GLOBAL ANALYSIS OF MENTAL HEALTH AND SUICIDE RATES

BALAKUMAR RAHUL

MS Health Informatics
INFO-B 512 Scientific & Clinical Data Management FA23
Luddy School of Informatics, Computing, and Engineering,
IUPUI, Indianapolis

Project Overview

This project conducted an in-depth examination of global mental health infrastructure data and suicide statistics across countries, age groups, and genders to uncover insightful patterns, trends, and solutions. A structured MySQL database was designed using a normalized entity-relationship diagram for efficient storage and analysis. Rigorous data analysis in Python combined with compelling visualizations revealed several key insights regarding correlates of suicide rates and high-risk demographic groups. The evidence gathered allows us to provide targeted, data-driven recommendations to key stakeholders for future suicide prevention and mental health promotion programs.

Data Visualization

My Approach

To perform data visualization for our project, I used Python libraries like Matplotlib, Seaborn, and Pandas as they yielded the most accurate and reliable results for our chosen dataset. The visualizations generated helped us to garner meaningful insights along with helping us understand the numerous trends and patterns among various age groups, genders, and years.

As with any visualization in general, the main points that were kept in mind were as follows:

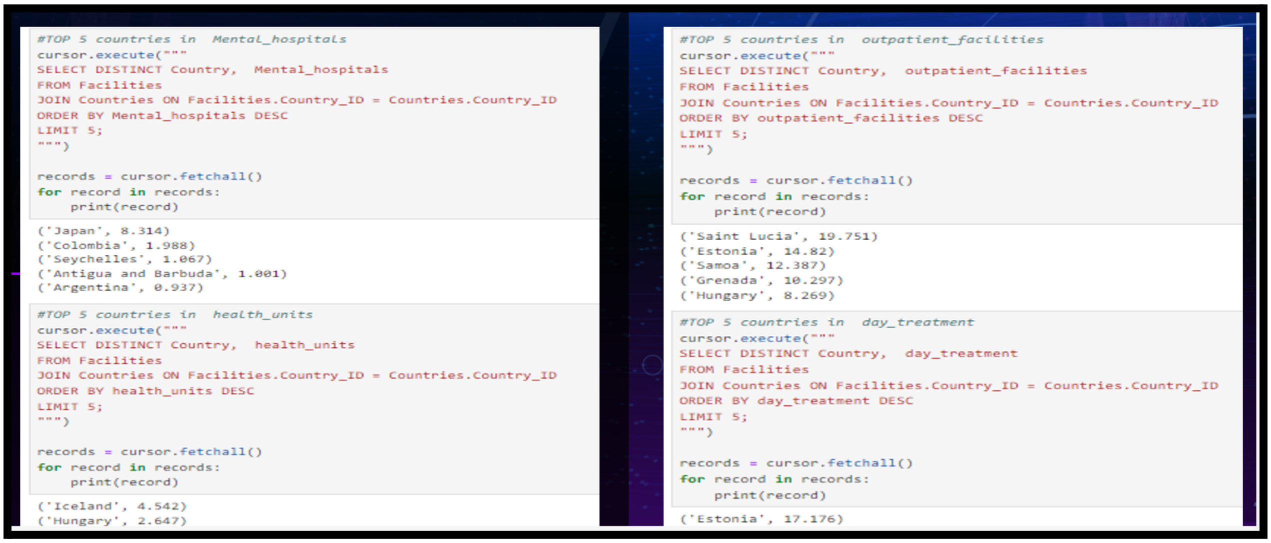
- ❖ *Data Exploration*: To understand the structure and content of the data.
- ❖ *Data Cleaning*: Handling any missing values, outliers, and possible inconsistencies.
- ❖ *Data Transformation:* Converting data types or creating new variables wherever deemed necessary.
- ❖ *Visualization Techniques:* Choosing the most appropriate visualization techniques based on the nature of the data.
- ❖ *Insights and Interpretation*: Deriving actionable insights from the visualizations and communicating the findings effectively.

For this project, to summarize and derive the best results, the visualization techniques I used are as follows:-

Correlation Matrices

Correlation matrices were used to establish relationships between the various attributes in the *facilities* table and the *human resources* table globally. A value closer to one means a strong positive correlation and a value closer to negative one means a strong negative correlation.

As shown in Fig. A1 (*see the Appendix*) the *facilities* table showed moderately high positive correlations between outpatient facilities and day treatment care centers (r = 0.59). There was also a positive correlation observed between residential facilities and outpatient facilities (r = 0.52). In addition, there were some strong negative correlations observed between mental hospitals and residential facilities (r = 0.01) as well as between health units and residential facilities (r = 0.09). This shows that countries are taking an integrated approach to building up their mental healthcare capacity. The strong negative correlations could mean that the countries are shifting resources away from more dated forms of care like mental hospitals and standalone health units towards more progressive residential facilities for those with mental health disorders.



```
#TOP 5 countries in  Mental_hospitals
cursor.execute("""
SELECT DISTINCT Country,  Mental_hospitals
FROM Facilities
JOIN Countries ON Facilities.Country_ID = Countries.Country_ID
ORDER BY Mental_hospitals DESC
LIMIT 5;
""")

records = cursor.fetchall()
for record in records:
    print(record)
```

```
('Japan', 8.314)
('Colombia', 1.988)
('Seychelles', 1.067)
('Antigua and Barbuda', 1.001)
('Argentina', 0.937)
```

```
#TOP 5 countries in  health_units
cursor.execute("""
SELECT DISTINCT Country,  health_units
FROM Facilities
JOIN Countries ON Facilities.Country_ID = Countries.Country_ID
ORDER BY health_units DESC
LIMIT 5;
""")

records = cursor.fetchall()
for record in records:
    print(record)
```

```
('Iceland', 4.542)
('Hungary', 2.647)
```

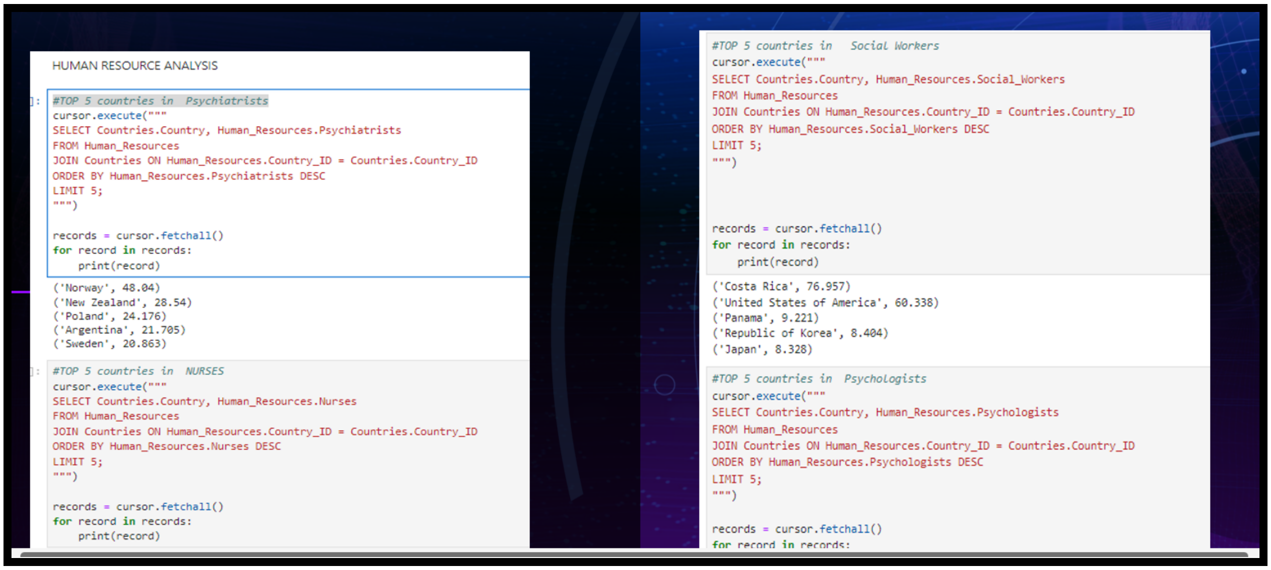
```
#TOP 5 countries in  outpatient_facilities
cursor.execute("""
SELECT DISTINCT Country,  outpatient_facilities
FROM Facilities
JOIN Countries ON Facilities.Country_ID = Countries.Country_ID
ORDER BY outpatient_facilities DESC
LIMIT 5;
""")

records = cursor.fetchall()
for record in records:
    print(record)
```

```
('Saint Lucia', 19.751)
('Estonia', 14.82)
('Samoa', 12.387)
('Grenada', 10.297)
('Hungary', 8.269)
```

```
#TOP 5 countries in  day_treatment
cursor.execute("""
SELECT DISTINCT Country,  day_treatment
FROM Facilities
JOIN Countries ON Facilities.Country_ID = Countries.Country_ID
ORDER BY day_treatment DESC
LIMIT 5;
""")

records = cursor.fetchall()
for record in records:
    print(record)
```

```
('Estonia', 17.176)
```

In the *human resources* table, a strong positive correlation was observed between social workers and psychologists (r = 0.76). Furthermore, there was a positive correlation between social workers and psychiatrists (r = 0.64). A reason could be that larger investments into mental healthcare allow for hiring more psychiatrists to support diagnosis and treatment planning along with social workers to provide counseling. The positive correlations indicate that as capacity expands in one area, there are greater resources to hire the other.



HUMAN RESOURCE ANALYSIS

```
#TOP 5 countries in  Psychiatrists
cursor.execute("""
SELECT Countries.Country, Human_Resources.Psychiatrists
FROM Human_Resources
JOIN Countries ON Human_Resources.Country_ID = Countries.Country_ID
ORDER BY Human_Resources.Psychiatrists DESC
LIMIT 5;
""")

records = cursor.fetchall()
for record in records:
    print(record)
```

```
('Norway', 48.04)
('New Zealand', 28.54)
('Poland', 24.176)
('Argentina', 21.705)
('Sweden', 20.863)
```

```
#TOP 5 countries in  NURSES
cursor.execute("""
SELECT Countries.Country, Human_Resources.Nurses
FROM Human_Resources
JOIN Countries ON Human_Resources.Country_ID = Countries.Country_ID
ORDER BY Human_Resources.Nurses DESC
LIMIT 5;
""")

records = cursor.fetchall()
for record in records:
    print(record)
```

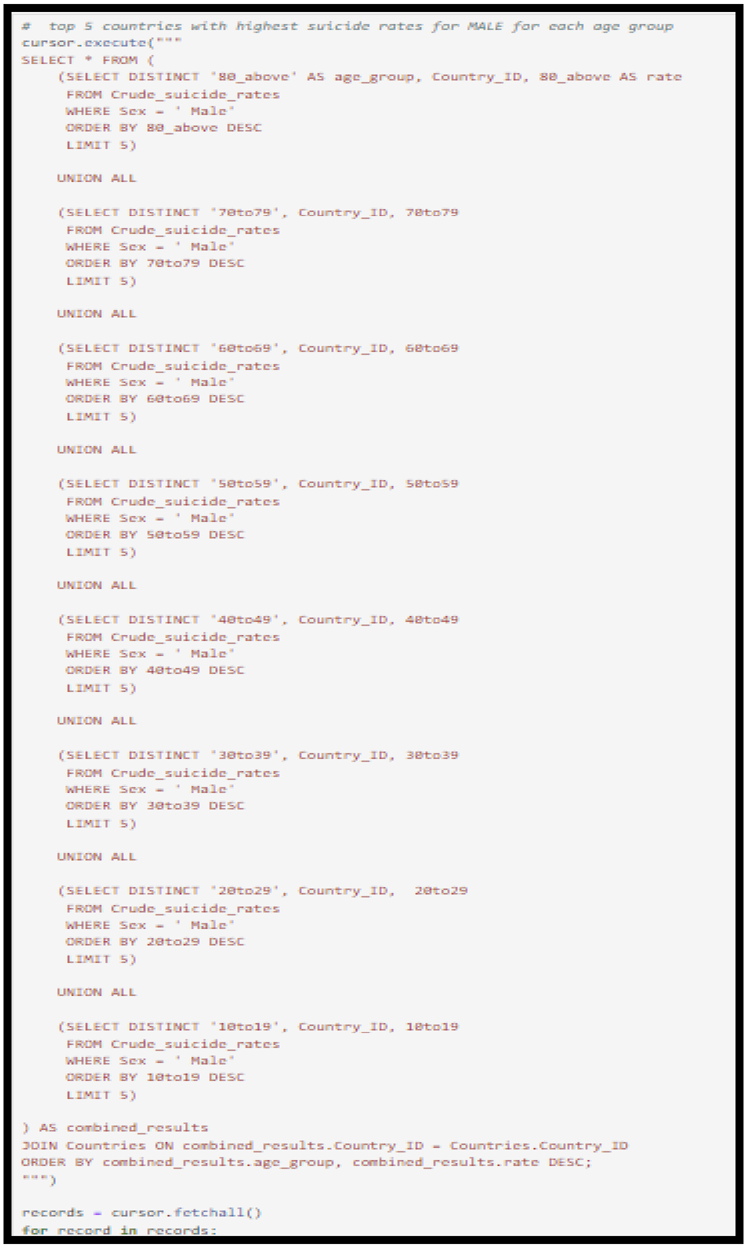
```
#TOP 5 countries in   Social Workers
cursor.execute("""
SELECT Countries.Country, Human_Resources.Social_Workers
FROM Human_Resources
JOIN Countries ON Human_Resources.Country_ID = Countries.Country_ID
ORDER BY Human_Resources.Social_Workers DESC
LIMIT 5;
""")



records = cursor.fetchall()
for record in records:
    print(record)
```

```
('Costa Rica', 76.957)
('United States of America', 60.338)
('Panama', 9.221)
('Republic of Korea', 8.404)
('Japan', 8.328)
```

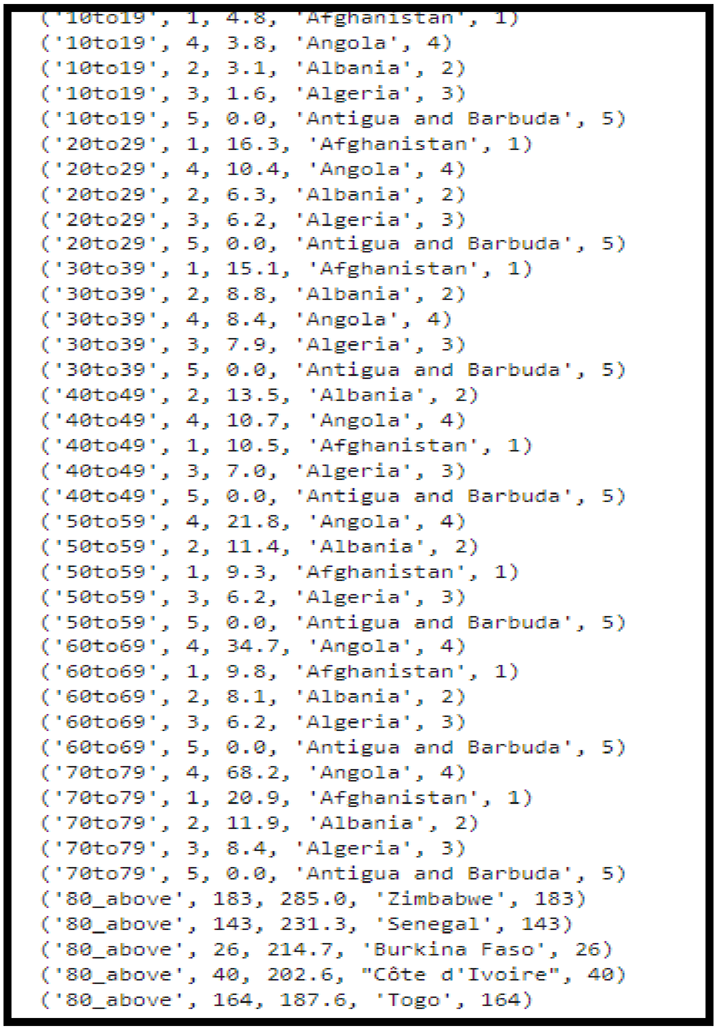
```
#TOP 5 countries in  Psychologists
cursor.execute("""
SELECT Countries.Country, Human_Resources.Psychologists
FROM Human_Resources
JOIN Countries ON Human_Resources.Country_ID = Countries.Country_ID
ORDER BY Human_Resources.Psychologists DESC
LIMIT 5;
""")

records = cursor.fetchall()
for record in records:
```

Line Charts

Line charts were used to visualize the trends among the various age groups. As per Fig. A2 *(see the appendix)*, globally, the trends indicate that suicide rates tend to increase with age from the 10-19 age group to the 80 above age group gradually in an ascending fashion, with the average suicide rate having risen around 5 in the 10-19 age group to around 40 in the 80 above age group. Furthermore, we can also infer that male suicide rates are significantly higher than female suicide rates across all age groups.



```
#  top 5 countries with highest suicide rates for MALE for each age group
cursor.execute("""
SELECT * FROM (
    (SELECT DISTINCT '80_above' AS age_group, Country_ID, 80_above AS rate
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 80_above DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '70to79', Country_ID, 70to79
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 70to79 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '60to69', Country_ID, 60to69
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 60to69 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '50to59', Country_ID, 50to59
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 50to59 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '40to49', Country_ID, 40to49
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 40to49 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '30to39', Country_ID, 30to39
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 30to39 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '20to29', Country_ID,  20to29
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 20to29 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '10to19', Country_ID, 10to19
     FROM Crude_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY 10to19 DESC
     LIMIT 5)

) AS combined_results
JOIN Countries ON combined_results.Country_ID = Countries.Country_ID
ORDER BY combined_results.age_group, combined_results.rate DESC;
""")

records = cursor.fetchall()
for record in records:
```



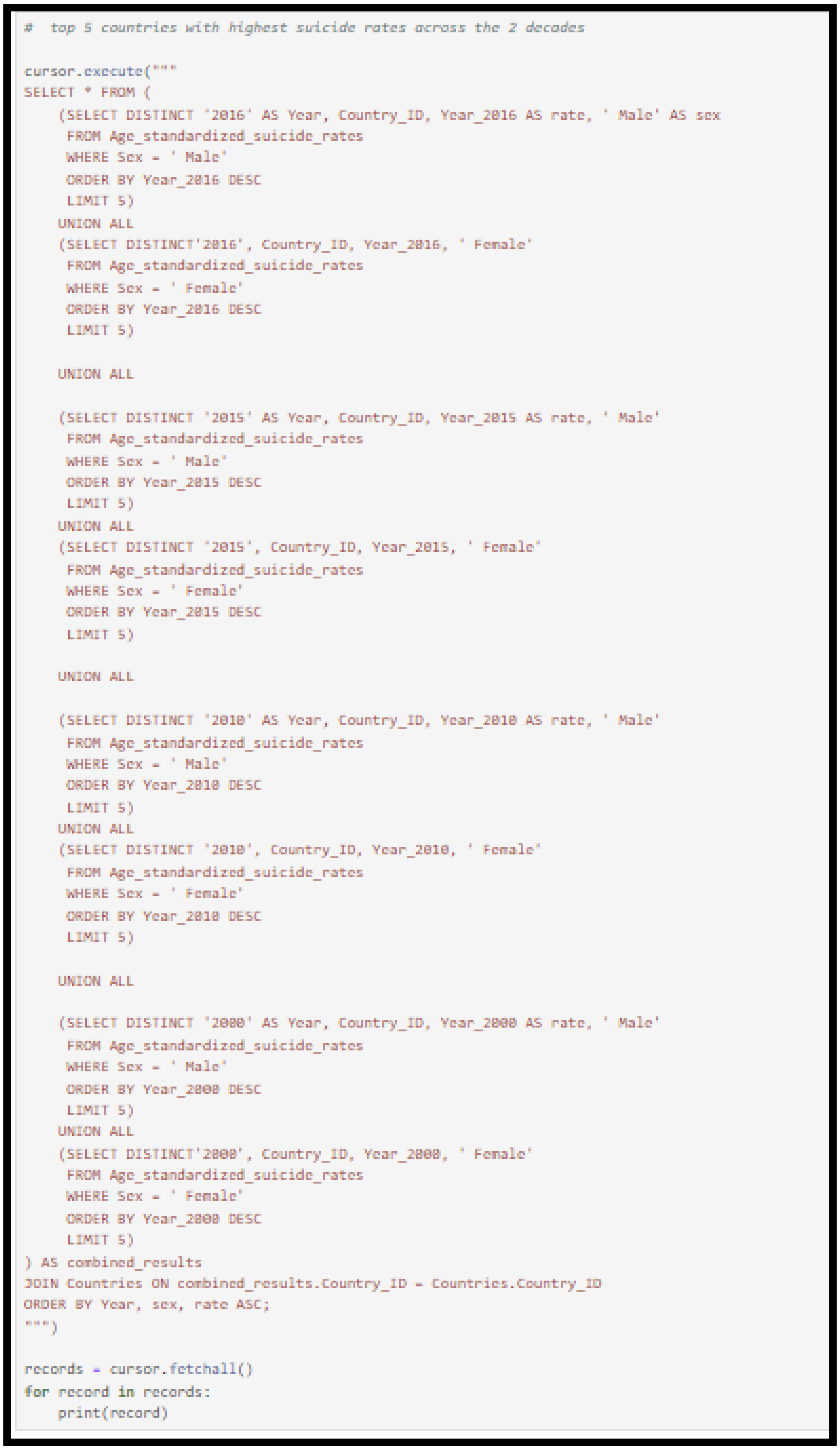
```
('10to19', 1, 4.8, 'Afghanistan', 1)
('10to19', 4, 3.8, 'Angola', 4)
('10to19', 2, 3.1, 'Albania', 2)
('10to19', 3, 1.6, 'Algeria', 3)
('10to19', 5, 0.0, 'Antigua and Barbuda', 5)
('20to29', 1, 16.3, 'Afghanistan', 1)
('20to29', 4, 10.4, 'Angola', 4)
('20to29', 2, 6.3, 'Albania', 2)
('20to29', 3, 6.2, 'Algeria', 3)
('20to29', 5, 0.0, 'Antigua and Barbuda', 5)
('30to39', 1, 15.1, 'Afghanistan', 1)
('30to39', 2, 8.8, 'Albania', 2)
('30to39', 4, 8.4, 'Angola', 4)
('30to39', 3, 7.9, 'Algeria', 3)
('30to39', 5, 0.0, 'Antigua and Barbuda', 5)
('40to49', 2, 13.5, 'Albania', 2)
('40to49', 4, 10.7, 'Angola', 4)
('40to49', 1, 10.5, 'Afghanistan', 1)
('40to49', 3, 7.0, 'Algeria', 3)
('40to49', 5, 0.0, 'Antigua and Barbuda', 5)
('50to59', 4, 21.8, 'Angola', 4)
('50to59', 2, 11.4, 'Albania', 2)
('50to59', 1, 9.3, 'Afghanistan', 1)
('50to59', 3, 6.2, 'Algeria', 3)
('50to59', 5, 0.0, 'Antigua and Barbuda', 5)
('60to69', 4, 34.7, 'Angola', 4)
('60to69', 1, 9.8, 'Afghanistan', 1)
('60to69', 2, 8.1, 'Albania', 2)
('60to69', 3, 6.2, 'Algeria', 3)
('60to69', 5, 0.0, 'Antigua and Barbuda', 5)
('70to79', 4, 68.2, 'Angola', 4)
('70to79', 1, 20.9, 'Afghanistan', 1)
('70to79', 2, 11.9, 'Albania', 2)
('70to79', 3, 8.4, 'Algeria', 3)
('70to79', 5, 0.0, 'Antigua and Barbuda', 5)
('80_above', 183, 285.0, 'Zimbabwe', 183)
('80_above', 143, 231.3, 'Senegal', 143)
('80_above', 26, 214.7, 'Burkina Faso', 26)
('80_above', 40, 202.6, "Côte d'Ivoire", 40)
('80_above', 164, 187.6, 'Togo', 164)
```

Pie Charts

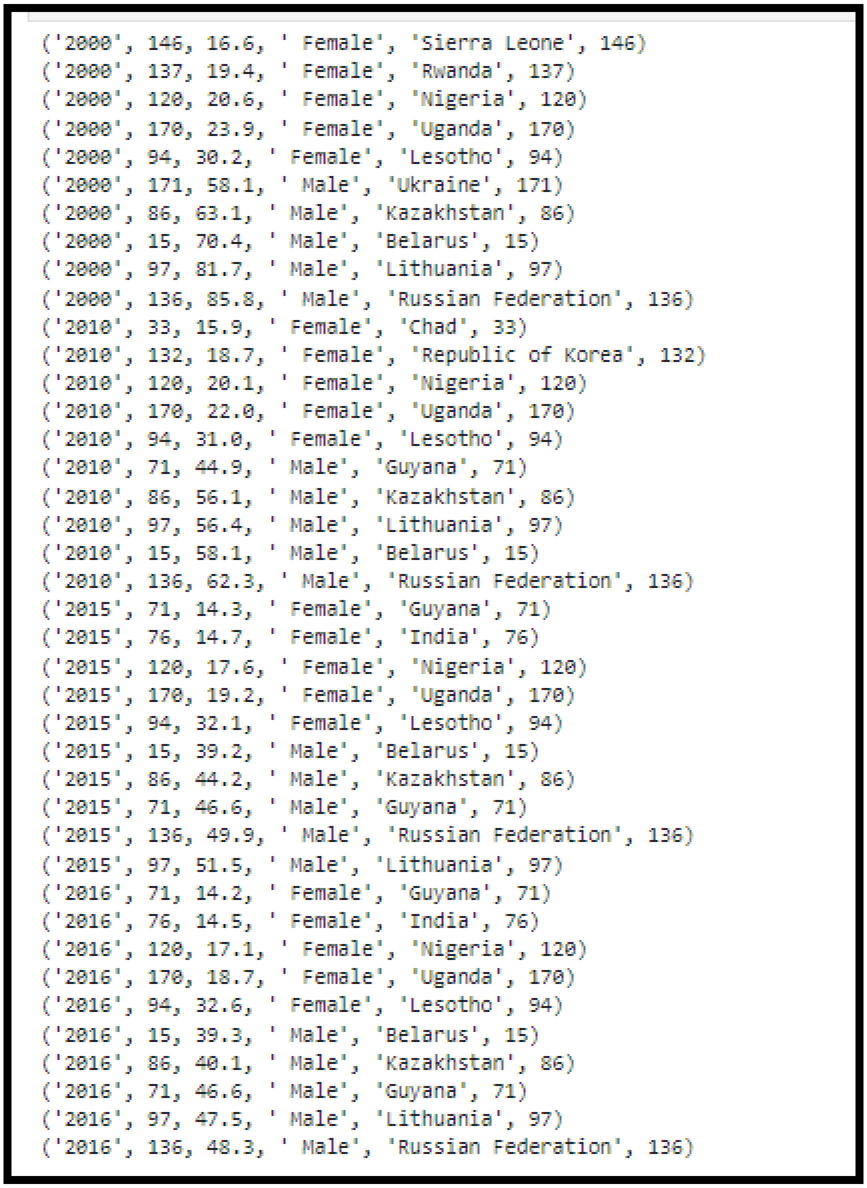
The pie charts in Fig. A3 *(see the appendix)* show that the distribution between the male and female suicide rates remains relatively stable over the fifteen years. It also shows that males account for almost 75% of suicides.

Box Plots

The box plots in Fig. A4 *(see the appendix)* help visualize the distribution of overall suicide rates from the different available years. We observe and see a slight gradual decline in average global suicide rates from 2000 to 2016.



```
#  top 5 countries with highest suicide rates across the 2 decades

cursor.execute("""
SELECT * FROM (
    (SELECT DISTINCT '2016' AS Year, Country_ID, Year_2016 AS rate, ' Male' AS sex
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2016 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT'2016', Country_ID, Year_2016, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2016 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '2015' AS Year, Country_ID, Year_2015 AS rate, ' Male'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2015 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT '2015', Country_ID, Year_2015, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2015 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '2010' AS Year, Country_ID, Year_2010 AS rate, ' Male'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2010 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT '2010', Country_ID, Year_2010, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2010 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '2000' AS Year, Country_ID, Year_2000 AS rate, ' Male'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2000 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT'2000', Country_ID, Year_2000, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2000 DESC
     LIMIT 5)
) AS combined_results
JOIN Countries ON combined_results.Country_ID = Countries.Country_ID
ORDER BY Year, sex, rate ASC;
""")

records = cursor.fetchall()
for record in records:
    print(record)
```



```
('2000', 146, 16.6, ' Female', 'Sierra Leone', 146)
('2000', 137, 19.4, ' Female', 'Rwanda', 137)
('2000', 120, 20.6, ' Female', 'Nigeria', 120)
('2000', 170, 23.9, ' Female', 'Uganda', 170)
('2000', 94, 30.2, ' Female', 'Lesotho', 94)
('2000', 171, 58.1, ' Male', 'Ukraine', 171)
('2000', 86, 63.1, ' Male', 'Kazakhstan', 86)
('2000', 15, 70.4, ' Male', 'Belarus', 15)
('2000', 97, 81.7, ' Male', 'Lithuania', 97)
('2000', 136, 85.8, ' Male', 'Russian Federation', 136)
('2010', 33, 15.9, ' Female', 'Chad', 33)
('2010', 132, 18.7, ' Female', 'Republic of Korea', 132)
('2010', 120, 20.1, ' Female', 'Nigeria', 120)
('2010', 170, 22.0, ' Female', 'Uganda', 170)
('2010', 94, 31.0, ' Female', 'Lesotho', 94)
('2010', 71, 44.9, ' Male', 'Guyana', 71)
('2010', 86, 56.1, ' Male', 'Kazakhstan', 86)
('2010', 97, 56.4, ' Male', 'Lithuania', 97)
('2010', 15, 58.1, ' Male', 'Belarus', 15)
('2010', 136, 62.3, ' Male', 'Russian Federation', 136)
('2015', 71, 14.3, ' Female', 'Guyana', 71)
('2015', 76, 14.7, ' Female', 'India', 76)
('2015', 120, 17.6, ' Female', 'Nigeria', 120)
('2015', 170, 19.2, ' Female', 'Uganda', 170)
('2015', 94, 32.1, ' Female', 'Lesotho', 94)
('2015', 15, 39.2, ' Male', 'Belarus', 15)
('2015', 86, 44.2, ' Male', 'Kazakhstan', 86)
('2015', 71, 46.6, ' Male', 'Guyana', 71)
('2015', 136, 49.9, ' Male', 'Russian Federation', 136)
('2015', 97, 51.5, ' Male', 'Lithuania', 97)
('2016', 71, 14.2, ' Female', 'Guyana', 71)
('2016', 76, 14.5, ' Female', 'India', 76)
('2016', 120, 17.1, ' Female', 'Nigeria', 120)
('2016', 170, 18.7, ' Female', 'Uganda', 170)
('2016', 94, 32.6, ' Female', 'Lesotho', 94)
('2016', 15, 39.3, ' Male', 'Belarus', 15)
('2016', 86, 40.1, ' Male', 'Kazakhstan', 86)
('2016', 71, 46.6, ' Male', 'Guyana', 71)
('2016', 97, 47.5, ' Male', 'Lithuania', 97)
('2016', 136, 48.3, ' Male', 'Russian Federation', 136)
```

Elective Questions

Q1: Demonstrate how you transformed a complex dataset into a meaningful visualization, including the underlying SQL query.

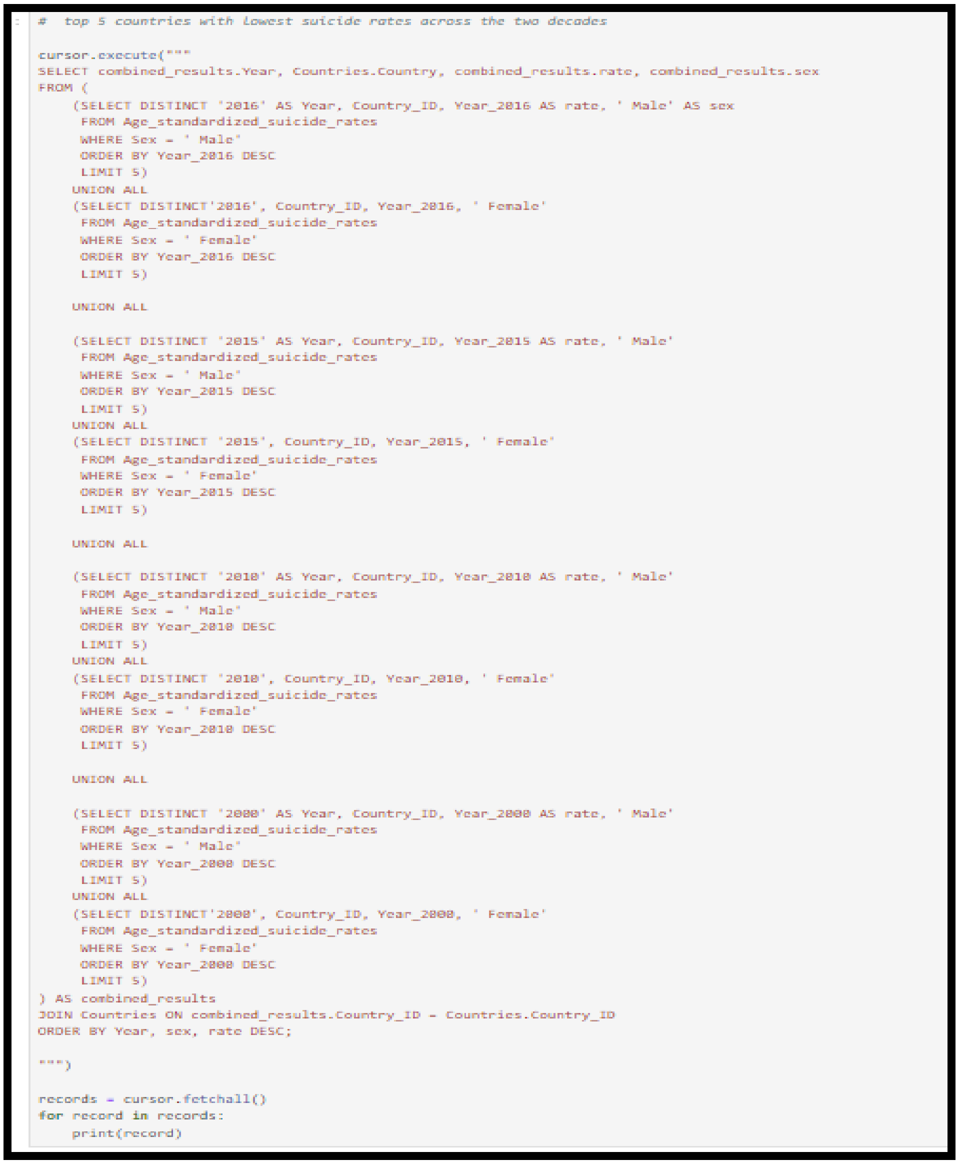
One of the key variables in my analysis was the *Age_standardized_sucide_rates* dataset as it contained suicide rates for males and females aggregated across all age groups for several years.

Visualizing this complex multivariate data allowed me to observe the trends and patterns over time more clearly.

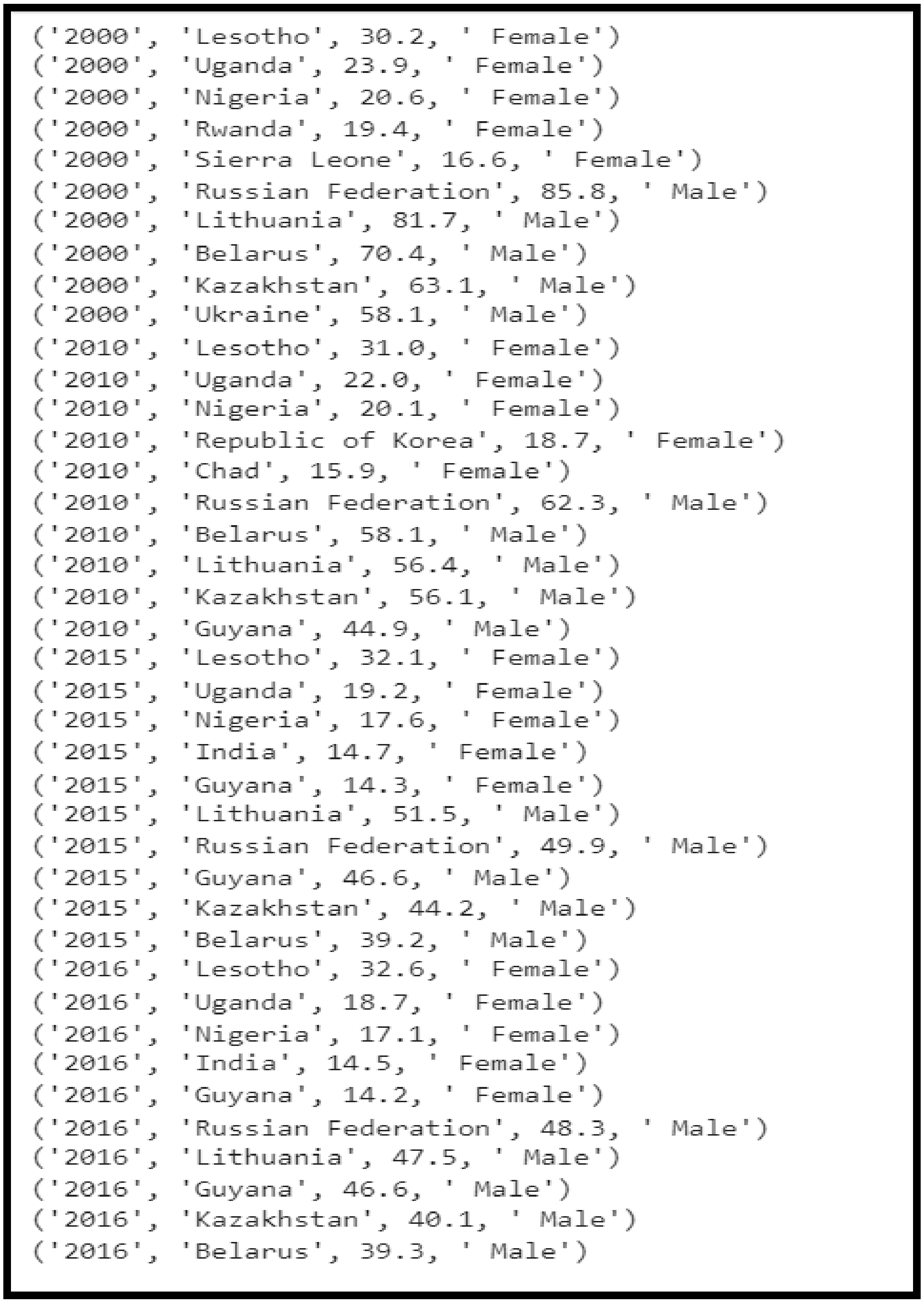
To transform this dataset into a meaningful visualization, I decided to create a box plot, as it excels in depicting distributions of numeric data like suicide rates.

The resulting box plot beautifully summarizes the distribution of global suicide rates over 4 years, with the median global suicide rates showing a gradual declining trend from 2000 to 2016. The interquartile range becoming smaller over time indicates lower variance and improved consistency globally among the countries.

Hence, through this transformation, I could surface the key trend of declining suicide rates over time in a very intuitive and easy-to-grasp visual chart.



```
#  top 5 countries with lowest suicide rates across the two decades

cursor.execute("""
SELECT combined_results.Year, Countries.Country, combined_results.rate, combined_results.sex
FROM (
    (SELECT DISTINCT '2016' AS Year, Country_ID, Year_2016 AS rate, ' Male' AS sex
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2016 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT'2016', Country_ID, Year_2016, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2016 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '2015' AS Year, Country_ID, Year_2015 AS rate, ' Male'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2015 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT '2015', Country_ID, Year_2015, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2015 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '2010' AS Year, Country_ID, Year_2010 AS rate, ' Male'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2010 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT '2010', Country_ID, Year_2010, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2010 DESC
     LIMIT 5)

    UNION ALL

    (SELECT DISTINCT '2000' AS Year, Country_ID, Year_2000 AS rate, ' Male'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Male'
     ORDER BY Year_2000 DESC
     LIMIT 5)
    UNION ALL
    (SELECT DISTINCT'2000', Country_ID, Year_2000, ' Female'
     FROM Age_standardized_suicide_rates
     WHERE Sex = ' Female'
     ORDER BY Year_2000 DESC
     LIMIT 5)
) AS combined_results
JOIN Countries ON combined_results.Country_ID = Countries.Country_ID
ORDER BY Year, sex, rate DESC;

""")

records = cursor.fetchall()
for record in records:
    print(record)
```



```
('2000', 'Lesotho', 30.2, ' Female')
('2000', 'Uganda', 23.9, ' Female')
('2000', 'Nigeria', 20.6, ' Female')
('2000', 'Rwanda', 19.4, ' Female')
('2000', 'Sierra Leone', 16.6, ' Female')
('2000', 'Russian Federation', 85.8, ' Male')
('2000', 'Lithuania', 81.7, ' Male')
('2000', 'Belarus', 70.4, ' Male')
('2000', 'Kazakhstan', 63.1, ' Male')
('2000', 'Ukraine', 58.1, ' Male')
('2010', 'Lesotho', 31.0, ' Female')
('2010', 'Uganda', 22.0, ' Female')
('2010', 'Nigeria', 20.1, ' Female')
('2010', 'Republic of Korea', 18.7, ' Female')
('2010', 'Chad', 15.9, ' Female')
('2010', 'Russian Federation', 62.3, ' Male')
('2010', 'Belarus', 58.1, ' Male')
('2010', 'Lithuania', 56.4, ' Male')
('2010', 'Kazakhstan', 56.1, ' Male')
('2010', 'Guyana', 44.9, ' Male')
('2015', 'Lesotho', 32.1, ' Female')
('2015', 'Uganda', 19.2, ' Female')
('2015', 'Nigeria', 17.6, ' Female')
('2015', 'India', 14.7, ' Female')
('2015', 'Guyana', 14.3, ' Female')
('2015', 'Lithuania', 51.5, ' Male')
('2015', 'Russian Federation', 49.9, ' Male')
('2015', 'Guyana', 46.6, ' Male')
('2015', 'Kazakhstan', 44.2, ' Male')
('2015', 'Belarus', 39.2, ' Male')
('2016', 'Lesotho', 32.6, ' Female')
('2016', 'Uganda', 18.7, ' Female')
('2016', 'Nigeria', 17.1, ' Female')
('2016', 'India', 14.5, ' Female')
('2016', 'Guyana', 14.2, ' Female')
('2016', 'Russian Federation', 48.3, ' Male')
('2016', 'Lithuania', 47.5, ' Male')
('2016', 'Guyana', 46.6, ' Male')
('2016', 'Kazakhstan', 40.1, ' Male')
('2016', 'Belarus', 39.3, ' Male')
```

Q3: Discuss any challenges faced while visualizing large datasets and how you overcame them.

There were some minor challenges that I had to overcome for the visualizations I created for this project. The major problems I encountered were with the line charts and pie charts.

Line Charts: The major challenge with plotting multiple line charts with a large underlying dataset resulted in slow render times. To overcome this, sampled using a filtered version with a smaller subset i.e., including only Male and Female data as the Both sexes column contained only the average and slowed the rendered process along with increasing redundancy. This fine-tuning helped in optimizing the visualization and gathering more meaningful insights.

Pie Charts: While trying to visualize Pie Charts, the high cardinality of country dimensions bloated the dataset size. This was mitigated by pre-aggregating the data by sex and the year fields to avoid handling the full granular dataset and aggregating many countries on the fly.

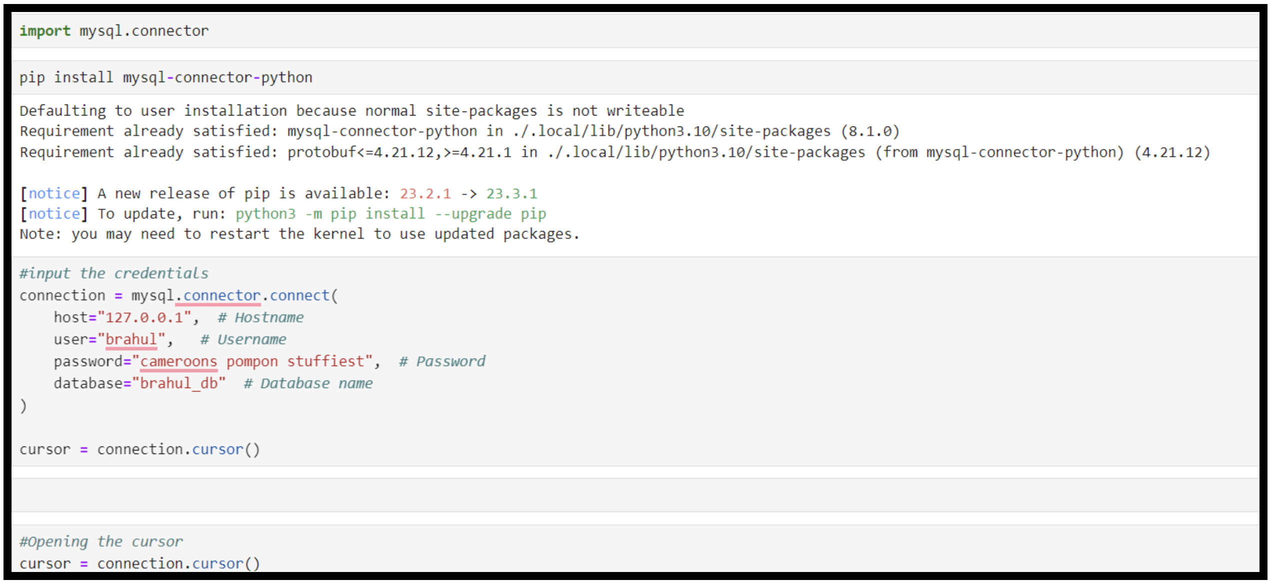
To summarize, some common optimization strategies I employed were sampling data subsets and pre-aggregating data.

.**Q4: Describe how you visualized analytics results directly. Detail any portal setup or direct database connections.**

To rapidly visualize results during analysis, I leveraged Jupyter Notebook that helped connect directly to the database i.e., PHPmyadmin. This helped in converting results directly into Panda data frames for visualization using Matplotlib and Seaborn. Furthermore, it also helped iterate on visualizations without the added latency of moving data across tools.

This direct connection to the database included many benefits such as:-

- ❖ Speed and flexibility for analyzing data as needed i.e., on an ad-hoc basis.
- ❖ Direct access to live data in the database for visualizations
- ❖ Visual data validation during the development cycles



```
import mysql.connector
```

```
pip install mysql-connector-python
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: mysql-connector-python in ./.local/lib/python3.10/site-packages (8.1.0)
Requirement already satisfied: protobuf<=4.21.12,>=4.21.1 in ./.local/lib/python3.10/site-packages (from mysql-connector-python) (4.21.12)

[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python3 -m pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

```
#input the credentials
connection = mysql.connector.connect(
    host="127.0.0.1",  # Hostname
    user="brahul",   # Username
    password="cameroons pompon stuffiest",  # Password
    database="brahul_db"  # Database name
)

cursor = connection.cursor()
```

```
#Opening the cursor
cursor = connection.cursor()
```

Q5: Illustrate how an optimized database design could enhance your visualization capabilities.

An optimized database design can help reduce redundancy through effective normalization and contribute towards faster query execution. Furthermore, a well-structured schema reduces the need for complex joins and prioritizes scalability and performance as the dataset grows. Efficient data retrieval paths are also established as a result of good normalization and appropriate indexing.

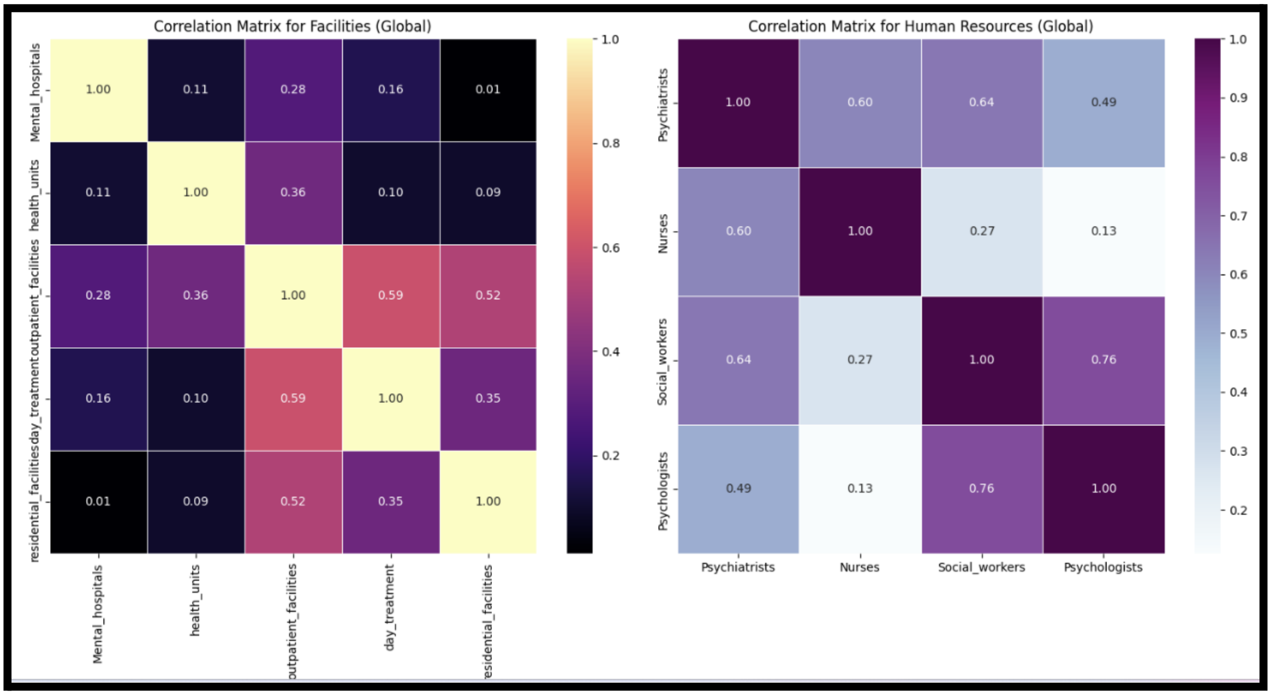
Impact on Visualization:-

- Normalization helps reduce the amount of data that needs to be processed, transmitted, and rendered, thus, improving efficiency
- Fine-tuning the queries helps enable faster generation of visualizations and analytics
- Simplified joins make it easier to retrieve data for visualization
- The streamlined data retrieval process ensures that visualization tools can quickly access the required data points
- Optimized storage design helps in faster data retrieval and transmission, making the rendering of visualizations more efficient

For the visualizations I used for this project, optimized data design helps in various ways:-

- Efficient data retrieval can help create *box plots* more quickly, as it aggregates data across multiple dimensions.
- Optimized data design ensures that data related to categories is instantly available for *pie charts*.
- Efficient relationships are crucial for visualizing trends over time, thus, helping support the creation of dynamic and responsive *line charts*.
- Normalization and proper indexing are essential to facilitate quick computation of *correlation matrices.*

APPENDIX



	Mental_hospitals	health_units	outpatient_facilities	day_treatment	residential_facilities
Mental_hospitals	1.00	0.11	0.28	0.16	0.01
health_units	0.11	1.00	0.36	0.10	0.09
outpatient_facilities	0.28	0.36	1.00	0.59	0.52
day_treatment	0.16	0.10	0.59	1.00	0.35
residential_facilities	0.01	0.09	0.52	0.35	1.00

	Psychiatrists	Nurses	Social_workers	Psychologists
Psychiatrists	1.00	0.60	0.64	0.49
Nurses	0.60	1.00	0.27	0.13
Social_workers	0.64	0.27	1.00	0.76
Psychologists	0.49	0.13	0.76	1.00

Fig. A1

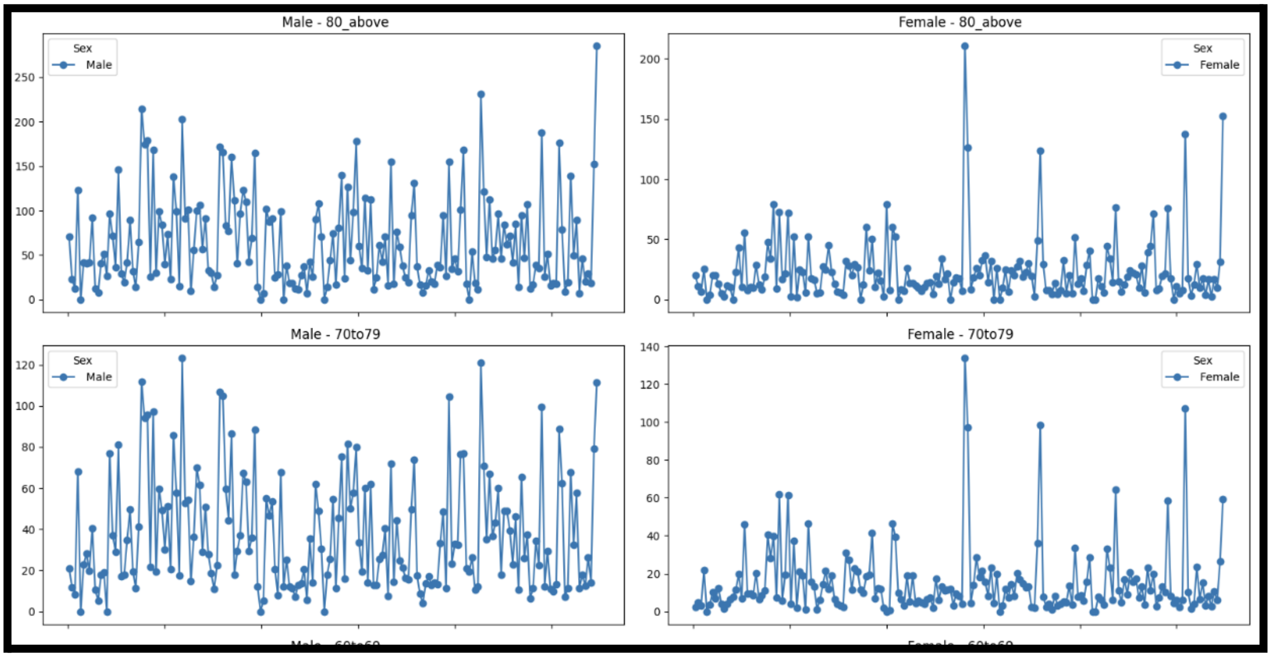


Fig. A2

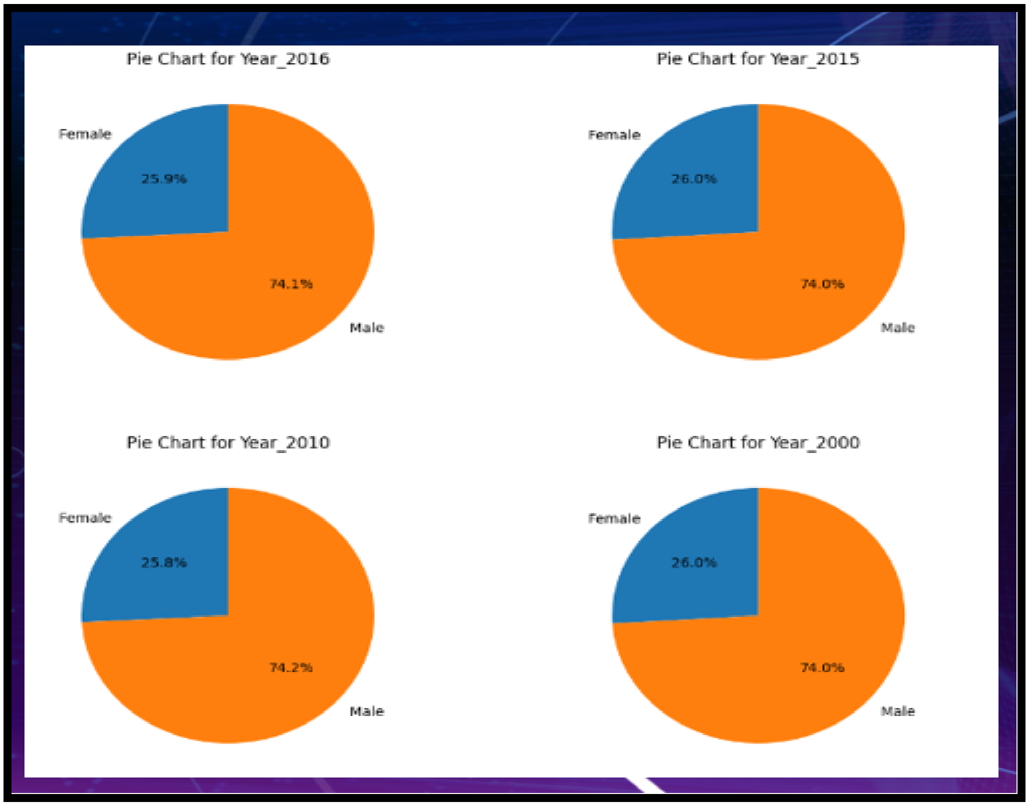

Pie Chart for Year_2016
Female
25.9%
74.1%
Male
Pie Chart for Year_2015
Female
26.0%
74.0%
Male
Pie Chart for Year_2010
Female
25.8%
74.2%
Male
Pie Chart for Year_2000
Female
26.0%
74.0%
Male

Fig. A3

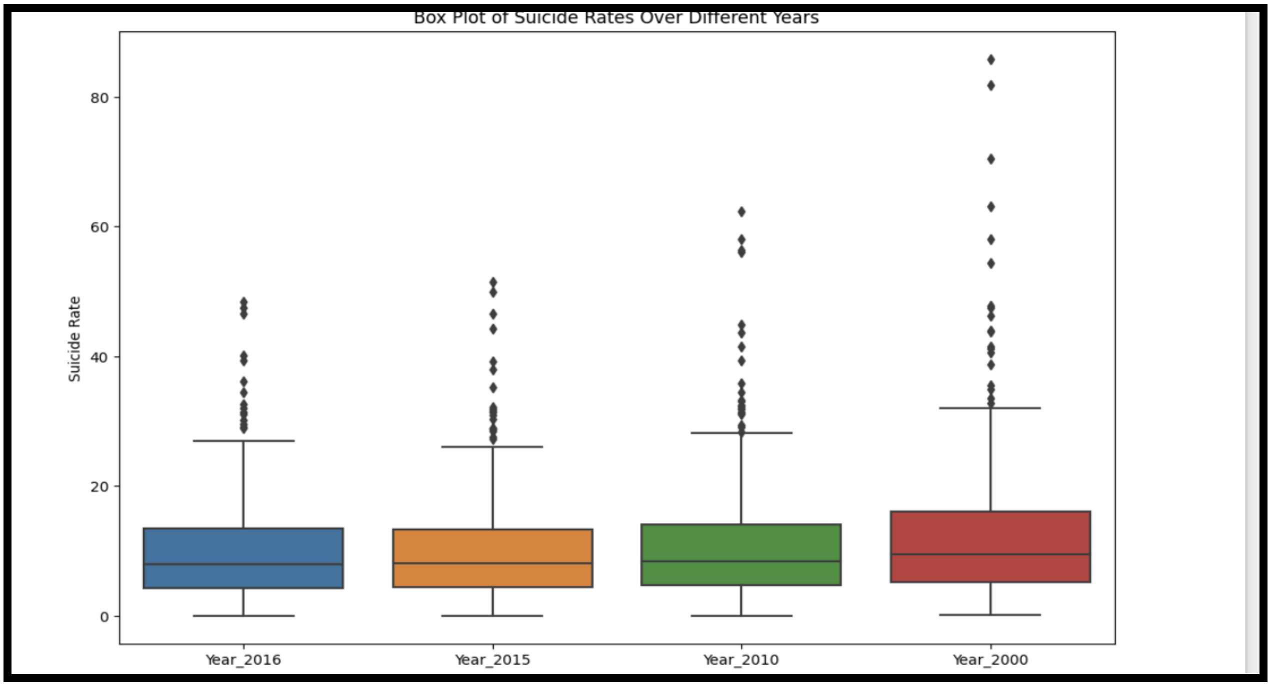

Box Plot of Suicide Rates Over Different Years
Suicide Rate
0
20
40
60
80
Year_2016
Year_2015
Year_2010
Year_2000

Fig. A4