

# BigtoData Analytics to support the decision making for solving the flooding problem in Bangkok

## การประมวลผลข้อมูล Bigdata เพื่อสนับสนุนการตัดสินใจวางแผนแก้ปัญหาน้ำท่วมในพื้นที่กรุงเทพฯ

FINISHED

### รายชื่อสมาชิกในกลุ่ม

- 62130700311 ธนพัฒน์ ฉัตราริน
- 62130700314 ลักษณ์ ศรีแสงธรรม
- 62130700315 วิชราภรณ์ งามแสง
- 62130700329 ธนวัฒน์ สุภานนท์

Took 3 sec. Last updated by anonymous at November 29 2020, 9:26:56 PM. (outdated)

### บทนำ

FINISHED

ประเทศไทยตั้งอยู่ในเขตมรสุม มีฝนตกปริมาณสูง มีลักษณะภูมิประเทศเป็นที่ราบลุ่มแม่น้ำ โดยมีแม่น้ำสำคัญคือ แม่น้ำเจ้าพระยา

กรุงเทพมหานครนั้น ตั้งอยู่ในพื้นที่ที่ราบลุ่มตอนปลายของแม่น้ำเจ้าพระยา ซึ่งอยู่ภายใต้อิทธิพลการขึ้น-ลงของระดับน้ำท่าทะเลในอ่าวไทย ด้วยวิธีชี้วัดในอดีตของประชาชนในพื้นที่กรุงเทพมหานคร นิยมเดินทางโดยใช้เรือ ส่งผลให้ลักษณะภูมิศาสตร์ของกรุงเทพมหานครมีแม่น้ำ ลำคลองจำนวนมากกระจายอยู่ตามพื้นที่ต่างๆ ซึ่งปัจจุบันแหล่งน้ำเหล่าท่าน้ำที่เป็นจุดหมายปริมาณน้ำลงไปสู่ทะเลทางอ่าวไทย

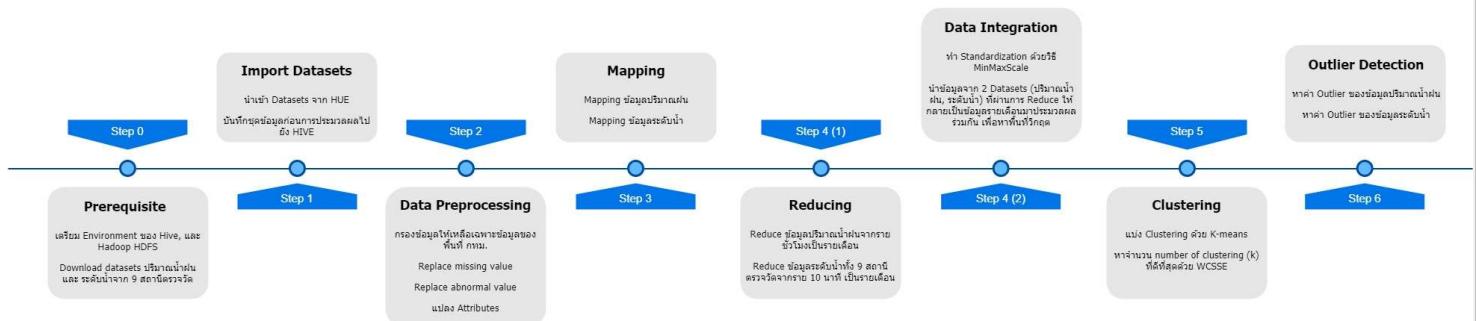
สิ่งที่น่าสนใจคือ ทำไม่เมืองที่มีแม่น้ำลำคลองน้อยในใหญ่มากรายกระจาดตัวอยู่ตามพื้นที่ต่างๆ แต่กลับประสบปัญหาน้ำท่วมซึ่งเมื่อมีฝนตก สิ่งนี้แสดงให้เห็นว่ากรุงเทพมหานครกำลังประสบปัญหาน้ำท่วมจากการระบายน้ำจากถนนลงไปยังแหล่งน้ำที่ยังไม่ได้ไม่ดีพอ ซึ่งหน่วยงานภาครัฐมีความพยายามในการแก้ไขปัญหาน้ำ ซึ่งวิธีแก้ไขปัญหาน้ำที่พบได้บ่อยมากที่สุดคือการทำความสะอาดท่อระบายน้ำ หรือการ “ลอกห้อ”

ทางกลุ่มผู้จัดทำเพื่อนำเสนอ ข้อมูลปริมาณน้ำฝน และ ข้อมูลระดับน้ำจากทั้ง 9 สถานีตรวจน้ำรอบพื้นที่กรุงเทพมหานคร มาประมวลผลและวิเคราะห์รวมกัน จะสามารถช่วยสนับสนุนการตัดสินใจได้ว่า ในแต่ละเดือนพื้นที่ไหนจะมีโอกาสออกในภาวะวิกฤตควรเริ่มดำเนินการ “ลอกห้อ” และพื้นที่ไหนที่บังสามารถระบายน้ำได้ดีอยู่ไม่ต้องการ “ลอกห้อ” เพื่อเป็นการวางแผนใช้งบให้เกิดประสิทธิภาพ

Took 0 sec. Last updated by anonymous at November 29 2020, 9:26:56 PM. (outdated)

### ขั้นตอนการประมวลผล

FINISHED



กรุณา Ctrl+Click ที่ link เพื่อดูรูปด้านฉบับ link (<https://raw.githubusercontent.com/TanaphatC/bigdata-flooding-project/master/images/processing.jpg>)

Took 0 sec. Last updated by anonymous at November 29 2020, 9:26:56 PM. (outdated)

### Step 0: Prerequisite

FINISHED

- 0.1) สร้าง Database 'flooding\_db' ที่ /Hive
- 0.2) สร้าง workspace folder '/user/zeppelin/flooding\_data' ที่ HDFS
- 0.3) เตรียม HIVE environment สำหรับการทำงาน

Took 0 sec. Last updated by anonymous at November 29 2020, 9:26:57 PM. (outdated)

## 0.1) Prerequisite - Prepare Hive Database at location '/hive'

FINISHED

Took 0 sec. Last updated by anonymous at November 29 2020, 9:44:59 PM. (outdated)

## 0.2) Prerequisite - Prepare Hadoop HDFS environment

FINISHED

/user/zeppelin/flooding\_data is existing

Took 5 sec. Last updated by anonymous at November 29 2020, 9:27:31 PM. (outdated)

## 0.3) Prerequisite - Prepare Hive environment

FINISHED

```
Drop table rainyHourly successful
Drop table waterLvBKK001 successful
Drop table waterLvBKK003 successful
Drop table waterLvBKK004 successful
Drop table waterLvBKK005 successful
Drop table waterLvBKK008 successful
Drop table waterLvBKK009 successful
Drop table waterLvBKK020 successful
Drop table waterLvBKK021 successful
Drop table waterLvCPY0015 successful
DataFrame[]
Drop table criticalScore successful
```

Took 8 sec. Last updated by anonymous at November 29 2020, 9:27:39 PM. (outdated)

## Prerequisite

FINISHED

0.4) Download ข้อมูลปริมาณน้ำฝนจาก Github และ upload ลงไปที่ folder '/user/zeppelin/flooding\_data' ใน Hadoop HDFS

ข้อมูลปริมาณน้ำฝนได้มากจากข้อมูลน้ำฝนรวมรายชั่วโมง(มิลลิเมตร) จากระบบ Telemetering กรมอุตุนิยมวิทยา(DGA Open Government, 2558)

ข้อมูลมี Meta Data ดังนี้

รายการข้อมูล	คำอธิบาย
Station_id	รหัสสถานี
Station_Name	ชื่อสถานี
Location	ที่ตั้ง
Province_Name	จังหวัด
Latitude	ละติจูด(N)
Longitude	ลองติจูด(E)
Date	วันที่
H00_01	ปริมาณฝนเวลา 00:01น.
H01_02	ปริมาณฝนเวลา 01:02น.
H02_03	ปริมาณฝนเวลา 02:03น.
H03_04	ปริมาณฝนเวลา 03:04น.
H04_05	ปริมาณฝนเวลา 04:05น.
H05_06	ปริมาณฝนเวลา 05:06น.
H06_07	ปริมาณฝนเวลา 06:07น.
H07_08	ปริมาณฝนเวลา 07:08น.
H08_09	ปริมาณฝนเวลา 08:09น.
H09_10	ปริมาณฝนเวลา 09:10น.
H10_11	ปริมาณฝนเวลา 10:11น.

รายการข้อมูล	คำอธิบาย
H11_12	ปริมาณฝนเวลา 11:12น.
H12_13	ปริมาณฝนเวลา 12:13น.
H13_14	ปริมาณฝนเวลา 13:14น.
H14_15	ปริมาณฝนเวลา 14:15น.
H15_16	ปริมาณฝนเวลา 15:16น.
H16_17	ปริมาณฝนเวลา 16:17น.
H17_18	ปริมาณฝนเวลา 17:18น.
H18_19	ปริมาณฝนเวลา 18:19น.
H19_20	ปริมาณฝนเวลา 19:20น.
H20_21	ปริมาณฝนเวลา 20:21น.
H21_22	ปริมาณฝนเวลา 21:22น.
H22_23	ปริมาณฝนเวลา 22:23น.
H23_24	ปริมาณฝนเวลา 23:24น.

Took 0 sec. Last updated by anonymous at November 29 2020, 9:27:39 PM. (outdated)

## Prerequisite

FINISHED

0.5) และ 0.6) - Download ข้อมูลระดับน้ำจากทั้ง 9 สถานีตรวจดูรอบพื้นที่กรุงเทพมหานคร จาก Github และ upload ลงไปที่ folder '/user/zeppelin/flooding\_data' ใน Hadoop HDFS

ข้อมูลระดับน้ำได้มากจากชุดข้อมูลระดับน้ำจากระบบโทรมาตรหัวประเทศไทยของสถาบันสารสนเทศทรัพยากรน้ำ (DGA Open Government, 2563)

ข้อมูลมี Meta Data ดังนี้

รายการข้อมูล	คำอธิบาย
code	รหัสสถานี
geocode	รหัสไปรษณีย์
name	ชื่อสถานีภาษาไทย
lat	ละติจูด
long	ลองจิจูด
prov_id	รหัสจังหวัด
tambon_name	ชื่อตำบลภาษาไทย
amphoe_name	ชื่ออำเภอภาษาไทย
province_name	ชื่อจังหวัดภาษาไทย
basin	ชื่อลุ่มน้ำภาษาไทย
name_e	ชื่อสถานีภาษาอังกฤษ
tambon_name_e	ชื่อตำบลภาษาอังกฤษ
amphoe_name_e	ชื่ออำเภอภาษาอังกฤษ
province_name_e	ชื่อจังหวัดภาษาอังกฤษ
wl_offset	ค่า offset ของระดับน้ำ

รายการข้อมูล	คำอธิบาย
station_type	ประเภทของสถานี
left_bank	ระดับดินซ้าย
right_bank	ระดับดินขวา

Took 0 sec. Last updated by anonymous at November 29 2020, 9:27:39 PM. (outdated)

จ้างของข้อมูลจาก:

FINISHED

DGA Open Government. (2558). ข้อมูลน้ำฝนรวมรายชั่วโมง ปี 2014 [ออนไลน์]. แหล่งที่มา:  
[https://opendata.data.go.th/en/dataset/item\\_e9ab5b50-d228-4953-ac6f-aeeed71ec751](https://opendata.data.go.th/en/dataset/item_e9ab5b50-d228-4953-ac6f-aeeed71ec751) [25 ตุลาคม 2563]

DGA Open Government. (2563). ข้อมูลระดับน้ำราย 10 นาที ปี 2014-2018 [ออนไลน์]. แหล่งที่มา:  
<https://www.data.go.th/en/dataset/set-of-water-level-by-station> (<https://www.data.go.th/en/dataset/set-of-water-level-by-station>) [17 ตุลาคม 2563]

Took 0 sec. Last updated by anonymous at November 29 2020, 9:27:40 PM. (outdated)

#### 0.4) Prerequisite - Download rainy rate datasets from Github to Hadoop HDFS

FINISHED

26550K .....	97%	159M 0s
26600K .....	97%	164M 0s
26650K .....	97%	147M 0s
26700K .....	97%	168M 0s
26750K .....	97%	167M 0s
26800K .....	98%	170M 0s
26850K .....	98%	144M 0s
26900K .....	98%	135M 0s
26950K .....	98%	164M 0s
27000K .....	98%	140M 0s
27050K .....	99%	133M 0s
27100K .....	99%	112M 0s
27150K .....	99%	154M 0s
27200K .....	99%	70.2M 0s
27250K .....	99%	60.3M 0s
27300K .....	99%	171M 0s
27350K .....	100%	141M=0.6s

Took 10 sec. Last updated by anonymous at November 29 2020, 9:27:50 PM. (outdated)

#### 0.5) Prerequisite - Download datasets from Github to Hadoop HDFS part 1

FINISHED

5850K .....	90%	20.1M 0s
5900K .....	91%	85.2M 0s
5950K .....	91%	158M 0s
6000K .....	92%	44.9M 0s
6050K .....	93%	51.5M 0s
6100K .....	94%	35.0M 0s
6150K .....	94%	83.2M 0s
6200K .....	95%	58.9M 0s
6250K .....	96%	37.5M 0s
6300K .....	97%	67.5M 0s
6350K .....	97%	21.0M 0s
6400K .....	98%	92.8M 0s
6450K .....	99%	111M 0s
6500K .....	100%	45.6M=0.2s

2020-11-29 14:28:16 (30.8 MB/s) - 'water\_level\_station\_BKK008.csv' saved [6694110/6694110]

Download water-level dataset of station 'BKK008' complete

Took 28 sec. Last updated by anonymous at November 29 2020, 9:28:18 PM. (outdated)

#### 0.6) Prerequisite - Download datasets from Github to Hadoop HDFS part 2

FINISHED

5800K .....	90%	154M 0s
5850K .....	91%	46.5M 0s
5900K .....	92%	45.1M 0s
5950K .....	93%	45.9M 0s
6000K .....	93%	45.7M 0s
6050K .....	94%	151M 0s
6100K .....	95%	53.2M 0s
6150K .....	96%	12.6M 0s

6150K ..... 96% 18.2M 0s  
6200K ..... 96% 95.0M 0s  
6250K ..... 97% 41.6M 0s  
6300K ..... 98% 179M 0s  
6350K ..... 99% 50.2M 0s  
6400K ..... 100% 38.1M=0.2s

2020-11-29 14:28:39 (29.6 MB/s) - 'water\_level\_station\_CPY015.csv' saved [6599582/6599582]

Download water-level dataset of station 'CPY015' complete

Took 23 sec. Last updated by anonymous at November 29 2020, 9:28:41 PM. (outdated)

## Step 1: Import Data

FINISHED

Took 0 sec. Last updated by anonymous at November 29 2020, 9:28:41 PM. (outdated)

### 1.1) นำเข้าข้อมูลปริมาณน้ำฝนจาก HUE สร้างเป็น Dataframe

FINISHED

### 1.2) นำเข้าข้อมูลระดับน้ำจาก 9 สถานีวัดจาก HUE และสร้างเป็น Dataframe ของแต่ละสถานีรัด โดยข้อมูลสถานีที่นำมาใช้ในการวิเคราะห์ครั้งนี้ได้แก่

รหัสสถานี	ชื่อสถานี	อำเภอ	ตำบล
BKK001	คลองลาดพร้าว ท้ายบตร.คลอง2	สายไหม	สายไหม
BKK003	คลองมหาสวัสดิ์ บางกรวย-สุนพัก	ตลึงชัน	ตลึงชัน
BKK004	คลองทวีวัฒนา ท้ายบตร.ทวีวัฒนา	ศalaธรรมสพน์	ทวีวัฒนา
BKK005	คลองภาษีเจริญ เพชรเกษม69	หลักสอง	บางแค
BKK008	คลองแสนแสบ บางกะปี	หัวหมาก	บางกะปี
BKK009	คลองล่าปลาทิว ลาดกระบัง	ลาดกระบัง	ลาดกระบัง
BKK020	คลองลาดพร้าว ปากคลอง2สายใต้	ลาดพร้าว	ลาดพร้าว
BKK021	คลองลาดพร้าว วัดบางบัว	อนุสาวรีย์	บางเขน
CPY015	สะพานกรุงเทพ	ดาวคะนอง	ธนบุรี

### 1.3) บันทึกข้อมูลก่อนการประมวลผลที่ HIVE

Took 0 sec. Last updated by anonymous at November 29 2020, 9:28:41 PM. (outdated)

## 1.1) Hourly rainfall rate dataset importing from HUE

SPARK JOBS FINISHED

Import 'Hourly Rainy Rate' dataset successful

Took 8 sec. Last updated by anonymous at November 29 2020, 9:28:50 PM. (outdated)

## 1.2) Hourly water level rate dataset importing from HUE

SPARK JOBS FINISHED

Import 'Water Level of station BKK001' dataset successful  
Import 'Water Level of station BKK003' dataset successful  
Import 'Water Level of station BKK004' dataset successful  
Import 'Water Level of station BKK005' dataset successful  
Import 'Water Level of station BKK008' dataset successful  
Import 'Water Level of station BKK009' dataset successful  
Import 'Water Level of station BKK020' dataset successful  
Import 'Water Level of station BKK021' dataset successful  
Import 'Water Level of station CPY015' dataset successful

Took 20 sec. Last updated by anonymous at November 29 2020, 9:29:10 PM. (outdated)

## 1.3) Save preprocessing datasets to HIVE

SPARK JOBS FINISHED

Save Hourly Rainy Rate before processing to HIVE successful

Save 'Water Level' dataset of 8 stations before processing to HIVE successful

Took 22 sec. Last updated by anonymous at November 29 2020, 9:29:32 PM. (outdated)

## 1.4) ทดสอบดึงข้อมูลปริมาณน้ำฝนจาก HIVE โดยใช้ SQL

## 1.5) ทดสอบดึงข้อมูลระดับน้ำจาก HIVE โดยใช้ SQL

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:32 PM. (outdated)

**1.4) Select 'Rainy Hourly Rate' from HIVE**

SPARK JOBS FINISHED

station_id	province_name_en	district	sub_district	d
4550001	Bangkok	Don Mueang	Sanam Bin	
4550001	Bangkok	Don Mueang	Sanam Bin	
4550001	Bangkok	Don Mueang	Sanam Bin	
4550001	Bangkok	Don Mueang	Sanam Bin	
4550001	Bangkok	Don Mueang	Sanam Bin	
4550001	Bangkok	Don Mueang	Sanam Bin	
4550001	Bangkok	Don Mueang	Sanam Bin	
4550001	Bangkok	Don Mueang	Sanam Bin	

Output is truncated to 102400 bytes. Learn more about ZEPPELIN\_INTERPRETER\_OUTPUT\_LIMIT

Took 1 sec. Last updated by anonymous at November 29 2020, 9:29:34 PM. (outdated)

**1.5) Select 'Water Level Rate' from HIVE**HIVE JOB (<http://hadoop-cluster-2020-msc-m.us-central1-f.c.abiding-orb-295915.internal:4040/jobs/job?id=32>) FINISHED

date	time	water_lv	station
2014-01-01 00:00:00.0	00:00:00	0.11	BKK001
2014-01-01 00:00:00.0	00:10:00	0.11	BKK001
2014-01-01 00:00:00.0	00:20:00	0.11	BKK001
2014-01-01 00:00:00.0	00:30:00	0.11	BKK001
2014-01-01 00:00:00.0	00:40:00	0.11	BKK001

Output is truncated to 1000 rows. Learn more about zeppelin.spark.maxResult

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:34 PM. (outdated)

**Step 2: Data Preprocessing**

FINISHED

**2.1) กรองข้อมูลให้เหลือแต่ข้อมูลที่ต้องการประมาณผล**

จากข้อมูลปริมาณน้ำฝนจากพื้นที่ทั่วประเทศไทยมากรองข้อมูลให้เหลือแต่ข้อมูลของกรุงเทพมหานคร โดยทำการ filter ด้วยเงื่อนไข

province\_name\_en = "Bangkok"

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:34 PM. (outdated)

**2.1) Data Preprocessing - Filter only Bangkok**

FINISHED

filter only Bangkok for Hourly Rainy Rate

## 2.2) และ 2.3) Replace missing value

FINISHED

จากข้อมูลปริมาณน้ำฝน กลุ่มผู้จัดทำพบข้อมูล NULL ในข้อมูลปริมาณน้ำฝนรายชั่วโมง ซึ่งข้อมูล NULL ได้กระจายตัวอยู่ตาม Column ชั่วโมงที่ 01:00 – 24:00 ดังนั้นกลุ่มผู้จัดทำจึงได้ทำการ Replace ข้อมูล NULL ด้วย “ค่าเฉลี่ยปริมาณน้ำฝนเฉพาะชั่วโมงที่ฝนตก” ในวันนั้น ตามทฤษฎี Statistical value of known data

$$\text{ค่าเฉลี่ยปริมาณน้ำฝนเฉพาะชั่วโมงที่ฝนตก} = \frac{\text{SUM(ปริมาณน้ำฝนทั้งวัน)}}{\text{COUNT(ชั่วโมงที่ฝนตก)}}$$

Took 0 sec. Last updated by anonymous at November 29 2020, 9:46:01 PM. (outdated)

## 2.2) Missing Value (Statistical value of known data); calculate average rainy rate of each that day

FINISHED

Calculate average rainy rate of each that day. The result of this step will be used for missing value replacement on next step

Took 2 sec. Last updated by anonymous at November 29 2020, 9:29:37 PM. (outdated)

## 2.3) Missing value (Statistical value of known data) - Replace NULL with average rainy rate of each that day

FINISHED

Replace missing value by using average rainy rate of each that day successful

Took 2 sec. Last updated by anonymous at November 29 2020, 9:29:39 PM. (outdated)

## 2.4) Assert เพื่อตรวจสอบให้แน่ใจว่าไม่มีค่า NULL เกิดขึ้นแล้วในข้อมูลปริมาณน้ำฝนรายชั่วโมง ตั้งแต่ Column ชั่วโมงที่ 01:00 till 24:00

FINISHED

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:39 PM. (outdated)

## 2.4) Missing value - Check rainy\_rate since 01:00 till 24:00 should not be NULL

FINISHED

Assert not null to ensure that all of missing values were replaced successful

Took 2 sec. Last updated by anonymous at November 29 2020, 9:29:42 PM. (outdated)

## 2.5) การแปลง Attribute

FINISHED

เนื่องจากข้อมูลระดับน้ำจากสถานีมีข้อมูลวันที่เป็นชนิด Date อยู่ในรูปแบบ “ปี-เดือน-วัน” ทางกลุ่มผู้จัดทำเห็นว่าหากไม่แปลง Attribute จะทำให้ยากต่อการประมวลผลข้อมูล จึงทำการแปลงข้อมูล Date ออกมาระบบ 3 Columns คือ “วัน”, “เดือน” และ “ปี”

เนื่องจากข้อมูลเดือนมีรูปแบบเป็นตัวเลข (Integer) ทำให้ไม่สามารถสื่อความหมายได้ดีนักเมื่อนำมาไปแสดงผลเป็นกราฟ ทางกลุ่มผู้จัดทำจึงทำการแปลงเดือนจากตัวเลข (Integer) เป็นตัวหนังสือ (String) โดยใช้ Arrays 12 ช่อง

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:43 PM. (outdated)

## 2.5) Preprocessing - Extract calendar column to day, month, year column on water-level dataset

SPARK JOBS FINISHED

Separate calendar column of Water-Level of station BKK001 successful

Separate calendar column of Water-Level of station BKK003 successful

Separate calendar column of Water-Level of station BKK004 successful

Separate calendar column of Water-Level of station BKK005 successful

Separate calendar column of Water-Level of station BKK008 successful

Separate calendar column of Water-Level of station BKK009 successful

Separate calendar column of Water-Level of station BKK020 successful

Separate calendar column of Water-Level of station BKK021 successful

Separate calendar column of Water-Level of station CPY015 successful

Took 1 sec. Last updated by anonymous at November 29 2020, 9:29:44 PM. (outdated)

## 2.6) Replace abnormal value

FINISHED

จากข้อมูลระดับน้ำจาก 9 สถานีรอดพบร่วมค่าระดับน้ำที่ผิดปกติอยู่คือ -999 ซึ่งกลุ่มผู้จัดได้ Replace ด้วยค่า 0.0

ระดับน้ำ > 0.0 : ระดับน้ำในแหล่งน้ำ สูงกว่า ระดับน้ำทะเล

ระดับน้ำ = 0.0 : ระดับน้ำในแหล่งน้ำ เท่ากับ ระดับน้ำทะเล

ระดับน้ำ < 0.0 : ระดับน้ำในแหล่งน้ำ ต่ำกว่า ระดับน้ำทะเล

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:45 PM. (outdated)

## 2.6) Cleansing Data - Replease abnormal value -999 of water-level

SPARK JOBS FINISHED

Replace abnormal value in Water-Level of station BKK001 successful  
Replace abnormal value in Water-Level of station BKK003 successful  
Replace abnormal value in Water-Level of station BKK004 successful  
Replace abnormal value in Water-Level of station BKK005 successful  
Replace abnormal value in Water-Level of station BKK008 successful  
Replace abnormal value in Water-Level of station BKK009 successful  
Replace abnormal value in Water-Level of station BKK020 successful  
Replace abnormal value in Water-Level of station BKK021 successful  
Replace abnormal value in Water-Level of station CPT015 successful

Took 5 sec. Last updated by anonymous at November 29 2020, 9:29:50 PM. (outdated)

## Step 3: Mapping

FINISHED

### 3.1) Map ข้อมูลจาก ผลกระทบปริมาณน้ำฝน และค่าเฉลี่ยน้ำฝนของแต่ละวัน

RainyRate: Array จำนวน 24 ช่อง โดยเก็บข้อมูลปริมาณน้ำฝนจากชั่วโมงที่ 1-24

DayRainyRate: ผลกระทบปริมาณน้ำฝนจาก Array จำนวน 24 ช่อง หรือผลกระทบปริมาณน้ำฝนในวันนั้นนั้นเอง

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:50 PM. (outdated)

### 3.1) Mapping - Finding summary and average rainy rate each that day

SPARK JOBS FINISHED

StationId	Location	Day	MonthNo	MonthName	Year	RainyRate	DayRainyRate
4550006	Khlong Sam Wa/Sam...	1	5	May	2014	[0.0, 0.0, 2.13, ...]	113.21
4550014	Nong Chok/Krathum...	8	5	May	2014	[0.0, 0.0, 0.0, 3...]	61.06
4550003	Lak Si/Thung Song...	8	5	May	2014	[0.0, 4.54, 3.95,...]	58.82
4550004	Bang Khen/Tha Raeng	8	5	May	2014	[0.0, 0.0, 4.4, 2...]	58.12
4550004	Bang Khen/Tha Raeng	1	5	May	2014	[0.0, 0.0, 1.14, ...]	56.15
4550025	Bang Khun Thian/S...	1	5	May	2014	[0.0, 0.0, 4.67, ...]	56.01
4550023	Lat Krabang/Lam P...	1	5	May	2014	[0.0, 0.0, 0.0, 3...]	55.55
4550007	Nong Chok/Khu Fan...	8	5	May	2014	[0.0, 0.0, 0.0, 3...]	53.81
4550024	Bang Bon/Bang Bon	8	5	May	2014	[3.34, 0.1, 0.72,...]	52.82
4550022	Prawet/Nong Bon	8	5	May	2014	[3.88, 2.26, 4.72...]	52.53
4550026	Bang Khun Thian/T...	1	5	May	2014	[0.0, 0.0, 0.0, 2...]	51.51
4550015	Phaya Thai/Sam Se...	1	5	May	2014	[0.0, 0.0, 0.0, 4...]	51.07
4550005	Khlong Sam Wa/Sam...	8	5	May	2014	[0.0, 2.19, 0.61,...]	50.6
4550003	Lak Si/Thung Song...	1	5	May	2014	[0.0, 0.0, 0.28, ...]	49.84
4550008	Khlong Toei/Khlon...	1	5	May	2014	[0.0, 0.0, 0.0, 0...]	49.51

Took 4 sec. Last updated by anonymous at November 29 2020, 9:29:54 PM. (outdated)

## Step 4: Reducing

FINISHED

การ Reduce จะถูกแบ่งออกเป็น 2 Phases ได้แก่

Phase #1: Reduce ข้อมูลปริมาณน้ำฝนรายวัน และ ข้อมูลระดับน้ำรายวัน ในหากลายเป็นข้อมูลรายเดือน

(input) ข้อมูลปริมาณน้ำฝนรายวัน => Reduce => (output) ข้อมูลปริมาณน้ำฝนรายเดือน

(input) ข้อมูลระดับน้ำรายวัน => Reduce => (output) ข้อมูลระดับน้ำรายเดือน

Phase #2: นำ outputs จาก Reduce Phase #1 มา integrate กัน และทำการ Reduce อีกครั้งเพื่อได้ข้อมูลสรุปปริมาณน้ำฝนเทียบกับระดับน้ำของแต่ละพื้นที่

(input) ข้อมูลปริมาณน้ำฝนรายเดือน, ข้อมูลระดับน้ำรายเดือน => Reduce => (output) ข้อมูลสรุปปริมาณน้ำฝนเทียบกับระดับน้ำของแต่ละพื้นที่

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:54 PM. (outdated)

### 4.1) Reduce ข้อมูลปริมาณน้ำฝน โดยการหา Summary ด้วยเงื่อนไข รหัสสถานี, พื้นที่, เดือน และปี ตั้ง code ต่อไปนี้

FINISHED

```
.groupBy(["StationId", "Location", "MonthName", "MonthNo", "Year"])
```

Output จากการ summary จะถูกเก็บลงใน column sum(DayRainyRate)

Took 0 sec. Last updated by anonymous at November 29 2020, 9:29:54 PM. (outdated)

## 4.1) Reduce - Summary rainy rate by month and year

SPARK JOBS FINISHED



settings ▾

StationId	Location	MonthName	MonthNo	Year	☰
4550019	Lat Krabang/Khlong Sam Prawet	February	2	2014	
4550025	Bang Khun Thian/Samae Dam	January	1	2014	
4550004	Bang Khen/Tha Raeng	May	5	2014	
4550018	Saphan Sung/Saphan Sung	May	5	2014	
4550016	Lat Krabang/Lam Pla Thio	March	3	2014	
4550004	Bang Khen/Tha Raeng	January	1	2014	
4550026	Bang Khun Thian/Tha Kham	March	3	2014	
4550026	Bang Khun Thian/Tha Kham	April	4	2014	
4550027	Khlong Toei/Khlong Toei	July	7	2014	
4550014	Nong Chok/Krathum Rai	January	1	2014	
4550001	Don Mueang/Sanam Bin	March	3	2014	
4550018	Saphan Sung/Saphan Sung	April	4	2014	
4550024	Bang Bon/Bang Bon	April	4	2014	

DataFrame[StationId: bigint, Location: string, MonthName: string, MonthNo: string, Year: string, sum(DayRainyRate): double]

Took 13 sec. Last updated by anonymous at November 29 2020, 9:30:07 PM. (outdated)

## 4.2) Reduce ข้อมูลระดับน้ำจากทั้ง 9 สถานีโดยการหา ค่าเฉลี่ย (mean) ของระดับน้ำในแต่ละเดือน

FINISHED

(input) ข้อมูลระดับน้ำจากทั้ง 9 สถานี => Reduce => (output) ข้อมูลค่าเฉลี่ยระดับน้ำในแต่ละเดือนทั้งหมด 9 สถานี

ทำการ union ข้อมูลระดับน้ำเฉลี่ยในแต่ละเดือนของทั้ง 9 สถานี รวมเป็น 1 dataset

(input) ข้อมูลระดับน้ำเฉลี่ยในแต่ละเดือนของทั้ง 9 สถานี => Reduce => ข้อมูลระดับน้ำเฉลี่ยในแต่ละเดือนของสถานีทั้งหมดในพื้นที่กรุงเทพฯ

Output:

- Station: รหัสสถานี
- 1: ค่าเฉลี่ยระดับน้ำในเดือนมกราคม
- 2: ค่าเฉลี่ยระดับน้ำในเดือนกุมภาพันธ์
- ...
- 12: ค่าเฉลี่ยระดับน้ำในเดือนธันวาคม

Took 0 sec. Last updated by anonymous at November 29 2020, 9:30:07 PM. (outdated)

## 4.2) Reduce - Find average of water-level dataset from 9 stations and union them all to one finally dataset.

SPARK JOBS FINISHED

Station 'BKK001' was completely reduced  
 Station 'BKK003' was completely reduced  
 Station 'BKK004' was completely reduced  
 Station 'BKK005' was completely reduced  
 Station 'BKK008' was completely reduced  
 Station 'BKK009' was completely reduced  
 Station 'BKK020' was completely reduced  
 Station 'BKK021' was completely reduced  
 Station 'CPY015' was completely reduced



settings ▾

Station	1	2	3	4	5	☰

BKK001	-0.033736559139785 224	-0.094890873015872 66	-0.067181899641576 14	-0.121002314814813 37	-0.100056003584229 3	-4.8 E-4
BKK003	0.7388037634408569	0.7790749007936503	0.8023745519713245	0.6482222222222206	0.5563821684587792	0.39 6
BKK004	0.5165412186379946	0.5463070436507858	0.5184923835125449	0.4258680555555585	0.3643862007168473	0.23 4
BKK005	0.2557974910394269 5	0.2699975198412655	0.254348118279568	0.1880324074074083	0.1551164874551973	0.08 8
BKK008	0.0533221326164879	0.0252802579365078	0.0042921146953404	-0.0976597222222222	-0.176823476702511	-0.1

Took 4 min 39 sec. Last updated by anonymous at November 29 2020, 9:34:46 PM. (outdated)

#### 4.3) นำผลลัพธ์การ Reduce จากข้อ 4.1 และ 4.2 มา integrate กันด้วย relationship ระหว่าง 2 fields ดังนี้

FINISHED

Field	Description	Dataset	Example Data
StationId	รหัสสถานีตรวจวัดปริมาณน้ำฝน	ข้อมูลปริมาณน้ำฝนจากขั้นตอนที่ 4.1	4550019, 4550025, 4550008
StationCode	รหัสสถานีตรวจวัดระดับน้ำ	ข้อมูลค่าเฉลี่ยระดับน้ำจากขั้นตอนที่ 4.2	BKK001, BKK004, BKK021, CPY015

โดยทางกลุ่มได้นำข้อมูลเพื่อจับคู่ว่า สถานีตรวจวัดปริมาณน้ำฝน ใดอยู่ในพื้นที่เดียวกันกับ สถานีตรวจวัดระดับน้ำ และทำเป็น Table Mapping ไว้ดังนี้

รหัสสถานีตรวจวัดระดับน้ำ	ชื่อสถานีตรวจวัดระดับน้ำ	รหัสสถานีตรวจวัดปริมาณน้ำฝนที่อยู่ในพื้นที่
BKK001	คลองลาดพร้าว ท้ายปต.คลอง2	4550004, 4550005, 4550006, 4550020
BKK004	คลองทรีรัตน์ ท้ายปต.ทรีรัตน์	4550007
BKK005	คลองภาษีเจริญ เพชรเกษม69	4550021, 4550024, 4550025, 4550026
BKK008	คลองแ سنแสน บางกะปิ	4550003, 4550017, 4550018, 4550022, 4550027
BKK009	คลองลำปลาทิว ลาดกระบัง	4550002, 4550009, 4550012, 4550013, 4550014, 4550016, 4550019, 4550023
BKK021	คลองลาดพร้าว วัดบางบัว	4550001, 4550008, 4550010, 4550011
CPY015	สะพานกรุงเทพ	4550015, 4550028

Took 0 sec. Last updated by anonymous at November 29 2020, 9:34:46 PM. (outdated)

#### เมื่อนำข้อมูลมา plot กราฟด้วย แกน X = ปริมาณน้ำฝน, แกน Y = ระดับน้ำจากสถานี โดยจัดกลุ่มตามพื้นที่ จะได้กราฟดังนี้

FINISHED

จะสังเกตได้ว่าข้อมูลที่ plot ในกราฟนั้นดูยากเนื่องจาก ลักษณะข้อมูลที่มีความแตกต่างกันมากเกินไป

- ข้อมูลปริมาณน้ำฝนมีช่วงข้อมูลอยู่ที่ 0.0 – 200.00
- ข้อมูลระดับน้ำมีช่วงตั้งแต่ -2.00 – 2.00

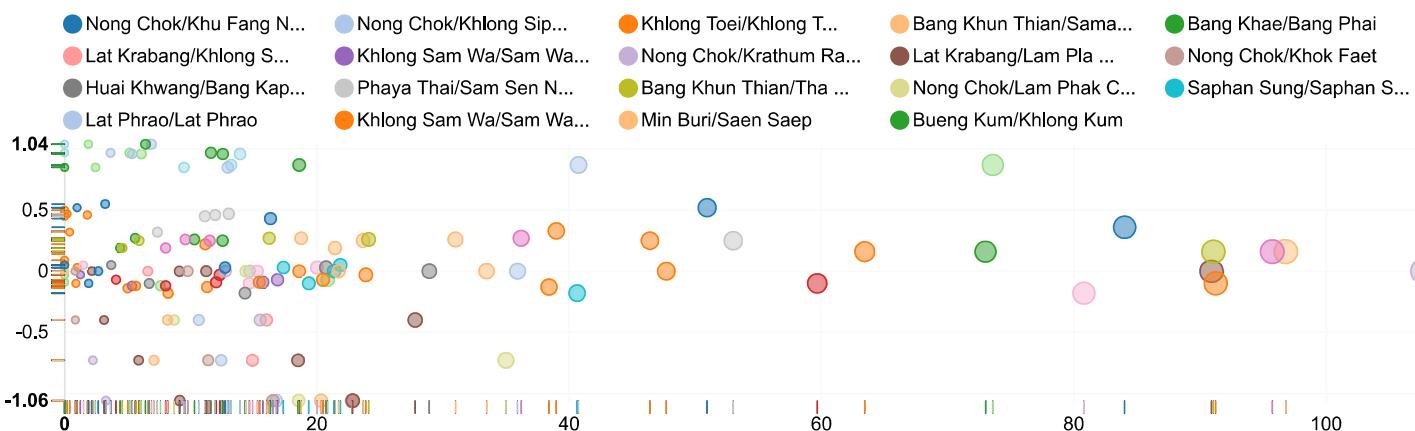
Took 0 sec. Last updated by anonymous at November 29 2020, 9:34:47 PM. (outdated)

#### 4.3) Reduce - 2 Sources integration. Rainy rate data & Water-level data

SPARK JOBS FINISHED



settings ▾



Took 2 min 10 sec. Last updated by anonymous at November 29 2020, 9:36:57 PM. (outdated)

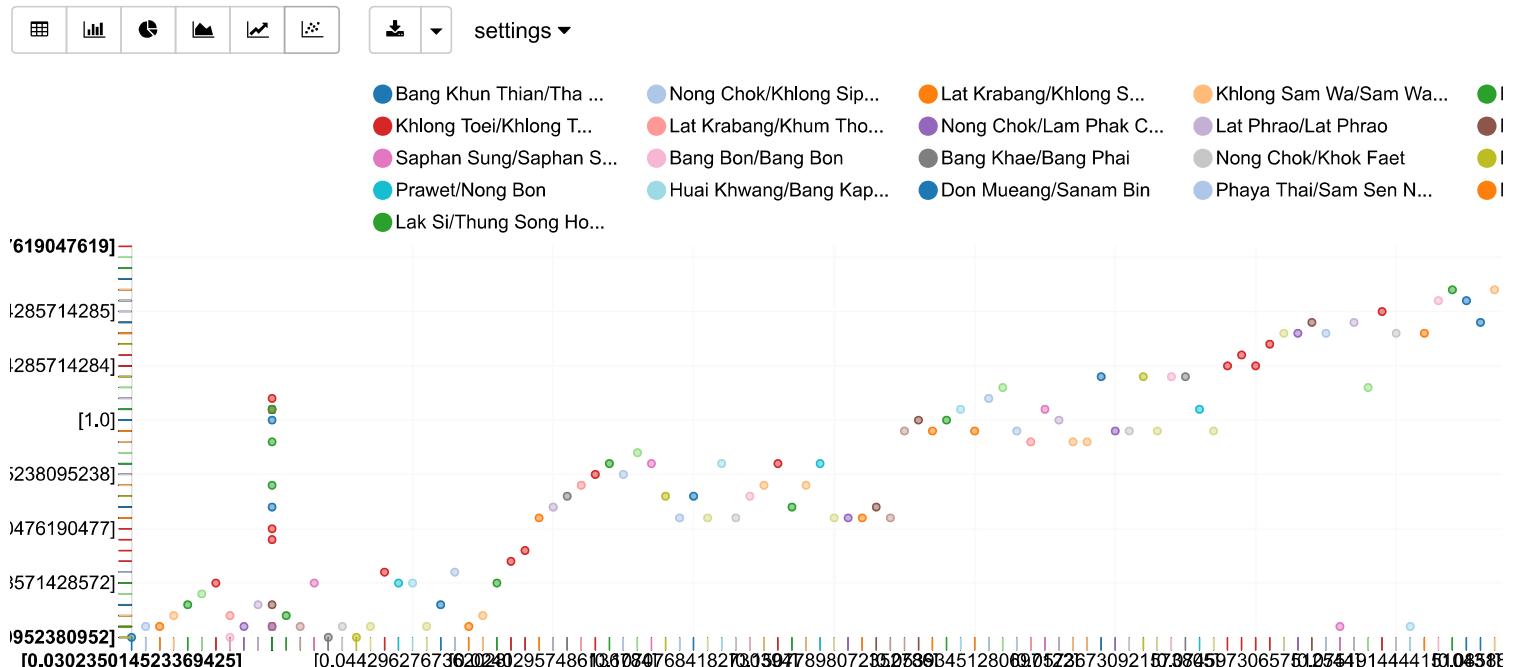
### การทำ Standardization ข้อมูล “ปริมาณน้ำฝน” และ “ระดับน้ำ” เพื่อให้ข้อมูลทั้ง 2 อุปกรณ์ในช่วงเดียวกันสามารถเปรียบเทียบกันได้ โดยใช้ MinMaxScaler

หลักจากทำ Standardization ด้วย MinMaxScaler พบว่าเมื่อในข้อมูลไป plot กราฟแล้วง่ายต่อความเข้าใจมากขึ้น โดยสามารถสรุปการวิเคราะห์ข้อมูลได้ดังนี้ ดังนี้

Pattern	Problem	Locations	Graph's Position
ปริมาณน้ำฝน สูง ระดับน้ำสูง	<b>พื้นที่ไม่มีปัญหา</b> แม้ฝนตกหนักแต่ยังสามารถระบายน้ำลงไปสู่แหล่งน้ำได้ดี	<ul style="list-style-type: none"> <li>เขตคลองเตย/แขวงคลองเตย (ในบางเดือน)</li> <li>เขตหนองจอก/แขวงคุ้งคิตติ์เหมือน</li> <li>เขตบางขุนเทียน/แขวงแสมดำและท่าข้าม</li> <li>เขตลาดพร้าว/แขวงลาดพร้าว</li> <li>เขตบางบอน/บางบอน</li> <li>เขตสัมพันธวงศ์/แขวงทุ่งสองห้อง</li> <li>อื่นๆ</li> </ul>	มุมขวาบนของกราฟ
ปริมาณน้ำฝน สูง ระดับน้ำต่ำ	<b>พื้นที่มีปัญหา</b> เนื่องจากไม่สามารถระบายน้ำฝนลงไปยังแหล่งน้ำได้ อาจเกิดจากปัญหาท่อระบายน้ำตัน	<ul style="list-style-type: none"> <li>เขตคลองสามวา/แขวงสามวาตะวันตก</li> <li>เขตหนองจอก / แขวงกระทุมราย</li> <li>เขตลาดกระบัง / แขวงลำล้าทิว</li> <li>เขตหนองจอก / แขวงโคลแฝด</li> <li>อื่นๆ</li> </ul>	มุมขวาล่างของกราฟ
ปริมาณน้ำฝน ต่ำ ระดับน้ำสูง	<b>พื้นที่เฝ้าระวัง</b> เนื่องจากฝนตกน้อย แต่ระดับน้ำสูง บ่อบอกถึงปัจจัยอื่น เช่นภาวะน้ำทะเลหมุนหรือน้ำเหนือ	<ul style="list-style-type: none"> <li>เขตเมืองกุ้ม / แขวงคลองกุ้ม</li> <li>เขตลาดกระบัง / แขวงคลองสามประเวศ</li> <li>เขตหนองจอก / แขวงลำพักซี</li> <li>เขตคลองเตย/แขวงคลองเตย (ในบางเดือน)</li> <li>อื่นๆ</li> </ul>	มุมซ้ายล่างของกราฟ
ปริมาณน้ำฝน ต่ำ ระดับน้ำต่ำ	<b>พื้นที่ไม่มีปัญหา</b> เนื่องจากฝนตกน้อย และ ระดับน้ำต่ำ	<ul style="list-style-type: none"> <li>เขตบางขุนเทียน / แขวงท่าข้าม (ในบางเดือน)</li> <li>เขตสะพานสูง / แขวงสะพานสูง</li> <li>เขตบางบอน / แขวงบางบอน</li> <li>เขตหนองจอก / แขวงคลองสิบสอง</li> <li>เขตบางแค / แขวงบางไผ</li> <li>อื่นๆ</li> </ul>	มุมซ้ายล่างของกราฟ

#### 4.3) Standardization MonthlyRainyRate and MonthlyWaterLevel by MinMaxScaler

### ■ SPARK JOBS FINISHED



Took 21 sec. Last updated by anonymous at November 29 2020, 9:37:18 PM (outdated)

4.4) Reduce เพื่อหาค่าเสียงเกิดปัณฑตหน้าทวมของแต่ละพื้นที่ โดยใช้สมการดังนี้

FINISHED

ค่าความเสี่ยง = ปริมาณน้ำฝน / ระดับน้ำ

หากผลลัพธ์ยังมีค่าสูง แสดงว่าพื้นที่นั้น ยังมีความเสี่ยงที่จะเกิดปัญหาน้ำท่วมสูงนั่นเอง

จากการฟังเสียงในห้องเรียนพุทธศาสนา มีนัยที่พูดถึงความเสียชั้นสูง ซึ่งสอดคล้องกับถูกกาลของประเทศไทยที่คุณภาพเริ่มต้นตั้งแต่เดือนพฤษภาคมไปจนถึงกลางเดือนตุลาคม (กรกฎาคม-กันยายน)

Took 0 sec. Last updated by anonymous at November 29 2020, 9:37:18 PM. (outdated)

อ้างอิงข้อมูลจาก:

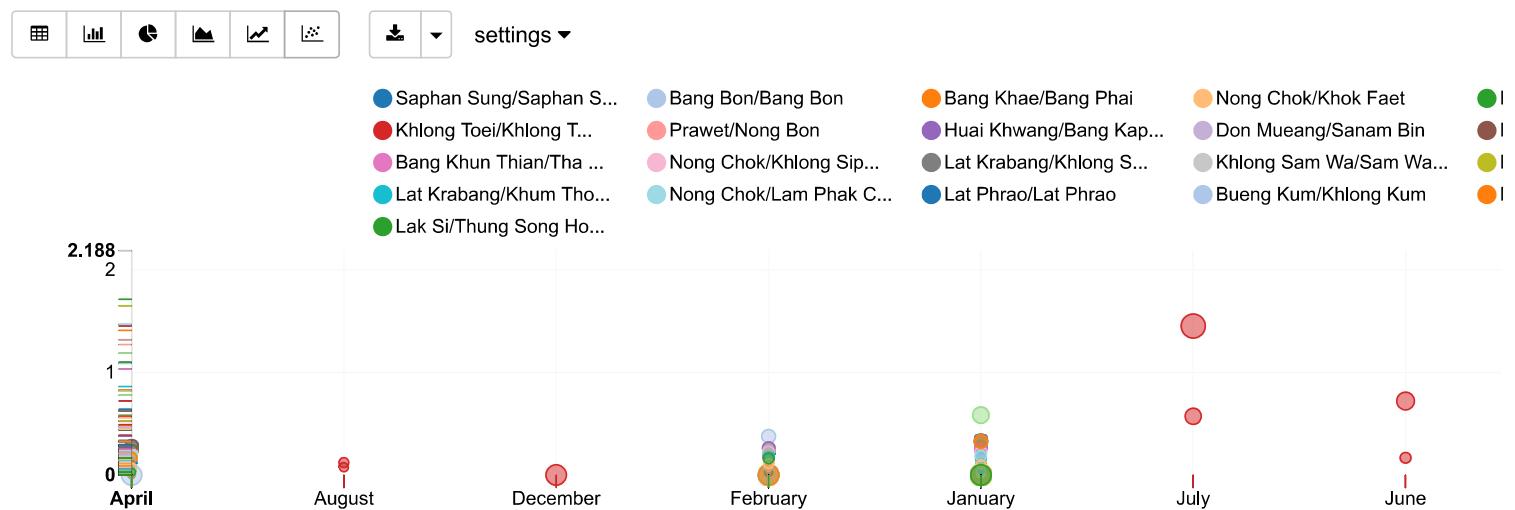
FINISHED

กรมอุตุนิยมวิทยา หนังสืออุตุนิยมวิทยา ถูกุกกาลของประเทศไทย [ออนไลน์]. แหล่งที่มา: <https://www.tmd.go.th/info/info.php?FileID=53> (<https://www.tmd.go.th/info/info.php?FileID=53>) [19 พฤษภาคม 2563]

Took 0 sec | Last updated by anonymous at November 29 2020 9:37:18 PM (outdated)

#### **4.4) Reduce - Finding critical area on each month**

SPARK JOBS FINISHED



## 4.5) บันทึกข้อมูลหลังจากการประมวลผลใน HIVE

บันทึกลงใน HIVE ตารางที่ชื่อว่า flooding\_db.criticalScore สำหรับเรียนรู้อย่างไร

Took 2 sec. Last updated by anonymous at November 29 2020, 9:37:27 PM. (outdated)

## Step 5: Clustering

FINISHED

### 5.1) จัดกลุ่มข้อมูลโดยการแบ่ง Clustering ด้วยวิธี K-means และทำการ validate จำนวน Clustering ด้วย WCSSE และ Elbow

จากการลองจัดกลุ่มข้อมูลด้วย K=2 จนถึง K=20 และหา WCSSE เพื่อ plot กราฟ

พบว่าความชันของกราฟคงที่ตั้งแต่ K=20 ไปลงมาจนถึง K=6 และ ความชันเริ่มเปลี่ยนแปลงตั้งแต่ K=5 เป็นต้นไป

ดังนั้นสรุปได้ว่าจำนวน Clusters ที่เหมาะสมที่สุดคือ 5 Clusters

Took 0 sec. Last updated by anonymous at November 29 2020, 9:37:27 PM. (outdated)

### 5.1) Clustering by using K-means & Validate Clustering by using WCSSE

SPARK JOBS FINISHED



Took 55 sec. Last updated by anonymous at November 29 2020, 9:38:22 PM. (outdated)

## Step 6: Outlier Detection

FINISHED

### 6.1) หา Outlier จากข้อมูล ผลกระทบปริมาณน้ำฝนรายเดือนแบ่งตามพื้นที่พบร่วม

ผลกระทบปริมาณน้ำฝน  $> 46.56$  = Upper Bound

ผลกระทบปริมาณน้ำฝน  $< -19.68$  = Lower Bound

### 6.2) หา Outlier จากข้อมูล ค่าเฉลี่ยระดับน้ำรายเดือนแบ่งตามพื้นที่พบร่วม

ผลกระทบปริมาณน้ำฝน  $> 0.819548917131$  = Upper Bound

ผลกระทบปริมาณน้ำฝน  $< -0.350834284283$  = Lower Bound

Took 0 sec. Last updated by anonymous at November 29 2020, 9:38:23 PM. (outdated)

### 6.1) Outlier Detection - Rainy Rate

SPARK JOBS FINISHED

```
+-----+-----+
|summary| sum(DayRainyRate)|
+-----+-----+
| min|      0.0|
| max|    151.48|
| 25%|      5.16|
| 75%| 21.72000000000002|
```

```
| mean| 21.69893333333333  
| stddev| 28.49236933398102  
+-----+
```

IQR: 16.56  
Upper Bound: 46.56 / Lower Bound: -19.68



settings ▾

StationId	Location	MonthName	MonthNo	Year	
4550018	Saphan Sung/Saphan Sung	May	5	2014	40.
4550018	Saphan Sung/Saphan Sung	April	4	2014	19.
4550018	Saphan Sung/Saphan Sung	February	2	2014	17.
4550018	Saphan Sung/Saphan Sung	March	3	2014	21.
4550018	Saphan Sung/Saphan Sung	January	1	2014	21.

Took 7 sec. Last updated by anonymous at November 29 2020, 9:38:30 PM. (outdated)

## 6.2) Outlier Detection - Water level

SPARK JOBS FINISHED

```
+-----+  
|summary|      inYearAverage|  
+-----+  
|   min| -0.21126636949209213|  
|   max|  1.048509819739572|  
| 25%|  0.08805941624708387|  
| 75%|  0.38065521660052815|  
|   mean|  0.30638183549919973|  
| stddev|  0.37306760760637947|  
+-----+
```

IQR: 0.292595800353  
Upper Bound: 0.819548917131 / Lower Bound: -0.350834284283

```
+-----+  
|Station|      inYearAverage|Outlier|  
+-----+  
| BKK001|  0.08805941624708387| normal|  
| BKK003|  0.60226148740185961| normal|  
+-----+
```

Took 3 min 45 sec. Last updated by anonymous at November 29 2020, 9:42:15 PM. (outdated)

## งานวิจัยที่เกี่ยวข้อง

FINISHED

จากแผนปฏิบัติการป้องกันและแก้ไขปัญหาน้ำท่วมกรุงเทพมหานคร พบว่าสาเหตุของน้ำท่วมเกิดได้จากการทางธรรมชาติเช่น น้ำฝน, น้ำทุ่ง, น้ำเนิน, น้ำทะเล หนอง และระดับน้ำในแม่น้ำเจ้าพระยา และสาเหตุทางกายภาพได้แก่ ปัญหาผังเมือง, ปัญหาการระบายน้ำ และปัญหาแฟ่นดินทรุด

จากหัวข้อ 7.2 แผนป้องกันน้ำท่วมนี้องจากฝันตอพนวฯ ภารกิจที่สำนักการระบายน้ำได้ทำมากที่สุดในปีได้แก่

- การตรวจสอบเครื่องสูบน้ำ
- การดำเนินการเปิดทางน้ำให้ลงคลอง
- การทำความสะอาดหอรระบายน้ำ

ชี้ 3 กิจกรรมดังกล่าวจะทำติดต่อกันเป็นระยะเวลาทั้งหมด 8 เดือน (มี.ค. - ต.ค.)

ดังนั้นหากมีการน้ำข้อมูลสถิติปริมาณน้ำฝนและระดับน้ำมาทำการวิเคราะห์แบบแยกตามพื้นที่และแยกตามช่วงเวลา เพื่อประกอบการตัดสินใจวางแผนแก้ไข ปัญหาน้ำท่วม จะส่งผลให้กระบวนการแก้ปัญหาน้ำท่วมเกิดประสิทธิภาพสูงสุด

อ้างอิง:

สำนักการระบายน้ำ. (2561). แผนปฏิบัติการป้องกันและแก้ไขปัญหาน้ำท่วมกรุงเทพมหานคร เมื่อจากน้ำฝนและน้ำทุ่น ประจำปี 2561. [ออนไลน์]. แหล่งที่มา: [http://203.155.220.119/News\\_dds/magazine/Plan61/05.pdf](http://203.155.220.119/News_dds/magazine/Plan61/05.pdf) ([http://203.155.220.119/News\\_dds/magazine/Plan61/05.pdf](http://203.155.220.119/News_dds/magazine/Plan61/05.pdf))

