hw3

task2c

```
>>> task2c_output.count()
85136
>>> task2c_output.take(20)
[('n346594', ['N346594']), ('dpe3045', ['DPE3045']), ('gmz3750', ['GMZ3750', 'GMZ3750']), ('44884', ['44884']), ('xz876g', ['XZ876G']), ('gmu429
6', ['GMU4296']), ('gej8235', ['GEJ8235']), ('42972jw', ['42972JW']), ('66951', ['66951']), ('ggs5172', ['GGS5172', 'GGS5172']), ('51329a', ['51
329A']), ('31695jz', ['31695JZ']), ('79638ka', ['79638KA']), ('88720mb', ['88720MB']), ('14307lv', ['14307LV']), ('ewt1353', ['EWT1353']), ('gxc
7520', ['GXC7520']), ('t639084c', ['T639084C']), ('gtj6780', ['GTJ6780']), ('401zgu', ['401ZGU'])]
>>> task2c_cluster.count()
3
>>> task2c_cluster.collect()
[('jnp981', ['JNP981&', 'JNP981']), ('ns', ['N/S', 'N/S', 'NS']), ('l21687', ['L.21687', 'L21687', 'L21687'])]
>>>
```

task2c_result = task2c_output

task2d

```
>>> task2d_output.count()
85134
>>> task2d_output.take(20)
[('346465594659n34', ['N346594']), ('262624dp2gdpp26', ['GDP2624']), ('11111m1mb511651', ['65111MB']), ('76g876xz8z87', ['XZ876G']), ('ikiinekin
pik', ['PIKINE']), ('296429gmumu4u42', ['GMU4296']), ('235823ej8gejj82', ['GEJ8235']), ('2972jw42972j972', ['42972JW']), ('00j0jm340400634', ['6
3400JM']), ('172517ggsgs5s51', ['GGS5172', 'GGS5172']), ('16k2164926ka921', ['49216KA']), ('1693165jz69595j', ['31695JZ']), ('38k6387968ka963',
['79638KA']), ('0mb20m720872887', ['88720MB']), ('07l1433074307lv', ['14307LV']), ('135353ewtt13wt1', ['EWT1353']), ('158515f51fpfpf5', ['FPF515
8']), ('2433933mc43993m', ['24393MC']), ('179798fxrr17xr1', ['FXR1798']), ('033103151332510', ['1510332'])]
>>> task2d_cluster.count()
3
>>> task2d_cluster.collect()
[('', ['N/S', 'HX', 'S5', 'N/S', 'NS']), ('168216687l21', ['L.21687', 'L21687', 'L21687']), ('981jnpnp9p98', ['JNP981&', 'JNP981'])]
>>>
```

task2d_result = task2d_output

task2e

```
>>> task2e_output.count()
8994
>>> task2e_cluster.count()
1455
```

task2e_result = task2e_cluster

task2f

```
>>> task2f_output.count()
6866
>>> task2f_cluster.count()
2299
```

task2f_result = task2f_cluster

task2h

```
>>> task2h_output.count()
8994
>>> task2h_cluster.count()
15
>>> task2h_cluster.take(20)
[('nicholas st', ['ST NICHOLAS', 'ST NICHOLAS', 'ST NICHOLAS', 'ST NICHOLAS', 'ST NICHOLAS', 'ST NICHOLAS', 'ST NICHOLAS', 'ST NICHOLAS', 'ST NI
CHOLAS', 'ST NICHOLAS', 'NICHOLAS ST', 'NICHOLAS ST', 'NICHOLAS ST', 'NICHOLAS ST', 'NICHOLAS ST', 'NICHOLAS ST', 'NICHOLAS ST',
'NICHOLAS ST', 'NICHOLAS ST']), ('main ny ri st', ['MAIN ST RI NY', 'MAIN ST RI NY', 'MAIN ST RI NY NY']), ('100 street', ['100 STREET', '-100 S
TREET', '100 STREET']), ('blvd city coop', ['CO-OP CITY BLVD', 'CO-OP CITY BLVD', 'CO-OP CITY BLVD', 'COOP CITY BLVD']), ('21 east st', ['EAST 2
1 ST', 'EAST 21 ST ST', 'EAST 21 ST ST', 'EAST 21 ST']), ('ave', ['AVE', 'AVE AVE']), ('herbert st', ['HERBERT ST', 'HERBERT ST+', 'HERBERT ST']
), ('ave johns st', ['AVE ST JOHNS', 'ST JOHNS AVE']), ('91 st west', ['WEST 91 ST', 'WEST 91 ST ST']), ('ave new utrech', ['NEW UTRECH AVE', 'N
EW UTRECH AVE', 'NEW UTRECH AVE', 'NEW UTRECH AVE+']), ('10044 main ny st', ['MAIN ST NY NY 10044', 'MAIN ST NY NY 10044', 'MA
IN ST NY NY 10044', 'MAIN ST NY NY 10044', 'MAIN ST NY NY 10044', 'MAIN ST NY NY 10044']), ('18 ave', ['18 AVE', '18 AVE', '18 AVE', '18 AVE', '18
AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE
', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE+', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '
18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 AVE', '18 A
VE', '18 AVE', '18 AVE']), ('101 e st', ['E 101 ST', 'E 101 ST', 'E 101 ST', 'E 101 ST ST']), ('11 avenue', ['AVENUE 11', '11 AVENUE']), ('park
prospect sw', ['PROSPECT PARK S/W', 'PROSPECT PARK S/W', 'PROSPECT PARK S/W', 'PROSPECT PARK S/W', 'PROSPECT PARK SW', 'PROSPECT PARK SW', 'PROS
PECT PARK SW', 'PROSPECT PARK SW', 'PROSPECT PARK SW'])]
>>>
```

task2h_result = task2h_cluster