**Q1**

**a)**Filtered out count attribute with select attributes to get

|  |  |  |  |
| --- | --- | --- | --- |
| **major** | **status** | **age** | **gpa** |
| French | M.A | over 30 | 2.8\_3.2 |
| cs | junior | 16...20 | 3.2\_3.6 |
| physics | M.S | 26...30 | 3.2\_3.6 |
| engineering | Ph.D | 26...30 | 3.6\_4.0 |
| philosophy | Ph.D | 26...30 | 3.2\_3.6 |
| French | senior | 16...20 | 3.2\_3.6 |
| chemistry | junior | 21...25 | 3.6\_4.0 |
| cs | senior | 16...20 | 3.2\_3.6 |
| philosophy | M.S | over 30 | 3.6\_4.0 |
| French | junior | 16...20 | 2.8\_3.2 |
| philosophy | junior | 26...30 | 2.8\_3.2 |
| philosophy | M.S | 26...30 | 3.2\_3.6 |
| French | junior | 16...20 | 3.2\_3.6 |
| math | senior | 16...20 | 3.6\_4.0 |
| cs | junior | 16...20 | 3.2\_3.6 |
| philosophy | Ph.D | 26...30 | 3.6\_4.0 |
| philosophy | senior | 26...30 | 2.8\_3.2 |
| French | Ph.D | over 30 | 2.8\_3.2 |
| engineering | junior | 21...25 | 3.2\_3.6 |
| math | Ph.D | 26...30 | 3.2\_3.6 |
| chemistry | junior | 16...20 | 3.6\_4.0 |
| engineering | junior | 21...25 | 3.2\_3.6 |
| French | M.S | over 30 | 3.2\_3.6 |
| philosophy | junior | 21...25 | 2.8\_3.2 |
| math | junior | 16...20 | 3.6\_4.0 |

**b)** Confidence value =0.9 results In

W-Apriori

=======

Minimum support: 0.1 (3 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 3

Best rules found:

Hence no rules are found with such a high value for confidence.

**c)**Confidence values=0.7

W-Apriori

Apriori

=======

Minimum support: 0.1 (3 instances)

Minimum metric <confidence>: 0.7

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 3

Best rules found:

**1. status=Ph.D 5 ==> age=26...30 4 conf:(0.8)**

**2. age=over 30 4 ==> major=French 3 conf:(0.75)**

Hence proper rules are generated with confidence 0.7

**Q2**

1. Filtering out CAR attribute gives:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **APPARTEMENT** | **VILLA** | **POOR** | **AVERAGE** | **RICH** |
| true | false | false | false | false |
| false | false | false | true | true |
| false | true | false | true | true |
| true | false | false | false | false |
| true | true | false | true | false |
| true | true | false | false | true |
| true | false | true | false | true |
| true | false | false | false | false |
| false | true | true | false | true |
| true | true | true | false | false |
| false | false | false | true | false |
| false | true | false | false | false |
| true | false | true | true | false |
| true | false | true | true | false |
| false | false | false | false | true |
| false | false | true | false | false |
| true | false | false | true | true |
| false | true | false | true | true |
| true | false | false | false | false |
| true | true | false | true | false |
| true | true | false | false | true |
| true | false | true | false | true |
| true | false | false | false | false |
| false | true | true | false | true |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |

b,c,d) With default confidence value there are no association rules found. Possibly because the default is too high for the dataset

e) With confidence value of 0.6 you get the association rules

[APARTMENT]🡺[RICH](confidence=0.667)

It implies if you have an apartment you are probably rich with a confidence of 66.7%

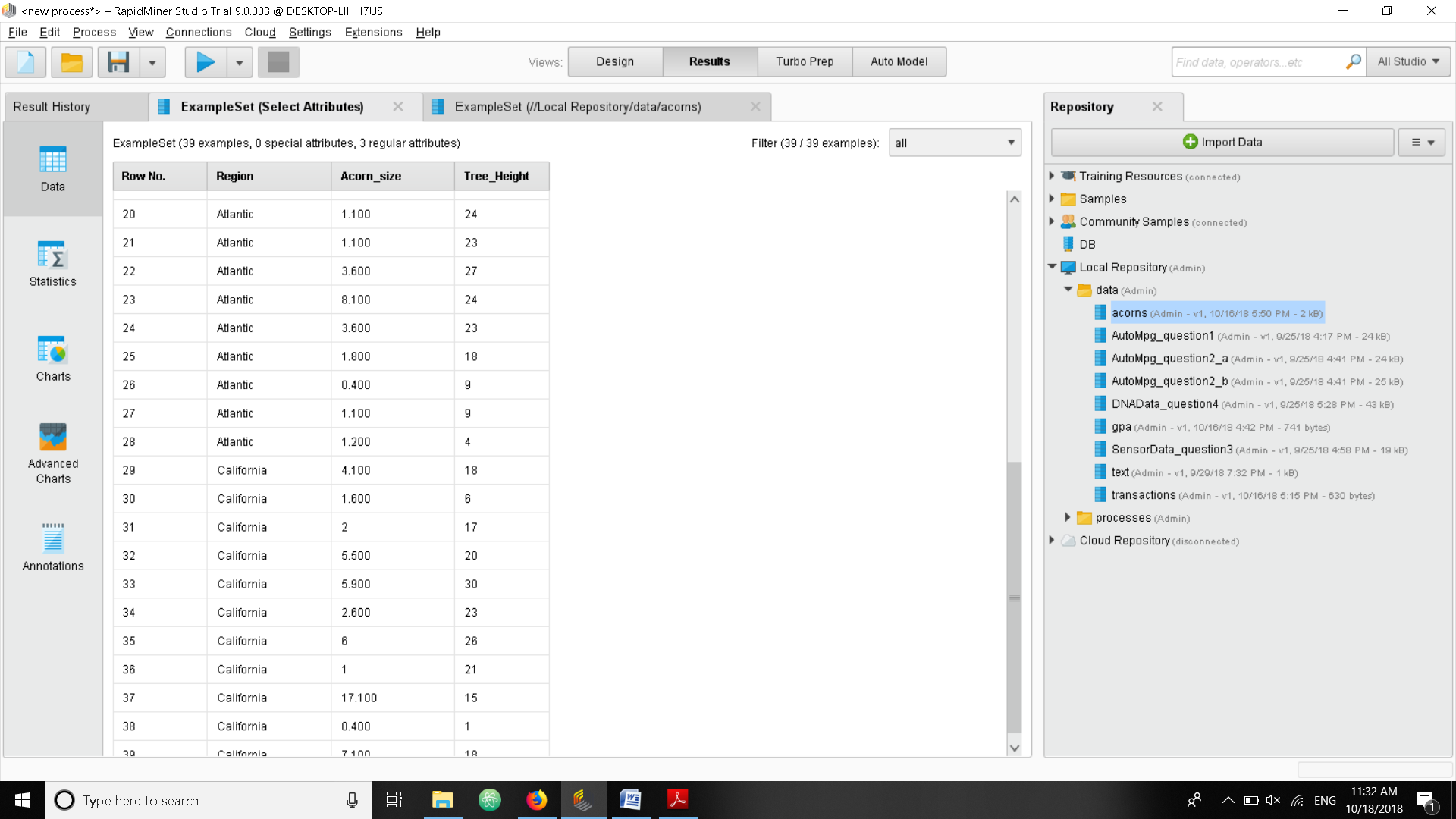
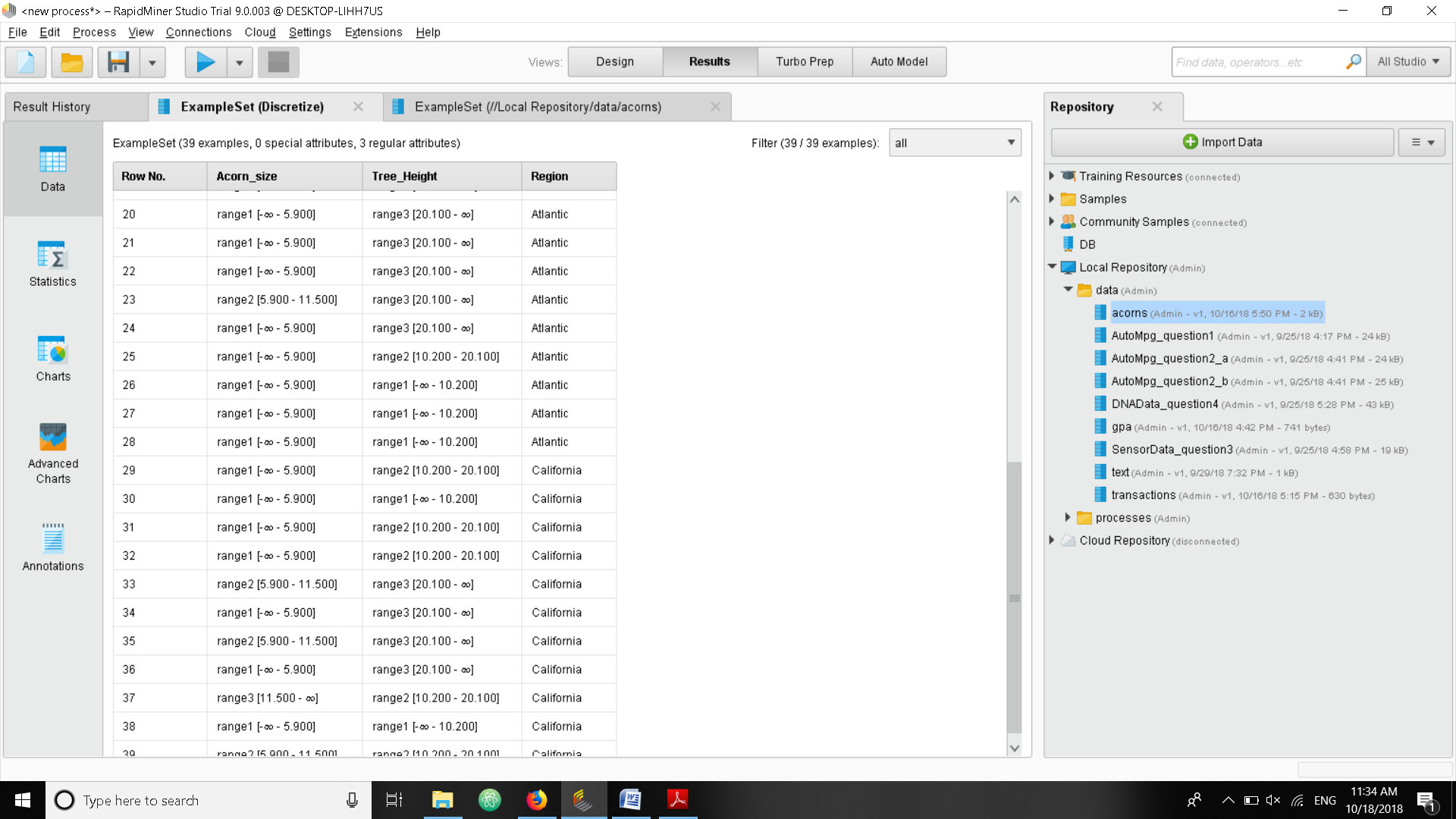
[VILLA]🡺[RICH](Confidence=0.600)

It implies if you have a villa you are rich with a confidence of 60%

**Q3**

a)with filter (sample of) dataset is:

1. b)



b)Applying binning we get----------------------🡪

c)Applying W-apriori algorithm we get no frequent itemsets because of the high default value.

d,e)With confidence of 0.4 we get 2 rules

Best rules found:

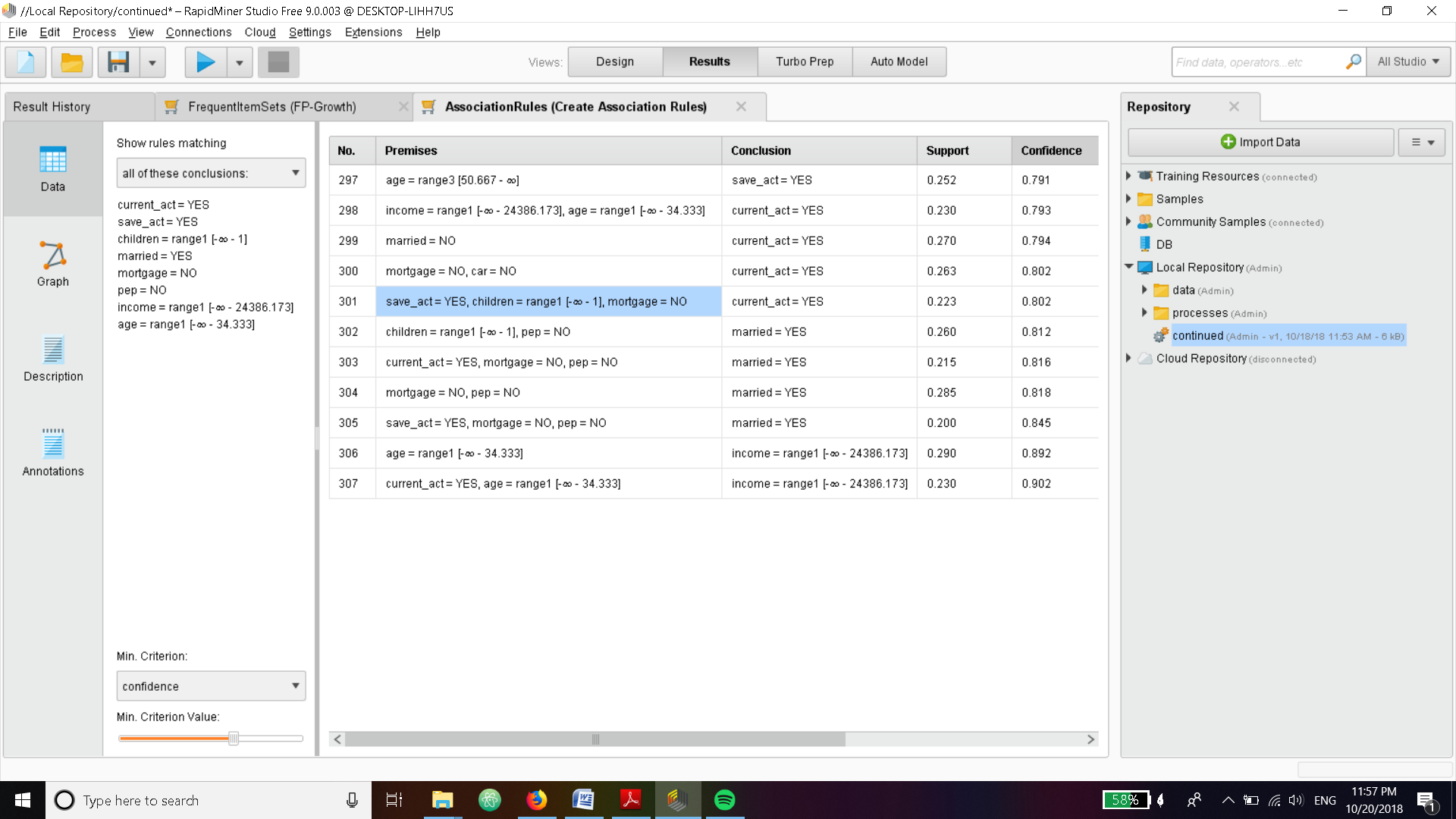
1. Region=California 11 ==> Tree\_Height=range2 [10.200 - 20.100] 5 conf:(0.45)

2. Tree\_Height=range2 [10.200 - 20.100] 11 ==> Region=California 5 conf:(0.45)

This confidence is lower than 50 % so one cant confidently decide if the rule holds or not.

**Q4)**

After applying the steps we get the top rules as



The Highlighted rule (301) looks interesting for the bank. If a customer has one child a saving account and no mortgages he will most likely also have a current account with the bank. The bank can incentivize tactics to target such customers since they bring a lot of business to bank. The bank can also say not to invest in advertising mortgages for such customers.

The rule 299 states that unmarried customers are very likely to have current\_acct. Bank can capitalize on this by creating schemes based on this data to maximize gain based on interests.