

Y
N

SC663401 Basic Programming for Data Science and Data Visualization

PROJECT

GROUP YOKENATNATYOKE

GOAL

"เปรียบเทียบความถี่ระหว่างจำนวนครั้งที่ผู้ชมเข้าชมเว็บไซต์กับจำนวนประกาศขาย/เช่า
ของที่อยู่อาศัยในพื้นที่กรุงเทพมหานครและปริมณฑล"

signs of the island being brought back to life anyway. I had

AIR MAIL

13

MONTH

MONDAY

TUES

Main Data

- 1 **Volume of Sale Listing by Location** : จำนวนประกาศขายที่อยู่อาศัยประเภท จัดสรร
แยกตามประเภทอสังหาริมทรัพย์ในพื้นที่ โดยข้อมูลมาจกประกาศขายในโลกออนไลน์
ของประเทศไทย ที่ผ่านกระบวนการจัดการข้อมูลของบาเนีย
- 2 **Volume of Rental Listing by Location** : จำนวนประกาศเช่าที่อยู่อาศัยประเภท
จัดสรร แยกตามประเภทอสังหาริมทรัพย์ในพื้นที่ โดยข้อมูลมาจกประกาศขายในโลก
ออนไลน์ของประเทศไทย ที่ผ่านกระบวนการจัดการข้อมูลของบาเนีย
- 3 **Customer Interest by Location** : จำนวนครั้งที่มีผู้เข้าชมโครงการผ่านเว็บไซต์ทั้งหมด
บนเว็บ Baania.com

Supplementary Data

- 1 **Residential Project Data** : ข้อมูลรายละเอียดโครงการที่อยู่อาศัยประเภท จัดสรร ผู้พัฒนา พื้นที่โครงการ และสิ่งอำนวยความสะดวก ของประเทศไทย จากการรวบรวมโดย บาเนีย
- 2 **Residential Unittype Data** : ข้อมูลรายละเอียดแบบบ้านในโครงการที่อยู่อาศัยประเภท จัดสรร จากการรวบรวมโดย บาเนีย
- 3 **Living Score By Location** : ข้อมูลคะแนนการเดินทางเข้าถึงสถานที่ใช้ชีวิตประจำวันของพื้นที่ต่างๆ โดยคำนวณจาก 3 รูปแบบการเดินทาง และ 5 ประเภทของสถานที่เป้าหมายเพื่อการดำเนินชีวิตในแต่ละวัน
- 4 **Eating Score By Location** : ข้อมูลคะแนนการกินของพื้นที่ต่างๆ ซึ่งวิเคราะห์จากปัจจัยด้านราคา ด้านคุณภาพร้าน/บริการ และด้านความหลากหลายของอาหารในพื้นที่ โดยจะแบ่งการวิเคราะห์ออกเป็น การกินช่วงเช้ากับการกินช่วงค่ำ จัดทำโดย Baania data team

Data Preparation


- 1 นำไฟล์ข้อมูลลงใน google drive
- 2 เชื่อม google drive กับ colab ของเรา
- 3 เรียกใช้ file ทั้งหมดที่เราต้องการใช้
- 4 ทำการ clean ข้อมูลทุกตารางเพื่อไม่ให้มีค่าสูญหาย
- 5 ทำการเชื่อมตารางทั้งหมด
- 6 เลือกแสดงแค่ข้อมูลในเขตกรุงเทพฯและปริมณฑล 5 จังหวัด ได้แก่ กรุงเทพฯ, สมุทรปราการ, นนทบุรี, ปทุมธานี และ นครปฐม

Problem



Problem 1

เกิดคอลัมน์ใหม่และค่าสูญหายจำนวนมากจากการรวมตารางโดยใช้ Foreign key คือ 'province_name_en', 'district_name_en', propertytype_name_en' เชื่อมกับ 'listing_province_name_en', 'listing_district_name_en', 'project_propertytype_name_en'

(49299, 63)																										
	project_id	name_en	propertytype_id	propertytype_name_en	price_min	subdistrict_id	subdistrict_name_en	neighborhood_id	neighborhood_name_en	latitude	...	access_bus	access_train	acccme_en	latitude	...	access_bus	access_train	access_total	drive_total	interest	volume_of_project	listing_province_name_en	listing_district_name_en	project_propertytype_name_en	volume_of_sale
0	project_4270	Paholyothin Park Condominium	2	Condo	3650000.0	3912.0	SAM SEN NAI	B1161	THOTHOBO	13.786388	...	100.0	44.0	HOB0	13.786388	...	100.0	44.0	89.0	98.0	4084.0	1103.0	NaN	NaN	NaN	NaN
1	project_4270	Paholyothin Park Condominium	2	Condo	3650000.0	3912.0	SAM SEN NAI	B1161	THOTHOBO	13.786388	...	100.0	44.0	HOB0	13.786388	...	100.0	44.0	89.0	98.0	4084.0	1103.0	NaN	NaN	NaN	NaN
2	project_9490	Diya Valley Sriracha	1	Detached House	2390000.0	4701.0	BANG PHRA	B0752	NA PHRAO	13.173645	...	0.0	0.0	HRAO	13.173645	...	0.0	0.0	0.0	78.0	0.0	0.0	NaN	NaN	NaN	NaN
3	project_9490	Diya Valley Sriracha	1	Detached House	2390000.0	4701.0	BANG PHRA	B0752	NA PHRAO	13.173645	...	0.0	0.0	HRAO	13.173645	...	0.0	0.0	0.0	78.0	0.0	0.0	NaN	NaN	NaN	NaN
4	project_9490	Diya Valley Sriracha	1	Detached House	2390000.0	4701.0	BANG PHRA	B0752	NA PHRAO	13.173645	...	0.0	0.0	HRAO	13.173645	...	0.0	0.0	0.0	78.0	0.0	0.0	NaN	NaN	NaN	NaN
5 rows x 63 columns																										
																										

Cause of the Problem

ชื่อในคอลัมน์ของตาราง table_merge4 และ ชื่อในคอลัมน์ของตาราง opendata_volume_listing_clean ที่ใช้เป็น Foreign key เป็นคนละชื่อกัน แต่มีความหมายเดียวกัน

```
[110] 1 table_merge4.columns
```

```
Index(['project_id', 'name_en', 'propertytype_id', 'propertytype_name_en',  
      'price_min', 'subdistrict_id', 'subdistrict_name_en', 'neighborhood_id',  
      'neighborhood_name_en', 'latitude', 'longitude', 'district_id',  
      'district_name_en', 'province_id', 'province_name_en', 'count_elevator',  
      'count_elevator_service', 'count_floor', 'count_parking',  
      'facility_clubhouse', 'facility_fitness', 'facility_meeting',  
      'facility_park', 'facility_playground', 'facility_pool',  
      'facility_security', 'unittype_id', 'name_th', 'area_total_min_wa',  
      'area_usable_min', 'count_airconditioner', 'count_elevator_unittype',  
      'count_floor_internal', 'count_parking_unittype', 'count_room_bath',  
      'count_room_bed', 'count_room_dinning', 'count_room_guest',  
      'count_room_kitchen', 'count_room_living', 'count_room_maid',  
      'count_room_storage', 'count_room_utility', 'eating_daytime',  
      'eating_daytime_price', 'eating_daytime_quality',  
      'eating_daytime_variety', 'eating_nighttime', 'eating_nighttime_price',  
      'eating_nighttime_quality', 'eating_nighttime_variety', 'walk_total',  
      'access_boat', 'access_bus', 'access_train', 'access_total',  
      'drive_total', 'interest', 'volume_of_project'],  
      dtype='object')
```

```
[111] 1 opendata_volume_listing_clean.columns
```

```
Index(['listing_province_name_en', 'listing_district_name_en',  
      'project_propertytype_name_en', 'volume_of_sale'],  
      dtype='object')
```


Cause of the Problem

ชื่อในคอลัมน์ของตาราง table_merge5 และ ชื่อในคอลัมน์ของตาราง opendata_volume_listing_rent_clean ที่ใช้เป็น Foreign key เป็นคนละชื่อกัน แต่มีความหมายเดียวกัน

```
[121] 1 table_merge5.columns
```

```
Index(['project_id', 'name_en', 'propertytype_id', 'propertytype_name_en',  
      'price_min', 'subdistrict_id', 'subdistrict_name_en', 'neighborhood_id',  
      'neighborhood_name_en', 'latitude', 'longitude', 'district_id',  
      'district_name_en', 'province_id', 'province_name_en', 'count_elevator',  
      'count_elevator_service', 'count_floor', 'count_parking',  
      'facility_clubhouse', 'facility_fitness', 'facility_meeting',  
      'facility_park', 'facility_playground', 'facility_pool',  
      'facility_security', 'unittype_id', 'name_th', 'area_total_min_wa',  
      'area_usable_min', 'count_airconditioner', 'count_elevator_unittype',  
      'count_floor_internal', 'count_parking_unittype', 'count_room_bath',  
      'count_room_bed', 'count_room_dinning', 'count_room_guest',  
      'count_room_kitchen', 'count_room_living', 'count_room_maid',  
      'count_room_storage', 'count_room_utility', 'eating_daytime',  
      'eating_daytime_price', 'eating_daytime_quality',  
      'eating_daytime_variety', 'eating_nighttime', 'eating_nighttime_price',  
      'eating_nighttime_quality', 'eating_nighttime_variety', 'walk_total',  
      'access_boat', 'access_bus', 'access_train', 'access_total',  
      'drive_total', 'interest', 'volume_of_project', 'volume_of_sale'],  
      dtype='object')
```



```
1 opendata_volume_listing_rent_clean.columns
```

```
Index(['listing_province_name_en', 'listing_district_name_en',  
      'project_propertytype_name_en', 'volume_of_rental'],  
      dtype='object')
```


Solution

แก้ปัญหาโดยการเปลี่ยนชื่อคอลัมน์ของตาราง `opendata_volume_listing_clean` ให้เหมือนกันกับตาราง `table_merge4`

	<code>listing_province_name_en</code>	<code>listing_district_name_en</code>	<code>project_propertytype_name_en</code>	<code>volume_of_sale</code>
0	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	643
1	BANGKOK	PHRA NAKHON	Twin house	19
2	BANGKOK	WANG THONGLANG	Condo	279



	<code>province_name_en</code>	<code>district_name_en</code>	<code>propertytype_name_en</code>	<code>volume_of_sale</code>
0	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	643
1	BANGKOK	PHRA NAKHON	Twin house	19
2	BANGKOK	WANG THONGLANG	Condo	279

Solution

แก้ปัญหาโดยการเปลี่ยนชื่อคอลัมน์ของตาราง `opendata_volume_listing_rent_clean` ให้เหมือนกับตาราง `table_merge5`

	<code>listing_province_name_en</code>	<code>listing_district_name_en</code>	<code>project_propertytype_name_en</code>	<code>volume_of_rental</code>
0	SAMUT PRAKAN	BANG PHLI	Townhome	18
1	BANGKOK	BANGKOK NOI	Condo	32
2	BANGKOK	YAN NAWA	Townhome	11



	<code>province_name_en</code>	<code>district_name_en</code>	<code>propertytype_name_en</code>	<code>volume_of_rental</code>
0	SAMUT PRAKAN	BANG PHLI	Townhome	18
1	BANGKOK	BANGKOK NOI	Condo	32
2	BANGKOK	YAN NAWA	Townhome	11

Problem 2

จากผลลัพธ์ที่ได้ของตาราง `table_merge_6` ในคอลัมน์ `volume_of_sale` และ `volume_of_rental` จะเห็นว่าเป็นค่าสูญหายทั้งหมด

name_en	neighborhood_id	neighborhood_name_en	latitude	...	walk_total	access_boat	access_bus	access_train	access_total	drive_total	interest	volume_of_project	volume_of_sale	volume_of_rental
EN NAI	B1161	THOTHOBO	13.786388	...	98.0	100.0	100.0	44.0	89.0	98.0	4084.0	1103.0	NaN	NaN
EN NAI	B1161	THOTHOBO	13.786388	...	98.0	100.0	100.0	44.0	89.0	98.0	4084.0	1103.0	NaN	NaN
PHRA	B0752	NA PHRAO	13.173645	...	22.0	0.0	0.0	0.0	0.0	78.0	0.0	0.0	NaN	NaN
PHRA	B0752	NA PHRAO	13.173645	...	22.0	0.0	0.0	0.0	0.0	78.0	0.0	0.0	NaN	NaN
PHRA	B0752	NA PHRAO	13.173645	...	22.0	0.0	0.0	0.0	0.0	78.0	0.0	0.0	NaN	NaN
...
CHIPAT	B0967	FUTURE PARK RUNGSIT	13.993865	...	95.0	0.0	100.0	0.0	40.0	100.0	13995.0	1260.0	NaN	NaN
CHIPAT	B0967	FUTURE PARK RUNGSIT	13.993865	...	95.0	0.0	100.0	0.0	40.0	100.0	13995.0	1260.0	NaN	NaN
CHIPAT	B0967	FUTURE PARK RUNGSIT	13.993865	...	95.0	0.0	100.0	0.0	40.0	100.0	13995.0	1260.0	NaN	NaN
UANG	B1049	ON NUT	13.707605	...	93.0	87.0	92.0	39.0	80.0	88.0	2263.0	602.0	NaN	NaN
UANG	B1049	ON NUT	13.707605	...	93.0	87.0	92.0	39.0	80.0	88.0	2263.0	602.0	NaN	NaN

Cause of the Problem

เนื่องจากค่าในคอลัมน์ province name_en ในตาราง table_merge4 และตาราง opendata_volume_listing_clean มีรูปแบบไม่เหมือนกัน ทำให้ผลลัพธ์ที่ได้ออกมาในคอลัมน์ volume_of_sale เป็น ค่าสูญหาย

```
[116] 1 table_merge4.province_name_en.unique()
```

```
array(['Bangkok', 'Chon Buri', 'Phuket', 'Chachoengsao', 'Khon Kaen',  
      'Chiang Mai', 'Samut Prakan', 'Nonthaburi', 'Pathum Thani',  
      'Phetchaburi', 'Nakhon Ratchasima', 'Songkhla', 'Rayong',  
      'Nakhon Pathom', 'Udon Thani', 'Prachuap Khiri Khan',  
      'Nakhon Si Thammarat', 'Chiang Rai'], dtype=object)
```



```
1 opendata_volume_listing_clean.province_name_en.unique()
```

```
array(['SAMUT PRAKAN', 'BANGKOK', 'PATHUM THANI', 'NONTABURI',  
      'NAKHON PATHOM'], dtype=object)
```


Cause of the Problem

เนื่องจากค่าในคอลัมน์ province_name_en ในตาราง table_merge5 และตาราง opendata_volume_listing_rent_clean มีรูปแบบไม่เหมือนกัน ทำให้ผลลัพธ์ที่ได้ออกมาในคอลัมน์ volume_of_rental เป็น ค่าสูญหาย

```
[128] 1 table_merge5.province_name_en.unique()
```

```
array(['Bangkok', 'Chon Buri', 'Phuket', 'Chachoengsao', 'Khon Kaen',  
      'Chiang Mai', 'Samut Prakan', 'Nonthaburi', 'Pathum Thani',  
      'Phetchaburi', 'Nakhon Ratchasima', 'Songkhla', 'Rayong',  
      'Nakhon Pathom', 'Udon Thani', 'Prachuap Khiri Khan',  
      'Nakhon Si Thammarat', 'Chiang Rai'], dtype=object)
```



```
1 opendata_volume_listing_rent_clean.province_name_en.unique()
```

```
array(['SAMUT PRAKAN', 'BANGKOK', 'NONTABURI', 'PATHUM THANI',  
      'NAKHON PATHOM'], dtype=object)
```


Solution

แก้ไขโดยเปลี่ยนชื่อจังหวัดในคอลัมน์ `province_name_en` ของทั้ง 2 ตาราง ให้เหมือนกับตาราง `table_merge5` และ 6 ตามลำดับ

```
1 opendata_volume_listing_clean.loc[opendata_volume_listing_clean['province_name_en'] == 'SAMUT PRAKAN','province_name_en'] = 'Samut Prakan'
2 opendata_volume_listing_clean.loc[opendata_volume_listing_clean['province_name_en'] == 'BANGKOK','province_name_en'] = 'Bangkok'
3 opendata_volume_listing_clean.loc[opendata_volume_listing_clean['province_name_en'] == 'PATHUM THANI','province_name_en'] = 'Pathum Thani'
4 opendata_volume_listing_clean.loc[opendata_volume_listing_clean['province_name_en'] == 'NONTABURI','province_name_en'] = 'Nonthaburi'
5 opendata_volume_listing_clean.loc[opendata_volume_listing_clean['province_name_en'] == 'NAKHON PATHOM','province_name_en'] = 'Nakhon Pathom'
```

ตาราง `opendata_volume_listing_clean`

```
1 opendata_volume_listing_rent_clean.loc[opendata_volume_listing_rent_clean['province_name_en'] == 'SAMUT PRAKAN','province_name_en'] = 'Samut Prakan'
2 opendata_volume_listing_rent_clean.loc[opendata_volume_listing_rent_clean['province_name_en'] == 'BANGKOK','province_name_en'] = 'Bangkok'
3 opendata_volume_listing_rent_clean.loc[opendata_volume_listing_rent_clean['province_name_en'] == 'PATHUM THANI','province_name_en'] = 'Pathum Thani'
4 opendata_volume_listing_rent_clean.loc[opendata_volume_listing_rent_clean['province_name_en'] == 'NONTABURI','province_name_en'] = 'Nonthaburi'
5 opendata_volume_listing_rent_clean.loc[opendata_volume_listing_rent_clean['province_name_en'] == 'NAKHON PATHOM','province_name_en'] = 'Nakhon Pathom'
```

ตาราง `opendata_volume_listing_rent_clean`

Problem

ข้อมูลมีจำนวนมากหลังจากทำการ merge ตาราง ทำให้โปรแกรมไม่สามารถประมวลผลได้



```
1 print(table_merge4.shape)
```

```
(313777, 59)
```



```
1 print(table_merge5.shape)
```

```
(5198242, 60)
```

```
[ ] 1 # merge
    2 table_merge6 = table_merge5.merge(opendata_volume_listing_re
    3 print(table_merge6.shape)
```

```
1 print(table_merge6.shape)
```

Missing/Null Checking

เซสชันของคุณหลังจากใช้ RAM ที่มีอยู่ทั้งหมด หาก
คุณสนใจที่จะเข้าถึงรันไทม์ที่ใช้ RAM สูง คุณอาจ
ต้องการลองใช้ [Colab Pro](#)

[ดูบันทึกการรันไทม์](#)

Problem

ผลลัพธ์ที่ได้จากการรวมตาราง table_merge5 จะเห็นว่ามีแถวที่ค่าในตารางซ้ำกันเยอะ ยกเว้นคอลัมน์ interest ,volume_of_project และ volume_of_sale

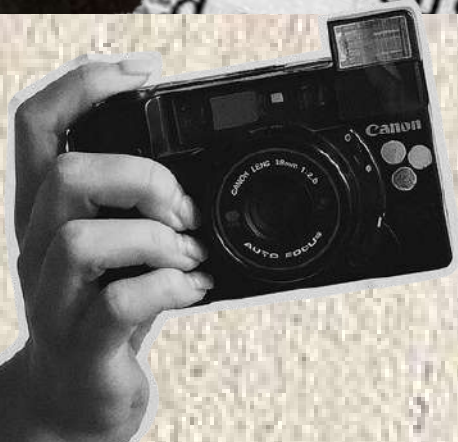
	project_id	name_en	propertytype_id	propertytype_name_en	price_min	subdistrict_id	subdistrict_name_en	neighborhood_id	neighborhood_name_en	latitude	...	walk_total	access_boat
0	project_4270	Paholyothin Park Condominium	2	Condo	3650000.0	3912.0	SAM SEN NAI	B1161	THOTHOBO	13.786388	...	98.0	100.0
1	project_4270	Paholyothin Park Condominium	2	Condo	3650000.0	3912.0	SAM SEN NAI	B1161	THOTHOBO	13.786388	...	98.0	100.0
2	project_4270	Paholyothin Park Condominium	2	Condo	3650000.0	3912.0	SAM SEN NAI	B1161	THOTHOBO	13.786388	...	98.0	100.0
3	project_4270	Paholyothin Park Condominium	2	Condo	3650000.0	3912.0	SAM SEN NAI	B1161	THOTHOBO	13.786388	...	98.0	100.0
4	project_4270	Paholyothin Park Condominium	2	Condo	3650000.0	3912.0	SAM SEN NAI	B1161	THOTHOBO	13.786388	...	98.0	100.0
...
11808614	project_7949	The Tropical Condominium	2	Condo	999000.0	3970.0	SUAN LUANG	B1049	ON NUT	13.707605	...	93.0	87.0
11808615	project_7949	The Tropical Condominium	2	Condo	999000.0	3970.0	SUAN LUANG	B1049	ON NUT	13.707605	...	93.0	87.0
11808616	project_7949	The Tropical Condominium	2	Condo	999000.0	3970.0	SUAN LUANG	B1049	ON NUT	13.707605	...	93.0	87.0
11808617	project_7949	The Tropical Condominium	2	Condo	999000.0	3970.0	SUAN LUANG	B1049	ON NUT	13.707605	...	93.0	87.0
11808618	project_7949	The Tropical Condominium	2	Condo	999000.0	3970.0	SUAN LUANG	B1049	ON NUT	13.707605	...	93.0	87.0

11808619 rows x 61 columns

Cause of the Problem

ตาราง opendata interest_clean มีการเก็บ transction_year และ transction_month คือเก็บข้อมูลจำนวนครั้งที่มีผู้เข้าชมโครงการบนเว็บไซต์และจำนวนโครงการที่มีขายในพื้นที่ ตลอดทุกเดือน จึงทำให้ข้อมูลเกิดความซ้ำซ้อนกันตอนรวมตาราง

	transction_year	transction_month	province_name_en	district_name_en	propertytype_name_en	volume_of_project	interest
35	2019	4	Bangkok	KHLONG TOEI	Condo	45	159
89	2019	7	Bangkok	KHLONG TOEI	Condo	89	318
932	2019	12	Bangkok	KHLONG TOEI	Condo	18	36
1224	2019	5	Bangkok	KHLONG TOEI	Condo	105	263
1302	2019	9	Bangkok	KHLONG TOEI	Condo	142	482
1373	2019	2	Bangkok	KHLONG TOEI	Condo	322	623
1377	2019	11	Bangkok	KHLONG TOEI	Condo	120	217
1514	2019	1	Bangkok	KHLONG TOEI	Condo	54	99
1703	2019	6	Bangkok	KHLONG TOEI	Condo	88	195
20	2020	6	Bangkok	KHLONG TOEI	Condo	129	174
418	2020	7	Bangkok	KHLONG TOEI	Condo	13	28
627	2020	1	Bangkok	KHLONG TOEI	Condo	225	443
758	2020	3	Bangkok	KHLONG TOEI	Condo	72	245
1011	2020	5	Bangkok	KHLONG TOEI	Condo	92	172
1550	2020	2	Bangkok	KHLONG TOEI	Condo	27	53
327	2021	10	Bangkok	KHLONG TOEI	Condo	5	14



Cause of the Problem

3

ตาราง opendata_volume_listing_clean มีการเก็บ transction_year และ transction_month คือเก็บข้อมูลจำนวนประกาศขาย ตลอดทุกเดือน จึงทำให้ข้อมูลเกิดความซ้ำซ้อนกันตอนรวมตาราง

	transction_year	transction_month	province_name_en	district_name_en	propertytype_name_en	volume_of_sale
1365	2019	8	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	128
1844	2019	1	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	38
199	2019	5	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	65
1338	2019	9	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	186
292	2019	11	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	860
1782	2019	10	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	284
1718	2019	2	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	46
561	2019	4	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	73
1678	2019	12	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	54
1269	2019	6	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	147
915	2019	3	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	117
863	2019	7	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	70
1714	2020	5	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	44
1579	2020	1	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	127
2059	2020	6	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	37
1559	2020	3	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	64
1542	2020	8	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	110
637	2020	2	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	63
584	2020	10	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	106
571	2020	12	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	603
241	2020	9	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	69
13	2020	7	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	52
2043	2020	11	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	134
797	2020	4	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	55
1685	2021	8	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	687
1625	2021	4	SAMUT PRAKAN	MUEANG SAMUT PRAKAN	Condo	299

Cause of the Problem

ตาราง opendata_volume_listing_rent_clean มีการเก็บ transction_year และ transction_month คือเก็บข้อมูลจำนวนประกาศเช่า ตลอดทุกเดือน จึงทำให้ข้อมูลเกิดความซ้ำซ้อนกันตอนรวมตาราง

	transction_year	transction_month	province_name_en	district_name_en	propertytype_name_en	volume_of_rental
93	2019	9	SAMUT PRAKAN	BANG PHLI	Townhome	11
1541	2019	11	SAMUT PRAKAN	BANG PHLI	Townhome	19
357	2020	9	SAMUT PRAKAN	BANG PHLI	Townhome	13
933	2020	8	SAMUT PRAKAN	BANG PHLI	Townhome	12
0	2021	3	SAMUT PRAKAN	BANG PHLI	Townhome	18
1369	2021	4	SAMUT PRAKAN	BANG PHLI	Townhome	28
1111	2021	1	SAMUT PRAKAN	BANG PHLI	Townhome	32
717	2021	9	SAMUT PRAKAN	BANG PHLI	Townhome	29
525	2021	12	SAMUT PRAKAN	BANG PHLI	Townhome	24
505	2021	10	SAMUT PRAKAN	BANG PHLI	Townhome	16
433	2021	2	SAMUT PRAKAN	BANG PHLI	Townhome	22
361	2021	11	SAMUT PRAKAN	BANG PHLI	Townhome	70
359	2021	7	SAMUT PRAKAN	BANG PHLI	Townhome	12
119	2021	5	SAMUT PRAKAN	BANG PHLI	Townhome	18
1664	2021	8	SAMUT PRAKAN	BANG PHLI	Townhome	25
1900	2021	6	SAMUT PRAKAN	BANG PHLI	Townhome	21
801	2022	1	SAMUT PRAKAN	BANG PHLI	Townhome	29

Solution

ตาราง `opendata_interest_clean` จะทำการจัดกลุ่มโดย คอลัมน์ `'province_name_en'`, `'district_name_en'` และ `'propertytype_name_en'` แล้วหาผลรวมของคอลัมน์ `interest` และ `volume_of_project`

```
1 # ทำการจัดกลุ่มโดย 'province_name_en','district_name_en','propertytype_name_en' // เพื่อที่จะให้ key ที่เราใช้ในการเชื่อมตาราง มีผลลัพธ์ ที่เป็นเอกลักษณ์เฉพาะ เพื่อลดความซ้ำซ้อนของข้อมูล
2 opendata_interest_clean = opendata_interest_clean.groupby(['province_name_en','district_name_en','propertytype_name_en'])[['interest','volume_of_project']].sum()
3 opendata_interest_clean = opendata_interest_clean.reset_index()
4 opendata_interest_clean
```

	province_name_en	district_name_en	propertytype_name_en	interest	volume_of_project
0	Bangkok	BANG BON	Condo	239	40
1	Bangkok	BANG BON	Detached House	1447	162
2	Bangkok	BANG BON	Home office	113	40
3	Bangkok	BANG BON	Townhome	1831	319
4	Bangkok	BANG BON	Twin house	23	11
...
311	Samut Prakan	PHRA SAMUT CHEDI	Commercial	3787	88
312	Samut Prakan	PHRA SAMUT CHEDI	Condo	71	37
313	Samut Prakan	PHRA SAMUT CHEDI	Detached House	6254	570
314	Samut Prakan	PHRA SAMUT CHEDI	Townhome	30158	1749
315	Samut Prakan	PHRA SAMUT CHEDI	Twin house	10879	646

316 rows × 5 columns

Solution

ตาราง `opendata_volume_listing_clean` จะทำการจัดกลุ่มโดย คอลัมน์ `'province_name_en'`, `'district_name_en'` และ `'propertytype_name_en'` แล้วหาผลรวมของคอลัมน์ `volume_of_sale`

```
[148] 1 opendata_volume_listing_clean = opendata_volume_listing_clean.groupby(['province_name_en', 'district_name_en', 'propertytype_name_en'])['volume_of_sale'].sum()  
      2 opendata_volume_listing_clean = opendata_volume_listing_clean.reset_index()  
      3 opendata_volume_listing_clean #รวมการให้คะแนนในจังหวัดกรุงเทพ
```

	province_name_en	district_name_en	propertytype_name_en	volume_of_sale
0	Bangkok	BANG BON	Condo	12
1	Bangkok	BANG KAPI	Condo	8519
2	Bangkok	BANG KAPI	Townhome	91
3	Bangkok	BANG KHAE	Condo	1942
4	Bangkok	BANG KHAE	Townhome	39
...
122	Samut Prakan	MUEANG SAMUT PRAKAN	Detached House	399
123	Samut Prakan	MUEANG SAMUT PRAKAN	Townhome	385
124	Samut Prakan	PHRA PRADAENG	Condo	315
125	Samut Prakan	PHRA SAMUT CHEDI	Condo	166
126	Samut Prakan	PHRA SAMUT CHEDI	Townhome	16

127 rows × 4 columns

Solution

ตาราง `opendata_volume_listing_rent_clean` จะทำการจัดกลุ่มโดย คอลัมน์ `province_name_en`, `'district_name_en'` และ `'propertytype_name_en'` แล้วหาผลรวมของคอลัมน์ `volume_of_rental`

```
[163] 1 opendata_volume_listing_rent_clean = opendata_volume_listing_rent_clean.groupby(['province_name_en', 'district_name_en', 'propertytype_name_en'])['volume_of_rental'].sum()
      2 opendata_volume_listing_rent_clean = opendata_volume_listing_rent_clean.reset_index()
      3 opendata_volume_listing_rent_clean #รวมการให้คะแนนในจังหวัดกรุงเทพ
```

	province_name_en	district_name_en	propertytype_name_en	volume_of_rental
0	Bangkok	BANG KAPI	Condo	5251
1	Bangkok	BANG KAPI	Detached House	34
2	Bangkok	BANG KAPI	Townhome	146
3	Bangkok	BANG KHAE	Condo	1493
4	Bangkok	BANG KHEN	Condo	2855
...
87	Samut Prakan	BANG PHLI	Detached House	109
88	Samut Prakan	BANG PHLI	Townhome	399
89	Samut Prakan	BANG SAO THONG	Condo	417
90	Samut Prakan	MUEANG SAMUT PRAKAN	Condo	14499
91	Samut Prakan	PHRA PRADAENG	Condo	585

92 rows × 4 columns

Draft of Dashboard

Score card

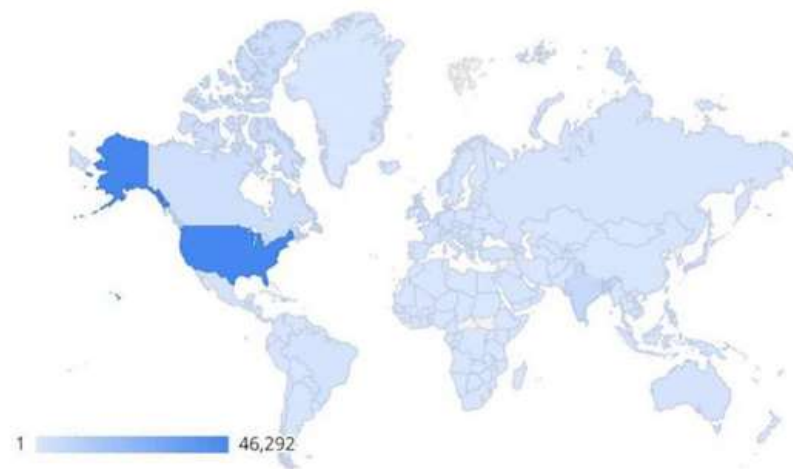
Score card

Score card

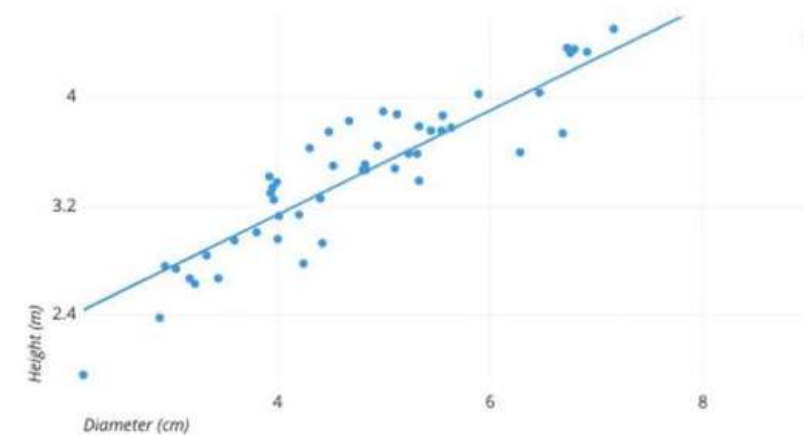
Score card

Score card

Geo chart



๑.๙๖๑๖ ๙๖ ๑.๙๖๑๖๙๖



Table

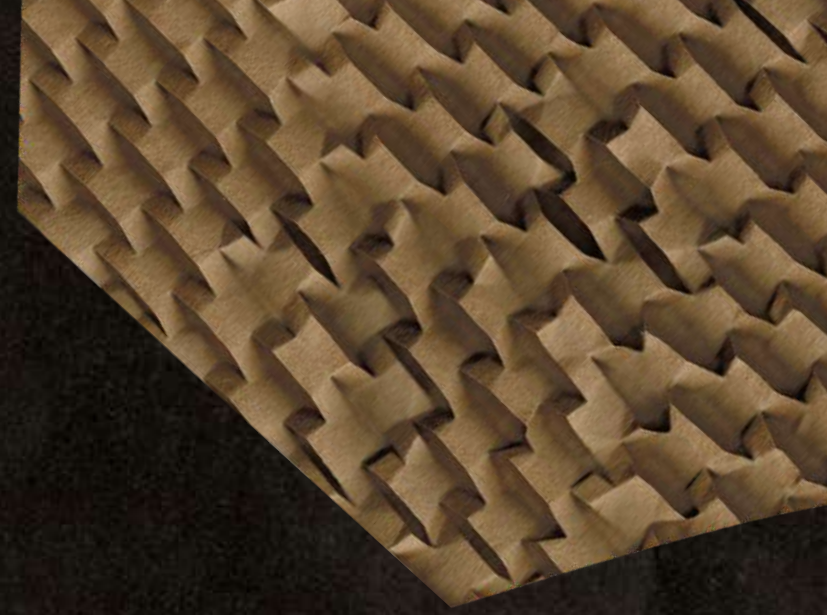
๑.๙๖๑๖๙๖ walk / access / drive

GROUP MEMBERS

633020446-7 PORNNAKANOK PHUPHATHAM

633020552-8 NATTHANICH HIRUNCHAVAROD





DashBoard

YOKENAT001MISSIONIMPOSSIBLE

