COVID-19 Global Data Analysis

Data Visualization and Storytelling Through Statistical Analysis

Student: Dumindu Thushan Abhayawickrama
Student ID: CL/BSCDS/CMU/09/79
Module: CIS5022 - Data Visualization and Storytelling
Programme: B.Sc. (Hons) in Data Science
Institution: ICBT Campus, Sri Lanka

Presentation Agenda

Introduction (3 mins)

- → Dataset background and context
- → Main research questions
- → Analysis objectives

Data & Methods (3 mins)

- → Data source and structure
- → Preprocessing approach
- → Tools and technologies

Key Insights (6 mins)

- → Pandemic wave patterns
- → Seasonal and geographic trends
- → Correlation analysis
- → Outlier impact assessment

Recommendations (3 mins)

- → Strategic takeaways
- → Actionable insights
- → Future implications

Research Questions & Theme

Main Research Question

"How do COVID-19 transmission patterns vary across countries and time, and what factors drive these differences?"

Sub-Questions for Analysis

- → What seasonal and temporal patterns emerge across different regions?
- → How strong is the relationship between cases and deaths across countries?
- → What role do outliers play in understanding pandemic dynamics?
- → Which variables most influence transmission patterns?

🎯 Analysis Theme: "Data-Driven Pandemic Intelligence"

Transforming raw COVID-19 data into actionable insights for public health decision-making through statistical analysis and pattern recognition.

Dataset Overview

14,610
Total Records

10
Countries Analyzed

4 Years
Time Period (2020-2023)

0%
Missing Data

Countries in Analysis

Key Variables

- → Daily new cases and cumulative totals
- → Daily deaths and cumulative mortality
- → Population and temporal features
- → Derived metrics: growth rates, seasonality indicators

Data Processing & Methodology

🔧 Data Cleaning

0% missing data validation, date formatting, duplicate removal

📊 Statistical Analysis

IQR outlier detection, correlation analysis, Z-score normalization

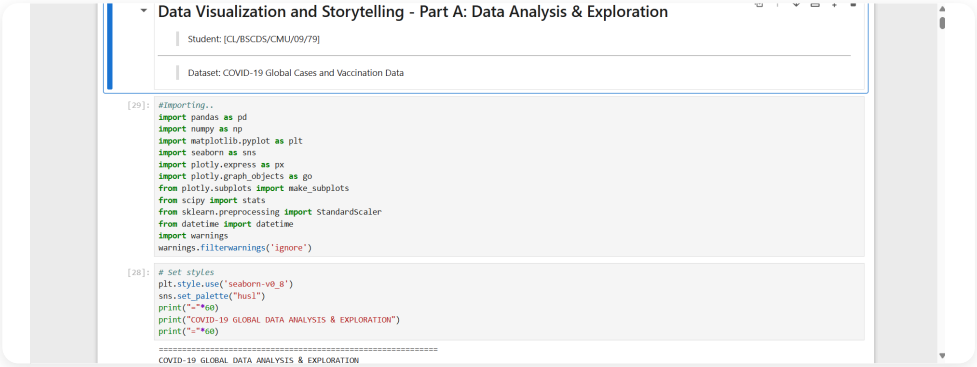
📈 Trend Analysis

7-day moving averages, seasonal decomposition, wave detection

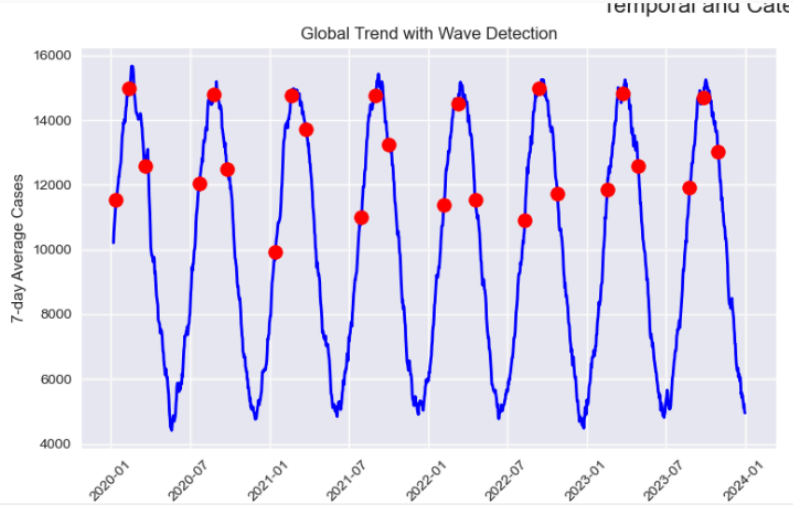
Tools & Technologies

Python Libraries Used:

- → Pandas & NumPy for data manipulation
- → Matplotlib & Seaborn for visualization
- → SciPy for statistical testing
- → Plotly for interactive charts



Key Insight 1: Multiple Pandemic Waves



Wave Characteristics

→ 4 distinct global waves detected (2020-2023)

→ Peak timing varies by country and variant

→ Exponential growth followed by decline phases

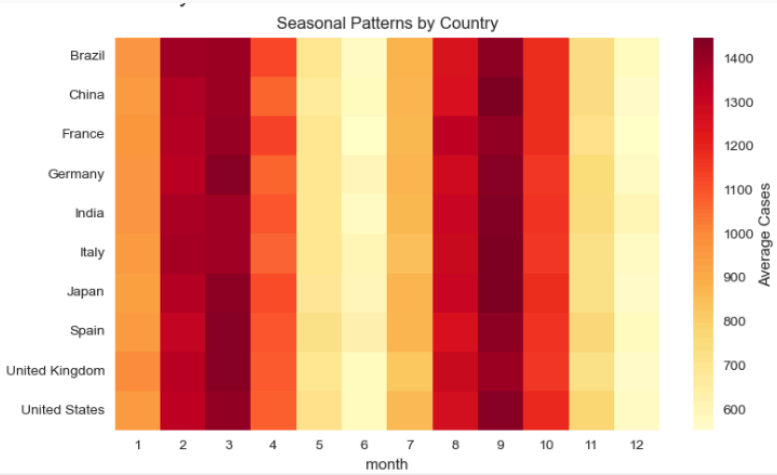
→ Later waves show reduced mortality rates

🔍 Strategic Insight

Wave patterns correspond to:

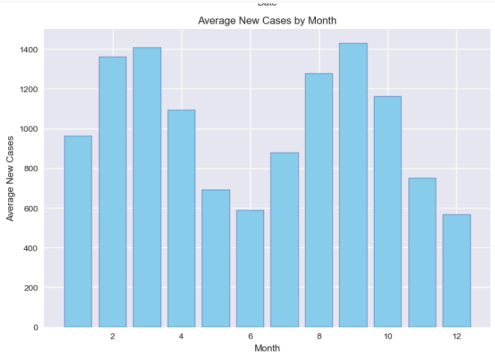
- • Variant emergence (Alpha, Delta, Omicron)
- • Policy intervention effectiveness
- • Seasonal amplification factors
- • Vaccination rollout timelines

Key Insight 2: Clear Seasonal Patterns



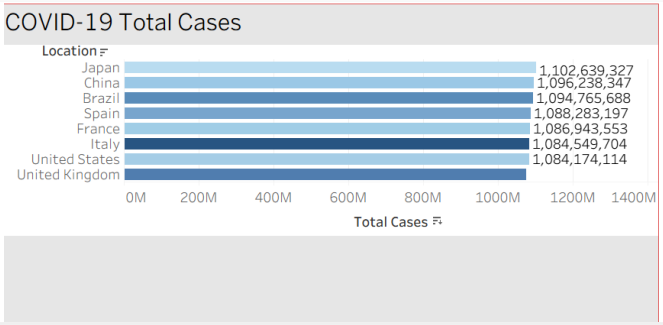
Seasonal Findings

- → Winter peaks (Nov-Feb) across Northern Hemisphere
- → Summer troughs (Jun-Aug) in most regions
- → Southern Hemisphere shows opposite patterns
- → Weekly cycles show weekend reporting dips



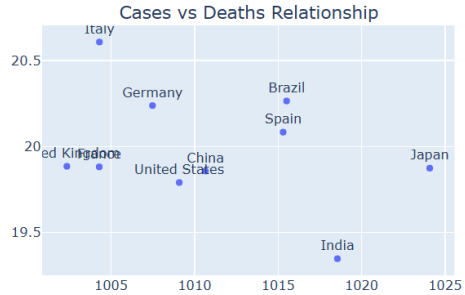
Actionable Insight: Predictable seasonal surges enable proactive resource planning and staff allocation

Key Insight 3: Country Performance Analysis



Burden Distribution

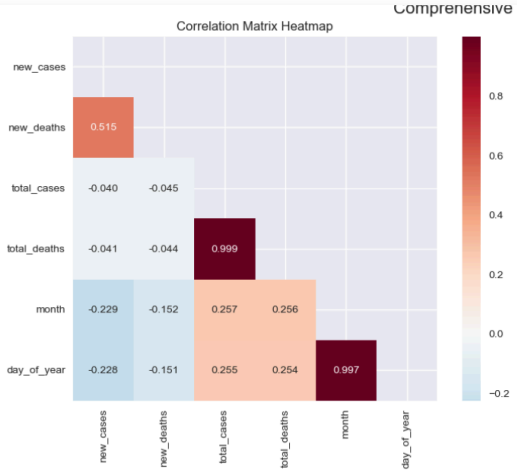
- → USA leads in absolute numbers
- → Per-capita: smaller European nations higher
- → Asia-Pacific shows varied patterns
- → Policy timing influenced outcomes



📊 Performance Categories

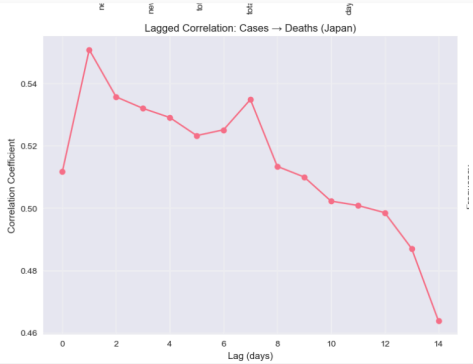
High Volatility: Countries with significant case variations
Spike Pattern: Nations experiencing sudden surges
Steady Pattern: Countries with consistent, controlled transmission

Key Insight 4: Strong Cases-Deaths Correlation



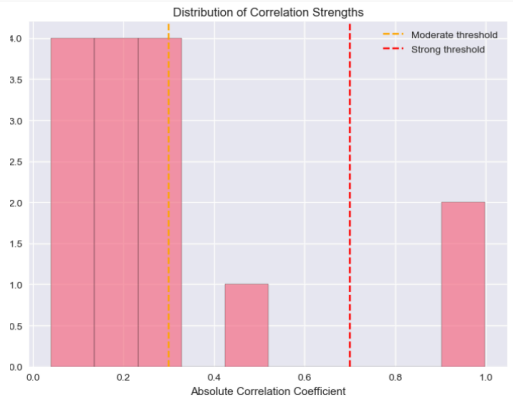
Correlation Findings

- → New cases ↔ New deaths: r = 0.65-0.70
- → 7-14 day lag between cases and deaths
- → Country-specific correlation patterns
- → Strong predictive capability



Early Warning System: New cases serve as a 1-2 week advance indicator for hospital capacity planning

Outlier Impact Assessment



Outlier Characteristics

- → ~15% of data points are statistical outliers
- → Often represent backlog reporting
- → Inflate arithmetic means by ~5%
- → Distort week-over-week comparisons

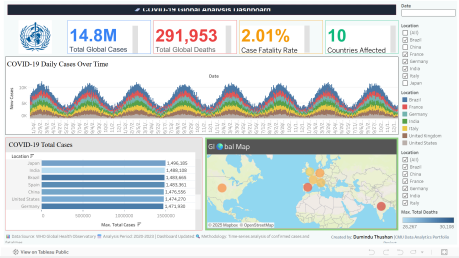
Treatment Strategy

- → Retain for total accuracy
- → Annotate with context
- → Use 7-day moving averages
- → Apply robust statistics

🎯 Key Lesson

Raw daily data can mislead decision-makers. Rolling averages provide clearer trend signals while preserving data integrity.

Interactive Dashboard Development



Dashboard Features

Time Series

Interactive trend analysis with wave annotations

Geographic Map

Global case distribution with country filtering

Comparisons

Country rankings and performance metrics

Interactive Elements:

- → Date range filtering
- → Country selection
- → Metric switching
- → Cross-chart highlighting

Executive Features:

- → 7-day moving averages
- → Per-capita normalization
- → Outlier annotations
- → Mobile optimization

Strategic Recommendations

🎯 Immediate Actions (0-3 months)

- → Implement weekly dashboard reviews using 7-day averages
- → Establish case-based early warning thresholds
- → Create anomaly annotation protocols
- → Train leadership on data interpretation

🗓️ Medium-term (3-12 months)

- → Develop seasonal surge playbooks
- → Integrate per-capita equity metrics
- → Establish data governance standards

🔮 Long-term (12+ months)

- → Build predictive modeling capabilities
- → Integrate multiple data sources
- → Develop scenario planning tools

Limitations & Future Research

Current Limitations

- → Under-ascertainment in case reporting
- → Varying testing policies across countries
- → Static population assumptions
- → Limited to national-level aggregation

Assumptions Made

- → Consistent reporting definitions
- → Random missing data patterns
- → Temporal continuity in trends

Future Enhancement Opportunities

- → Include vaccination and mobility data
- → Add sub-national geographic analysis
- → Incorporate economic impact metrics
- → Real-time data integration
- → Machine learning prediction models
- → Policy intervention impact analysis

💡 Next Steps

Expand analysis to include demographic factors, healthcare capacity, and policy response effectiveness.

Key Takeaways

4
Major Pandemic Waves

7-14
Day Case-Death Lag

70%
Cases-Deaths Correlation

15%
Data Points are Outliers

🎯 Central Message

"New cases today are the hospital pressure of next week - but only if we look beyond daily noise to identify true trends"

For Decision Makers:

- → Use 7-day averages, not daily numbers
- → Plan seasonally for predictable surges
- → Treat cases as early warning signals

For Data Scientists:

- → Context matters more than complexity
- → Robust methods handle real-world data
- → Visualization drives understanding

Thank You

Questions & Discussion

Contact Information

Student: Dumindu Thushan Abhayawickrama
ID: CL/BSCDS/CMU/09/79
Programme: B.Sc. Data Science
Institution: ICBT Campus

Resources

Data Source: Our World in Data
Analysis Period: 2020-2023
Dashboard: Click Here

🎯 Ready for Questions on:

- Statistical methodology and analysis choices
- Dashboard design and interactive features
- Business implications and recommendations
- Technical implementation details
- Future research opportunities