

Data Science with Big Data Analysis

01418565

ไกรพิชญ์ กิติทัศน์เศรษฐี 5714450043

รังษกฤช มัญานันท์ 6014450017

วุฒิชัย ชุ่มเย็น 6014450076



THE EXPERT OF SMART HOME SYSTEM
ផ្សេខិត្តមានបច្ចនាការវឌ្ឍនីយ៍

Customer Service
086 322 8675

HOME

ABOUT US

PRODUCTS

SITE REFERENCE

CONTACT US

Enter Keyword



AUTO DOOR
SLIDING AND SWING DOOR



AUTO GATE
REMOTE CONTROL GATE



CARPARK SYSTEM
INTELLIGENT BARRIER



ACCESS CONTROL
KEYCARD/FINGER/FACE SCAN



SECURITY SYSTEM
HIP SECURITY SYSTEM



CCTV SYSTEM
HIP/AVTECH/INNEKT

Raw data

- รายการลูกค้าที่ติดต่อ
- รายได้รวมแต่ละเดือนประจำปี

2017,2018 ของบริษัท

ชื่อ	วันที่ปรับเปลี่ยน	ชนิด	ขนาด
5 พฤษภาคม 2560	8/12/2561 11:51	ไฟล์เดอร์แฟ้ม	
6 มิถุนายน 2560	8/12/2561 10:58	ไฟล์เดอร์แฟ้ม	
7 กรกฎาคม 60	8/12/2561 11:02	ไฟล์เดอร์แฟ้ม	
8 สิงหาคม 2560	8/12/2561 11:08	ไฟล์เดอร์แฟ้ม	
9 กันยายน 2560	8/12/2561 11:12	ไฟล์เดอร์แฟ้ม	
10 ตุลาคม 2560	8/12/2561 11:17	ไฟล์เดอร์แฟ้ม	
11 พฤศจิกายน 2560	8/12/2561 11:21	ไฟล์เดอร์แฟ้ม	
12 ธันวาคม 2560	8/12/2561 11:24	ไฟล์เดอร์แฟ้ม	
ไม่ใช้แล้ว กำลังจะลบ	15/11/2561 19:52	ไฟล์เดอร์แฟ้ม	
~\$4 เมษายน 2560.xlsx	13/7/2560 10:19	ไฟล์รีกซีต Microsoft E...	1 KB
~\$มกราคม 2560.xlsx	15/2/2560 14:57	ไฟล์รีกซีต Microsoft E...	1 KB
1 มกราคม 2560.xlsx	24/5/2560 16:59	ไฟล์รีกซีต Microsoft E...	42 KB
2 กุมภาพันธ์ 2560.xlsx	24/5/2560 15:55	ไฟล์รีกซีต Microsoft E...	51 KB
3 มีนาคม 2560.xlsx	29/6/2560 11:22	ไฟล์รีกซีต Microsoft E...	52 KB
4 เมษายน 2560.xlsx	28/6/2560 15:27	ไฟล์รีกซีต Microsoft E...	40 KB

Cleaning Data

ปัญหาที่เจอ

1. ชื่อพนักงานสะกดไม่เหมือนกัน เช่น แจ็ค, แจ็ค, Jack เป็นต้น
2. วันที่คนละ format
3. ไฟล์ excel ที่บันทึกในแต่ละเดือน บันทึกไม่เหมือนกัน เช่น บางคนบันทึก row เดียว บางคนผasan row
4. ไม่ระบุสถานที่ติดตั้งที่ชัดเจน
5. กรอกข้อมูลสลับกัน เช่น กรอกข้อมูล เบอร์โทร กับ สถานที่ติดตั้งสลับกัน
6. ข้อมูลที่ขายได้ หรือ ไม่ได้ ใช้วิธีระบุไม่เหมือนกัน เช่น คนหนึ่งใช้ตัวหนังสือสีเขียว อีกคนทำแบบ row สีเขียว

Raw (Before)

ข้อมูลที่ได้รับการป้องกัน โปรดระวังไฟล์ที่มาจากอินเทอร์เน็ตอาจมีไวรัส ถ้าคุณไม่จำเป็นต้องแก้ไขไฟล์ จะปลอดภัยกว่าหากอยู่ในมุมมองที่ได้รับการป้องกัน เปิดใช้งานการแก้ไข

A2	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
วัน/เดือน/ปี	ชื่อ-ที่อยู่ ลูกค้า	เบอร์ติดต่อ	สถานที่ติดตั้ง	ชนิดของงาน				เลขที่ใบเสนอราคา	พนักงานขาย	รายละเอียดการติดตาม					
				งานศิลปกรรม	งานประตูรั้ว	งานอโตตอร์	งานอื่น ๆ								
1/1/60	คณฤทธิ์ / wiwornh@gmail.com	089-794-4183	พุทธมงคล สาย 1			เสนอราคางาน เฟรมเลื่อนคู่		QT-050160-12	โน	คุยกับลูกค้าทางไลน์ ลูกค้าแจ้งว่ามีลูกค้า หน้างาน ยังไม่พร้อม ถ้าพร้อมแล้วจะแจ้ง 14/3/60 งานอีกในวันนี้จะลาก่อนผู้รับเหมาหลักมี มือหากับเจ้าของบ้านอยู่งานจึงยังไม่เดินต่อเลย					
4/1/60	คุณมนภา / ID : wannapha2525	096-845-5601							นิชา	จบเจ้าอื่นไปแล้ว					
4/1/60	คุณโนร์	02-564-7932-40 ต่อ 346	ศูนย์วิทยาศาสตร์เมืองไทย	เสนอราคางาน ทำบานพับตัวแบบ บานทึกข้อมูล				QT-040160-02	เบลล์	22/2/60 ศูนย์นี้บ่มีภาระหนักแล้ว แต่มีงานที่ ศูนย์ลงบันธ์ต่อ					
4/1/60	คุณเกตตี้	096-924-4910	ม.ธรรมศาสตร์ รังสิต			เสนอราคางาน เฟรมเดี่ยวขนาด 1.00*2.10 ม.		QT-040160-04	โน	31/01/60 โทรติดตามลูกค้า ลูกค้าขอใบเสนอ ราคากล่องรึ ขอ@Line ลูกค้าเพื่อคุยกับต่อเนื่อง ตามผลต่อ 16/2/60 ติดตั้งเจ้าอื่นไปแล้วบอกถูก กว่า					
4/1/60	คุณวินัย / คุณเจริญ hr@trimoro.com/winai@trimoro.com	091-881-9586/ 081-668-3706	พระราม 3			เสนอราคางาน Auto Swing Door (บาน เหล็กเฉียว) ขนาด 0.85*1.95			เบลล์	สรปงานแล้วรอติดตั้ง					

Cleaning (After)

มุมมองที่ได้รับการป้องกัน โปรดระวังไฟฟ้าที่มาจากการอิเล็กทรอนิกส์และอาจมีไวรัส ถ้าคุณไม่จำเป็นต้องแก้ไขไฟฟ้า จะปลอดภัยกว่าหากอยู่ในมุมมองที่ได้รับการป้องกัน เปิดใช้งานการแก้ไข

A1

X ✓ fx พนักงานขาย

1	พนักงานขาย	วัน/เดือน/ปี	ชื่อ-ที่อยู่ ลูกค้า	เบอร์ติดต่อ	สถานที่ตั้ง	งานคีย์การ์ด	งานประตูรั้ว	งานอ็อโต้ดอร์	งานอื่นๆ	เลขที่ใบเสนอราคา	รายละเอียดการติดตาม
2	Bell	4/1/60	คุณโนร์	02-564-7932-40 ต่อ 346	ศูนย์วิทยาศาสตร์มหาวิทยาลัย	เสนอรำคາคือร่องบานบัวแบบบันทึกช้อมูล				QT-040160-02	22/2/60 ศูนย์นั่งประชุมห้องแม่สัว แรมเมจันท์คุณโนร์
3	Bell	4/1/60	คุณวินัย / คุณเจษบhorn hr@trimoro.com/winai@trimoro.com	091-881-9586/ 081-668-3706	พระราม 3			เสนอรำคາคือร่องบานบัวแบบบันทึกช้อมูล			สรุปงานเด้วรอดิตติง
4	Bell	4/1/60	คุณนิรันดร์ suport55@hotmail.com / khuan513@gmail.com / nivarth@gmail.com	086-374-9647	ทุ่งสงขลาฯ 4	เสนอรำคາคือร่องบานบัวแบบบันทึกช้อมูล				QT-050160-09 QT-050160-11	ลูกค้าติดตั้งแล้ว
5	Bell	4/1/60	คุณชวัญ / khwan.bsc@gmail.com	097-131-0221	เมกาบางนา	เสนอรำคາคือร่องสแกนน้ำทึบแบบบันทึกช้อมูล				QT-050160-17	22/2/60 ลูกค้าไม่รับสาย
6	Bell	5/1/60	คุณป้อม / pramonsut@gmail.com	081-815-3382	สามาถห้าหัวหมาก	เสนอรำคາคือร่องบานบัวแบบบันทึกช้อมูล 2 ชุด				QT-050160-18	26/1/60 ลูกค้าแจ้งว่าไม่ต้องการสาย
7	Bell	5/1/60	คุณชนวนท์ / it.sahafavms@gmail.com	02-510-0051 ต่อ 451-452	รามอินทรา กม.8	เสนอรำคາคือร่องบานบัวแบบบันทึกช้อมูล				QT-050160-14	26/1/60 ลูกค้าแจ้งว่าต้องซ่อนบานบัวต่อไปใหม่
8	Bell	5/1/60	คุณยอด	02-168-7038,39	สุขุมวิท 13	เสนอรำคາคือร่องบานบัว บานไม้บานเดียว				QT-050160-16	22/2/60 คุณยอดไม่รู้ไฟตัดต่อไปใหม่
9	Bell	5/1/60	คุณไอล์ฟ	086-332-6894	ดอนเมือง	ลูกค้าอยากรื้อเปลี่ยนสแกนน้ำทึบแบบบันทึกช้อมูล max bt99					ไม่ได้ดำเนินเสนอรำคາเพราะว่าลูกค้าต้องการรุ่น
10	Bell	5/1/60	คุณธนาวดี / thanawut.ruin@gmail.com	088-212-6699	ทองหล่อ	เสนอรำคາคือร่องสแกนน้ำทึบแบบบันทึกช้อมูล				QT-060160-24	26/1/60 ลูกค้าแจ้งว่าติดตั้งแล้ว
11	Bell	6/1/60	คุณพุฒากร mt@mdrahotel.com	099-458-9696	โรงแรมอินทรา					QT-060160-22 / QT-060160-23	3/2/1960 ตั้งของชำร่วยที่ติดไปแล้ว
12	Bell	6/1/60	คุณพราพรรณ	081-811-7518	หัวลำโพง					QT-100160-40	22/2/60 ลูกค้าพิจารณาห้องน้ำจ่ายไม่เรียบร้อย
13	Bell	6/1/60	คุณพัชรา / mamua@windowslive.com	085-168-4271	วงศ์สว่าง 11	เสนอรำคາคือร่องบานบัวและเครื่องสแกนน้ำทึบแบบบันทึกช้อมูล				QT-090160-30	26/1/60 ลูกค้ารอนายสุรุ่ย
14	Bell	6/1/60	คุณสุ / Chm2552@yahoo.com	081-307-6534	ลาดพร้าว	เสนอรำคາคือร่องบานบัวและเครื่องสแกนน้ำทึบแบบบันทึกช้อมูล				QT-090160-27	26/1/60 ลูกค้าติดเข้าอื่นแล้ว แจ้งว่าเข้าอื่นราคายัง
15	Bell	9/1/60	คุณสุกิจ ID: JOOBOO	098-929-6710	รัชดา 166				บานเปิดอยู่เลื่อนซ้าย ขนาด 0.90*4.50 ม	QT-100160-32	
16	Bell	9/1/60	คุณอีต	084-018-7815	กำแพงเพชร					QT-090160-25	จบแล้ว
17	Bell	9/1/60	คุณชัยวัฒน์ / chaiyawatc@apcom.org	091-715-3634	สุขุมวิท	เสนอรำคາคือร่องสแกนน้ำทึบบานไม้บานเดียว				QT-090160-31	26/01/60 ลูกค้าร้องขอันยาติดตั้งหน้า
18	Bell	9/1/60	คุณพัทัศน์	089-063-7771	นواเทลส์ สะพานสูง	เสนอรำคາคือร่องบานบัวและเครื่องสแกนน้ำทึบแบบบันทึกช้อมูล 3 ชุด				QT-200160-87 QT-200160-88	22/2/2560 ลูกค้าห้องเข้าของงานบูรณะ ประมาณ
19	Bell	10/1/60	คุณทรงกรานต์ it_songkran@epicureans.co.th	088-088-1993	ทองหล่อ	เครื่องสแกนคือร่องบานบัวและบันทึกช้อมูล				QT-100160-33	26/1/60 รอเข้ามาพิจารณา
20	Bell	10/1/60	คุณben / sevenbuildingglass@gmail.com / ID : JANE_KKK	085-3240904	ดี.เค.รัชฐ์กิจ				เสนอรำคາบานเฟรมบานซ้าย ขนาด 1.00*2.10 ม.	QT-100160-37 / QT-100160-37-1	ตรวจสอบความเสียหายที่บานบัวและบันทึกช้อมูล
21	Bell	10/1/60	คุณโภสกhan / LuciferSound4@gmail.com	083-260-0879 / 02-860-7012	กรุงธนบุรี ช.6 อาคารสำนักงาน 8 ชั้น				เสนอรำคາบานเฟรมบานซ้าย ขนาด 1.00*2.10 ม.	QT-100160-35 / QT-110160-44	7/02 ໂປຣຄາມໃນเสนอรำคາ ลูกค้าแจ้งว่าหน้างานยังไม่
22	Bell	10/1/60	คุณไก่	089-020-7757	คลอง 15 นครนายก				เสนอรำคາบานเฟรมเดียว ขนาด 1.20*2.00 ม.	QT-100160-36 / QT-100160-39	7/02 ลูกค้าเปลี่ยนเนื่องจากบันทึกตั้งไม่พอ
23	Bell	10/1/60	คุณภูมิชัย / phuit@whitegroup.co.th	088-003-5263	โรงแรมลาลก้าวหน้าไทย	เสนอรำคາคือร่องบานบัว ทั้งหมด 13 ชุด				QT-110160-41	22/2/60 ลูกค้าติดตั้งเรียบร้อยแล้ว
24	Bell	10/1/60	คุณจันทร์ / gentechthai@hotmail.com	096-440-2445	เมืองทอง	เสนอรำคາคือร่องบานบัวและบันทึกช้อมูล 3 ชุด				QT-100160-38	26/1/60 ลูกค้าไม่รับสาย
25	Bell	11/1/60	คุณวิภาดา / asawa0001@yahoo.com	081-550-0065	พระราม 2				เสนอรำคາบานเฟรมเดียว ขนาด 0.90*1.80 ม.	QT-110160-43	7/02 ลูกค้าไม่ติดตั้งแล้ว
26	Bell	11/1/60	คุณกิตติศักดิ์	086-792-1199	เอกมัย 10				เสนอรำคາบานเฟรมเดียว ขนาด 10-15 ชุด.	QT-120160-46	7/02 ติดตั้งเข้าอื่นไปแล้ว

DataFrame

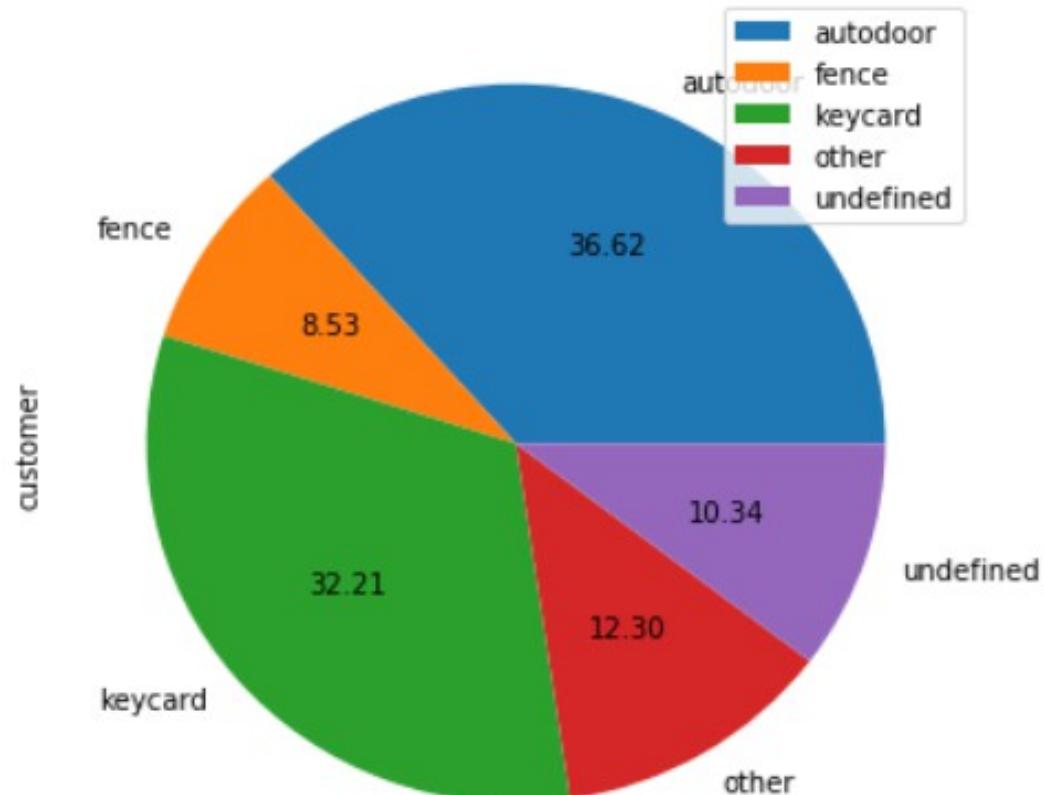
	staff	customer	latitude	longitude	success	day	month	year	job_type	date	dayofweek
0	เบล	คุณโนร์	14.095901	100.361614	False	4	1	2017	keycard	2017-01-04	Wednesday
1	เบล	คุณวินัย / คุณเจี๊ยบ hr@trimoro.com/winai@tri...	13.964549	100.274953	True	4	1	2017	autodoor	2017-01-04	Wednesday
2	เบล	คุณนิวัฒน์ supatt55@hotmail.com / khuan.s13@gm...	13.913982	99.803157	True	4	1	2017	keycard	2017-01-04	Wednesday
3	เบล	คุณชวัญ / khwan.bsc@gmail.com	13.597682	100.564099	False	4	1	2017	keycard	2017-01-04	Wednesday
4	เบล	คุณโ้อดิต / pramonsut@gmail.com	13.778840	100.604973	False	5	1	2017	keycard	2017-01-05	Thursday
5	เบล	คุณชนันท์ / it.sahafavms@gmail.com	13.819130	100.640444	False	5	1	2017	keycard	2017-01-05	Thursday
6	เบล	คุณเอ oy	13.726560	100.560886	False	5	1	2017	keycard	2017-01-05	Thursday
7	เบล	คุณไอซ์	13.921335	100.557837	False	5	1	2017	keycard	2017-01-05	Thursday
8	เบล	คุณธนวุฒิ / thanawut.ruin@gmail.com	13.726560	100.560886	False	5	1	2017	keycard	2017-01-05	Thursday
9	เบล	คุณคุณاجر mt@indrahotel.com	13.761010	100.524447	False	6	1	2017	autodoor	2017-01-06	Friday

1. วิเคราะห์การติดต่อของลูกค้า

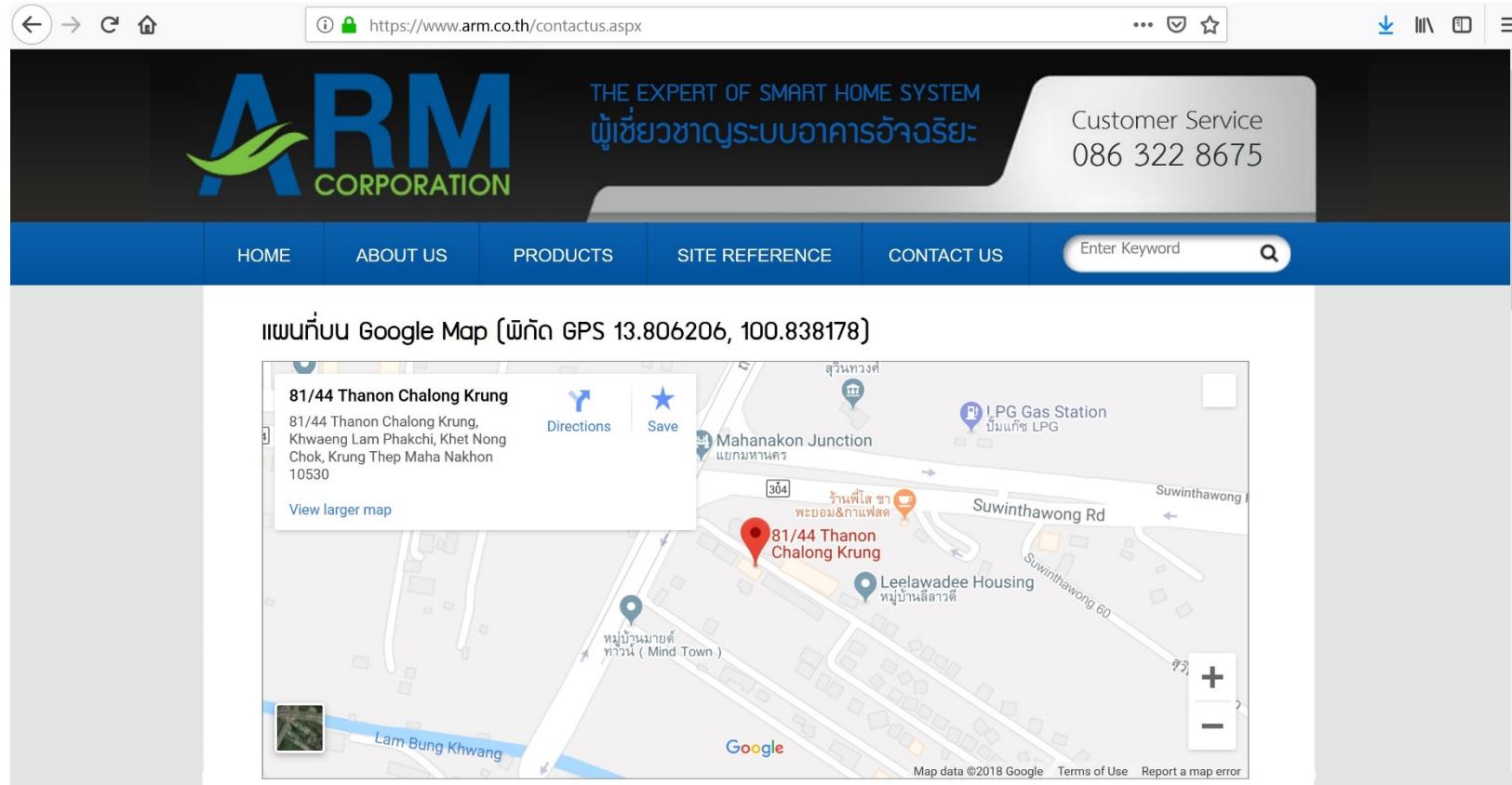
1.1 วิเคราะห์ตามหมวดสินค้า

ตารางสรุปลูกค้าติดต่อแยกตามหมวดสินค้าทั้งหมด

		customer		
year	2017	2018	TOTAL	
job_type				
autodoor	407	488	854	
fence	137	74	203	
keycard	406	356	732	
other	180	149	325	
undefined	144	121	257	
TOTAL	1224	1079	2161	



1.2 วิเคราะห์ตามระยะทางของลูกค้า



1.2 วิเคราะห์ตามระยะทางของลูกค้า



```
In [6]: import pandas as pd
import math
from geopy.distance import vincenty

arm_lo = (13.806206, 100.838178)
cust_lo = (13.822045, 100.578)
dist = vincenty(arm_lo, cust_lo)
print(dist)
```

28.185040806976655 km

1.2 วิเคราะห์ตามระยะทางของลูกค้า

```
In [7]: # clean na ใส่ที่ตั้งบริษัท Arm เข้าไปแทนเพื่อลบให้ได้ 0 km
values = {'latitude': 13.806206, 'longitude': 100.838178}
df.fillna(value=values,inplace=True)

# clean latitude error ใส่ที่ตั้งบริษัท Arm เข้าไปแทนเพื่อลบให้ได้ 0 km
df.loc[df['latitude']>90, 'latitude'] = 13.806206
df.loc[df['longitude']>90, 'longitude'] = 100.838178

df['distance'] = df.apply(lambda x: math.floor(vincenty((13.806206, 100.838178), (x[2], x[3])).meters/1000), axis = 1)
df
```

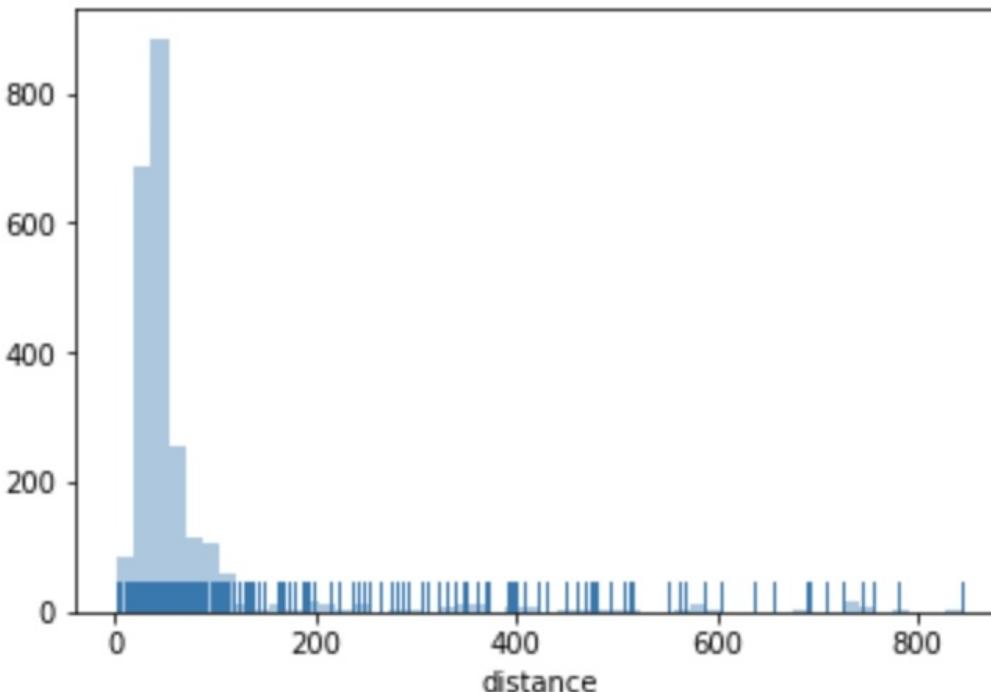
Out[7]:

	staff	customer	latitude	longitude	success	day	month	year	job_type	date	dayofweek	distance
0	เบล	คุณโนร์	14.095901	100.361614	False	4	1	2017	keycard	2017-01-04	Wednesday	60
1	เบล	คุณวินัย / คุณเจี๊ยบ hr@trimoro.com/winai@tri...	13.964549	100.274953	True	4	1	2017	autodoor	2017-01-04	Wednesday	63
2	เบล	คุณนิวัฒน์ supatt55@hotmail.com / khuan.s13@gm...	13.913982	99.803157	True	4	1	2017	keycard	2017-01-04	Wednesday	112
3	เบล	คุณชวัญ / khwan.bsc@gmail.com	13.597682	100.564099	False	4	1	2017	keycard	2017-01-04	Wednesday	37
4	เบล	คุณโอดี้ / pramonsut@gmail.com	13.778840	100.604973	False	5	1	2017	keycard	2017-01-05	Thursday	25
5	เบล	คุณชานนท์ / it.sahafavms@gmail.com	13.819130	100.640444	False	5	1	2017	keycard	2017-01-05	Thursday	21
6	เบล	คุณเอ oy	13.726560	100.560886	False	5	1	2017	keycard	2017-01-05	Thursday	31
7	เบล	คุณไอซ์	13.921335	100.557837	False	5	1	2017	keycard	2017-01-05	Thursday	32
8	เบล	คุณธนวุฒิ / thanawut.ruin@gmail.com	13.726560	100.560886	False	5	1	2017	keycard	2017-01-05	Thursday	31
9	เบล	คุณคุณกร mt@indrahotel.com	13.761010	100.524447	False	6	1	2017	autodoor	2017-01-06	Friday	34

1.2 ວິເຄາະທີ່ຕາມຮະຍະທາງຂອງລູກຄ້າ

```
In [21]: import seaborn as sns  
sns.distplot(series_dist,norm_hist=False,kde=False,rug=True)  
print(series_dist.quantile([0.5,0.8,0.9]))
```

```
0.5      53.0  
0.8      78.0  
0.9     112.0  
Name: distance, dtype: float64
```



1.3 วิเคราะห์การติดต่อของลูกค้าแยกตามวัน

จำนวนลูกค้าที่ติดต่อทั้งหมดแยกตามวัน

dayofweek	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
customer	550	530	591	552	476	60	54

จำนวนวันที่ทำงาน

workingday

dayofweek

Friday	86
Monday	85
Saturday	30
Sunday	26
Thursday	88
Tuesday	87
Wednesday	91

1.3 วิเคราะห์การติดต่อของลูกค้าแยกตามวัน

dayofweek	customer	workingday	customerperday
Wednesday	591	91	6.494505
Monday	550	85	6.470588
Thursday	552	88	6.272727
Tuesday	530	87	6.091954
Friday	476	86	5.534884
Sunday	54	26	2.076923
Saturday	60	30	2.000000

สรุปวิเคราะห์การติดต่อของลูกค้า

1. ลูกค้ากลุ่มใหญ่ ประมาณ 36% ติดต่อเกี่ยวกับสินค้าชนิด autodoor
2. กลุ่มลูกค้า เกิน 80% จะอยู่ห่างไม่เกิน 78 km.

2. วิเคราะห์รายได้

Raw

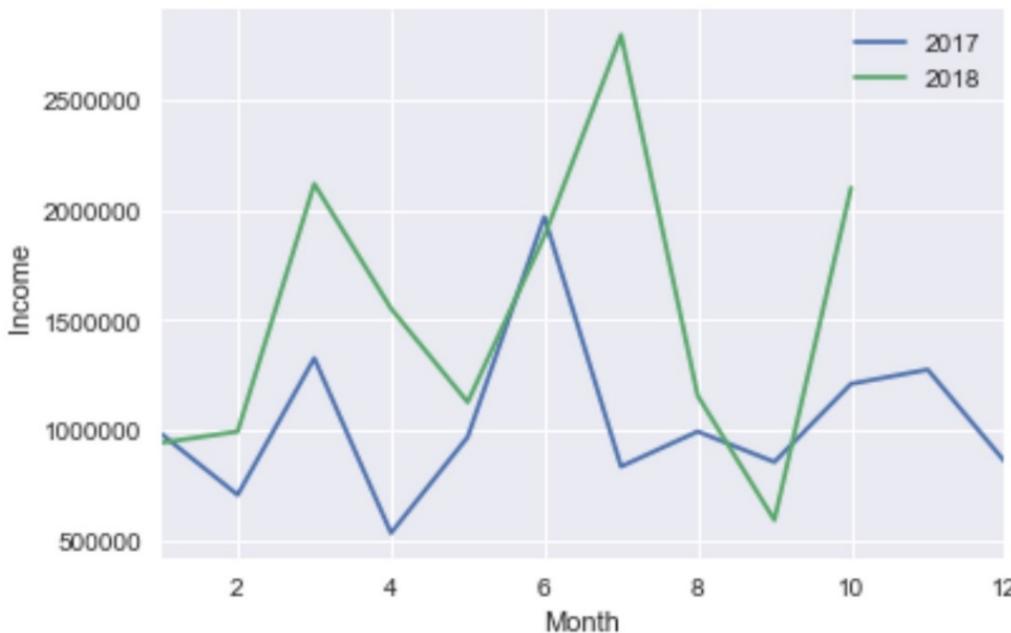
- ได้รับข้อมูลเป็นไฟล์ pdf
- เป็นข้อมูลรายได้ที่ต้องยื่นส่งสรรพากร
- ข้อมูลปี 2017-2018
- หมายเหตุข้อมูลปี 2018 มี 11 เดือน

สรุปยอด

	2560	2561
มกราคม	983670	941800
กุมภาพันธ์	705546	992976
มีนาคม	1326058	2117658
เมษายน	532450	1551795
พฤษภาคม	968117	1126477
มิถุนายน	1968708	1879349
กรกฎาคม	834100	2793935
สิงหาคม	993250	1155700
กันยายน	854669	591260
ตุลาคม	1209784	2099361
พฤษจิกายน	1274075	2186598
ธันวาคม	858750	
รวม	12509177	17436909

2.1 ວິເຄຣະໜ່າຍໄດ້ປະຈຳປີ

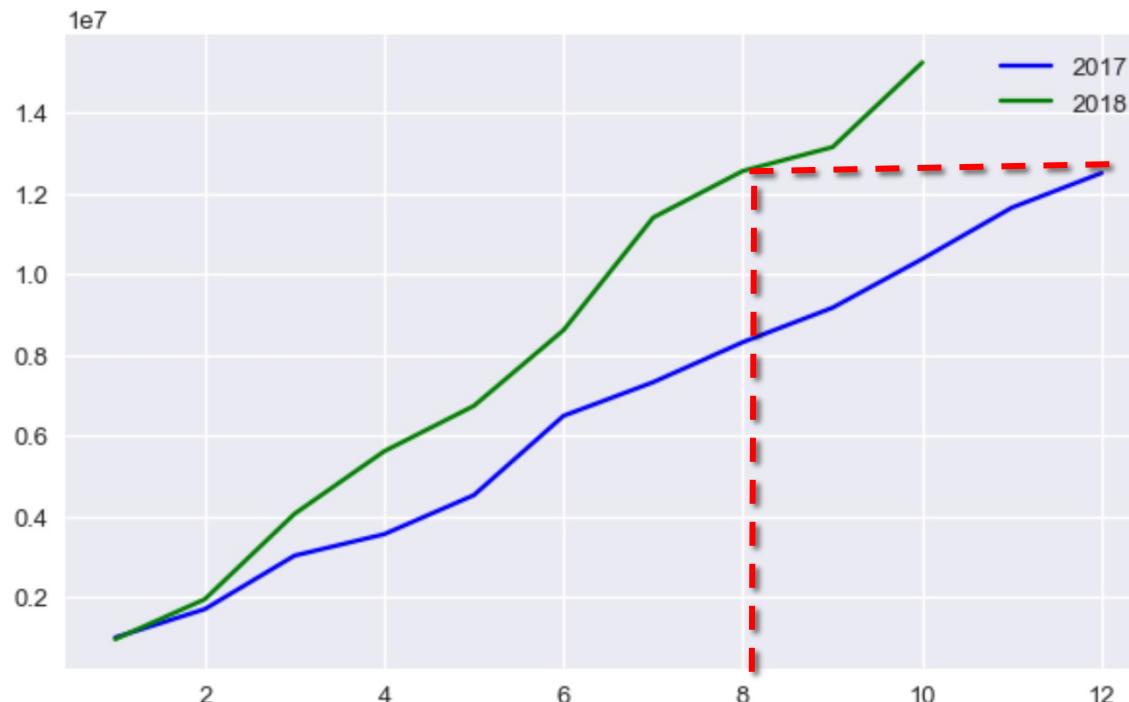
```
In [7]: fig, ax = plt.subplots()
dfincome.groupby('year').plot(kind='line',x='month',y='income',ax=ax, legend=False)
plt.xlabel('Month')
plt.ylabel('Income')
ax.legend(['2017', '2018']);
plt.show()
```



ການແສດງຮາຍໄດ້ແບກຕາມເດືອນ

2.1 ວິເຄາະຫໍ່ຮ່າຍໄດ້ປະຈຳປີ

```
In [34]: dfincome17 = dfincome17.loc[dfincome['year'] == 2017]
dfincome17['acc'] = dfincome17['income'].cumsum(axis = 0)
dfincome18 = dfincome18.loc[dfincome['year'] == 2018]
dfincome18['acc'] = dfincome18['income'].cumsum(axis = 0)
fig, ax = plt.subplots()
ax.plot(dfincome17['month'],dfincome17['acc'],color='blue')
ax.plot(dfincome18['month'],dfincome18['acc'],color='green')
ax.legend(['2017', '2018']);
```

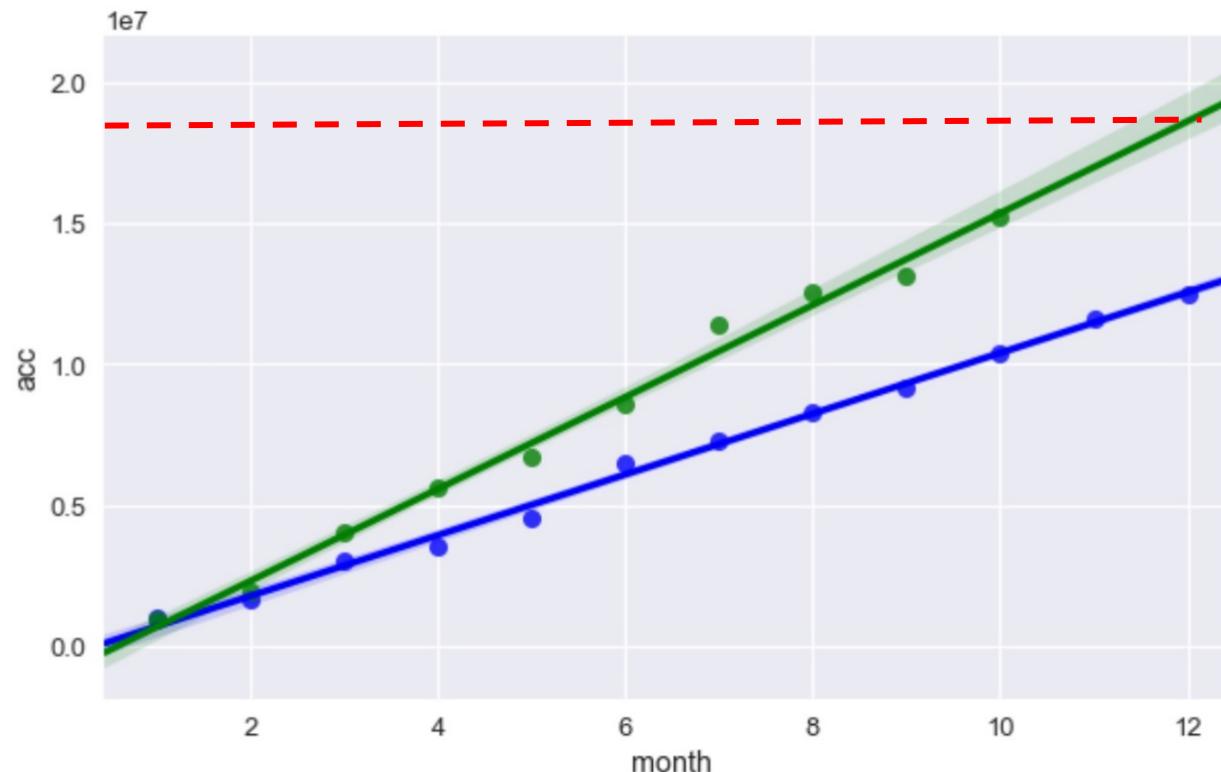


ການແສດງຮ່າຍໄດ້ສະສນ

2.1 විකරාහ්‍රය තිබුප්‍රස්ථාපී

ทำ Regression เพื่อทำนายผลรายได้สิ้นปี 2018

```
In [35]: ax = sns.regplot(x="month", y="acc", data=dfincome17,color='blue')
ax = sns.regplot(x="month", y="acc", data=dfincome18,color='green')
```



ทำนายผลรายได้สิ้นปี 2018 තිබු
ප්‍රමාණ 18.2 ලාංඡනบาท

2.2 วิเคราะห์แยกตามพนักงาน

ตารางแสดงผลการขายของพนักงาน

2017

	staff	False	True	TOTAL	SuccessRate
0	Bell	369	45	414	0.108696
1	Jack	224	22	246	0.089431
2	Moi	97	19	115	0.165217
3	NanNan	38	3	39	0.076923
4	Nicha	65	8	72	0.111111
5	Oil	6	1	7	0.142857
6	Pai	259	58	310	0.187097
7	Roml	38	5	43	0.116279
8	Ta	64	3	67	0.044776
9	TOTAL	1079	159	1224	0.129902

2018

	staff	False	True	TOTAL	SuccessRate
0	Jack	384	89	458	0.194323
1	Na	23	4	27	0.148148
2	NanNan	130	44	173	0.254335
3	Pai	470	68	522	0.130268
4	TOTAL	924	194	1079	0.179796

*หมายเหตุ ปี 2018 มีจำนวนเดือน 10 เดือน

2.3 ວິເຄາະທີ່ Success rate ຂອງໜົດສິນຄ້າ

	2017					2018					
	job_type	False	True	TOTAL	SuccessRate		job_type	False	True	TOTAL	SuccessRate
0	autodoor	359	50	407	0.122850	0	autodoor	439	58	488	0.118852
1	fence	125	12	137	0.087591	1	fence	66	8	74	0.108108
2	keycard	372	37	406	0.091133	2	keycard	302	57	356	0.160112
3	other	123	57	180	0.316667	3	other	79	74	149	0.496644
4	undefined	139	5	144	0.034722	4	undefined	114	7	121	0.057851
5	TOTAL	1079	159	1224	0.129902	5	TOTAL	924	194	1079	0.179796

สรุปวิเคราะห์รายได้

1. จากราฟรายได้สะสม แสดงให้เห็นว่า ปี 2018 เติบโตกว่า ปี 2017 อย่างชัดเจน
2. ได้ทำการทำนายผลของรายได้สิ้นปี 2018 จากการ Regression
3. จากตารางผลการขายของพนักงาน แสดงให้เห็น Success rate ของปี 2018 ดีขึ้น 5%
4. สินค้า autodoor มีจำนวนลูกค้าเพิ่มขึ้นต่างจากสินค้าชนิดอื่นๆ ที่ลดลง
5. ผลทำนายจาก Regression จะได้ยอดขายปี 2018 เท่ากับ 18.2 ล้าน

3. Machine Learning (Decision tree)

```
# จัด Category ให้ระยะทางเป็น 3 ระยะคือ Near, Mid, Far
df.loc[(df['distance']>0)&(df['distance']<=30), 'dist'] = 'Near'
df.loc[(df['distance']>30)&(df['distance']<=60), 'dist'] = 'Mid'
df.loc[(df['distance']>60), 'dist'] = 'Far'

#สร้างผลเฉลยให้เป็น 0 กับ 1
df.loc[(df['success']==True), 'buy'] = 1
df.loc[(df['success']==False), 'buy'] = 0

dataset = df.drop(columns=['latitude', 'longitude', 'day', 'month', 'year', 'date', 'distance'], axis="columns")
#กรอง Nan ออกและเอา job_type ที่ไม่ระบุออก
dataset = dataset.dropna()
dataset = dataset[dataset['job_type'] != 'undefined'];
#เรียง column ใหม่
dataset = dataset[['staff', 'job_type', 'dayofweek', 'dist', 'buy']]
dataset.reset_index(drop=True, inplace=True)
print('dataset=' + str(len(dataset)) + ' rows')
dataset.head()
```

dataset=2230 rows

	staff	job_type	dayofweek	dist	buy
0	Bell	keycard	Wednesday	Mid	0.0
1	Bell	autodoor	Wednesday	Far	1.0
2	Bell	keycard	Wednesday	Far	1.0
3	Bell	keycard	Wednesday	Mid	0.0
4	Bell	keycard	Thursday	Near	0.0

3. Machine Learning (Decision tree)

```
from sklearn.model_selection import train_test_split
from sklearn import tree
from sklearn import preprocessing
from sklearn.preprocessing import LabelEncoder
from sklearn.feature_extraction import DictVectorizer

X_feature = dataset[['staff', 'job_type', 'dayofweek', 'dist']]
X_dict = X_feature.T.to_dict().values()

# เปลี่ยน string ที่เป็น category ให้成 numpy array
vect = DictVectorizer(sparse=False)
X_vector = vect.fit_transform(X_dict)
y = dataset['buy'][:].values

# แบ่ง train, test
X_train, X_test, y_train, y_test = train_test_split( X_vector, y, test_size = 0.3, random_state = 100)

clf = tree.DecisionTreeClassifier(criterion='entropy')
clf = clf.fit(X_train,y_train)
```

3. Machine Learning (Decision tree)

```
from sklearn.metrics import accuracy_score

predict = clf.predict(X_test)
accuracy = accuracy_score(y_test, predict)

print ('Accuracy is:' + str(accuracy))
```

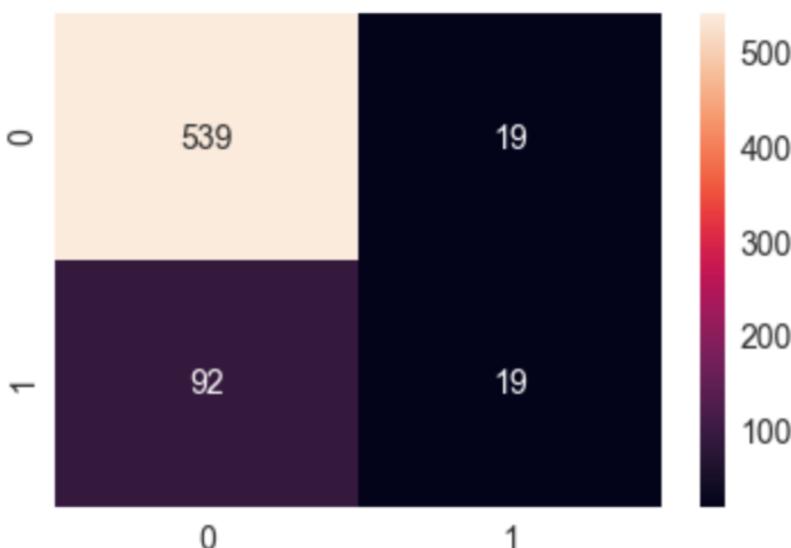
Accuracy is:0.8340807174887892

3. Machine Learning (Decision tree)

```
: import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix

cm = confusion_matrix(y_test,predict)
sns.set(font_scale=1.4)
sns.heatmap(cm, annot=True,annot_kws={"size": 14},fmt="d")

: <matplotlib.axes._subplots.AxesSubplot at 0x1a1991e828>
```





A large, three-dimensional, blue sign with white text that reads "Thank You". The letters are rounded and have a slight shadow, giving them a 3D appearance. The sign is suspended by a thin brown string from a small metal hook at the top center. The background is plain white.
