

## STT 811

### In-Class Assignment 4

On D2L is the complaints dataset. Using the strategy described in the slides, modify the dataset by

- (a) Summarizing the number of complaints by customer and year.

```
complaintsSummary <- sqldf("SELECT *,COUNT(*) as ComplaintCount
                             FROM complaints
                             GROUP BY customerID, ComplaintYear
                             ORDER BY ComplaintYear, customerID")
```

- (b) Add a column for predictive year

```
predictions <- sqldf("SELECT *, ComplaintYear+1 as PredictiveYear
                      FROM complaintsSummary")
```

- (c) Add a column which includes all complaints made by the customer on or before the year field (not predictive year).

```
ComCount <- sqldf("SELECT CustomerID, PredictiveYear,
                    SUM (SUM (ComplaintCount)) OVER (ORDER BY PredictiveYear) AS
ComplaintHist
                  FROM predictions
                  GROUP BY PredictiveYear
                  ORDER BY PredictiveYear")
```

```
View(ComCount)
```