

นำเสนอ

Project

นำเสนอด้วย I love อ.ต้น



วัตถุประสงค์

เพื่อดูว่าเบตต่างๆใน กทม. ที่มีปริมาณชุมชน
แօอัด ส่งผลให้มีประชากรเป็นชาวต่างชาติ
เป็นจำนวนมากหรือไม่ โดยจัดกลุ่มเป็นเบตที่
มีชุมชนแօอัดน้อย และกลุ่มที่มีชุมชนแօอัด
มาก

แหล่งข้อมูล

ข้อมูลจำนวนประชากรในกรุงเทพมหานคร จำแนกตามอายุและเขต
ปี 56

อ้างอิง : <https://data.go.th/dataset/poppulationbyage56>
จาก : กองสารสนเทศภูมิศาสตร์ สำนักยุทธศาสตร์และประเมินผล

ข้อมูลจำนวนประชากรในกรุงเทพมหานคร จำแนกตามอายุและเขต
ปี 57

อ้างอิง : <https://data.go.th/dataset/poppulationbyage57>
จาก : กองสารสนเทศภูมิศาสตร์ สำนักยุทธศาสตร์และประเมินผล

ข้อมูลจำนวนประชากรในกรุงเทพมหานคร จำแนกตามอายุและเขต
ปี 59

อ้างอิง : <https://data.go.th/dataset/poppulationbyage59>
จาก : กองสารสนเทศภูมิศาสตร์ สำนักยุทธศาสตร์และประเมินผล

จำนวนชุมชนต่างๆในกรุงเทพมหานคร

อ้างอิง : <https://webportal.bangkok.go.th>
จาก : กองยุทธศาสตร์บริหารจัดการ

ตารางที่ 1

ข้อมูลจำนวนประชากรในกรุงเทพมหานคร จำแนกตามอายุและเขตปี 56

In [43]: data0

Out[43]:

	id	dcode	sex	dcode_sex	age0_56	age1_56	age2_56	age3_56	age4_56	age5_56	age6_56	age7_56	age8_56	age9_56	age10_56	age11_56	age12_
0	1	1033	1	1033_1	446	548	465	506	532	486	529	490	558	505	584	546	578
1	2	1033	2	1033_2	391	491	493	491	489	494	535	502	501	542	515	501	592
2	3	1018	1	1018_1	278	289	269	292	281	314	340	289	322	336	342	321	345
3	4	1018	2	1018_2	240	291	269	264	266	315	332	313	302	346	358	361	384
4	5	1046	1	1046_1	981	1161	1163	1071	1158	1191	1291	1289	1259	1294	1275	1165	1193
...
95	96	1003	2	1003_2	1016	1079	1082	1060	1108	1155	1213	1147	1234	1176	1188	1189	1172
96	97	1041	1	1041_1	390	487	472	443	466	484	500	557	531	522	562	545	588
97	98	1041	2	1041_2	392	449	415	461	440	487	541	574	551	530	525	535	531
98	99	1017	1	1017_1	337	401	326	328	352	358	400	409	399	405	369	407	369
99	100	1017	2	1017_2	306	343	354	338	358	384	384	371	397	393	362	378	388

100 rows × 111 columns

ตารางที่ 2

ข้อมูลจำนวนประชากรในกรุงเทพมหานคร จำแนกตามอายุและเขตปี 57

In [44]: data1

Out[44]:

	id	dcode	sex	dcode_sex	age0_57	age1_57	age2_57	age3_57	age4_57	age5_57	age6_57	age7_57	age8_57	age9_57	age10_57	age11_57	age12_
0	1	1033	1	1033_1	447	499	557	473	491	521	472	526	486	549	503	579	540
1	2	1033	2	1033_2	379	413	492	485	478	485	488	532	492	500	540	514	488
2	3	1018	1	1018_1	236	301	277	273	287	282	313	335	283	312	327	334	322
3	4	1018	2	1018_2	206	262	293	257	268	259	316	334	309	313	370	357	350
4	5	1046	1	1046_1	1029	1079	1191	1184	1120	1173	1223	1292	1293	1270	1290	1271	1195
...
95	96	1003	2	1003_2	938	1062	1084	1080	1093	1110	1162	1225	1152	1231	1203	1192	1200
96	97	1041	1	1041_1	398	412	488	468	446	472	476	509	534	519	508	547	547
97	98	1041	2	1041_2	364	402	450	398	463	445	482	534	562	536	515	527	534
98	99	1017	1	1017_1	333	362	403	340	336	359	365	406	404	406	404	365	400
99	100	1017	2	1017_2	339	349	342	359	341	361	376	368	362	386	389	365	378

100 rows × 111 columns

ตารางที่ 3

ข้อมูลจำนวนประชากรในกรุงเทพมหานคร จำแนกตามอายุและเขตปี 59

In [45]:

data2

Out[45]:

						age0_59	age1_59	age2_59	age3_59	age4_59	age5_59	age6_59	age7_59	age8_59	age9_59	age10_59	age11_59	age12_
0	1	1033	1	1033_1	374	411	478	488	543	451	479	506	468	506	470	525	502	
1	2	1033	2	1033_2	341	397	411	393	488	470	470	474	465	513	475	493	531	
2	3	1018	1	1018_1	210	210	255	295	275	255	271	267	304	336	277	292	324	
3	4	1018	2	1018_2	177	225	232	269	293	264	260	255	326	334	336	329	365	
4	5	1046	1	1046_1	993	1146	1150	1135	1248	1276	1180	1225	1232	1324	1314	1288	1335	
...	
95	96	1003	2	1003_2	840	983	1022	1070	1090	1133	1124	1109	1194	1234	1160	1258	1217	
96	97	1041	1	1041_1	370	385	423	407	485	471	445	465	460	484	509	502	481	
97	98	1041	2	1041_2	350	395	411	422	466	405	442	425	474	516	537	513	511	
98	99	1017	1	1017_1	327	343	366	350	407	356	344	358	349	377	388	387	408	
99	100	1017	2	1017_2	314	327	373	336	351	379	355	351	357	367	351	376	368	

100 rows × 111 columns

ตารางที่ 4

จำนวนชุมชนต่างๆในกรุงเทพมหานคร

In [164]: `data3 #ค่า slum ถ้าหาก มีค่าที่มากกว่า 12 แปลว่า ชุมชนนี้เป็น more ถ้าหากน้อยกว่า 12 แปลว่าชุมชนนี้เป็น nomore`

Out[164]:

	dcode	เขต	slum	morethan&nomorethan
0	1033	คลองเตย	23	more
1	1018	คลองสาน	35	more
2	1046	คลองสามวา	0	nomore
3	1043	คันนายาว	14	more
4	1030	จตุจักร	21	more
5	1035	จอมทอง	3	nomore
6	1036	ดอนเมือง	16	more
7	1026	ดินแดง	8	nomore
8	1002	ดุสิต	9	nomore
9	1019	ตลิ่งชัน	8	nomore
10	1048	ทวีวัฒนา	0	nomore
11	1049	ทุ่งครุ	1	nomore
12	1015	ธนบุรี	44	more
13	1020	บางกอกน้อย	33	more
14	1016	บางกอกใหญ่	30	more
15	1006	บางกะปิ	8	nomore
16	1021	บางขุนเทียน	3	nomore
17	1005	บางเขน	7	nomore
18	1031	บางคอแหลม	27	more
19	1040	บางแค	23	more
20	1029	บางซื่อ	47	more
21	1047	บางนา	9	nomore
22	1050	บางบอน	2	nomore

โดยทำการกำหนดว่า

ถ้าค่า slum < 12 แปลว่าเป็น nomore

ถ้าค่า slum > 12 แปลว่าเป็น more

เลือกคอลัมน์ที่ต้องการใช้

เลือกข้อมูลที่ต้องการใช้

```
In [47]: data0_RE = data0[['id','dcode','dcode_sex','othernationalities_56','sex']]  
data0_RE
```

Out[47]:

	id	dcode	dcode_sex	othernationalities_56	sex
0	1	1033	1033_1	1018	1
1	2	1033	1033_2	645	2
2	3	1018	1018_1	920	1
3	4	1018	1018_2	753	2
4	5	1046	1046_1	162	1
...
95	96	1003	1003_2	70	2
96	97	1041	1041_1	150	1
97	98	1041	1041_2	118	2
98	99	1017	1017_1	383	1
99	100	1017	1017_2	272	2

100 rows × 5 columns

ทำการเลือกคอลัมน์ที่ต้องการใช้โดยใช้คำสั่ง Dataframe โดยเลือกปี 56 เป็น ID,dcode_sex,othernationalities_56 และ sex ส่วนปี 57 และ 59 เลือกเพียงแค่ dcode_sex และ othernationalities ของแต่ละปี

```
In [48]: data1_RE = data1[['dcode_sex','othernationalities_57']]  
data1_RE
```

Out[48]:

	dcode_sex	othernationalities_57
0	1033_1	1026
1	1033_2	633
2	1018_1	917
3	1018_2	716
4	1046_1	184
...
95	1003_2	78
96	1041_1	158
97	1041_2	124
98	1017_1	402
99	1017_2	275

100 rows × 2 columns

```
In [49]: data2_RE = data2[['dcode_sex','othernationalities_59']]  
data2_RE
```

Out[49]:

	dcode_sex	othernationalities_59
0	1033_1	1032
1	1033_2	628
2	1018_1	815
3	1018_2	680
4	1046_1	230
...
95	1003_2	106
96	1041_1	166
97	1041_2	147
98	1017_1	474
99	1017_2	311

100 rows × 2 columns

รวมตราง

รวมตาราง						
merged_table_dcode_sex = data0.RE.merge(data1.RE,how='left',left_on='dcode_sex',right_on='dcode_sex') #ครั้งที่ 1						
	id	dcode	dcode_sex	othenationalities_56	sex	othenationalities_57
0	1	1033	1033_1	1018	1	1026
1	2	1033	1033_2	645	2	633
2	3	1018	1018_1	920	1	917
3	4	1018	1018_2	753	2	716
4	5	1046	1046_1	162	1	184
...
95	96	1003	1003_2	70	2	78
96	97	1041	1041_1	150	1	158
97	98	1041	1041_2	118	2	124
98	99	1017	1017_1	383	1	402
99	100	1017	1017_2	272	2	275

merged_table_dcode_sex2 = merged_table_dcode_sex.merge(data2_RE,how='left',left_on='dcode_sex',right_on='dcode_sex')							
#ANSWER 2							
	id	dcode	dcode_sex	othernationalities_56	sex	othernationalities_57	othernationalities_59
0	1	1033	1033_1	1018	1	1026	1032
1	2	1033	1033_2	645	2	633	628
2	3	1018	1018_1	920	1	917	815
3	4	1018	1018_2	753	2	716	680
4	5	1046	1046_1	162	1	184	230
...
95	96	1003	1003_2	70	2	78	106
96	97	1041	1041_1	150	1	158	166
97	98	1041	1041_2	118	2	124	147
98	99	1017	1017_1	383	1	402	474
99	100	1017	1017_2	272	2	275	311

merged_table_dcode_sex3 = merged_table_dcode_sex2.merge(data3,how='left',left_on='dcode',right_on='dcode')												
merged_table_dcode_sex3 # ให้ค่าส่วนตัวมากกว่า 12 แปลว่า บุขชุมชนเป็นบุขชุมชนแออัด(More) ด้านอยู่กว่า 12 จะไม่เป็นบุขชุมชนแออัด(Nomore)												
Out[129]:		id	dcode	dcode_sex	othernationalities_56	sex	othernationalities_57	othernationalities_59	เขต	slum	morethan&nomorethan	
0	1	1033	1033_1	1018	1	1026	1032	คลองเตย	23	more		
1	2	1033	1033_2	645	2	633	628	คลองเตย	23	more		
2	3	1018	1018_1	920	1	917	815	คลองสาน	35	more		
3	4	1018	1018_2	753	2	716	680	คลองสาน	35	more		
4	5	1046	1046_1	162	1	184	230	คลองสามวา	0	nomore		
...		
95	96	1003	1003_2	70	2	78	106	หนองจอก	0	nomore		
96	97	1041	1041_1	150	1	158	166	หลักสี่	17	more		
97	98	1041	1041_2	118	2	124	147	หลักสี่	17	more		
98	99	1017	1017_1	383	1	402	474	ทวายดาว	18	more		
99	100	1017	1017_2	272	2	275	311	ทวายดาว	18	more		

```
In [130]: merged_table_dcode_sex3.shape  
Out[130]: (100, 10)  
  
In [131]: merged_table_dcode_sex3.isnull().any()  
Out[131]: id          False  
dcode         False  
dcode_sex      False  
othernationalities_56  False  
sex           False  
othernationalities_57  False  
othernationalities_59  False  
ഏବ  False  
slum          False  
morethan&nomorethan   False  
dtype: bool
```

ใช้คำสั่ง df.merged ในการ
รวมตาราง
โดยใช้ dcode_sex เป็น
index
พอเชื่อมตารางเสร็จจึงทำการ
เช็คແກวและคอลัมน์ของข้อมูล
และใช้คำสั่ง isnull().any()
เพื่อดูว่าคอลัมน์และແກวไหนมี
ค่า nan อยู่

Data minning

data minning
ประชากรปี 56

In [17]: merged_table_dcode_sex3

Out[17]:

	id	dcode	dcode_sex	othenationalities_56	sex	othenationalities_57	othenationalities_59	เขต	slum	morethan&nomorethan
0	1	1033	1033_1	1018	1	1026	1032	คลองเตย	23	more
1	2	1033	1033_2	645	2	633	628	คลองเตย	23	more
2	3	1018	1018_1	920	1	917	815	คลองสาน	35	more
3	4	1018	1018_2	753	2	716	680	คลองสาน	35	more
4	5	1046	1046_1	162	1	184	230	คลองสามวา	0	nomore
...
95	96	1003	1003_2	70	2	78	106	หนองจอก	0	nomore
96	97	1041	1041_1	150	1	158	166	หลักสี่	17	more
97	98	1041	1041_2	118	2	124	147	หลักสี่	17	more
98	99	1017	1017_1	383	1	402	474	ท่าข้าว	18	more
99	100	1017	1017_2	272	2	275	311	ท่าข้าว	18	more

100 rows x 10 columns

In [18]: A0 = merged_table_dcode_sex3.drop(columns=['id','dcode','sex','othenationalities_57','othenationalities_59','เขต'])
A0 # นำคอลัมน์ othenationalities_56 มาใช้ในการทำนายประชากรที่อยู่ในเขต ว่าต้องแยกอัตโนมัติเป็นกลุ่มๆ ไม่แออัดอัตโนมัติ

Out[18]:

	dcode_sex	othenationalities_56	slum	morethan&nomorethan
0	1033_1	1018	23	more
1	1033_2	645	23	more
2	1018_1	920	35	more
3	1018_2	753	35	more
4	1046_1	162	0	nomore
...
95	1003_2	70	0	nomore
96	1041_1	150	17	more
97	1041_2	118	17	more
98	1017_1	383	18	more
99	1017_2	272	18	more

100 rows x 4 columns

จากการรวมตาราง merged_table_dcode_sex3 จึงนำข้อมูลที่ต้องการพิจารณาออกแบบมาทำเป็นตารางใหม่ โดยในตารางจะมี dcode_sex,othenationalities_56,slum และ morethan&nomorethan

*โดยทำการกำหนดว่า

ถ้าค่า slum < 12 แปลว่าเป็นน้ำมันไม่แออัด(nomore)

ถ้าค่า slum > 12 แปลว่าเป็นน้ำมันแออัด(more)

*ทำแบบเดิมกับปี 57 และปี 59

In [19]: target = merged_table_dcode_sex3['sex']
target

Out[19]:

0	1
1	2
2	1
3	2
4	1
..	..
95	2
96	1
97	2
98	1
99	2

Name: sex, Length: 100, dtype: int64

In [20]: `A0['morethan&nomorethan'] = A0['morethan&nomorethan'].map({'more': 1, 'nomore': 2})`
`A0 #ใช้คำสั่ง map เพื่อเปลี่ยนข้อมูลในคอลัมน์ morethan&nomorethan ให้เป็นตัวเลข โดยให้ more = 1 และ nomore = 2`

Out[20]:	dcode_sex	othernationalities_56	slum	morethan&nomorethan
0	1033_1	1018	23	1
1	1033_2	645	23	1
2	1018_1	920	35	1
3	1018_2	753	35	1
4	1046_1	162	0	2
...
95	1003_2	70	0	2
96	1041_1	150	17	1
97	1041_2	118	17	1
98	1017_1	383	18	1
99	1017_2	272	18	1

100 rows x 4 columns

```
In [21]: X = A0.values
```

```
Out[21]: array([['1033_1', 1018, 23, 1],  
   ['1033_2', 645, 23, 1],  
   ['1018_1', 920, 35, 1],  
   ['1018_2', 753, 35, 1],  
   ['1046_1', 162, 0, 2],  
   ['1046_2', 105, 0, 2],  
   ['1043_1', 157, 14, 1],  
   ['1043_2', 103, 14, 1],  
   ['1030_1', 457, 21, 1],  
   ['1030_2', 342, 21, 1],  
   ['1035_1', 735, 3, 2],  
   ['1035_2', 564, 3, 2],  
   ['1036_1', 222, 16, 1],  
   ['1036_2', 137, 16, 1],  
   ['1026_1', 459, 8, 2],  
   ['1026_2', 403, 8, 2],
```

```
In [22]: A0['morethan&nomorethan'].value_counts()
```

```
Out[22]: 2    56  
        1    44  
Name: morethan&nomorethan, dtype: int64
```

In [23]: `y = target.values`

`map({})`เป็นฟังก์ชันที่ใช้ในการจัดการกับ collection จำพวก array,list ด้วยค่า ข้อมูลในตาราง `morethan&nomorethan` เป็นภาษาที่เป็น string จึงทำการเปลี่ยนค่า `morethan=1` และ `nomorethan=2` ให้เป็นค่า array

ให้ x ที่เป็น Feature=ตารางข้างต้น ตามด้วย .values
เป็นคำสั่งสำหรับดึงค่าใน Key ทั้งหมดของ Dictionary ออกมา ดังภาพ

จาก output22 ที่ใช้คำสั่ง `.value_count()` เพื่อนับจำนวนข้อมูล
ได้ว่า มีชุมชนที่ไม่แออัด=56 และมีจำนวนชุมชนแออัด=44

`y = target.values` โดยให้ `y` ที่มีค่าเป็นเพศแทนเป็นคำตอบ
*โดย 1=เพศชาย และ 2=เพศหญิง

```
train_test_split

In [24]: from sklearn.model_selection import train_test_split

In [25]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=6)

In [26]: X_train.shape

Out[26]: (50, 4)

    train - validation แบ่งอีกรอบ

In [104]: X_train2, X_val, y_train2, y_val = train_test_split(X_train, y_train, test_size=0.5, random_state=6)

In [105]: X_train2.shape #เหลือแค่ 35 และ 4 คอลัมน์ ที่จะนำมาทำต้นไม้

Out[105]: (25, 4)

import

In [106]: from sklearn.tree import DecisionTreeClassifier

Define

In [107]: mytree56 = DecisionTreeClassifier(max_leaf_nodes=5,criterion='entropy') #กำหนดให้มีมากสุด 5 node

Train

In [108]: mytree56.fit(X_train2,y_train2)

Out[108]: DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='entropy',
                                max_depth=None, max_features=None, max_leaf_nodes=5,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, presort='deprecated',
                                random_state=None, splitter='best')
```

เริ่มจากการเลือกตัวตันแบบโดยใช้ฟังก์ชัน `train_test_split` แบบคืนค่า 4 ตัวแปร ประกอบด้วย `train_x`, `train_y` และ `test_x`, `test_y` วิธีการนี้ ข้อมูล `train_x` จะประกอบด้วยตัวแปรอิสระที่ถูกแบ่งออกมาใช้สำหรับการสอน ส่วน `train_y` คือตัวแปรตามที่ถูกแบ่งออกมาใช้สำหรับการสอนในทำนองเดียวกัน ตัวแปร `test_x` คือตัวแปรอิสระ และ `test_y` คือตัวแปรตาม ที่ถูกแบ่งออกมาสำหรับการทดสอบโมเดลที่สร้างขึ้นโดยให้กลุ่ม `train` สำหรับให้ ML เรียนรู้ และให้กลุ่ม `test` สำหรับ ML ลองทำนาย `target variable`

โดยก็จะกำหนดค่า `Test_size = 0.5` หรือประมาณ 50 เปอร์เซ็นของข้อมูล และ `random_state = 6` คือการสุ่มที่ไม่ใช่การสุ่มที่แท้จริงก็จะสุ่มได้ตำแหน่งที่ 6 เท่ากัน และยังทำการ Train แบ่งอีกรอบ เพื่อให้จำนวนข้อมูลลดลง ก็จะได้ 25 และ เพราะใช้ `Test_size = 0.5` เมื่อันเดิม

Evaluate วัดผล

```
In [109]: from sklearn.metrics import accuracy_score
```

Test 56

```
In [110]: tree56_result = mytree56.predict(X_val) # เอาข้อมูลเรามา Test
```

```
In [111]: accuracy_score(y_val, tree56_result ) #ความแม่นยำของคันนี้ได้ 40 เปอร์เซ็น
```

```
Out[111]: 0.52
```

Evaluate วัดผล

```
In [51]: from sklearn.metrics import accuracy_score
```

Test 57

```
In [52]: tree57_result = mytree57.predict(X_val1) # เอาข้อมูลเรามา Test
```

```
In [53]: accuracy_score(y_val1, tree57_result ) #ความแม่นยำของคันนี้ได้ 50 เปอร์เซ็น
```

```
Out[53]: 0.52
```

Evaluate วัดผล

```
In [119]: from sklearn.metrics import accuracy_score
```

Test 59

```
In [120]: tree59_result = mytree59.predict(X_val2) # เอาข้อมูลเรามา Test
```

```
In [121]: accuracy_score(y_val2, tree59_result ) #ความแม่นยำของคันนี้ได้ 50 เปอร์เซ็น
```

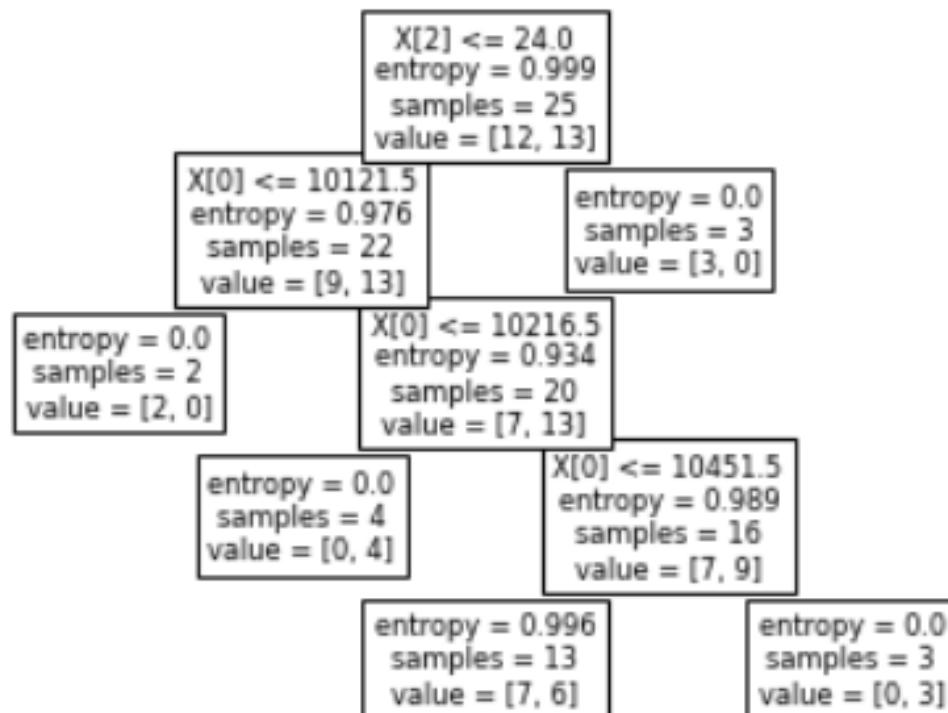
```
Out[121]: 0.52
```

Decision Tree

```
In [112]: from sklearn.tree import plot_tree
```

```
In [113]: plot_tree(mytree56)
```

```
Out[113]: [Text(167.4, 195.696, 'X[2] <= 24.0\nentropy = 0.999\nsamples = 25\nvalue = [12, 13]'),  
Text(111.60000000000001, 152.208, 'X[0] <= 10121.5\nentropy = 0.976\nsamples = 22\nvalue = [9, 13]'),  
Text(55.80000000000004, 108.72, 'entropy = 0.0\nsamples = 2\nvalue = [2, 0]'),  
Text(167.4, 108.72, 'X[0] <= 10216.5\nentropy = 0.934\nsamples = 20\nvalue = [7, 13]'),  
Text(111.60000000000001, 65.232, 'entropy = 0.0\nsamples = 4\nvalue = [0, 4]'),  
Text(223.20000000000002, 65.232, 'X[0] <= 10451.5\nentropy = 0.989\nsamples = 16\nvalue = [7, 9]'),  
Text(167.4, 21.744, 'entropy = 0.996\nsamples = 13\nvalue = [7, 6]'),  
Text(279.0, 21.744, 'entropy = 0.0\nsamples = 3\nvalue = [0, 3]'),  
Text(223.20000000000002, 152.208, 'entropy = 0.0\nsamples = 3\nvalue = [3, 0]')]
```

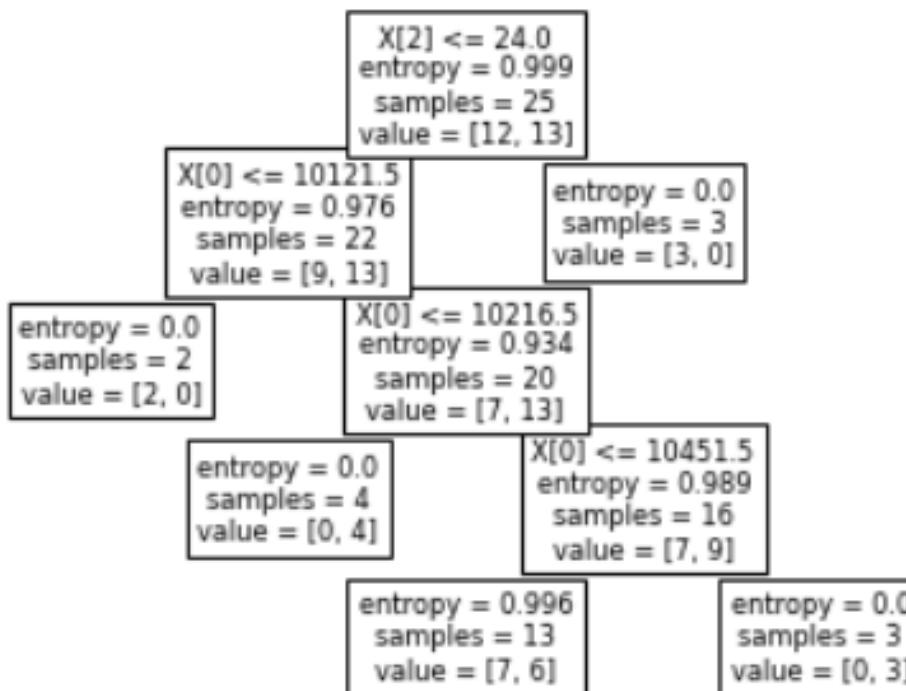


Decision Tree

```
In [54]: from sklearn.tree import plot_tree
```

```
In [55]: plot_tree(mytree57)
```

```
Out[55]: [Text(167.4, 195.696, 'X[2] <= 24.0\nentropy = 0.999\nsamples = 25\nvalue = [12, 13]'),  
Text(111.60000000000001, 152.208, 'X[0] <= 10121.5\nentropy = 0.976\nsamples = 22\nvalue = [9, 13]'),  
Text(55.80000000000004, 108.72, 'entropy = 0.0\nsamples = 2\nvalue = [2, 0]'),  
Text(167.4, 108.72, 'X[0] <= 10216.5\nentropy = 0.934\nsamples = 20\nvalue = [7, 13]'),  
Text(111.60000000000001, 65.232, 'entropy = 0.0\nsamples = 4\nvalue = [0, 4]'),  
Text(223.20000000000002, 65.232, 'X[0] <= 10451.5\nentropy = 0.989\nsamples = 16\nvalue = [7, 9]'),  
Text(167.4, 21.744, 'entropy = 0.996\nsamples = 13\nvalue = [7, 6]'),  
Text(279.0, 21.744, 'entropy = 0.0\nsamples = 3\nvalue = [0, 3]'),  
Text(223.20000000000002, 152.208, 'entropy = 0.0\nsamples = 3\nvalue = [3, 0]')]
```

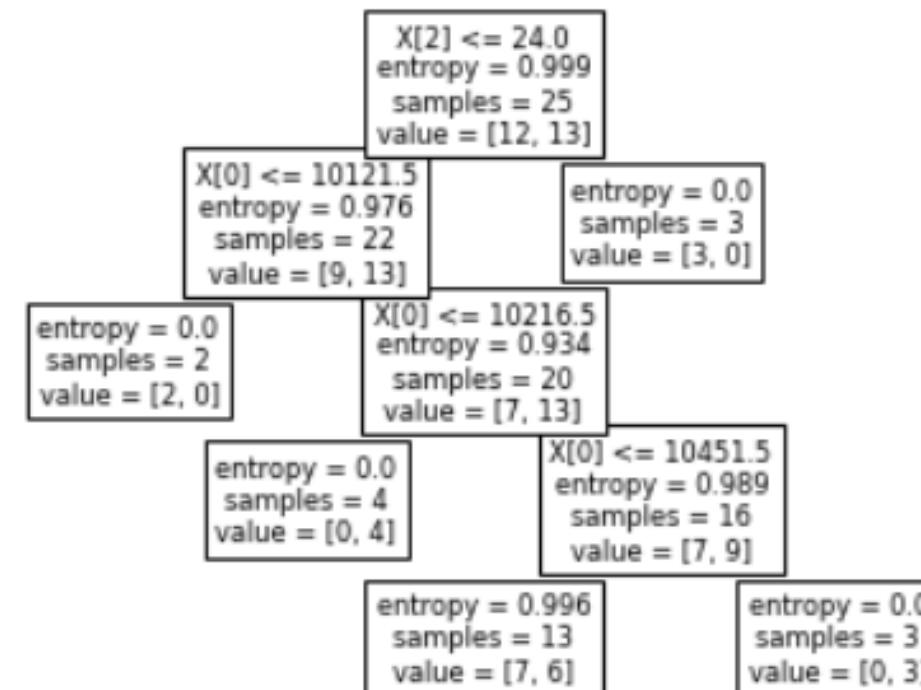


Decision Tree

```
In [122]: from sklearn.tree import plot_tree
```

```
In [123]: plot_tree(mytree59)
```

```
Out[123]: [Text(167.4, 195.696, 'X[2] <= 24.0\nentropy = 0.999\nsamples = 25\nvalue = [12, 13]'),  
Text(111.60000000000001, 152.208, 'X[0] <= 10121.5\nentropy = 0.976\nsamples = 22\nvalue = [9, 13]'),  
Text(55.80000000000004, 108.72, 'entropy = 0.0\nsamples = 2\nvalue = [2, 0]'),  
Text(167.4, 108.72, 'X[0] <= 10216.5\nentropy = 0.934\nsamples = 20\nvalue = [7, 13]'),  
Text(111.60000000000001, 65.232, 'entropy = 0.0\nsamples = 4\nvalue = [0, 4]'),  
Text(223.20000000000002, 65.232, 'X[0] <= 10451.5\nentropy = 0.989\nsamples = 16\nvalue = [7, 9]'),  
Text(167.4, 21.744, 'entropy = 0.996\nsamples = 13\nvalue = [7, 6]'),  
Text(279.0, 21.744, 'entropy = 0.0\nsamples = 3\nvalue = [0, 3]'),  
Text(223.20000000000002, 152.208, 'entropy = 0.0\nsamples = 3\nvalue = [3, 0]')]
```



VISUALIZATION

- เริ่มจาก Import `from matplotlib import pyplot as plt`
- สร้างตารางใหม่และเลือกข้อมูลที่ต้องนำมาใช้คือ ชื่อตาราง `Plt_0` ใช้ข้อมูล `sex,othernationalties_56,othernationalties_57,othernationalties_59`

Visualization

In [75]: `from matplotlib import pyplot as plt`

In [76]: `merged_table_dcode_sex3`

Out[76]:

	<code>id</code>	<code>dcode</code>	<code>dcode_sex</code>	<code>othernationalities_56</code>	<code>sex</code>	<code>othernationalities_57</code>	<code>othernationalities_59</code>	<code>เขต</code>	<code>slum</code>	<code>morethan&nomorethan</code>
0	1	1033	1033_1	1018	1	1026	1032	คลองเตย	23	more
1	2	1033	1033_2	645	2	633	628	คลองเตย	23	more
2	3	1018	1018_1	920	1	917	815	คลองสาน	35	more
3	4	1018	1018_2	753	2	716	680	คลองสาน	35	more
4	5	1046	1046_1	162	1	184	230	คลองสามวา	0	nomore
...
95	96	1003	1003_2	70	2	78	106	หนองจอก	0	nomore
96	97	1041	1041_1	150	1	158	166	หลักสี่	17	more
97	98	1041	1041_2	118	2	124	147	หลักสี่	17	more
98	99	1017	1017_1	383	1	402	474	ทวยขวาง	18	more
99	100	1017	1017_2	272	2	275	311	ทวยขวาง	18	more

100 rows × 10 columns

In [77]: `Plt_0 = merged_table_dcode_sex3[['sex','othernationalities_56','othernationalities_57','othernationalities_59']]`
`Plt_0`

Out[77]:

	<code>sex</code>	<code>othernationalities_56</code>	<code>othernationalities_57</code>	<code>othernationalities_59</code>
0	1	1018	1026	1032
1	2	645	633	628
2	1	920	917	815
3	2	753	716	680
4	1	162	184	230
...
95	2	70	78	106
96	1	150	158	166
97	2	118	124	147
98	1	383	402	474
99	2	272	275	311

100 rows × 4 columns

- Plt_0_sex1 คือ ตารางข้อมูลที่เลือกแค่ เพศ ชายหรือ 1
- Plt_0_sex2 คือ ตารางข้อมูลที่เลือกแค่ เพศ หญิงหรือ 2
- จึงใช้คำสั่ง Df.describe() เพื่อต้องการเลือกดูค่าทางสถิติและจากนั้นนำข้อมูลที่เป็นค่า Mean ของประชากรปี 56,57,59 มาดูเปรียบเทียบประชากรว่าต่างกันมากหรือไม่

In [78]:
Plt_0_sex1 = Plt_0[Plt_0['sex'] == 1]
Plt_0_sex1

Out[78]:

	sex	othernationalities_56	othernationalities_57	othernationalities_59
0	1	1018	1026	1032
2	1	920	917	815
4	1	162	184	230
6	1	157	162	696
8	1	457	473	505
10	1	735	715	595
12	1	222	229	277
14	1	459	450	401
16	1	707	705	475
18	1	232	231	234
20	1	180	192	247
22	1	310	319	337
24	1	808	767	646
26	1	349	337	302
28	1	344	346	300
30	1	418	427	488
32	1	390	406	417
34	1	335	340	393
36	1	716	693	583
38	1	527	528	542
40	1	454	449	416

In [80]:
Plt_0_sex2 = Plt_0[Plt_0['sex'] == 2]
Plt_0_sex2

Out[80]:

	sex	othernationalities_56	othernationalities_57	othernationalities_59
1	2	645	633	628
3	2	753	716	680
5	2	105	127	170
7	2	103	109	630
9	2	342	343	361
11	2	564	551	483
13	2	137	137	146
15	2	403	395	355
17	2	375	362	276
19	2	184	187	205
21	2	176	170	209
23	2	260	259	271
25	2	667	637	550
27	2	315	307	283
29	2	305	290	281
31	2	355	362	393
33	2	367	376	393
35	2	240	240	291
37	2	561	536	444
39	2	479	478	495

In [79]:
Plt_0_sex1.describe()

Out[79]:

	sex	othernationalities_56	othernationalities_57	othernationalities_59
count	50.0	50.000000	50.000000	50.000000
mean	1.0	502.180000	505.240000	505.360000
std	0.0	353.677398	354.905538	341.48958
min	1.0	104.000000	119.000000	148.000000
25%	1.0	270.250000	272.000000	300.250000
50%	1.0	380.000000	387.000000	408.500000
75%	1.0	649.500000	651.000000	590.250000
max	1.0	2036.000000	2074.000000	2076.000000

In [81]:
Plt_0_sex2.describe()

Out[81]:

	sex	othernationalities_56	othernationalities_57	othernationalities_59
count	50.0	50.000000	50.000000	50.000000
mean	2.0	384.500000	383.260000	384.900000
std	0.0	268.343387	266.669467	247.579734
min	2.0	70.000000	78.000000	106.000000
25%	2.0	240.250000	234.000000	240.500000
50%	2.0	312.000000	309.500000	305.500000
75%	2.0	465.000000	464.000000	443.250000
max	2.0	1558.000000	1576.000000	1526.000000

เพศ 1

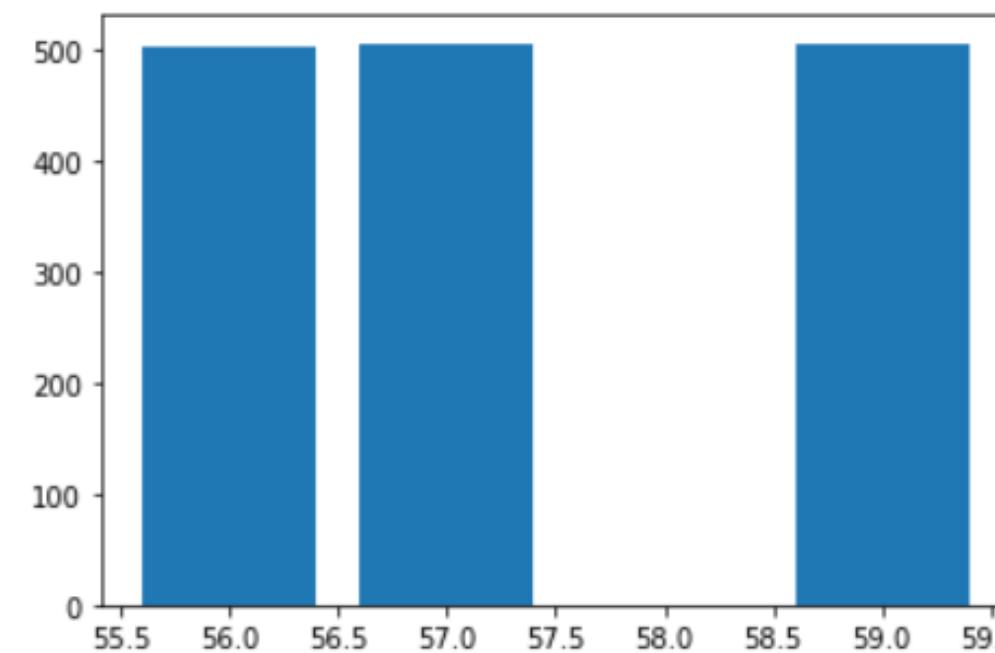
1. ประชากรสัญชาติอื่น ปี 56 เฉลี่ยมีจำนวน 502.180000
2. ประชากรสัญชาติอื่น ปี 57 เฉลี่ยมีจำนวน 505.240000
3. ประชากรสัญชาติอื่น ปี 59 เฉลี่ยมีจำนวน 505.360000

เพศ 2

1. ประชากรสัญชาติอื่น ปี 56 เฉลี่ยมีจำนวน 384.500000
2. ประชากรสัญชาติอื่น ปี 57 เฉลี่ยมีจำนวน 383.260000
3. ประชากรสัญชาติอื่น ปี 59 เฉลี่ยมีจำนวน 384.900000

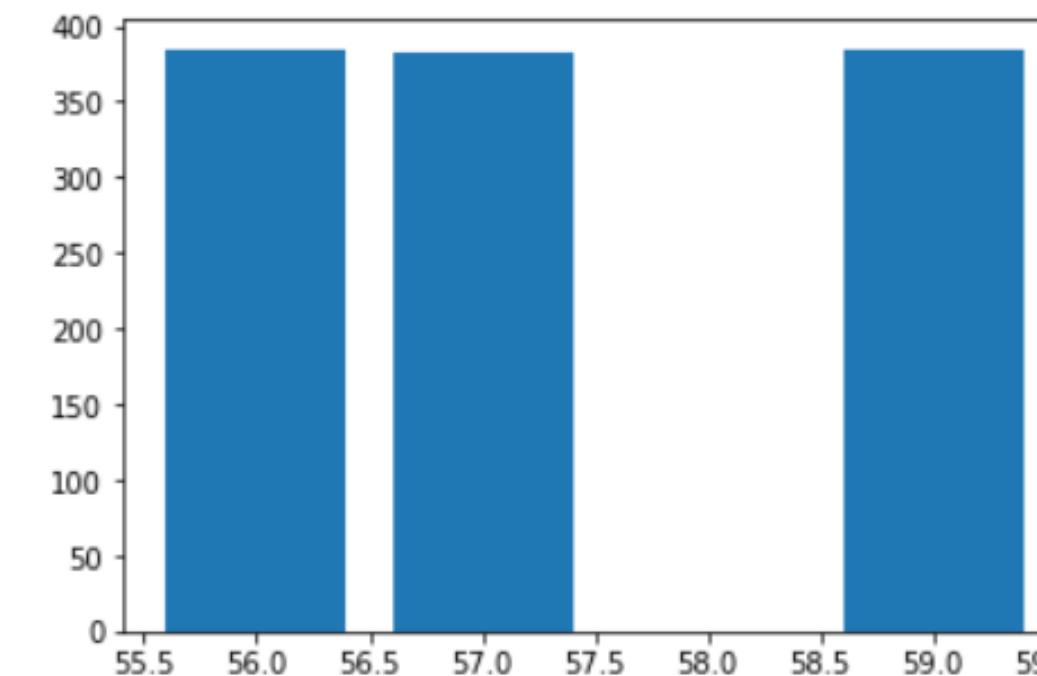
```
In [83]: plt.bar([56,57,59],men_means1)
```

```
Out[83]: <BarContainer object of 3 artists>
```



```
In [84]: plt.bar([56,57,59],women_means2)
```

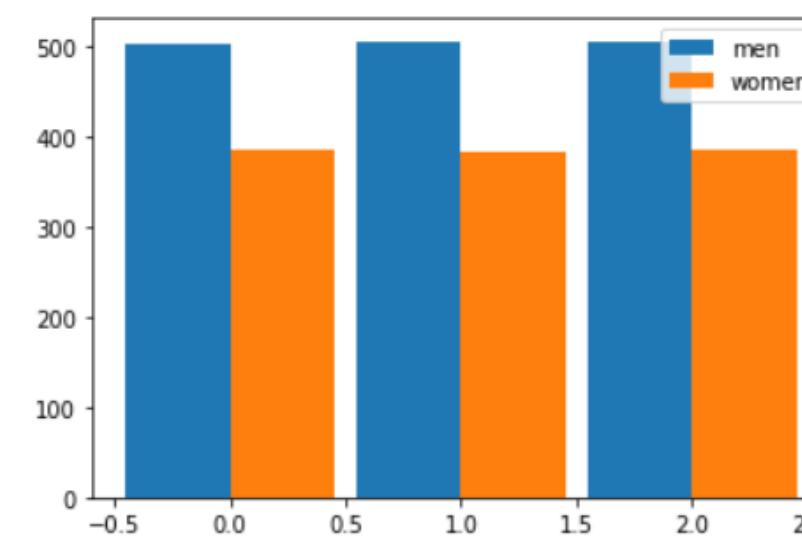
```
Out[84]: <BarContainer object of 3 artists>
```



```
In [85]: width =0.45
```

```
plt.bar(np.array([0,1,2]) - width/2, men_means1, width=width, label = 'men')
plt.bar(np.array([0,1,2]) + width/2, women_means2, width=width, label = 'women')
plt.legend()
```

```
Out[85]: <matplotlib.legend.Legend at 0x7f8b7499b9d0>
```



จากการภาพแห่งสรุปได้ว่าชาวต่างชาติส่วนใหญ่ที่เข้ามาอยู่ใน
กม. ส่วนใหญ่เป็นเพศชายในทุกๆ ปี

สรุปจากราฟ

ค่า slum น้อยกว่า 12 จะเห็นว่าทั้ง 3 ปีมีชาวต่างชาติในเขตนั้นๆ ใกล้เคียงกัน โดยส่วนใหญ่ต่ำกว่า 2,000 คน

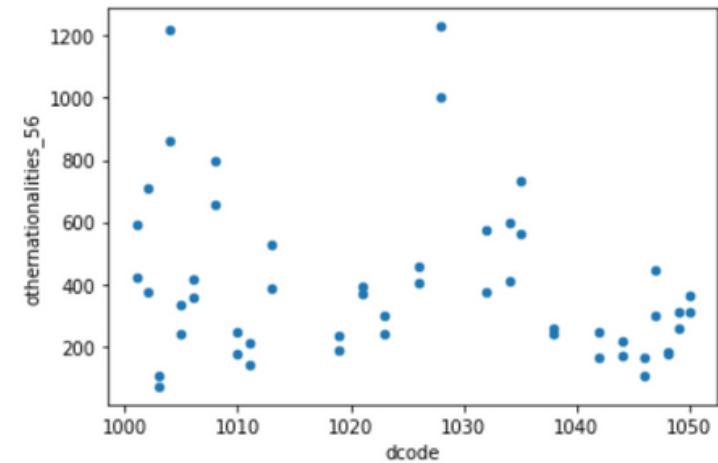
In [92]: Plt_0_slum_1 = Plt_0_slum[Plt_0_slum['slum'] < 12]
Plt_0_slum_1

Out[92]:

	dcode	othenationalities_56	othenationalities_57	othenationalities_59	slum
4	1046	162	184	230	0
5	1046	105	127	170	0
10	1035	735	715	595	3
11	1035	564	551	483	3
14	1026	459	450	401	8
15	1026	403	395	355	8
16	1002	707	705	475	9
17	1002	375	362	276	9
18	1019	232	231	234	8
19	1019	184	187	205	8
20	1048	180	192	247	0
21	1048	176	170	209	0
22	1049	310	319	337	1
23	1049	260	259	271	1
30	1006	418	427	488	8
31	1006	355	362	393	8
32	1021	390	406	417	3
33	1021	367	376	393	3
34	1005	335	340	393	7
35	1005	240	240	291	7
42	1047	447	452	456	9

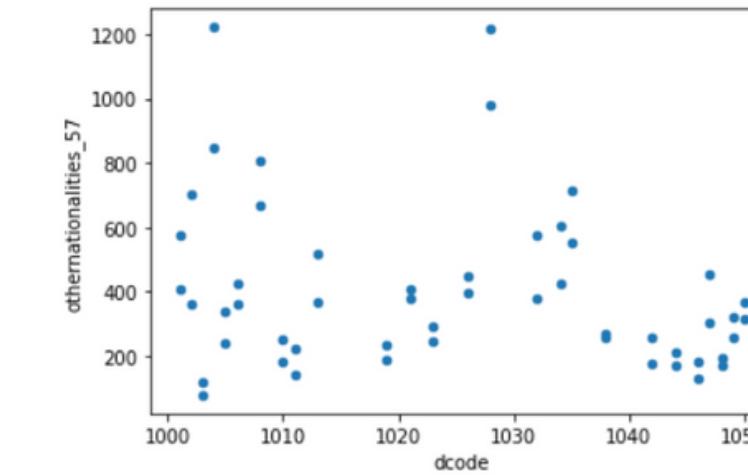
In [96]: Plt_0_slum_1.plot.scatter('dcode','othenationalities_56')

Out[96]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8b6fb1390>



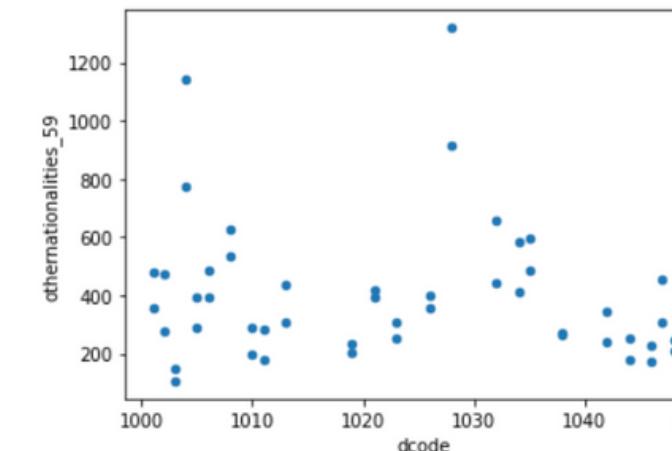
In [99]: Plt_0_slum_1.plot.scatter('dcode','othenationalities_57')

Out[99]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8b6fb19c50>



In [100]: Plt_0_slum_1.plot.scatter('dcode','othenationalities_59')

Out[100]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8b6fad2d10>



สรุปจากการเเสดง

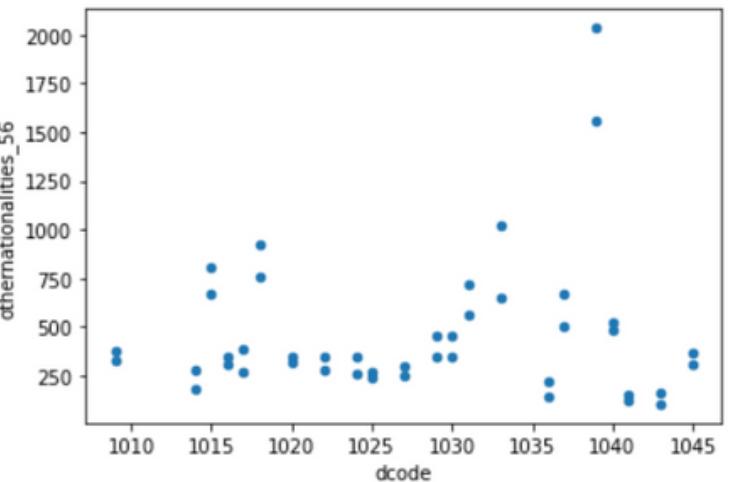
ค่า slum มากกว่า 12 จะเห็นว่าในปี 56 มีชาวต่างชาติส่วนใหญ่อยู่ในเขต
แต่ละเขตมากกว่า 2,000 คน

```
In [98]: Plt_0_slum_2 = Plt_0_slum[Plt_0_slum['slum'] > 12 ]  
Plt_0_slum_2
```

Out[98]:	dcode	othenationalities_56	othenationalities_57	othenationalities_59	slum
0	1033	1018	1026	1032	23
1	1033	645	633	628	23
2	1018	920	917	815	35
3	1018	753	716	680	35
6	1043	157	162	696	14
7	1043	103	109	630	14
8	1030	457	473	505	21
9	1030	342	343	361	21
12	1036	222	229	277	16
13	1036	137	137	146	16
24	1015	808	767	646	44
25	1015	667	637	550	44
26	1020	349	337	302	33
27	1020	315	307	283	33
28	1016	344	346	300	30
29	1016	305	290	281	30
36	1031	716	693	583	27
37	1031	561	536	444	27
38	1040	527	528	542	23
39	1040	479	478	495	23

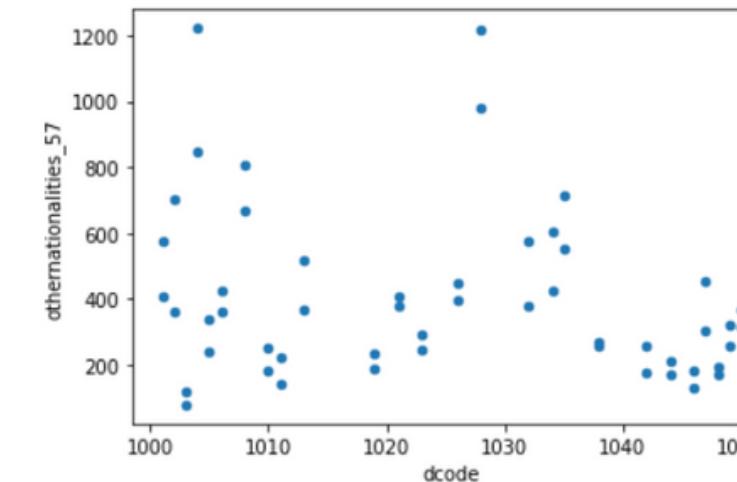
```
In [101]: Plt_0_slum_2.plot.scatter('dcode','othenationalities_56')
```

```
Out[101]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8b6f9f5090>
```



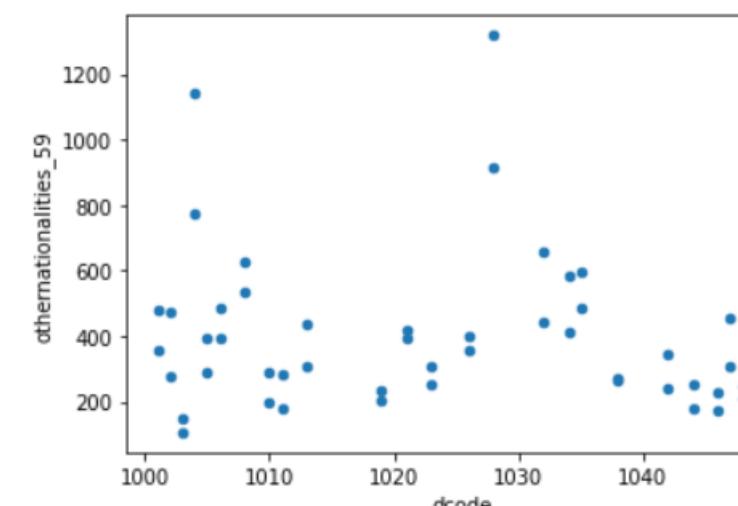
```
In [102]: Plt_0_slum_1.plot.scatter('dcode','othenationalities_57')
```

```
Out[102]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8b6f9dd150>
```



```
In [103]: Plt_0_slum_1.plot.scatter('dcode','othenationalities_59')
```

```
Out[103]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8b6f946b10>
```



สรุป

ชาวต่างชาติส่วนใหญ่อาศัยอยู่ในเขตที่มีชุมชนแออัดอยู่มากและส่วนใหญ่เป็นเพศชาย จึงสรุปได้ว่าจำนวนชาวต่างชาติกับจำนวนชุมชนแออัดในเบตันน์มีความสัมพันธ์กันในทิศทางเดียวกัน

รายชื่อสมาชิกกลุ่ม

นางสาวนวลแพร พนาวัฒนวงศ์	รหัส 623020525-0
นายชัยกร จันทร์ลุน	รหัส 623020517-9
นายช่วงชัย จันทร์มนี	รหัส 623020519-5
นายอภิวัฒน์ เหลี่ยมสิงขร	รหัส 623020545-4
นายกิตติคุณ เกียรติศักดิ์ศรี	รหัส 623020760-0