

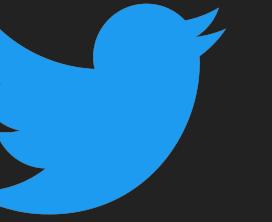


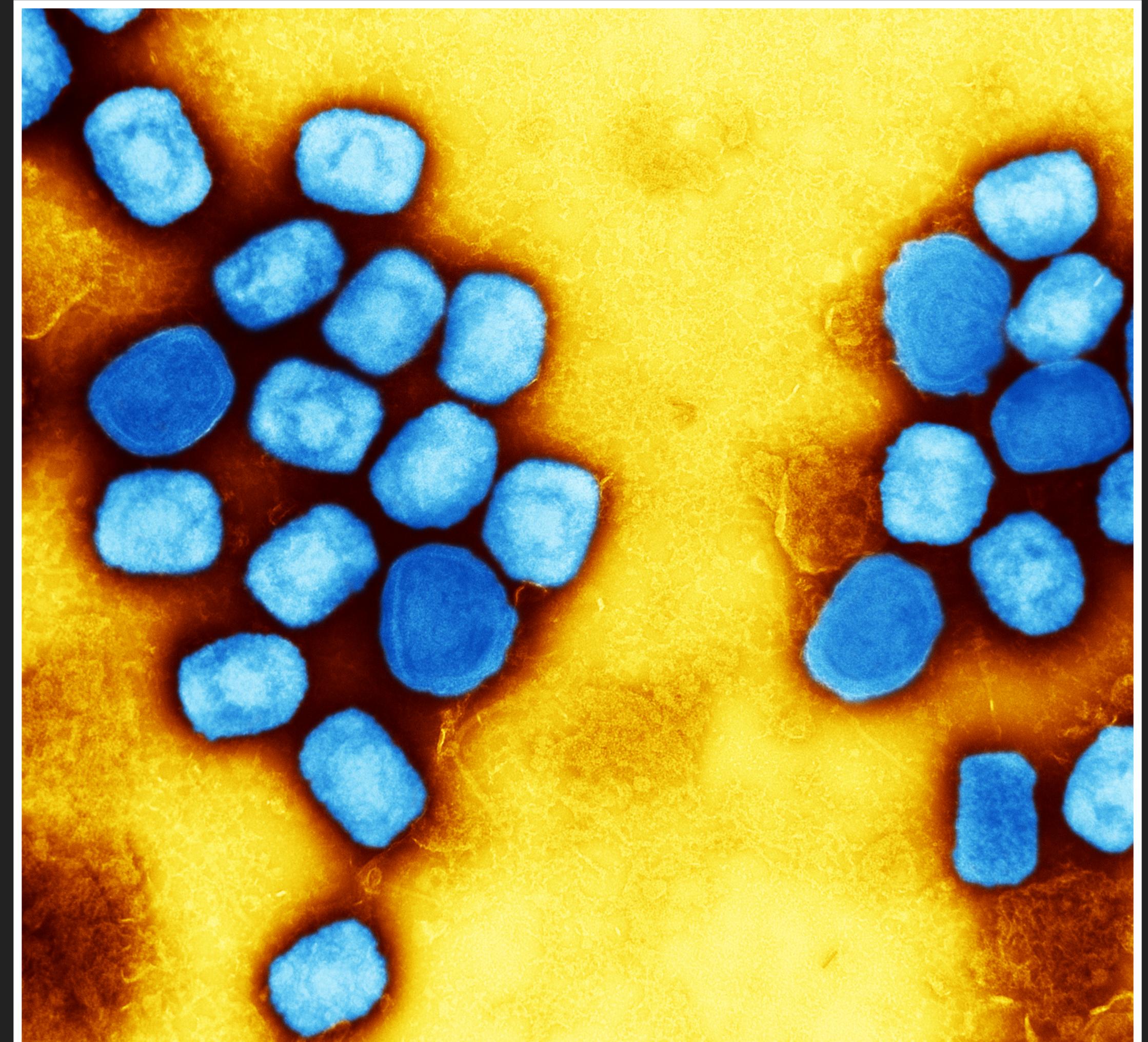
DATA ENGINEERING

MONKEYPOX TWEETS DASHBOARD

Sandra Paredes

INTRODUCTION

- **Motivation:** Twitter may reveal insights into how to address monkeypox. [1] 
- **Pipeline:** To monitor chatter on Twitter, we built a proof of concept dashboard.
- **Impact Hypothesis:** Inform communication strategy on monkeypox to avoid stigma.



Colorized transmission electron micrograph of monkeypox virus particles (blue) cultivated and purified from cell culture . Photo by NIH/NIAID

TWITTER DATA

DATA INGESTION

- ▶ **Dataset**
 - ▶ Kaggle Coronavirus TweetIDs [2]
 - ▶ Required hydration to retrieve tweet information [3]
 - ▶ Upload to Google Sheets
 - ▶ **EDA**
 - ▶ Removed duplicates, emojis, special characters
 - ▶ Filtered for English and 2022
 - ▶ $n \approx 28,500$



DATA STORAGE

SQL Database

- ▶ SQLite3
- ▶ Created database and tables
- ▶ SQL query to Pandas DataFrame



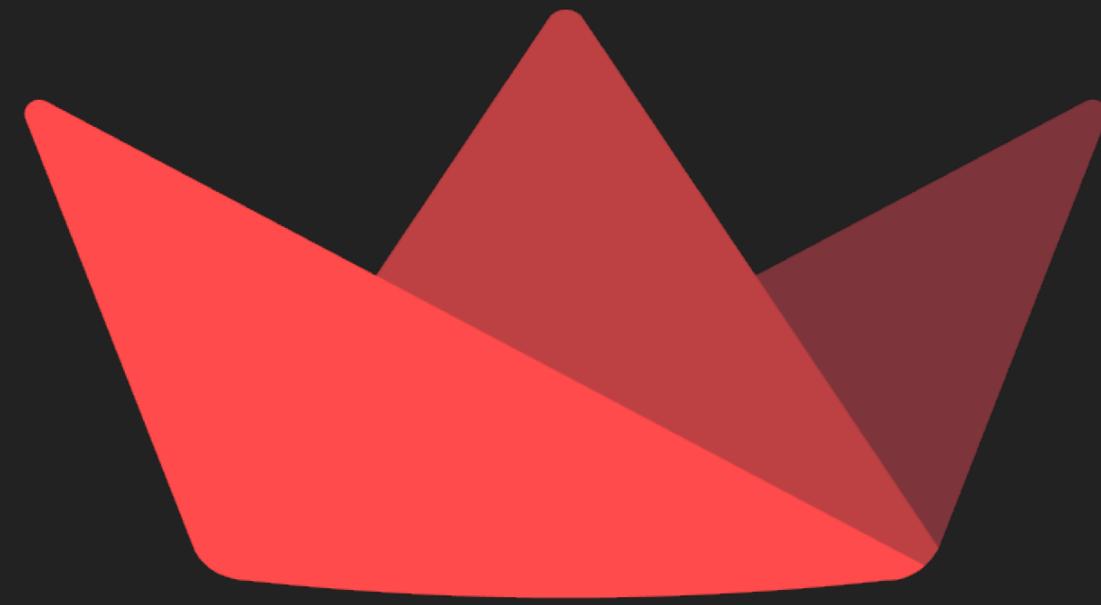
DATA PROCESSING

Google Colaboratory

- ▶ PySpark for word count
- ▶ NLP topic modeling
- ▶ Word cloud
- ▶ Tweets by date
- ▶ Case count by state [4]



DEPLOYMENT



Streamlit

CONCLUSIONS

Insights

- ▶ Location data not in standard format.
- ▶ Hashtags provided less information than anticipated.
- ▶ Majority of fields had sparse data.

MONKEYPOX

Planning to attend a festival or an event?

Events where attendees are outdoors, fully clothed and unlikely to share skin-to-skin contact are safer.

An indoor event where there is minimal clothing and where there is direct, skin-to-skin contact has more risk.



www.cdc.gov/monkeypox



CR333824-B

FUTURE WORK

Data Ingestion

- ▶ Twitter API
- ▶ Third-party, i.e, PhantomBuster

Processing

- ▶ Conda environment
- ▶ Clean and map location data



Photo by Tom Parks on Unsplash

APPENDIX

- ▶ Summary, data, and slides: github.com/slp22/data-engineering-project
- ▶ Monkeypox Tweet Dashboard slp22-mpx-app.streamlitapp.com



APPENDIX: SOURCES

1. Reducing Stigma in Monkeypox Communication and Community Engagement:
<https://www.cdc.gov/poxvirus/monkeypox/resources/reducing-stigma.html>
2. Monkeypox Outbreak – Nine States, May 2022: https://www.cdc.gov/mmwr/volumes/71/wr/mm7123e1.htm?s_cid=mm7123e1_w
3. Kaggle - Twitter Dataset on the 2022 MonkeyPox Outbreak: <https://www.kaggle.com/datasets/thakurnirmalya/monkeypox2022tweets>
4. CDC 2022 Outbreak Cases and Data: <https://www.cdc.gov/poxvirus/monkeypox/response/2022/index.html>