

Solution :

<https://github.coecis.cornell.edu/cg633/ORIE5270>

Problem1:

(c)

Average min max
[(3.051006711409397, 2.0, 4.4)]

(d)

Class	Average	min	max
Iris-setosa	(3.416326530612245,	2.3,	4.4)
Iris-versicolor	(2.7700000000000005,	2.0,	3.4)
Iris-virginica	(2.9739999999999998,	2.2,	3.8)

From the result we know that the Sepal Width of Iris-versicolor and Iris-virginica are below the total average of Sepal Width, and Sepal Width of Iris-setosa is beyond that level

Problem2:

The two output is below:

```
[('Paul Novak', '2009-22-11'), ('Paul Novak', '2009-29-11'), ('Terry Neils', '2009-28-11'), ('Terry Neils', '2009-29-11'), ('Jack Fonda', '2009-02-12')]  
[('Paul Novak', '2009-22-11'), ('Paul Novak', '2009-29-11'), ('Terry Neils', '2009-28-11'), ('Terry Neils', '2009-29-11'), ('Jack Fonda', '2009-02-12'), ('Tom Willis', None)]
```

The results show that

(1). Two results except second output last line are the same, because the property of join, that take the same intersection of two tables and join with them in the same CustomerID. Since in table2 CustomerID 1 and 2 appear twice, so when join with table1, it will have two line for a customer.

(2). inner join does not contain 'Tom Willis', since table 2 do not have customer 'Tom Willis' information. Therefore, in the inner join table, it takes the intersection of two tables.

While in the left out join, means whatever the element in table2, table2 should maintain all the information, however table2 does not have customer4 information, so last line ('Tom Willis', None) second element is None.