

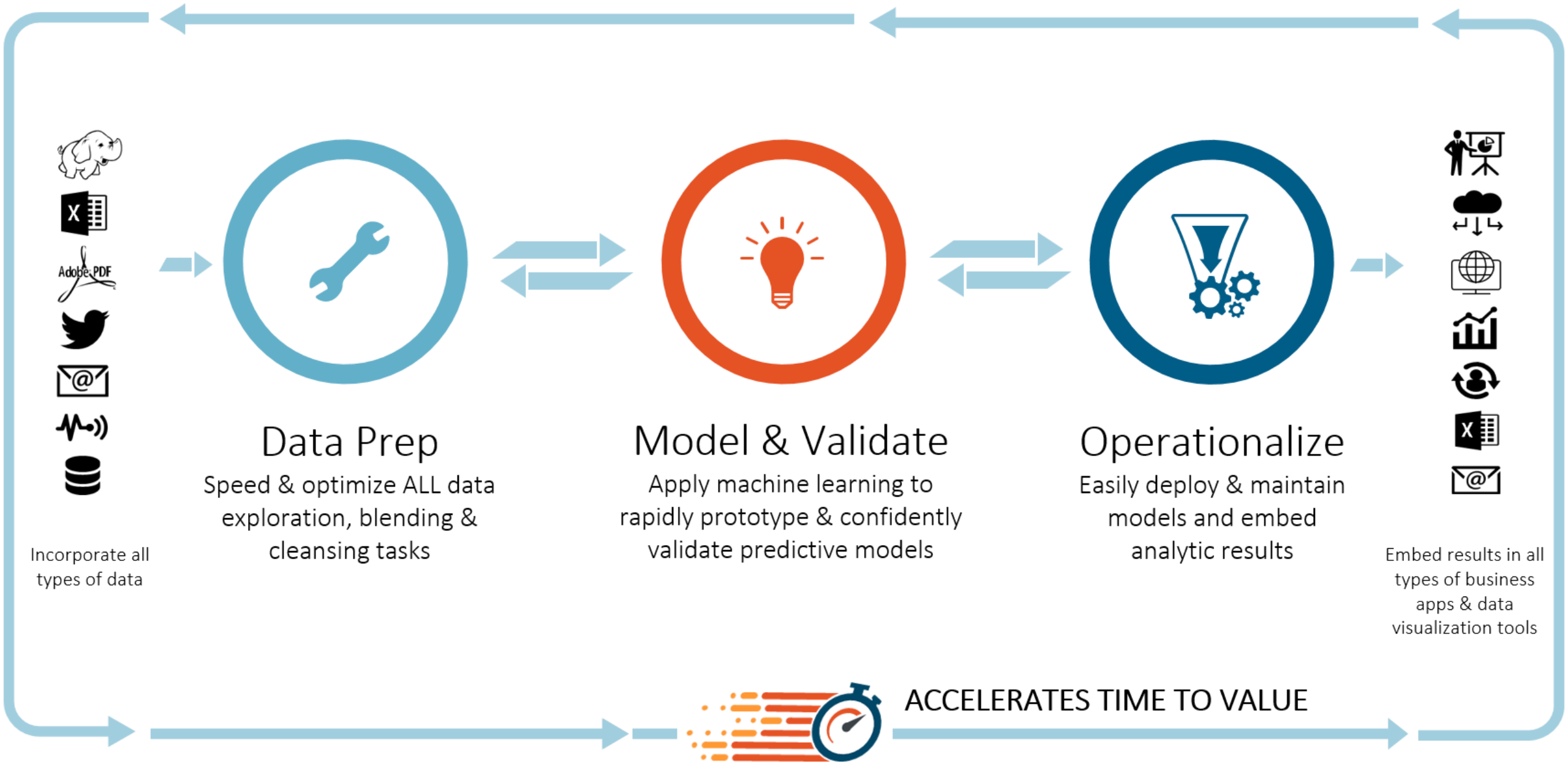
Twitter Crawler and Analysis

By

Sarawoot Kongyoung

sansarn@gmail.com

Data Science Process:



Data Preparation



Twitter คือ เป็นบริการเครือข่ายสังคมออนไลน์ที่จำพวกไม่โครบล็อก โดยผู้ใช้งานสามารถ
ส่งข้อความยาวไม่เกิน 280 ตัวอักษร ว่าตนเองกำลังทำอะไรอยู่ โดยเรียกการส่ง
ข้อความนี้ว่า ทวิต (อังกฤษ: Tweet) ซึ่งแปลว่า เสียงนกร้อง

Twitter Trend

ความนิยมสำหรับคุณ · [ปรับเปลี่ยน](#)

[#บุพเพสันนิวาส](#)

@mktmag, @thidakarn และอีก 1 คนกำลังทวีตเกี่ยวกับสิ่งนี้

[#DragRaceTH](#)

5,146 ทวีต

[#themasksinger4](#)

@DekpurjurPINNOA กำลังทวีตเกี่ยวกับสิ่งนี้

[#Aline_A](#)

5,027 ทวีต

[#thairath](#)

@Thairath_News กำลังทวีตเกี่ยวกับสิ่งนี้

[#Ch7news](#)

[#นาฬิกาหรร](#)

@Mr_sumeth และ @MatichonOnline กำลังทวีตเกี่ยวกับสิ่งนี้

[#พุทธทอล์คพุทธโธ](#)

30.4K ทวีต

[#ดาวจรัสฟ้า](#)

2,041 ทวีต

Pop.ssense.in.th

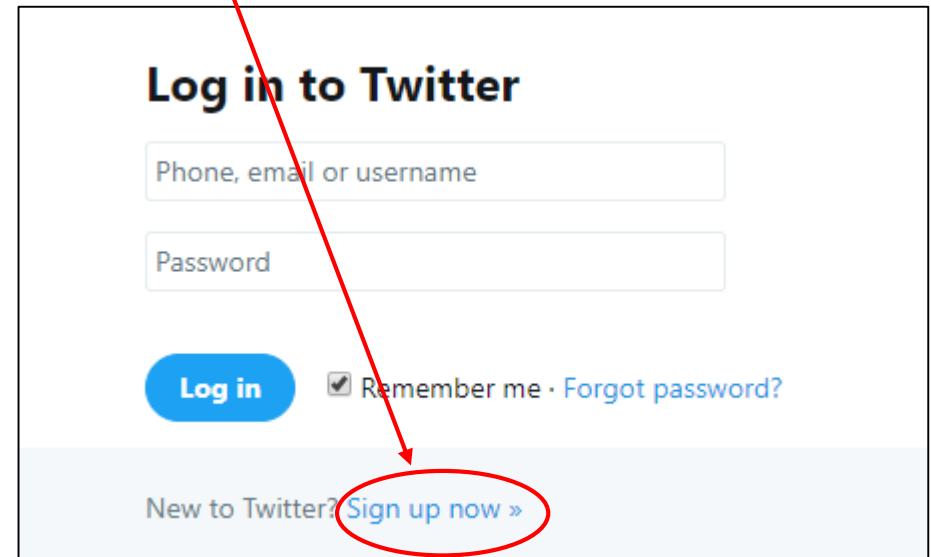
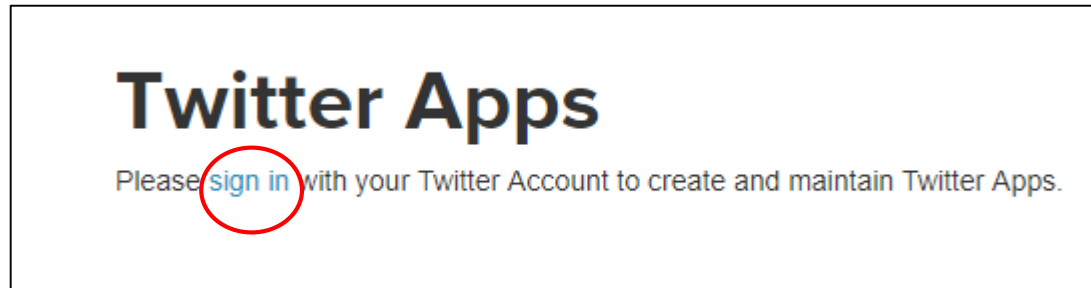
Related TagCloud

โป๊ปธนวรรธ โป๊ปธนวรรธ
หลวงสรศักดิ์ bellacampen คมแฝก นายณภัทร
โป๊ปเบลล่า
เบลล่าราณี **บุพเพ**
ดีสิบเดย์ popezaap
ดุสดพิณสุด **สันนิวาส**
เบลล่า popezapp จ้อย ออเจ้าทั้งพระนคร
รีวิวเซเว่น

Twitter API Registration

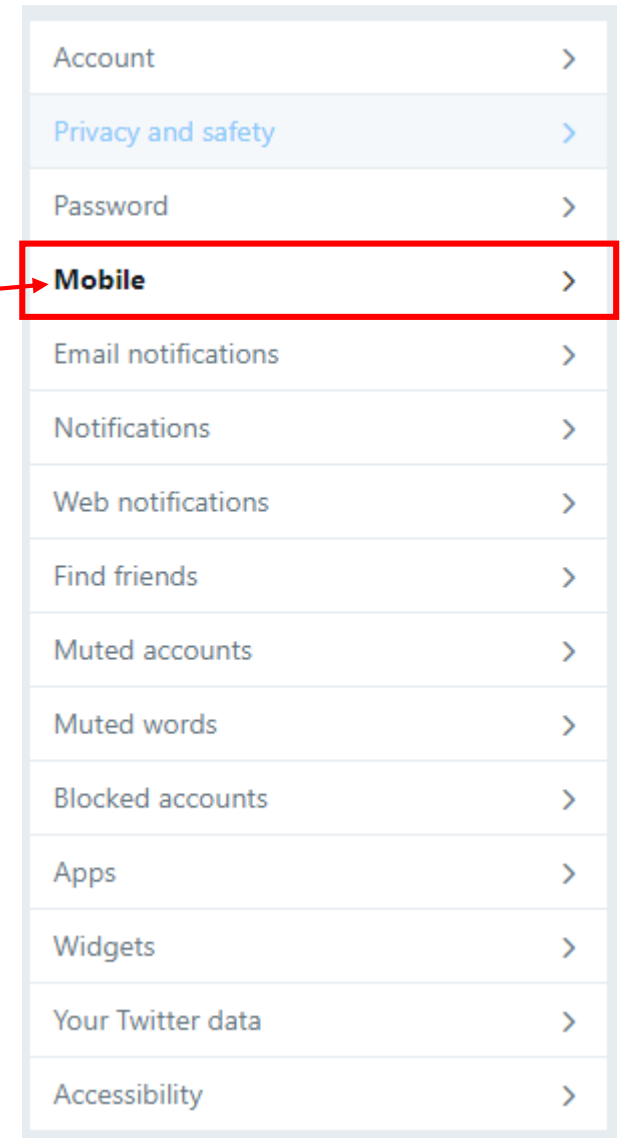
การลงทะเบียนผู้พัฒนา

- ไปที่ <https://apps.twitter.com/>
- Login ด้วย twitter account ถ้าไม่มีให้ สร้าง account ใหม่



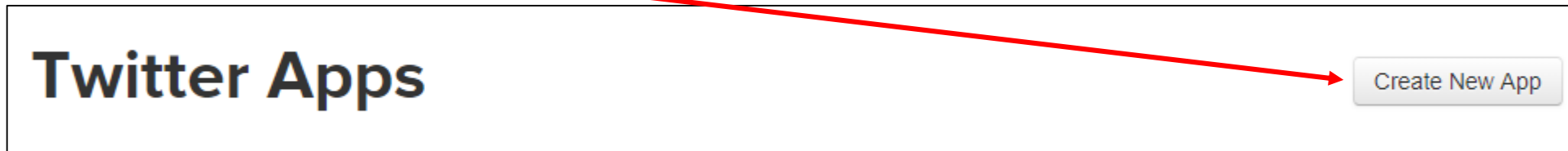
การลงทะเบียนผู้พัฒนา

- ไปที่ <https://twitter.com/settings/account>
- ไปที่ Mobile เพื่อเพิ่มเบอร์โทรศัพท์ในการยืนยันตัวตน



การสร้าง Twitter API KEY

- ไปที่ <https://apps.twitter.com/> หลังจากที่ได้ login และ ลงทะเบียนเบอร์มือถือเรียบร้อยแล้ว
- กด Create New App เพื่อสร้าง API KEY



การสร้าง Twitter API KEY

Create an application

กรอกข้อมูลให้ครบ

The screenshot shows the 'Create an application' form on the Twitter developer portal. It is divided into two main sections: 'Application Details' and 'Developer Agreement'. The 'Application Details' section contains four input fields: 'Name', 'Description', 'Website', and 'Callback URL'. Red arrows point from the Thai text 'กรอกข้อมูลให้ครบ' to each of these four fields. The 'Name' field contains 'MyTraningCrawler', the 'Description' field contains 'Create application for workshop', and the 'Website' field contains 'http://pop.ssense.in.th'. The 'Callback URL' field is empty. Below the 'Application Details' section is the 'Developer Agreement' section, which has a checked checkbox and the text 'Yes, I have read and agree to the [Twitter Developer Agreement](#)'. At the bottom of the form is a button labeled 'Create your Twitter application', which is highlighted with a red rectangle. A red arrow points from the Thai text 'Click เพื่อสร้าง App' to this button.

Application Details

Name *

MyTraningCrawler

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Create application for workshop

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

http://pop.ssense.in.th

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the here. To restrict your application from using callbacks, leave this field blank.

Developer Agreement

☒ Yes, I have read and agree to the [Twitter Developer Agreement](#).

Create your Twitter application


Click เพื่อสร้าง App

การสร้าง Twitter API KEY

Click เพื่อสร้าง API KEY

mytweetpy995

Details Settings **Keys and Access Tokens** Permissions

 Test for tweetpy
<http://www.sansarn.com>

Organization

Information about the organization or company associated with your application. This information is optional.

Organization	None
Organization website	None

Application Settings

Your application's Consumer Key and Secret are used to [authenticate](#) requests to the Twitter Platform.

Access level	Read and write (modify app permissions)
Consumer Key (API Key)	cxYrPjo1JvLYo9yAhoqdtiEPD (manage keys and access tokens)
Callback URL	None
Callback URL Locked	No
Sign in with Twitter	Yes
App-only authentication	https://api.twitter.com/oauth2/token
Request token URL	https://api.twitter.com/oauth/request_token
Authorize URL	https://api.twitter.com/oauth/authorize
Access token URL	https://api.twitter.com/oauth/access_token

การสร้าง Twitter API KEY

MyTraningCrawler

[Test OAuth](#)[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key) NzDRw1V12GnxXrMrk5sOPgZ6s

Consumer Secret (API Secret) 9aMDJXSFBu1P3k4Llq1uYMUJcfymQbmSDELIWmNQfRulsAe1HI

Access Level Read and write ([modify app permissions](#))

Owner 9meo

Owner ID 92082793

KEY ที่ใช้สำหรับ Twitter API

Application Actions

[Regenerate Consumer Key and Secret](#)[Change App Permissions](#)

Your Access Token

You haven't authorized this application for your own account yet.

By creating your access token here, you will have everything you need to make API calls right away. The access token generated will be assigned your application's current permission level.

Token Actions

[Create my access token](#)

Click เพื่อสร้าง Access token

การสร้าง Twitter API KEY

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token

92082793-
o2WvZU1FK4LqQNYqgmh1ouW9yXCgPLmtY0bBM2y6S

Access Token Secret

2UmBa0omH2dYc5crqXg175SL5n3WLeC0bDJtFnDf3HiX

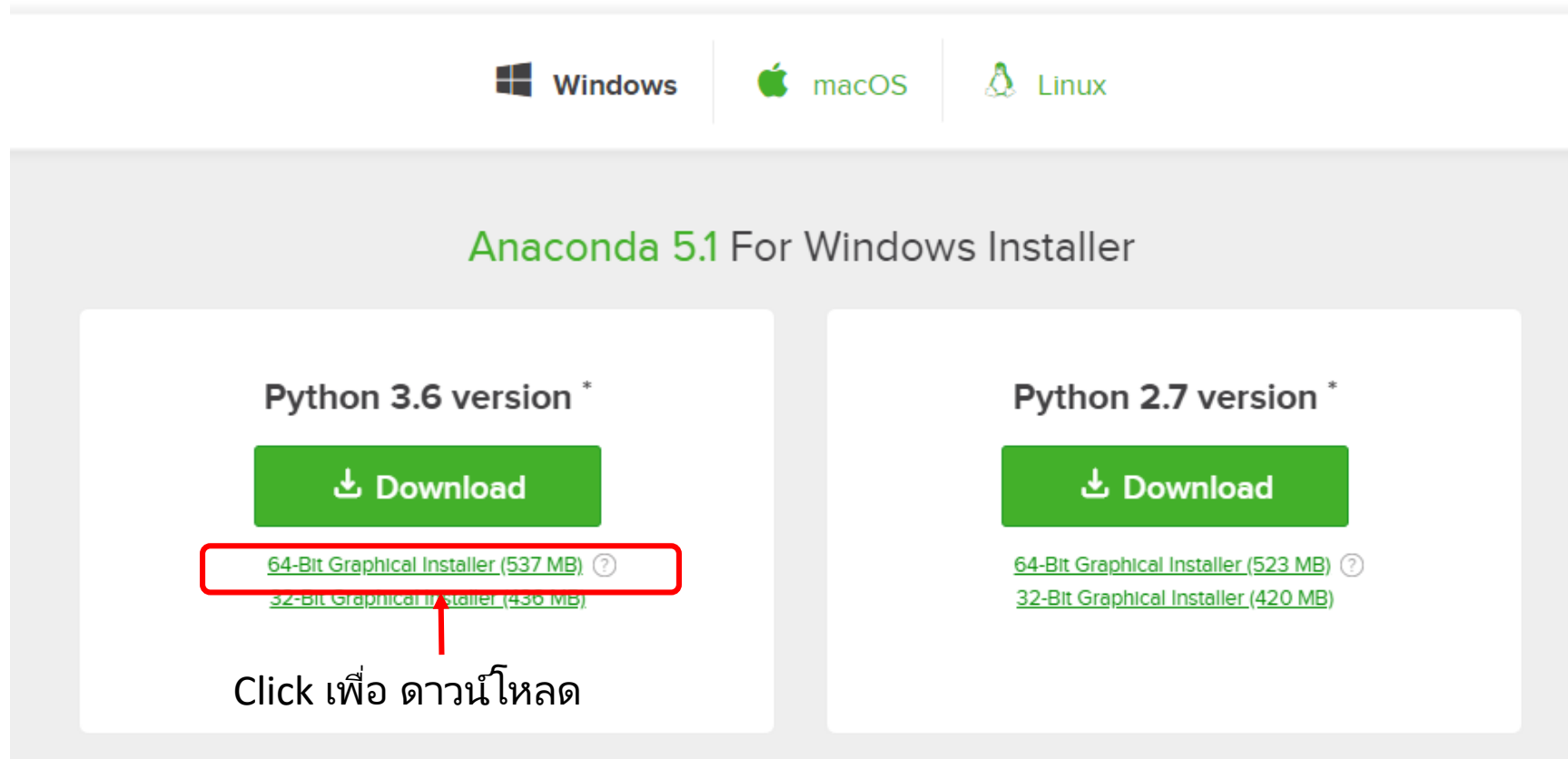
KEY ที่ใช้สำหรับ Twitter API



Twitter API by
Tweepy library

Install Anaconda

- ไปที่ <https://www.anaconda.com/download/> เพื่อ download anaconda

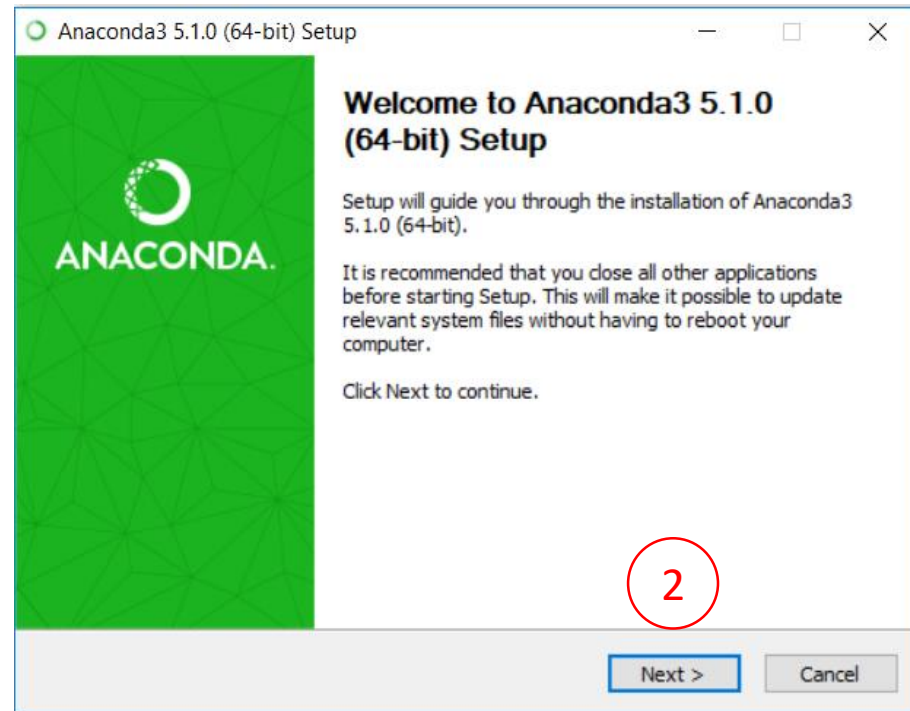


Install Anaconda

- คลิกเพื่อติดตั้ง

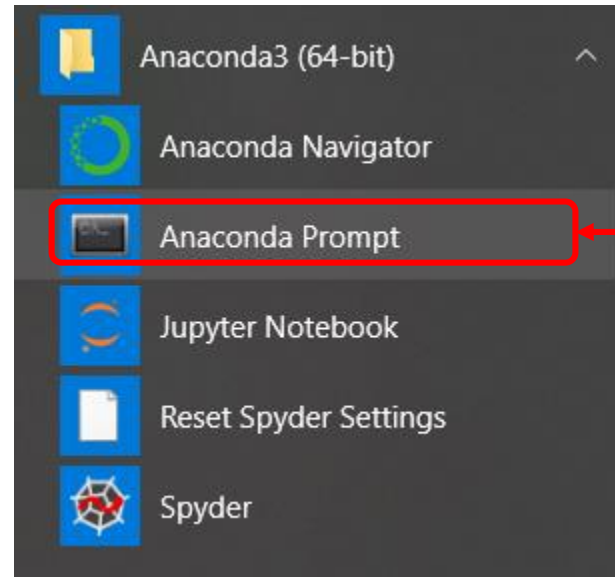
1

Anaconda3-5.1.0-Windows-x86_64



ติดตั้ง Library Tweepy

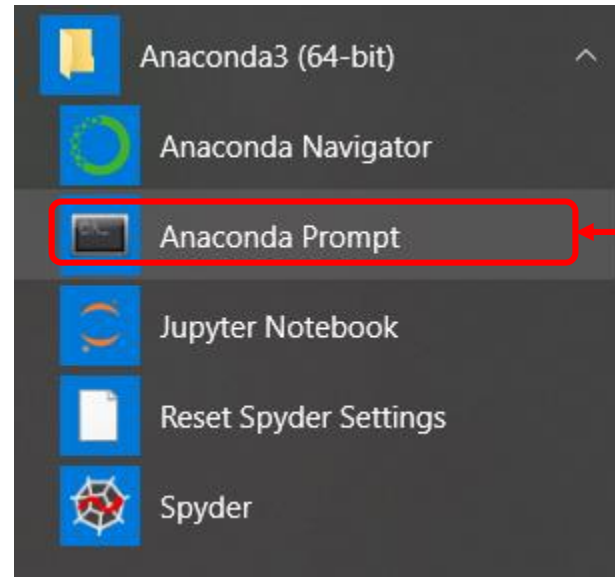
- ไปที่ Start menu->



Click เลือก เพื่อพิมพ์คำสั่ง

ติดตั้ง Library Tweepy

- ไปที่ Start menu->



Click เลือก เพื่อพิมพ์คำสั่ง

```
(base) C:\Users\Nectec>conda install -c conda-forge tweepy  
Solving environment: /
```

พิมพ์คำสั่ง

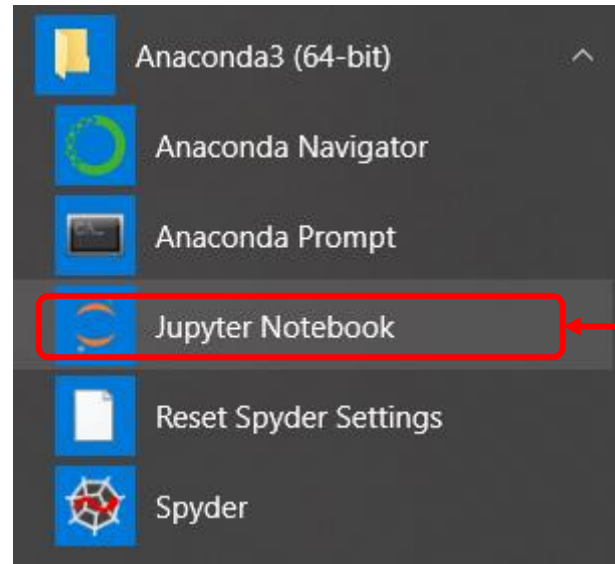
conda install -c conda-forge tweepy
รอจนติดตั้งเสร็จ

Jupyter Notebook

Python Editor


Jupyter notebook

- Jupyter notebook เป็นโปรแกรมที่ใช้สำหรับสร้างโปรแกรมด้วยภาษา Python เนื่องจากสะดวกและใช้งานง่ายสามารถเรียกใช้งานข้ามเครื่องได้ จึงทำให้ Jupyter ได้รับความนิยมเป็นอย่างมากในการใช้พัฒนาโปรแกรม
- เริ่มต้นใช้งาน Jupyter notebook




Click เลือก เพื่อเรียกใช้งาน

Jupyter notebook

 jupyter Logout

Files Running Clusters

Select items to perform actions on them. Upload New 

Click เพื่อเข้าทำงานใน Documents

<input type="checkbox"/> 0	/	Name	Last Modified
<input type="checkbox"/>	3D Objects		1 เดือนที่แล้ว
<input type="checkbox"/>	Anaconda3		7 ชั่วโมงที่แล้ว
<input type="checkbox"/>	Contacts		1 เดือนที่แล้ว
<input type="checkbox"/>	Desktop		15 วันที่แล้ว
<input type="checkbox"/>	Documents		3 นาทีที่แล้ว
<input type="checkbox"/>	Downloads		12 นาทีที่แล้ว
<input type="checkbox"/>	eclipse-workspace		16 วันที่แล้ว
<input type="checkbox"/>	Favorites		1 เดือนที่แล้ว
<input type="checkbox"/>	Links		1 เดือนที่แล้ว
<input type="checkbox"/>	Music		1 เดือนที่แล้ว
<input type="checkbox"/>	OneDrive		1 เดือนที่แล้ว
<input type="checkbox"/>	Pictures		1 เดือนที่แล้ว
<input type="checkbox"/>	Saved Games		1 เดือนที่แล้ว
<input type="checkbox"/>	Searches		1 เดือนที่แล้ว
<input type="checkbox"/>	Videos		16 วันที่แล้ว

Jupyter notebook

Files Running Clusters

Select items to perform actions on them.

0 / Documents

- ..
- Custom Office Templates
- My Tableau Repository
- Python Scripts

Click เพื่อสร้าง Folder

1

Upload New

Notebook:
Python 3

Other:
Text File
Folder
Terminal

Files Running Clusters

Rename Move

1 / Documents

- ..
- Custom Office Templates
- My Tableau Repository
- Python Scripts
- Untitled Folder

Click เพื่อเปลี่ยนชื่อ Folder

2

3

4

Rename directory

Enter a new directory name:

MyCrawler

5

Cancel Rename

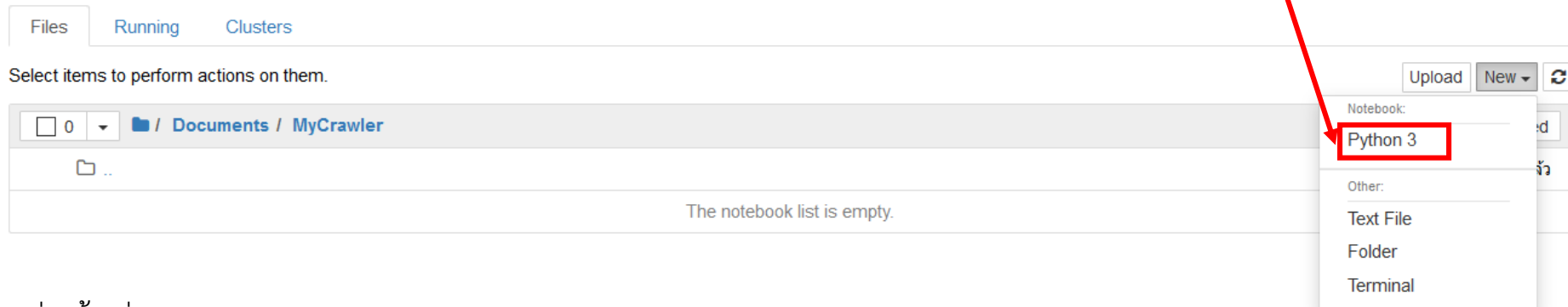
ไม่กี่วินาทีที่แล้ว

Jupyter notebook

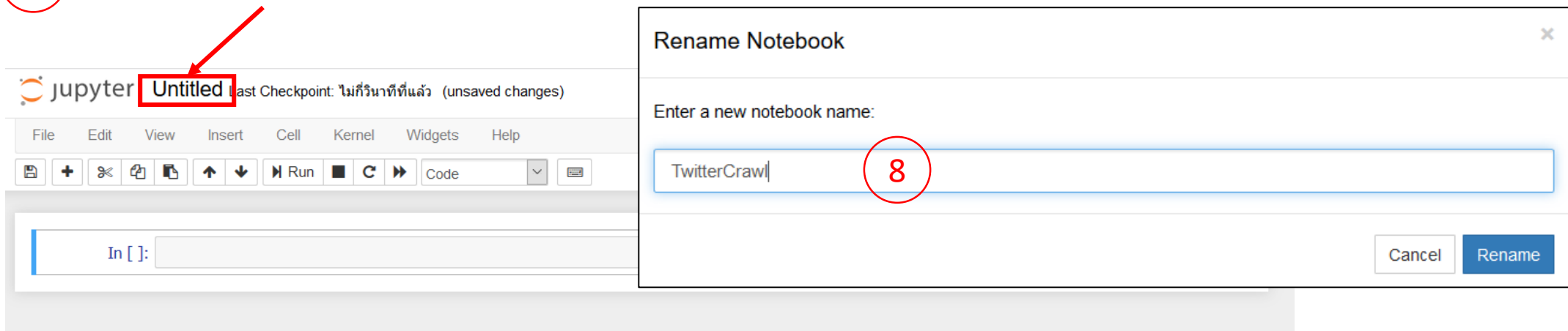
- Click ไปใน Folder MyCrawler

Click เพื่อสร้าง Python Application

6



7 Click เพื่อตั้งชื่อ Application



Twitter API

สร้างโปรแกรมเพื่อเรียกใช้งาน Twitter API เพื่อใช้ในการเก็บข้อมูล กด Shift+Enter เพื่อ Run คำสั่งในแต่ละส่วน

In [1]: `import tweepy
import re`

#เรียก Library ที่จำเป็นสำหรับการใช้งาน

In [2]: `consumer_key = "
consumer_secret = "
access_token = "
access_token_secret = "`

#ใส่ API KEY ที่ได้สร้างไว้

In [3]: `# OAuth process, using the keys and tokens
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

Creation of the actual interface, using authentication
api = tweepy.API(auth)`

#Code เพื่อ login ใช้งาน API

Twitter API

```
In [4]: data=[]
query = "บุพเพสันนิวาส" # Keyword ที่ต้องการ เก็บ
maxtweet=100 # จำนวน Tweet ที่ต้องการเก็บ
for tweet in tweepy.Cursor(api.search,
                             q=query,
                             count=100,
                             result_type="recent",
                             include_entities=True,
                             lang="th").items():
    text = tweet.text.replace("\n","")
    if text not in data:
        data.append(text) #เก็บข้อมูลไว้ในตัวแปร data
    if len(data) == maxtweet:
        break
```

#ส่วนของโปรแกรมสำหรับเก็บข้อมูลตาม keyword ที่ต้องการ

In [5]: data

#เรียกดูข้อมูลที่เก็บได้

[illegible]

Thai Word Segmentation

- WordCut

- Copy file “bigthai.txt” และ “wordcut.py” ไปยัง Documents-> MyCrawler



#ส่วนของโปรแกรมเพื่อเตรียมพร้อมสำหรับ WordCut

```
In [7]: from wordcut import Wordcut
custom_word = ['ทองหยิบ','ทองหยอด','ฝอยทอง','ออเจ้า','นากิน','ยี่งัก','อยุธยา','หลวงศรียศ','หมูโสร่ง','กระโปรง'] #เพิ่มคำศัพท์เพื่อให้ตัดคำได้ถูกต้อง
if __name__ == '__main__':
    with open('bigthai.txt', encoding="UTF-8") as dict_file:
        word_list = list(set([w.rstrip() for w in dict_file.readlines()])) #Load คำศัพท์
        word_list.extend(custom_word)
        wordcut = Wordcut(word_list)
        print(wordcut.tokenize("หมูโสร่ง พอทำเอง ทำไมกลายเป็นหมูกระโปรงล่ะ")) #ทดสอบตัดคำ

['หมูโสร่ง', '', 'พอ', 'ทำ', 'เอง', '', 'ทำไม', 'กลายเป็น', 'หมู', 'กระโปรง', 'ล่ะ']
```

Thai Word Segmentation

- LexToPlus

#ส่วนของโปรแกรมเพื่อเตรียมพร้อมสำหรับ LexToPlus

```
In [9]: import requests
def lextoseg(text):

    url = 'http://sansarn.com/api/lextoplus.php?key=482768c6f8b1efbfb8aac542a4533bc4&text='+text # url สำหรับ เรียกใช้งาน LexToPlus API
    response = requests.post(url)
    return response.json()['tokens']
print(lextoseg("หมูโสร่ง พอทำเอง ทำไมกลายเป็นหมูกระโปรงล่ะ")) #ทดสอบตัดคำ

['หมู', 'โสร่ง', '', 'พอ', 'ทำ', 'เอง', '', 'ทำไม', 'กลายเป็น', 'หมู', 'กระโปรง', 'ล่ะ']
```

Remove Emoji 😊😊

```
In [10]: def removeEmoji(text):
emoji_pattern = re.compile("[
    u\"\\U0001F600-\\U0001F64F\" # emoticons
    u\"\\U0001F300-\\U0001F5FF\" # symbols & pictographs
    u\"\\U0001F680-\\U0001F6FF\" # transport & map symbols
    u\"\\U0001F1E0-\\U0001F1FF\" # flags (iOS)
    u\"\\U00002702-\\U000027B0\"
    u\"\\U000024C2-\\U0001F251\"
    \"]+", flags=re.UNICODE)
return emoji_pattern.sub(r'',text)
```

Data Cleansing

#ทำความสะอาดข้อมูลและตัดคำ

```
In [*]: import re
res = []
for tw in data:
    tw=tw.replace("","").replace("","").replace("&quot;","").replace('...','').replace('?', '')
    tw = re.sub('[!":~.-]', '', tw) #Remove Escape Character
    tw = removeEmoji(tw) #Remove Emoji
    seg2 = []
    for s in tw.split():
        if s.startswith('@') or s.startswith('#') or s.startswith('RT@') or s=='RT' :
            seg2.append(s.replace('RT','')) #Remove RT
        elif s.lower().startswith('http'): # Remove URL
            continue
        else:
            seg2.extend(lextoseg(s)) ← #หากต้องการใช้ WordCut ให้เปลี่ยนเป็น seg2.extend(wordcut.tokenize(s))
    res.extend(seg2)
```

#ทำความสะอาดข้อมูลและรวมคำพร้อมทั้ง remove คำทั่วไปออก

```
In [12]: stoplist = ['ที่', 'นี้', 'ไม่', 'ให้', 'ได้'] ← #รายการคำทั่วไป
res = ''.join(res).replace('อ๋อ เจ้า', 'อ๋อเจ้า').split()
res = ''.join(res).replace('ขุน ศรี ยศ', 'ขุนศรียศ').split()
res = ''.join(res).replace('หลวง ศรี ยศ', 'หลวงศรียศ').split()
res = ''.join(res).replace('แม่กา ระ เกด', 'แม่การะเกด').split()
res = [x for x in res if x not in stoplist]
```

Write Data to text file

In [13]: `import codecs`

```
file = codecs.open("tweetstext.txt", "w", "utf-8")
file.write("keyword\n")
for r in res:
    if len(r.strip()) > 2:
        file.write(r+"\n")
file.flush()
file.close()
```

#ชื่อไฟล์ที่ต้องการ save

รายการของคำใน Text File เพื่อนำไปสร้าง Tag Cloud

keyword
@liurangsiyanon
ทุกสิ่ง
แม้การเกิด
อยู่
ตัว
หมดแล้ว
หมอน
ถือ
เสื้อ
ก๊าก
ใส่
เสื้อ
แขน
ยาว
ใส่
เหยย
ยยยย
เกิน
คุณ
พี่
@HOLYJJOIPPOI
ความ
เพื่อนสนิท
อะ
แบบ
ฮั่น
แห่นะ
รู้
มึง
คิด
อะไร
อยู่
#บุพเพสันนิวาส
@Moremimpawee



Analytic and Visualization Tool

Tableau

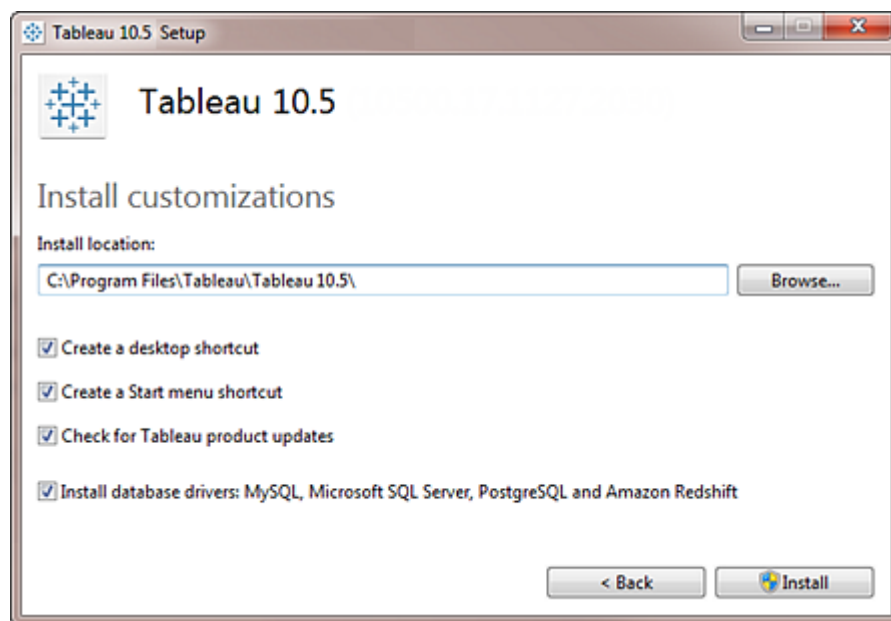
- **Tableau** เป็นโปรแกรมสำหรับการวิเคราะห์ข้อมูล และหาคำตอบในเชิงธุรกิจอย่างรวดเร็ว โดยการใชื Tableau จะเป็นการช่วยให้ผู้ใช้งานมีความเข้าใจในข้อมูลมากขึ้น รวมถึงผู้ใช้งาน สามารถนำข้อมูลที่มีอยู่มาวิเคราะห์และแสดงผลอย่างมีประสิทธิภาพ
- ดาวน์โหลดและติดตั้ง Tableau ได้ที่
<https://www.tableau.com/support/releases/desktop/10.5.2>

Tableau Desktop 10.5.2

DOWNLOAD DESKTOP 10.5.2

Tableau

- ติดตั้ง Tableau



Activate Tableau Desktop

- หลังจากที่ได้ติดตั้งโปรแกรม Tableau Desktop แล้ว ให้เปิดโปรแกรมขึ้นมาจากนั้นกรอกข้อมูลเพื่อลงทะเบียนใช้งานโปรแกรม



Tableau 10.5
Desktop app

Tableau Registration

×

Almost there
Have a product key? [Activate Tableau](#)

First name

Last name

Business email

jjones@mycompany.com

Organization

Country/Region

United States

State

--

ZIP code

e.g. 98103

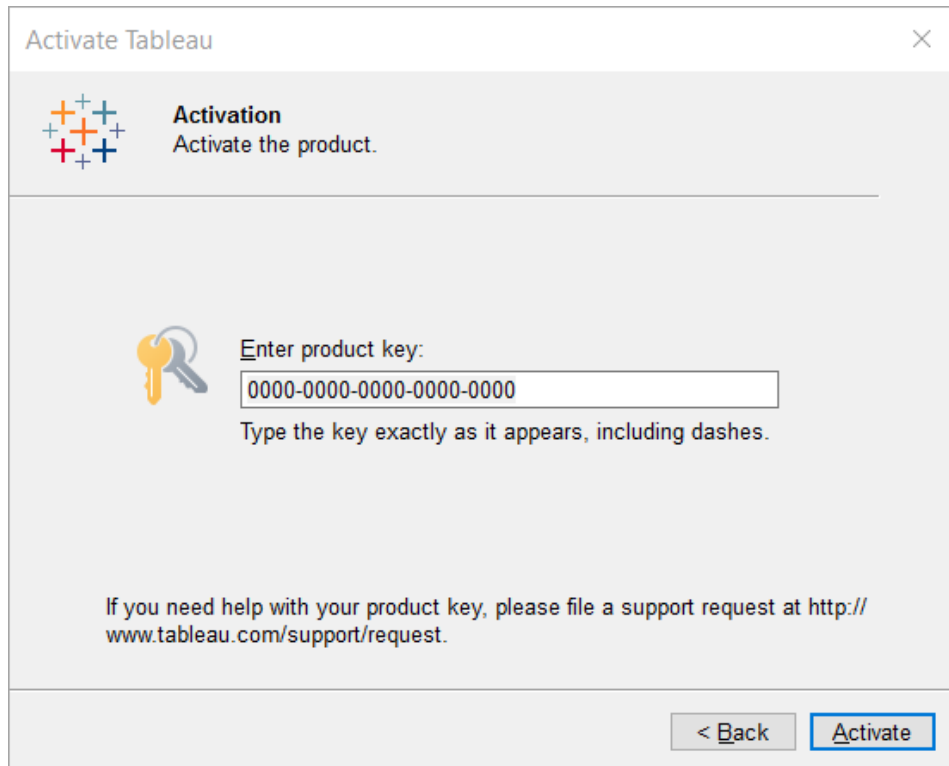
Phone (e.g. (201) 555-5555)

Start trial now

[We respect your privacy](#) | [Having trouble?](#)

Activate Tableau Desktop

- หลังจากลงทะเบียนแล้ว ระบบจะส่ง product key ไปยัง email ที่ได้ลงทะเบียนไว้
- Click Activate เพื่อใส่ค่า product key



Activate Tableau

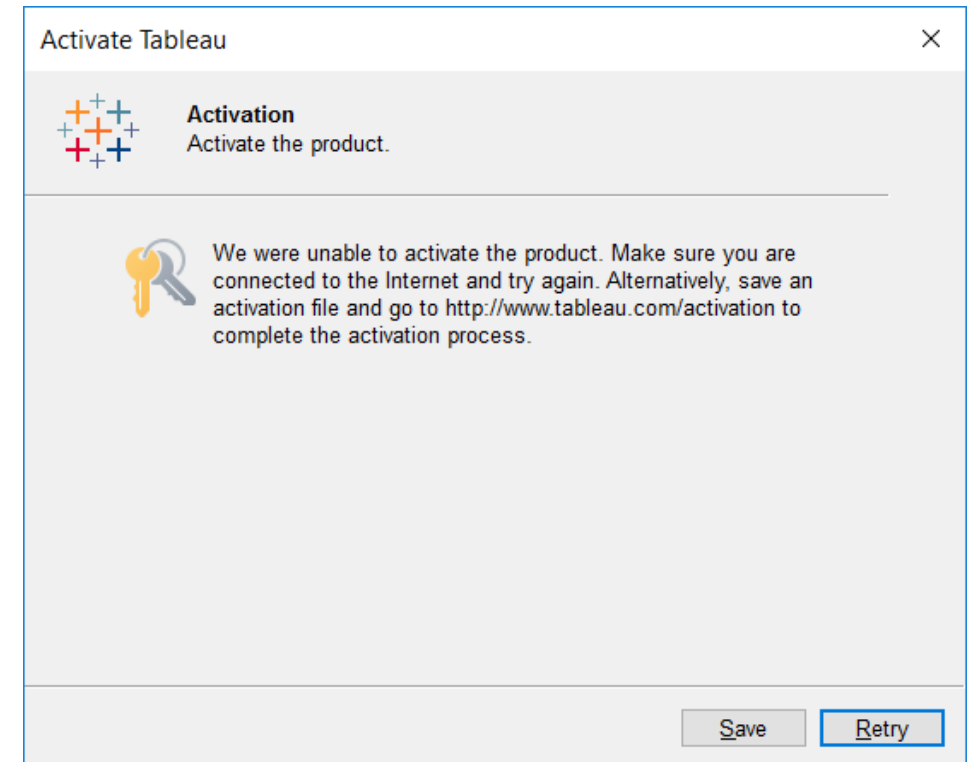
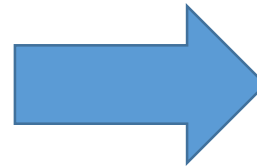
Activation
Activate the product.

Enter product key:
0000-0000-0000-0000-0000
Type the key exactly as it appears, including dashes.

If you need help with your product key, please file a support request at <http://www.tableau.com/support/request>.

< Back Activate

Click Activate

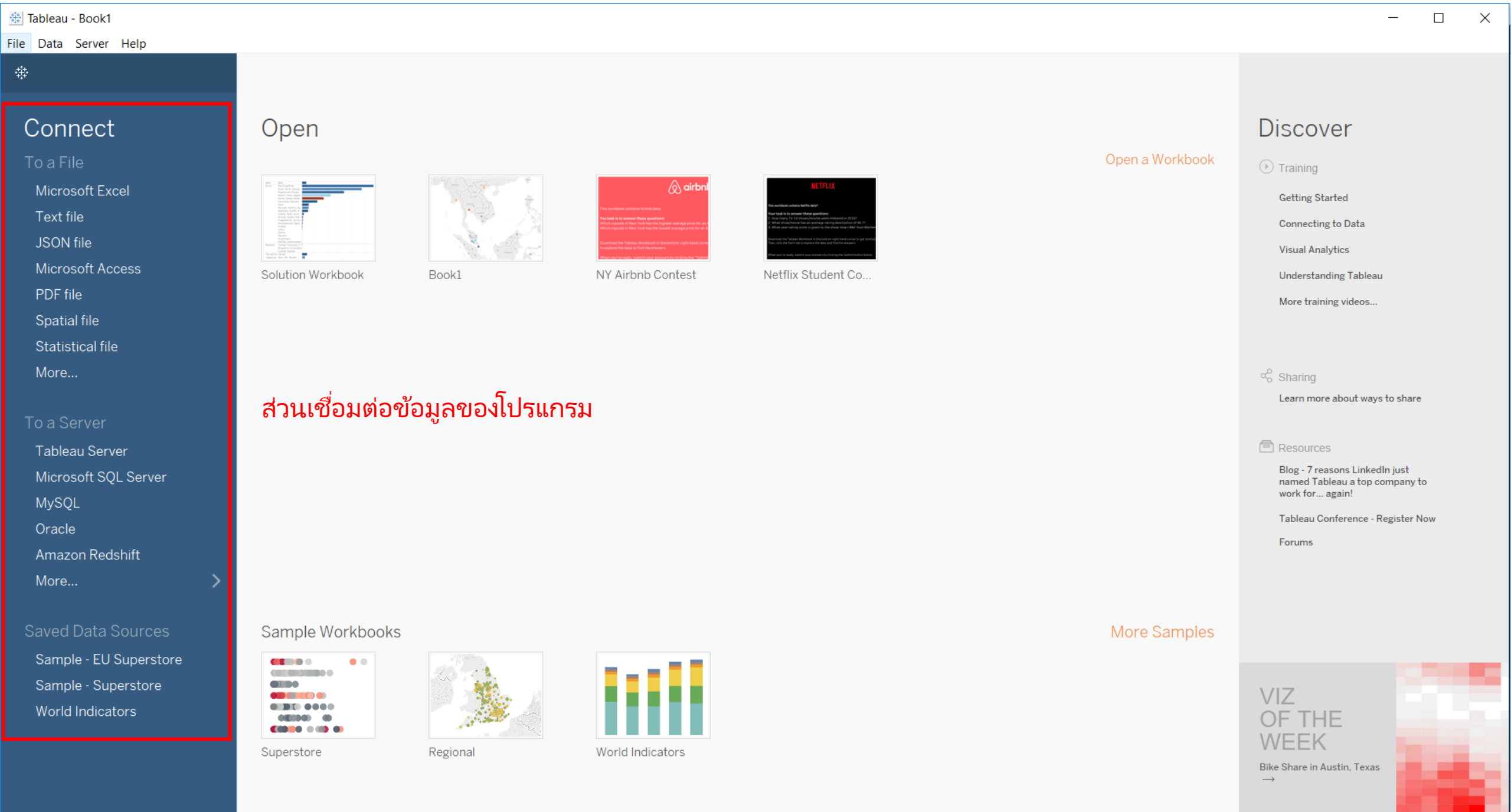


Activate Tableau

Activation
Activate the product.

We were unable to activate the product. Make sure you are connected to the Internet and try again. Alternatively, save an activation file and go to <http://www.tableau.com/activation> to complete the activation process.

Save Retry



Connect

To a File

- Microsoft Excel
- Text file
- JSON file
- Microsoft Access
- PDF file
- Spatial file
- Statistical file
- More...

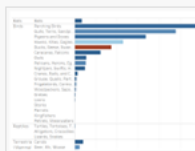
To a Server

- Tableau Server
- Microsoft SQL Server
- MySQL
- Oracle
- Amazon Redshift
- More...

Saved Data Sources

- Sample - EU Superstore
- Sample - Superstore
- World Indicators

Open



Solution Workbook



Book1



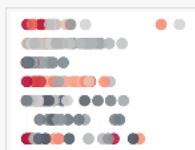
NY Airbnb Contest



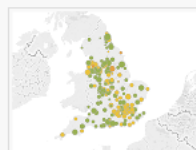
Netflix Student Co...

ส่วนเชื่อมต่อข้อมูลของโปรแกรม

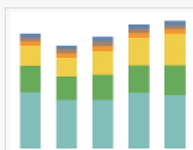
Sample Workbooks



Superstore



Regional



World Indicators

Open a Workbook

Discover

Training

- Getting Started
- Connecting to Data
- Visual Analytics
- Understanding Tableau
- More training videos...

Sharing

Learn more about ways to share

Resources

- Blog - 7 reasons LinkedIn just named Tableau a top company to work for... again!
- Tableau Conference - Register Now
- Forums

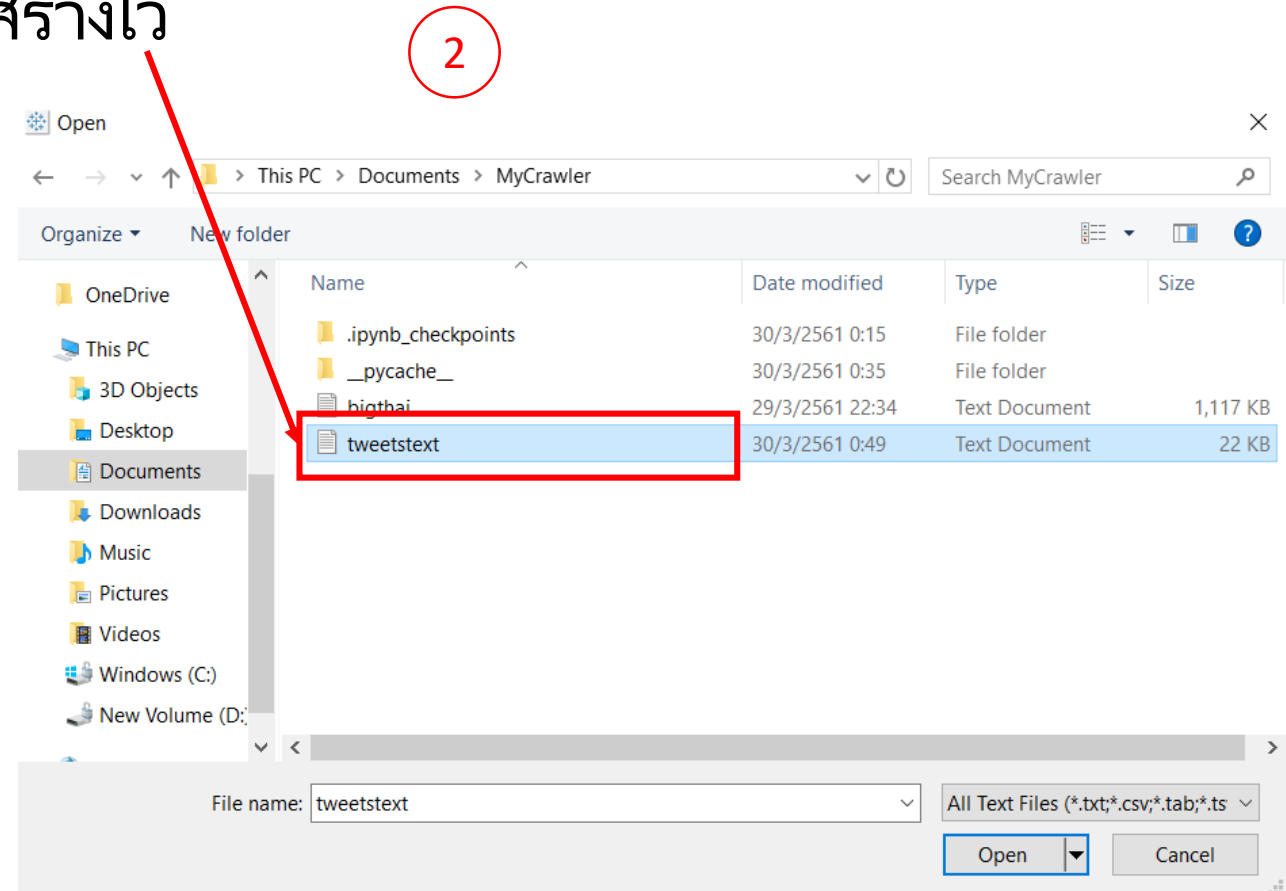
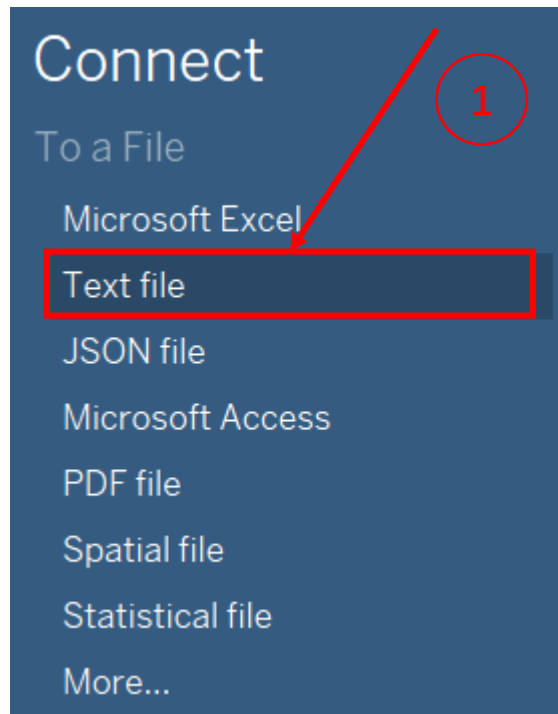
More Samples

VIZ
OF THE
WEEK

Bike Share in Austin, Texas
→

สร้าง Word Cloud ด้วย Tableau

- เลือกเพื่อเชื่อมต่อกับไฟล์ที่ได้สร้างไว้



สร้าง Word Cloud ด้วย Tableau

The screenshot shows the Tableau interface with the following elements:

- Connections:** A list containing 'tweetstext' (Text File).
- Files:** A list containing 'bigthai.txt' and 'tweetstext.txt'. There is a checkbox for 'Use Data Interpreter' which is unchecked.
- Worksheet:** The main area shows a table with the following data:

keyword
@liurangsiyanon
ทุกสิ่ง
แม่เกาะเกิด
อยู่
ตัว
หมดแล้ว
หมอน
ถือ
เสื้อ

Red annotations highlight key steps:

- A red arrow points from the text 'ข้อมูลที่เตรียมไว้' (Data prepared) to the 'keyword' column header.
- A red circle with the number '3' and the text 'คลิกเพื่อสร้าง Tag Cloud' (Click to create Tag Cloud) points to the 'Tag Cloud' button in the bottom toolbar.

สร้าง Word Cloud ด้วย Tableau

คลิกและลาก
Keyword ไปยัง
“Text” เพื่อแสดง
คำที่เตรียมไว้

1

The screenshot shows the Tableau interface with the following components:

- Data Source:** tweetstext
- Dimensions:** keyword (highlighted with a red box and arrow labeled '1')
- Measures:** Number of Records, Measure Values
- Marks:** Text (selected)
- Columns:** keyword (selected)
- Rows:** keyword (selected)

The main view displays a word cloud of the text data, with words of varying sizes and colors representing their frequency. The words are arranged in a grid-like pattern, with the most frequent words being the largest and most prominent.

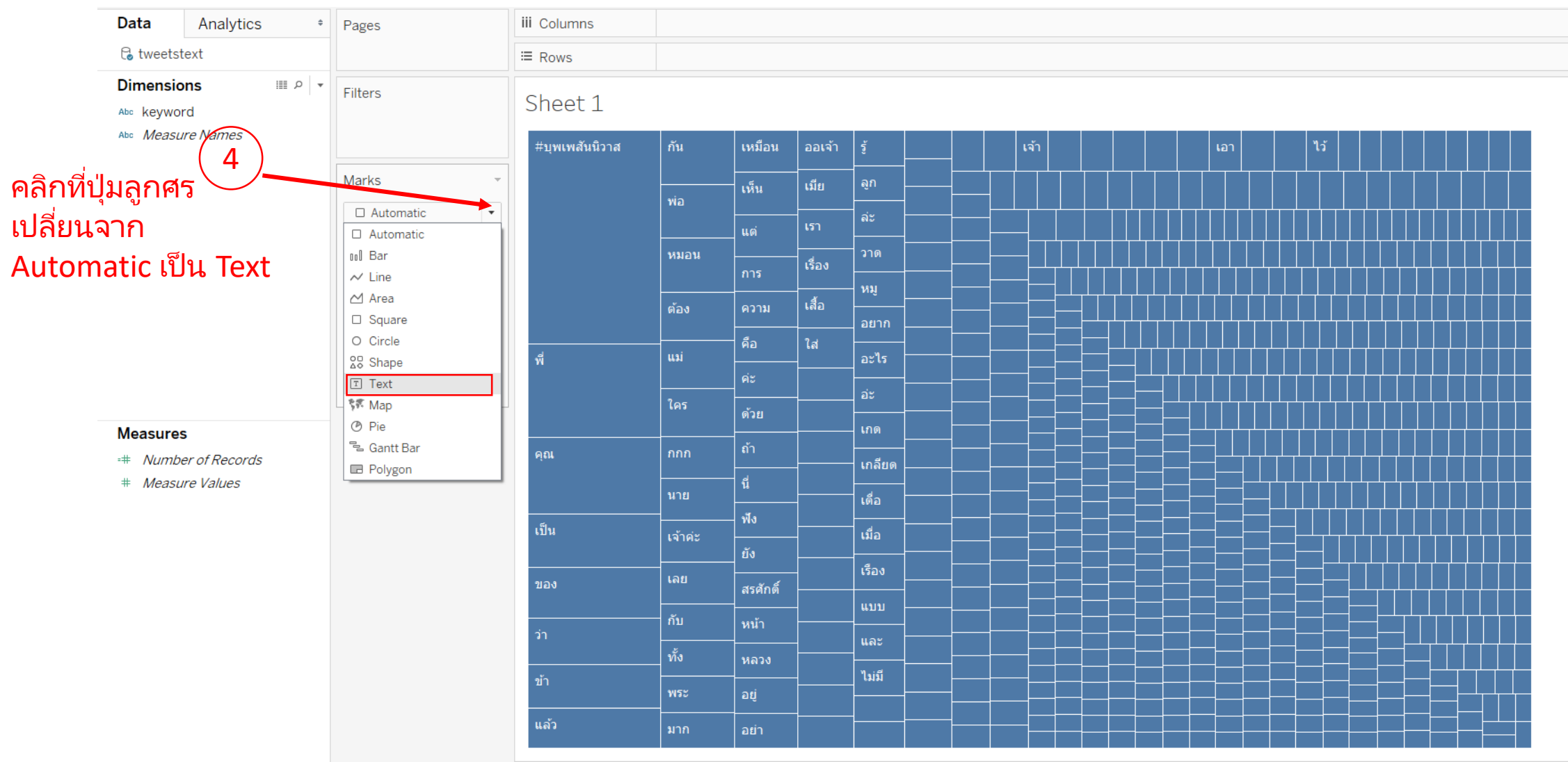
สร้าง Word Cloud ด้วย Tableau

[illegible]

สร้าง Word Cloud ด้วย Tableau

[illegible]

สร้าง Word Cloud ด้วย Tableau



สร้าง Word Cloud ด้วย Tableau

tweetstext

Dimensions

keyword

Measure Names

Measures

Number of Records

Measure Values

Columns

Rows

Filters

Marks

Text

Color

Size

Text

Detail

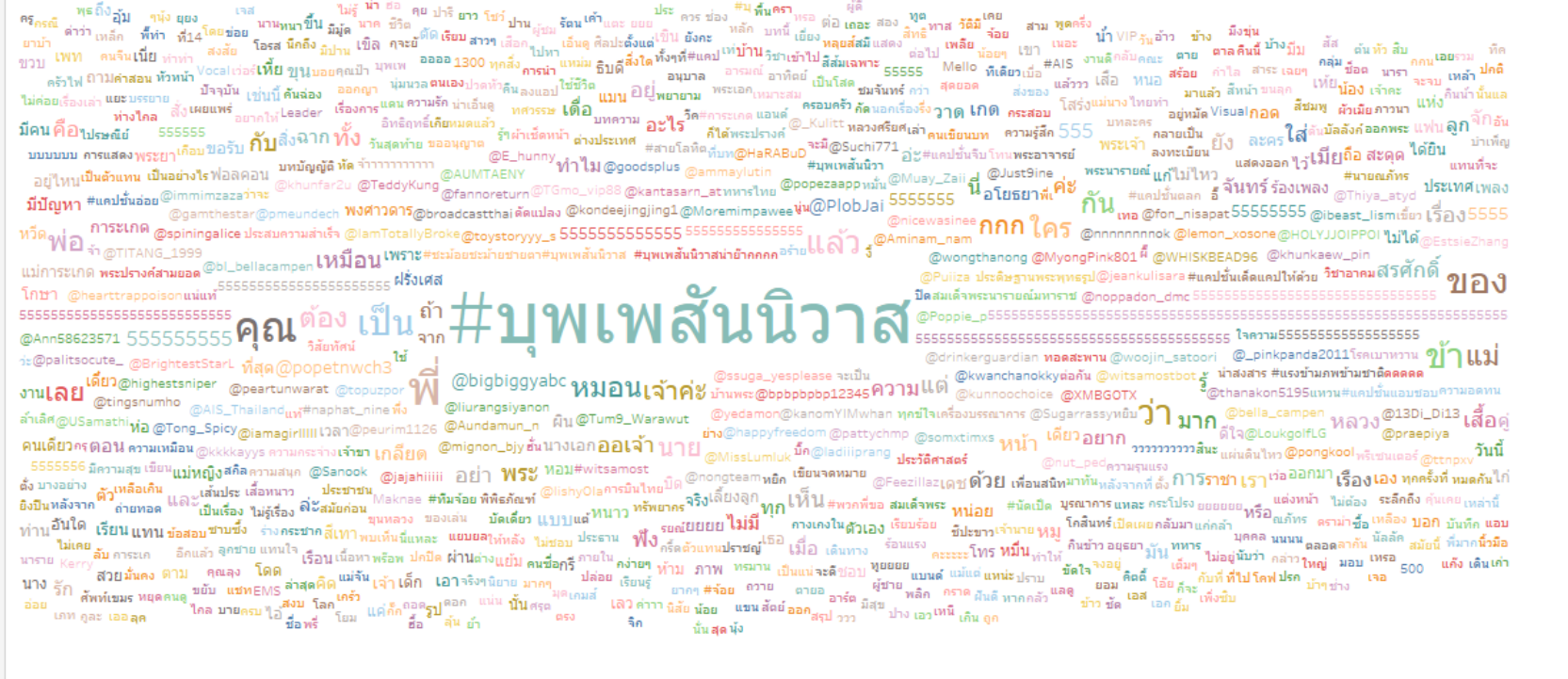
Tooltip

keyword

CNT(keyword)

keyword

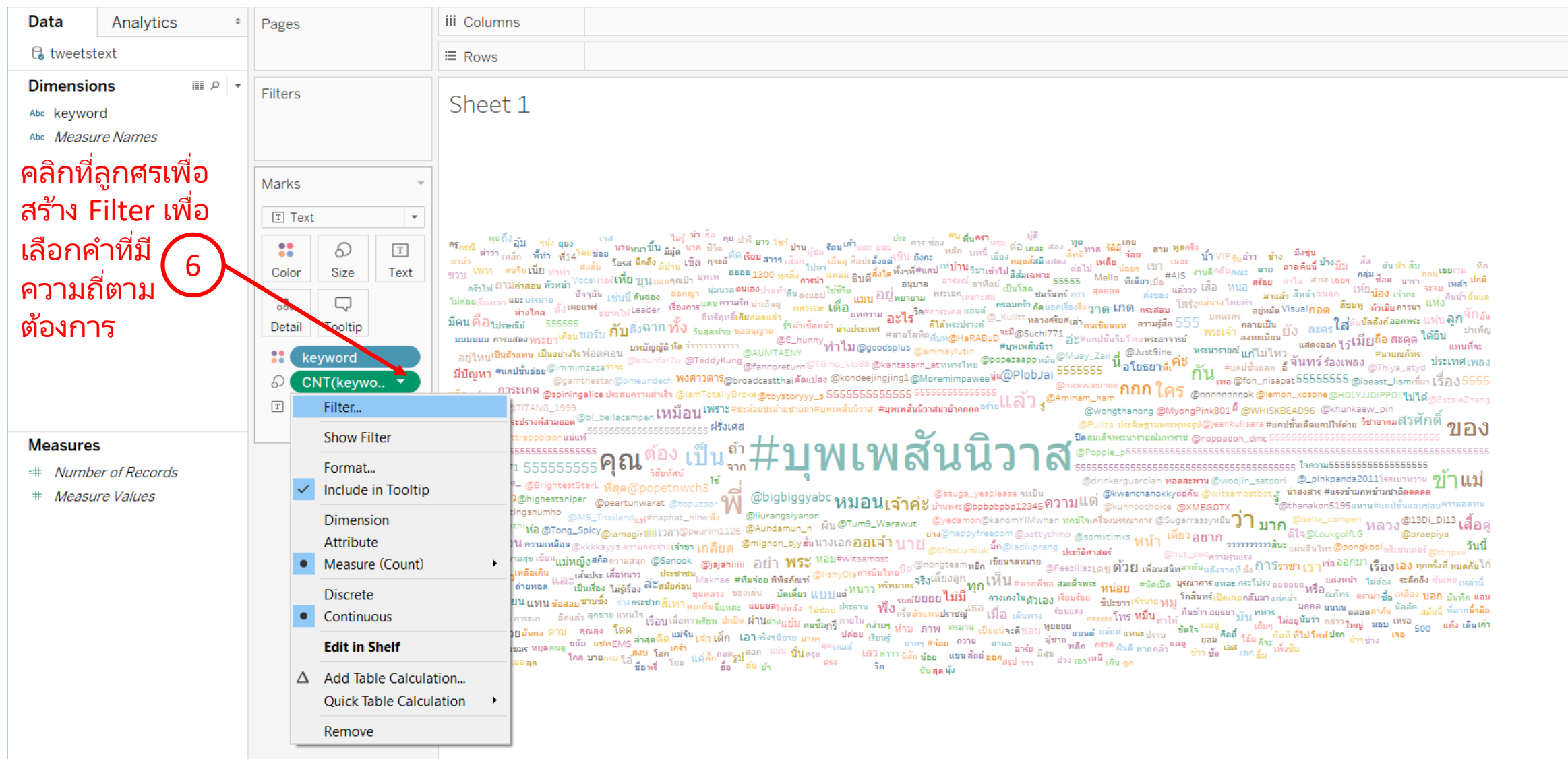
Sheet 1



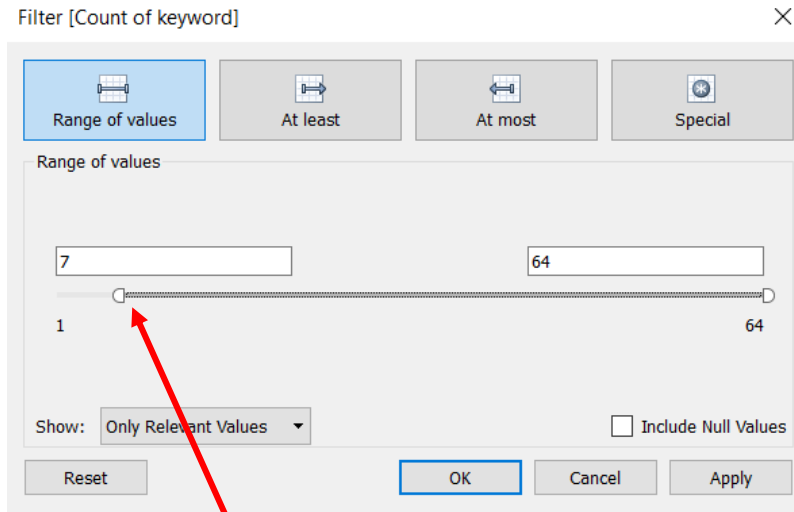
คลิกและลาก Keyword อีกครั้งไปยัง “Color” กำหนดสีให้กับคำ

5

สร้าง Word Cloud ด้วย Tableau



สร้าง Word Cloud ด้วย Tableau



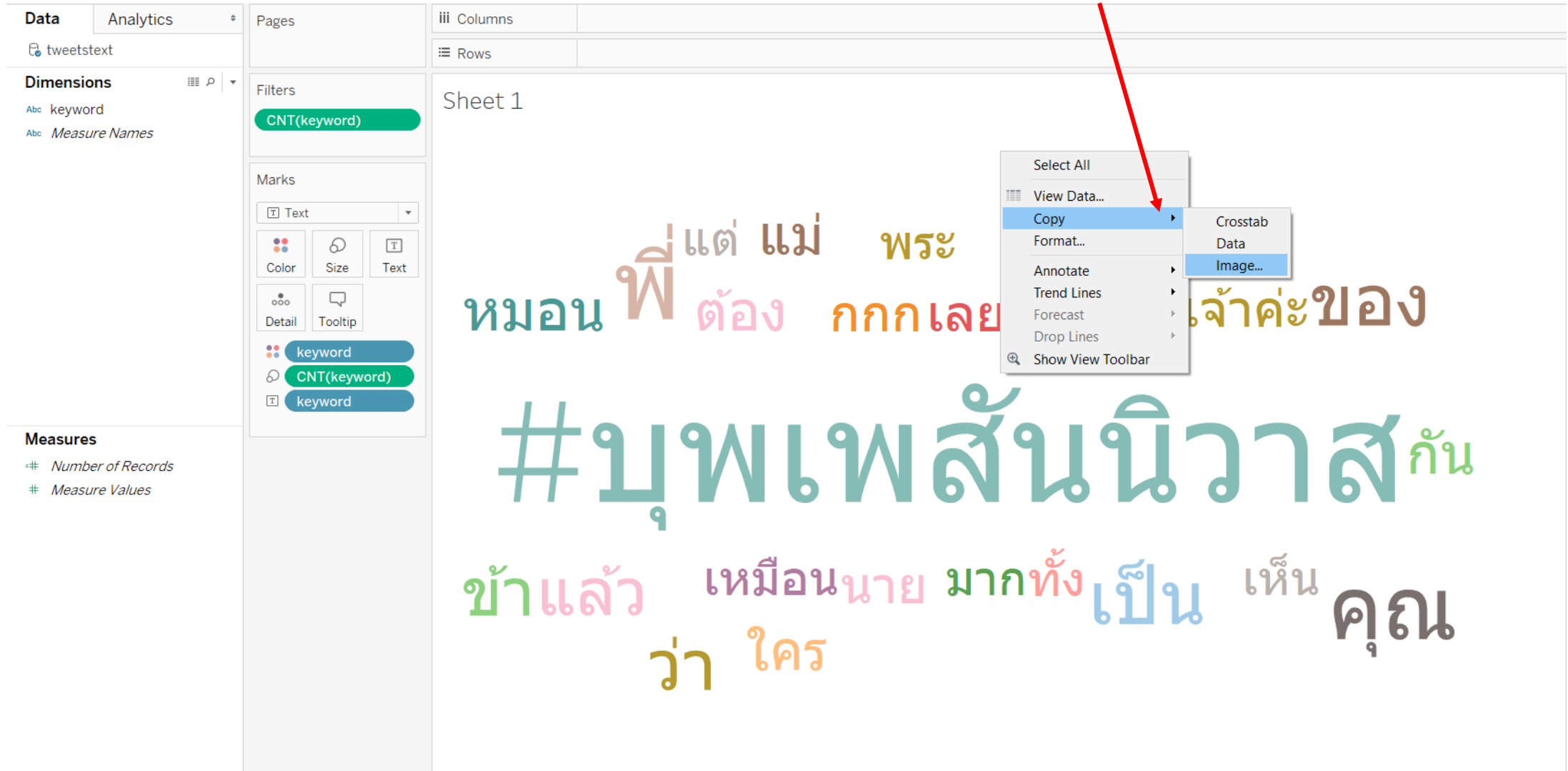
ปรับขนาดตาม
ต้องการ

Tag Cloud ที่ได้จากโปรแกรม Tableau

แต่ แม่ พระ
หมอน พี่ ต้อง กกกเลย กับพ่อเจ้าค่ะของ
#บุพเพสันนิวาสกัน
ข้าแล้ว เหมือนนาย มากทั้ง เป็น เห็น คุณ
ว่า ใคร

Copy Word Cloud

คลิกขวาเลือก Copy -> Image จากนั้นสามารถ
Paste หรือ Ctrl+v เพื่อ Copy รูปภาพของ Word
Cloud ไปยังโปรแกรมที่ต้องการ



แบบฝึกหัด

- สร้าง Tag Cloud โดยดึง #hashtag จากข้อมูลที่เก็บได้จาก Twitter โดยเลือกใช้ Keyword เพื่อค้นหาได้ตามต้องการ
- เพิ่มจำนวน tweets ที่เก็บจาก 100 เป็น 2000