

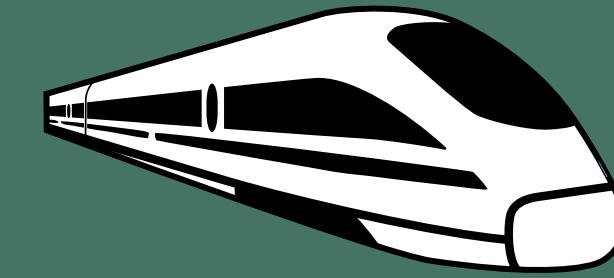
# Project

# NUMBERONE

20



# ข้อมูลที่ได้มาจากการ 5 แหล่ง



ข้อมูล  
กลุ่มข้อมูล

สถานีรับส่งผู้โดยสารการรถไฟ  
องค์กร : การรถไฟแห่งประเทศไทย

ดาวน์โหลดกันหมด

[https://data.go.th/dataset/  
station](https://data.go.th/dataset/station)

ข้อมูลสถานีรับส่งผู้โดยสาร  
การรถไฟแห่งประเทศไทย

ข้อมูล  
กลุ่มข้อมูล

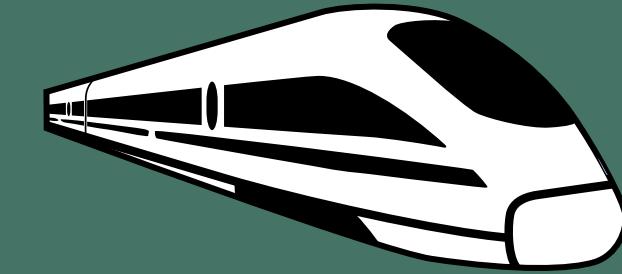
ขบวนรถไฟฟ้าหัวรับ-ส่งผู้โดยสาร  
องค์กร : การรถไฟแห่งประเทศไทย

ดาวน์โหลดกันหมด

[https://data.go.th/dataset/  
train](https://data.go.th/dataset/train)

ข้อมูลขบวนรถไฟฟ้าหัวรับ-ส่ง  
ผู้โดยสาร

# ข้อมูลที่ได้มาจากการ 5 แหล่ง



**สถานีรถไฟ**

สถานีรถไฟ (ข้อมูล ณ สิงหาคม 2563)

ข้อมูลและทรัพยากร

RailStation 13 down

สกัดคุณภาพ โครงสร้างพื้นฐาน

ข้อมูลเพิ่มเติม

ผลิต สร้างในระบบเมื่อ ปรับปรุงครั้งล่าสุดในระบบเมื่อ

ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร ไม่มีคำว่ารายสื่อขององค์กรนี้

สังคม

บุคคลของคุณ

<https://datagov.mot.go.th/dataset/train-station>

ข้อมูล railstation

ธนาคารแห่งประเทศไทย

เกี่ยวกับ ธปท. นโยบายการเงิน

หน้าหลัก > สถิติ > การรับถึงข้อมูลกับ ธปท. > TempClose

<https://www.bot.or.th/Thai/Statistics/DataManagementSystem/TempClose>

ข้อมูลชื่อตำบล อำเภอ และจังหวัด

ขนาดตัวอักษร : A- A A+ | C C C

กระทรวงการท่องเที่ยวและกีฬา  
Ministry of Tourism & Sports

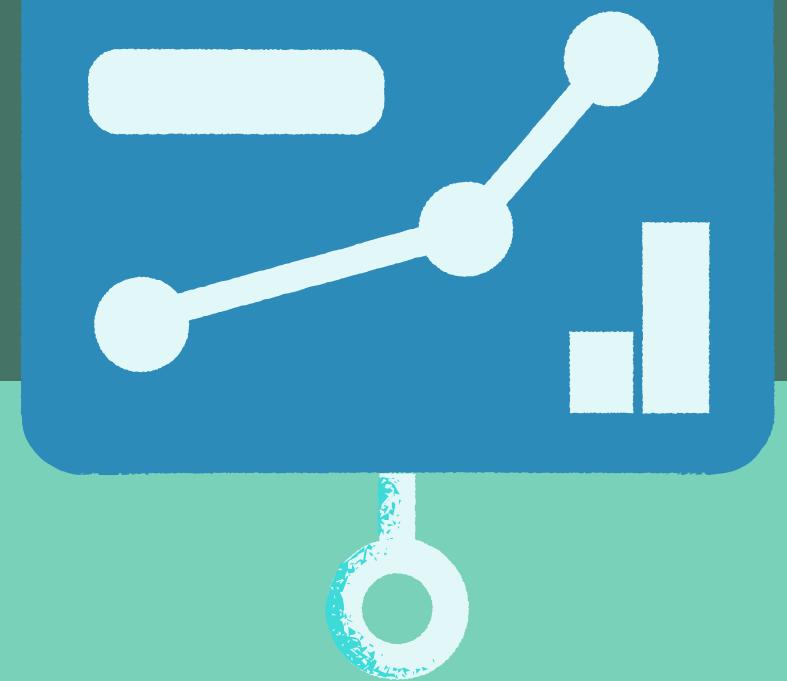
Home / สถิติ按年份 / สถิติด้านการท่องเที่ยว ปี 2563 / สถานการณ์การท่องเที่ยวในประเทศไทย รายจังหวัด ปี 2563

สถานการณ์การท่องเที่ยวในประเทศไทย รายจังหวัด ปี 2563

[https://mots.go.th/more\\_news\\_new.php?cid=594](https://mots.go.th/more_news_new.php?cid=594)

สถานการณ์การท่องเที่ยวในประเทศไทย รายจังหวัด ปี 2563

# Challenge



1. ทำนายประเภทไฟที่จะเลือกใช้บริการ โดยใช้ระยะทางระหว่างสถานี จำนวนนักท่องเที่ยวในจังหวัดต้นทาง จำนวนนักท่องเที่ยวในจังหวัดปลายทาง
2. แบ่งกลุ่มข้อมูลที่มี ลงจิจูด(column : X) และ ละติจูด (column : Y) ที่ใกล้กันจำนวน 6 กลุ่ม
3. หาจังหวัดที่มีรถไฟข้ามมากที่สุด



# Preprocessing

## ตารางที่ 1 ตาราง ข้อมูลสถานีรับส่งผู้โดยสารการรถไฟแห่งประเทศไทย

```
data_st = pd.read_csv(os.path.join(path,'data_station.csv')) #กำหนดตัวแปร data_st เป็นข้อมูลของสถานีรถไฟแต่ละสถานี
data_st #เรียกดูตารางของสถานีรถไฟแต่ละสถานี
```

	<b>id</b>	<b>station_code</b>	<b>name</b>	<b>en_name</b>	<b>th_short</b>	<b>en_short</b>	<b>chname</b>	<b>controldivision</b>	<b>exact_km</b>	<b>exact_distance</b>	<b>km</b>	<b>class</b>	<b>lat</b>	<b>long</b>	<b>act</b>
0	1	1001	กรุงเทพ	Bangkok	กท.	BKK	曼谷	14	0	0	0	11	13.739277	100.517391	1
1	2	1002	ยมราช	Yommaraj	ยช.	YMR	勇马拉	14	2	2	2	99	0	0	1
2	3	1004	สามเสน	Sam Sen	สส.	SAM	三森	14	5	4	4	1	13.779293	100.529235	1
3	4	1006	ประดิพัทธ์	Pra Dipat	ปธ.	PDP	帕涤潘	14	6	6	6	99	0	0	1
4	5	1007	ชุมทาง บางซื่อ	Bang Sue Junction	บช.	BSJ	邦苏总 站	14	7	7	7	1	13.8043	100.54005	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
677	678	5025	บางโหรัด	Bang Thorat	งโ.	BTR	邦托拉	0	14	-13	13	98	0	0	1
678	679	5026	บ้าน กาหลง	Ban Ka Long	กท.	BLG	班卡龙	0	16	-15	15	98	0	0	1
679	680	5027	บ้านนา ขาวง	Ban Na Khwang	อัม.	BNW	班那煌	0	18	-17	17	98	0	0	1
680	681	5028	บ้านนา โคก	Ban Na Khok	กโ.	BNH	班那柯	0	20	-19	19	98	0	0	1
681	682	5029	เขตเมือง	Ked Muang	ต.	KEM	可孟	0	24	-23	23	99	0	0	1

# Preprocessing

## ตารางที่ 2 ตารางข้อมูลขบวนรถไฟฟ้าหัวรับรับ-ส่งผู้โดยสาร

```
data_t = pd.read_csv(os.path.join(path, 'data_train.csv')) #กำหนดตัวแปร data_t เป็นข้อมูลของรถไฟ  
data_t #เรียกดูตารางข้อมูลรถไฟ
```

	Zone	goback	Num	IdTrain	NameFee	StStart	NameSt	StFinal	NameSt_L	NameStEng	NameStEng_L	Service
0	1	1	0.0	997	เรือ	1001	กรุงเทพ	3047	บ้านพลูตาหลวง	Bangkok	Ban Plu Ta Luang	False
1	1	1	1.0	1049	ธรรมดा	1001	กรุงเทพ	1007	ชุมทางบางซื่อ	Bangkok	Bang Sue Junction	True
2	1	1	1.0	379	ธรรมดा	3001	มัคกะสัน	3015	หัวตะเข้	Makkasan	Hua Takhe	False
3	1	1	2.0	1053	ธรรมดा	1001	กรุงเทพ	1007	ชุมทางบางซื่อ	Bangkok	Bang Sue Junction	True
4	1	1	4.0	283	ธรรมดा	1001	กรุงเทพ	3047	บ้านพลูตาหลวง	Bangkok	Ban Plu Ta Luang	False
...	...	...	...	...	...	...	...	...	...	...	...	...
229	5	2	26.0	260	ธรรมดा	4077	น้ำตก	4002	ธนบุรี	Nam Tok	Thon Buri	False
230	5	2	27.0	252	ธรรมดा	4142	ประจวบคีรีขันธ์	4002	ธนบุรี	Prachuap Khiri Khan	Thon Buri	False
231	5	2	28.0	254	ธรรมดा	4203	หลังสวน	4002	ธนบุรี	Lang Suan	Thon Buri	False
232	5	2	29.0	258	ธรรมดा	4077	น้ำตก	4002	ธนบุรี	Nam Tok	Thon Buri	False
233	5	2	30.0	486	ท่องถิ่น	4077	น้ำตก	4020	ชุมทางหนองปลาดุก	Nam Tok	Nong Pla Duk Junction	False

# Preprocessing

## ตารางที่ 3 ตารางข้อมูล railstation

```
data_rst = pd.read_csv(os.path.join(path,'railstation_new.csv')) #กำหนดตัวแปร data_rst ให้เป็นร่างร็อกไฟ  
data_rst #เรียกดูตารางของร่างร็อกไฟ
```

	OBJECTID	NAMET	NAMEE	ProvName	AmpName	TamName	X	Y
0	1	บ้านເກົ່າປິ່ງ	BAN KO PRING UNMANNED STATION		ທຸງສົງ	NaN	99.644994	8.213274
1	2	ຫ້ວຍຍອດ	HUAI YOT RAILWAY STATION		ຫ້ວຍຍອດ		99.637987	7.785456
2	3	ໂຮງຮຽນການບິນ	FLYING TRAINING SCHOOL UNMANNED STATION		ກຳພັງແສນ		99.956660	14.088246
3	4	ບ້ານປັກແຮດ	BAN PLUK RAET UNMANNED STATION		ທ່າເຮືອ		100.725785	14.530267
4	5	ເດັ່ນຂໍ້	DEN CHAI RAILWAY STATION	PHRAE	ເດັ່ນຂໍ້	DEN CHAI	100.046225	17.980781
...	...	...	...	...	...	...	...	...
677	678	ຄູນຢ່າຊກາຮຈັງຫວັດສະແກ້ວ	THE GOVERNMENT COMPLEX SA KAEO UNMANNED STATION		ເມືອງສະແກ້ວ		102.138802	13.799641
678	679	ທຸງທອງ	THUNG THONG UNMANNED STATION		ທ່ານ່ວງ		99.638000	13.980168
679	680	ຕລາດທນອນຄາຍ	Talat Nong Khai		ເມືອງທນອນຄາຍ		102.725416	17.878784
680	681	ແກຣນົດ ດາແນລ	GRAND CANEL UNMANNED STATION		ດອນເມືອງ		100.608267	13.941516
681	682	ຫຼັກໂພງ	BANGKOK RAILWAY STATION	BANGKOK	ປາທຸນວັນ	RONG MUEANG	100.516732	13.744026

# Preprocessing

## ตารางที่ 4 ตารางข้อมูลชื่อตำบล อําเภอ และจังหวัด

```
data_location = pd.read_csv(os.path.join(path, 'Location_Postal_Code.csv')) ## นำเข้าข้อมูลชื่อตำบล อําเภอ และจังหวัด
data_location ##เรียกดูข้อมูลชื่อตำบล อําเภอ และจังหวัด
```

	Location Code	จังหวัด	เขต/อำเภอ	แขวง/ตำบล	รหัสไปรษณีย์
0	100101	กรุงเทพ	พระนคร	พระบรมมหาราชวัง	10200
1	100102	กรุงเทพ	พระนคร	วังบูรพาภิรมย์	10200
2	100103	กรุงเทพ	พระนคร	วัดราชนพิธ	10200
3	100104	กรุงเทพ	พระนคร	สำราญราษฎร์	10200
4	100105	กรุงเทพ	พระนคร	ศาลาเจ้าพ่อเสือ	10200
...	...	...	...	...	...
7421	961203	นราธิวาส	จะแนะ	ผดุงมาตร	96220
7422	961204	นราธิวาส	จะแนะ	ช้างเผือก	96220
7423	961301	นราธิวาส	เจาะไกรทอง	จุบ	96130
7424	961302	นราธิวาส	เจาะไกรทอง	บุกิต	96130
7425	961303	นราธิวาส	เจาะไกรทอง	มะรือโนบอค	96130



# Preprocessing

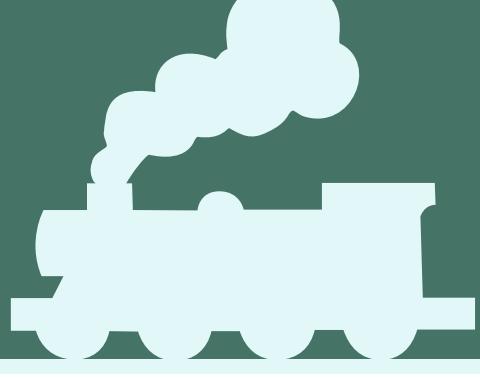
## ตารางที่ 5 ตารางข้อมูลสถานการณ์การท่องเที่ยวในประเทศไทย รายจังหวัด ปี 2563

```
data_tourist = pd.read_csv(os.path.join(path,'สถานการณ์การท่องเที่ยวในประเทศไทย รายจังหวัด ปี 2563.csv')) ## นำเข้าข้อมูลสถานการณ์การท่องเที่ยวในประเทศไทย รายจังหวัด ปี 2563  
data_tourist ##เรียกดูข้อมูล
```

	ภาค	จังหวัด	จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)
0	ภาคกลาง	กรุงเทพ	23617914.0
1	ภาคกลาง	ลพบุรี	1659868.0
2	ภาคกลาง	พระนครศรีอยุธยา	3560866.0
3	ภาคกลาง	สระบุรี	1673819.0
4	ภาคกลาง	ชัยนาท	394900.0
...	...	...	...
72	ภาคตะวันออกเฉียงเหนือ	อุบลราชธานี	1537178.0
73	ภาคตะวันออกเฉียงเหนือ	สกลนคร	856974.0
74	ภาคตะวันออกเฉียงเหนือ	ยโสธร	303461.0
75	ภาคตะวันออกเฉียงเหนือ	อ่างทอง	153692.0
76	ภาคตะวันออกเฉียงเหนือ	หนองบัวลำภู	191384.0

77 rows × 3 columns





# ເຊື່ອມ 2 ຕາຮາງເຂົ້າດ້ວຍກັນ

10

```
table_l1 = data_st.merge(data_RST, how='right', left_on='name', right_on='NAMET') #ทำการเชื่อมตาราง 2 ตาราง คือ data_RST และ data_st  
table_l1
```

# เชื่อม 3 ตารางเข้าด้วยกัน

```
table_l2 = table_l1.merge(data_location,how='left',left_on='AmpName',right_on='เขต/อำเภอ') #ทำการเชื่อมตาราง 3 ตาราง คือ data_rst data_st และ data_location
table_l2
```

	id	station_code	name	en_name	th_short	en_short	chname	controldivision	exact_km	exact_distance	km	class	lat	long	ad
0	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
1	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
2	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
3	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
4	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7060	138.0	1233.0	แกรนด์ คานาล	Grand Canal	คาน.	GRC	格兰卡 内	14.0	24.0	24.0	24.0	99.0	0	0	1.



# เชื่อม 4 ตารางเข้าด้วยกัน

```
table_l3 = table_l2.merge(data_tourist,how='left',left_on='จังหวัด',right_on='จังหวัด') #ทำการเชื่อมตาราง 4 ตาราง คือ data_rst data_st data_location เชื่อมกับ data_tourist
table_l3
```

	id	station_code	name	en_name	th_short	en_short	chname	controldivision	exact_km	exact_distance	km	class	lat	long	ac
0	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
1	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
2	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
3	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
4	565.0	4281.0	ห้วยยอด	Huai Yot	ยอด.	HYO	辉勇	42.0	817.0	-800.0	800.0	3.0	7.786067	99.638224	1.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7030	138.0	1233.0	แกรนด์ คานาเล	Grand Canal	คาน.	GRC	格兰卡 内	14.0	24.0	24.0	24.0	99.0	0	0	1.

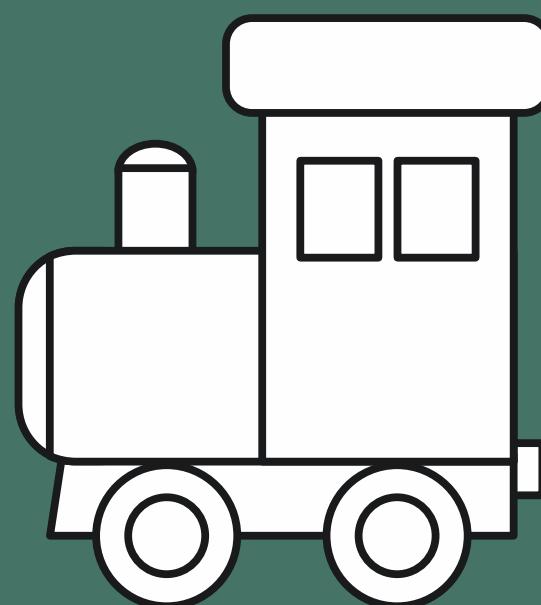


# เลือกคอลัมน์ของสถานีปลายทาง

```
table_l = table_l3[['station_code','name','exact_km','km','class','X','Y','จังหวัด','ภาค','จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)']] ##เลือกคอลัมน์ที่ต้องการ
table_l ##แสดงข้อมูลของตารางที่เลือก
```

	station_code	name	exact_km	km	class	X	Y	จังหวัด	ภาค	จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)
0	4281.0	ห้วยยอด	817.0	800.0	3.0	99.637987	7.785456	ตรัง	ภาคใต้	648092.0
1	4281.0	ห้วยยอด	817.0	800.0	3.0	99.637987	7.785456	ตรัง	ภาคใต้	648092.0
2	4281.0	ห้วยยอด	817.0	800.0	3.0	99.637987	7.785456	ตรัง	ภาคใต้	648092.0
3	4281.0	ห้วยยอด	817.0	800.0	3.0	99.637987	7.785456	ตรัง	ภาคใต้	648092.0
4	4281.0	ห้วยยอด	817.0	800.0	3.0	99.637987	7.785456	ตรัง	ภาคใต้	648092.0
...	...	...	...	...	...	...	...	...	...	...
7030	1233.0	แกรนด์ คานแนล	24.0	24.0	99.0	100.608267	13.941516	กรุงเทพ	ภาคกลาง	23617914.0
7031	1001.0	กรุงเทพ	0.0	0.0	11.0	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
7032	1001.0	กรุงเทพ	0.0	0.0	11.0	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
7033	1001.0	กรุงเทพ	0.0	0.0	11.0	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
7034	1001.0	กรุงเทพ	0.0	0.0	11.0	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0

7035 rows × 10 columns



# เชื่อม 2 ตารางเข้าด้วยกัน



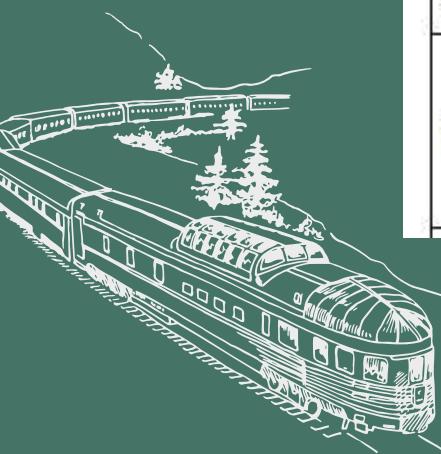
```
table_st1 = data_t.merge(data_rst,how='left',left_on='NameSt',right_on='NAMET') #ทำการเชื่อมตาราง 2 ตาราง คือ data_rst และ data_t
table_st1
```

	Zone	goback	Num	IdTrain	NameFee	StStart	NameSt	StFinal	NameSt_L	NameStEng	NameStEng_L	Service	OBJECTID	NAMET	NAMEE
0	1	1	0.0	997	เรือ	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGK RAILWA STATIO
1	1	1	1.0	1049	ธรรมดा	1001	กรุงเทพ	1007	ชุมทาง บางซื่อ	Bangkok	Bang Sue Junction	True	682.0	กรุงเทพ	BANGK RAILWA STATIO
2	1	1	1.0	379	ธรรมดा	3001	มัจฉาชีวิน	3015	หัวตะเข้	Makkasan	Hua Takhe	False	140.0	มัจฉาชีวิน	MAKKA RAILWA STATIO
3	1	1	2.0	1053	ธรรมดा	1001	กรุงเทพ	1007	ชุมทาง บางซื่อ	Bangkok	Bang Sue Junction	True	682.0	กรุงเทพ	BANGK RAILWA STATIO
4	1	1	4.0	283	ธรรมดा	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGK RAILWA STATIO
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
229	5	2	26.0	260	ธรรมดा	4077	น้ำตก	4002	ธนบุรี	Nam Tok	Thon Buri	False	610.0	น้ำตก	NAM T RAILWA STATIO

# เชื่อม 3 ตารางเข้าด้วยกัน

```
table_st2 = table_st1.merge(data_location, how='left', left_on='AmpName', right_on='เขต/อำเภอ') #ทำการเชื่อมตาราง 3 ตาราง คือ data_t data_rst เชื่อมกับ data_location
table_st2
```

	Zone	goback	Num	IdTrain	NameFee	StStart	NameSt	StFinal	NameSt_L	NameStEng	NameStEng_L	Service	OBJECTID	NAMET	NAMEE	Prov
0	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
1	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
2	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
3	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
4	1	1	1.0	1049	ธรรมชาติ	1001	กรุงเทพ	1007	ชุมทาง บางซื่อ	Bangkok	Bang Sue Junction	True	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2395	5	2	30.0	486	ท่องถิน	4077	น้ำตก	4020	ชุมทาง หนองปลา ดุก	Nam Tok	Nong Pla Duk Junction	False	610.0	น้ำตก	NAM TOK RAILWAY STATION	



# เชื่อม 4 ตารางเข้าด้วยกัน

table\_st3 = table\_st2.merge(data\_tourist, how='left', left\_on='จังหวัด', right\_on='จังหวัด') #ทำการเชื่อมตาราง 4 ตาราง คือ data\_t data\_rst data\_location เชื่อมกับ data\_tourist

	Zone	goback	Num	IdTrain	NameFee	StStart	NameSt	StFinal	NameSt_L	NameStEng	NameStEng_L	Service	OBJECTID	NAMET	NAMEE	Prov
0	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
1	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
2	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
3	1	1	0.0	997	เร็ว	1001	กรุงเทพ	3047	บ้านพลูตา หลวง	Bangkok	Ban Plu Ta Luang	False	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
4	1	1	1.0	1049	ธรรมดा	1001	กรุงเทพ	1007	ชุมทาง บางซื่อ	Bangkok	Bang Sue Junction	True	682.0	กรุงเทพ	BANGKOK RAILWAY STATION	BAN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2393	5	2	30.0	486	ห้องถีน	4077	น้ำตก	4020	ชุมทาง หนองปลา ดุก	Nam Tok	Nong Pla Duk Junction	False	610.0	น้ำตก	NAM TOK RAILWAY STATION	

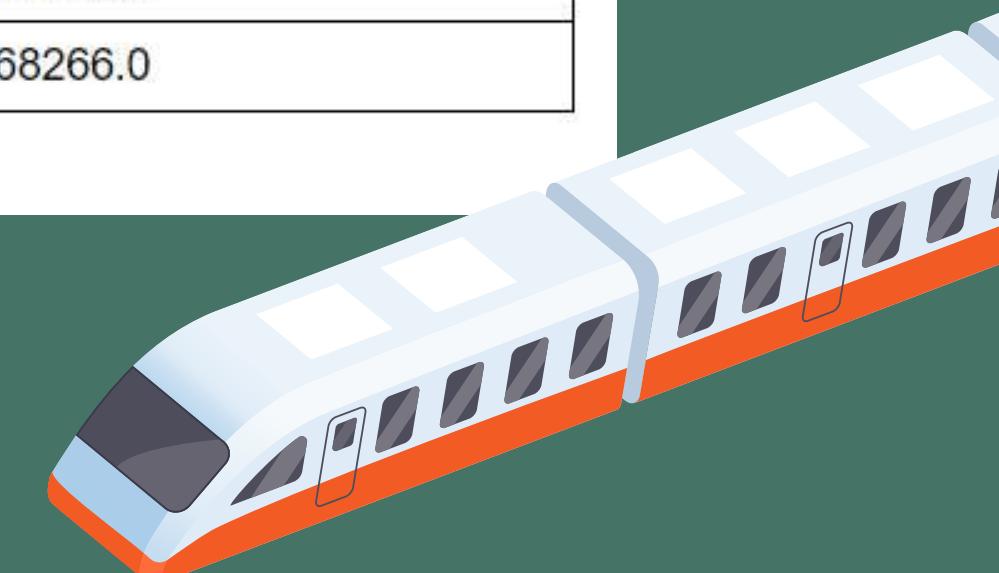


# เลือกคอลัมน์ของสถานีต้นทาง

```
table_st = table_st3[['Zone','NameFee','StStart','NameSt','StFinal','NameSt_L','X','Y','จังหวัด','ภาค','จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)']] ##เลือกคอลัมน์ที่ต้องการ
table_st ##แสดงข้อมูลของตารางที่เลือก
```

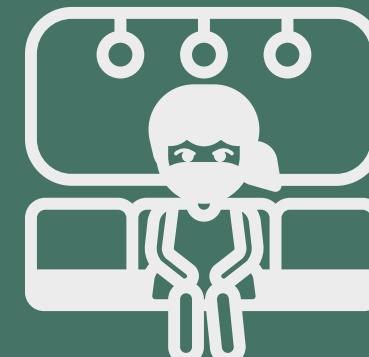
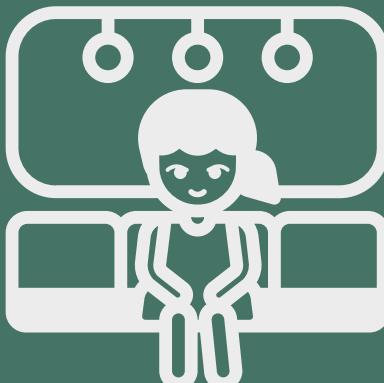
	<b>Zone</b>	<b>NameFee</b>	<b>StStart</b>	<b>NameSt</b>	<b>StFinal</b>	<b>NameSt_L</b>	<b>X</b>	<b>Y</b>	<b>จังหวัด</b>	<b>ภาค</b>	<b>จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)</b>
0	1	เริ่ว	1001	กรุงเทพ	3047	บ้านพลูตาหลวง	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
1	1	เริ่ว	1001	กรุงเทพ	3047	บ้านพลูตาหลวง	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
2	1	เริ่ว	1001	กรุงเทพ	3047	บ้านพลูตาหลวง	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
3	1	เริ่ว	1001	กรุงเทพ	3047	บ้านพลูตาหลวง	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
4	1	ธรรมชาติ	1001	กรุงเทพ	1007	ชุมทางบางซื่อ	100.516732	13.744026	กรุงเทพ	ภาคกลาง	23617914.0
...	...	...	...	...	...	...	...	...	...	...	...
2393	5	ห้องถีน	4077	น้ำตก	4020	ชุมทางหนองปลาดุก	99.067848	14.232451	กาญจนบุรี	ภาคตะวันตก	6068266.0
2394	5	ห้องถีน	4077	น้ำตก	4020	ชุมทางหนองปลาดุก	99.067848	14.232451	กาญจนบุรี	ภาคตะวันตก	6068266.0
2395	5	ห้องถีน	4077	น้ำตก	4020	ชุมทางหนองปลาดุก	99.067848	14.232451	กาญจนบุรี	ภาคตะวันตก	6068266.0
2396	5	ห้องถีน	4077	น้ำตก	4020	ชุมทางหนองปลาดุก	99.067848	14.232451	กาญจนบุรี	ภาคตะวันตก	6068266.0
2397	5	ห้องถีน	4077	น้ำตก	4020	ชุมทางหนองปลาดุก	99.067848	14.232451	กาญจนบุรี	ภาคตะวันตก	6068266.0

2398 rows × 11 columns



# ເຊື່ອມ 2 ຕາຮາງເຂົ້າດ້ວຍກັນ

	Zone	NameFee	StStart	NameSt	StFinal	NameSt_L	X_x	Y_x	ຈັງหวັດ_x	ກາດ_x	ຈຳນວນຜູ້ ເຢີມເຢືນ ທຶນໝາດ (ຄນ)_x	station_code	name	exact_km	km	class
0	1	ເຮົາ	1001	ກຣຸງເທິພ	3047	ນ້ຳນັກງານ ພລູດາ ຫລວງ	100.516732	13.744026	ກຣຸງເທິພ	ກາຄ ກລາງ	23617914.0	3047.0	ນ້ຳ ພລູດາ ຫລວງ	184.0	184.0	4.0
1	1	ເຮົາ	1001	ກຣຸງເທິພ	3047	ນ້ຳນັກງານ ພລູດາ ຫລວງ	100.516732	13.744026	ກຣຸງເທິພ	ກາຄ ກລາງ	23617914.0	3047.0	ນ້ຳ ພລູດາ ຫລວງ	184.0	184.0	4.0
2	1	ເຮົາ	1001	ກຣຸງເທິພ	3047	ນ້ຳນັກງານ ພລູດາ ຫລວງ	100.516732	13.744026	ກຣຸງເທິພ	ກາຄ ກລາງ	23617914.0	3047.0	ນ້ຳ ພລູດາ ຫລວງ	184.0	184.0	4.0
3	1	ເຮົາ	1001	ກຣຸງເທິພ	3047	ນ້ຳນັກງານ ພລູດາ ຫລວງ	100.516732	13.744026	ກຣຸງເທິພ	ກາຄ ກລາງ	23617914.0	3047.0	ນ້ຳ ພລູດາ ຫລວງ	184.0	184.0	4.0
4	1	ເຮົາ	1001	ກຣຸງເທິພ	3047	ນ້ຳນັກງານ ພລູດາ ຫລວງ	100.516732	13.744026	ກຣຸງເທິພ	ກາຄ ກລາງ	23617914.0	3047.0	ນ້ຳ ພລູດາ ຫລວງ	184.0	184.0	4.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21971	5	ທອງຄືນ	4077	ນ້ຳຕົກ	4020	ຫນອງປລາ ດຸກ	99.067848	14.232451	ການຟັນບັນຍີ	ກາຄ ຕະວັນ ຕົກ	6068266.0	4020.0	ຫນ ທາງ ຫນອງ ປລາ ດຸກ	80.0	64.0	3.0

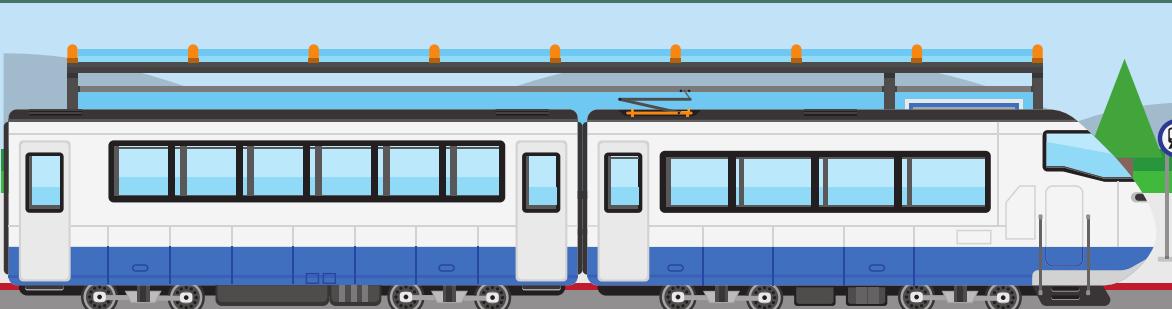


# Classification

```
table_challenge = table_mix[['class', 'NameFee', 'exact_km', 'km', 'จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_x', 'จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_y']] ### เลือกคอลัมน์ที่ต้องใช้ดอน challenge
table_challenge #เรียกดูตาราง
```

	class	NameFee	exact_km	km	จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_x	จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_y
0	4.0	เร็ว	184.0	184.0	23617914.0	6964743.0
1	4.0	เร็ว	184.0	184.0	23617914.0	6964743.0
2	4.0	เร็ว	184.0	184.0	23617914.0	6964743.0
3	4.0	เร็ว	184.0	184.0	23617914.0	6964743.0
4	4.0	เร็ว	184.0	184.0	23617914.0	6964743.0
...	...	...	...	...	...	...
21971	3.0	ห้องถีน	80.0	64.0	6068266.0	1430360.0
21972	3.0	ห้องถีน	80.0	64.0	6068266.0	1430360.0
21973	3.0	ห้องถีน	80.0	64.0	6068266.0	1430360.0
21974	3.0	ห้องถีน	80.0	64.0	6068266.0	1430360.0
21975	3.0	ห้องถีน	80.0	64.0	6068266.0	1430360.0

```
table_challenge['NameFee'] = table_challenge['class']+ " "+table_challenge['NameFee'] ### รวม column ของ class กับ NameFee
```



# Classification

```
table_challenge = table_challenge.drop('class',axis=1) ##ลบ column : class
```

table\_challenge ### เช็คตารางดูอีกรอบ

	NameFee	exact_km	km	จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_x	จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_y
0	4.0 เร็ว	184.0	184.0	23617914.0	6964743.0
1	4.0 เร็ว	184.0	184.0	23617914.0	6964743.0
2	4.0 เร็ว	184.0	184.0	23617914.0	6964743.0
3	4.0 เร็ว	184.0	184.0	23617914.0	6964743.0
4	4.0 เร็ว	184.0	184.0	23617914.0	6964743.0
...	...	...	...	...	...
21971	3.0 ห้องถีน	80.0	64.0	6068266.0	1430360.0
21972	3.0 ห้องถีน	80.0	64.0	6068266.0	1430360.0
21973	3.0 ห้องถีน	80.0	64.0	6068266.0	1430360.0
21974	3.0 ห้องถีน	80.0	64.0	6068266.0	1430360.0
21975	3.0 ห้องถีน	80.0	64.0	6068266.0	1430360.0

```
X = table_challenge[['exact_km', 'km', 'จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_x', 'จำนวนผู้เยี่ยมเยือนทั้งหมด (คน)_y']] ### กำหนด X และ y
y = table_challenge[['NameFee']]
```



# Classification Decision Tree

```
#define
mytree1 = DecisionTreeClassifier(max_depth=5,random_state=13)
#train
mytree1.fit(X_train2,y_train2)
#test
mytree1_result = mytree1.predict(X_val)
#accuracy_score
accuracy_score(y_val, mytree1_result)
```

0.7866504114599208

```
#define
mytree3 = DecisionTreeClassifier(criterion='entropy',splitter='random',random_state=13)
#train
mytree3.fit(X_train2,y_train2)
#test
mytree_result3 = mytree3.predict(X_val)
#accuracy_score
accuracy_score(y_val, mytree_result3)
```

0.8716854617494666

```
#define
mytree2 = DecisionTreeClassifier(splitter='random',random_state=13)
#train
mytree2.fit(X_train2,y_train2)
#test
mytree_result2 = mytree2.predict(X_val)
#accuracy_score
accuracy_score(y_val, mytree_result2)
```

0.8716854617494666

```
#define
mytree4 = DecisionTreeClassifier(criterion='gini',max_leaf_nodes=50,splitter='random',random_state=13)
#train
mytree4.fit(X_train2,y_train2)
#test
mytree_result4 = mytree4.predict(X_val)
#accuracy_score
accuracy_score(y_val, mytree_result4)
```

0.7647058823529411



# Classification

## KNN

### KNN 1

```
# Define
neigh1 = KNeighborsClassifier(n_neighbors=4, weights='uniform') # ถ้ามีบ้านที่ใกล้สุด 4 คนและทุกๆคนเท่ากัน
#Train
neigh1.fit(X_train2,y_train2)
#Test
KNN1_result = neigh1.predict(X_val)
accuracy_score(y_val, KNN1_result) ##ค่าความแม่นยำ
```

0.8473026516306005



# Classification

## KNN

### KNN 2

```
# Define
neigh2 = KNeighborsClassifier(n_neighbors=30, weights='distance') # ถ้ามีเพื่อนบ้าน 30 คน และเพื่อนบ้านที่ใกล้ล
#Train
neigh2.fit(X_train2,y_train2)
#Test
KNN2_result = neigh2.predict(X_val)
accuracy_score(y_val, KNN2_result) ##ค่าความแม่นยำ
```

0.8661993294727217



# Classification

## KNN

### KNN 3

```
# Define  
neigh3 = KNeighborsClassifier(n_neighbors=1) # เลือกเพื่อนบ้านที่ใกล้ที่สุดแค่คนเดียว  
# Train  
neigh3.fit(X_train2,y_train2)  
# Test  
KNN3_result = neigh3.predict(X_val)  
accuracy_score(y_val, KNN3_result) ##ค่าความแม่นยำ
```

0.8512648582749162



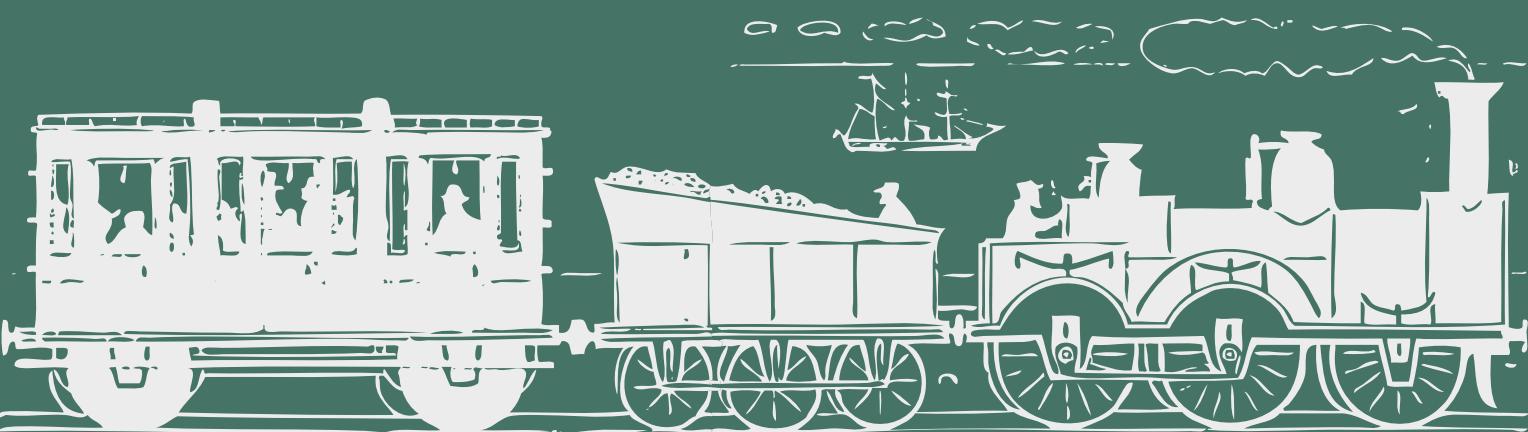
# Classification

## ANN

### ANN 1

```
#Train  
perceptron1.fit(X_train2,y_train2)  
#Test  
perceptron1_result = perceptron1.predict(X_val)  
accuracy_score(y_val,perceptron1_result) ##ค่าความแม่นยำ
```

0.44163364827796403



# Classification

## ANN

### ANN 2

```
perceptron2 = MLPClassifier(random_state=6, max_iter=2000,hidden_layer_sizes=10,learning_rate_init=0.005) #กำหนด max_iter=2000 และ Learning_rate_init=0.005
#Train
perceptron2.fit(X_train2,y_train2)
#Test
perceptron2_result = perceptron2.predict(X_val)
accuracy_score(y_val,perceptron2_result) ##ค่าความแม่นยำ
```

0.43736665650716244



# Classification

## ANN

### ANN 3

```
perceptron3 = MLPClassifier(random_state=6, max_iter=20000,hidden_layer_sizes=10,learning_rate_init=0.001) #กำหนด max_iter=20000 และ Learning_rate_init=0.001
#Train
#Train
perceptron3.fit(X_train2,y_train2)
#Test
perceptron3_result = perceptron3.predict(X_val)
accuracy_score(y_val,perceptron3_result) ##ค่าความแม่นยำ
```

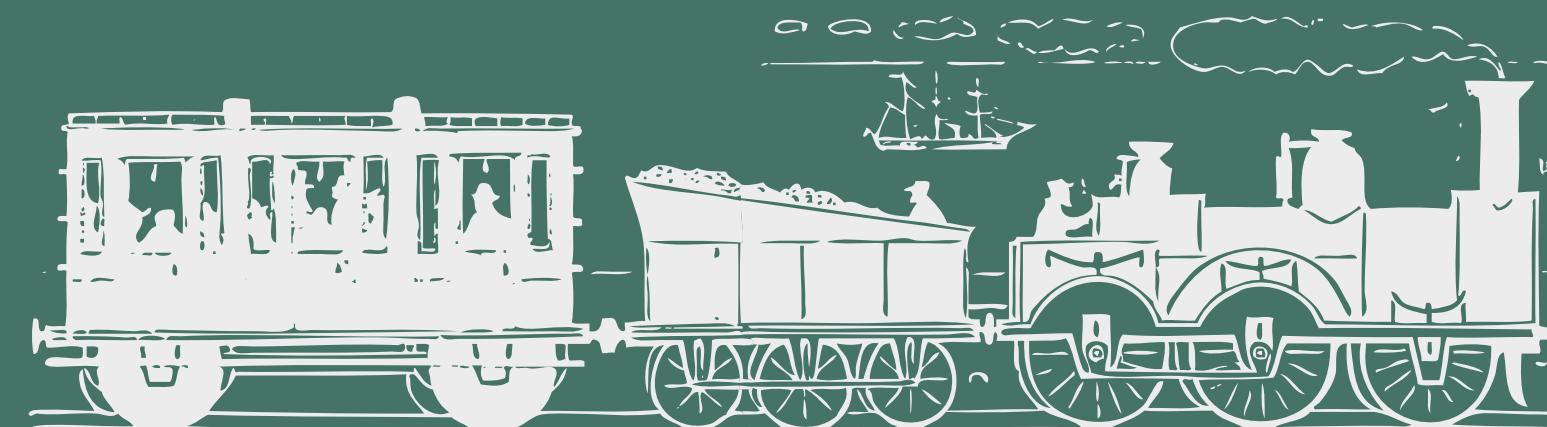
0.4535202682109113



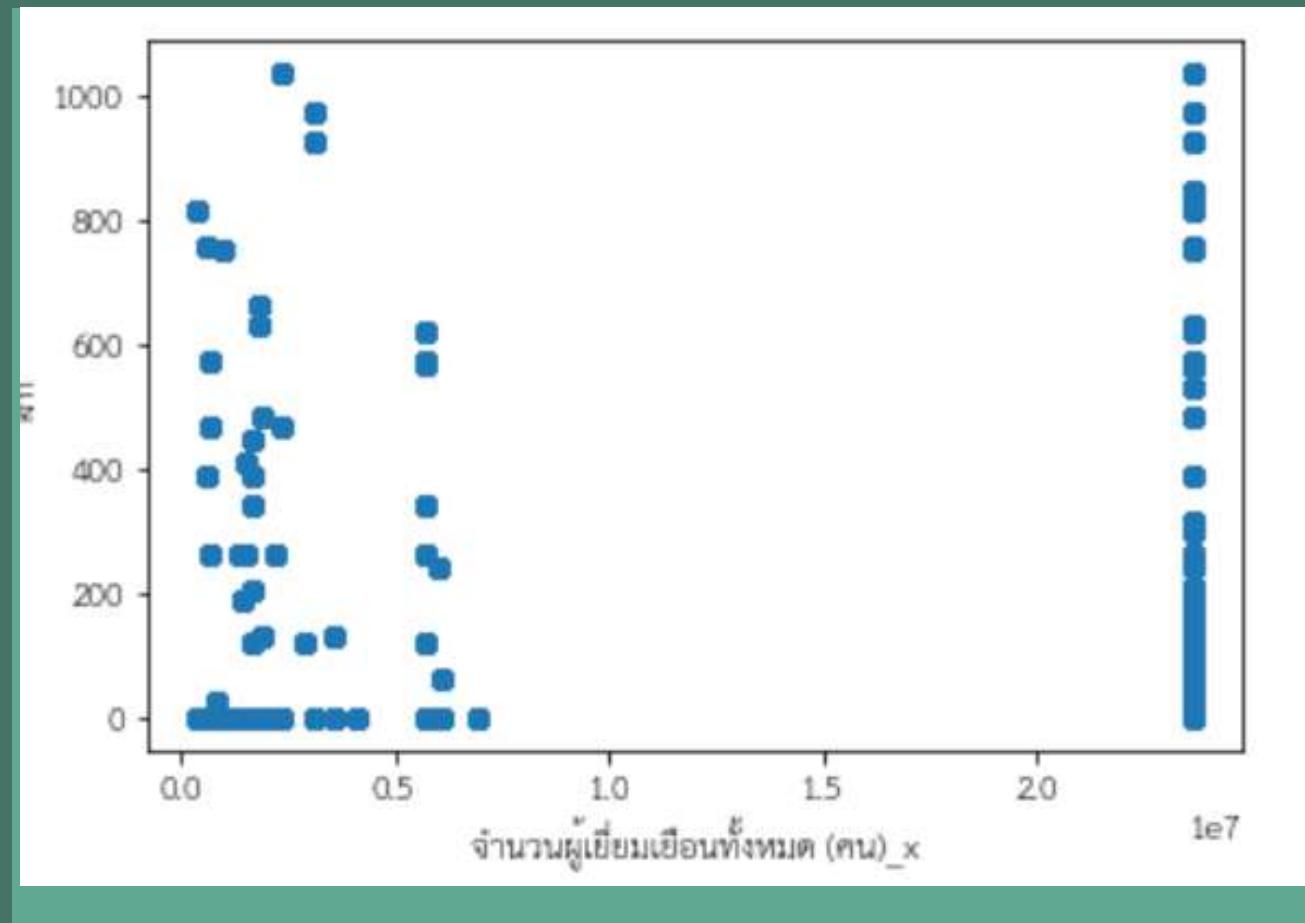
# Classification

## Retain & Evaluate

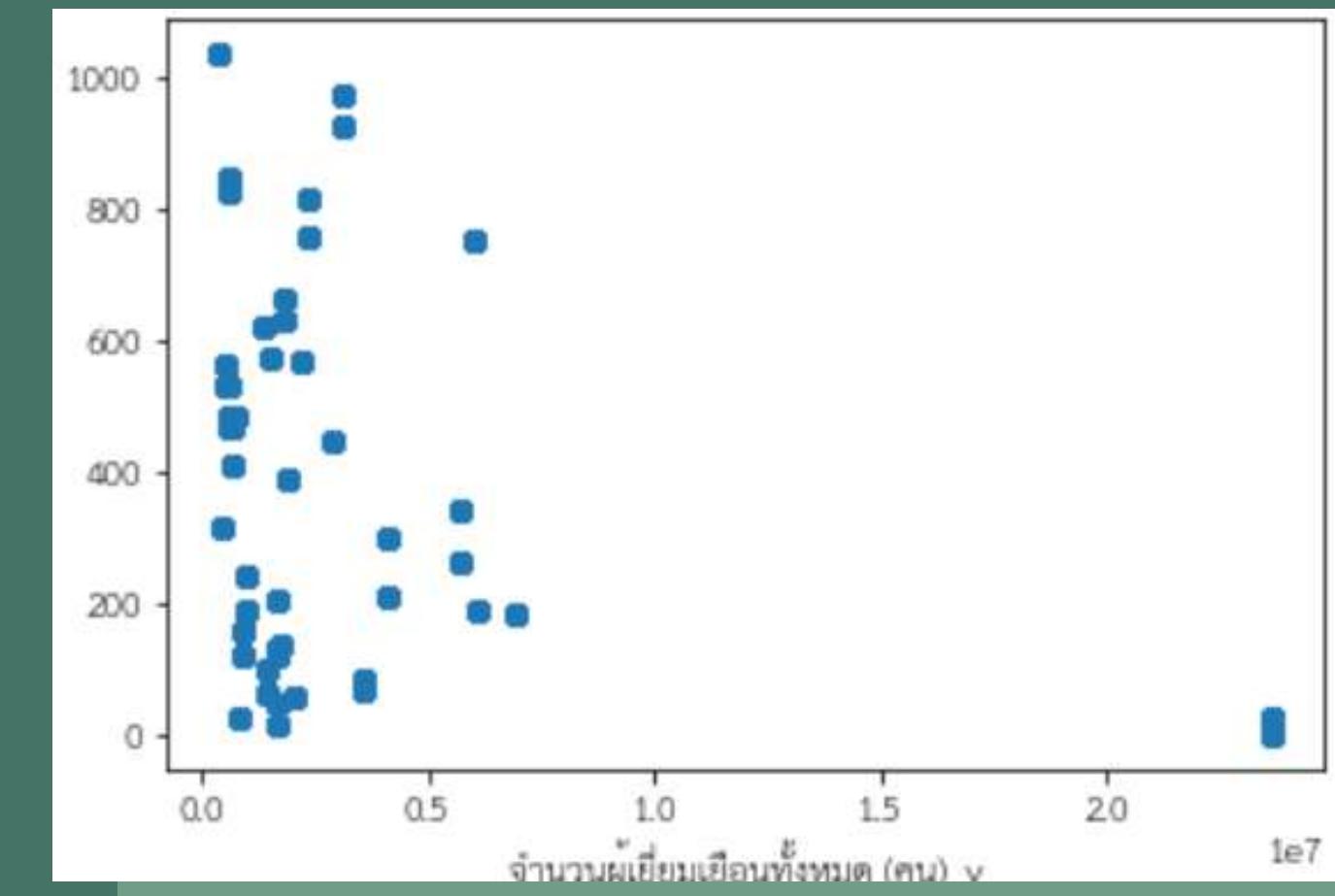
```
mytree_result_final = DecisionTreeClassifier(splitter='random',random_state=13) ## เลือก Decision Tree
#Train
mytree_result_final.fit(X_train,y_train)
#Test
mytreefinal_result = mytree.predict(X_test)
accuracy_score(y_test,mytreefinal_result) ##ค่าความแม่นยำ
0.8749885688157293
```



# Data Visualization

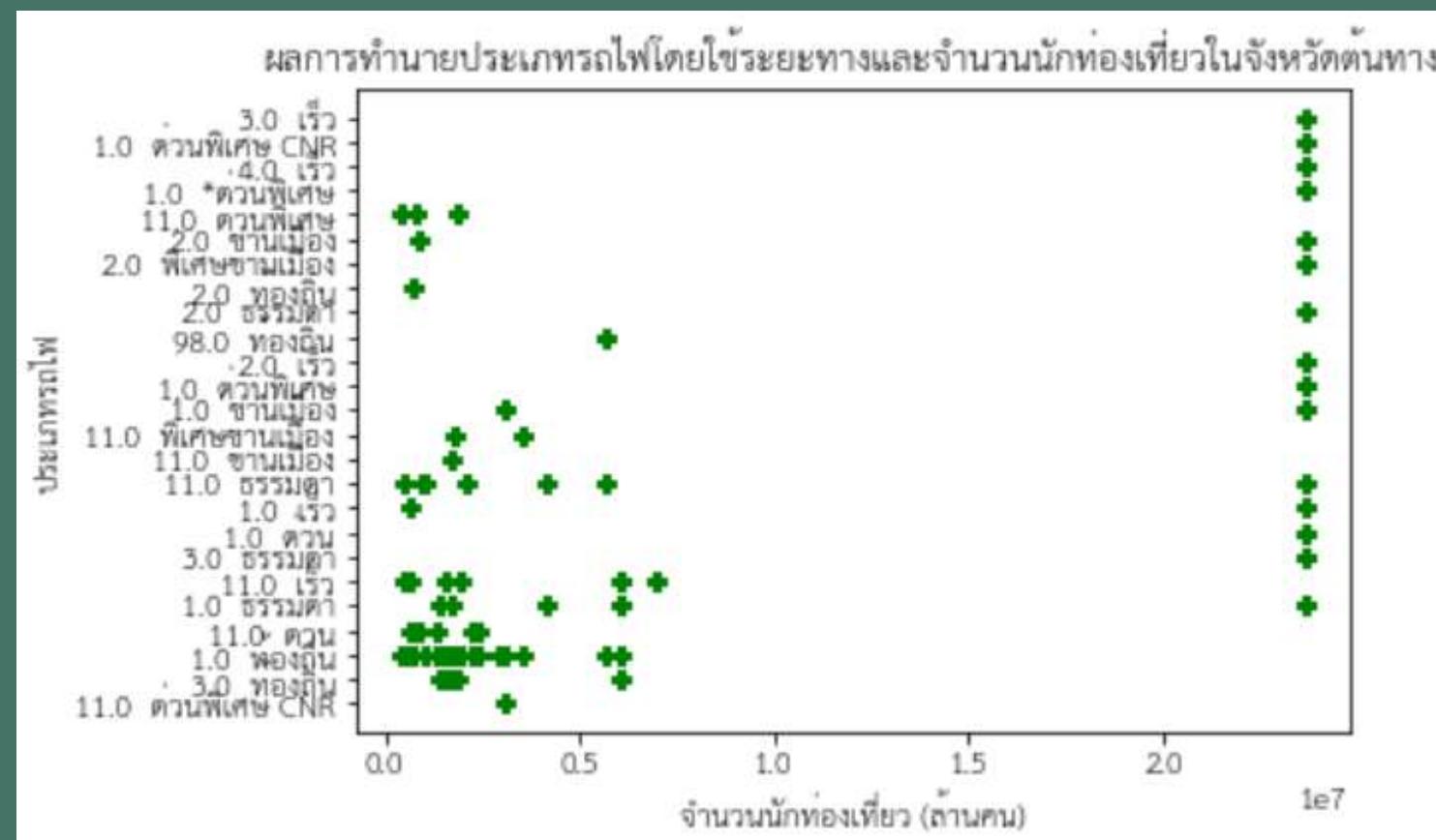


กราฟการกระจายของจำนวนนักท่อง  
เที่ยวตั้นทางกับระยะทาง

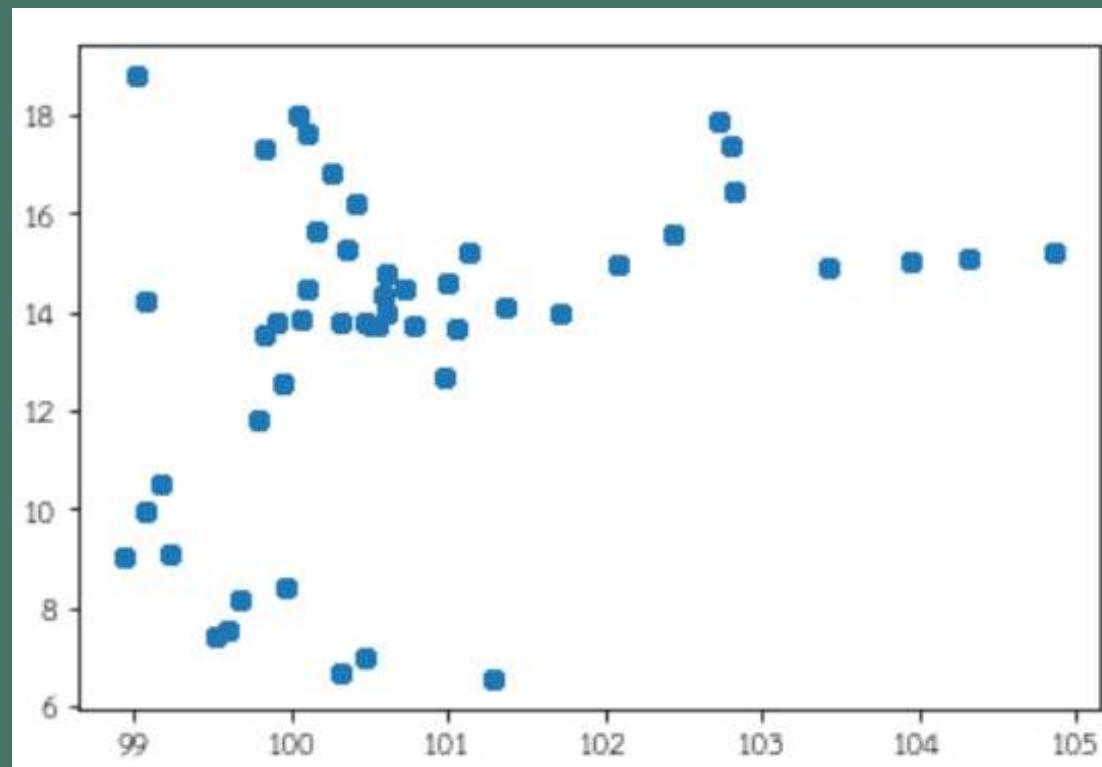


กราฟการกระจายของจำนวนนักท่อง  
เที่ยวปลายทางกับระยะทาง

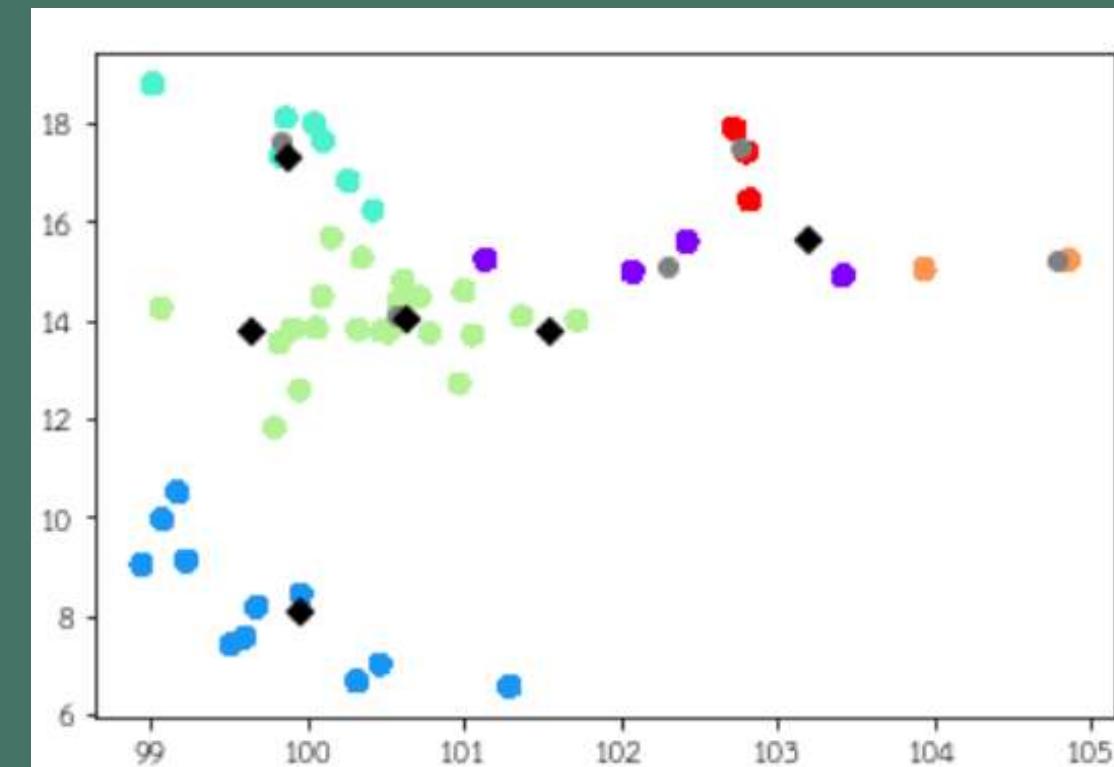
# Data Visualization



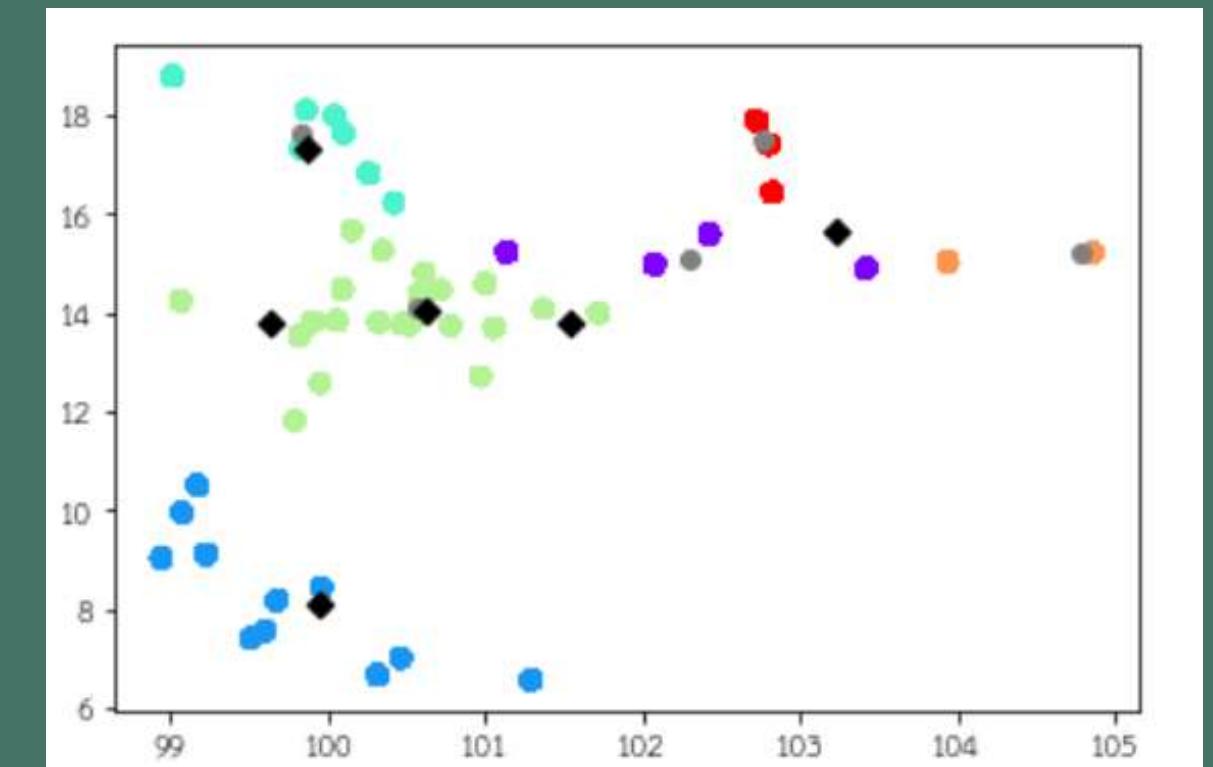
# K-means



การกระจาย



สถานีต้นทาง



สถานีปลายทาง



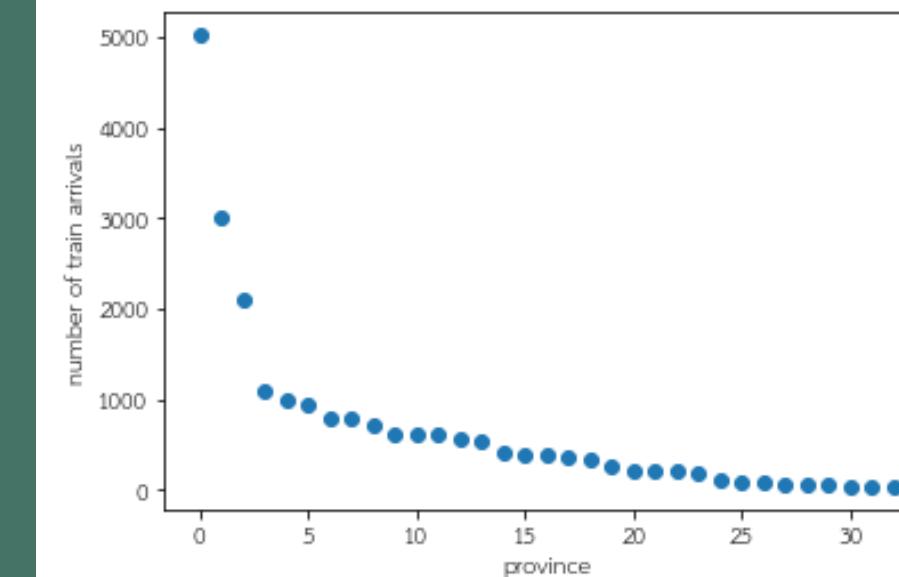
# จังหวัดที่มีรถไฟฟ้าเข้ามากที่สุด

```
data = table_mix[['จังหวัด_y','StFinal','ภาค_y']] ##เลือกคอลัมน์ที่ต้องการ
```

```
Top_Prov = data.groupby('จังหวัด_y').count().sort_values(by=['StFinal'],ascending=False)
al
Top_Prov ##เรียงกดตามจำนวน
```

	StFinal	ภาค_y
จังหวัด_y		
กรุงเทพ	5068	5068
นครราชสีมา	2855	2855
อุบลราชธานี	2176	2176
ลพบุรี	1606	1606
พิษณุโลก	980	980
สระบุรี	938	938
สุรินทร์	882	882
หนองคาย	784	784

```
plt.scatter(range(len(Top_Prov.iloc[-33:,1])),Top_Prov.iloc[-33:,1]) ## Plot กราฟ
plt.xlabel('province'); ##ใช้ชื่อแกน X
plt.ylabel('number of train arrivals'); ##ใช้ชื่อแกน Y
```



จากตารางจะเห็นได้ว่า กรุงเทพ เป็นจังหวัดที่มีรถไฟฟ้าเข้ามากที่สุด

# Thank You

