



# คาดการณ์เวลาที่ใช้ เพื่อแก้ไขปัญหาเสร็จสิ้น

Data Science and  
Data Engineering  
2110446 (2022/2)

# Member

Group : คอยคิง



6331308021

ชาบู

รัตนจรัสกุล

6331315321

ธีรโชค

เดชสุวรรณกิจ

6331337121

ภูรินทร์

แต่งศรีวรรณ

6331341621

วัชรวิชญ์

วิวัฒน์สุขไพศาล

# Objective

- เพื่อให้ผู้เจ้งปัญหาผ่าน Traffy fondu ทราบระยะเวลาที่ใช้ในการแก้ไขปัญหาอย่างคร่าวๆ ได้กันทีหลังแจ้งปัญหา โดยการใช้ ML ในการคำนวณ
- วิเคราะห์ข้อมูล และหา Insight ที่นำเสนอจากข้อมูล Traffy fondu

# Data

**Traffy Fondu** [https://www.traffy.in.th/?page\\_id=27351](https://www.traffy.in.th/?page_id=27351)

[สำหรับผู้นำข้อมูลไปพัฒนาต่อ�อด](#)

สามารถขอข้อมูลในระบบ เพื่อนำไปจัดทำสถิติหรือ  
วิเคราะห์ข้อมูล ผ่าน API (ข้อมูลจะอัปเดตทุกๆ 60 นาที)

หมายเหตุ : หากท่านได้มีการนำข้อมูลหรือผลการวิเคราะห์ไปเผยแพร่ต่อ ขอความ  
กรุณาใส่ลิงค์ [traffy.in.th](http://traffy.in.th) หรือหากเป็นรูปภาพ กรุณาใส่รูปໄกีของ Traffy  
Fondues ด้วย เพื่อเป็นการระบุแหล่งข้อมูลต้นทาง และเป็นการให้เครดิตกับพัฒนา<sup>者</sup>  
[Traffy Fondues /ขอบคุณครับ](#)

**Resource URL**

**JSON format** (อัปเดตทุกๆ 60 นาที)  
<https://publicapi.traffy.in.th/share/teamchadchart/search>

**CSV format** (อัปเดตทุกๆ 60 นาที)  
<https://publicapi.traffy.in.th/share/teamchadchart/download>



หมายเหตุ : หากต้องการข้อมูลกึ่งหมด สามารถดาวน์โหลดไฟล์ CSV ได้ที่นี่ (ขนาดไฟล์  
ประมาณ 280 MB)

**CSV format** (อัปเดตทุกๆ 3:00 u.)  
[https://publicapi.traffy.in.th/dump-csv-chadchart/bangkok\\_traffy.csv](https://publicapi.traffy.in.th/dump-csv-chadchart/bangkok_traffy.csv)

# Data

ticket_id	type	organization	comment	photo	photo_after	coords	address	subdistrict
2021-9LHDHM6	{}		ไม่มีภาพ	<a href="https://storage.googleapis.com/trashbin/100.4866113.79386">https://storage.googleapis.com/trashbin/100.4866113.79386</a>			1867 จรัญสนิทวงศ์ แขวง บางพลัด	
2021-FYJTFP	{ความสะอาด}	เขตบางซื่อ	ขยะเย袍ะ	<a href="https://storage.googleapis.com/trashbin/100.5308413.81865">https://storage.googleapis.com/trashbin/100.5308413.81865</a>			12/14 ถนน กรุงเทพฯ- นนทบุรี แขวง บางซื่อ เขต นนทบุรี	
2021-8GKAR9	{สายไฟ}	ยังไม่มีหน่วยงานรับผิดชอบ	1. เก้าวัลย์งดบนสายไฟ ผูกกลัวว่า <a href="https://storage.googleapis.com/trashbin/100.5768513.79704">https://storage.googleapis.com/trashbin/100.5768513.79704</a>				335/31 ลาดพร้าว แขวง สามเสนนอกราช	
			1 ช่องทางท่านหลังศินพื้นที่จากการปรับปรุงสะพานลอย 2 3 ปากซอย สาทร 12 ถนน สาทร เหนือ					
2021-AFPUXZ	{ถนน,สะพาน}		น้ำท่วม เนื่องจากน้ำท่วมแล้วฝนตกและทะลุเข้าบ้าน	<a href="https://storage.googleapis.com/trashbin/100.5291613.72338">https://storage.googleapis.com/trashbin/100.5291613.72338</a>			37 10 ซอย สีลม 9 แขวง สีลม สีลม	
2021-CGPMUN	{น้ำท่วม,ร่องเรียน}	เขตประเวศ,ฝ่ายโยธา เขตประเวศ	น้ำท่วมเวลาฝนตกและทะลุเข้าบ้าน	<a href="https://storage.googleapis.com/trashbin/100.6670913.67891">https://storage.googleapis.com/trashbin/100.6670913.67891</a>			189 เจริญพระเกียรติ ๑.๙ แขวง หนองบอน	
2021-8Z8JQ3	{ถนน,ร่องเรียน,น้ำท่วม,ความปลอดภัย}		1) พื้นที่น้ำข้ามหน่วยงานรับผิดชอบ เช่น ถนนสีลม บริเวณแยกสีลม-สุขุมวิท	<a href="https://storage.googleapis.com/trashbin/100.8619013.72804">https://storage.googleapis.com/trashbin/100.8619013.72804</a>			3/8 บุนทาง- ล่าต่ออยู่ดึง แขวง บุนทาง	
2021-GAFN8W	{คลօป}		น้ำท่วม เนื่องจากน้ำท่วมแล้วฝนตกและทะลุเข้าบ้าน	<a href="https://storage.googleapis.com/trashbin/100.6229813.74227">https://storage.googleapis.com/trashbin/100.6229813.74227</a>			50 พระราม ที่ 9 แขวง สาทรใต้ สาทร	
2021-AFJYNF	{ถนน,น้ำท่วม}		น้ำท่วม เอ่อล้นถนน น้ำระบายน้ำไม่ทัน	<a href="https://storage.googleapis.com/trashbin/100.7159213.73614">https://storage.googleapis.com/trashbin/100.7159213.73614</a>			3 ถนน พัฒนาชัยนาท แขวง คลองสองดันบุน	
2021-HAJULK	{ร่องเรียน,ถนน,น้ำท่วม}		มีการยกถนนในซอยอยู่เบื้องหลัง ถนนสีลม	<a href="https://storage.googleapis.com/trashbin/100.7166113.79225">https://storage.googleapis.com/trashbin/100.7166113.79225</a>			95 ซอย 12 มีนบุรี กรุงเทพมหานคร มีนบุรี	
2021-Q7XMCN	{}		งานไม่เสร็จ ควรเก็บท่อHDPE ทั้งหมด	<a href="https://storage.googleapis.com/trashbin/100.5077913.79675">https://storage.googleapis.com/trashbin/100.5077913.79675</a>			1 ถนน จรัญสนิทวงศ์ แขวง บนางอ้อ	

district	province	timestamp	state	star	count_reopen	last_activity
บางพลัด	กรุงเทพมหานคร	2021-09-01 10:44:55.353	กำลังดำเนินการ			2022-02-22 04:59:58.622268+00
บางซื่อ	กรุงเทพมหานคร	2021-09-03 12:51:09.453	เสร็จสิ้น			2022-06-04 15:34:14.609206+00
ห้วยขวาง	จังหวัดกรุงเทพฯ	2021-09-19 06:47:50.488	กำลังดำเนินการ			2022-02-22 04:30:07.293416+00
บางรัก	กรุงเทพมหานคร	2021-09-19 07:40:30.000	กำลังดำเนินการ			2022-02-22 04:30:04.507406+00
ประเวศ	กรุงเทพมหานคร	2021-09-19 14:56:08.924	เสร็จสิ้น	4		2022-06-21 08:21:09.532782+00
ลาดกระบัง	กรุงเทพมหานคร	2021-09-20 00:24:47.881	กำลังดำเนินการ			2022-02-22 04:29:57.80173+00
สวนหลวง	กรุงเทพมหานคร	2021-09-21 05:22:55.500	กำลังดำเนินการ			2022-02-22 04:29:54.729153+00
ลาดกระบัง	กรุงเทพมหานคร	2021-09-22 05:06:09.308	กำลังดำเนินการ			2022-02-22 04:29:51.833505+00
มีนบุรี	กรุงเทพมหานคร	2021-09-23 06:25:55.715	กำลังดำเนินการ			2022-02-22 04:29:49.030483+00
บางพลัด	กรุงเทพมหานคร	2021-09-24 04:15:40.312	กำลังดำเนินการ			2022-02-22 04:29:46.037649+00

# Assumption

- timestamp คือเวลาที่ปัญหาถูกเพิ่มเข้ามาในระบบ
- last\_activity คือเวลาที่ปัญหาถูกเปลี่ยนแปลง เช่น เจ้าหน้าที่เข้ามาอัพเดตข้อมูลใดๆ ให้ผู้แจ้งปัญหาทราบ
- state ประกอบด้วย "รอรับเรื่อง" "กำลังดำเนินการ" "เสร็จสิ้น"
- state เปลี่ยนแปลงจาก "รอรับเรื่อง" → "กำลังดำเนินการ" → "เสร็จสิ้น" ในทิศทางนี้เท่านั้น



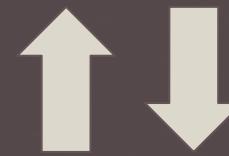
# Traffy Fondué Data from API

DE  
**Spark**

Visualization  
**Power BI**  
↓  
**spatial & time**



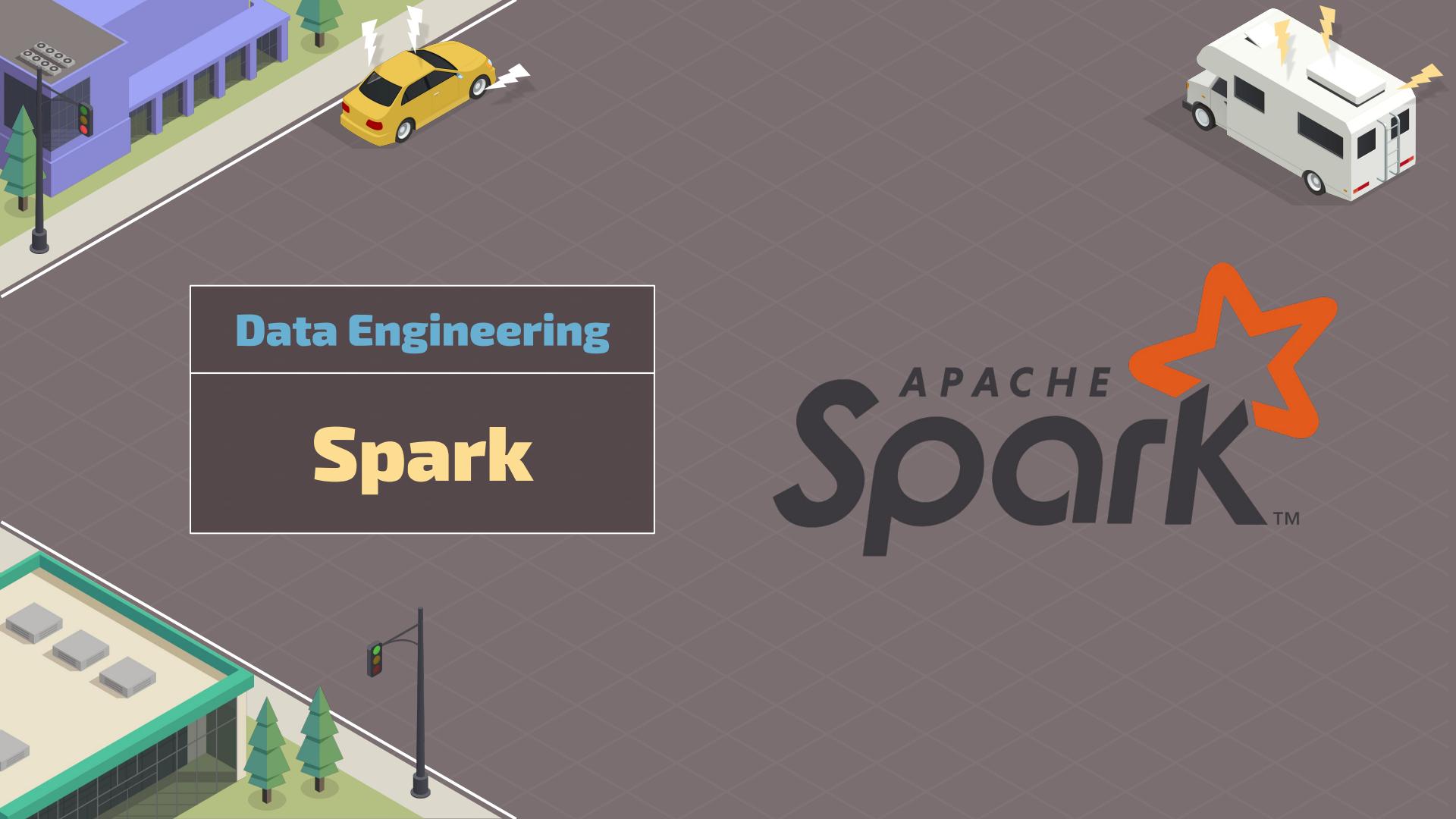
AI/ML  
**Traditional ML**



AI/ML  
**MLflows**

↓  
**FastAPI**





Data Engineering

Spark



# Data Preparation

## Spark SQL Data Preparation

```
[ ] from pyspark import SparkFiles  
  
# Data from https://www.traffy.in.th/?page\_id=27351  
# Update every 3hr  
  
# url = 'https://publicapi.traffy.in.th/dump-csv-chadchart/bangkok\_traffy.csv'  
# spark.sparkContext.addFile(url)  
  
# df = spark.read.option("delimiter", ",").option("multiline", "true").option("quote", "'").option("header", "true").option("escape", "\\").option("escape", "'").csv("file://" + SparkFiles.get("bangkok_traffy.csv"))  
# df.show()  
  
[ ] # download data from google drive  
# https://drive.google.com/file/d/1Fd1R\_ZUB1RGuYwx4AH52u1VKWEbJpV1L/view?usp=share\_link  
# data at 12-May-2023 21:46  
!gdown 1Fd1R_ZUB1RGuYwx4AH52u1VKWEbJpV1L  
  
Downloading...  
From: https://drive.google.com/uc?id=1Fd1R\_ZUB1RGuYwx4AH52u1VKWEbJpV1L  
To: /content/bangkok_traffy.csv  
100% 357M/357M [00:01<00:00, 209MB/s]
```

```

path = '/content/bangkok_traffy.csv'

df = spark.read.option("delimiter", ",").option("multiline", "true").option("quote", "'").option("header", "true").option("escape", "\\").option("escape", "'").csv(path)
df.show()

+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| ticket_id | type | organization | comment | photo | photo_after | coords | address | subdistrict |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 2021-9LHDM6 | {} | null | ไม่มีภาพ | https://storage.g... | null | 100.48661,13.79386 | 1867 จรัญสิทธิวงศ์... | บางแพด | | |
| 2021-FYJTEP | {ความสะอาด} | เดบานาซื้อ | ขอบเมือง | https://storage.g... | null | 100.53084,13.81865 | 12/14 กถนน กรุงเทพ... | นป. |
| 2021-8GKAR9 | {สายไฟ} | ชั้นไม้มีหมู่บ้าน... | 1. เดลวัลย์ดอนบน... | https://storage.g... | null | 100.57685,13.79704 | 33/51 ถนน คลองพารา ฯ... | สามเสนนอ | 
| 2021-AFPUXZ | {ถนน,สะพาน} | null | 1 ข้อความท้าตัวเอง... | https://storage.g... | null | 100.52916,13.72338 | 37 10 ซอย สีลม 9 ... | สีลม |
| 2021-CGPMUN | [น้ำท่วม,ร่องเรียน] | เขตประเวศ,ฝ้ายโภช... | [น้ำท่วมเวลาฝนตกและ... | https://storage.g... | https://storage.g... | 100.66709,13.67891 | 189 เลขบ้านพะเกียร... | หนองบอน |
| 2021-8Z8JQ3 | {ถนน,ร่องเรียน,น้ำ... | null | 1 พื้นที่น้ำขาด... | https://storage.g... | null | 100.86190,13.72804 | 3/8 ถนนทอง-สานต์อ่อน... | ชุมทอง |
| 2021-G4FN8W | {คลอง} | null | น้ำภายในคลอง | https://storage.g... | null | 100.62298,13.74227 | 50 พระราชนครินทร์ 9... | สวนหลวง |
| 2021-AFJYNF | {ถนน,น้ำท่วม}| null | น้ำท่วม เอ่อลงบัน... | https://storage.g... | null | 100.71592,13.73614 | 3 กถนน พัฒนาชัย 3... | คลองส่งน้ำทุ่น |
| 2021-HAJULK | {ร่องเรียน,ถนน,น้ำ... | null | มีการยกถนนในช่วงเย็น... | https://storage.g... | null | 100.71661,13.79225 | 95 ซอย 12 มีนบุรี... | มีนบุรี |
| 2021-Q7XMCN | {} | null | งานไม่เสร็จ ควรเก็บ... | https://storage.g... | null | 100.50779,13.79675 | 1 กถนน จรัญสิทธิวงศ์... | บางอ้อ |
| 2021-7XATFA | {สะพาน} | เขตสาทร สะพานลอบมีน้ำรุ่นไก... | https://storage.g... | null | 100.52649,13.72060 | 191/1 ถนน สาทรเหนือ... | สาทร |
| 2021-79YL3M | {ถนน} | null | 1. ช่องระบายน้ำไม่ได้... | https://storage.g... | null | 100.73357,13.91732 | 147 ซอย ราชพฤกษ์ไม่... | สามวาด้วนมอง |
| 2021-9U2NJT | [น้ำท่วม] | [เดบานาซื้อ,ฝ้ายโภช... | น้ำท่วม | https://storage.g... | https://storage.g... | 100.53009,13.81959 | 12/14 กถนน คลองหนอง... | null |
| 2021-DVJPAV | {} | null |
| 2021-F7A2PK | {ห้องน้ำ,ความสะอาด} | null | ไม่มีภาพ | null | null | null | null | null |
| 2021-DHG7W7 | {ทางเท้า} | null | บ้านแพด | บ้านแพด | กรุงเทพมหานคร | 2021-09-01 10:44:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:59:... |
| 2021-DANRZQ | {ทางเท้า} | null | null | null | กรุงเทพมหานคร | 2021-09-03 12:51:... | เสร็จแล้ว | null | null | 2022-06-04 15:34:... |
| 2021-GANMJA | {} | null | สามเสนนอ | หัวขวยหัว/จังหวัดกรุงเทพมหานคร | 2021-09-19 06:47:... | [กำลังดำเนินการ] | [กบ.] | null | 2022-02-22 04:30:... |
| 2021-94U766 | {คลอง,ความสะอาด} | null | มีนบุรี | สีลม | บ้านรัก | กรุงเทพมหานคร | 2021-09-19 07:40:... | [กำลังดำเนินการ] | null | 2022-02-22 04:30:... |
| 2021-3MBGD7 | {} | null | หนองบอน | ประเวศ | กรุงเทพมหานคร | 2021-09-19 14:56:... | เสร็จแล้ว | 4 | null | 2022-06-21 08:21:... |
| 2021-9U2NJT | {} | null | ชุมทาง | ลากคลอง | กรุงเทพมหานคร | 2021-09-20 00:24:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | สวนหลวง | สวนหลวง | กรุงเทพมหานคร | 2021-09-21 05:22:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | คลองส่งน้ำทุ่น | คลองส่งน้ำทุ่น | กรุงเทพมหานคร | 2021-09-22 05:06:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | มีนบุรี | มีนบุรี | กรุงเทพมหานคร | 2021-09-23 06:25:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | บางอ้อ | บ้านแพด | กรุงเทพมหานคร | 2021-09-23 10:25:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | บ้านนา | สาทร | กรุงเทพมหานคร | 2021-09-24 04:15:... | [กำลังดำเนินการ] | [กบ.] | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | สามวาด้วนมอง | คลองสามวา | กรุงเทพมหานคร | 2021-09-28 23:01:... | [กำลังดำเนินการ] | [กบ.] | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | ถนนพญาไท | ราชปรี | กรุงเทพมหานคร | 2021-10-14 10:45:... | เสร็จแล้ว | null | null | 2022-09-08 08:35:... |
| 2021-9U2NJT | {} | null | บางรัก | ภาษีเจริญ | กรุงเทพมหานคร | 2021-11-07 05:39:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | ถนนพญาไท | ภาษีเจริญ | กรุงเทพมหานคร | 2021-11-12 02:19:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | ถนนพญาไท | ราชดำเนิน | กรุงเทพมหานคร | 2021-11-14 10:21:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | ถนนปรานี/ถนนปรานีสุดรุ่งท่า | กรุงเทพมหานคร | 2021-11-16 04:16:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | ปากคลองภาษีเจริญ | ภาษีเจริญ | กรุงเทพมหานคร | 2021-11-23 02:03:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | สวนหลวง | สวนหลวง | กรุงเทพมหานคร | 2021-11-25 03:43:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |
| 2021-9U2NJT | {} | null | วัดคุณยวดี | ถนนธนบุรี | กรุงเทพมหานคร | 2021-11-25 09:16:... | [กำลังดำเนินการ] | null | null | 2022-02-22 04:29:... |

```

```
df.count()
```

```
270632
```

```
df.printSchema()
```

```
root
|-- ticket_id: string (nullable = true)
|-- type: string (nullable = true)
|-- organization: string (nullable = true)
|-- comment: string (nullable = true)
|-- photo: string (nullable = true)
|-- photo_after: string (nullable = true)
|-- coords: string (nullable = true)
|-- address: string (nullable = true)
|-- subdistrict: string (nullable = true)
|-- district: string (nullable = true)
|-- province: string (nullable = true)
|-- timestamp: string (nullable = true)
|-- state: string (nullable = true)
|-- star: string (nullable = true)
|-- count_reopen: string (nullable = true)
|-- last_activity: string (nullable = true)
```

```
# drop unused column
cols = ['ticket_id','photo', 'photo_after']
df = df.drop(*cols)
```

# Convert to proper data type

```
from pyspark.sql.functions import col
cols = ['star', 'count_reopen']
for c in cols:
    df = df.withColumn(c, col(c).cast('int'))

cols = ['timestamp', 'last_activity']
for c in cols:
    df = df.withColumn(c, col(c).cast('timestamp'))

from pyspark.sql.functions import split, regexp_replace
cols = ['type']
for c in cols:
    df = df.withColumn(c, split(regexp_replace(col(c), "[{}]", ""), ","))

cols = ['organization', 'coords']
for c in cols:
    df = df.withColumn(c, split(col(c), ","))

from pyspark.sql.functions import col, round

# Split the coords array into separate latitude and longitude columns
df = df.withColumn("latitude", col("coords").getItem(0).cast("double")) \
        .withColumn("latitude", round(col("latitude"), 2)) \
        .withColumn("longitude", col("coords").getItem(1).cast("double")) \
        .withColumn("longitude", round(col("longitude"), 2)) \
        .drop("coords")
```

```
[ ] df.printSchema()
```

```
root
|-- type: array (nullable = true)
|   |-- element: string (containsNull = false)
|-- organization: array (nullable = true)
|   |-- element: string (containsNull = false)
|-- comment: string (nullable = true)
|-- address: string (nullable = true)
|-- subdistrict: string (nullable = true)
|-- district: string (nullable = true)
|-- province: string (nullable = true)
|-- timestamp: timestamp (nullable = true)
|-- state: string (nullable = true)
|-- star: integer (nullable = true)
|-- count_reopen: integer (nullable = true)
|-- last_activity: timestamp (nullable = true)
|-- latitude: double (nullable = true)
|-- longitude: double (nullable = true)
```

```
[ ] df.show()
```

type	organization	comment	address	subdistrict	district	province	timestamp	state	star	count_reopen	last_activity	latitude	longitude
[]	null	ไม่มีภาค 1867 จังหวัดทั่วไป...	บางพลัด	บางพลัด	กรุงเทพมหานคร	2021-09-01 10:44:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:59:...  100.49   13.79					
[ความต้องการ]	[เชียงใหม่ชื่อ]	เชียงใหม่ 12/14 ถนน กรุงเทพ...	null	null	กรุงเทพมหานคร	2021-09-03 12:51:...  เผรีสัมภានnull	null	2022-06-04 15:34:...  100.53   13.82					
[ถนน, สะพาน]	[ชั้นในเมืองงาน...]	1. เอกวิธีของถนน... 335/1 ถนนพ...	สามเสนอกอก	ห้วยขวาง จังหวัดกรุงเทพมหานคร	2021-09-19 06:47:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:30:...  100.58   13.8						
[ถนน, สะพาน]		null 1 ช่องทางเท้าล่วง... 37 10 ซอย สีลม 9 ...	สีลม	บางรัก	กรุงเทพมหานคร	2021-09-19 07:40:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:30:...  100.53   13.72					
[น้ำท่วม, ร่องเรือน]	[เขตปะรจำ, ฝ่ายใต...	น้ำท่วมใจกลางเมือง... 189 เลี้นพรหมเกียร...	หนองบอน	ประจวบ	กรุงเทพมหานคร	2021-09-19 14:56:...  เผรีสัมภ...	null	2022-06-21 08:21:...  100.67   13.68					
[ถนน, ร่องเรือน, ...]	null 1	พื้นที่น้ำขาดดอ...	3/8 หมู่บ้าน ล่าด้ออ...	หมู่บอน	ลาดกระบัง	กรุงเทพมหานคร	2021-09-20 00:24:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.86   13.73				
[คลอง]	null 1	น้ำบันไดที่ไหลลง...	50 พระราม ที่ 9 แยก...	สวนหลวง	สวนหลวง	กรุงเทพมหานคร	2021-09-21 05:22:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.62   13.74				
[ถนน, น้ำท่วม]	null 1	น้ำท่วม เกิดอุบัติ...	3 ถนน พลเอกชัย 3...	คลองสองต้นน้ำ	ลาดกระบัง	กรุงเทพมหานคร	2021-09-22 05:06:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.72   13.74				
[ร่องเรือน, ถนน, ...]	null 1	น้ำการตอกถนนในพื้นที่...	95 ซอย 12 บ้านทึ...	บ้านทึ	บ้านทึ	กรุงเทพมหานคร	2021-09-23 06:25:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.72   13.79				
[]		บ้านบันไดเดียว ครอง...	1 กอน จังหวัดทั่วไป...	บางอ้อ	บางพลัด	กรุงเทพมหานคร	2021-09-24 04:15:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.51   13.8				
[สะพาน]	[เขตสัมพันธ์]	[สะพานและริบบินร่อง... 191/1 กอน สำราญ...	บ้านนาวา	สีทร	กรุงเทพมหานคร	2021-09-26 05:03:...  เผรีสัมภានnull	null	2022-06-04 01:17:...  100.53   13.72					
[ถนน]		null 1. ช่องทางริมแม่น้ำ...	147 ซอย ราชวิถี...	สันวลาจะนองออก	คลองสานฯ	กรุงเทพมหานคร	2021-09-28 23:01:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.73   13.92				
[น้ำท่วม]	[เขตปะรจำ, ฝ่าย...	น้ำท่วม 12/14 ถนน กรุงเทพ...	กนก...	กนก	กรุงเทพมหานคร	2021-10-14 10:45:...  เผรีสัมภ...	null	2022-09-08 08:35:...  100.53   13.82					
[]	null 1	และส่วนที่ 5 ถนน รากน้ำ แขวง...	ถนน พญาไท	ราชเทวี	กรุงเทพมหานคร	2021-11-07 05:39:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.54   13.76					
[พื้นที่น้ำท่วม, ควร...	null 1	น้ำท่วมน้ำที่น้ำท่า...	19/31 ซอย เพชรเกษ...	บางหว้า	ภาษีเจริญ	กรุงเทพมหานคร	2021-11-12 02:19:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.45   13.72				
[ทางเท้า]	null 1	ทางท่า 542/26 ถนน เพชรเกษ...	ถนนพญาไท	ราชเทวี	กรุงเทพมหานคร	2021-11-14 10:21:... ก้าลังค่าบ้านมาก null	null	2022-02-22 04:29:...  100.54   13.75					

# Drop null

## Drop null

```
[ ] # count null
    import pyspark.sql.functions as F

    df_agg = df.agg(*[F.count(F.when(F.isnull(c), c)).alias(c) for c in df.columns])

[ ] df_agg.show()

+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|type|organization|comment|address|subdistrict|district|province|timestamp|state|  star|count_reopen|last_activity|latitude|longitude|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 97|      1002|     2648|     2648|       72|      74|      25|       0|      0|171624|     117695|        0|        0|        0|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

```
[ ] df.filter("last_activity is NULL").show()

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|type|organization|comment|address|subdistrict|district|province|timestamp|state|star|count_reopen|last_activity|latitude|longitude|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

```
[ ] # drop rows where last_activity = null
    df = df.na.drop(subset=["last_activity"])

[ ] df.count()

270632
```

# Filter latitude and longitude

```
[ ] # Select only latitude and longitude inside Bangkok area  
df = df.filter((col("latitude") >= 100.32) & (col("latitude") <= 100.96) & \  
                (col("longitude") >= 13.40) & (col("longitude") <= 13.96))  
  
df.count()  
  
270299
```

# Calculate date\_to\_state

```
[ ] from pyspark.sql.functions import datediff  
  
# calculate date_to_state  
df = df.withColumn("date_to_state", datediff(col("last_activity"), col("timestamp")))  
df.show()  
  
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
|      type| organization| comment| address| subdistrict| district| province| timestamp| state|star|count_reopen| last_activity|latitude|longitude|date_to_state|  
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
|      []|       null|  นิมิตรฯ/1867 จังหวัดเชียงใหม่...|  บ้านพ่อดี|  บ้านพ่อดี|  กรุงเทพมหานคร| 2021-09-01 10:44:...| [ก้าวเดินบันทึก]| null|2022-02-22 04:59:...| 100.49| 13.79| 174| |
|[ความส่วนตัว][เขียนจดหมาย] ชัยมงคล/12/14 ถนน กรุงเทพฯ...|  null|  กกรุงเทพมหานคร|2021-09-03 12:51:...| เรื่ิซึสึ|null| null|2022-06-04 15:34:...| 100.53| 13.82| 274|  
|[สำนักฯ][ถึงไม่มีทบทวนงานฯ] | 1. เยาวลั่งอ่อนนุ่ม...|335/31 ลาดพร้าว แขวง...|  สวนแสงนุ่ม|  ท่าขี้ยววัง|  จังหวัดเชียงใหม่|  กรุงเทพมหานคร|2021-09-19 06:47:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:30:...| 100.58| 13.8| 156|  
|[ถนน, สะพาน]|  null|1 ช่องทางเดินที่ดี...|37/8 ซอย สีลม 9 แขวง...|  สีลม|  บางรัก|  กรุงเทพมหานคร|2021-09-19 07:40:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:30:...| 100.53| 13.72| 156|  
|[น้ำท่วม, ร่องรั้น][เขตประเวศ, ฝ่ายฯ...]|  น้ำท่วมทางเดินดี...|189 เอเชี่ยนพรีเมี่ยม...|  ถนนอนุฯ|  ประเวศ|  กรุงเทพมหานคร|2021-09-19 14:56:...| เรื่ิซึสึ 4| null|2022-06-21 08:21:...| 100.67| 13.68| 275|  
|[ถนน, ร่องรั้น, ...]|  null|1 น้ำท่วมเดินดี...|3/8 ถนนสีลมฯ...|  ชุมทาง|  ลาดพร้าว|  กรุงเทพมหานคร|2021-09-20 08:24:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.86| 13.73| 155|  
|[คลอง]|  null|  น้ำท่วมเดินดีในคลอง|58 พระราม ที่ 9 แขวง...|  สวนหลวง|  สวนหลวง|  กรุงเทพมหานคร|2021-09-21 05:22:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.62| 13.74| 154|  
|[ถนน, น้ำท่วม]|  null|  น้ำท่วม เดินดี...|3 ถนน พัฒนาอุปนิสัย 3 แขวง...|  คลองสอดเด่นนุ่ม|  ลาดพร้าว|  กรุงเทพมหานคร|2021-09-22 05:06:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.72| 13.74| 153|  
|[ร่องรั้น, ถนน, ...]|  null|  มีการยกถนนให้เดิน...|95 ซอย 12 มีนรี...|  มีนรี|  มีนรี|  กรุงเทพมหานคร|2021-09-23 06:25:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.72| 13.79| 152|  
|[ถนน]|  null|  ราษฎร์เดินดี ควรเดิน...|1 ถนน จังหวัดเชียงใหม่...|  บ้านขาว|  บ้านขาว|  กรุงเทพมหานคร|2021-09-24 04:15:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.51| 13.8| 151|  
|[สะพาน][เขตสัมพุทธฯ][สะพานแยกรั้วบึงฯ...]| 191/1 ถนน สะพานเดิน...|  ถนนขาว|  สะพาน|  กรุงเทพมหานคร|2021-09-26 05:03:...| เรื่ิซึสึ|null| null|2022-06-06 01:17:...| 100.53| 13.72| 253|  
|[ถนน]|  null| 1. สะพานขาวเดินดี...|147 ซอย ราชบูรณะ...|  ถนนขาว|  ถนนขาว|  กรุงเทพมหานคร|2021-09-28 23:01:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.73| 13.92| 147|  
|[น้ำท่วม][เขตด้ำดื่น, ฝ่ายฯ...]|  น้ำท่วม|12/14 ถนน กรุงเทพฯ...|  null|  null|  กรุงเทพมหานคร|2021-10-14 10:45:...| เรื่ิซึสึ|null| null|2022-09-08 08:35:...| 100.53| 13.82| 329|  
|[ถนน]|  null|  แสงสว่าง 5 ถนน งามวงศ์ฯ...|  ถนนพญาไท|  ราชเทวี|  กรุงเทพมหานคร|2021-11-07 05:39:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.54| 13.76| 187|  
|[ท่อระบายน้ำ, ดาดฯ...]|  null|  มีท่อระบายน้ำดี...|19/31 ซอย เพชรเกษม...|  ถนนพญาไท|  ราชเทวี|  กรุงเทพมหานคร|2021-11-12 02:19:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.45| 13.72| 102|  
|[ทางเดิน]|  null|  ทางเดิน|542/26 ถนน เพชรบุรี...|  ถนนพญาไท|  ราชเทวี|  กรุงเทพมหานคร|2021-11-14 10:21:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.54| 13.75| 100|  
|[ทางเดิน]|  null|  ทางเดิน|326 ถนน พัฒนาอุปนิสัย...|  ถนนพญาไท|  ราชเทวี|  กรุงเทพมหานคร|2021-11-16 04:16:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.51| 13.74| 98|  
|[ถนน]|  null|  ขอร่วมเป็นเพื่อนฯ...|206/52 ถนน เพชรบุรี...|  ถนนพญาไท|  ราชเทวี|  กรุงเทพมหานคร|2021-11-23 02:03:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.46| 13.72| 91|  
|[คลอง, ความสะอาด]]|  null|  มีร่องน้ำเดินดีมาก...|17/1 ซอย อ่อนนุ่ม...|  ถนนขาว|  ถนนขาว|  กรุงเทพมหานคร|2021-11-25 03:43:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.65| 13.71| 89|  
|[ถนน]|  null|  บริเวณเดินดีดี...|59/21 ซอย วัดก้อนดี...|  ถนนขาว|  ถนนขาว|  กรุงเทพมหานคร|2021-11-25 09:16:...|[ก้าวเดินบันทึก]| null|2022-02-22 04:29:...| 100.49| 13.74| 89|  
only showing top 20 rows
```

# Prepare data for visualization

```
from pyspark.sql.functions import explode

df_vis = df.select(explode('type').alias('type_exploded'), "*").drop('type','organization','comment')
df_vis.show()

for column in df_vis.columns:
    df_vis = df_vis.withColumn(column, regexp_replace(col(column), '\n', ''))

df_vis.write.option("header", "true").option("quoteAll", "true").option("lineSep", "\n").csv(path="/content/spark_output/vis")
```

type_exploded	address	subdistrict	district	province	timestamp	state star count_reopen	last_activity latitude longitude date_to_state
1867 จรัญสินธวงศ์...  บางแพล็ด  บางแพล็ด  กรุงเทพมหานคร 2021-09-01 10:44:... ก้าลังดำเนินการ null  null 2022-02-22 04:59:...  100.49  13.79  174							
ความสะอาด 12/14 ถนน กรุงเทพ...  null  null  กรุงเทพมหานคร 2021-09-03 12:51:...  เสร็จสิ้น null  null 2022-06-04 15:34:...  100.53  13.82  274							
สาไฟ 335/31 ลาดพร้าว ...  สามเสนออก  ห้วยขวาง  จังหวัดกรุงเทพมหานคร 2021-09-19 06:47:... ก้าลังดำเนินการ null  null 2022-02-22 04:30:...  100.58  13.8  156							
ถนน 37 10 ซอย สีลม 9 ...  สีลม  บางรัก  กรุงเทพมหานคร 2021-09-19 07:40:... ก้าลังดำเนินการ null  null 2022-02-22 04:30:...  100.53  13.72  156							
สะพาน 37 10 ซอย สีลม 9 ...  สีลม  บางรัก  กรุงเทพมหานคร 2021-09-19 07:40:... ก้าลังดำเนินการ null  null 2022-02-22 04:30:...  100.53  13.72  156							
น้ำท่วม 189 เฉลิมพระเกียร...  หนองบอน  ประเวศ  กรุงเทพมหานคร 2021-09-19 14:56:...  เสร็จสิ้น  4  null 2022-06-21 08:21:...  100.67  13.68  275							
ร่องเรียน 189 เฉลิมพระเกียร...  หนองบอน  ประเวศ  กรุงเทพมหานคร 2021-09-19 14:56:...  เสร็จสิ้น  4  null 2022-06-21 08:21:...  100.67  13.68  275							
ถนน 3/8 ชุมทาง-ล่าด้อย...  ชุมทาง  ลาดกระบัง  กรุงเทพมหานคร 2021-09-20 00:24:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.86  13.73  155							
ร่องเรียน 3/8 ชุมทาง-ล่าด้อย...  ชุมทาง  ลาดกระบัง  กรุงเทพมหานคร 2021-09-20 00:24:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.86  13.73  155							
น้ำท่วม 3/8 ชุมทาง-ล่าด้อย...  ชุมทาง  ลาดกระบัง  กรุงเทพมหานคร 2021-09-20 00:24:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.86  13.73  155							
ความปลอดภัย 3/8 ชุมทาง-ล่าด้อย...  ชุมทาง  ลาดกระบัง  กรุงเทพมหานคร 2021-09-20 00:24:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.86  13.73  155							
คลอง 50 พระราม ที่ 9 ...  สวนหลวง  สวนหลวง  กรุงเทพมหานคร 2021-09-21 05:22:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.62  13.74  154							
ถนน 3 ถนน พัฒนาชัยท 3...  คลองสองต้นบุญ  ลาดกระบัง  กรุงเทพมหานคร 2021-09-22 05:06:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.72  13.74  153							
น้ำท่วม 3 ถนน พัฒนาชัยท 3...  คลองสองต้นบุญ  ลาดกระบัง  กรุงเทพมหานคร 2021-09-22 05:06:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.72  13.74  153							
ร่องเรียน 95 ซอย 12 มีนบุรี...  มีนบุรี  มีนบุรี  กรุงเทพมหานคร 2021-09-23 06:25:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.72  13.79  152							
ถนน 95 ซอย 12 มีนบุรี...  มีนบุรี  มีนบุรี  กรุงเทพมหานคร 2021-09-23 06:25:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.72  13.79  152							
น้ำท่วม 95 ซอย 12 มีนบุรี...  มีนบุรี  มีนบุรี  กรุงเทพมหานคร 2021-09-23 06:25:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.72  13.79  152							
1 ถนน จรัญสินธวงศ...  บางอ้อ  บางแพล็ด  กรุงเทพมหานคร 2021-09-24 04:15:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.51  13.8  151							
สะพาน 191/1 ถนน สารทaben...  ยานนาวา  สารทaben  กรุงเทพมหานคร 2021-09-26 05:03:...  เสร็จสิ้น null  null 2022-06-06 01:17:...  100.53  13.72  253							
ถนน 147 ซอย รามอุปถัม...  สามวาจัณออก/คลองสามวา  กรุงเทพมหานคร 2021-09-28 23:01:... ก้าลังดำเนินการ null  null 2022-02-22 04:29:...  100.73  13.92  147							

only showing top 20 rows

# Focus on State

```
[ ] df_state_1 = df.groupby('state').count()  
df_state_1.show()
```

state	count
รอรับเรื่อง	18501
กำลังดำเนินการ	49431
เสร็จสิ้น	202367

```
from pyspark.sql.functions import min, avg, max

df_state_2 = df.groupby('state').agg(min('date_to_state').alias('min_date_to_state'),
                                     avg('date_to_state').alias('avg_date_to_state'),
                                     max('date_to_state').alias('max_date_to_state'))
df_state_2.show(20, False)
```

state	min_date_to_state	avg_date_to_state	max_date_to_state
รองรับเรื่อง	0	11.590508621155614	336
กำลังดำเนินการ	0	85.57648034634137	491
เสร็จสิ้น	0	43.693438159383696	464

```
from pyspark.sql.functions import year, month

df_state_3 = df.groupBy(year('timestamp').alias('year'), month('timestamp').alias('month')).count().orderBy('year', 'month')
df_state_3.show()

+---+---+---+
|year|month|count|
+---+---+---+
|2021|   9|   12|
|2021|  10|    1|
|2021|  11|    7|
|2021|  12| 162|
|2022|   1| 160|
|2022|   2| 172|
|2022|   3| 108|
|2022|   4|  24|
|2022|   5| 2214|
|2022|   6|61602|
|2022|   7|40821|
|2022|   8|26959|
|2022|   9|22306|
|2022|  10|13504|
|2022|  11|11471|
|2022|  12|11371|
|2023|   1|18572|
|2023|   2|18559|
|2023|   3|18193|
|2023|   4|17588|
+---+---+---+
only showing top 20 rows
```

3

```

from pyspark.sql.functions import year, month, count

df_state_4 = df.groupBy(year('timestamp').alias('year'), month('timestamp').alias('month'), 'state')\
    .agg(count('date_to_state').alias('row_count'),\
        min('date_to_state').alias('min_date_to_state'),\
        avg('date_to_state').alias('avg_date_to_state'),\
        max('date_to_state').alias('max_date_to_state'))\
    .orderBy('state', 'year', 'month')
df_state_4.show(100, False)

```

year	month	state	row_count	min_date_to_state	avg_date_to_state	max_date_to_state
2021	9	กทลังค่าเพิ่มการ	9	147	155.33333333333334	174
2021	11	กทลังค่าเพิ่มการ	7	89	96.57142857142857	107
2021	12	กทลังค่าเพิ่มการ	154	54	72.57792207792208	491
2022	1	กทลังค่าเพิ่มการ	141	1	40.94326241134752	436
2022	2	กทลังค่าเพิ่มการ	135	0	19.637037837037036	34
2022	3	กทลังค่าเพิ่มการ	56	0	7.910714285714286	14
2022	4	กทลังค่าเพิ่มการ	3	0	0.0	0
2022	5	กทลังค่าเพิ่มการ	300	0	189.08666666666667	346
2022	6	กทลังค่าเพิ่มการ	10050	0	171.91771144278687	344
2022	7	กทลังค่าเพิ่มการ	6378	0	152.61602383192223	314
2022	8	กทลังค่าเพิ่มการ	3721	0	127.17038430529428	280
2022	9	กทลังค่าเพิ่มการ	2680	0	108.67276119402985	251
2022	10	กทลังค่าเพิ่มการ	1814	0	92.31256890848952	216
2022	11	กทลังค่าเพิ่มการ	1675	0	77.09611940298508	191
2022	12	กทลังค่าเพิ่มการ	1930	0	59.17305699481865	158
2023	1	กทลังค่าเพิ่มการ	3349	0	33.534786503433864	130
2023	2	กทลังค่าเพิ่มการ	4016	0	20.433764940239044	99
2023	3	กทลังค่าเพิ่มการ	4317	0	11.68566133889275	71
2023	4	กทลังค่าเพิ่มการ	5562	0	4.575332614167565	41
2023	5	กทลังค่าเพิ่มการ	3134	0	1.4757498404594767	10
2022	2	รอรับเรื่อง	8	0	0.0	0
2022	3	รอรับเรื่อง	34	0	1.6470588235294117	56
2022	4	รอรับเรื่อง	20	0	0.0	0
2022	5	รอรับเรื่อง	123	0	13.544715447154472	303
2022	6	รอรับเรื่อง	1511	0	58.0	336
2022	7	รอรับเรื่อง	2260	0	21.00796460176991	311
2022	8	รอรับเรื่อง	1680	0	15.564285714285715	259
2022	9	รอรับเรื่อง	1511	0	9.44540039708802	241

2022	10	รอรับเรื่อง	1090	0	7.377064220183486	216
2022	11	รอรับเรื่อง	1164	0	6.591924398625429	186
2022	12	รอรับเรื่อง	1120	0	5.74375	157
2023	1	รอรับเรื่อง	1472	0	3.1345108695652173	118
2023	2	รอรับเรื่อง	1442	0	2.622746185852982	94
2023	3	รอรับเรื่อง	1696	0	1.9127358490566038	68
2023	4	รอรับเรื่อง	1953	0	1.3230926779313876	40
2023	5	รอรับเรื่อง	1417	0	0.5695130557515878	10
2021	9	ເລື່ອງສິນ	3	253	267.3333333333333	275
2021	10	ເລື່ອງສິນ	1	329	329.0	329
2021	12	ເລື່ອງສິນ	8	166	235.75	456
2022	1	ເລື່ອງສິນ	19	29	186.73684210526315	464
2022	2	ເລື່ອງສິນ	29	0	85.20689655172414	434
2022	3	ເລື່ອງສິນ	18	4	84.55555555555556	278
2022	4	ເລື່ອງສິນ	1	6	6.0	6
2022	5	ເລື່ອງສິນ	1791	0	82.32049134561697	346
2022	6	ເລື່ອງສິນ	50041	0	72.98057592773925	343
2022	7	ເລື່ອງສິນ	32183	0	61.78821116738651	313
2022	8	ເລື່ອງສິນ	21558	0	48.15859541701457	283
2022	9	ເລື່ອງສິນ	18115	0	39.79718465360199	251
2022	10	ເລື່ອງສິນ	10600	0	31.815849056603774	220
2022	11	ເລື່ອງສິນ	8632	0	27.25023169601483	188
2022	12	ເລື່ອງສິນ	8321	0	22.91227015983656	161
2023	1	ເລື່ອງສິນ	13751	0	15.372991055195985	128
2023	2	ເລື່ອງສິນ	13101	0	10.895122509732081	98
2023	3	ເລື່ອງສິນ	12180	0	8.58128078817734	71
2023	4	ເລື່ອງສິນ	10073	0	5.713690062543433	40
2023	5	ເລື່ອງສິນ	1942	0	2.6261589938208	11

4

```

from pyspark.sql.functions import year, month, count, sum, col

df_count = df.groupBy(year('timestamp').alias('year'), month('timestamp').alias('month'), 'state') \
    .agg(count('date_to_state').alias('row_count'))

df_total_count = df_count.groupBy('year', 'month') \
    .agg(sum('row_count').alias('total_count'))

df_ratio = df_count.join(df_total_count, ['year', 'month']) \
    .withColumn('state_ratio', col('row_count') / col('total_count')) \

df_state_5 = df_ratio.orderBy('year', 'month', 'state')
df_state_5.show(100, False)

```

year	month	state	row_count	total_count	state_ratio
2021 9	กำลังดำเนินการ 9	12	0.75		
2021 9	เสร็จสิ้น 3	12	0.25		
2021 10	เสร็จสิ้น 1	1	1.0		
2021 11	กำลังดำเนินการ 7	7	1.0		
2021 12	กำลังดำเนินการ 154	162	0.9506172839506173		
2021 12	เสร็จสิ้น 8	162	0.04938271604938271		
2022 1	กำลังดำเนินการ 141	160	0.88125		
2022 1	เสร็จสิ้น 19	160	0.11875		
2022 2	กำลังดำเนินการ 135	172	0.7848837209302325		
2022 2	รอรับเรื่อง 8	172	0.046511627906976744		
2022 2	เสร็จสิ้น 29	172	0.1686046511627907		
2022 3	กำลังดำเนินการ 56	108	0.5185185185185185		
2022 3	รอรับเรื่อง 34	108	0.3148148148148148		
2022 3	เสร็จสิ้น 18	108	0.166666666666666666		
2022 4	กำลังดำเนินการ 3	24	0.125		
2022 4	รอรับเรื่อง 20	24	0.8333333333333334		
2022 4	เสร็จสิ้น 1	24	0.041666666666666664		
2022 5	กำลังดำเนินการ 300	2214	0.13550135501355012		
2022 5	รอรับเรื่อง 123	2214	0.0555555555555555		
2022 5	เสร็จสิ้น 1791	2214	0.8089430894308943		
2022 6	กำลังดำเนินการ 10050	61602	0.16314405376448818		

2022 6	กำลังดำเนินการ 10050	61602	0.16314405376448818	
2022 6	รอรับเรื่อง 1511	61602	0.024528424401805138	
2022 6	เสร็จสิ้น 50041	61602	0.8123275218337067	
2022 7	กำลังดำเนินการ 6378	40821	0.15624311016388623	
2022 7	รอรับเรื่อง 2260	40821	0.055363660860831436	
2022 7	เสร็จสิ้น 32183	40821	0.7883932289752823	
2022 8	กำลังดำเนินการ 3721	26959	0.13802440743351013	
2022 8	รอรับเรื่อง 1680	26959	0.06231685151526392	
2022 8	เสร็จสิ้น 21558	26959	0.7996587410512259	
2022 9	กำลังดำเนินการ 2680	22306	0.12014704563794494	
2022 9	รอรับเรื่อง 1511	22306	0.06773962162646821	
2022 9	เสร็จสิ้น 18115	22306	0.8121133327355868	
2022 10	กำลังดำเนินการ 1814	13504	0.13433056872037916	
2022 10	รอรับเรื่อง 1090	13504	0.0807168246445497	
2022 10	เสร็จสิ้น 10600	13504	0.784952606635071	
2022 11	กำลังดำเนินการ 1675	11471	0.1460203992677186	
2022 11	รอรับเรื่อง 1164	11471	0.18147328044634295	
2022 11	เสร็จสิ้น 8632	11471	0.7525063202859384	
2022 12	กำลังดำเนินการ 1930	11371	0.1697300149503122	
2022 12	รอรับเรื่อง 1120	11371	0.09849617447893765	
2022 12	เสร็จสิ้น 8321	11371	0.7317738105707502	
2023 1	กำลังดำเนินการ 3349	18572	0.18032522076243807	
2023 1	รอรับเรื่อง 1472	18572	0.0792590997200862	
2023 1	เสร็จสิ้น 13751	18572	0.7404156795175533	
2023 2	กำลังดำเนินการ 4016	18559	0.21639096934102053	
2023 2	รอรับเรื่อง 1442	18559	0.07769815184007758	
2023 2	เสร็จสิ้น 13101	18559	0.7059108788189019	
2023 3	กำลังดำเนินการ 4317	18193	0.23728906722365745	
2023 3	รอรับเรื่อง 1696	18193	0.09322266805914363	
2023 3	เสร็จสิ้น 12180	18193	0.6694882647171989	
2023 4	กำลังดำเนินการ 5562	17588	0.3162383443256766	
2023 4	รอรับเรื่อง 1953	17588	0.11104161928587673	
2023 4	เสร็จสิ้น 10073	17588	0.5727200363884467	
2023 5	กำลังดำเนินการ 3134	6493	0.4826736485445865	
2023 5	รอรับเรื่อง 1417	6493	0.21823502233174188	
2023 5	เสร็จสิ้น 1942	6493	0.29909132912367165	

```
df_state_6 = df_ratio.orderBy('state', 'year', 'month')
```

```
df_state_6.show(100, False)
```

year	month	state	row_count	total_count	state_ratio
2021	9	ปราจีนบุรี	9	12	0.75
2021	11	ปราจีนบุรี	7	7	1.0
2021	12	ปราจีนบุรี	154	162	0.9506172839506173
2022	1	ปราจีนบุรี	141	160	0.88125
2022	2	ปราจีนบุรี	135	172	0.7848837209302325
2022	3	ปราจีนบุรี	56	188	0.5185185185185185
2022	4	ปราจีนบุรี	3	24	0.125
2022	5	ปราจีนบุรี	300	2214	0.13550135501355012
2022	6	ปราจีนบุรี	10050	61602	0.16314405376448818
2022	7	ปราจีนบุรี	6378	48821	0.15624311016388623
2022	8	ปราจีนบุรี	3721	26959	0.13802440743351013
2022	9	ปราจีนบุรี	2680	22306	0.12014784563794494
2022	10	ปราจีนบุรี	1814	13504	0.13433056872837916
2022	11	ปราจีนบุรี	1675	11471	0.1460203992677186
2022	12	ปราจีนบุรี	1930	11371	0.1697300149503122
2023	1	ปราจีนบุรี	3349	18572	0.18032522076243807
2023	2	ปราจีนบุรี	4016	18559	0.21639096934102053
2023	3	ปราจีนบุรี	4317	18193	0.23728906722365745
2023	4	ปราจีนบุรี	5562	17588	0.3162383443256766
2023	5	ปราจีนบุรี	3134	6493	0.4826736485445865
2022	2	รองรับเรื่อง	8	172	0.046511627906976744
2022	3	รองรับเรื่อง	34	108	0.3148148148148148
2022	4	รองรับเรื่อง	20	24	0.8333333333333334
2022	5	รองรับเรื่อง	123	2214	0.05555555555555555
2022	6	รองรับเรื่อง	1511	61602	0.024528424401805138
2022	7	รองรับเรื่อง	2260	48821	0.055363660860831436
2022	8	รองรับเรื่อง	1680	26959	0.06231685151526392
2022	9	รองรับเรื่อง	1511	22306	0.06773962162646821
2022	10	รองรับเรื่อง	1090	13504	0.08071682464454977
2022	11	รองรับเรื่อง	1164	11471	0.10147328044634295
2022	12	รองรับเรื่อง	1120	11371	0.09849617447893765
2023	1	รองรับเรื่อง	1472	18572	0.0792590997200862
2023	2	รองรับเรื่อง	1442	18559	0.07769815184007758
2023	3	รองรับเรื่อง	1696	18193	0.09322266805914363

2023	4	รองรับเรื่อง	1953	17588	0.11104161928587673
2023	5	รองรับเรื่อง	1417	6493	0.21823502233174188
2021	9	เลี้ยวซ้าย	3	12	0.25
2021	10	เลี้ยวซ้าย	1	1	1.0
2021	12	เลี้ยวซ้าย	8	162	0.04938271604938271
2022	1	เลี้ยวซ้าย	19	160	0.11875
2022	2	เลี้ยวซ้าย	29	172	0.1686046511627907
2022	3	เลี้ยวซ้าย	18	108	0.16666666666666666
2022	4	เลี้ยวซ้าย	1	24	0.041666666666666664
2022	5	เลี้ยวซ้าย	1791	2214	0.8089430894308943
2022	6	เลี้ยวซ้าย	50041	61602	0.8123275218337067
2022	7	เลี้ยวซ้าย	32183	48821	0.7883932289752823
2022	8	เลี้ยวซ้าย	21558	26959	0.7996587410512259
2022	9	เลี้ยวซ้าย	18115	22306	0.8121133327355868
2022	10	เลี้ยวซ้าย	10600	13504	0.784952606635071
2022	11	เลี้ยวซ้าย	8632	11471	0.7525063202859384
2022	12	เลี้ยวซ้าย	8321	11371	0.7317738105707502
2023	1	เลี้ยวซ้าย	13751	18572	0.7404156795175533
2023	2	เลี้ยวซ้าย	13101	18559	0.7059108788189019
2023	3	เลี้ยวซ้าย	12180	18193	0.6694882647171989
2023	4	เลี้ยวซ้าย	10073	17588	0.5727200363884467
2023	5	เลี้ยวซ้าย	1942	6493	0.29909132912367165

# Focus on Type

```
from pyspark.sql.functions import explode, size

df_exploded = df.select(explode('type').alias('type_exploded'), 'state', 'date_to_state')

df_grouped = df_exploded.groupBy('type_exploded', 'state') \
    .agg(count('date_to_state').alias('row_count'),
         min('date_to_state').alias('min_date_to_state'),
         avg('date_to_state').alias('avg_date_to_state'),
         max('date_to_state').alias('max_date_to_state')) \
    .orderBy('type_exploded')

df_type_1 = df_grouped
df_type_1.show(100, False)
```

type_exploded	state	row_count	min_date_to_state	avg_date_to_state	max_date_to_state
	รองรับเรื่อง	4414	0	12.421159945627549	330
	กำลังดำเนินการ	8608	0	109.98013475836431	491
	เสร็จสิ้น	49221	0	47.83348570320044	456
PM2.5	รองรับเรื่อง	848	0	9.441037735849056	96
PM2.5	กำลังดำเนินการ	227	0	7.488986784140969	81
PM2.5	รองรับเรื่อง	56	0	2.125	27
กำรเดินทาง	รองรับเรื่อง	84	0	3.6785714285714284	156
กำรเดินทาง	กำลังดำเนินการ	154	0	75.75974025974025	286
กำรเดินทาง	เมื่อสักวัน	300	0	37.79666666666667	301
กีดขวาง	กำลังดำเนินการ	1419	0	52.29245947850599	343
กีดขวาง	รองรับเรื่อง	1807	0	5.69894853348989	324
กีดขวาง	เสร็จสิ้น	12463	0	25.20999893284121	328
คนงานชั่วคราว	เสร็จสิ้น	782	0	15.801790281329923	265
คนงานชั่วคราว	กำลังดำเนินการ	90	0	40.622222222222	276
คนงานชั่วคราว	รองรับเรื่อง	95	0	2.6218526315789475	99
คล่อง	รองรับเรื่อง	343	0	10.434402332361516	292
คล่อง	กำลังดำเนินการ	1521	0	52.30996469428008	338
คล่อง	เมื่อสักวัน	6791	0	42.706670593432484	341
ความไม่สงบภัย	กำลังดำเนินการ	5581	0	78.41300842142985	436
ความไม่สงบภัย	รองรับเรื่อง	1942	0	11.2605561277034	317
ความไม่สงบภัย	เมื่อสักวัน	13077	0	47.45981494226505	341
ความสะอาด	กำลังดำเนินการ	1241	0	32.05560032232071	328
ความสะอาด	เสร็จสิ้น	18276	0	27.6913985554826	322
ความสะอาด	รองรับเรื่อง	381	0	7.687664041994751	259

พัฒนาดูแล	ตรวจสอบ	10270	0	27.0913985554826	522
ตรวจสอบ	รองรับเรื่อง	381	0	7.687664041994751	259
ตรวจสอบ	รองรับเรื่อง	2236	0	6.040250447227192	311
ตรวจสอบ	เสร็จสิ้น	8283	0	33.44138597126645	335
ตรวจสอบ	กำลังดำเนินการ	1829	0	83.4794969928923	344
ดำเนิน	กำลังดำเนินการ	935	0	57.319786096256685	339
ดำเนิน	รองรับเรื่อง	213	0	20.492957746478872	263
ดำเนิน	เสร็จสิ้น	6949	0	38.60526694488416	327
ถนน	กำลังดำเนินการ	16873	0	73.11266520476501	436
ถนน	เมื่อสักวัน	55220	0	43.26043100325969	462
ถนน	รองรับเรื่อง	4768	0	10.37312146107382	315
ทางท่อ	รองรับเรื่อง	1283	0	11.155884645362432	321
ทางท่อ	เมื่อสักวัน	23726	0	43.885998314088	464
ทางท่อ	กำลังดำเนินการ	6711	0	96.73476382059306	344
ห้องน้ำ	รองรับเรื่อง	374	0	27.788770053475936	315
ห้องน้ำ	เสร็จสิ้น	10812	0	42.253144654088054	337
ห้องน้ำ	กำลังดำเนินการ	2759	0	69.40703153161419	344
น้ำท่วม	เสร็จสิ้น	17096	0	42.00848151614413	329
น้ำท่วม	รองรับเรื่อง	348	0	9.827586206896552	259
น้ำท่วม	กำลังดำเนินการ	1698	0	75.4734982321555	343
ป้าย	รองรับเรื่อง	549	0	4.302367941712204	186
ป้าย	เมื่อสักวัน	4789	0	11.435790352892043	195
ป้าย	กำลังดำเนินการ	1202	0	19.92928452579035	181
ป้ายจราจร	รองรับเรื่อง	24	0	1.875	17
ป้ายจราจร	กำลังดำเนินการ	48	0	13.625	128
ป้ายจราจร	เมื่อสักวัน	187	0	20.44385026737968	187
ร้องเรียน	กำลังดำเนินการ	1846	0	13.02979414951246	182
ร้องเรียน	รองรับเรื่อง	786	0	2.464376590330789	119
ร้องเรียน	เมื่อสักวัน	4949	0	12.747570850202429	275
สบายน้ำ	รองรับเรื่อง	113	0	2.849557522123894	93
สบายน้ำ	เมื่อสักวัน	827	0	13.107617896009673	181
สบายน้ำ	กำลังดำเนินการ	254	0	13.098425196850394	134
สบายน้ำ	กำลังดำเนินการ	3469	0	99.59267800518882	342
สบายน้ำ	รองรับเรื่อง	1048	0	15.62029007633588	336
สบายน้ำ	เมื่อสักวัน	7741	0	54.16522413124919	434
ส้วมร่องดitch	รองรับเรื่อง	557	0	1.8402154398563735	178
ส้วมร่องดitch	กำลังดำเนินการ	1046	0	38.618546845124285	305
ส้วมร่องดitch	เมื่อสักวัน	2577	0	34.69499417927823	284
สายไฟ	รองรับเรื่อง	2297	0	22.609055289508053	320
สายไฟ	กำลังดำเนินการ	2807	0	112.8717491984325	340
สายไฟ	เมื่อสักวัน	4537	0	55.98302843288516	327
หลังคา	เสร็จสิ้น	197	0	14.314720812182742	159
หลังคา	รองรับเรื่อง	17	0	9.176470588235293	119
หลังคา	กำลังดำเนินการ	67	0	22.07462685671643	121
เส้นทาง	เสร็จสิ้น	752	0	34.097074468805104	339
เส้นทาง	รองรับเรื่อง	88	0	2.70454545454546	62
เส้นทาง	กำลังดำเนินการ	174	0	46.310344827586286	326
เสียงรบกวน	เสร็จสิ้น	7011	0	27.979902866923407	321
เสียงรบกวน	รองรับเรื่อง	526	0	5.159695817490494	257
เสียงรบกวน	กำลังดำเนินการ	1313	0	30.011422919345013	319
แสงสว่าง	เสร็จสิ้น	9766	0	86.62830227319272	341
แสงสว่าง	เมื่อสักวัน	11766	0	78.5097392486827	336
แสงสว่าง	รองรับเรื่อง	1137	0	15.944591029023746	330

1

# Focus on Organization

```
from pyspark.sql.functions import explode, size

df_exploded = df.select(explode('organization').alias('organization_exploded'), 'state', 'date_to_state')

df_grouped = df_exploded.groupBy('organization_exploded', 'state') \
    .agg(count('date_to_state').alias('row_count'),
         min('date_to_state').alias('min_date_to_state'),
         avg('date_to_state').alias('avg_date_to_state'),
         max('date_to_state').alias('max_date_to_state')) \
    .filter(col('state') == 'ເຈົ້າຈັ້ນ') \
    .orderBy('row_count', ascending=False)

df_org_1 = df_grouped
df_org_1.show(100, False)
```

organization_exploded	state	row_count	min_date_to_state	avg_date_to_state	max_date_to_state
ກ່ອນກາງທຳຫໄດ (ນາຍຈຸ່າທີ່ ວິໄລຍະຈົກ)	ເຈົ້າຈັ້ນ	32475	0	168.22729022324866	340
ກ່ອນກາງທຳຫວິນລົດ (ນາຍອຣັບ ເຈັດສິງ)	ເຈົ້າຈັ້ນ	32027	0	41.60272668567645	343
ກ່ອນກາງທຳຫົ່ວໂມງ	ເຈົ້າຈັ້ນ	27292	0	56.427595330490405	335
ກ່ອນກາງທຳຫລາກ (ນາຍສູງສັນຕິ ກິດຕືກກາງ)	ເຈົ້າຈັ້ນ	21676	0	38.814774181751694	346
ສໍາທິກຳກາງໃຫ້ກົມບ.	ເຈົ້າຈັ້ນ	19246	0	99.2743946794139	464
ກ່ອນກາງທຳຫົ່ວໂມງ (ນາຍອຣັບ ໂພນິພິບີ)	ເຈົ້າຈັ້ນ	18863	0	20.788262736574247	346
ສໍາທິກຳກາງຄ່າງແລ້ວຂ້າດີ (Royal Thai Police)	ເຈົ້າຈັ້ນ	16745	0	31.593968348768825	346
ກ່ອນຫຼັກຖານຄ່າງຄ່າງຄົນນາງ (ບັນ.ນ.)	ເຈົ້າຈັ້ນ	16734	0	31.431337396616458	346
ກ່ອນຫຼັກຖານທີ່ມີ (ນາຍອຣັບ ໂພນິພິບີ)	ເຈົ້າຈັ້ນ	16215	0	50.4027758014061	343
ກາງໄຟຟ້ານຄົນລວງ MEA	ເຈົ້າຈັ້ນ	11501	0	99.30623424850883	343
ເອກະລາດກົມ	ເຈົ້າຈັ້ນ	10064	0	93.30067567568	339
ສໍາທິກຳກາງຈາກວັນແຂວງລວງ ກຽມທ່ານຫານຕົກ (ສຈ.ສ.)	ເຈົ້າຈັ້ນ	9795	0	85.59979581419891	462
ເອກະປະເຈດ	ເຈົ້າຈັ້ນ	8823	0	27.61453820514564	336
ເອກະນາງເຄດ	ເຈົ້າຈັ້ນ	7848	0	19.80020387359837	341
ກອ.ເອດອອຈິກ (ນາຍພຣະເລີດ)	ເຈົ້າຈັ້ນ	7752	0	118.14138286893704	311
ກອ.ເອດກະວຽກ (ນາຍອອນຈິດ)	ເຈົ້າຈັ້ນ	7248	0	29.636451434878587	336
ເອກະນາງຂາຍ	ເຈົ້າຈັ້ນ	6960	0	23.70948275862069	328
ເອກະວິນາ	ເຈົ້າຈັ້ນ	6813	0	42.871715837369734	348
ເອກະຄາດຮັບນິງ	ເຈົ້າຈັ້ນ	6642	0	35.331249623607345	334
ເອກະນາງປິຕີ	ເຈົ້າຈັ້ນ	6615	0	40.167498110355254	342
ເອກະຄອບລົງຍອ	ເຈົ້າຈັ້ນ	6097	0	33.09348860095129	343
ກອ.ເອດວາງຜົນ (ນາຍອອນຈິດ)	ເຈົ້າຈັ້ນ	5822	0	16.63947987217451	341
ເບີດສ່ວນລົງລວງ	ເຈົ້າຈັ້ນ	5751	0	151.14432272648236	336
ສໍາທິກຳກາງຄ່າງແລ້ວຂ້າດີ ສໍາກຳກາງໂຍ້ດ້າ (ສກນ.ສນ.ຍ.) ການຍ.	ເຈົ້າຈັ້ນ	5545	0	128.969782434625	335
ກອ.ເອດວັດນາ (ນ.ສ.ສົກ.ທີ່)	ເຈົ້າຈັ້ນ	5475	0	45.91050283105025	340
ກອ.ເອດວັດນາ (ນ.ສ.ທີ່)	ເຈົ້າຈັ້ນ	5431	0	22.8031678042434947	328
ກອ.ເອດວັດນາກະບົວນິ	ເຈົ້າຈັ້ນ	5237	0	32.6681306091273	328
ເອກະນາງທຶນ	ເຈົ້າຈັ້ນ	5162	0	21.023440526927548	329
ກອ.ເອດວັດນາກະບົວນິ (ນາຍໄພຣັດ)	ເຈົ້າຈັ້ນ	5070	0	42.37888678509985	342
ເອກະຫ຾ກົມ	ເຈົ້າຈັ້ນ	4966	0	30.71236407571487	309
ສໍາທິກຳກາງນາຍ້າ ການຍ.	ເຈົ້າຈັ້ນ	4916	0	69.9953213995118	343
ກອ.ເອດວັດນົກລວງ (ນາຍສູງສັນ)	ເຈົ້າຈັ້ນ	4731	0	173.48869160853943	336
ເອກະນາງປິຕີ	ເຈົ້າຈັ້ນ	4721	0	44.35500953187884	329
ເອກະປົມກົມ	ເຈົ້າຈັ້ນ	4662	0	139.49141999141999	331

1

# Focus on District

## Focus on district

```
[ ] df_grouped = df.groupBy('district', 'state') \
    .agg(count('date_to_state').alias('row_count'),
        min('date_to_state').alias('min_date_to_state'),
        avg('date_to_state').alias('avg_date_to_state'),
        max('date_to_state').alias('max_date_to_state')) \
    .orderBy('district')

df_dist_1 = df_grouped
df_dist_1.show(100, False)
```

district	state	row_count	min_date_to_state	avg_date_to_state	max_date_to_state
null	ก้าสังดำเนินการ	4	1147	195.75	275
null	ออร์บีริอง	4	0	0.75	2
null	ເສົ້າຈີ່ນ	35	0	85.62857142857143	329
กรุงเทพมหานคร	ອอร์บีริอง	3	0	0.0	0
คลองสาน	ເສົ້າຈີ່ນ	2228	0	46.864981256732495	341
คลองสาน	ອอร์บีริอง	260	0	14.457692307692307	321
คลองสาน	ກໍາລັງດ້າເນີນກາງ	433	0	101.64665127020785	329
คลองสานวາ	ກໍາລັງດ້າເນີນກາງ	1214	0	78.34184514003294	325
คลองสานวາ	ເສົ້າຈີ່ນ	3726	0	44.94793440687066	334
คลองสานวາ	ອອຣິບິເຮືອ	247	0	2.8421052631578947	178
คลองเตย	ເສົ້າຈີ່ນ	6068	0	33.200065919578115	343
คลองเตย	ກໍາລັງດ້າເນີນກາງ	1378	0	87.12917271407838	338
คลองเตย	ອອຣິບິເຮືອ	996	0	11.264056224899598	304
ศันนาຍາວ	ເສົ້າຈີ່ນ	1969	0	23.39207719654647	386
ศันนาຍາວ	ອອຣິບິເຮືອ	108	0	8.083333333333334	224
ศันนาຍາວ	ກໍາລັງດ້າເນີນກາງ	823	0	50.308626974483595	341
จุดธาร	ກໍາລັງດ້າເນີນກາງ	3593	0	84.1839688282772	343
จุดธาร	ອອຣິບິເຮືອ	1383	0	9.843094721619668	314
จุดธาร	ເພື້ອຈີ່ນ	9960	0	93.36295180722891	339
จอมทอง	ກໍາລັງດ້າເນີນກາງ	723	0	83.46749654218534	334
จอมทอง	ເສົ້າຈີ່ນ	3140	0	44.09299363057325	335
จอมทอง	ອອຣິບິເຮືອ	1895	0	14.222054514260176	220

```
# rank by avg_date_to_state

df_grouped = df.groupBy('district', 'state') \
    .agg(count('date_to_state').alias('row_count'),
        min('date_to_state').alias('min_date_to_state'),
        avg('date_to_state').alias('avg_date_to_state'),
        max('date_to_state').alias('max_date_to_state')) \
    .filter((col('state') == 'ເຕີບສັນ') & (col('row_count') > 50)) \
    .orderBy('avg_date_to_state')

df_dist_2 = df_grouped
df_dist_2.show(100, False)
```

district	state	row_count	min_date_to_state	avg_date_to_state	max_date_to_state
ຈາກຖົງຮຽນ	ເຕີບສັນ	2075	0	15.88867469879518  326	
ນາງຄອນຫລວມ	ເຕີບສັນ	2466	0	16.714517437145176  328	
ນາງແດ	ເຕີບສັນ	7803	0	20.16403947199795  341	
ສັກພ່າງວັດ	ເຕີບສັນ	1074	0	20.45996275052143  346	
ນາງທຸນທີ່ຍິນ	ເຕີບສັນ	5134	0	21.030775223996883  329	
ທ່ອງຮ	ເຕີບສັນ	2723	0	21.808666911494676  341	
ດັນນາຍາວ	ເຕີບສັນ	1969	0	23.39207719654647  306	
ນາງເກມ	ເຕີບສັນ	6863	0	23.74165816698237  315	
ກຳຊີ້ວິຊູ	ເຕີບສັນ	3421	0	23.775796550716166  329	
ດົມເບື້ອງ	ເຕີບສັນ	3146	0	25.87730451366815  321	
ລາດພວັນ	ເຕີບສັນ	4303	0	26.587961887055542  462	
ປະເວສ	ເຕີບສັນ	8844	0	27.502826775214835  336	
ພະໄວງົງ	ເຕີບສັນ	3193	0	27.619793297839024  331	
ທະນອແຂນ	ເຕີບສັນ	2725	0	28.319633027522936  346	
ອຸນົກ	ເຕີບສັນ	3979	0	30.464438301880673  336	
ນຶື່ງນຸ່ງ	ເຕີບສັນ	4885	0	30.623541453428864  304	
ປິອນປານວັດຖຸພາຍ	ເຕີບສັນ	2698	0	30.872127501853225  341	
ພະບານດ	ເຕີບສັນ	3751	0	31.358304452146093  329	
ທ້າວັດນາ	ເຕີບສັນ	1968	0	32.864837398373986  323	
ຄວາມເດຍ	ເຕີບສັນ	5068	0	33.200065919578115  343	
ນາງນອນ	ເຕີບສັນ	2427	0	33.206015657189944  336	
ທັບສິ້ນ	ເຕີບສັນ	1832	0	33.79415461973601  334	
ນາງພັດ	ເຕີບສັນ	3425	0	33.88905109489051  335	
ລາດກະບັງ	ເຕີບສັນ	6612	0	35.64745916515427  334	
ນິ້ນບຸງ	ເຕີບສັນ	3560	0	36.15308988764045  326	

ຕິດແດງ	ເຕີບສັນ	4419	0	37.0296447159991  337
ສາກ	ເຕີບສັນ	4280	0	37.31892523364486  464
ຈຳຍົງ	ເຕີບສັນ	4556	0	37.55267778753292  321
ປຸ່ມວັນ	ເຕີບສັນ	4557	0	38.70133859995611  331
ຍານນາວາ	ເຕີບສັນ	2653	0	38.74104787033547  337
ຕລື່ອນ	ເຕີບສັນ	2866	0	38.945917655268666  341
ພຸກາທ	ເຕີບສັນ	4023	0	39.22694506587124  334
ສາຍໄທນ	ເຕີບສັນ	3796	0	39.33587987355111  335
ນາງກປີ	ເຕີບສັນ	6544	0	40.00030562347188  342
ນາງານາ	ເຕີບສັນ	3816	0	40.107180293501045  321
ດຸດິດ	ເຕີບສັນ	2193	0	41.6124031007752  434
ວັກວາງລາງ	ເຕີບສັນ	4284	0	42.26117982873454  336
ວັກນາ	ເຕີບສັນ	6386	0	42.65393047290949  324
ນາງໝໍອ	ເຕີບສັນ	4712	0	44.05241935483871  325
ຈອນທອງ	ເຕີບສັນ	3140	0	44.09299363057325  335
ຄວາມສຳວັດ	ເຕີບສັນ	3726	0	44.947933440687066  334
ຫ້ວຍຫວາງ	ເຕີບສັນ	3475	0	46.80374100719425  337
ຄວາມສຳນັກ	ເຕີບສັນ	2228	0	46.846901256732495  341
ນາງກອກໃຫຍ່	ເຕີບສັນ	1597	0	53.7407639323732  334
ທັນອອກ	ເຕີບສັນ	3250	0	68.95046153846154  324
ສະພານສູງ	ເຕີບສັນ	3255	0	61.83348694316436  343
ນາງກອກນ້ອຍ	ເຕີບສັນ	4345	0	77.15742232451093  343
ຈຸດວັກ	ເຕີບສັນ	9960	0	93.36295180722891  339
ນາງກັກ	ເຕີບສັນ	4255	0	130.99553466509988  338
ສຸວະຫວະງ	ເຕີບສັນ	5741	0	151.23480229925102  335

2

# Focus on Star

```
[ ] df_grouped = df.groupBy('star', 'state') \
    .agg(count('date_to_state').alias('row_count'),
        min('date_to_state').alias('min_date_to_state'),
        avg('date_to_state').alias('avg_date_to_state'),
        max('date_to_state').alias('max_date_to_state')) \
    .filter(col('state') == "ເສດຖະກິນ") \
    .orderBy('star')
```

```
df_star_1 = df_grouped
df_star_1.show(100, False)
```

star	state	row_count	min_date_to_state	avg_date_to_state	max_date_to_state
null	ເສດຖະກິນ	104732	0	45.65467096971317	464
1	ເສດຖະກິນ	13772	0	50.977708393842576	434
2	ເສດຖະກິນ	4629	0	51.606826528407865	325
3	ເສດຖະກິນ	10394	0	48.483740619588225	343
4	ເສດຖະກິນ	21080	0	41.2600569259962	342
5	ເສດຖະກິນ	47760	0	36.55674204355109	343



AI/ML

**Traditional ML**

# **Use to Train**

latitude  
longitude  
comment length  
type  
date\_to\_state

# **Unuse**

subdistrict  
district  
province  
organization  
star  
count\_reopen  
last\_activity  
timestamp

```
from pyspark.sql.functions import col, explode, when, length

df = df.drop('star', 'state', 'address', 'last_activity', 'subdistrict', 'district', 'province', 'organization', 'timestamp', 'count_reopen')

df = df.withColumn('comment_len', length(col('comment')))
df = df.drop('comment')

# Drop rows with null 'latitude', 'longitude', 'timestamp' and 'date_to_state'
df = df.dropna(subset=['date_to_state', 'type', 'latitude', 'longitude'])

# Explode 'type' and 'organization' columns
df = df.dropna(subset=['type'])
df = df.select('*', explode('type').alias('type_exploded'))
df = df.drop('type')
df = df.withColumn("type_exploded", when(col("type_exploded") == "", "unknown").otherwise(col("type_exploded")))
```

+-----+   type   organization   comment   address subdistrict district   province   timestamp   state star count_reopen   last_activity latitude longitude date_to_state
[ความสะอาด]   [เจดนาชีวี]   กะบะยะ[12/14 ตอน กrustoph...   null   null กรุงเทพมหานคร 2021-09-03 12:51:... [เสริชสื้น null]   null 2022-06-04 15:34:...  100.53   13.82   274
[น้ำท่วม, ร่องเรียบ]   [เขตประเวศ, ฝั่งธน., ฝั่งฯ...  น้ำท่วมเวลาฟลักพล...[189 เมื่อพึ่งก่อต...   หนองบอน   ประเวศ กรุงเทพมหานคร 2021-09-19 14:56:... [เสริชสื้น  4   null 2022-06-21 08:21:...  100.67   13.68   275
[สพน.]   [เขตสาทร]   สะพานอลย์เปรี้ยว...[191/1 ถนน สหัสสรฯ...   ถนนนราฯ   สาทร กรุงเทพมหานคร 2021-09-20 05:03:... [เสริชสื้น null]   null 2022-06-06 01:17:...  100.53   13.72   253
[น้ำท่วม, ถนน]   [เขตคลองเตย, ฝ่าย...]   น้ำท่วม[12/14 ถนน กrustoph...   null   null กรุงเทพมหานคร 2021-10-14 10:45:... [เสริชสื้น null]   null 2022-09-08 08:35:...  100.53   13.82   329
[น้ำท่วม, ถนน]   [เขตคลองเตย, ฝ่าย...]   ช่องดักพร้าวทัน...[189 เมื่อพึ่งก่อต...   ถนนนราฯ   สาทร กรุงเทพมหานคร 2021-12-09 12:29:... [เสริชสื้น  5   null 2022-08-12 07:18:...  100.59   13.82   246
[][][เขตคลองเตย, ฝ่าย...]   ช่องในเมืองที่ยวานฯ[627 ถนนส่วนสูง แขวง...   คลองสีคุต   คลองสีคุต กรุงเทพมหานคร 2021-12-17 08:46:... [เสริชสื้น  5   null 2022-03-14 12:09:...  100.59   13.81   456
[ความสะอาด]   [เขตประเวศ, ฝ่าย...]   ถนนของคนที่ลืม...[110 ซอย แม่น้ำ...   หนองบอน   ประเวศ กรุงเทพมหานคร 2021-12-18 14:50:... [เสริชสื้น  2   null 2022-06-14 07:36:...  100.65   13.67   178
[][][เขตประเวศ, ฝ่าย...]   ถนนสุดฯ ที่ พม...[208/22 ถ. พล. ที่ดินมาก...   ประเวศ   ประเวศ กรุงเทพมหานคร 2021-12-21 23:03:... [เสริชสื้น  2   null 2022-06-24 06:32:...  100.66   13.73   185
[ท่อระบายน้ำ]   [เขตประเวศ, ฝ่าย...]   ช่องเส้นเรือท่อระบ...[70 ซอย เลขในประเทศไทย...   หนองบอน   ประเวศ กรุงเทพมหานคร 2021-12-22 10:15:... [เสริชสื้น  5   null 2022-06-20 13:12:...  100.65   13.68   180
[][][เขตประเวศ, ฝ่าย...]   แม่เหล็กที่พังค่า...[55 ถนน ศรีด่านพงษ์...   หนองบอน   ประเวศ กรุงเทพมหานคร 2021-12-23 10:26:... [เสริชสื้น null]   null 2022-09-28 08:35:...  100.65   13.69   279
[ความสะอาด]   [เขตประเวศ, ฝ่าย...]   ป้ายหกสีแห่งชาติ...[22/1 ซอย อ่อนนุช...   ประเวศ   ประเวศ กรุงเทพมหานคร 2021-12-28 03:59:... [เสริชสื้น  4   null 2022-07-12 06:50:...  100.69   13.72   196
[ถนน, ทางเท้า]   [เขตคลองเตย, ฝ่ายเบซ...   บริเวณหน้าร้านวิสา...   สาทร กรุงเทพมหานคร 2022-01-02 10:53:... [เสริชสื้น null]   null 2022-06-08 05:46:...  100.54   13.71   157
[][][เขตสาทร, ฝ่ายเหน็...   บ่อเดื่อยาห์จอดรถแท...[184 ซอย จันท...   ห้องน้ำดอน   สาทร กรุงเทพมหานคร 2022-01-14 01:17:... [เสริชสื้น null]   null 2022-06-22 08:06:...  100.53   13.71   159
[][][เขตสาทร, ฝ่ายเหน็...   บ่อเดื่อยาห์จอดรถแท...[184 ซอย จันท...   ห้องน้ำดอน   สาทร กรุงเทพมหานคร 2022-01-14 01:18:... [เสริชสื้น null]   null 2022-06-22 08:10:...  100.53   13.71   159
[][][เขตประเวศ, ฝ่ายอeste...   ออกก้าวให้ส่วนลด...[59...   หนองบอน   ประเวศ กรุงเทพมหานคร 2022-01-15 12:52:... [เสริชสื้น null]   null 2022-06-20 07:12:...  100.66   13.69   156
[ถนน, สพน.]   [เขตคลองเตย, ฝ่ายกลาง...]   ถนนจัมโบ้หัวสุวัฒนา...[1357/1-4 ซอย จันท...   ถนน...   สาทร กรุงเทพมหานคร 2022-01-16 10:57:... [เสริชสื้น null]   null 2023-01-29 03:27:...  100.53   13.71   378
[][][เขตสาทร, ฝ่ายเหน็...   ป้ายชื่อ จันท...[13 ซอย จันท...   ห้องน้ำดอน   สาทร กรุงเทพมหานคร 2022-01-16 11:02:... [เสริชสื้น null]   null 2022-06-09 03:49:...  100.53   13.71   144
[สพน.]   [เขตสาทร, ฝ่ายเหน็...   ริมถนน ตรงสะพานสา...[3 ซอย จันท...   ห้องน้ำดอน   สาทร กรุงเทพมหานคร 2022-01-16 11:08:... [เสริชสื้น null]   null 2022-06-23 08:06:...  100.53   13.71   158
[จราจร, สะพาน, ถนน]   [เขตสาทร, ฝ่ายเหน็...   ริมถนน ตรงสะพานสา...[3 ซอย จันท...   ห้องน้ำดอน   สาทร กรุงเทพมหานคร 2022-01-16 11:14:... [เสริชสื้น null]   null 2022-06-22 08:18:...  100.53   13.71   157

↓

+-----+   latitude longitude date_to_state comment_len type_exploded
100.53   13.72   253   46   สพน.
100.65   13.68   180   62   ท่อระบายน้ำ
100.53   13.71   378   94   ถนน
100.53   13.71   378   94   ถนน
100.52   13.77   127   345   unknown
100.62   13.75   8   24   unknown
100.52   13.78   0   7   น้ำท่วม
100.7   13.68   136   119   ความสะอาด
100.66   13.68   95   14   unknown
100.61   13.84   5   37   น้ำท่วม
100.53   13.82   22   275   ทางเท้า
100.54   13.77   17   144   unknown
100.64   13.71   256   41   unknown
100.5   13.74   27   143   สพน.
100.5   13.74   27   143   ต้นไม้
100.62   13.78   28   165   น้ำท่วม
100.56   13.82   122   115   ทางเท้า
100.56   13.82   122   115   สายไฟ
100.64   13.76   34   86   คลอง
100.44   13.68   88   92   น้ำท่วม

```
from pyspark.ml import Pipeline
from pyspark.ml.feature import StringIndexer, VectorAssembler, OneHotEncoder
from pyspark.ml.regression import RandomForestRegressor
from pyspark.ml.evaluation import RegressionEvaluator

# Create StringIndexer for categorical columns
indexers = [StringIndexer(inputCol=column, outputCol=column+"_index", handleInvalid="keep").fit(df) for column in ['type_exploded']]

# OneHotEncoder for categorical columns
encoders = [OneHotEncoder(inputCol=indexer.getOutputCol(), outputCol=indexer.getOutputCol()+"_encoded") for indexer in indexers]

# VectorAssembler to combine all features
assembler = VectorAssembler(inputCols=[encoder.getOutputCol() for encoder in encoders] + ['latitude', 'longitude', 'comment_len'], outputCol="features", handleInvalid="skip")

# Regression model
rf = RandomForestRegressor(labelCol='date_to_state', featuresCol='features')

# Define a pipeline model
pipeline = Pipeline(stages=indexers + encoders + [assembler, rf])

# Split the data
(train_data, test_data) = df.randomSplit([0.8, 0.2], seed = 2023)

with mlflow.start_run():
    rf.setParams(maxDepth=maxDepth,numTrees=numTrees,maxBins=maxBins)
    # Train the model
    model = pipeline.fit(train_data)

    # Make predictions
    predictions = model.transform(test_data)

    # Evaluate the model
    evaluator = RegressionEvaluator(labelCol="date_to_state", predictionCol="prediction", metricName="rmse")
    rmse = evaluator.evaluate(predictions)
```

AI/ML

# MLflows & FastAPI

# MLflow

```
1 import mlflow  
2 import mlflow.spark  
3 mlflow.set_experiment('Datasci_project')
```

```
mlflow.spark.log_model(model, "random_forest_model")  
mlflow.log_param("maxDepth", maxDepth)  
mlflow.log_param("numTrees", numTrees)  
mlflow.log_param("maxBins", maxBins)  
mlflow.log_metric("rmse", rmse)
```

```
from pyngrok import ngrok
NGROK_AUTH_TOKEN = "2P6uaojOMwpxXazvH0bMhq10A0f_5zLWDu55i7pVhsiqqPHp7"
ngrok.set_auth_token(NGROK_AUTH_TOKEN)
ngrok.kill()
ngrok_tunnel = ngrok.connect(addr="5000",proto="http",bind_tls=True)
print("MLflow Tracking UI:",ngrok_tunnel.public_url)
```

```
. !mlflow ui #need to run this before clicking the link
```

# FastAPI

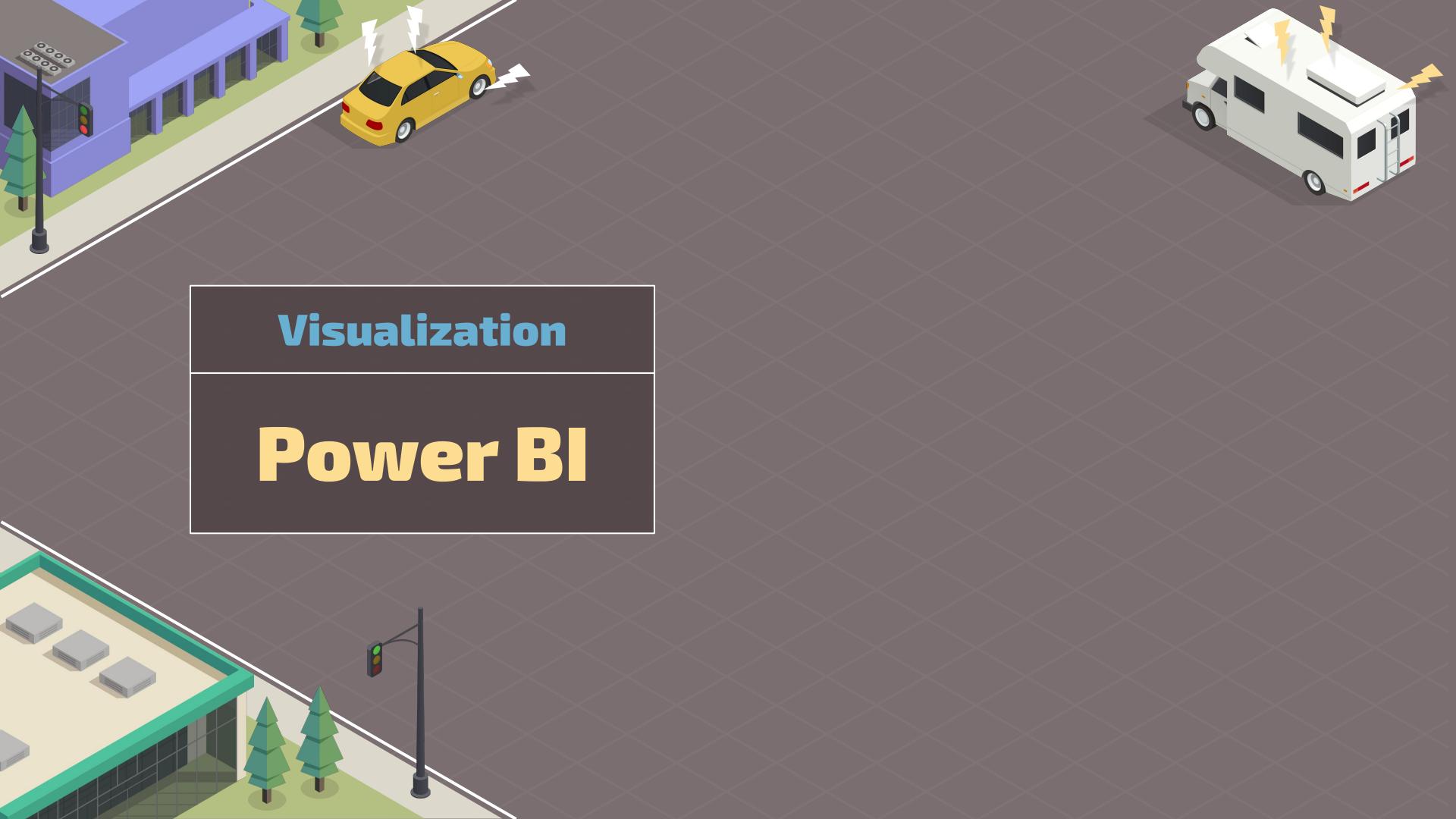
```
from typing import List
class model_input(BaseModel):
    latitude : float
    longitude : float
    comment_len : int
    ptype: List[str]
```

```
logged_model = 'runs:/d0c3014811dc43608ab1fd9c42a9969c/random_forest_model' #change model to deploy here
loaded_model = mlflow.spark.load_model(logged_model)
```

```
@app.post("/predict")
def predict(input_parameters : model_input):
    input_data = input_parameters.json()
    input_dictionary = json.loads(input_data)
    rows = [
        {
            'latitude': input_dictionary['latitude'],
            'longitude': input_dictionary['longitude'],
            'comment_len': input_dictionary['comment_len'],
            'type_exploded': ptype
        }
        for ptype in input_dictionary['ptype']
    ]

    # Create DataFrame from the list of dictionaries
    df = spark.createDataFrame(rows)

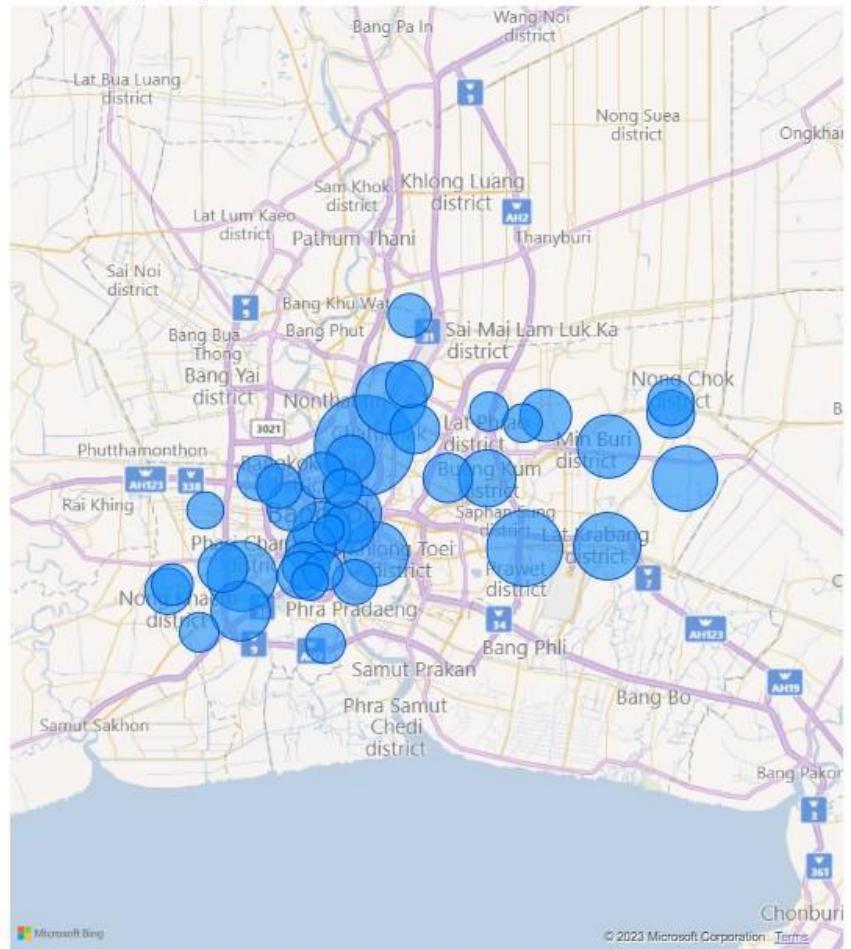
    pred = loaded_model.transform(df)
    mean_prediction = pred.selectExpr("mean(prediction)").collect()[0][0]
    return mean_prediction
```



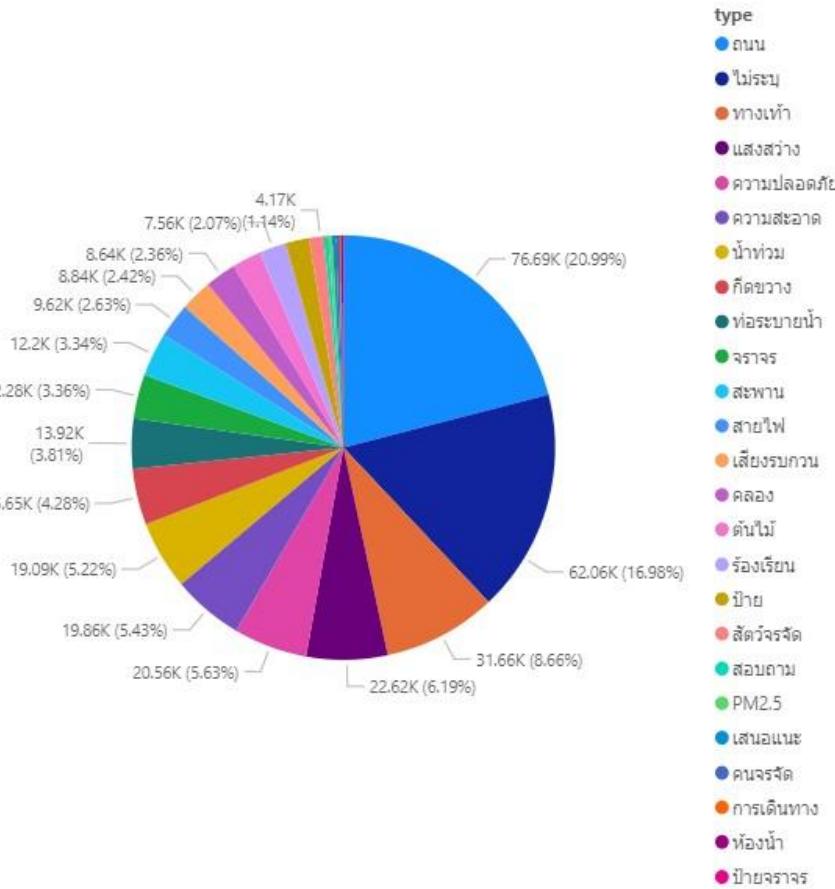
**Visualization**

**Power BI**

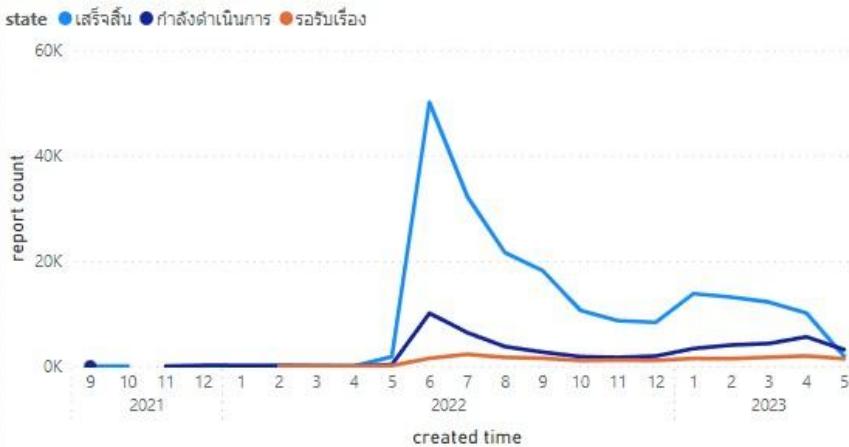
### amount of report by address



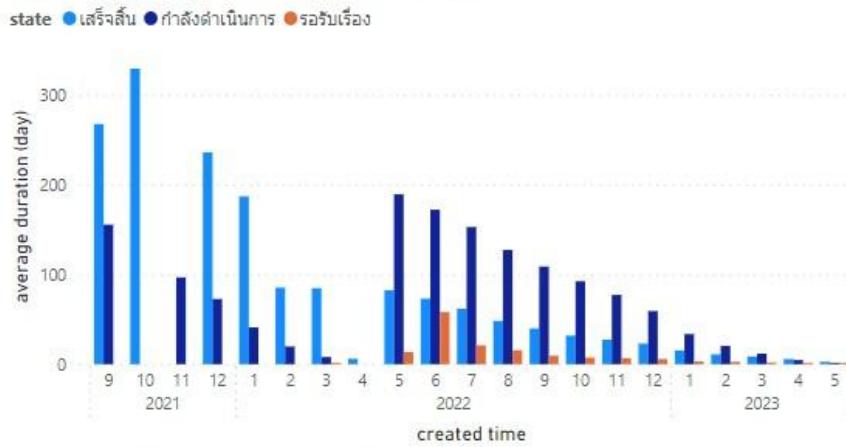
### type ratio



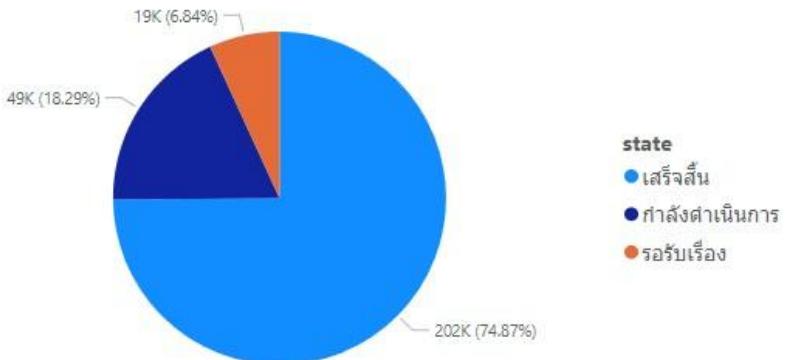
amount of report by report created time



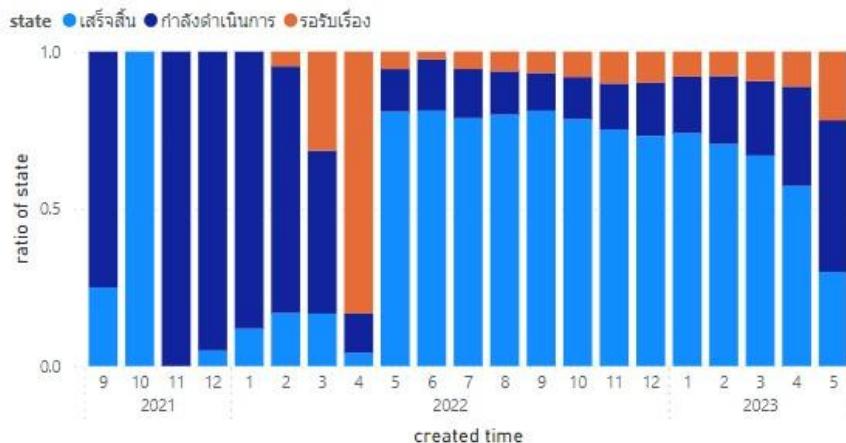
avg duration of report to reach a state by report created time

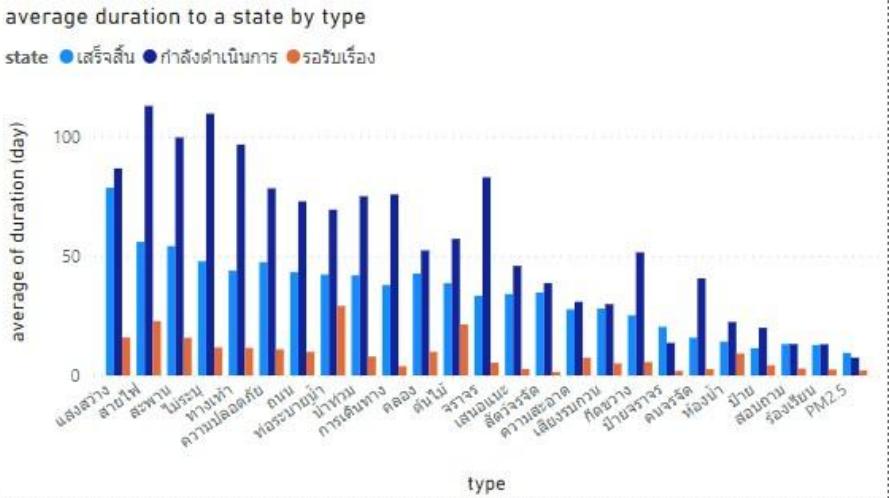
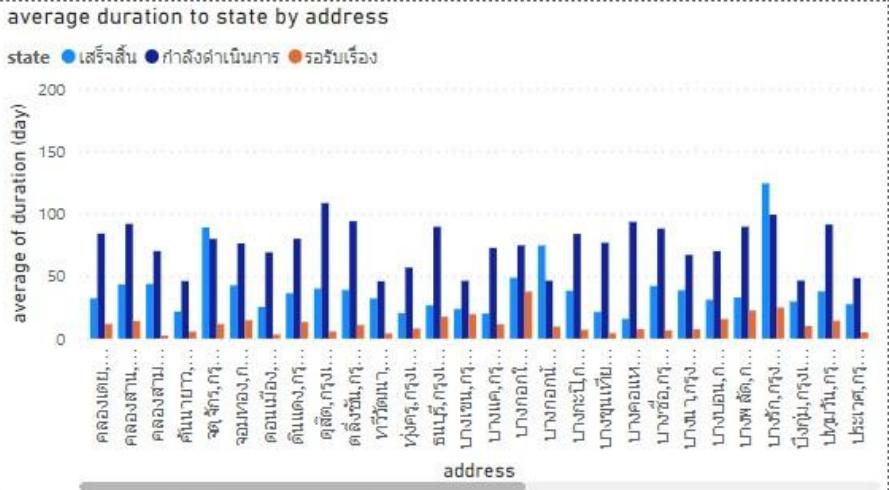
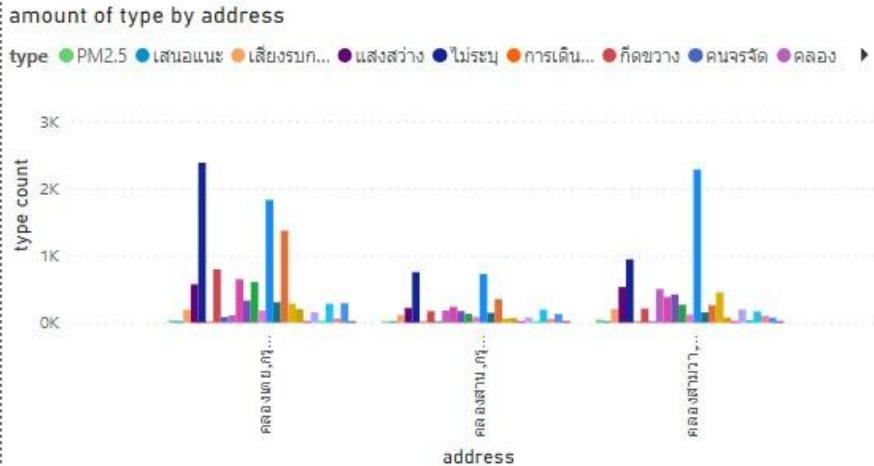
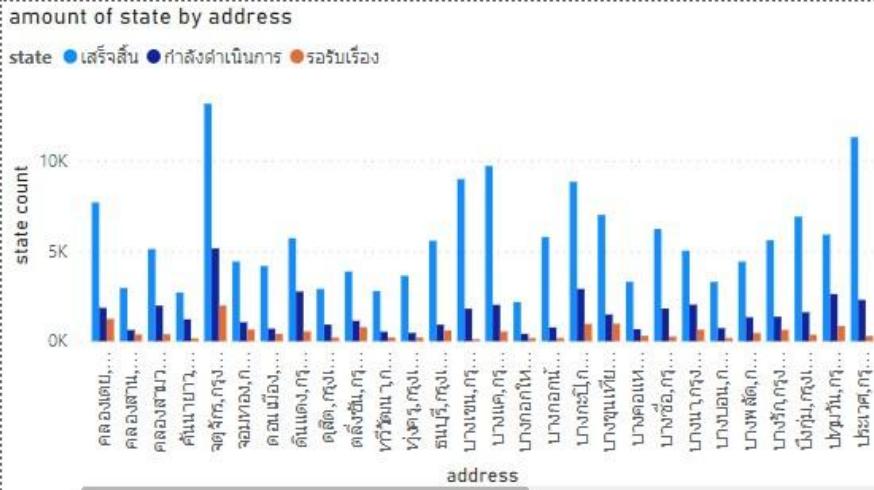


state ratio



ratio of state by report created time





# Conclusion

- การทำนายคลาดเคลื่อนจากของจริงโดยส่วนมากอยู่ที่ 1 เดือน
- สามารถนำไปคาดการณ์ว่า ณ สถานที่นี้ ปัญหา ประเภทนี้ จะใช้เวลา กี่วัน โดยเรียกใช้ fastapi
- พบรอว่าสถานะของปัญหาอยู่ในช่วงเวลา กำลังดำเนินการอยู่นาน

rounded_diff	count
0.0	11377
10.0	31499
20.0	47210
30.0	52877
40.0	42138
50.0	24355
60.0	14271
70.0	7154
80.0	3921
90.0	3142
100.0	2698
110.0	2312
120.0	2452
130.0	2361
140.0	2246
150.0	2260
160.0	2435
170.0	2413
180.0	2307
190.0	2132