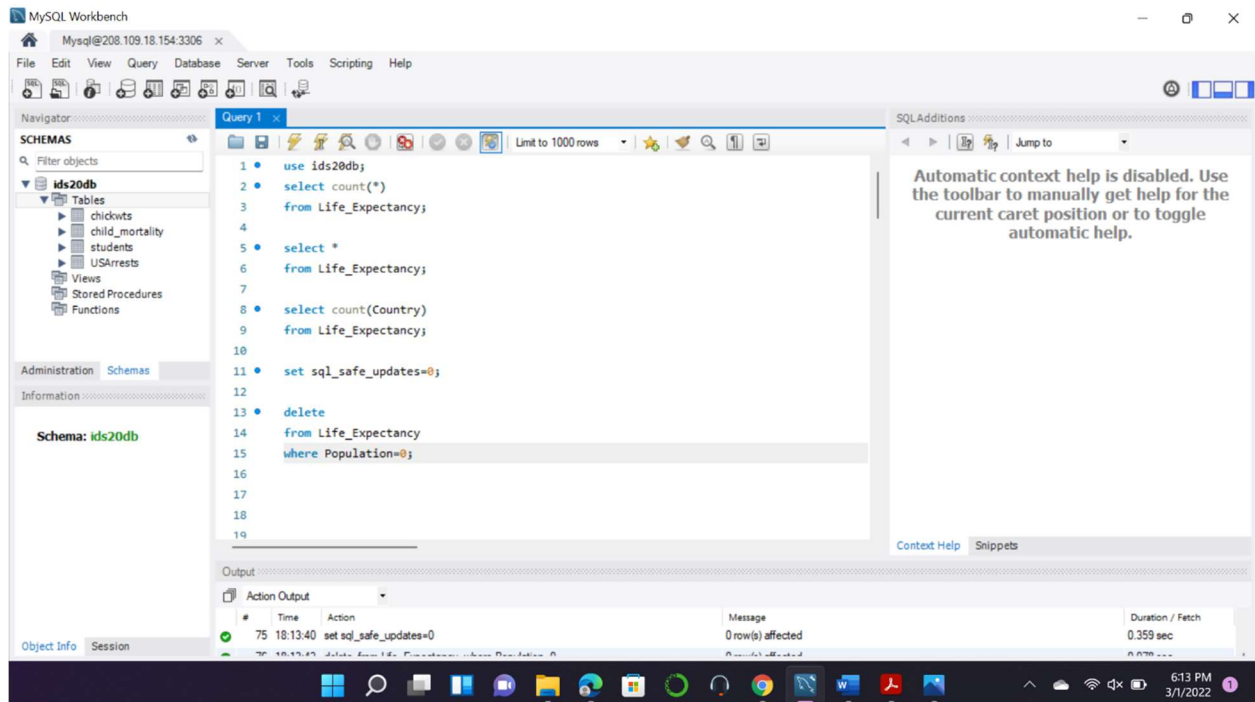
Operations:

Imported the Life_Expectancy file from the Professor GitHub repository and uploaded it to my MySQL Database.

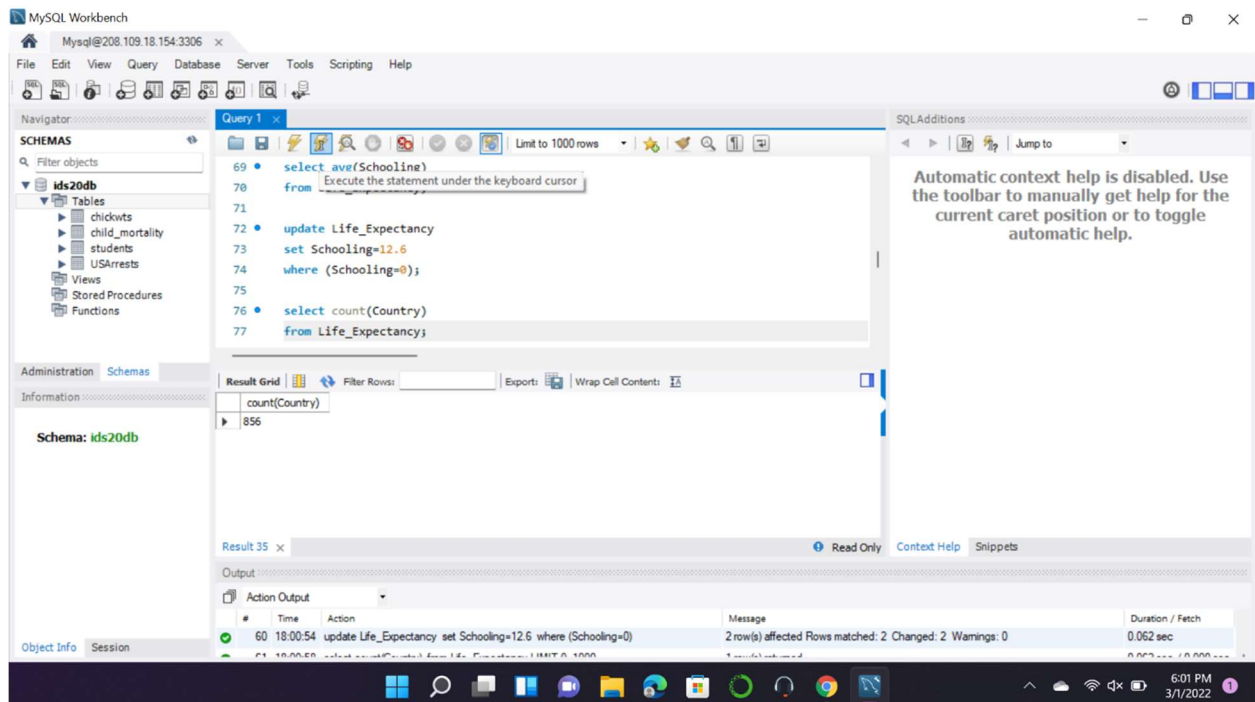


Delete all rows with population = 0. Perform further data cleaning for other attributes, as necessary.

Answer: Deleted all rows where the population = 0



Display total count of countries after data cleaning.



List of countries with the highest and lowest average mortality rates (years 2010-2015).

The screenshot shows the MySQL Workbench interface. The query editor contains the following SQL code:

```
83 from Life_Expectancy
84 group by Country
85 order by avg(Adult_Mortality) desc;
86
87 • select Country, avg(Adult_Mortality)
88 from Life_Expectancy
89 group by Country
90 order by avg(Adult_Mortality);
91
```

The Results Grid displays the following data:

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

The Output pane shows the execution message: "select Country, avg(Adult_Mortality) from Life_Expectancy group by Country order by... 145 row(s) returned".

The screenshot shows the MySQL Workbench interface. The query editor contains the following SQL code:

```
83 from Life_Expectancy
84 group by Country
85 order by avg(Adult_Mortality) desc;
86
87 • select Country, avg(Adult_Mortality)
88 from Life_Expectancy
89 group by Country
90 order by avg(Adult_Mortality);
91
```

The Results Grid displays the following data:

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

The Output pane shows the execution message: "select Country, avg(Adult_Mortality) from Life_Expectancy group by Country order by... 145 row(s) returned".

A OneDrive notification is visible in the bottom right corner: "Screenshot saved. The screenshot was added to your OneDrive."

List of countries with the highest and lowest average population (years 2010-2015).

MySQL Workbench interface showing a query in the `Life_Expectancy` table. The query is:

```
87 • select Country,avg(Adult_Mortality)
88   from Life_Expectancy
89   group by Country
90   order by avg(Adult_Mortality);
91
92 • select Country,avg(Population)
93   from Life_Expectancy
94   group by Country
95   order by avg(Population);
```

The `Result Grid` shows the results of the second query, ordered by average population:

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

The `Output` pane shows an error message:

```
114 23:33:33 select Country,avg(Population) from Life_Expectancy group by Country order by avg... Error Code: 1054. Unknown column 'Popopulation' in 'order clause' 0.062 sec
```

MySQL Workbench interface showing a query in the `Life_Expectancy` table. The query is:

```
92 • select Country,avg(Population)
93   from Life_Expectancy
94   group by Country
95   order by avg(Population);
96
97 • select Country,avg(Population)
98   from Life_Expectancy
99   group by Country
100  order by avg(Population) desc;
```

The `Result Grid` shows the results of the second query, ordered by average population in descending order:

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

The `Output` pane shows a successful message:

```
115 23:33:42 select Country,avg(Population) from Life_Expectancy group by Country order by avg... 145 row(s) returned 0.062 sec / 0.000 sec
```

List of countries with the highest and lowest average GDP (years 2010-2015).

This screenshot shows the MySQL Workbench interface. The 'Schemas' pane on the left displays the 'ids20db' database with various tables. The 'Query' editor in the center contains two SQL queries. The first query calculates the average population by country, and the second query calculates the average GDP by country. The 'Result Grid' shows the results of the first query, listing countries and their average GDP values. The 'Output' pane at the bottom shows the execution details of the first query.

```
97 • select Country,avg(Population)
98   from Life_Expectancy
99   group by Country
100  order by avg(Population) desc;
101
102 • select Country,avg(GDP)
103   from Life_Expectancy
104   group by Country
105  order by avg(GDP) desc;
```

Country	avg(GDP)
Switzerland	70817.14495
Luxembourg	63600.08135
Australia	61393.17152833334
Austria	40276.874234999996
Netherlands	34698.751229999994
Denmark	33730.662534999996
Canada	33583.381264999996

Result 84 x

Output

#	Time	Action	Message	Duration / Fetch
116	23:34:05	select Country,avg(Population) from Life_Expectancy group by Country order by avg...	145 row(s) returned	0.062 sec / 0.000 sec

This screenshot shows the MySQL Workbench interface. The 'Schemas' pane on the left displays the 'ids20db' database with various tables. The 'Query' editor in the center contains two SQL queries. The first query calculates the average GDP by country, and the second query calculates the average GDP by country. The 'Result Grid' shows the results of the first query, listing countries and their average GDP values. The 'Output' pane at the bottom shows the execution details of the first query.

```
102 • select Country,avg(GDP)
103   from Life_Expectancy
104   group by Country
105  order by avg(GDP) desc;
106
107 • select Country,avg(GDP)
108   from Life_Expectancy
109   group by Country
110  order by avg(GDP);
```

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

Result 85 x

Output

#	Time	Action	Message	Duration / Fetch
117	23:35:22	select Country,avg(GDP) from Life_Expectancy group by Country order by avg(GDP)...	145 row(s) returned	0.093 sec / 0.000 sec

List of countries with the highest and lowest average Schooling (years 2010-2015).

This screenshot shows the MySQL Workbench interface. The query editor contains the following SQL code:

```
109 group by Country
110 order by avg(GDP);
111
112 • select Country,avg(Schooling)
113 from Life_Expectancy
114 group by Country
115 order by avg(Schooling);
```

The 'Result Grid' tab is active, displaying the results of the query. The table has two columns: 'Country' and 'avg(Schooling)'. The results are sorted by 'avg(Schooling)' in ascending order.

Country	avg(Schooling)
Eritrea	5.05
Niger	5.066666666666666
Djibouti	6.1000000000000005
South Sudan	6.183333333333333
Central African Republic	6.966666666666666
Sudan	7.033333333333334
Chad	7.099999999999999

The 'Output' tab shows the execution details: 118 23:35:44 select Country,avg(Schooling) from Life_Expectancy group by Country order by avg(GDP... 145 row(s) returned. Duration / Fetch: 0.078 sec / 0.000 sec.

This screenshot shows the MySQL Workbench interface. The query editor contains the following SQL code:

```
112 • select Country,avg(Schooling)
113 from Life_Expectancy
114 group by Country
115 order by avg(Schooling);
116
117 • select Country,avg(Schooling)
118 from Life_Expectancy
119 group by Country
120 order by avg(Schooling) desc;
```

The 'Result Grid' tab is active, displaying the results of the query. The table has two columns: 'Country' and 'avg(Schooling)'. The results are sorted by 'avg(Schooling)' in descending order.

Country	avg(Schooling)
Australia	20.083333333333332
Iceland	18.783333333333335
Ireland	18.5
Denmark	18.2
Netherlands	17.766666666666666
Norway	17.583333333333332
Spain	17.216666666666665

The 'Output' tab shows the execution details: 119 23:37:05 select Country,avg(Schooling) from Life_Expectancy group by Country order by avg(... 145 row(s) returned. Duration / Fetch: 0.078 sec / 0.000 sec.

Which countries have the highest and lowest average alcohol consumption (years 2010-2015)?

The screenshot shows the MySQL Workbench interface with a query window titled 'Life_Expectancy'. The query is as follows:

```
117 • select Country,avg(Schooling)
118 from Life_Expectancy
119 group by Country
120 order by avg(Schooling) desc;
121
122 • select Country,avg(Alcohol)
123 from Life_Expectancy
124 group by Country
125 order by avg(Alcohol) desc;
```

The 'Result Grid' displays the results of the first query, showing the average schooling for various countries. The data is as follows:

Country	avg(Alcohol)
Belarus	13.378333333333332
Lithuania	12.416666666666666
Austria	10.685
Croatia	10.646666666666667
France	10.195
Bulgaria	10.020000000000001
Luxembourg	9.985
Poland	9.955555555555555

The 'Output' pane shows the execution of the query, indicating that 145 rows were returned.

The screenshot shows the MySQL Workbench interface with a query window titled 'Life_Expectancy'. The query is as follows:

```
122 • select Country,avg(Alcohol)
123 from Life_Expectancy
124 group by Country
125 order by avg(Alcohol) desc;
126
127 • select Country,avg(Alcohol)
128 from Life_Expectancy
129 group by Country
130 order by avg(Alcohol);
```

The 'Result Grid' displays the results of the first query, showing the average alcohol consumption for various countries. The data is as follows:

Country	avg(Alcohol)
Afghanistan	0.01
Tuvalu	0.01
Bangladesh	0.6033333333333333
Mauritania	0.6033333333333333
Eritrea	0.615
Pakistan	0.6233333333333333
Comoros	0.645
Indonesia	0.646

The 'Output' pane shows the execution of the query, indicating that 145 rows were returned.

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

The screenshot shows the MySQL Workbench interface. The left sidebar displays the 'SCHEMAS' panel with a tree view of the 'ids20db' database, including tables like 'chickwts', 'child_mortality', 'students', and 'USArrests'. The main editor window shows a SQL query in the 'Life_Expectancy' table:

```
148 • select Country,Population,Life_Expectancy
149 from Life_Expectancy
150 order by Life_Expectancy;
```

The 'Result Grid' below the query shows the following data:

Country	Population	Life_Expectancy
Haiti	9999617	36.3
Sierra Leone	645872	48.1
Sierra Leone	779162	48.1
Sierra Leone	6611692	48.9
Central African Republic	4448525	49.2
Angola	23369131	49.6
Sierra Leone	676613	49.7
Sierra Leone	6611692	49.9

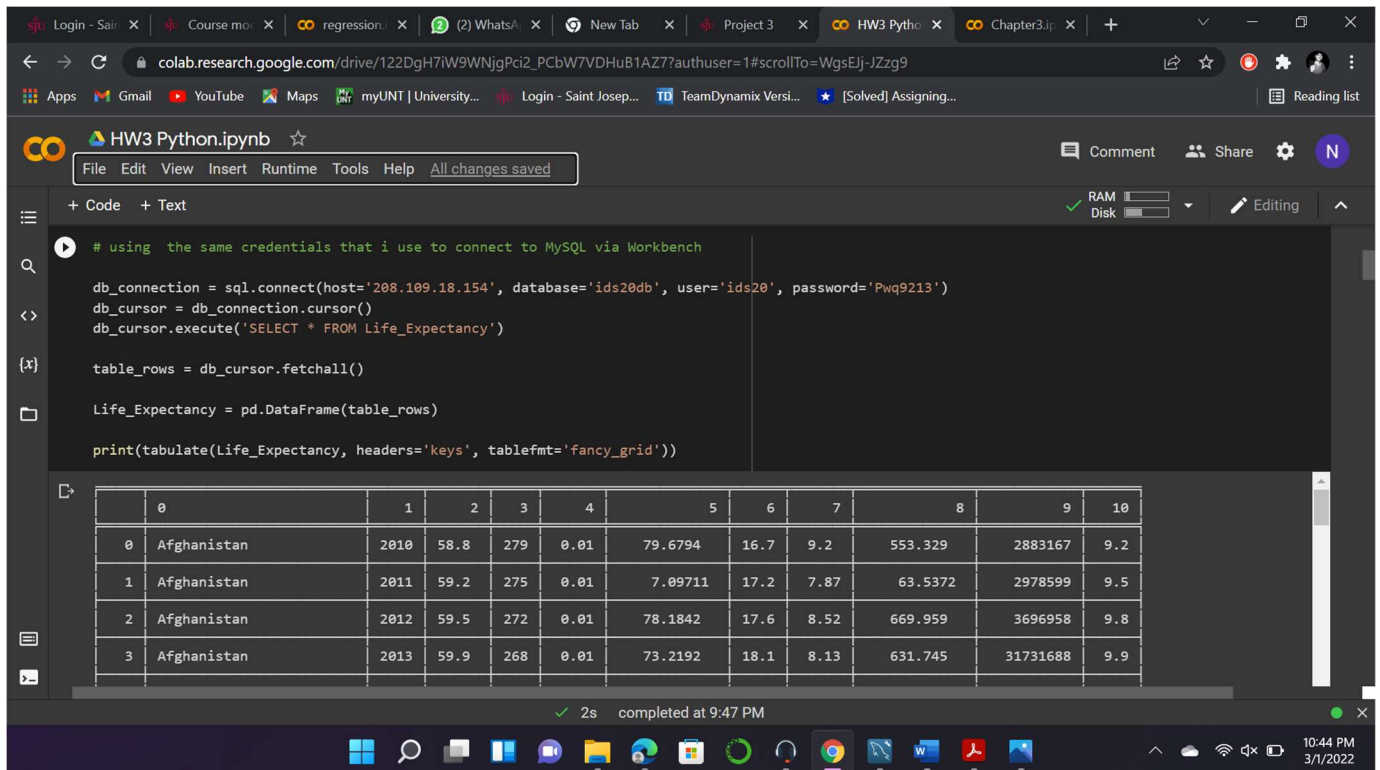
The bottom panel shows the 'Output' tab with the following message:

#	Time	Action	Message	Duration / Fetch
122	23:38:27	select Country,avg(Alcohol) from Life_Expectancy group by Country order by avg(Al...	145 row(s) returned	0.094 sec / 0.000 sec

The system tray at the bottom indicates the time is 11:52 PM on 3/1/2022.

Python:

Imported the cleaned life expectancy data file from MySQL into a data frame and created a Jupiter notebook in my Google collab.



```
# using the same credentials that i use to connect to MySQL via Workbench

db_connection = sql.connect(host='208.109.18.154', database='ids20db', user='ids20', password='Pwq9213')
db_cursor = db_connection.cursor()
db_cursor.execute('SELECT * FROM Life_Expectancy')

table_rows = db_cursor.fetchall()

Life_Expectancy = pd.DataFrame(table_rows)

print(tabulate(Life_Expectancy, headers='keys', tablefmt='fancy_grid'))
```

	0	1	2	3	4	5	6	7	8	9	10
0	Afghanistan	2010	58.8	279	0.01	79.6794	16.7	9.2	553.329	2883167	9.2
1	Afghanistan	2011	59.2	275	0.01	7.09711	17.2	7.87	63.5372	2978599	9.5
2	Afghanistan	2012	59.5	272	0.01	78.1842	17.6	8.52	669.959	3696958	9.8
3	Afghanistan	2013	59.9	268	0.01	73.2192	18.1	8.13	631.745	31731688	9.9

2s completed at 9:47 PM

Remaining all the questions I explained in my Jupiter notebook and uploaded them into my canvas.

Kindly refer to it.

