



**CMPT 733 Final Project**

---

**Global Data Analysis for  
Social Unrest Prediction**

---

Team ExaMine:

Raksha Harish, Tavleen Sahota, Harman Jit Singh

# PROJECT DEVELOPMENT

## PROBLEM STATEMENT

1

- 1.How can we predict social unrest events (protests, riots, etc.) across the globe based on factors like historical event patterns, poverty, hunger, crime rate, natural calamities, etc.?
- 2.What insights can be derived from the study of these factors?

2

## APPROACHES

- 1.Use Temporal Burst Patterns in Historical Events data to predict impactful unrest events
- 2.Use Additional Data Sources that capture social, political and economic contexts such as- poverty, hunger(food prices), crime rates, etc. to predict social unrest.

## CHALLENGES

3

- 1.Limited Domain Knowledge due to insufficient material for literature review
- 2.Data quota constraints when handling huge amounts of data
- 3.Integration of different datasets with inconsistent timelines and missing data for certain time periods

4

## ACTION PLAN

1. Data Collection
- 2.EDA
- 3.Visualizing data patterns
- 4.Integration of data sources
- 5.Feature Selection and Engineering
- 6.Building Models for prediction
7. Build an interactive UI to access real-time social unrest predictions

5

## FINAL PRODUCT

1. An Interactive Dashboard using Dash by plotly facilitating Data Exploration, Social Unrest Prediction and Insights about possible unrest events based on twitter hashtags.
- 2.Insights from Analysis

# DATA SCIENCE PIPELINE

## 1 - Data Collection



Datasets used :

- GDELT Global Events dataset (using Google BigQuery)
- ACLED (Armed Conflict Data)
- World Food Program and UN Global Criminal Rates datasets
- Latest Twitter Data

## 2 - Data Exploration



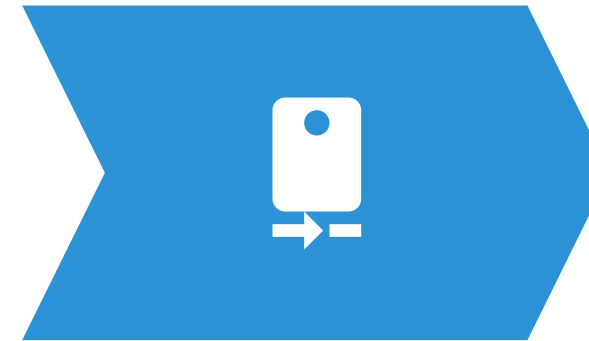
- Descriptive Statistics - Distribution Visualization, Outlier Detection, Correlation Analysis, etc.
- Temporal Analysis
- Event sequence/pattern analysis

## 3 - Data Manipulation



- Data Filtering - reduced 2.5TB to 1GB
- Data Aggregation
- ETL - One hot encoding, categorization into classes, splitting
- Feature selection and engineering
- Feature importance analysis for modeling and prediction

## 4 - Modelling



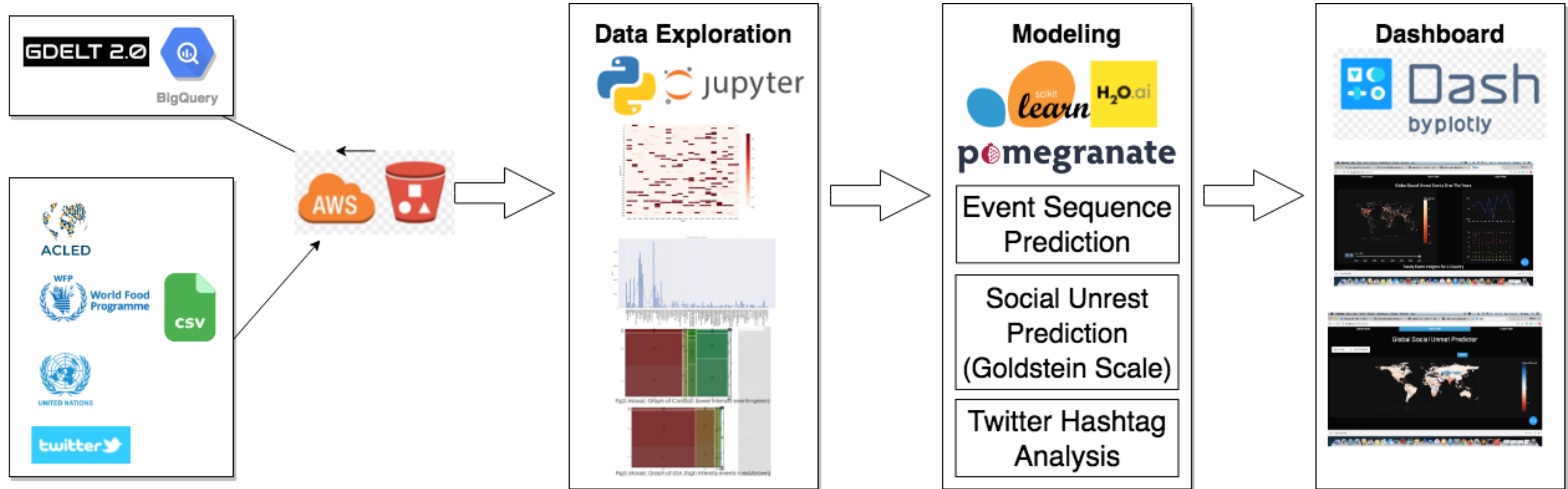
- Social Unrest Scale Prediction (Goldstein Scale):
  1. Linear Reg + Lasso
  2. Random Forest Regressor Model
  3. H2O ai flow + Gradient Boosting Machine
  4. H2O ai flow + Isolation Forest Model
- Even sequence prediction:
  1. Simulation using Markov chain and Prediction using Hidden Markov Model

## 5 - Data Product



- Interactive Dashboard allows for Data Exploration and Social Unrest Prediction.
- Dynamic analysis of twitter hashtag streams indicating social unrest
- Insights from EDA

# DATA AND TECHNOLOGIES



# METHODOLOGY

1. EDA + Feature Engineering - Exploring GDELT Dataset using Google BigQuery, and integrating with other datasets based on specific timeline to form an original bespoke dataset.
2. Modelling - Using the bespoke dataset to predict social unrest by creating models that predict the GoldsteinScale (range of -10 to +10 -> no impact to violent event scale). This is used to simulate the sequence of events observed for a specific timeline that could possibly lead to future social unrest, using Hidden Markov Model.

## Prediction of Goldstein Scale

1. Model-1: Linear Regression with Lasso - Accuracy of 59.01%
2. Model-2: Random Forest Regressor - Accuracy of 78.17%
3. Model-3: H2O ai framework with Gradient Boosting Machine - Accuracy of 86.31%
4. Model-4: H2O ai framework with Isolation Forest Model - Accuracy of 64.03%

CONCLUSION : Best models to predict Goldstein Scale for Social Unrest are Model-2 and Model-3

## Event Sequence Prediction

1. Simulating the sequence of events for a specific timeline based on the 5 types of events that have occurred previously - Markov Chain (MC) Simulation.
2. Using Hidden Markov Model (HMM) to predict the event sequence transition probability, firstly for MC simulated sequence, and then testing the model for a Random Event Sequence.



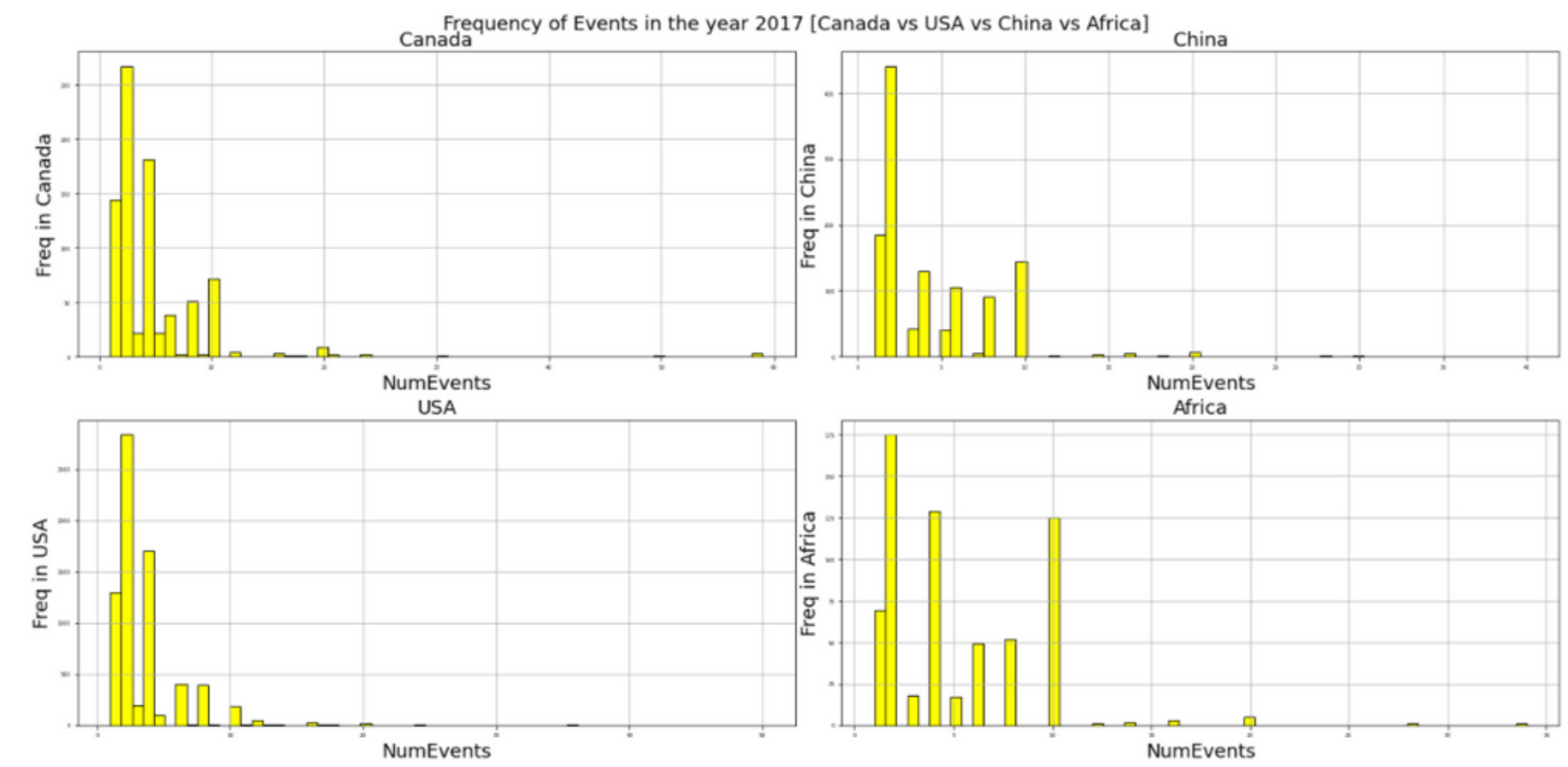
# METHODOLOGY (Contd)

- 3. Twitter feed analysis to detect social unrest in real time - Using tweepy and nltk packages with Twitter API for sentiment analysis to detect keywords (eg: protest, fight, violence, etc.) in tweets for trending twitter hashtags that indicate social unrest
- 4. Creating Dashboard - The final product created is a dynamic, user-interactive dashboard with functionalities to query and visualize data, and make predictions

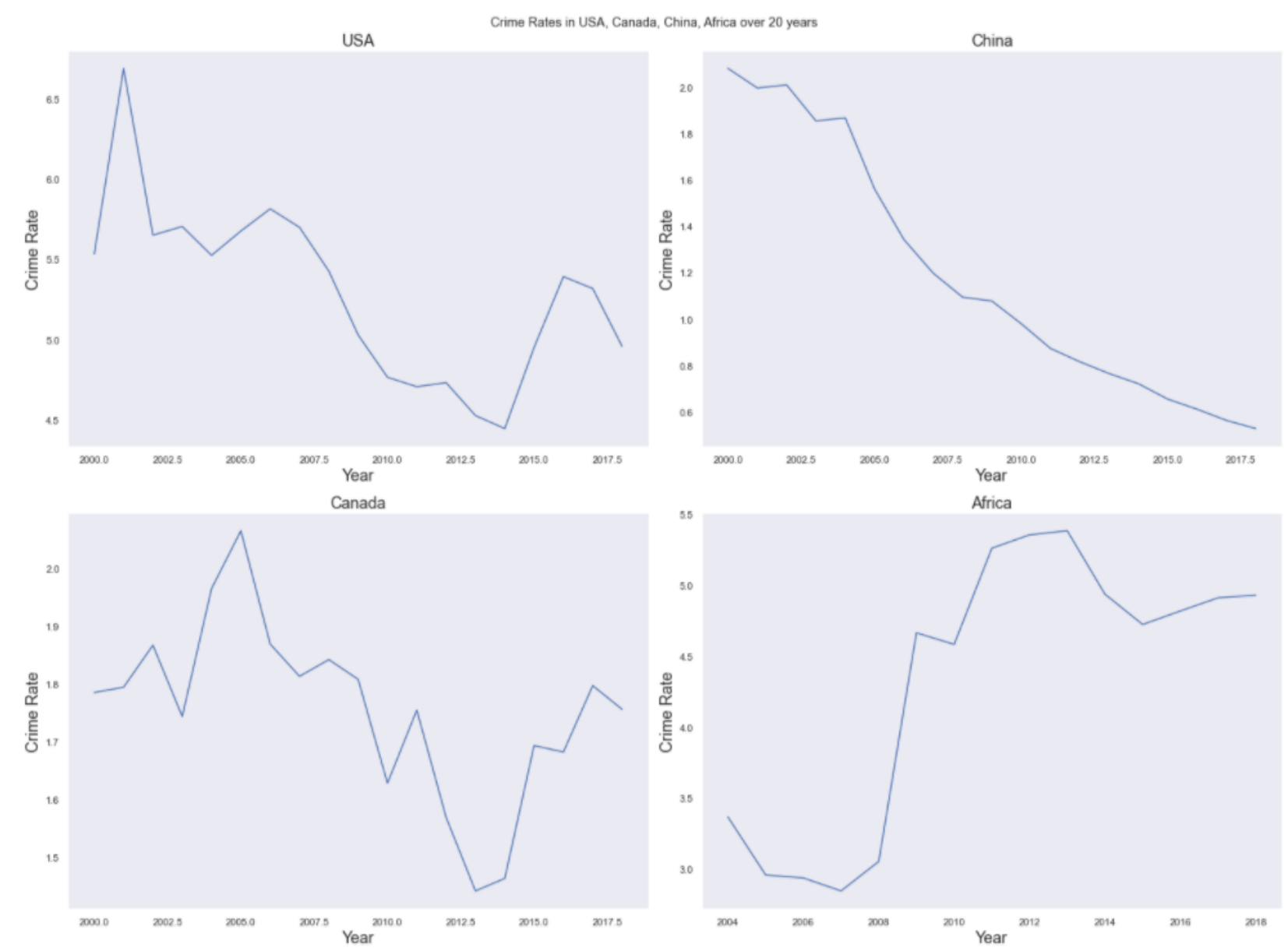
Twitter feed analysis to detect social unrest in real time	Creation of Dynamic, User-Interactive Dashboard
<div>1. For the countries with Goldstein Scale values above threshold, top 50 trending hashtags are identified and plotted based on the tweet volume.</div> <div>2. Hashtags and tweets analysis:<ul style="list-style-type: none"><li>For each hashtag, Tweets are searched, cleaned and analysed for unrest related keywords.</li><li>Hashtag with most unrest related keywords identified.</li><li>Related hashtags found for identified hashtag and plotted.</li></ul></div>	<div>1.Used Dash by plotly to make interactive dashboard allowing users to visualize key insights from historical data, predict future unrest and investigate Twitter feed related to unrest.</div> <div>2. Functionalities include:<ul style="list-style-type: none"><li>Time-Series Visualization (global &amp; country-specific)</li><li>Comparative Analysis of key features</li><li>Real-time Prediction of Likely Unrest Events and Goldstein Scales across the world for a specific date</li><li>Visualization of latest twitter hashtags associated with unrest</li></ul></div>

# RESULTS

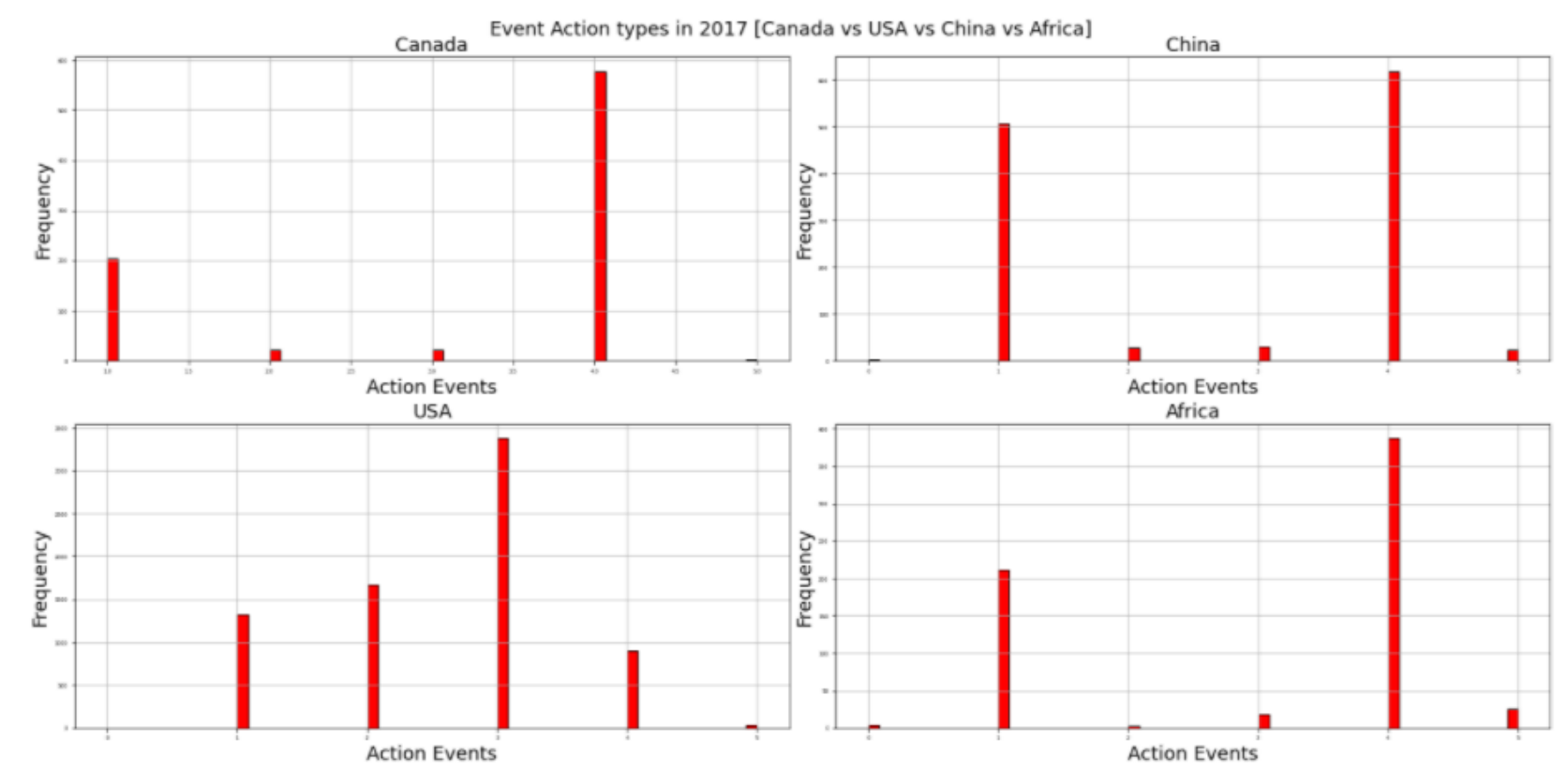
1



2



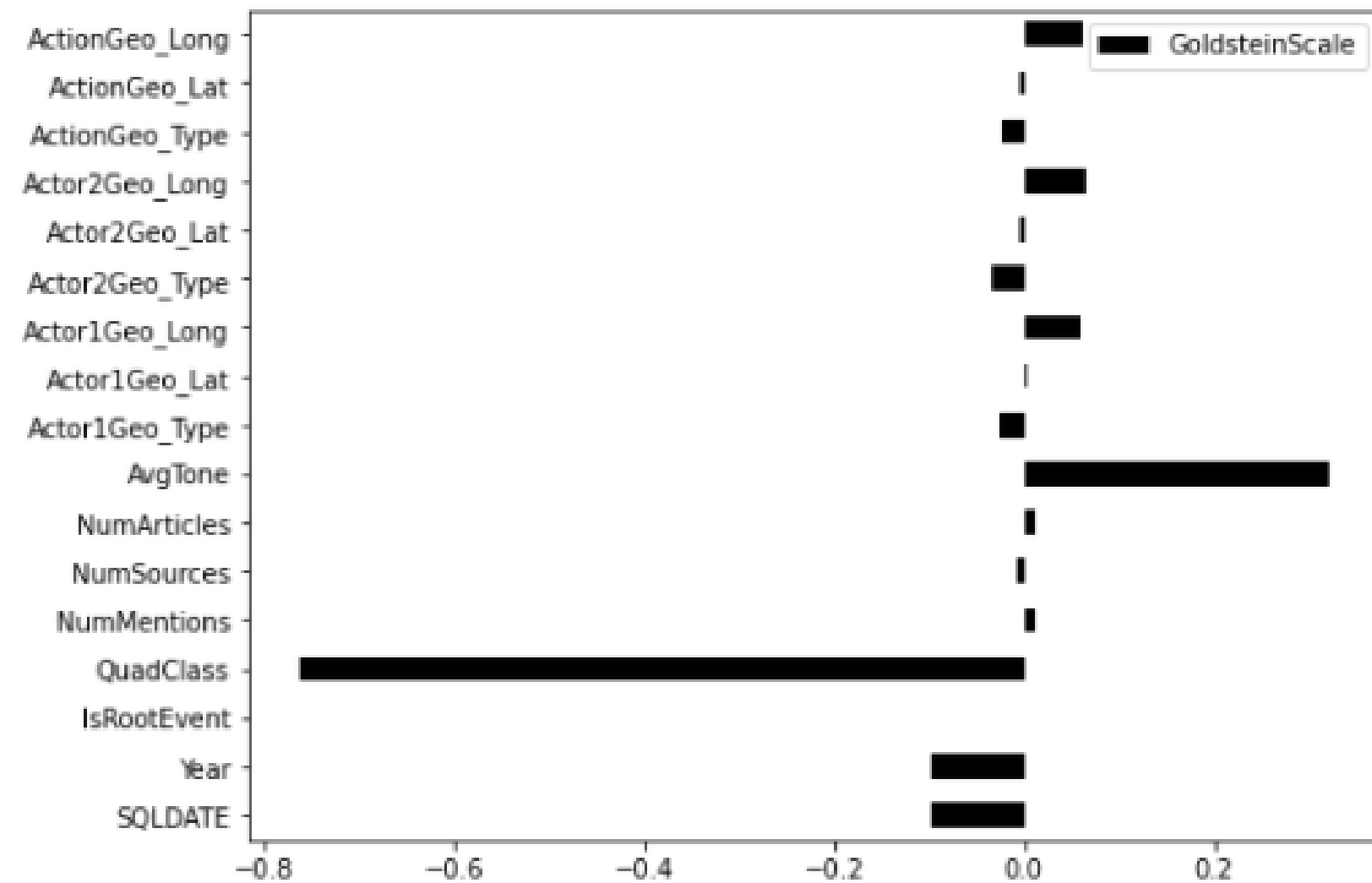
3



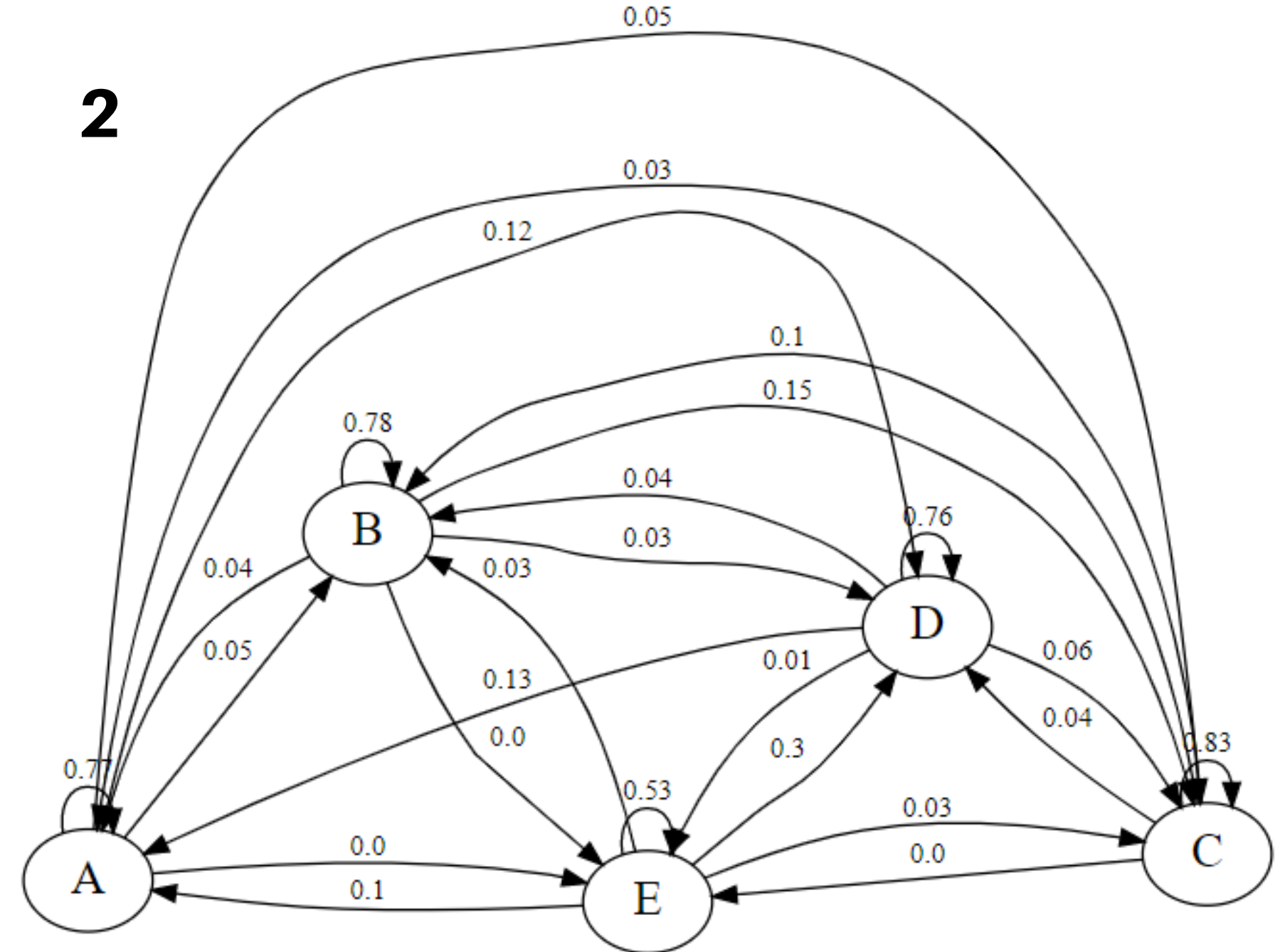
**Figures : Initial EDA for Frequency of regular events (1), Crime Rates (2) , and Frequency of occurrence of Action event types (3)**

## RESULTS (Contd)

1

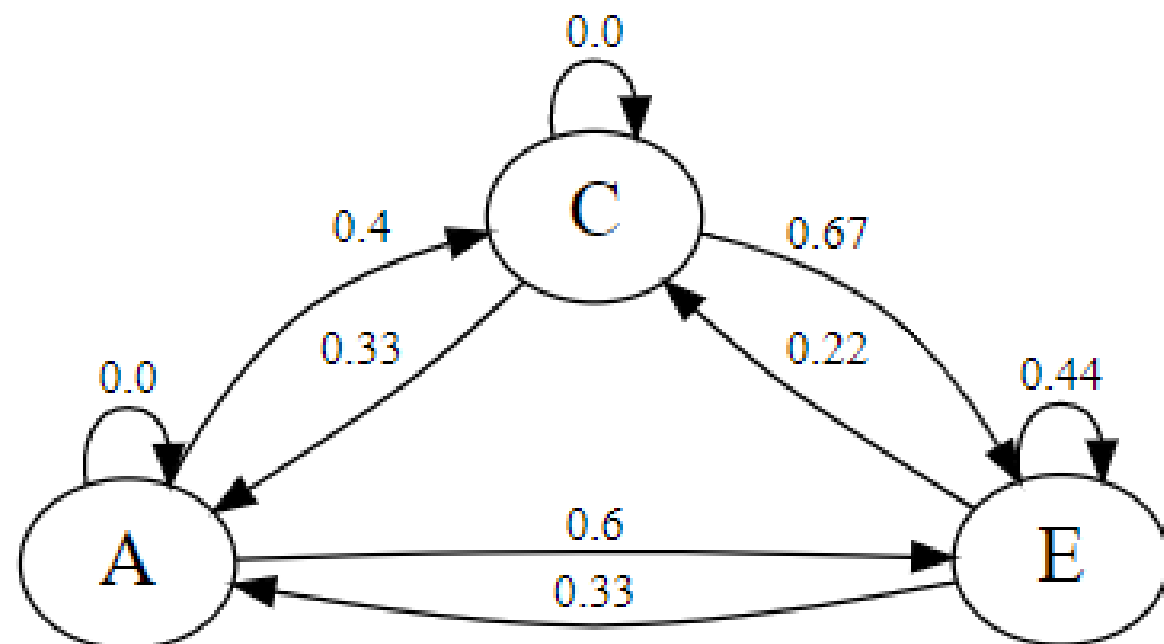


2



**Figures : Pearson Correlation for most relevant features (1), Initial MC simulation for event sequence (2) , HMM prediction for random input event sequence (3)**

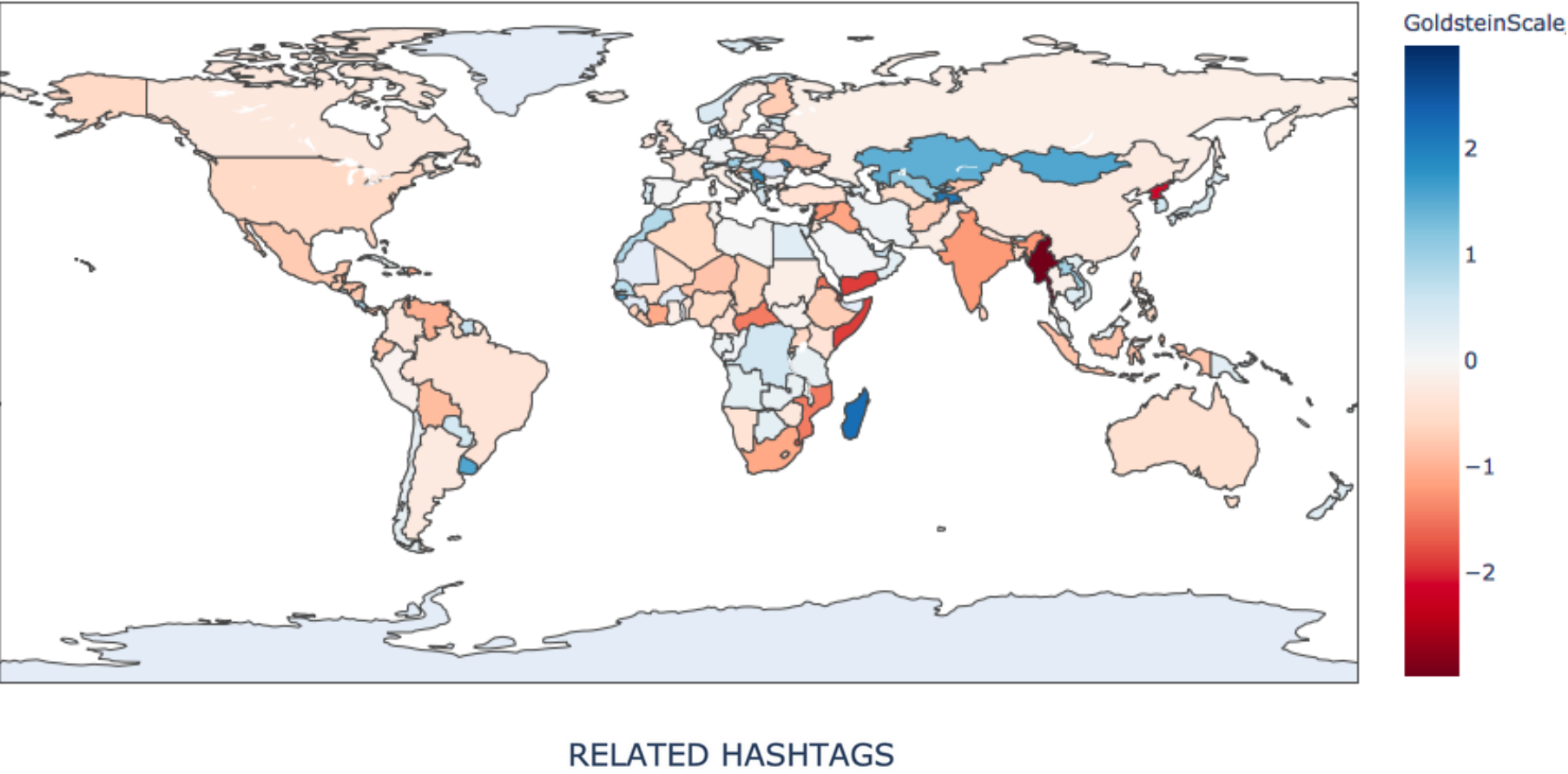
3 Graph for Random Sequence



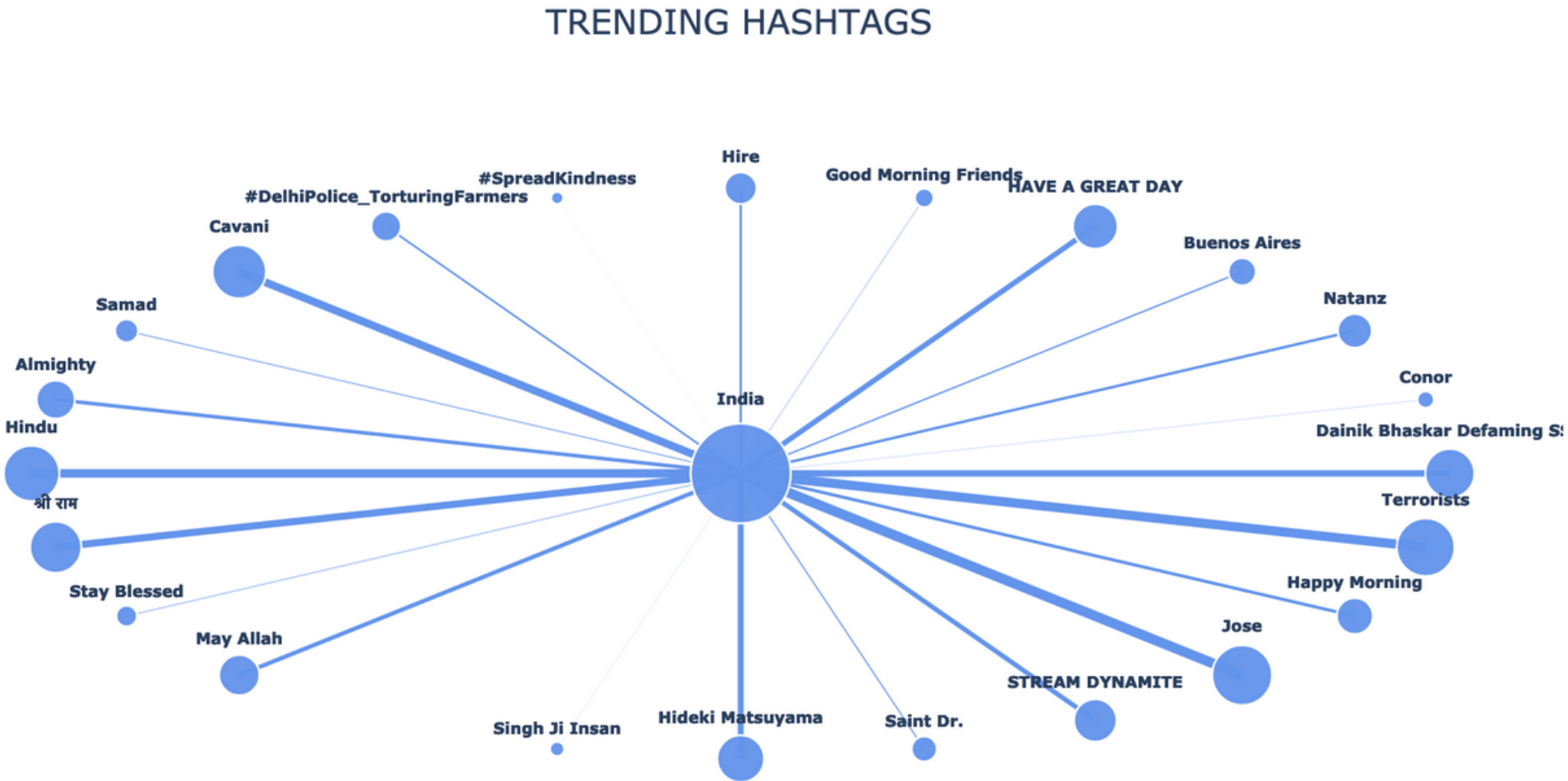


# RESULTS (Contd)

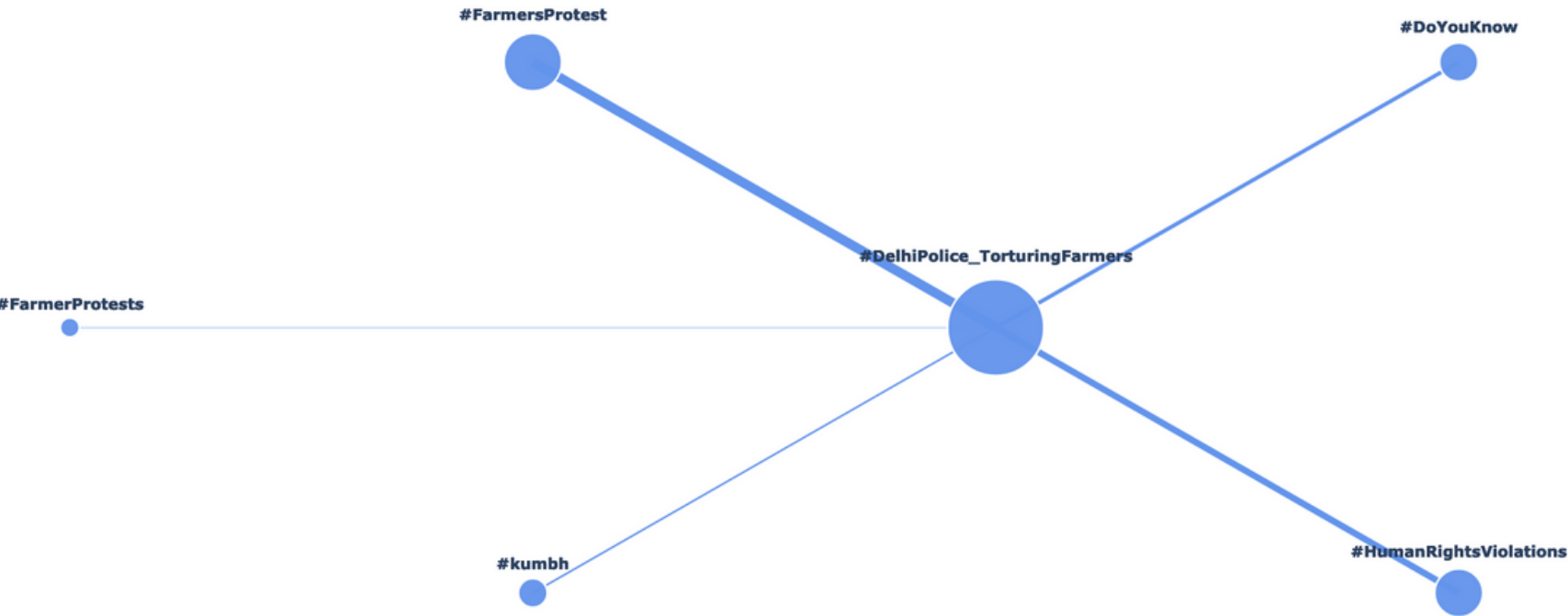
1



2

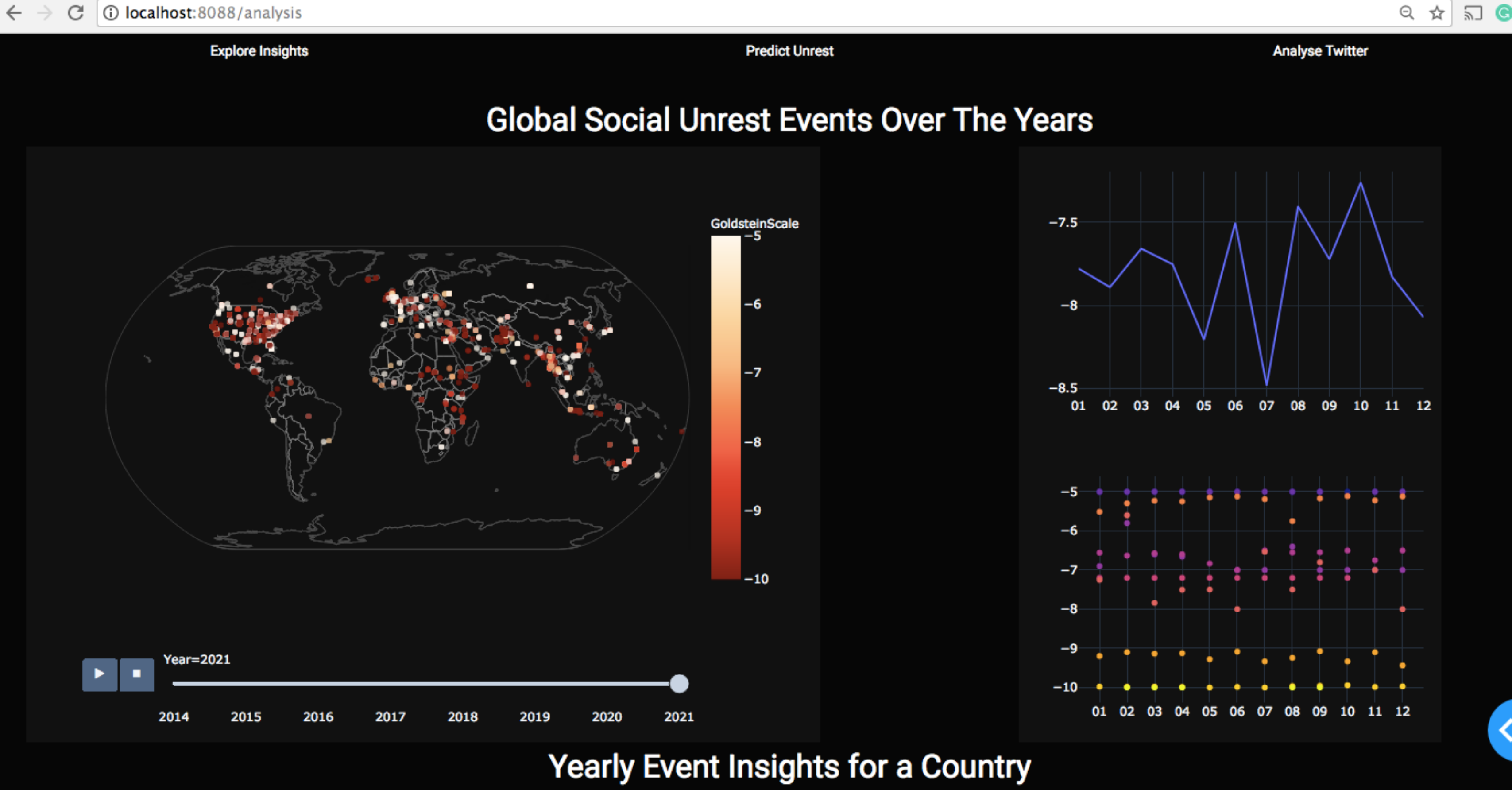


3

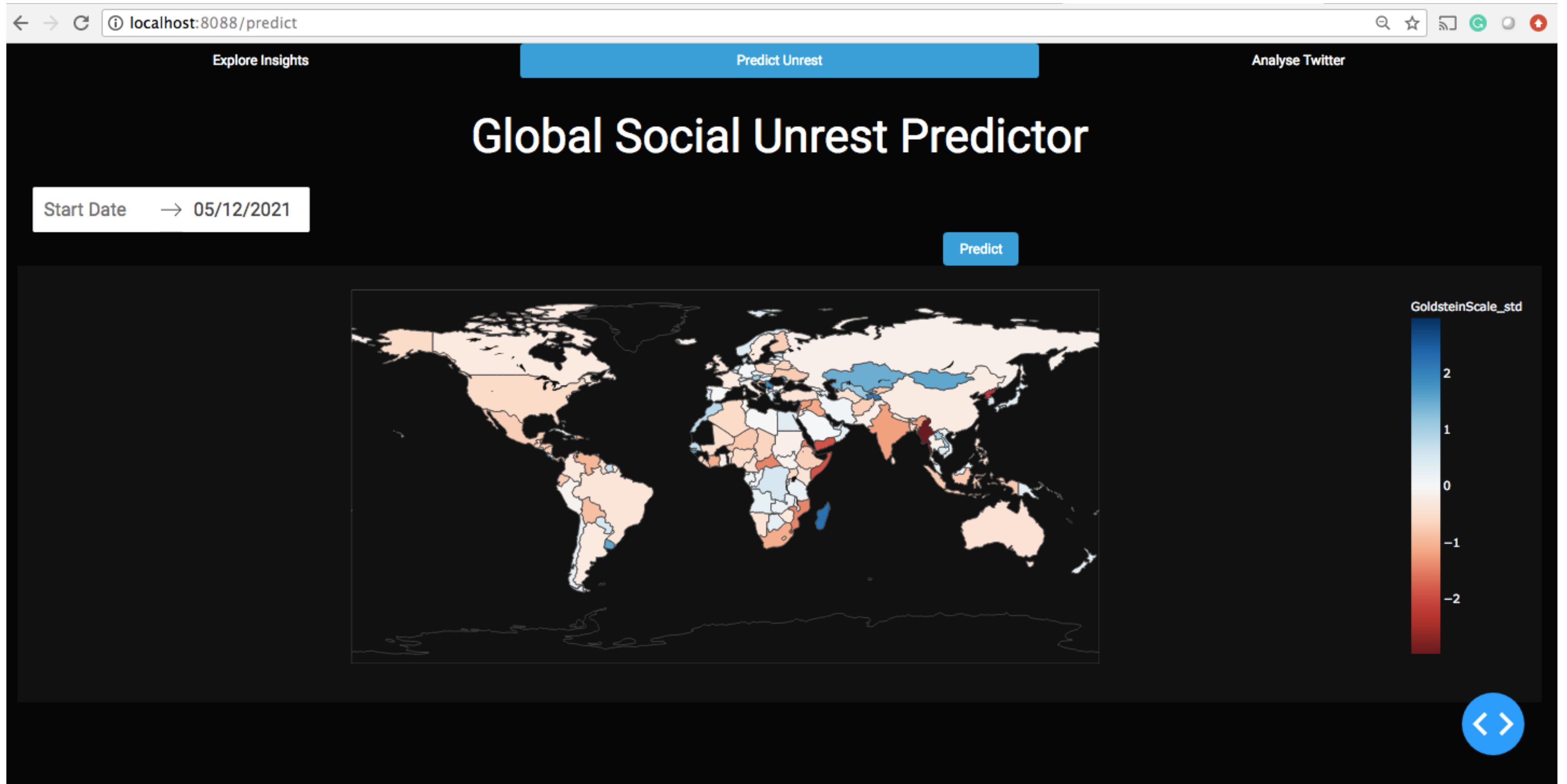


Figures : Social Unrest around the world (Mar-Apr, 2021) (1) , Trending Hashtags for India (2), Related hashtags to identified unrest related hashtag (3)

# DASHBOARD UI-1



# DASHBOARD UI-2



**THANK YOU**