Part 1

2. Use the Apriori algorithm to generate frequent itemsets from the input data, with a minimum support equal to 0.15. In your answer, comment on the number of frequent itemsets and their sizes.

0	0.518892	(16...20)
1	0.251889	(21...25)
2	0.200252	(26...30)
3	0.600756	(3.2_3.6)
4	0.338791	(3.6_4.0)
5	0.220403	(cs)
6	0.308564	(engineering)
7	0.603275	(junior)
8	0.202771	(senior)
9	0.336272	(16...20, 3.2_3.6)
10	0.172544	(16...20, 3.6_4.0)
11	0.220403	(cs, 16...20)
12	0.340050	(16...20, junior)
13	0.178841	(16...20, senior)
14	0.210327	(3.2_3.6, 21...25)
15	0.210327	(engineering, 21...25)
16	0.251889	(junior, 21...25)
17	0.220403	(cs, 3.2_3.6)
18	0.210327	(engineering, 3.2_3.6)
19	0.408060	(3.2_3.6, junior)
20	0.163728	(3.6_4.0, junior)
21	0.210327	(engineering, junior)
22	0.220403	(cs, 16...20, 3.2_3.6)
23	0.197733	(16...20, 3.2_3.6, junior)
24	0.210327	(engineering, 3.2_3.6, 21...25)
25	0.210327	(junior, 3.2_3.6, 21...25)
26	0.210327	(junior, engineering, 21...25)
27	0.210327	(engineering, 3.2_3.6, junior)
28	0.210327	(junior, engineering, 3.2_3.6, 21...25)

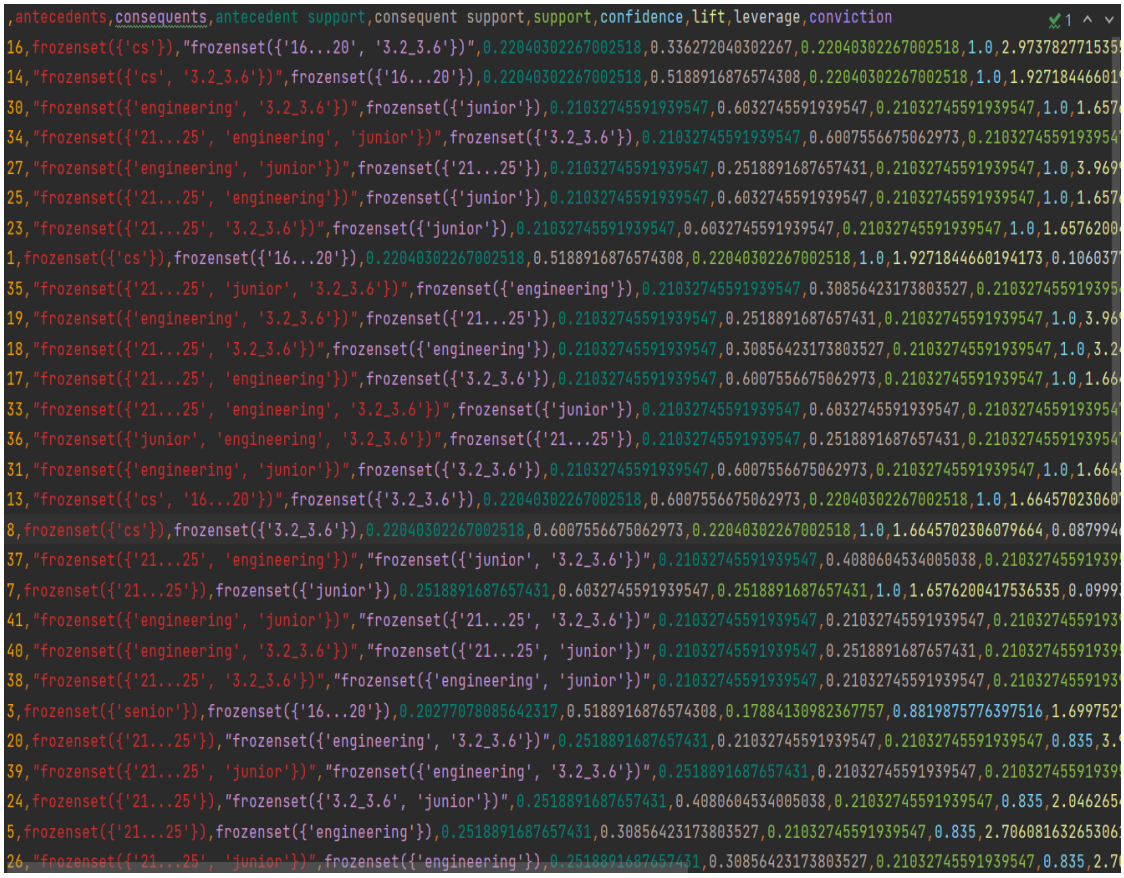
As can be seen from the result, The number of frequent items is 29. There are 9 frequent items that in size-1, 13 of that in size-2, 6 of that in size-3, and1 of that in size-4.

3.Does the attribute “count” have an impact on the Apriori algorithm’s results? Justifier your answer.

	support	itemsets
0	0.36	(16...20)
1	0.24	(2.8_3.2)
2	0.16	(21...25)
3	0.32	(26...30)
4	0.48	(3.2_3.6)
5	0.28	(3.6_4.0)
6	0.24	(French)
7	0.16	(M.S)
8	0.20	(Ph.D)
9	0.44	(junior)
10	0.16	(over 30)
11	0.28	(philosophy)
12	0.16	(senior)
13	0.20	(16...20, 3.2_3.6)
14	0.24	(16...20, junior)
15	0.16	(junior, 21...25)
16	0.16	(26...30, 3.2_3.6)
17	0.16	(26...30, Ph.D)
18	0.20	(26...30, philosophy)
19	0.20	(junior, 3.2_3.6)

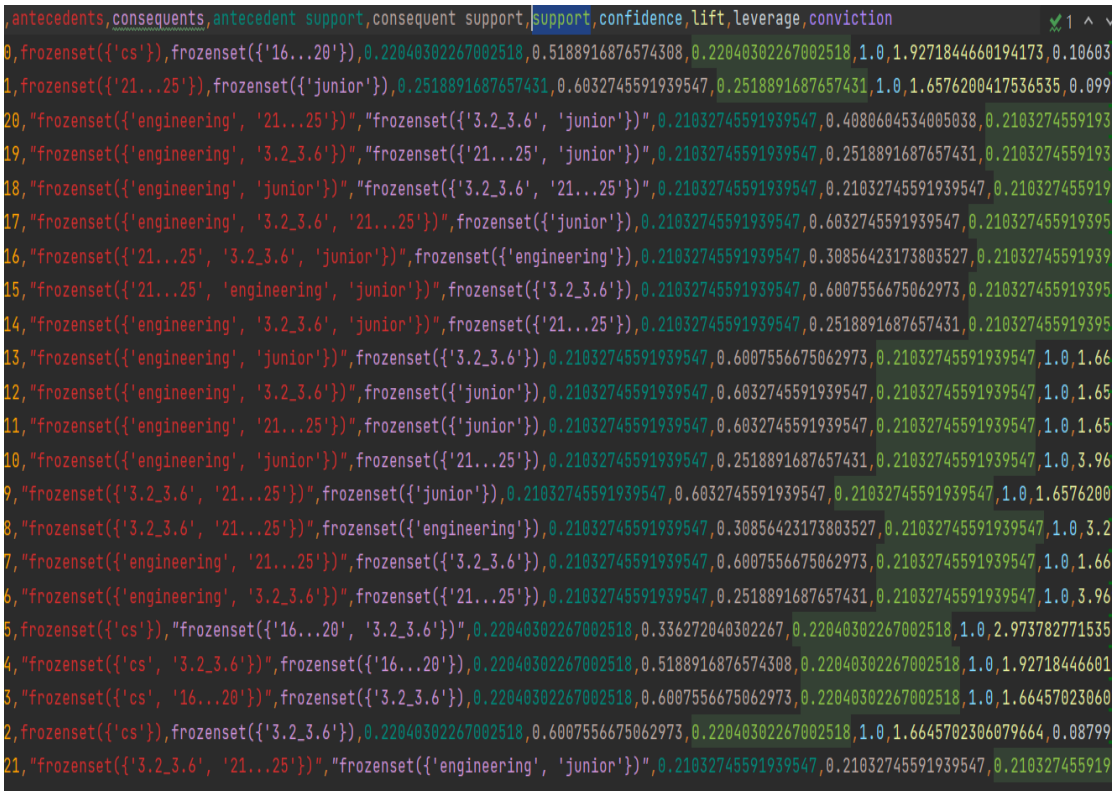
The attribute “count” does have an impact on the Apriori algorithm’s results. The different numbers of the same sample have an impact on the Association. For example, if we have 100 samples are [engineering, Ph.D, 26...30, 3.6_4.0], and 10 samples are [French, Ph.D,over 30,2.8_3.2], we will have a greater certainty that PHD is more related to engineering than French.

5.



```
,antecedents,consequents,antecedent support,consequent support,support,confidence,lift,leverage,conviction
16,frozenset({'cs'}),"frozenset({'16...20', '3.2_3.6'})",0.22040302267002518,0.336272040302267,0.22040302267002518,1.0,2.9737827715355
14,"frozenset({'cs', '3.2_3.6'})",frozenset({'16...20'}),0.22040302267002518,0.5188916876574308,0.22040302267002518,1.0,1.927184466017
30,"frozenset({'engineering', '3.2_3.6'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.21032745591939547,1.0,1.657
34,"frozenset({'21...25', 'engineering', 'junior'})",frozenset({'3.2_3.6'}),0.21032745591939547,0.6007556675062973,0.2103274559193954
27,"frozenset({'engineering', 'junior'})",frozenset({'21...25'}),0.21032745591939547,0.2518891687657431,0.21032745591939547,1.0,3.969
25,"frozenset({'21...25', 'engineering'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.21032745591939547,1.0,1.657
23,"frozenset({'21...25', '3.2_3.6'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.21032745591939547,1.0,1.6576200
1,frozenset({'cs'}),frozenset({'16...20'}),0.22040302267002518,0.5188916876574308,0.22040302267002518,1.0,1.927184466019413,0.106037
35,"frozenset({'21...25', 'junior', '3.2_3.6'})",frozenset({'engineering'}),0.21032745591939547,0.30856423173803527,0.2103274559193954
19,"frozenset({'engineering', '3.2_3.6'})",frozenset({'21...25'}),0.21032745591939547,0.2518891687657431,0.21032745591939547,1.0,3.96
18,"frozenset({'21...25', '3.2_3.6'})",frozenset({'engineering'}),0.21032745591939547,0.30856423173803527,0.21032745591939547,1.0,3.2
17,"frozenset({'21...25', 'engineering'})",frozenset({'3.2_3.6'}),0.21032745591939547,0.6007556675062973,0.21032745591939547,1.0,1.66
33,"frozenset({'21...25', 'engineering', '3.2_3.6'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.2103274559193954
36,"frozenset({'junior', 'engineering', '3.2_3.6'})",frozenset({'21...25'}),0.21032745591939547,0.2518891687657431,0.2103274559193954
31,"frozenset({'engineering', 'junior'})",frozenset({'3.2_3.6'}),0.21032745591939547,0.6007556675062973,0.21032745591939547,1.0,1.664
13,"frozenset({'cs', '16...20'})",frozenset({'3.2_3.6'}),0.22040302267002518,0.6007556675062973,0.22040302267002518,1.0,1.664570230607
8,frozenset({'cs'}),frozenset({'3.2_3.6'}),0.22040302267002518,0.6007556675062973,0.22040302267002518,1.0,1.6645702306079664,0.0879944
37,"frozenset({'21...25', 'engineering'})","frozenset({'junior', '3.2_3.6'})",0.21032745591939547,0.4080604534005038,0.2103274559193954
7,frozenset({'21...25'}),frozenset({'junior'}),0.2518891687657431,0.6032745591939547,0.2518891687657431,1.0,1.6576200417536535,0.0999
41,"frozenset({'engineering', 'junior'})","frozenset({'21...25', '3.2_3.6'})",0.21032745591939547,0.21032745591939547,0.2103274559193
40,"frozenset({'engineering', '3.2_3.6'})","frozenset({'21...25', 'junior'})",0.21032745591939547,0.2518891687657431,0.2103274559193954
38,"frozenset({'21...25', '3.2_3.6'})","frozenset({'engineering', 'junior'})",0.21032745591939547,0.21032745591939547,0.2103274559193
3,frozenset({'senior'}),frozenset({'16...20'}),0.20277078085642317,0.5188916876574308,0.17884130982367757,0.8819875776397516,1.699752
20,frozenset({'21...25'}),"frozenset({'engineering', '3.2_3.6'})",0.2518891687657431,0.21032745591939547,0.21032745591939547,0.835,3.
39,"frozenset({'21...25', 'junior'})","frozenset({'engineering', '3.2_3.6'})",0.2518891687657431,0.21032745591939547,0.2103274559193954
24,frozenset({'21...25'}),"frozenset({'3.2_3.6', 'junior'})",0.2518891687657431,0.4080604534005038,0.21032745591939547,0.835,2.046265
5,frozenset({'21...25'}),frozenset({'engineering'}),0.2518891687657431,0.30856423173803527,0.21032745591939547,0.835,2.706081632653061
26,"frozenset({'21...25', 'junior'})",frozenset({'engineering'}),0.2518891687657431,0.30856423173803527,0.21032745591939547,0.835,2.7
```

7.



```
,antecedents,consequents,antecedent support,consequent support,support,confidence,lift,leverage,conviction
0,frozenset({'cs'}),frozenset({'16...20'}),0.22040302267002518,0.5188916876574308,0.22040302267002518,1.0,1.927184466019413,0.10603
1,frozenset({'21...25'}),frozenset({'junior'}),0.2518891687657431,0.6032745591939547,0.2518891687657431,1.0,1.6576200417536535,0.099
20,"frozenset({'engineering', '21...25'})","frozenset({'3.2_3.6', 'junior'})",0.21032745591939547,0.4080604534005038,0.2103274559193
19,"frozenset({'engineering', '3.2_3.6'})","frozenset({'21...25', 'junior'})",0.21032745591939547,0.2518891687657431,0.2103274559193
18,"frozenset({'engineering', 'junior'})","frozenset({'3.2_3.6', '21...25'})",0.21032745591939547,0.21032745591939547,0.210327455919
17,"frozenset({'engineering', '3.2_3.6', '21...25'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.2103274559193954
16,"frozenset({'21...25', '3.2_3.6', 'junior'})",frozenset({'engineering'}),0.21032745591939547,0.30856423173803527,0.2103274559193
15,"frozenset({'21...25', 'engineering', 'junior'})",frozenset({'3.2_3.6'}),0.21032745591939547,0.6007556675062973,0.2103274559193954
14,"frozenset({'engineering', '3.2_3.6', 'junior'})",frozenset({'21...25'}),0.21032745591939547,0.2518891687657431,0.2103274559193954
13,"frozenset({'engineering', 'junior'})",frozenset({'3.2_3.6'}),0.21032745591939547,0.6007556675062973,0.21032745591939547,1.0,1.66
12,"frozenset({'engineering', '3.2_3.6'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.21032745591939547,1.0,1.65
11,"frozenset({'engineering', '21...25'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.21032745591939547,1.0,1.65
10,"frozenset({'engineering', 'junior'})",frozenset({'21...25'}),0.21032745591939547,0.2518891687657431,0.21032745591939547,1.0,3.96
9,"frozenset({'3.2_3.6', '21...25'})",frozenset({'junior'}),0.21032745591939547,0.6032745591939547,0.21032745591939547,1.0,1.6576200
8,"frozenset({'3.2_3.6', '21...25'})",frozenset({'engineering'}),0.21032745591939547,0.30856423173803527,0.21032745591939547,1.0,3.2
7,"frozenset({'engineering', '21...25'})",frozenset({'3.2_3.6'}),0.21032745591939547,0.6007556675062973,0.21032745591939547,1.0,1.66
6,"frozenset({'engineering', '3.2_3.6'})",frozenset({'21...25'}),0.21032745591939547,0.2518891687657431,0.21032745591939547,1.0,3.96
5,frozenset({'cs'}),"frozenset({'16...20', '3.2_3.6'})",0.22040302267002518,0.336272040302267,0.22040302267002518,1.0,2.973782771535
4,"frozenset({'cs', '3.2_3.6'})",frozenset({'16...20'}),0.22040302267002518,0.5188916876574308,0.22040302267002518,1.0,1.92718446601
3,"frozenset({'cs', '16...20'})",frozenset({'3.2_3.6'}),0.22040302267002518,0.6007556675062973,0.22040302267002518,1.0,1.664570230607
2,frozenset({'cs'}),frozenset({'3.2_3.6'}),0.22040302267002518,0.6007556675062973,0.22040302267002518,1.0,1.6645702306079664,0.08799
21,"frozenset({'3.2_3.6', '21...25'})","frozenset({'engineering', 'junior'})",0.21032745591939547,0.21032745591939547,0.210327455919
```

The age belong to‘21...25' may be related to 'junior' , all students who are ‘junior’ aging from 21 to 25.

The students who age belong to‘21...25' and major in engineer may be related to GPA of 3.2_3.6.

All cs students have a GPA between 3.2 and 3.6.

All cs students are between 16 and 20 years old.

All students with a GPA of 3.2 to 3.6 and ages 21-25 are engineering majors.

Part 2

1. a. Which attributes should be selected for data mining task?

ID should be deleted. Because everyone has a different ID, the association between ID and other attributes is meaningless.

b. Discretize the numeric attributes into 3 bins of equal width

Use pd.cut method.

2. Assume that the minimum support is equal to 20%. Use the FP-Growth algorithm to generate frequent all frequent itemsets. Comment in your report on the frequent itemsets, number, size, and

usefulness.

```
--------------------------------------------------------------------------------

      support                                                itemsets
0    0.500000                                                (FEMALE)
1    0.448333                                           (INNER_CITY)
2    0.500000                                                  (MALE)
3    0.506667                                                (NO_car)
4    0.241667                                       (NO_current_act)
..        ...                                                    ...
226  0.236667  (NO_mortgage, YES_save_act, YES_current_act, Y...
227  0.221667  (less than one child, NO_mortgage, YES_current...
228  0.223333  (less than one child, NO_mortgage, YES_save_ac...
229  0.216667  (YES_save_act, NO_pep, YES_current_act, YES_ma...
230  0.225000  (less than one child, YES_save_act, YES_curren...

[231 rows x 2 columns]
```

Number of frequent items:231

Size from 1 to 4.

```
0,0.5,frozenset({'FEMALE'})
2,0.5,frozenset({'MALE'})
```

Gender is usefulness.

5, Which confidence values that can return a set of rules of size at least equal to 10 (i.e., the number of rules). Explain in your report how did you identify these confidence values.

```
:,support,confidence,lift,leverage,conviction
'}),0.255,0.475,0.23,0.9019607843137255,1.898648090815274,0.10887500000000001,5.355000000000001
).892307692307692,1.8785425101214575,0.135625,4.874999999999997
)zenset({'YES_married'}),0.2366666666666666,0.66,0.2,0.8450704225352114,1.2804097311139566,0.04380
'ried'}),0.34833333333333333,0.66,0.285,0.8181818181818181,1.2396694214876032,0.055099999999999955,
,frozenset({'YES_married'}),0.2633333333333333,0.66,0.215,0.8164556962025317,1.2370540851553509,0.0
{'YES_married'}),0.32,0.66,0.26,0.8125,1.231060606060606,0.04879999999999998,1.8133333333333332
/e_act'})",frozenset({'YES_current_act'}),0.2783333333333333,0.7583333333333333,0.22333333333333333
'rent_act'}),0.3283333333333333,0.7583333333333333,0.2633333333333333,0.8020304568527918,1.05762258
,34,0.7583333333333333,0.27,0.7941176470588235,1.0471881060116355,0.012166666666666659,1.1738095238
'}),0.29,0.7583333333333333,0.23,0.7931034482758622,1.045850701023115,0.010083333333333336,1.1680555
,31833333333333336,0.69,0.25166666666666665,0.7905759162303664,1.1457621974353136,0.032016666666666
```

Value of confidence need to below than 0.7931034482758622

0.7931034482758622. Sort the rules in descending order of confidence and find the tenth confidence value. If we need to return a set of rules of size at least equal to 10, The value of confidence must be less than or equal to this value.

7.

['youth']→[‘low’](0.29,0.892)

Young people may have less income, this shows that young people may need credit services more, and banks may recommend credit cards to young people instead of wealth management products.

[elder people]→ [YES_save_act](0.252,0.791)

Similarly, it may be more effective to recommend wealth management products to the elderly.

[‘NO_mortgage', 'YES_car']→ [YES_save_act](0.235,0.727)

People who don’t have a mortgage and have a car are more likely to have saving action. It will be easier for banks to sell savings products to these people

[NO_mortgage]→[YES_save_act](0.45,0.690)

People who don’t have mortgages are likely to have deposits