NAVIGATING COVID-19 RESEARCH DATA: AN INTEGRATED ANALYSIS OF EPIDEMIOLOGY, GOVERNMENT POLICIES, AND PUBLIC SENTIMENT

INFOSCI 301 Final Project
Author: Shouzhifan Zhu
Instructor: Prof. Luyao Zhang

INTRODUCTION

This project analyzes COVID-19-related tweets from February 1 to 10, 2020, examining temporal and geographical trends. By integrating epidemiological data and government policies, we assess how public sentiment evolved. Using SHAP analysis, we identify key factors influencing tweet volumes, offering insights into global engagement and crisis communication. Our findings highlight the interaction between public discourse, policy responses, and pandemic progression, providing valuable lessons for future health crises.

RESEARCH QUESTION

• Temporal Trends: How does tweet frequency vary by day, week, and month? Are there spikes during key pandemic phases, such as lockdowns or major policy shifts?
• Geographical Patterns: Which countries and cities had the highest tweet activity? How do regional differences in tweet volume relate to COVID-19 severity and government responses?
• Policy and Sentiment Impact: How did government interventions affect public discourse on social media? What key factors, identified via SHAP, influenced tweet volume and sentiment trends over time?

RESULT & VISUALIZATION
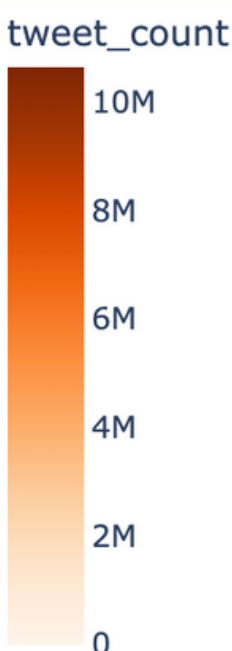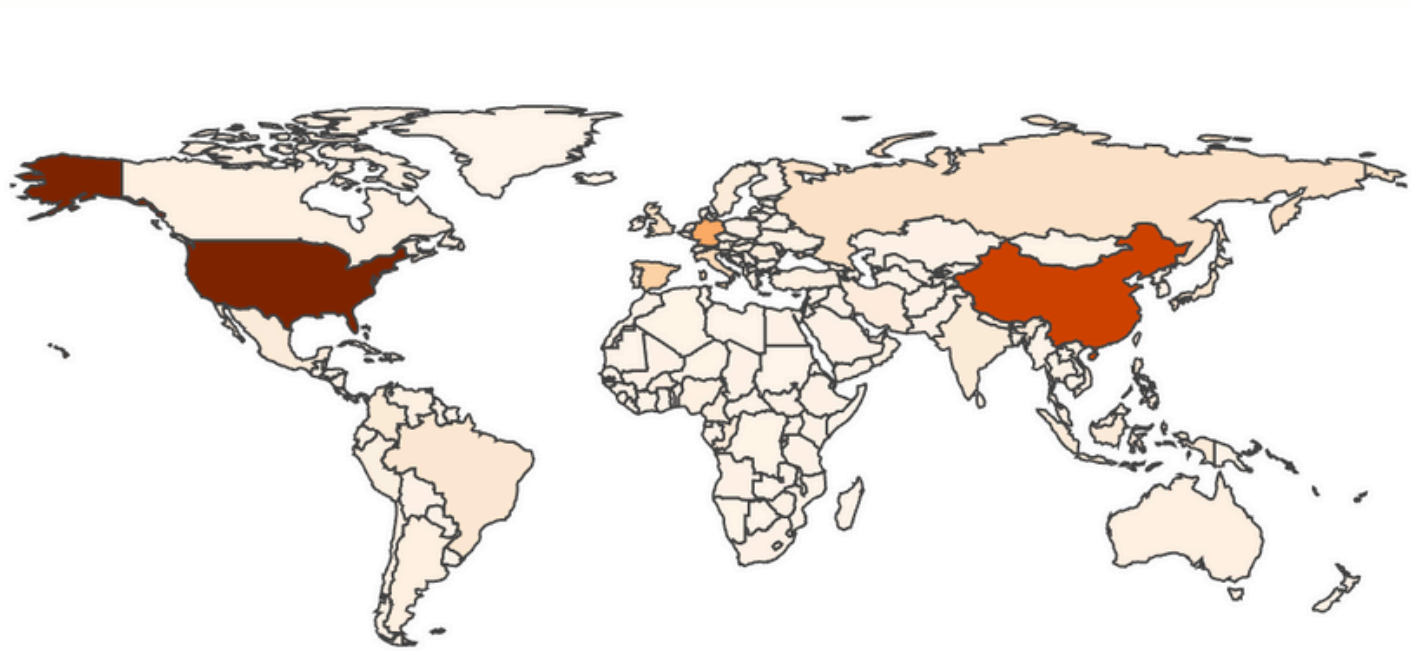
A map left visualizing COVID-19-related tweet volumes. A color gradient represents tweet counts per country:
- Darkest shades: Indicates highest tweet counts (e.g., USA, China).
- Medium shades: Moderately high tweet activity (e.g., India, Brazil, Russia).
- Lightest shades: Very low tweet activity (e.g., certain areas in Africa, and Central Asia).

There is a significant regional variation, influenced by population, internet penetration, and social media usage, reflecting global disparities in pandemic-related discussions.

RESULT & VISUALIZATION (CONTINUED)

1. **Is Weekend**: Higher positive SHAP values on weekends indicate a decrease in tweet volume.
2. **Weekday**: Values hover around zero, implying weekdays do not strongly influence tweet counts.
3. **Date (Ordinal)**: SHAP concentrated near zero, showing no strong day trend in tweet volume.
4. **Month**: Slight variations around zero; some months show mildly positive or negative contributions.
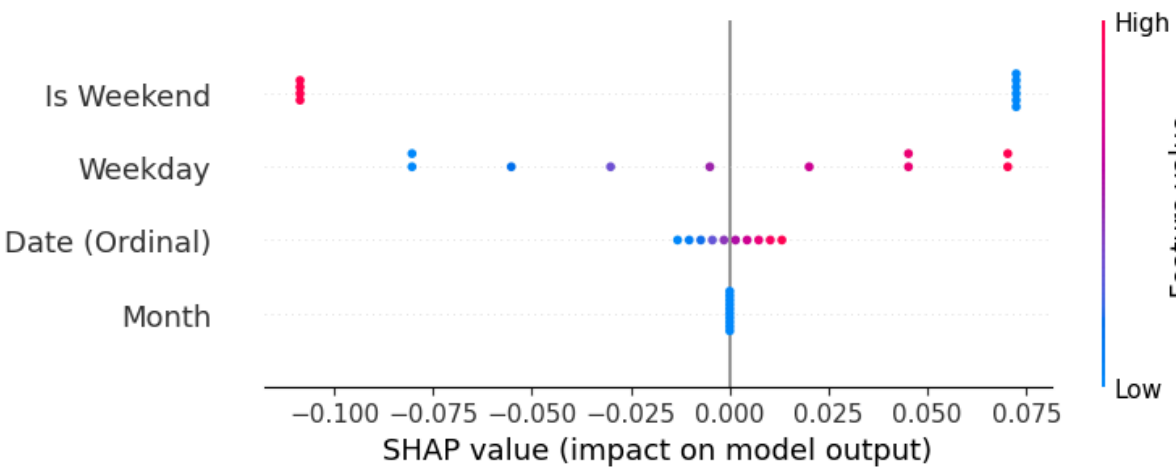
The SHAP plot right reveals how regions contribute to tweet volume:
- **Color Scale**: Red = High tweet volume, Blue = Low tweet volume.
- **USA**: Dominant influence (strongest red), indicating the highest impact on predictions.
- **China** (CHN): Also significant (red), reflecting large tweet counts.
- Spain (ESP), France (FRA), Italy (ITA): Moderate-high influence.
- Less Active Countries (e.g., Seychelles (SYC), Lesotho (LSO)): Minimal effect (blue).

Large, populous, and social-media-active nations (USA, China) drive the overall global tweet volume prediction.
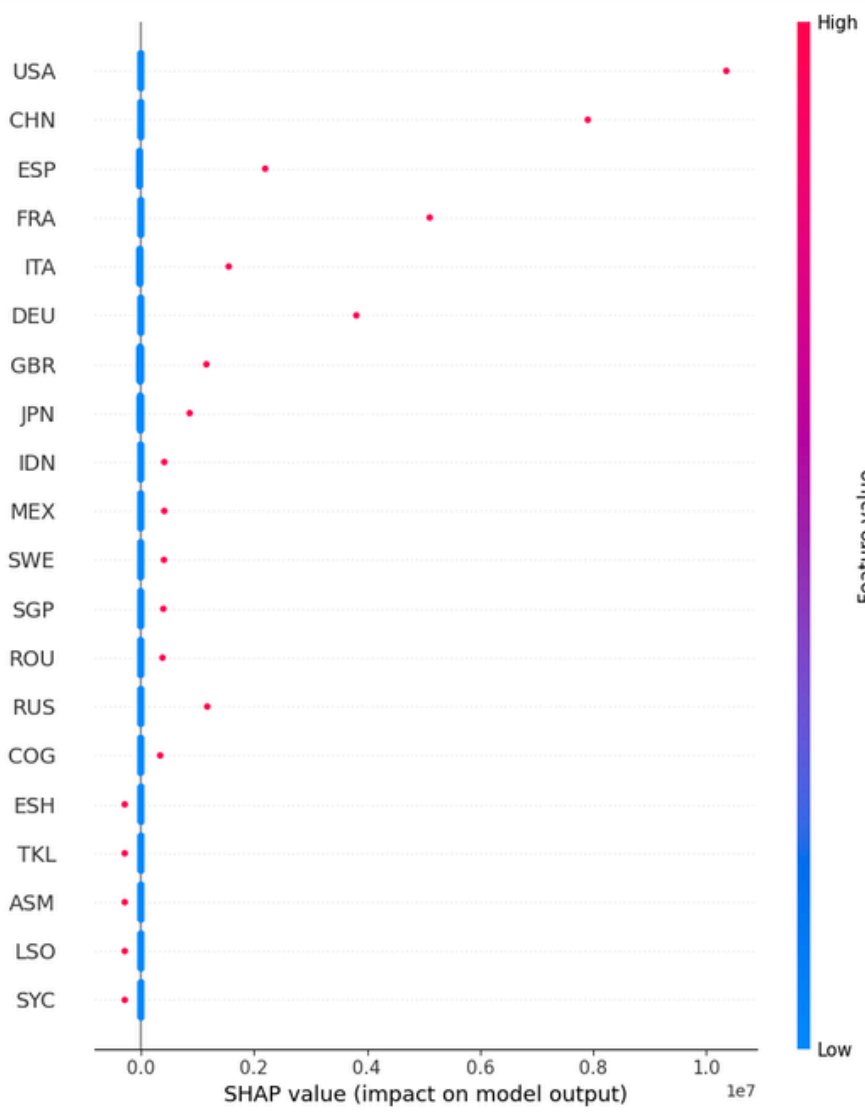
CONCLUSION

- Weekend Effect: Tweet activity does not spike on weekends, while weekdays and specific dates show more influence.
- Geographical Disparities: USA & China dominate tweet volume, while some regions show low engagement.
- SHAP Analysis Findings: Weekend factors and high-engagement countries are the strongest predictors of tweet volume.
- Global Social Media Trends: Population density, social media usage, and regional policies influence COVID-19 discussions.
- Impact: Results highlight the importance of real-time social media analysis for public health communication & policy strategies.

REFERENCES

Natural Earth. (2020). Admin 0 - Countries (1:110m scale). Retrieved from https://www.naturalearthdata.com/downloads/110m-cultural-vectors/110m-admin-0-countries/.

Qatar Computing Research Institute (QCRI). (2020). GeoCoV19: A Dataset of Hundreds of Millions of Multilingual COVID-19 Tweets with Location Information. Retrieved from https://crisisnlp.qcri.org/covid19.

Munzner, T. (2014). Visualization analysis and design. A K Peters/CRC Press.