

INTRODUCTION

BACKGROUND: DATA SOURCES AVAILABLE

Diagnosis on current available open sources as we have yet to receive data from the government.

SOURCE	WEBSITE/URL	NOTES ON DATA QUALITY
Kerala Open Data	https://kerala.data.gov.in/	<ul style="list-style-type: none">• 9,398 resources, 370 data catalogs, 32 departments• Seems impressive at first glance (for the sheer number of datasets available) but the quality of the datasets remains wanting.• API doesn't work
Map Kerala Alpha	https://map.opendatakerala.org/	<ul style="list-style-type: none">• Good source for downloadable (SHP) geofiles, but other layers take time to load and hard to see which layers are available at one glance
Open Data Kerala	https://opendatakerala.org/ https://github.com/opendatakerala	<ul style="list-style-type: none">• Github has quite a number of data but some in non-editable formats (pdfs, jpeg etc)
Kerala State Spatial Data Infrastructure	https://opensdi.kerala.gov.in/	<ul style="list-style-type: none">• 25 maps, 217 layers• Downloadable map layers available in images but many do not have shapefiles• Webapp allows for stacking of layers but not analysis• Contains overlaps/some data similarly found on KSDMA's platform (e.g. Flood 2018)
Kerala Data Portal	https://spatialagent.org/Kerala/	<ul style="list-style-type: none">• Most of the sublinks do not work, take an extremely long time to load, or do not have downloadable data

MACHINE LEARNING FOR DATA

DATA PRODUCT 1: Predictive Analysis for Air Quality in Kerala

About the Data

Historical Daily Ambient and Air Quality data obtained from kerala.data.gov.in.

Years available: 1987-2002, 2005-2015.

A total of 27 years is available for download

Some years are not well documented (for example, for the year 1987, only one record was noted)

Sampling Date	State	City/Town/	Location o	Type of Loc	SO2	NO2	RSPM/PM1	SPM
1/3/2008	Kerala	Kochi	Eloor	Industrial	2.9	10.7	56	81
1/7/2008	Kerala	Kochi	Eloor	Industrial	5.8	13.5	83	103
1/10/2008	Kerala	Kochi	Eloor	Industrial	NA	29.9	70	99
14/1/2008	Kerala	Kochi	Eloor	Industrial	8.4	16.9	77	118
17/1/2008	Kerala	Kochi	Eloor	Industrial	21.6	13.3	83	108
21/1/2008	Kerala	Kochi	Eloor	Industrial	26.5	7.6	51	77
24/1/2008	Kerala	Kochi	Eloor	Industrial	10.2	16.9	54	93
28/1/2008	Kerala	Kochi	Eloor	Industrial	2.7	39.3	48	96
31/1/2008	Kerala	Kochi	Eloor	Industrial	9.1	7	43	101
2/4/2008	Kerala	Kochi	Eloor	Industrial	8.9	10.3	34	56
2/7/2008	Kerala	Kochi	Eloor	Industrial	16.8	NA	40	65
2/11/2008	Kerala	Kochi	Eloor	Industrial	11.3	6.5	42	80
14/2/2008	Kerala	Kochi	Eloor	Industrial	14.3	9.8	40	55
18/2/2008	Kerala	Kochi	Eloor	Industrial	NA	13.8	50	72
21/2/2008	Kerala	Kochi	Eloor	Industrial	NA	5.5	46	68
25/2/2008	Kerala	Kochi	Eloor	Industrial	NA	6.3	71	100
27/2/2008	Kerala	Kochi	Eloor	Industrial	6.3	21.1	111	127
29/2/2008	Kerala	Kochi	Eloor	Industrial	13.6	5.4	88	152
18/3/2008	Kerala	Kochi	Eloor	Industrial	NA	6	53	95
25/3/2008	Kerala	Kochi	Eloor	Industrial	7.9	NA	66	118
28/3/2008	Kerala	Kochi	Eloor	Industrial	4.6	5.4	46	71

Snapshot of data columns and contents from year 2025

- Inconsistent date formats – had to manually extract and edit
- Values not consistently present for every year
- Reports
 - Sampling Data
 - State
 - City
 - Type of Location
 - SO2
 - NO2
 - RSPM10
 - SPM

MACHINE LEARNING FOR DATA

DATA PRODUCT 1: Predictive Analysis for Air Quality in Kerala

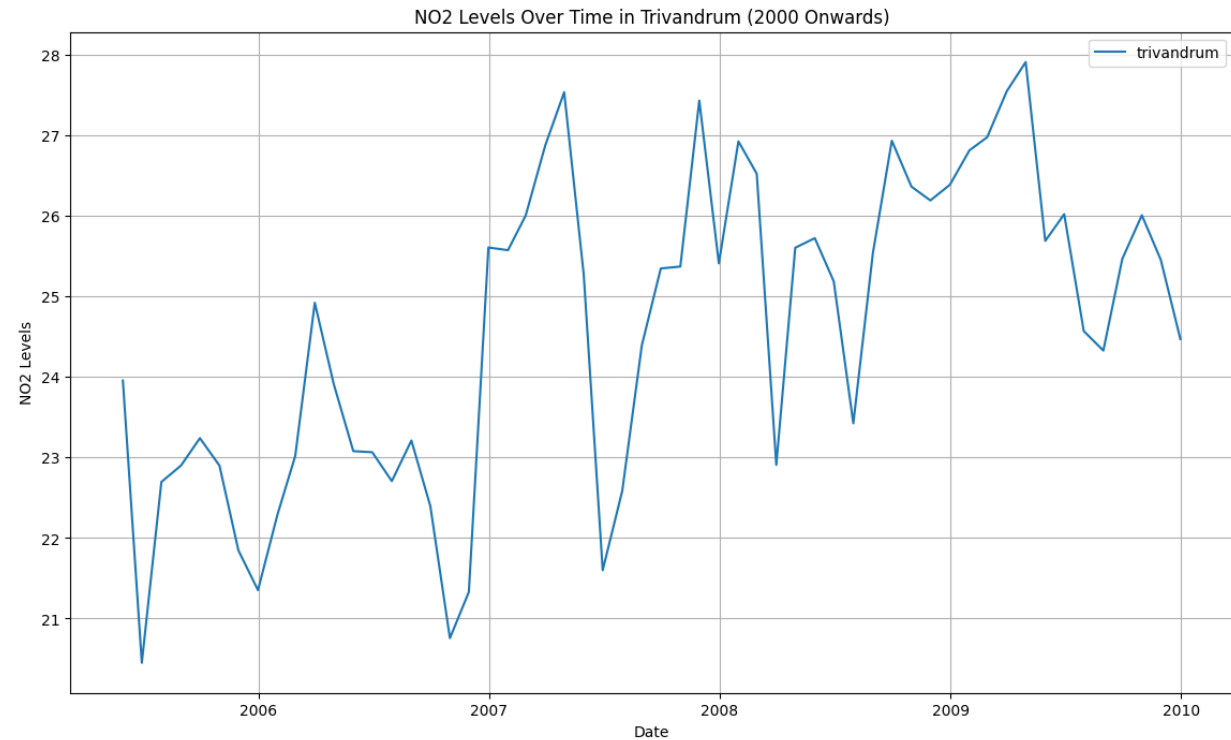
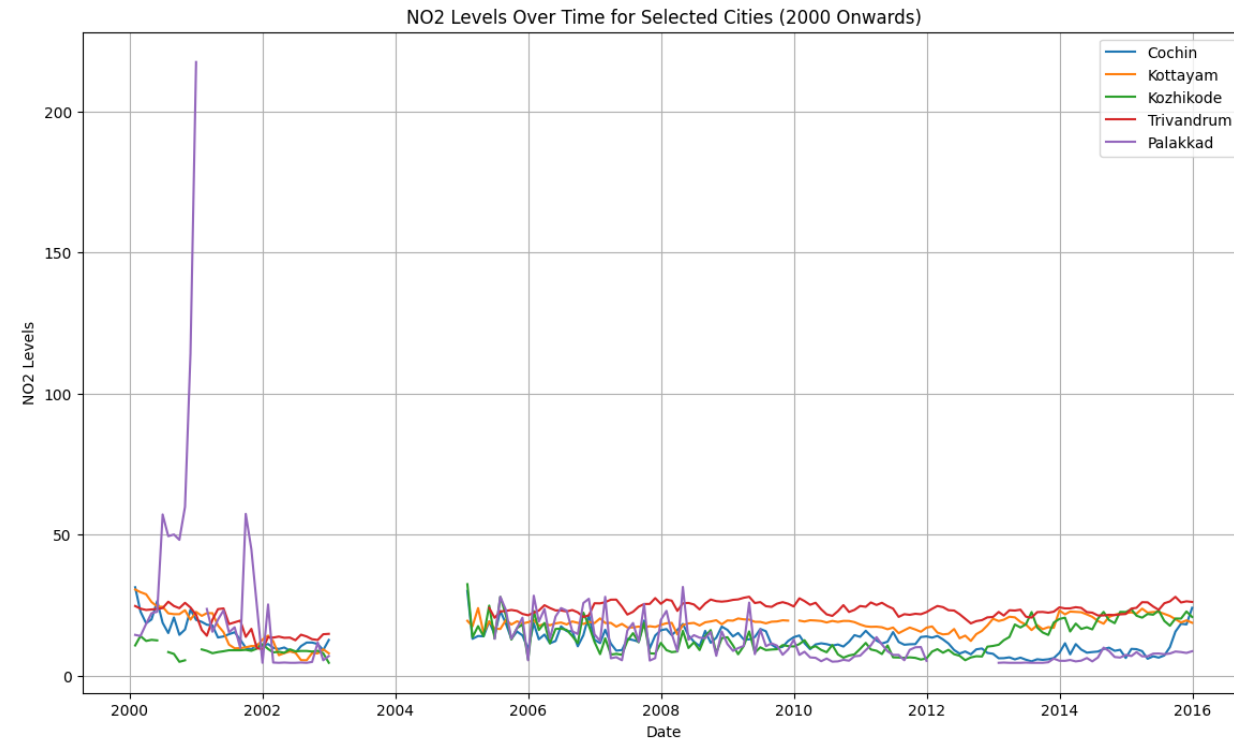
About the Data

Historical Daily Ambient and Air Quality data obtained from kerala.data.gov.in.

Years available: 1987-2002, 2005-2015.

A total of 27 years is available for download

Some years are not well documented (for example, for the year 1987, only one record was noted)



Visualizations

MACHINE LEARNING FOR DATA

DATA PRODUCT 1: Predictive Analysis for Air Quality in Kerala

About the model

Predicting NO2 levels for Palakkad from years 2016 to 2020 (where data is not available)

Training a Random Forest Regressor on the training data.

Things that can be improved further

Improving the model

Using other models and comparing

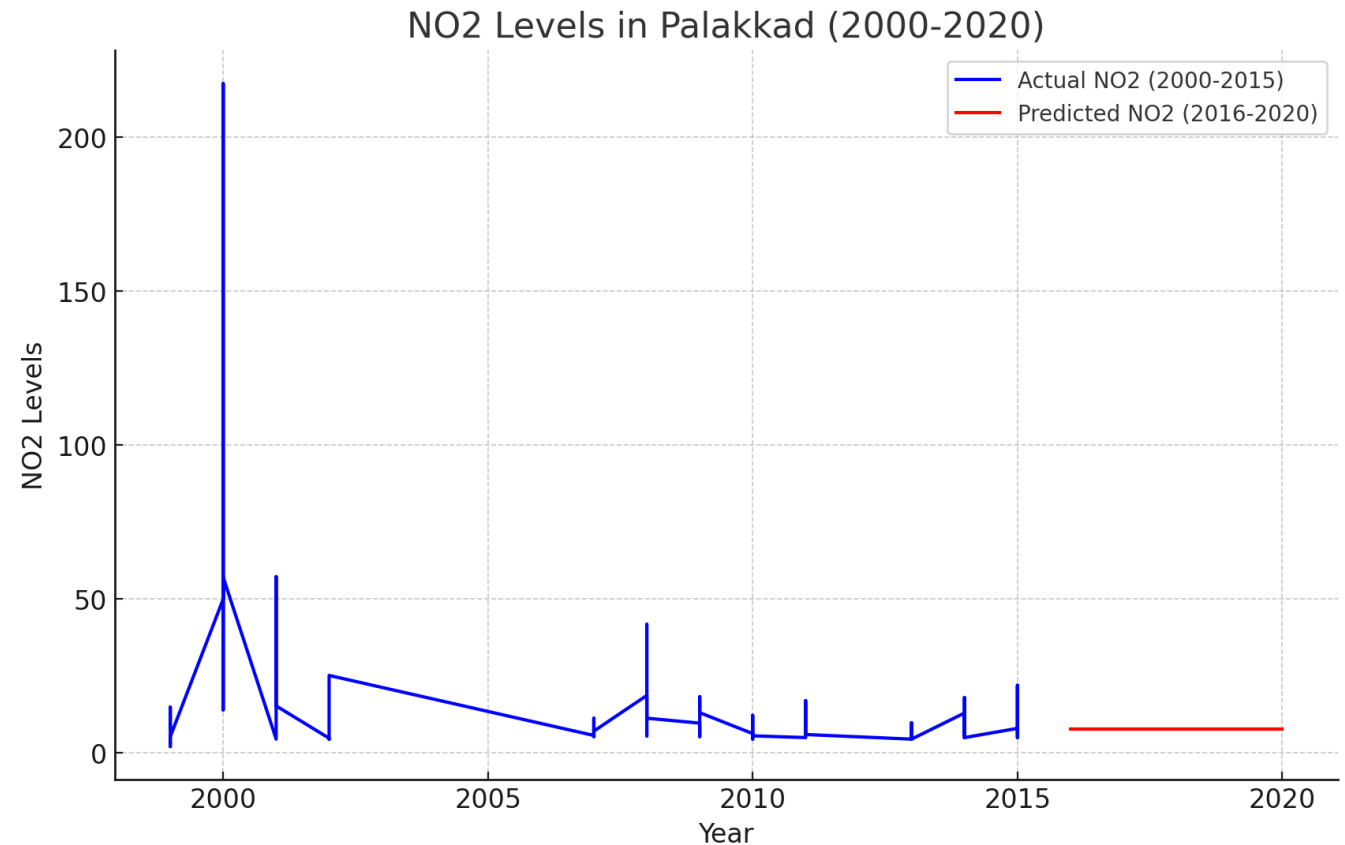
```
# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train the Random Forest Regressor
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Predict on the test set
y_pred = rf_model.predict(X_test)

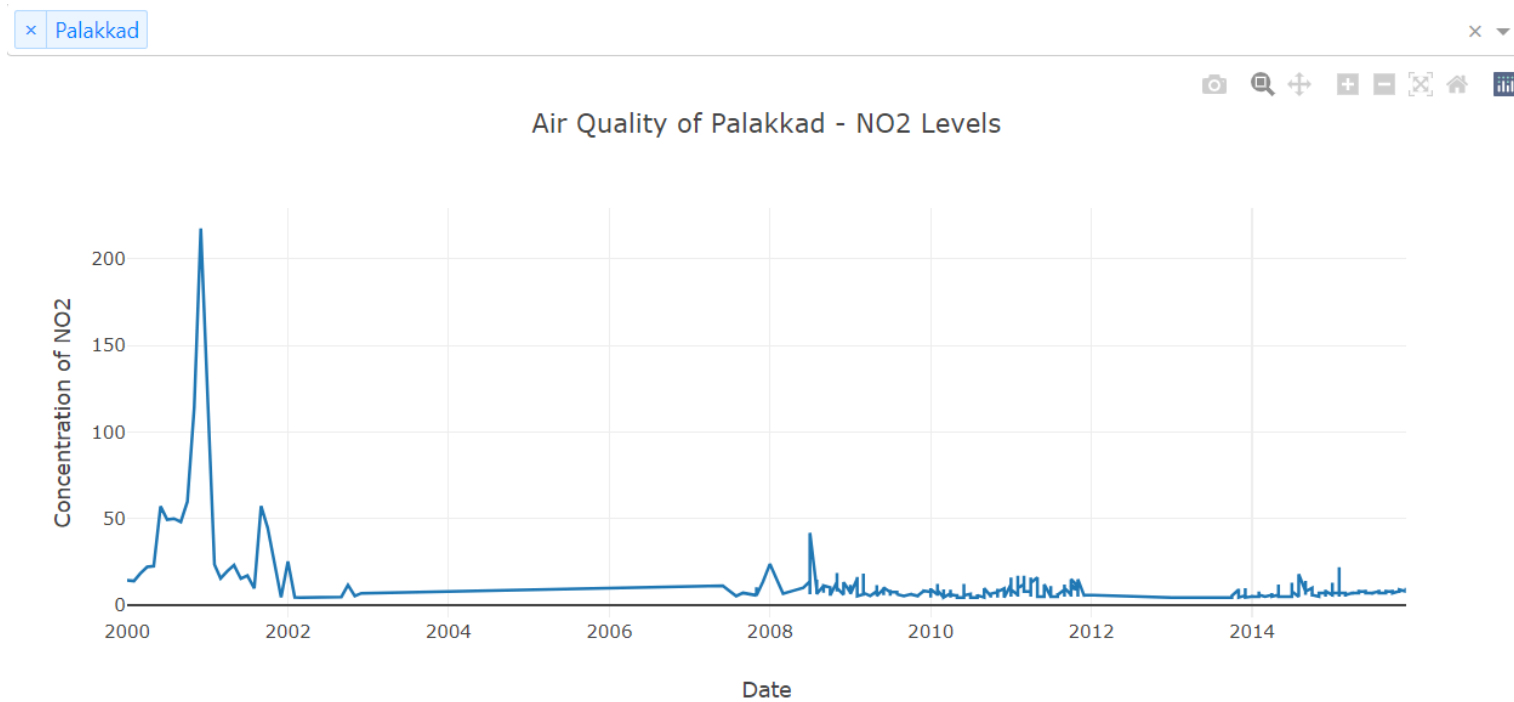
# Predict future NO2 levels (2015-2020)
future_dates = pd.date_range(start='2015-01-01', end='2020-12-01', freq='MS')
future_data = pd.DataFrame({'Parsed Date': future_dates})
future_data['Year'] = future_data['Parsed Date'].dt.year
future_data['Month'] = future_data['Parsed Date'].dt.month
```

Code Snapshot



VISUALIZATIONS ON WEB-BASED DASHBOARDS

DATA PRODUCT 2: MVP of data visualization dashboard



Data Source

These data were obtained from www.kerala.data.gov for the years 1987 to 2015. Some of the years in between were not available. The datasets were concatenated and cleaned for visualization clarity. You can use the dropdown bar to filter for different cities.

The graphs currently look slightly chaotic as the values have not been treated for yearly averaged readings. Hover over the lines to interactively track the dates. As the reported dates varied in formats, there may have been misinterpretation errors when preparing the data.

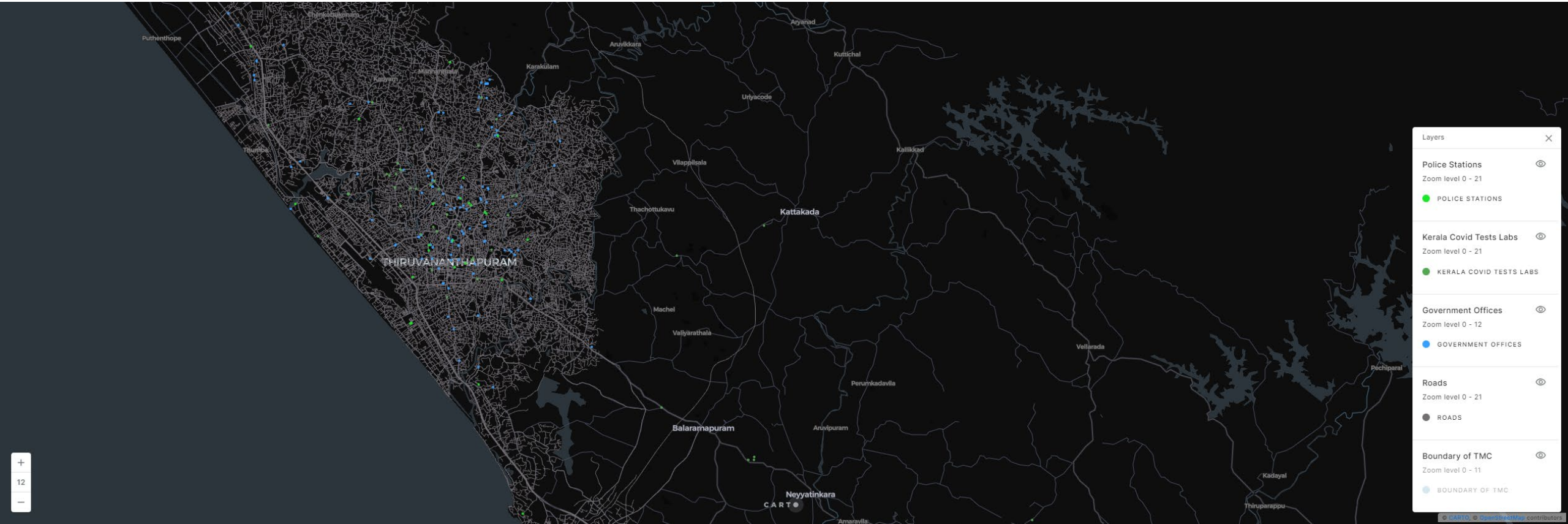
This page should be used as a proof-of-concept demo and if further analysis is required, data quality would need to be further improved. Data is filtered for year 2000 onwards due to inconsistent data reporting in earlier years.

Snip of live dashboard that visualizes Air Quality of different cities within Kerala.

Live link here: <https://kerala-air-quality-dee67b649d5a.herokuapp.com/>

WEB MAPPING FOR ANALYSIS

DATA PRODUCT 3: MVP of web-based geo analysis

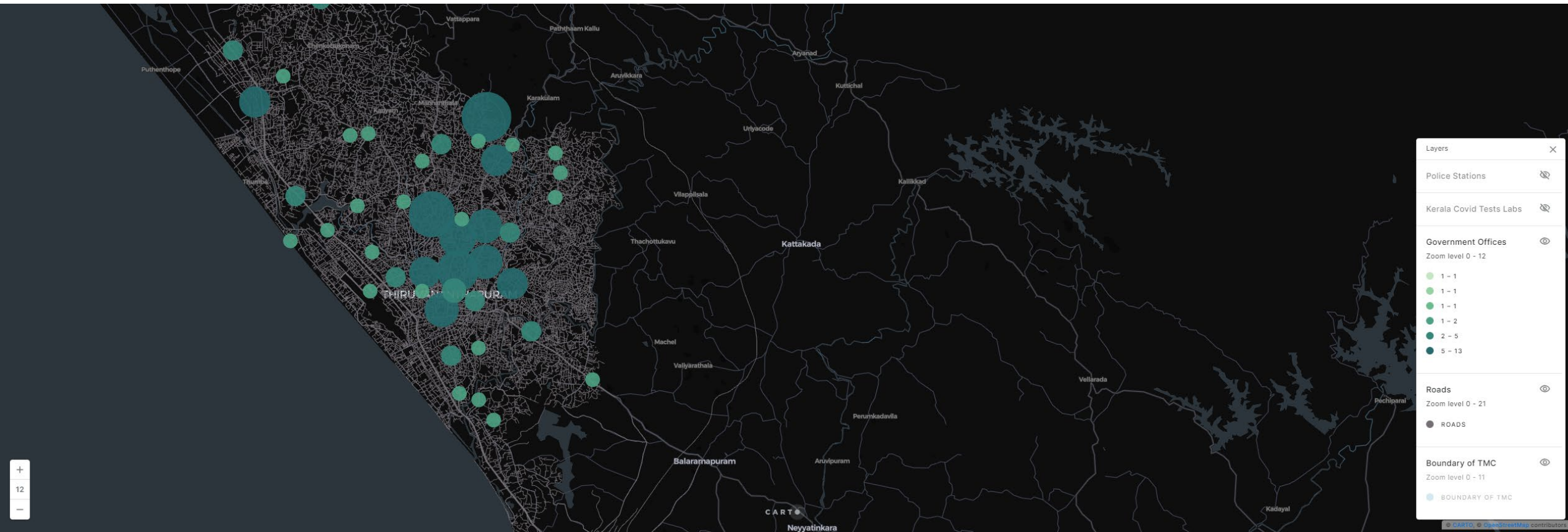


Map of Thiruvananthapuram Municipal Corporation

Live link here: <https://thunbergii.app.carto.com/map/30b39ddb-0000-4188-890c-bc8a712395b2>

WEB MAPPING FOR ANALYSIS

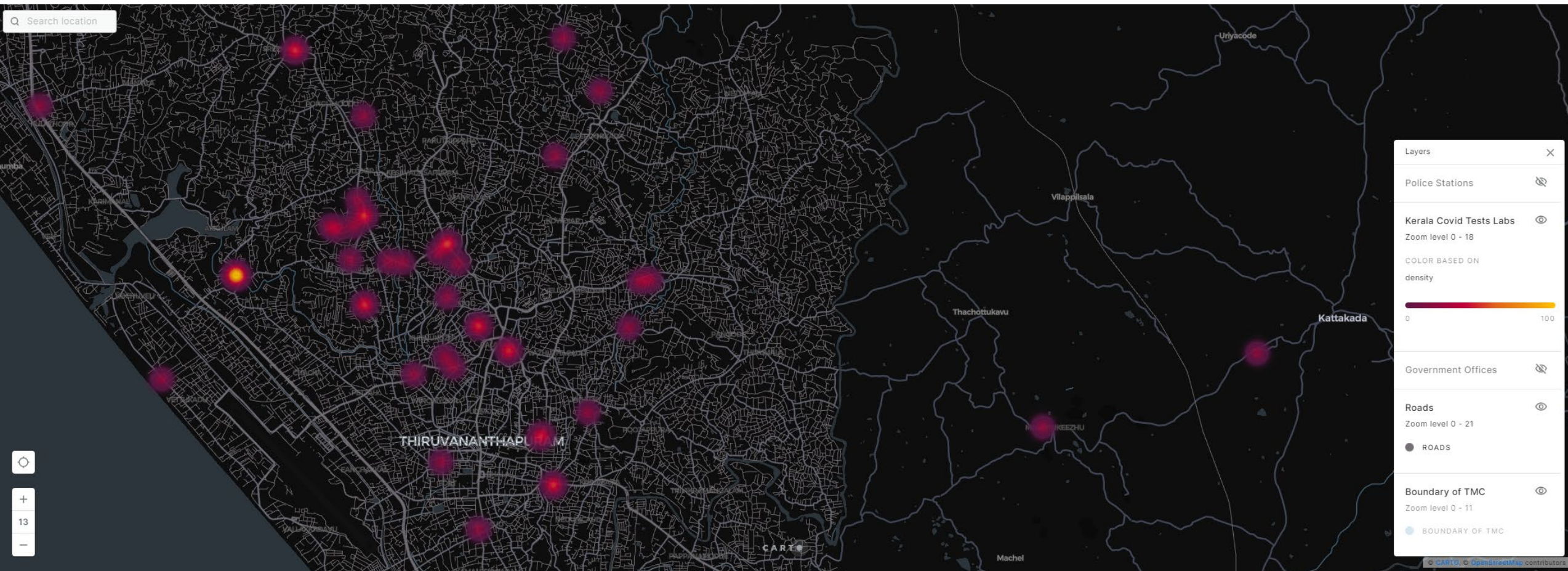
DATA PRODUCT 3: MVP of web-based geo analysis



Clusters, Police Stations

WEB MAPPING FOR ANALYSIS

DATA PRODUCT 3: MVP of web-based geo analysis



Heatmap of Kerala Covid Test Labs