

Forecasting CO₂ Emissions in Rwanda: A Time Series Analysis

Presented by:
Ayush Sati,
Chaitanya Chaniyara,
Neha Roy Choudhury,
Prasad Hadbe,
Vamsi Krishna Pirati



DATASET INSIGHTS

Source: Kaggle

ID_LAT_LON_YEAR_WEEK	latitude	longitude	year	week_no	SulphurDioxide_SO2_column_number_density	Cloud_sensor_zenith_angle	Cloud_solar_azimuth_angle	Cloud_solar_zenith_angle	emission
ID_-0.510_29.290_2019_00	-0.51	29.29	2019	0	-0.000108339721919	35.6324155330658	-138.7864227294920	30.752140045166000	3.7509942
ID_-0.510_29.290_2019_01	-0.51	29.29	2019	1	2.05267923327286E-05	39.55763284190910	-145.18392972572900	27.251779362626600	4.0251765
ID_-0.510_29.290_2019_02	-0.51	29.29	2019	2	0.00051414151676	30.401823012154800	-142.519544680814	26.19329569765290	4.231381
ID_-0.510_29.290_2019_03	-0.51	29.29	2019	3		24.380356674011	-132.66582821963800	28.829154729021800	4.3052855
ID_-0.510_29.290_2019_04	-0.51	29.29	2019	4	-7.87662604528522E-05	37.392979049682600	-141.50980529785200	22.20461235046390	4.347317
ID_-0.510_29.290_2019_05	-0.51	29.29	2019	5	0.0002943348954434	37.00702015182660	-137.38871898896600	21.157819651278400	4.3108187
ID_-0.510_29.290_2019_06	-0.51	29.29	2019	6	-0.0002852637844625	30.4915412902832	-128.196337890625	21.72120227813720	4.2693343
ID_-0.510_29.290_2019_07	-0.51	29.29	2019	7	-1.6218675227724E-05	35.84024825950980	-122.25069903060000	20.432005274889300	4.251361

- The dataset contains data from 497 unique locations
- These locations were clustered based on their latitude and longitude coordinates to group geographically close locations.
- Peaks and troughs in emissions were observed to occur simultaneously across multiple locations, suggesting a linear dependence between them.
- There was a sudden decrease in emissions during the 2020-21 period, which was attributed to the COVID-19 pandemic. This was treated as an outlier in the data.

DATASET DESCRIPTION

- Weekly data from Sentinel-5P was collected from January 2019 to November 2022, encompassing seven main features: Sulphur Dioxide, Carbon Monoxide, Nitrogen Dioxide, Formaldehyde, UV Aerosol Index, Ozone, and Cloud.
- **Sulphur Dioxide** - COPERNICUS/S5P/NRTI/L3_SO2
- **Carbon Monoxide** - COPERNICUS/S5P/NRTI/L3_CO
- **Nitrogen Dioxide** - COPERNICUS/S5P/NRTI/L3_NO2
- **Formaldehyde** - COPERNICUS/S5P/NRTI/L3_HCHO
- **UV Aerosol Index** - COPERNICUS/S5P/NRTI/L3_AER_AI
- **Ozone** - COPERNICUS/S5P/NRTI/L3_O3
- **Cloud** - COPERNICUS/S5P/OFFL/L3_CLOUD

Handeling COVID-19 Outlier

- **Data Splitting:** The data was split into two periods: pre-COVID and during-COVID
- **Weekly Average Emissions:** Weekly average emissions were calculated for both periods.
- **Emission Ratio:** The ratio of average emissions in the pre-COVID period to the COVID period was computed.
- **Adjusted COVID Emissions:** This ratio was applied to the COVID period emissions to adjust them and update the original DataFrame.

Dimensionality Reduction

- **Technique Used:** Singular Value Decomposition (SVD) was employed to reduce the dimensions of the location data.
- **Reduced Dimensions:** The data was reduced to 5 dimensions to simplify the analysis while retaining essential information.

Modeling

- **Auto ARIMA and Prophet Models:** Both Auto ARIMA and Prophet models were applied to each of the 5 components obtained from SVD.
- **Multivariate Analysis:** Although considered, multivariate analysis was rejected due to high multicollinearity and poor prediction performance of complex models. The top 20 highest correlated variables had correlations less than 0.1, which further supported this decision.

Model Validation and Selection

- **10-Fold Cross Validation:** Time series 10-fold cross-validation was performed on each component to ensure robust model evaluation.
- **Residual Diagnostics:** Residuals from the models were checked and found to be randomly distributed, indicating that they did not contain additional valuable information.
- **Evaluation Metrics:** Models were evaluated using metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).
- **Final Model Selection:** Based on the evaluation metrics, the best-performing model was selected for the final forecast.

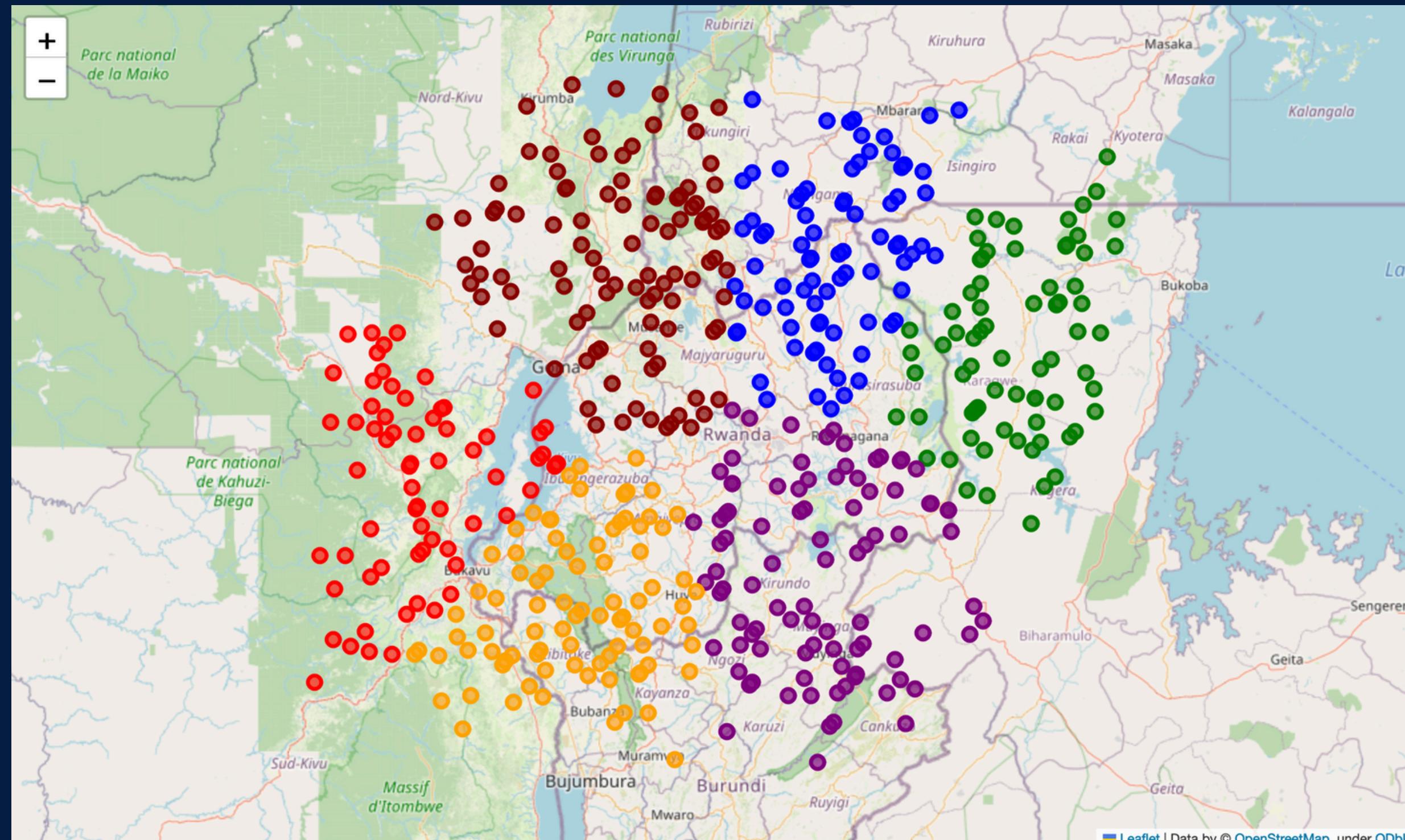
Forecast Horizon

The chosen model was used to forecast emissions for the next 14 days.

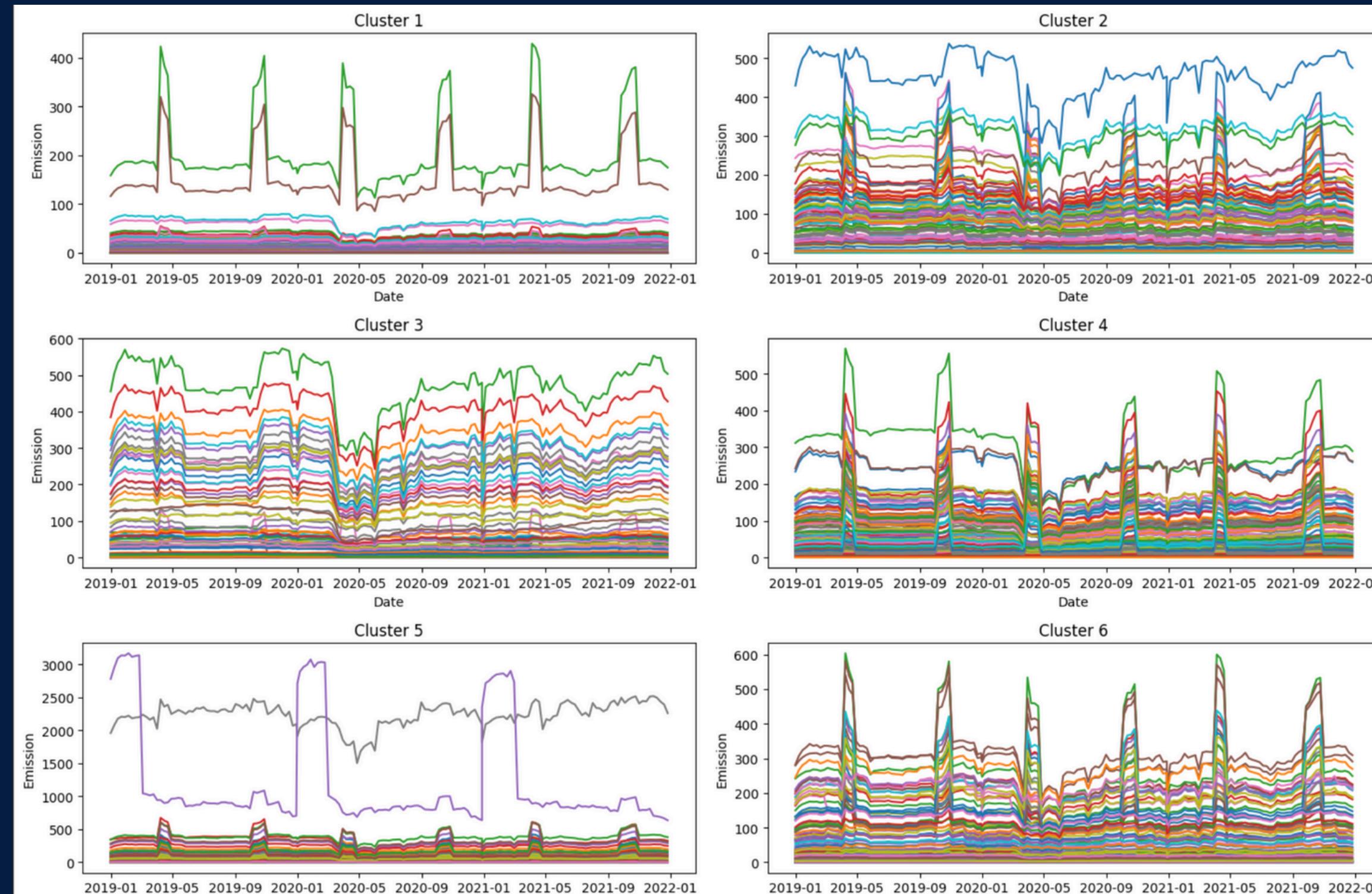
3D Scatter Plot of the locations



2D Representation of the Locations



TIME SERIES PLOT FOR ALL THE LOCATIONS BASED ON CLUSTER





Thank You!