



ລານບ້ອຍກລອຍໃຈ

Data Science Project

2110446 Data Science and Data Engineering (2022/2)

- Pras Pitasawad 6330305921
- Panithi Khamwangyang 6330301321
- Bhuribhat Ratanasanguanvongs 6330440921



Agenda

3 Project Objective

12 Visualization part

4 Dataset

19 Interesting Insight

5 Pipeline Overview

20 Demonstration

6 DE part - Airflow

9 ML part - Clustering



Project Objective

บริเวณชุมชนของผู้มีรายได้น้อยมีความสัมพันธ์กับบริเวณที่มีปัญหาประเภทต่างๆ หรือไม่ ?

Dataset

Traffy fondue

- Problem reported by user with problem coordinates, problem's types, and problem status
 - Dynamic
 - JSON -> convert to CSV

url : <https://publicapi.traffy.in.th/share/teamchadchart/search>

ส่วนที่ 1 ของข้อมูลทั้งหมดที่ได้รับจาก Google Cloud Storage คือ

```
type,org,description,ticket_id,coords,photo_url,after_photo,address,timestamp,problem_type_abdul,star,count_r  
สอบความ,กรุงเทพมหานคร,สอบความครับ มันใช่หน้าที่ของเทศกิจมั้ยครับ ที่ต้องมาเก็บป้ายหาเสียงของบรรดาพ่อค้าแม่ค้าเมืองครับ,2023-3DEYQ7,[ 'storage.googleapis.com/traffy_public_bucket/attachment/2023-05/844f7cc80c8578aca2eef202b167bc854cffd365.jpg,,  
กรุงเทพมหานคร 10600 ประเทศไทย,2023-05-18 05:44:27.850942+00,['สอบความ'],,0,,รอรับเรื่อง,2023-05-18 05:44:27.842205-  
.....
```

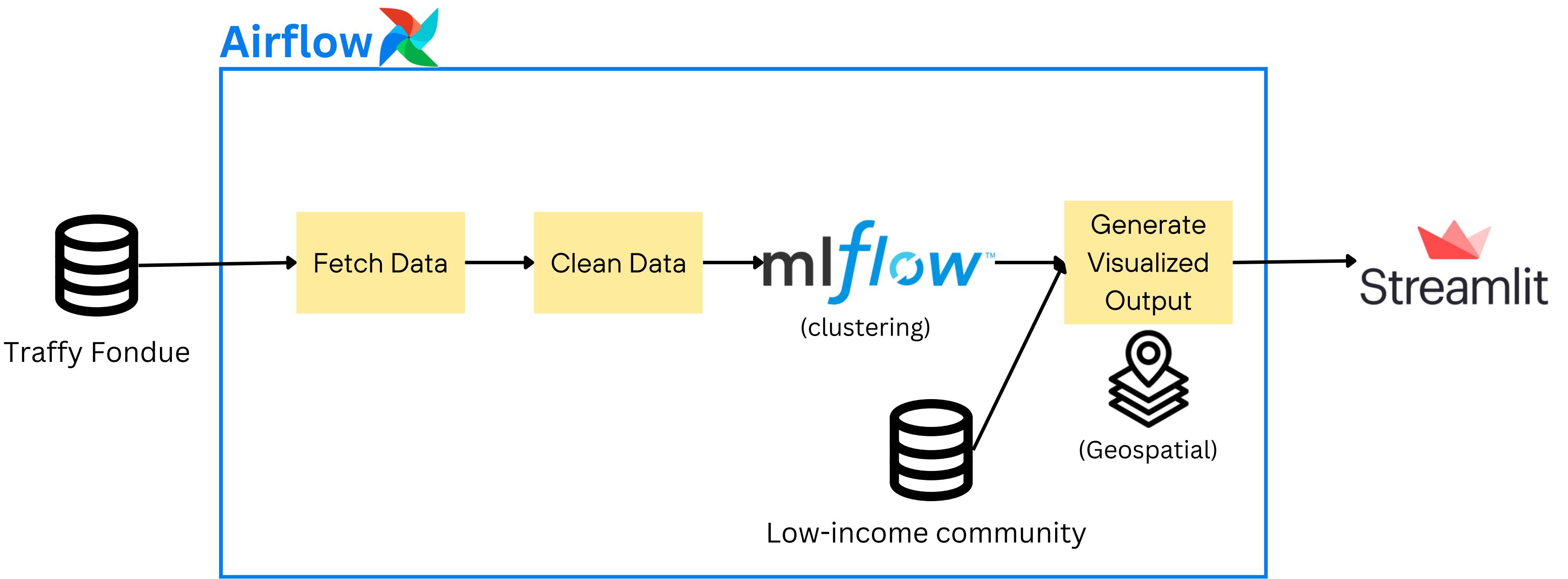
Low-income Community

- Low-income communities' description with coordinates
 - Static
 - CSV

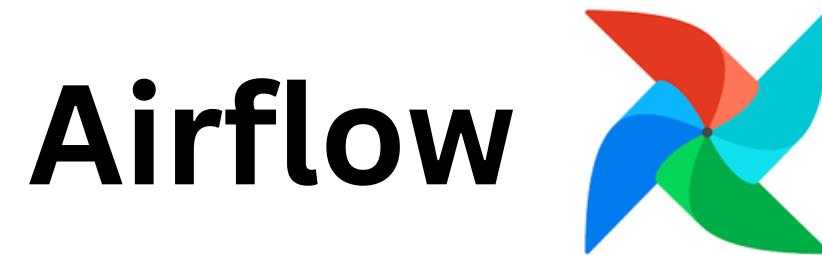
url : https://gdcatalog.nha.co.th/dataset/dataset11_01

COMM_ID	COMM_...	SHAPE_...	X	Y	Lat	Long	PROV_...	PROV_...	AMP_C...
100002	ริมคลองก...	2562	677065....	1520645...	13.749692	100.637...	10	กรุงเทพม...	6
100003	สาสาลีพ...	2562	678292....	1520800...	13.751023	100.64903	10	กรุงเทพม...	6
100004	วงศ์สอม (บ...	2562	679393....	1521497...	13.75725	100.659...	10	กรุงเทพม...	6

Pipeline Overview

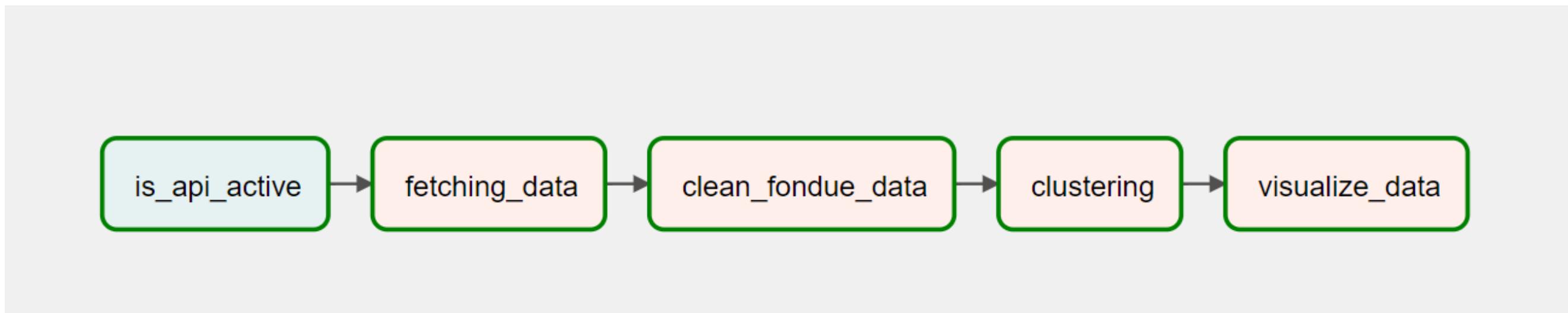


Data Engineering (DE)



- manage data pipeline
- monitor tasks in pipeline

DAG



Data Engineering (DE)

DAG: `fondu_dag` Visualize fondu data with poor people

Schedule: @daily | Next Run: 2023-05-18, 00:00:00

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details Code Audit Log

05/18/2023 07:20:46 AM 25 All Run Types All Run States Clear Filters

Auto-refresh

Duration May 14, 00:00

is_api_active fetching_data clean_fondu_data clustering visualize_data

defered failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status

» DAG `fondu_dag`

Details Graph

DAG Runs Summary

Total Runs Displayed	9
■ Total success	9
First Run Start	2023-05-18, 07:19:14 UTC
Last Run Start	2023-05-18, 07:19:16 UTC
Max Run Duration	00:02:03
Mean Run Duration	00:02:02
Min Run Duration	00:02:01

DAG Summary

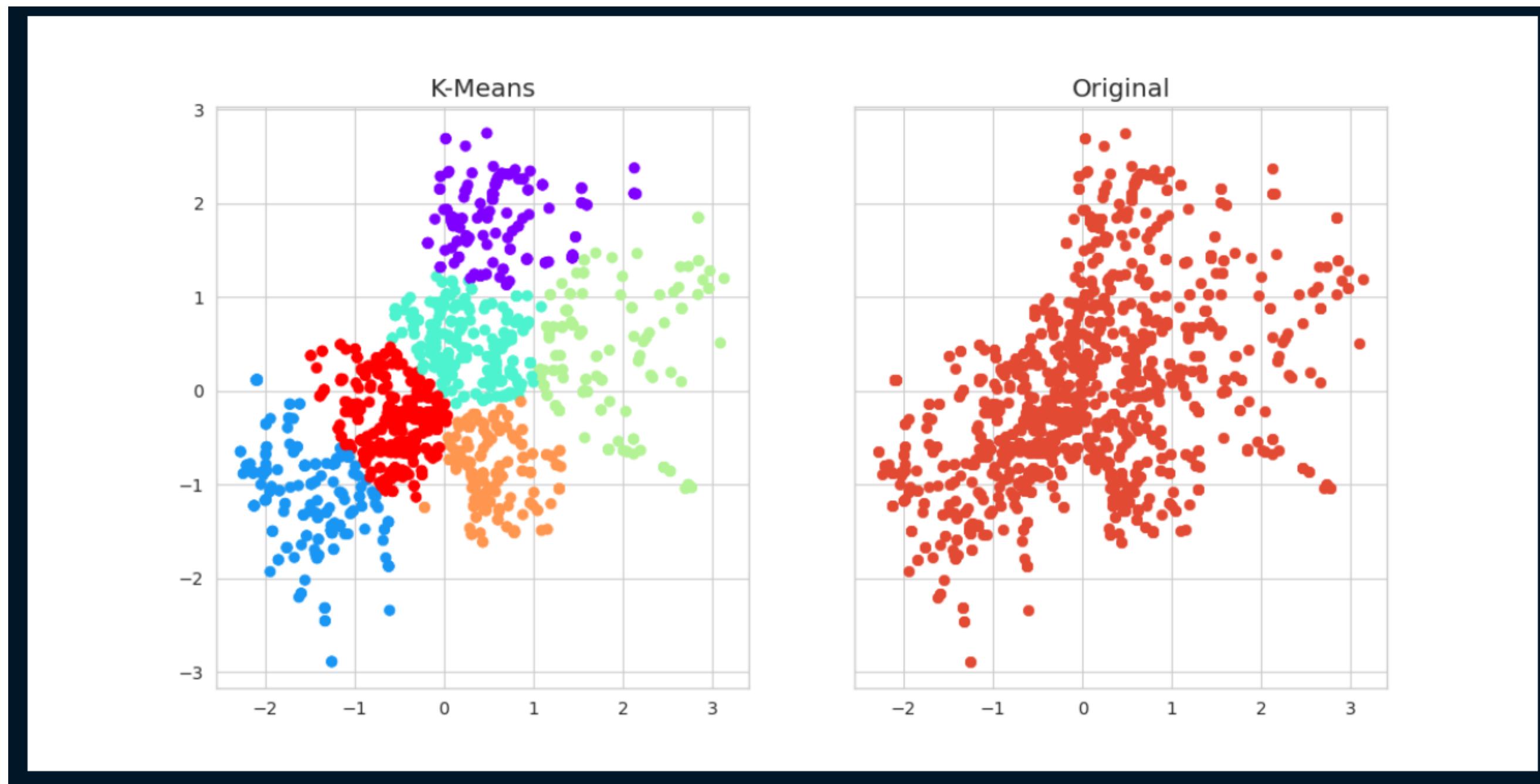
Total Tasks	5
HttpSensor	1
PythonOperators	4

Data Engineering (DE)



Machine Learning (ML)

K-Means Clustering



Machine Learning (ML)



finding optimal k using elbow method

mlflow 2.3.2 Experiments Models GitHub Docs

Experiments

Airflow DAG triggered: 2023-05-18 08:04:04.942629 [Provide Feedback](#)

Experiment ID: 1 Artifact Location: file:///opt/airflow/mlruns/1

Description Edit

Table view Chart view Sort: Created Refresh

Columns

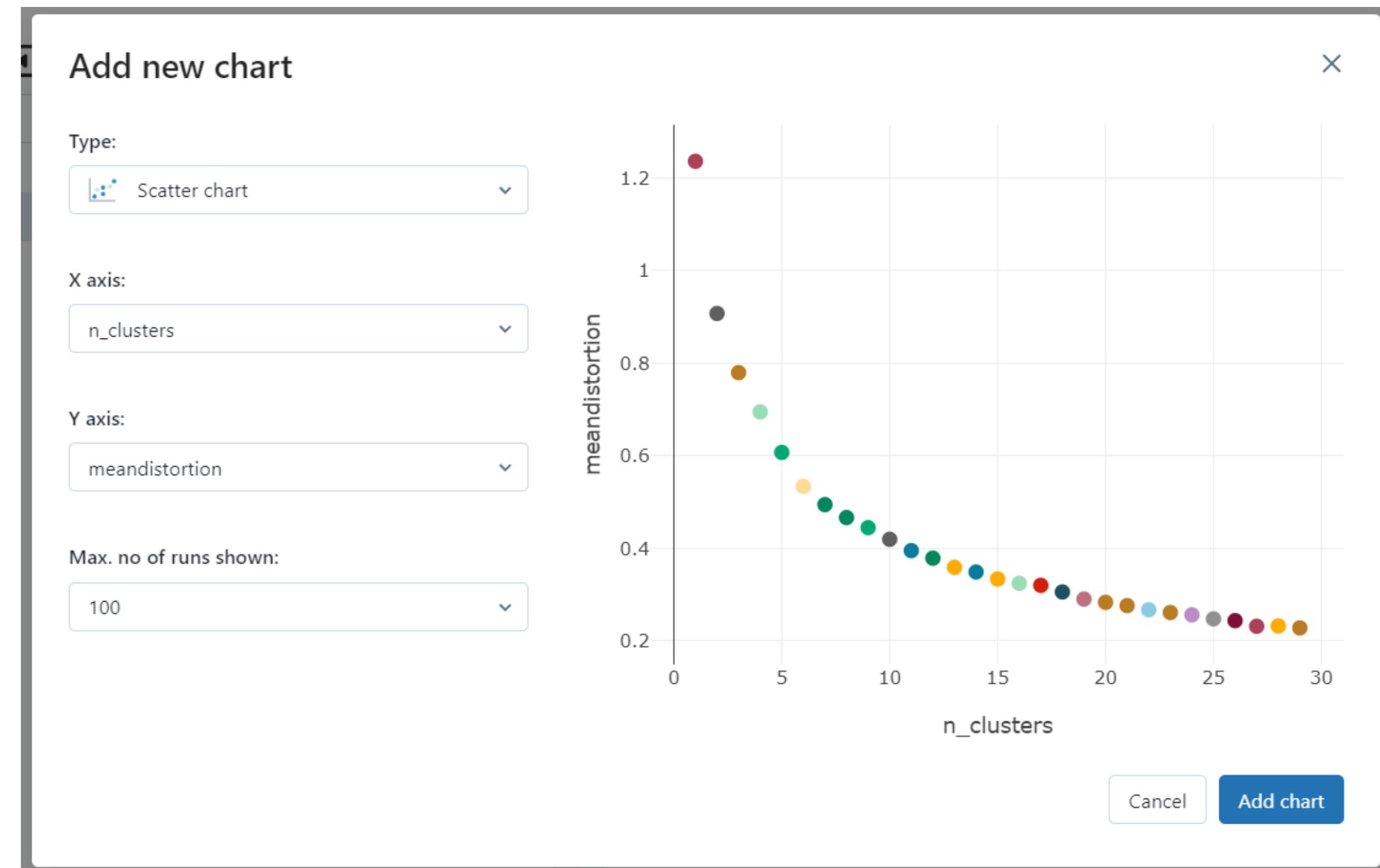
Time created: All time State: Active

	Run Name	Created	Duration	Source	Models
<input type="checkbox"/>	n_clusters = 26	36 minutes ago	3.2s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 25	36 minutes ago	3.1s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 24	37 minutes ago	3.4s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 23	37 minutes ago	3.2s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 22	37 minutes ago	3.8s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 21	37 minutes ago	3.2s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 20	37 minutes ago	3.1s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 19	37 minutes ago	3.1s	airflow	sklearn
<input type="checkbox"/>	n_clusters = 18	37 minutes ago	3.7s	airflow	sklearn

Machine Learning (ML)

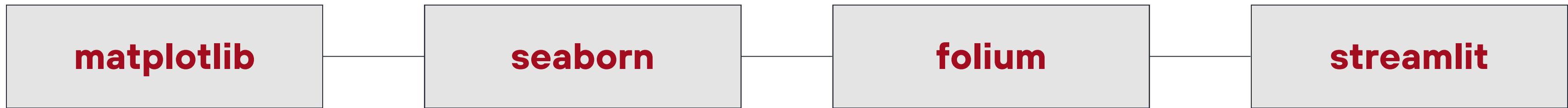


finding optimal k using elbow method



Visualization

using Geospatial and heatmap



Plot graph and save
figure in outputs

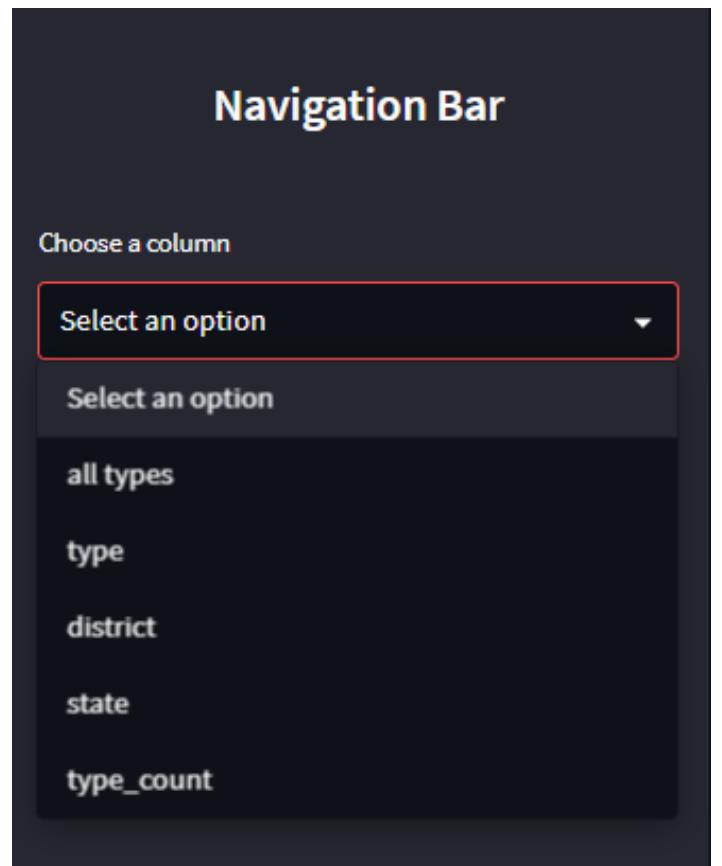
Plot graph and save
figure in outputs

Create interactive
map for geospatial
visualization

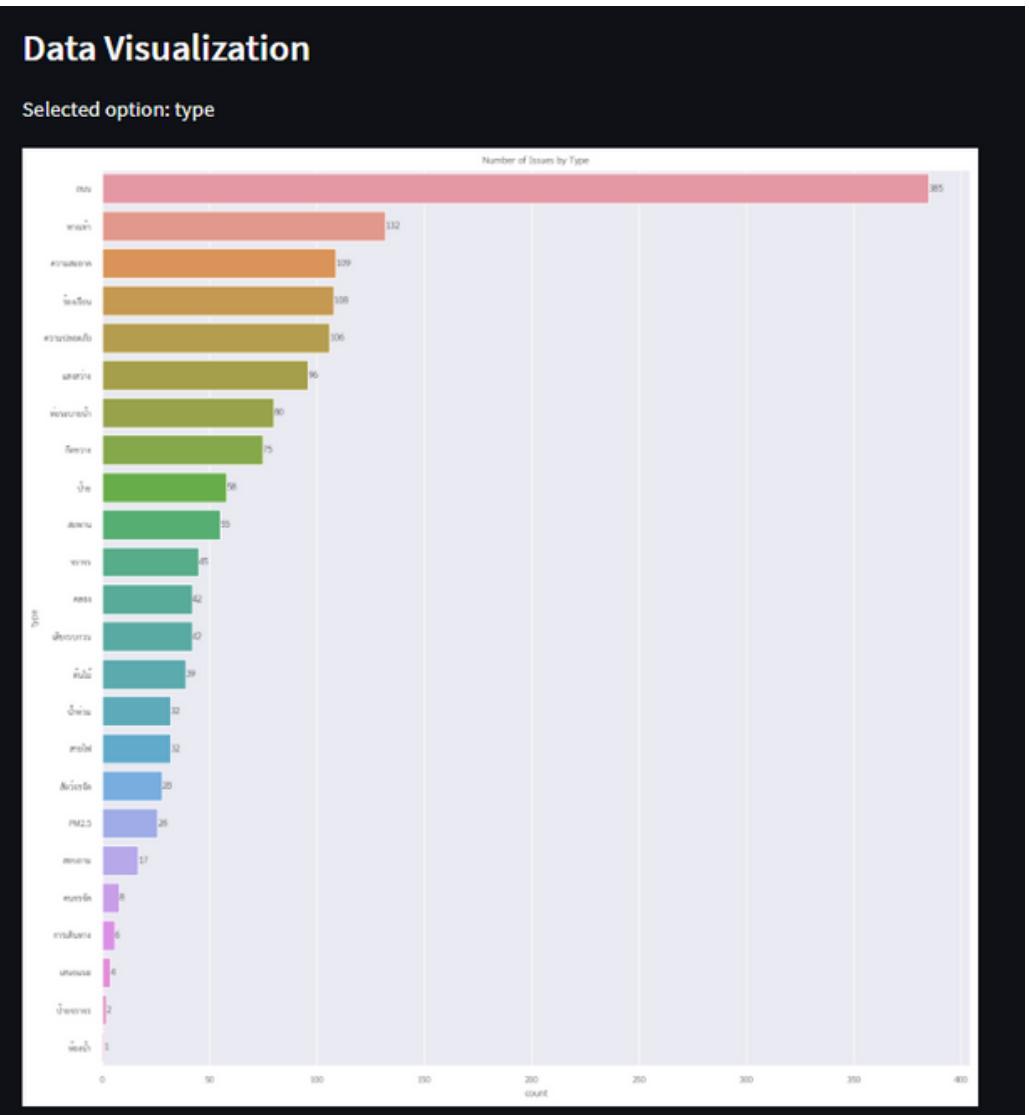
Create interactive
dashboard in webserver

Visualization

using Geospatial and heatmap



- Streamlit app in navigation bar
- Choose a column to display count plot



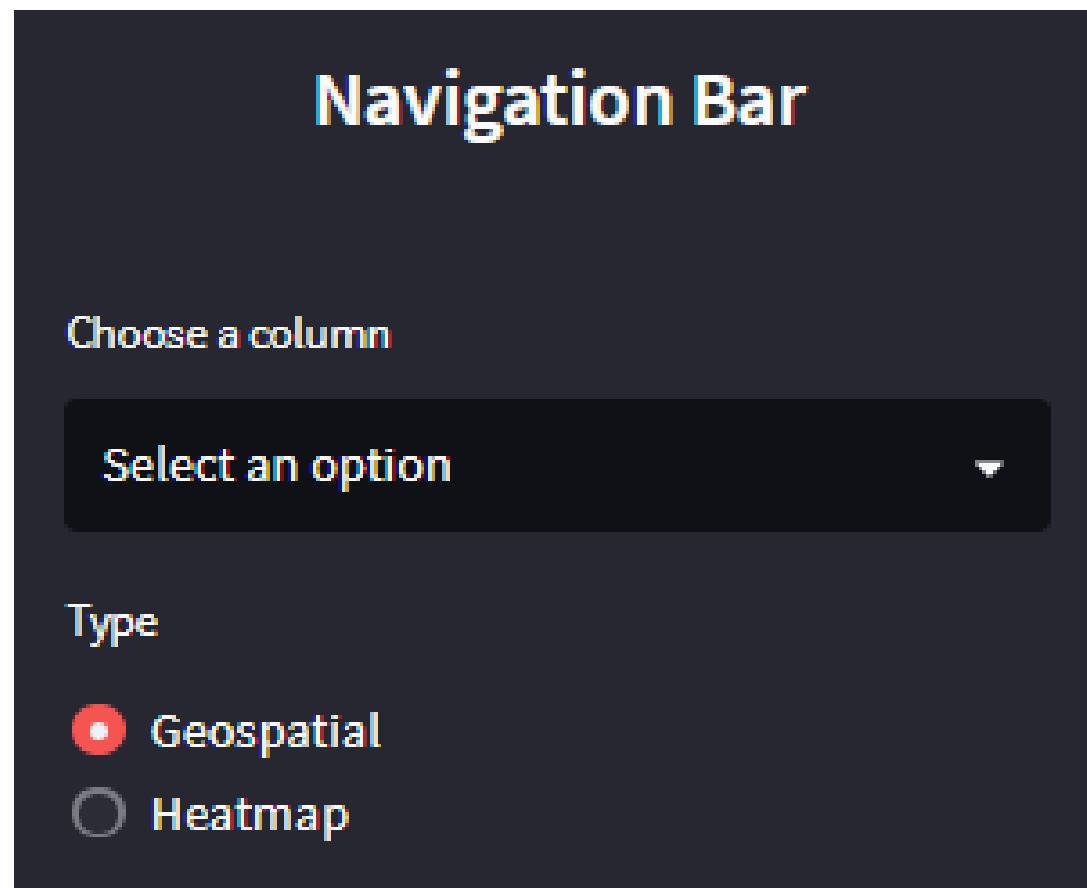
matplotlib

seaborn

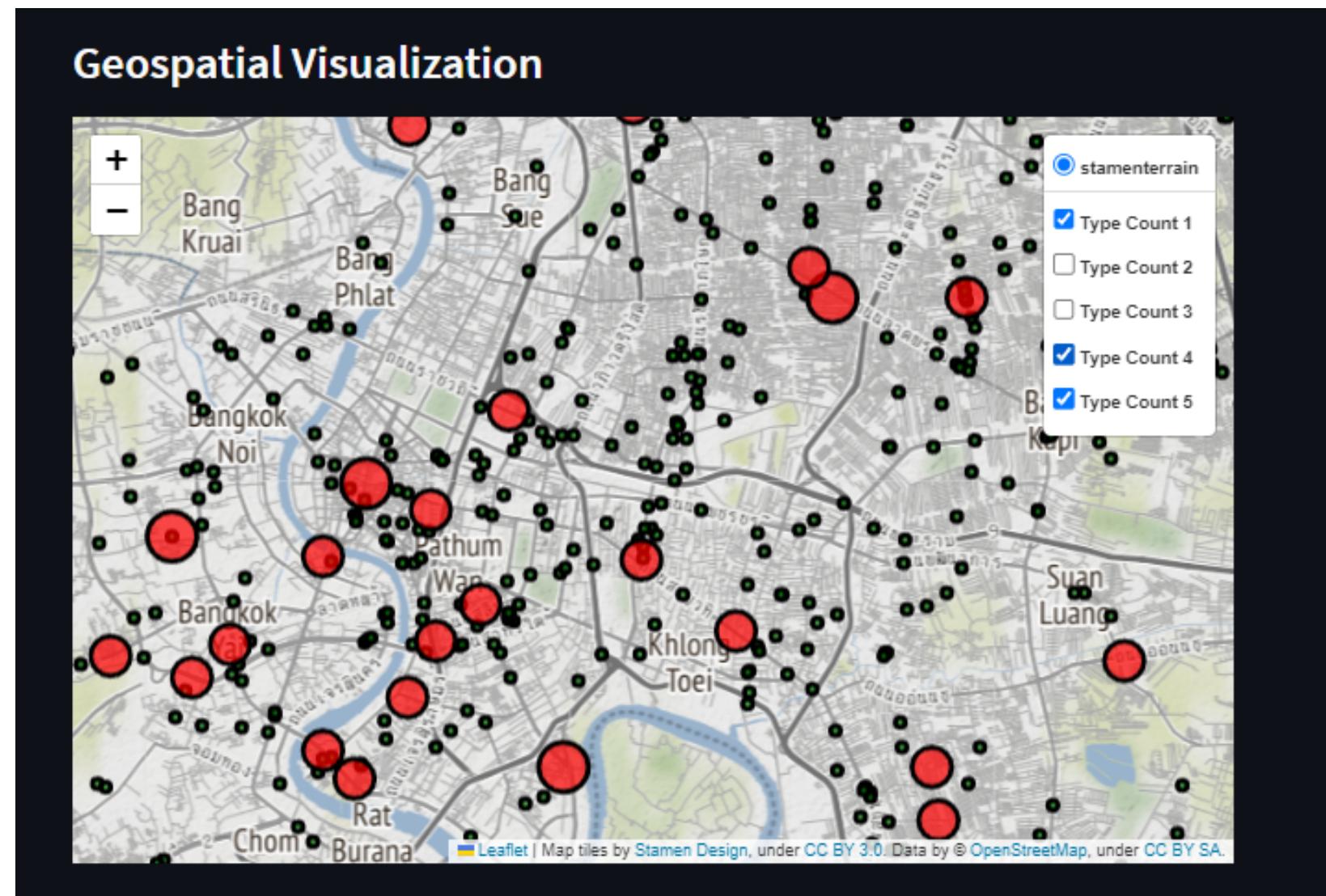
streamlit

Visualization

using Geospatial and heatmap



- Streamlit app in navigation bar
- Select Geospatial radio

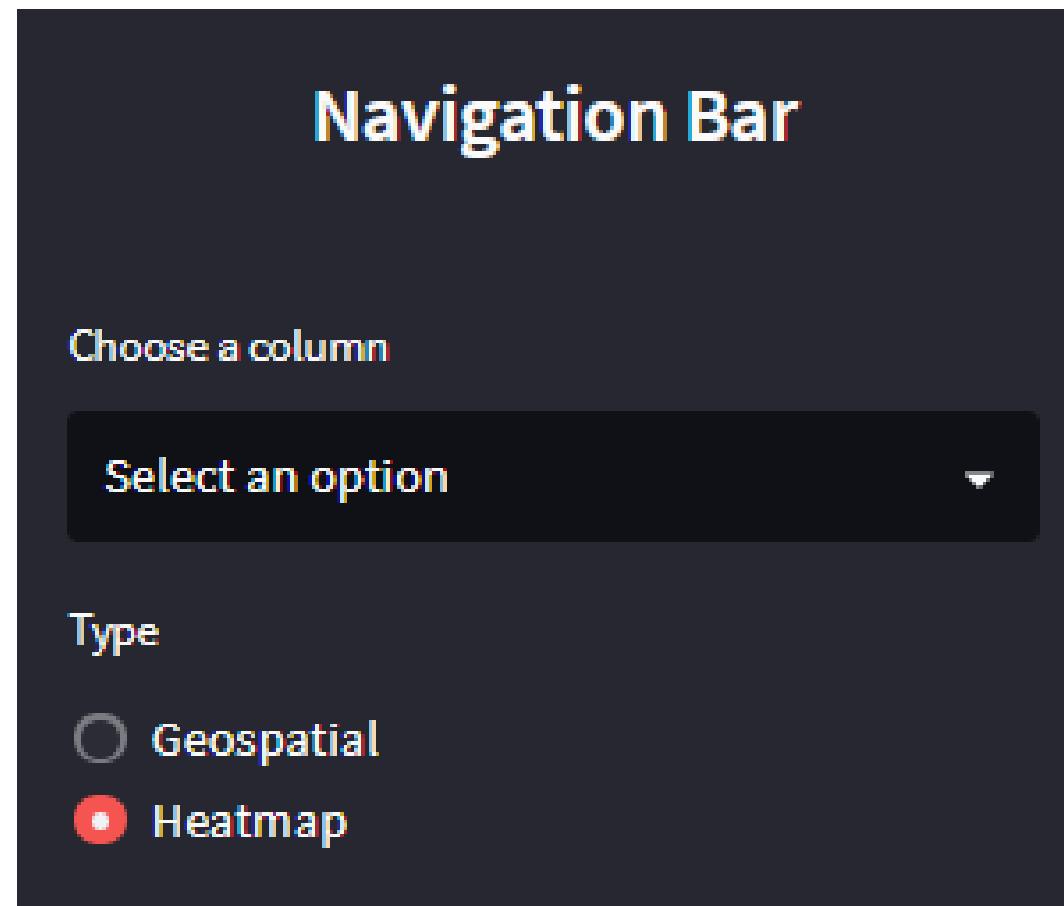


folium

streamlit

Visualization

using Geospatial and heatmap

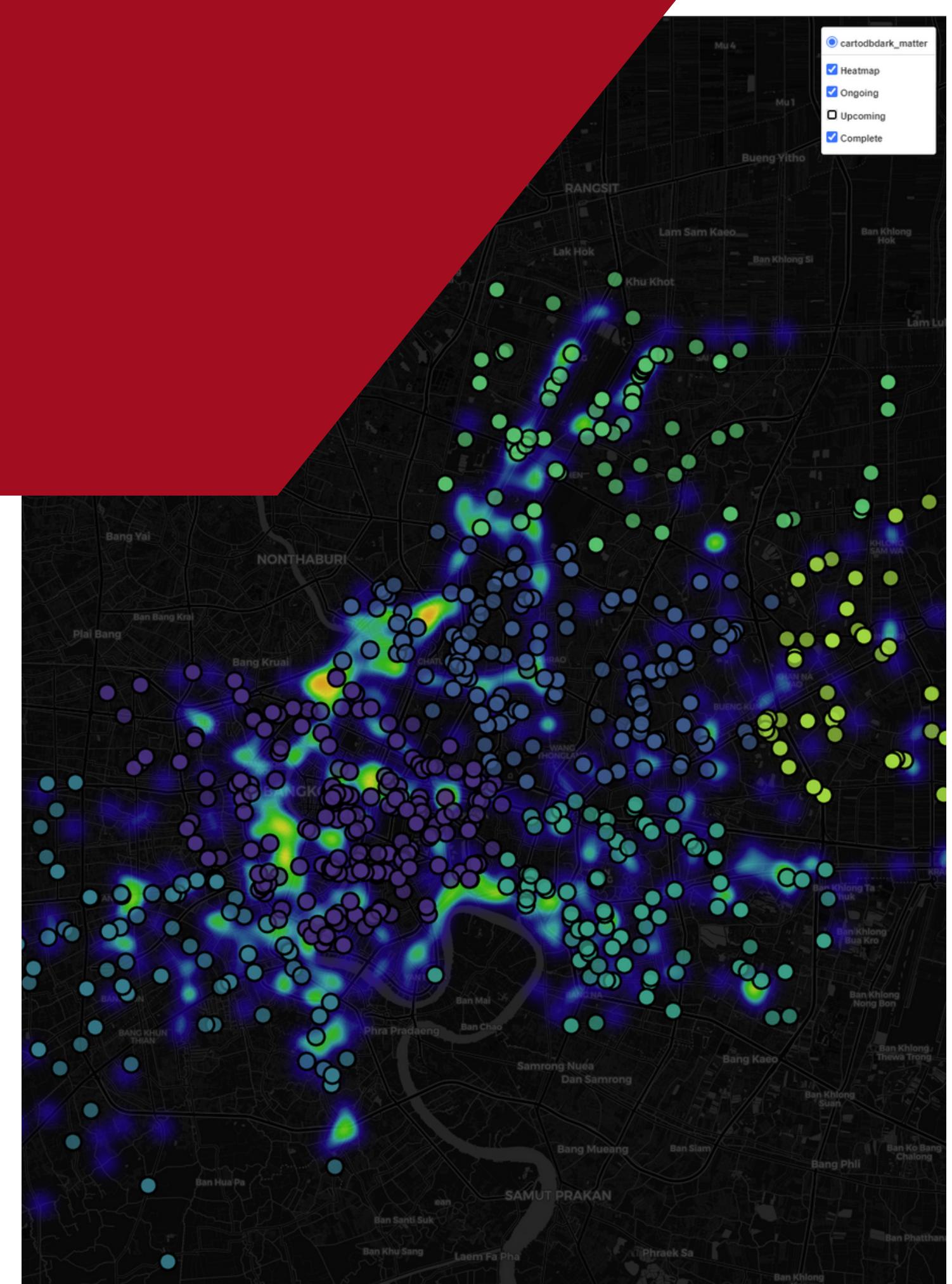
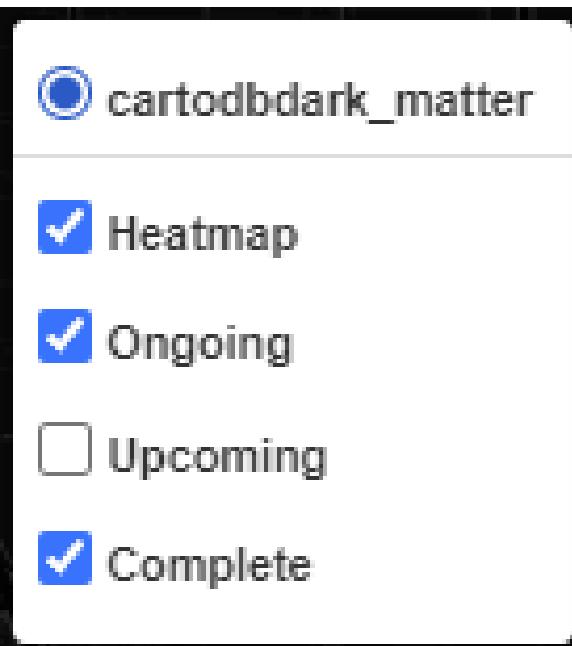


- Streamlit app in navigation bar
- Select Heatmap radio

Visualization

using Geospatial and heatmap

- Streamlit app in navigation bar
- Select Heatmap radio
- Filter in the control layer
 - heatmap
 - ongoing
 - upcoming
 - complete



Navigation Bar

Choose a column

Select an option

Type

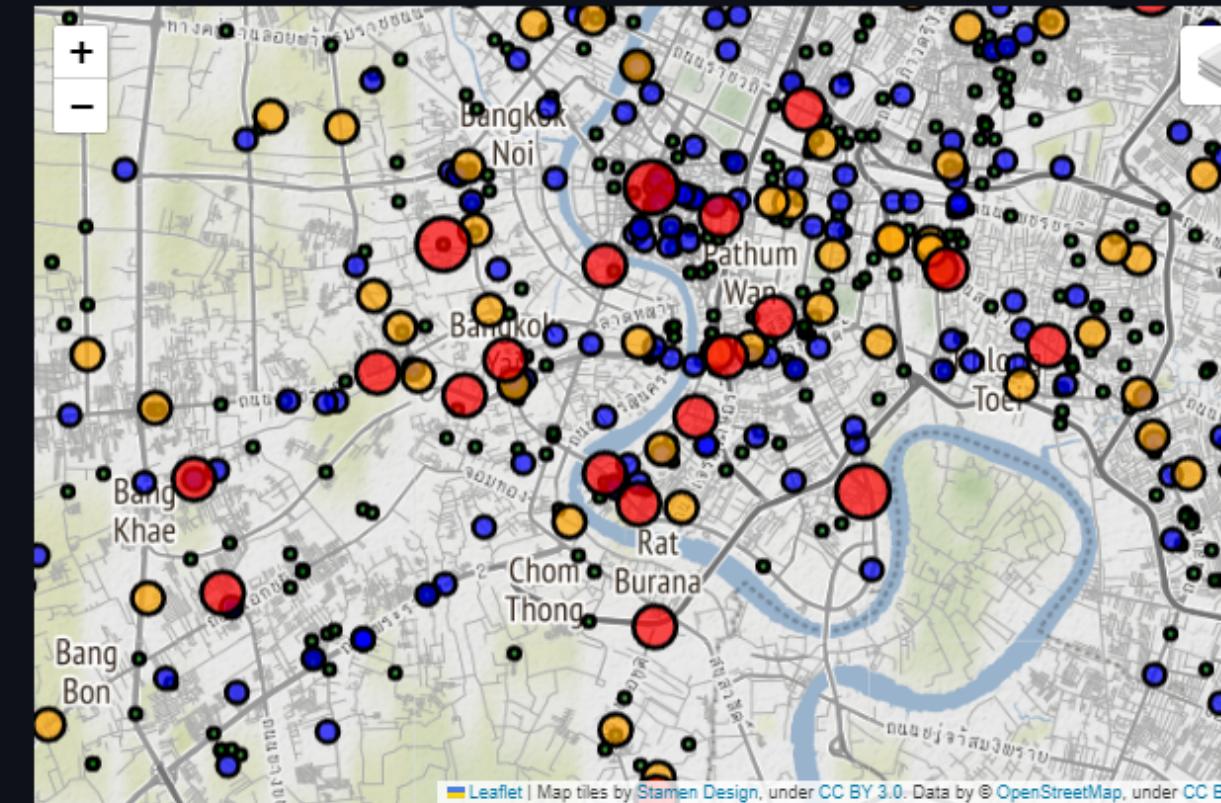
- Geospatial
- Heatmap

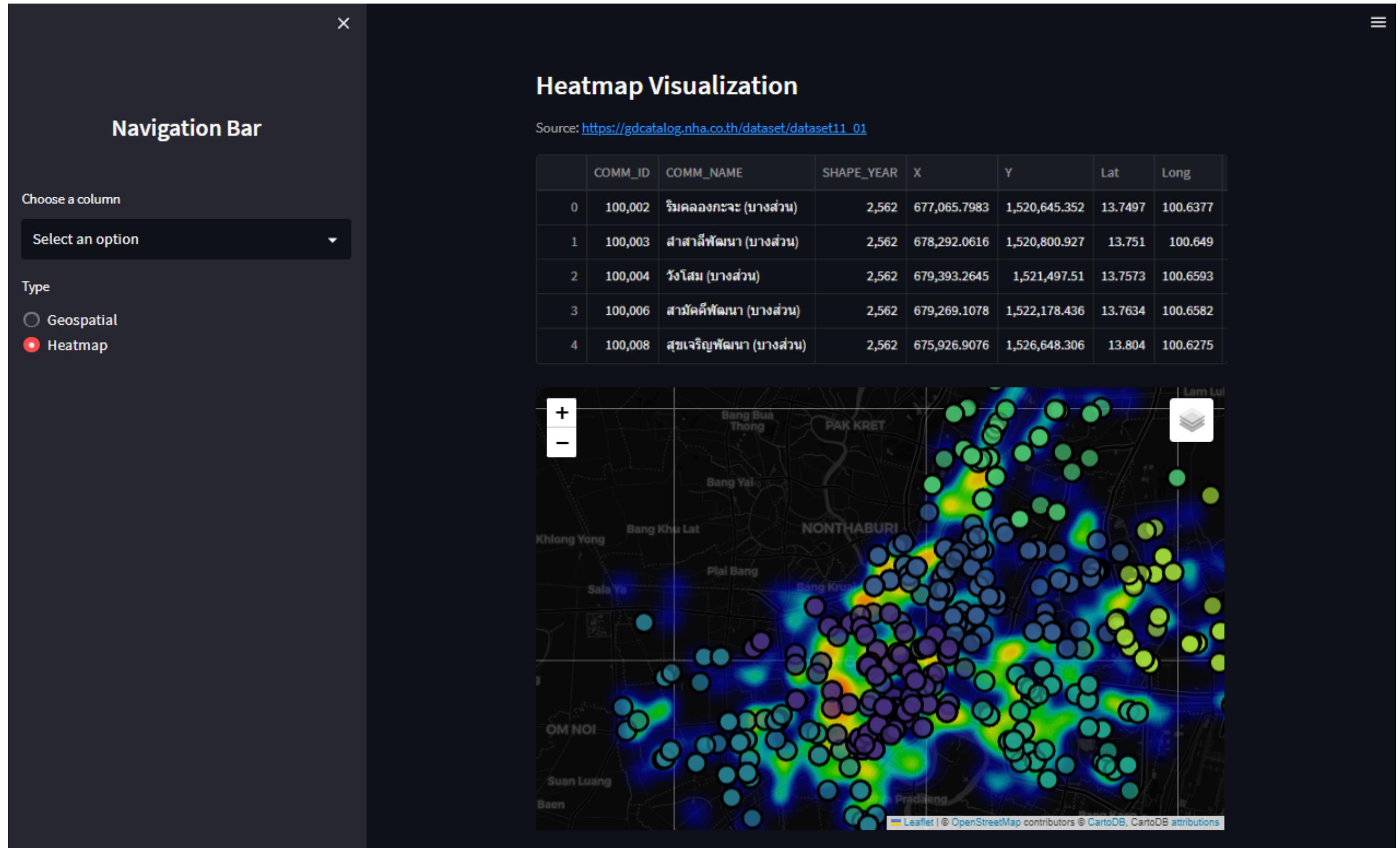
Traffy Fondué Clustering

Source: <https://share.traffy.in.th/teamchadchart>

	type	address
0	['ถนน']	29/11 ถ. เพชรเกษม แขวงหนองค้างพลู เขตหนองแขม กรุงเทพมหานคร 10160 ประเทศไทย
1	['ป้าย', 'การเดินทาง']	39 1 แขวง จตุจักร เขตจตุจักร กรุงเทพมหานคร 10900 ประเทศไทย
2	['ความสะอาด', 'ถนน']	34/63 ซอย เจริญกรุง 107 แยก 11 แขวงบางคอแหลม เขตบางคอแหลม กรุงเทพมหานคร 10150 ประเทศไทย
3	['ถนน']	ถนนศรีนครินทร์: 2998 ถ. พัฒนาการ แขวงสวนหลวง กรุงเทพมหานคร 10250 ประเทศไทย
4	['ถนน', 'ห้องน้ำยาไร้']	573/16 ถ. สุทธิสารวิวิจัย แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10400 ประเทศไทย

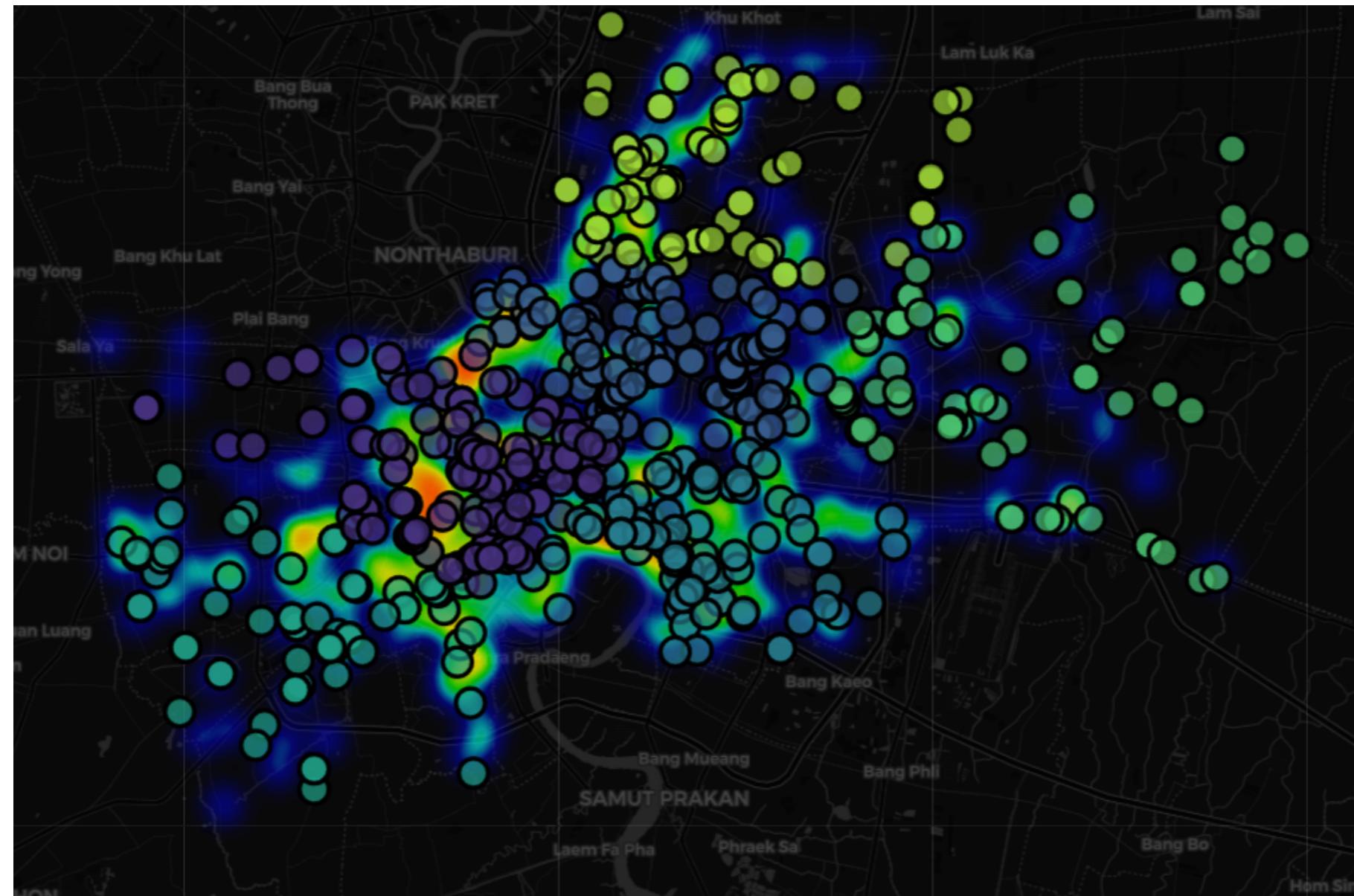
Geospatial Visualization



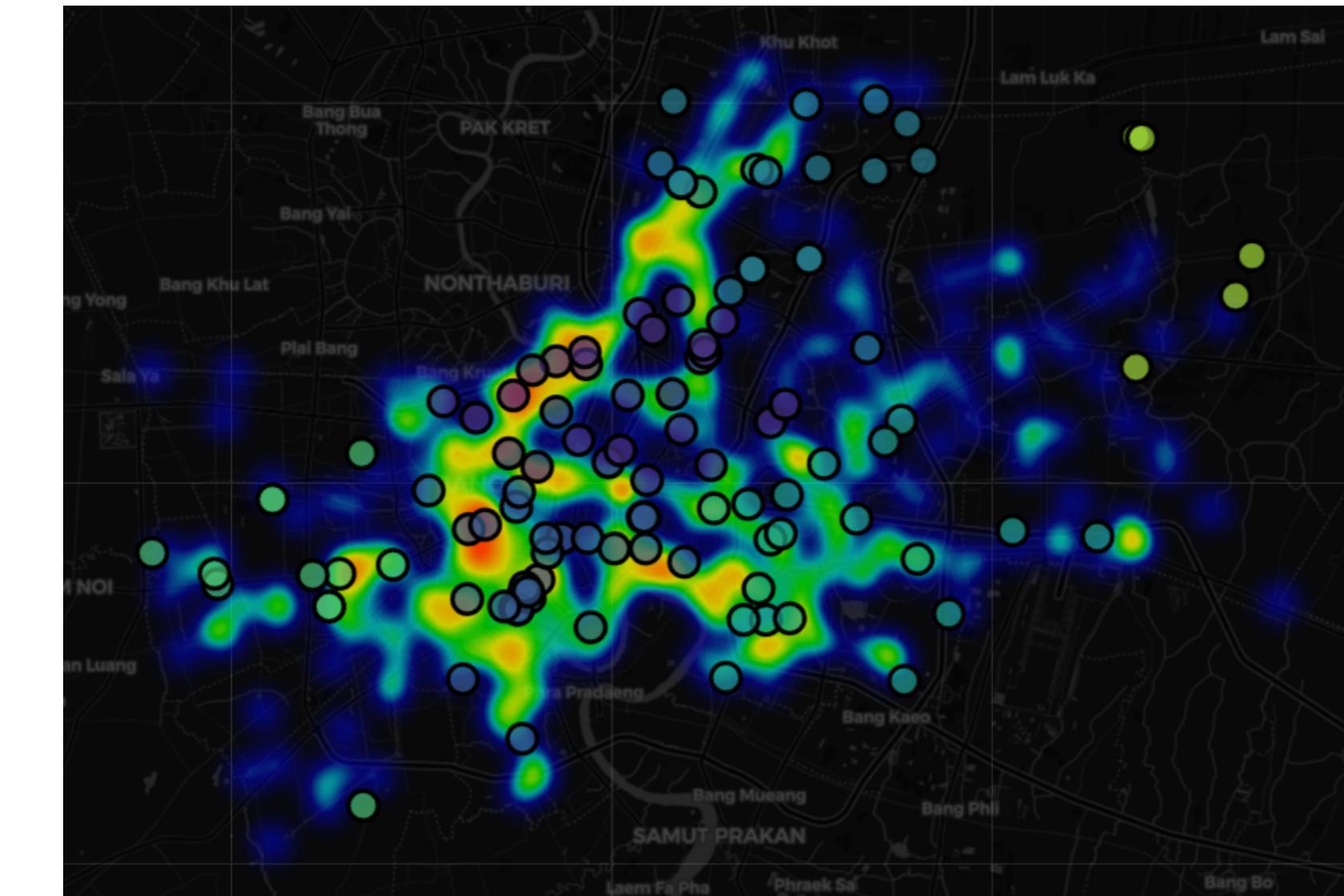


Insight

ถนน ทางเท้า ก่อระบายน้ำ vs low-income heatmap



ความสะอาด vs low-income heatmap



DEMONSTRATION





ลัมบ้อยกลอยใจ

Thank you

Group name: ลัมบ้อยกลอยใจ

Github: <https://github.com/Bhuribhat/Traffy-Clustering/tree/main>

