Presented by OLUWAFEMI ENIOLA

WORLD CHOLERA CASES

Project

Introduction

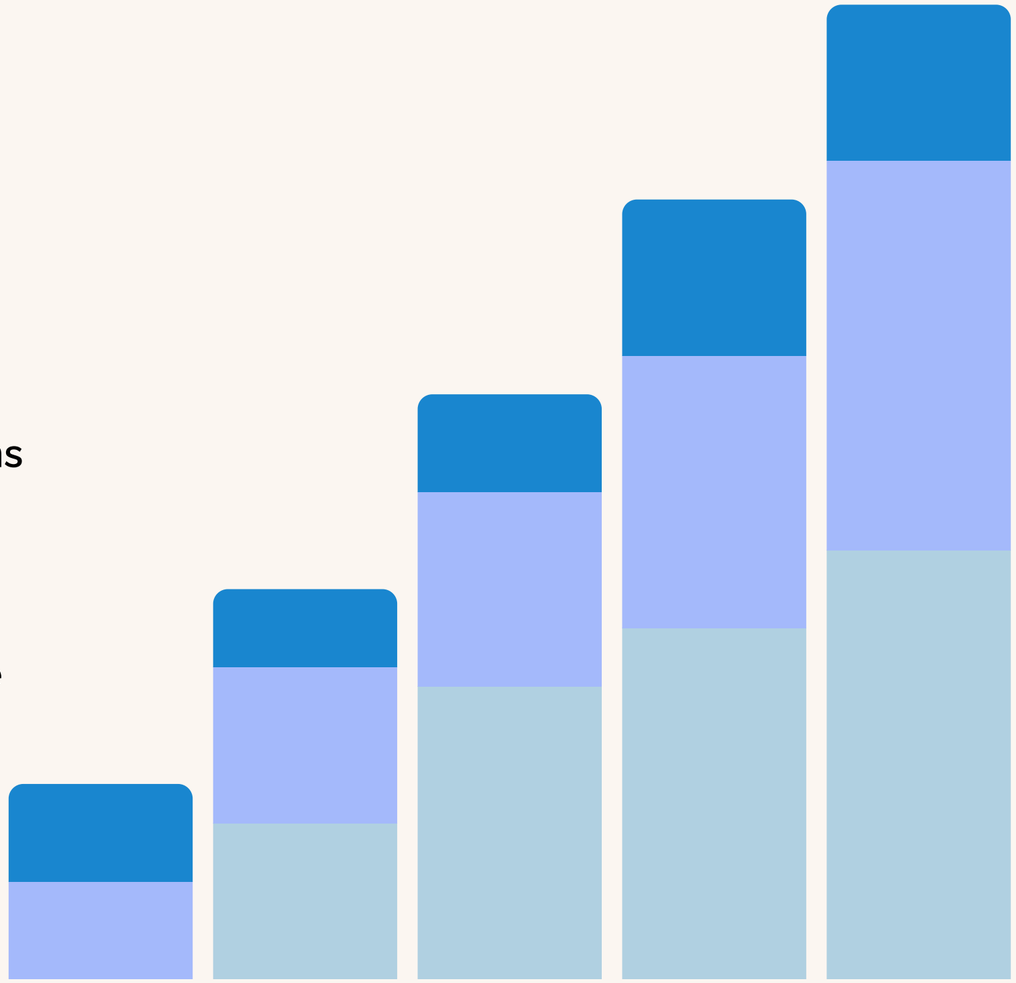
Cholera remains a critical global health challenge. This academic project analyzes a comprehensive dataset on cholera cases, along with environmental, socioeconomic, and health infrastructure factors. Using some data analysis tools, my goal is to uncover key trends, identify influencing factors, and provide actionable insights to better understand and combat cholera.

Methodology

My project followed a structured data analysis workflow to transform raw cholera data into actionable insights.

- Phase 1: Data Cleaning & Preparation
- Phase 2: Exploratory Data Analysis (EDA) & Initial Aggregations
- Phase 3: Data Visualization
- Phase 4: Insights & Reporting

This systematic approach ensured data integrity, comprehensive analysis, and effective communication of my findings to understand and combat cholera.



Data Cleaning & Preparation

Checking for missing values

```
[11]: cp.isnull().sum()

[11]: Country                                    0
      Region                                     0
      Year                                       0
      Water Source Type                          0
      Contaminant Level (ppm)                    0
      pH Level                                   0
      Turbidity (NTU)                            0
      Dissolved Oxygen (mg/L)                    0
      Nitrate Level (mg/L)                       0
      Lead Concentration (µg/L)                  0
      Bacteria Count (CFU/mL)                    0
      Water Treatment Method                   747
      Access to Clean Water (% of Population)    0
      Diarrheal Cases per 100,000 people         0
      Cholera Cases per 100,000 people           0
      Typhoid Cases per 100,000 people           0
      Infant Mortality Rate (per 1,000 live births)  0
```

investigating column with missing values

```
[13]: cp['Water Treatment Method']

[13]: 0       Filtration
      1          Boiling
      2              NaN
      3          Boiling
      4       Filtration
                 ...
      2995           NaN
      2996       Boiling
      2997       Boiling
      2998       Boiling
      2999       Boiling
      Name: Water Treatment Method, Length: 3000, dtype: object
```

Replacing missing values

```
cp.fillna({'Water Treatment Method':'No treatment'},inplace=True)

cp['Water Treatment Method']

0         Filtration
1            Boiling
2       No treatment
3            Boiling
4         Filtration
              ...
2995    No treatment
2996         Boiling
2997         Boiling
2998         Boiling
2999         Boiling
Name: Water Treatment Method, Length: 3000, dtype: object
```

```
[15]: cp.isnull().sum()

[15]: Country                                    0
      Region                                     0
      Year                                       0
      Water Source Type                          0
      Contaminant Level (ppm)                    0
      pH Level                                   0
      Turbidity (NTU)                            0
      Dissolved Oxygen (mg/L)                    0
      Nitrate Level (mg/L)                       0
      Lead Concentration (µg/L)                  0
      Bacteria Count (CFU/mL)                    0
      Water Treatment Method                     0
      Access to Clean Water (% of Population)    0
      Diarrheal Cases per 100,000 people         0
      Cholera Cases per 100,000 people           0
      Typhoid Cases per 100,000 people           0
      Infant Mortality Rate (per 1,000 live births)  0
      GDP per Capita (USD)                       0
      Healthcare Access Index (0-100)            0
      Urbanization Rate (%)                      0
      Sanitation Coverage (% of Population)      0
```

Result

Changing data type

```
Water Treatment Method                              object
Access to Clean Water (% of Population)            float64
Diarrheal Cases per 100,000 people                   int64
Cholera Cases per 100,000 people                     int64
Typhoid Cases per 100,000 people                     int64
Infant Mortality Rate (per 1,000 live births)      float64
GDP per Capita (USD)                                 int64
Healthcare Access Index (0-100)                    float64
Urbanization Rate (%)                              float64
Sanitation Coverage (% of Population)              float64
Rainfall (mm per year)                               int64
```

```
[6]: cp['Access to Clean Water (% of Population)']=cp['Access to Clean Water (% of Population)'].astype('int64')
     cp['Urbanization Rate (%)']=cp['Urbanization Rate (%)'].astype('int64')
     cp['Sanitation Coverage (% of Population)']=cp['Sanitation Coverage (% of Population)'].astype('int64')

[7]: cp.dtypes
```

```
Access to Clean Water (% of Population)            int64
Diarrheal Cases per 100,000 people                 int64
Cholera Cases per 100,000 people                   int64
Typhoid Cases per 100,000 people                   int64
Infant Mortality Rate (per 1,000 live births)    float64
GDP per Capita (USD)                               int64
Healthcare Access Index (0-100)                  float64
Urbanization Rate (%)                              int64
Sanitation Coverage (% of Population)              int64
```

Checking for inconsistencies

```
[16]: cp['Country'].unique()

[16]: array(['Mexico', 'Brazil', 'Indonesia', 'Nigeria', 'Ethiopia', 'China',
             'Bangladesh', 'India', 'USA', 'Pakistan'], dtype=object)

[17]: cp['Year'].unique()

[17]: array([2015, 2017, 2022, 2016, 2005, 2013, 2024, 2014, 2023, 2021, 2002,
             2018, 2011, 2001, 2000, 2019, 2020, 2004, 2007, 2003, 2009, 2010,
             2006, 2012, 2008], dtype=int64)
```

Exploratory Data Analysis (EDA)

Brief summary

```
[13]: cp.select_dtypes(include='object').describe()
```

[13]:

	Country	Region	Water Source Type	Water Treatment Method
count	3000	3000	3000	3000
unique	10	5	6	4
top	USA	East	River	Boiling
freq	319	625	538	777

```
[10]: cp.describe()
```

[10]:

	Year	Contaminant Level (ppm)	pH Level	Turbidity (NTU)	Dissolved Oxygen (mg/L)	Nitrate Level (mg/L)	Lead Concentration (µg/L)	Bacteria Count (CFU/mL)
count	3000.000000	3000.000000	3000.000000	3000.000000	3000.000000	3000.00000	3000.000000	3000.000000
mean	2012.012667	4.954390	7.255847	2.480023	6.492850	25.08025	10.047913	2488.477333
std	7.229287	2.860072	0.720464	1.419984	2.027966	14.50517	5.798238	1431.421553
min	2000.000000	0.000000	6.000000	0.000000	3.000000	0.05000	0.000000	0.000000
25%	2006.000000	2.560000	6.630000	1.257500	4.710000	12.52500	5.120000	1268.000000
50%	2012.000000	4.950000	7.280000	2.460000	6.490000	24.79000	10.065000	2469.000000
75%	2018.000000	7.400000	7.870000	3.660000	8.252500	37.91000	15.032500	3736.250000
max	2024.000000	10.000000	8.500000	4.990000	10.000000	49.99000	20.000000	4998.000000

Reviewing factors responsible for Cholera Outbreaks by Country

```
[7]: ccases=cp[cp['Year'].between(2019, 2024)].groupby('Country')
        .agg(avg_cases_per100k=('Cholera Cases per 100,000 people','mean'),
        Bacteria_count=('Bacteria Count (CFU/mL)','mean'),Turbidity=('Turbidity (NTU)','mean'),
        san_coverage_percent=('Sanitation Coverage (% of Population)','mean'),
        Contamination_level=('Contaminant Level (ppm)','mean'))
        .round(2).sort_values(by='avg_cases_per100k',ascending=False).reset_index()
ccases
```

[7]:

	Country	avg_cases_per100k	Bacteria_count	Turbidity	san_coverage_percent	Contamination_level
0	USA	25.50	2463.61	2.50	62.95	4.59
1	India	25.49	2339.66	2.33	61.65	5.00
2	Nigeria	24.43	2328.16	2.37	56.60	4.83
3	China	24.18	2613.93	2.59	66.24	5.00
4	Ethiopia	23.20	2664.06	2.45	60.30	5.41
5	Indonesia	22.81	2685.01	2.32	56.06	4.49
6	Pakistan	22.71	2490.12	2.38	63.45	4.57
7	Brazil	22.53	2382.75	2.79	60.76	4.51
8	Bangladesh	22.44	2497.77	2.56	59.30	4.78
9	Mexico	21.42	2423.27	2.31	55.63	5.09

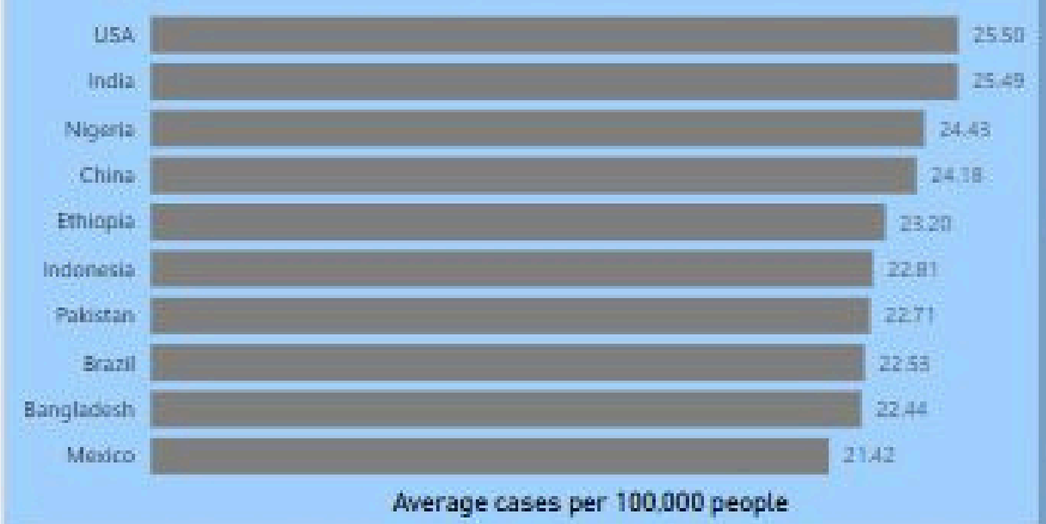
Data visualisation Insights and Recommendations

DASHBOARD

WORLD CHOLERA CASES

2019 - 2024

AVERAGE CASES PER 100K PEOPLE BY COUNTRY



AVG INFANT MORTALITY

51

RATE PER 1K LIVE BIRTHS

23

AVG CASES PER 100K PPL

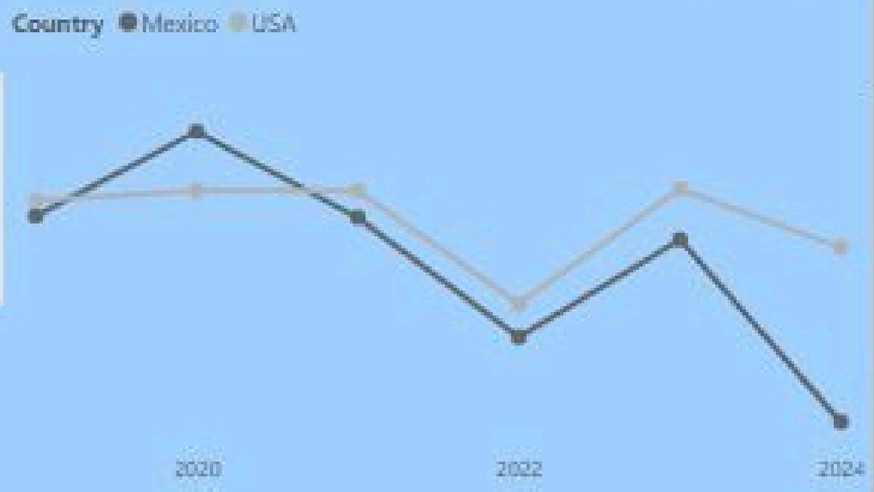
10

NO OF COUNTRIES

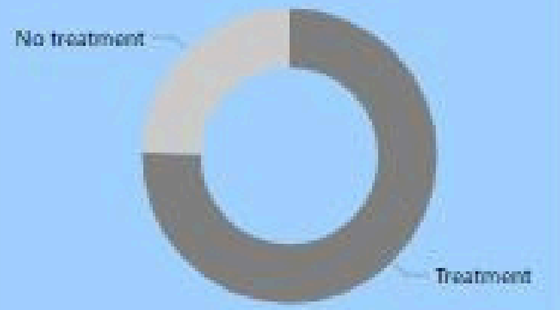
Year

All

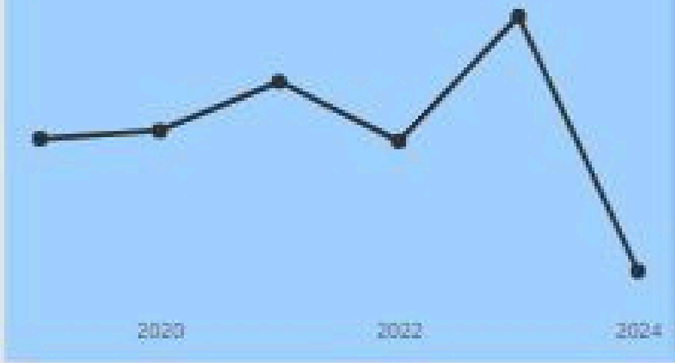
AVG CASES PER 100K PEOPLE BY COUNTRY/YEAR



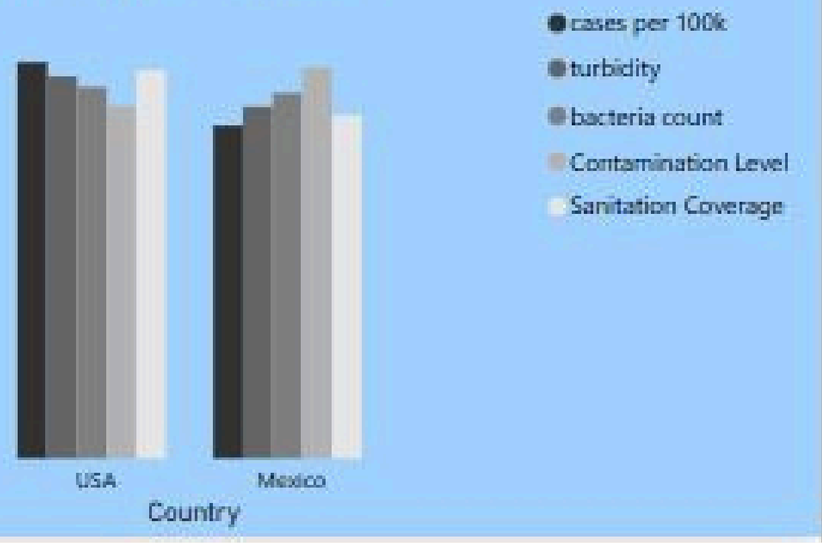
TREATMENT VS NO TREATMENT IN CHOLERA CASES



AVG INFANT MOTALITY RATE PER 1K LIVE BIRTHS BY YEAR

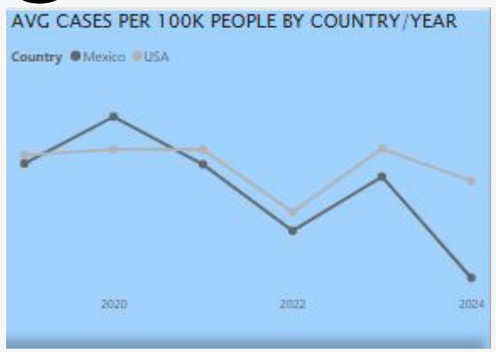


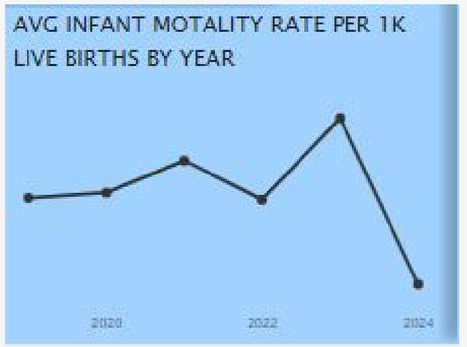
COUNTRY BY CHOLERA CAUSES



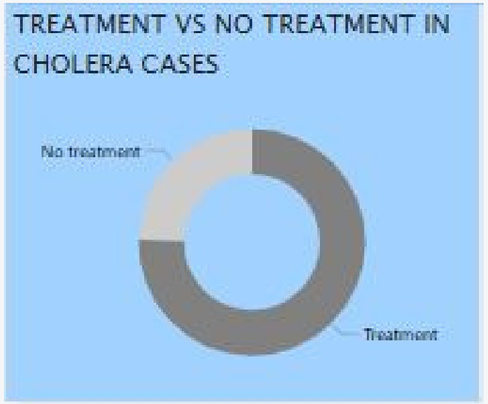
Insights

FROM OUR EDA, IT IS SEEN THAT IN THE LAST SIX YEARS INSPITE OF **USA** HAVING A HIGH SANITATION COVERAGE AND LOW CONTAMINATION LEVEL, IT CAN BE SAID THAT THE HIGH BACTERIA LEVEL AND TURBIDITY WERE DETERMINANTS TO BEING TOP OF THE LIST.

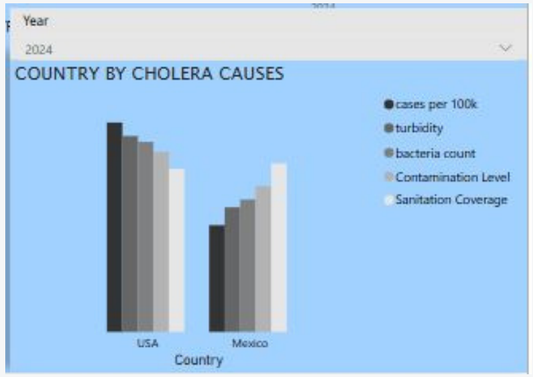


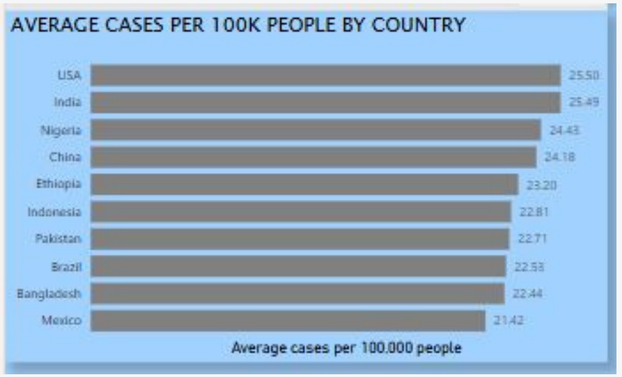


CHOLERA OUTBREAKS PLAYS A **SIGNIFICANT ROLE** IN INFANT MORTALITY RATE AS A DROP IN CHOLERA CASES IS ALSO A DROP IN INFANT MORTALITY RATE



ALSO IN **2024** THERE IS A GOOD AMOUNT OF DECREASE SEEN IN THE AVERAGE AMOUNT OF CASES IN BOTH COUNTRIES (USA AND MEXICO) WHICH IS DIRECTLY PROPORTIONAL TO THE FACTORS CAUSING CHOLERA





TREATED WATER STILL POSSESS A RISK OF CHOLERA OUTBREAK

Recommendations

- IMPLEMENTATION OF A SYSTEM THAT WILL EFFECTIVELY REDUCE BACTERIA COUNT AND TURBIDITY
- TACKLING CHOLERA OUTBREAKS EQUALS DECREASE IN INFANT MORTALITY RATE
- TREATMENT METHODS SHOULD BE REVIEWED
- STRATEGIES USED TO COMBAT CHOLERA OUTBREAK IN 2024 SHOULD BE REVIEWED AND IMPROVED ON IF POSSIBLE IN SUBSEQUENT YEARS