problem: improve bidding

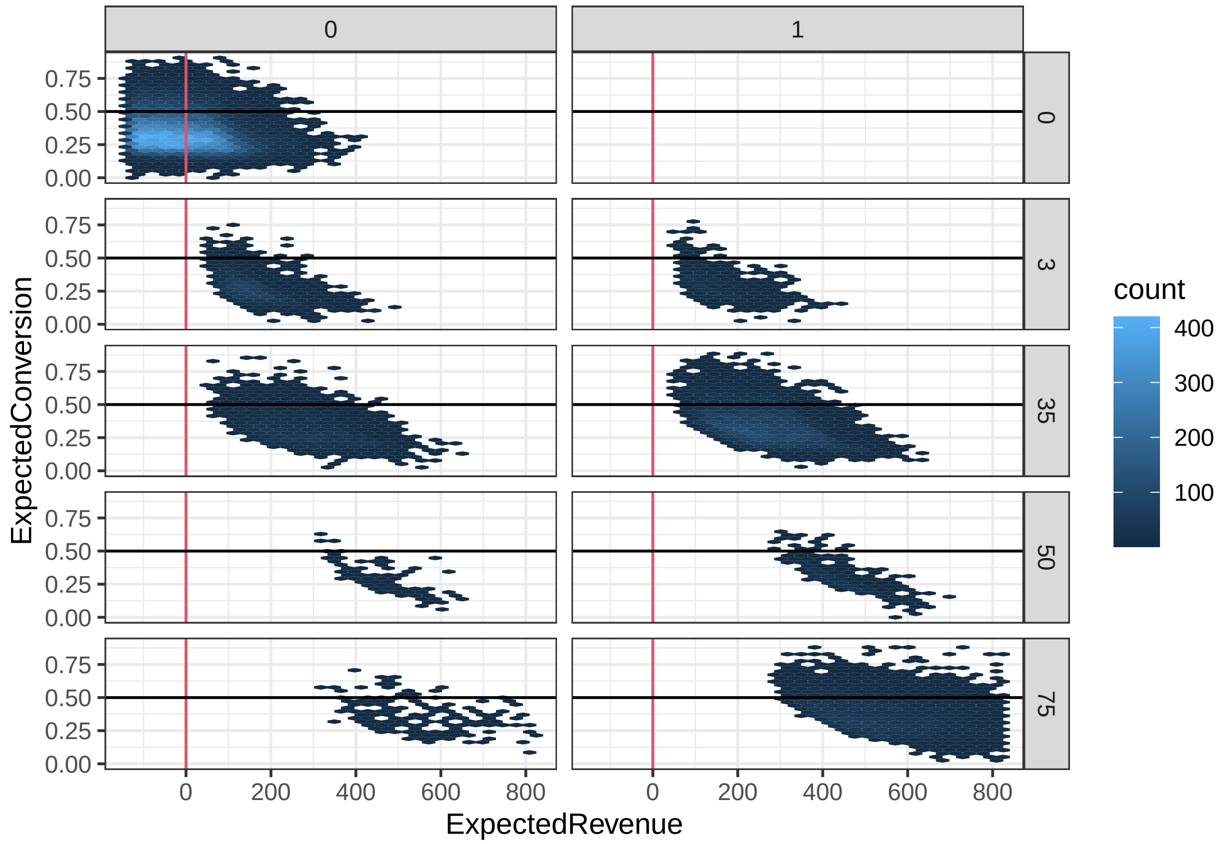
Derek Lukacsko

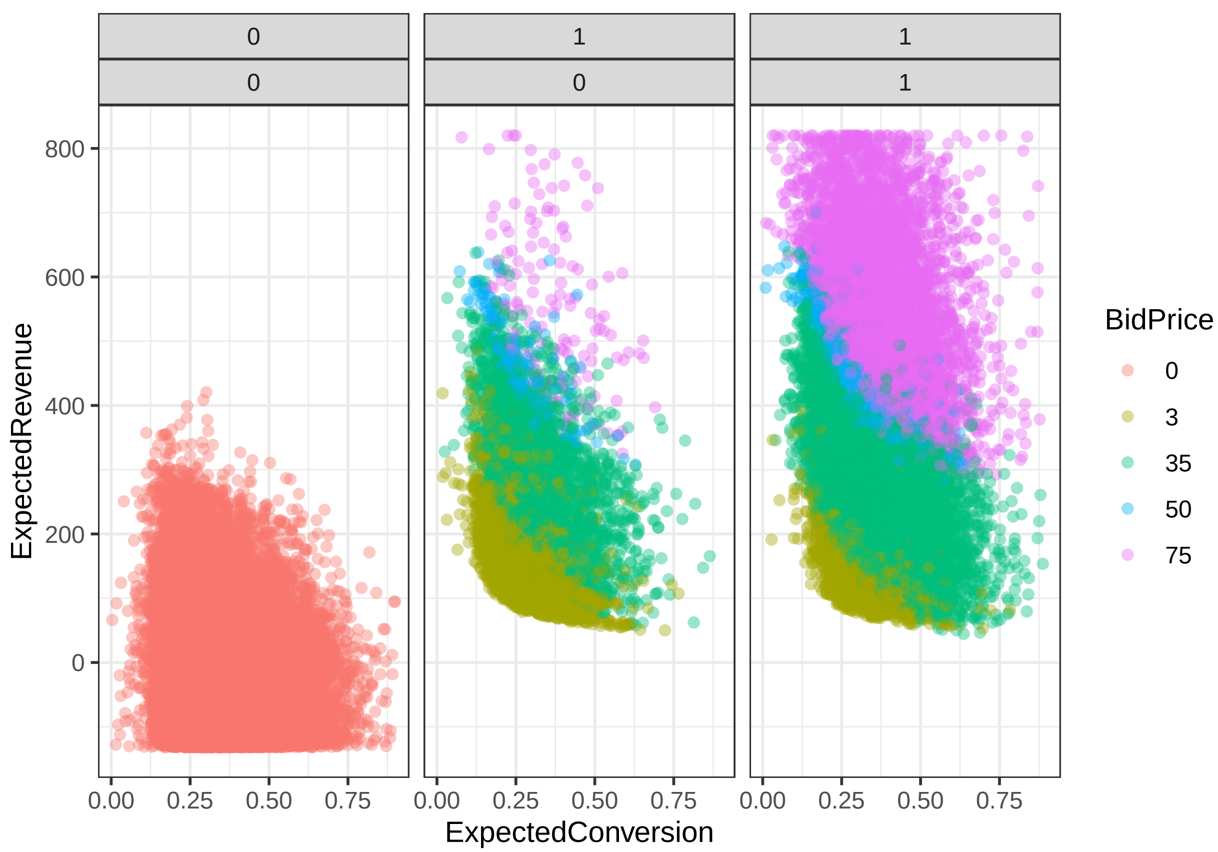
March 11 2021

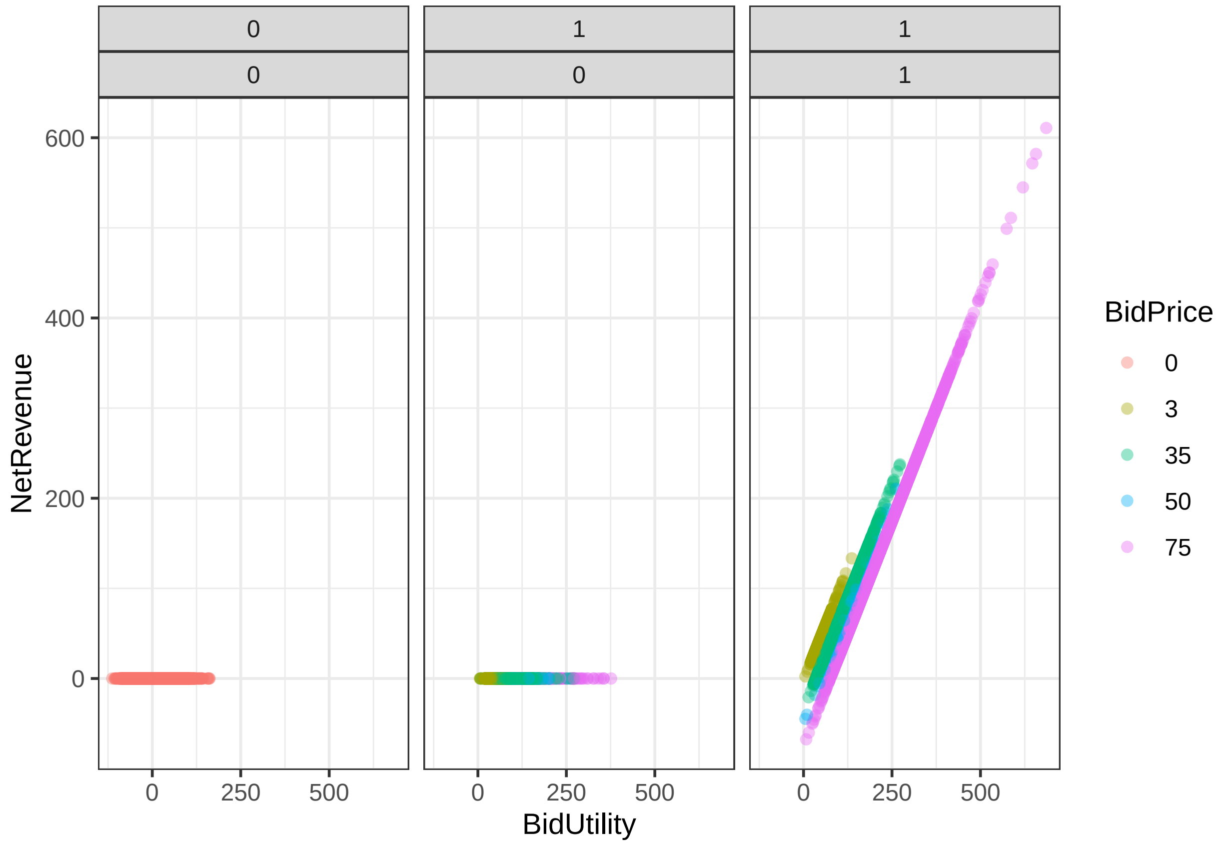
1. Setup

- General comments
  - This is a two stage problem: we don't know if bids will be accepted OR whether bids will convert to leads. An estimate was provided for the latter, but not the former. So I spent time thinking about this problem from a bayesian perspective but I think this wasn't productive because I did not (yet) bring back P(Bid X amount | X amount accepted) to the objective, maximizing revenue. I thought about if I had actual profit numbers for each bid, I could set this up as a two-stage supervised problem: (1) revenue <- BidUtility, (2) BidUtility <- BidAmount + priors. I could train a portion of the data where bids were made, test/dev it on the remainder, and fit it to the whole set after. I also thought about bootstrapping the estimated bid priors and expected lead acceptance. My biggest issue is that I don't see how I can provide a model that improves revenue without observed revenue values. I can explore the above (and bid optimization which I started skimming journal papers), but I don't see how to go beyond making the model merely analytic and how I can fully leverage the leads where bids weren't made. This is my first exposure to this type of problem. I limited myself to one day to work on this but look for the opprotunity to move forward with some advice. Look forward to hearing your feedback.

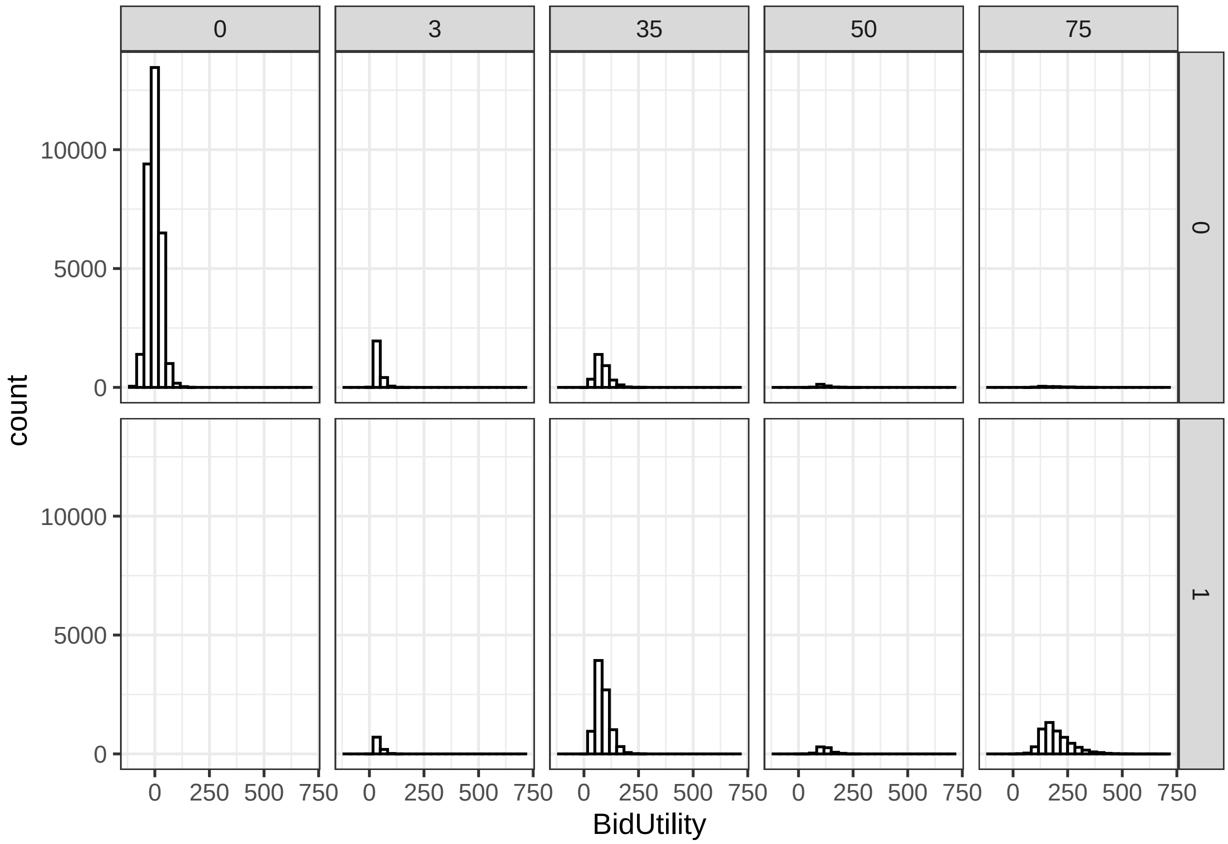
2. Looking around



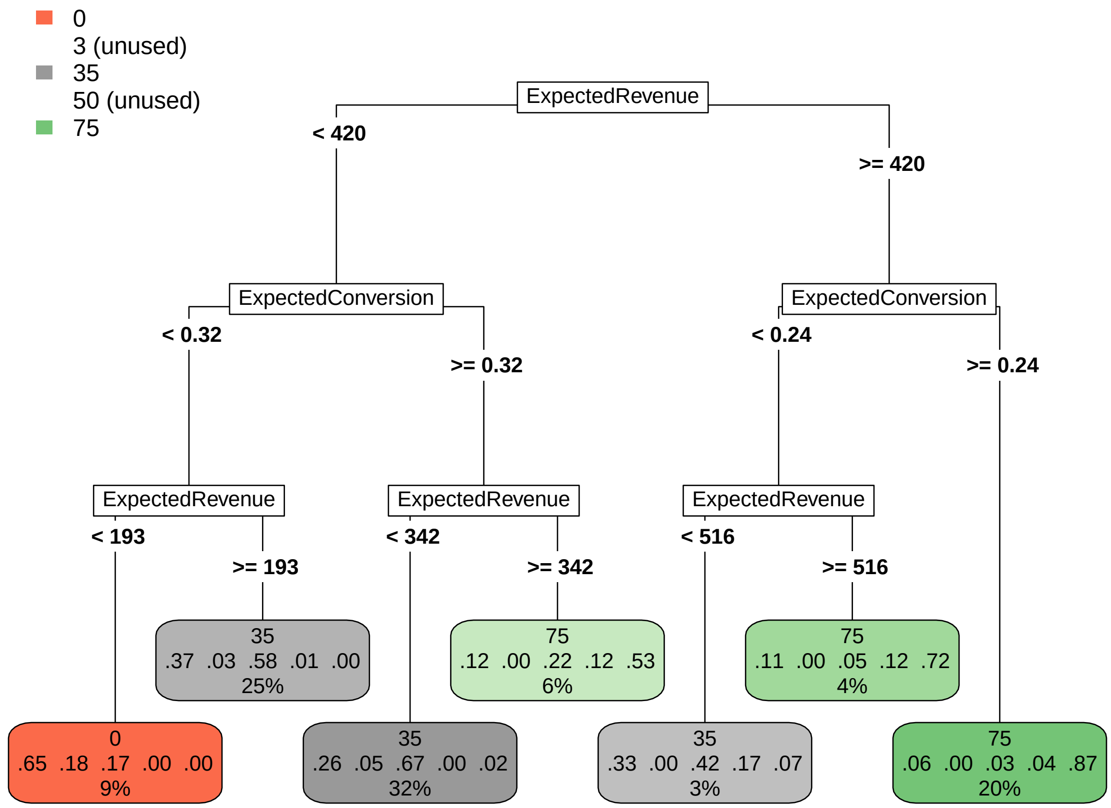


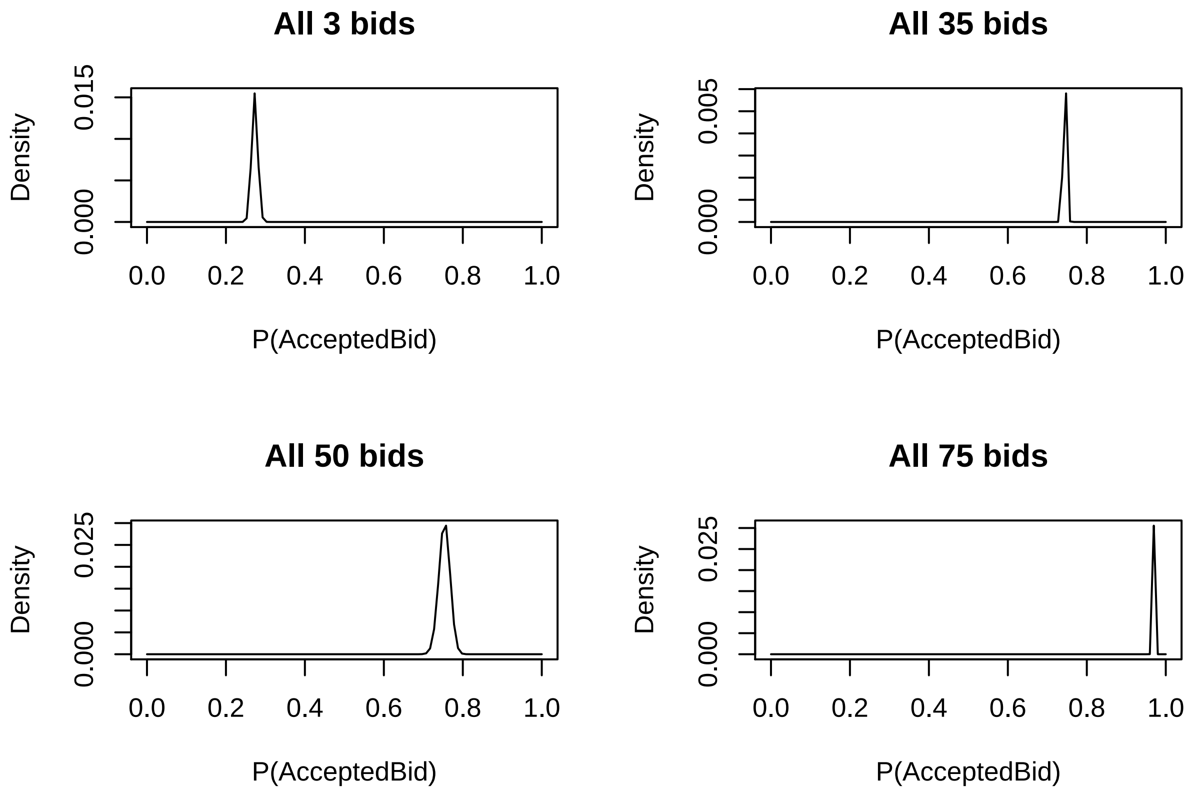


0
0
1
0
1
1
600
400
200
0
NetRevenue
0
250
500
0
250
500
0
250
500
BidUtility
BidPrice
0
3
35
50
75




```
## n= 21945
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 21945 13000 35 (0.27 0.042 0.41 0.032 0.25)
##    2) ExpectedRevenue< 4.2e+02 15851  7400 35 (0.34 0.058 0.53 0.018 0.055)
##      4) ExpectedConversion< 0.32 7500  4000 35 (0.44 0.074 0.47 0.011 0.002)
##        8) ExpectedRevenue< 1.9e+02 2052   730 0 (0.65 0.18 0.17 0 0) *
##        9) ExpectedRevenue>=1.9e+02 5448  2300 35 (0.37 0.032 0.58 0.015 0.0028) *
##      5) ExpectedConversion>=0.32 8351  3400 35 (0.24 0.043 0.59 0.024 0.1)
##       10) ExpectedRevenue< 3.4e+02 6972  2300 35 (0.26 0.052 0.67 0.0042 0.017) *
##       11) ExpectedRevenue>=3.4e+02 1379   640 75 (0.12 0 0.22 0.12 0.53) *
##    3) ExpectedRevenue>=4.2e+02 6094  1500 75 (0.099 0.00016 0.083 0.068 0.75)
##      6) ExpectedConversion< 0.24 1677   970 75 (0.21 0.0006 0.22 0.15 0.42)
##       12) ExpectedRevenue< 5.2e+02 760   440 35 (0.33 0.0013 0.42 0.17 0.07) *
##       13) ExpectedRevenue>=5.2e+02 917   260 75 (0.11 0 0.055 0.12 0.72) *
##      7) ExpectedConversion>=0.24 4417   560 75 (0.057 0 0.03 0.039 0.87) *
```


0
3 (unused)
35
50 (unused)
75
ExpectedRevenue
< 420
>= 420
ExpectedConversion
< 0.32
>= 0.32
ExpectedConversion
< 0.24
>= 0.24
ExpectedRevenue
< 193
>= 193
ExpectedRevenue
< 342
>= 342
ExpectedRevenue
< 516
>= 516
35
.37 .03 .58 .01 .00
25%
75
.12 .00 .22 .12 .53
6%
75
.11 .00 .05 .12 .72
4%
0
.65 .18 .17 .00 .00
9%
35
.26 .05 .67 .00 .02
32%
35
.33 .00 .42 .17 .07
3%
75
.06 .00 .03 .04 .87
20%

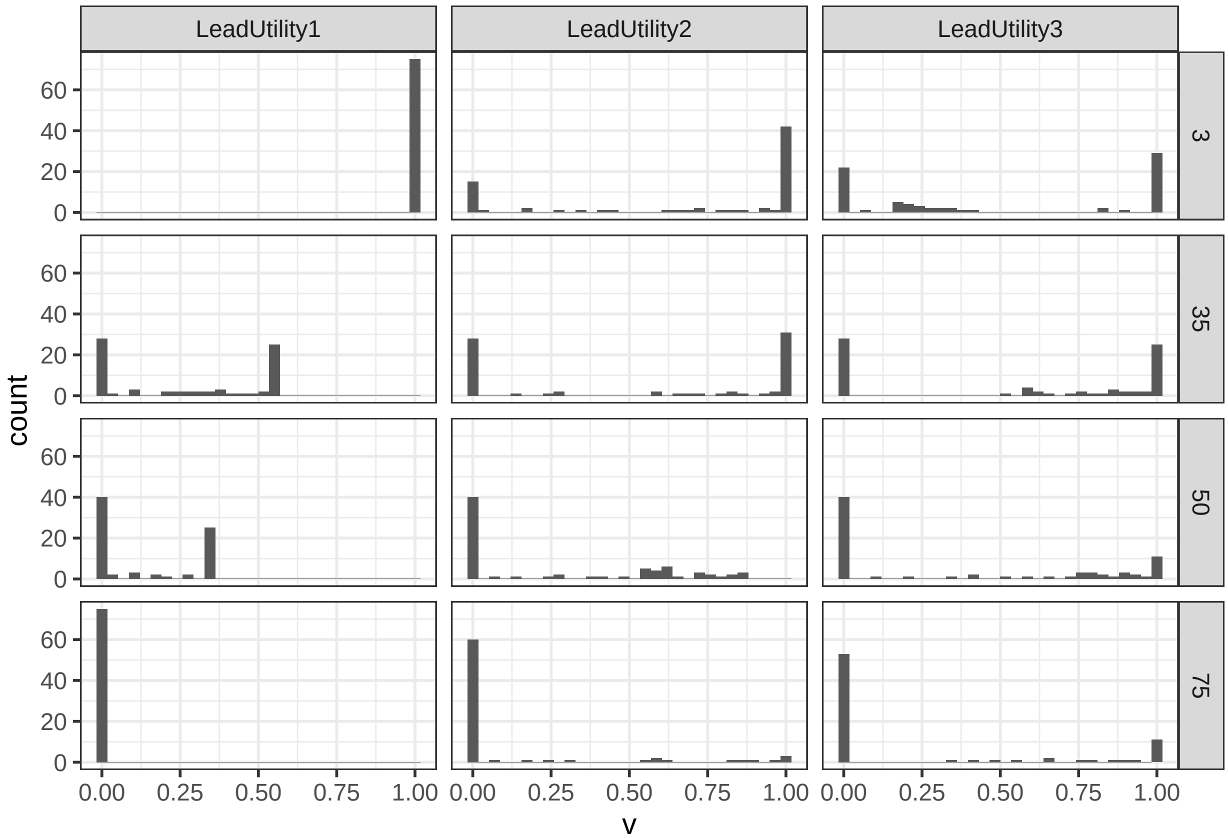


3. Solution attempt

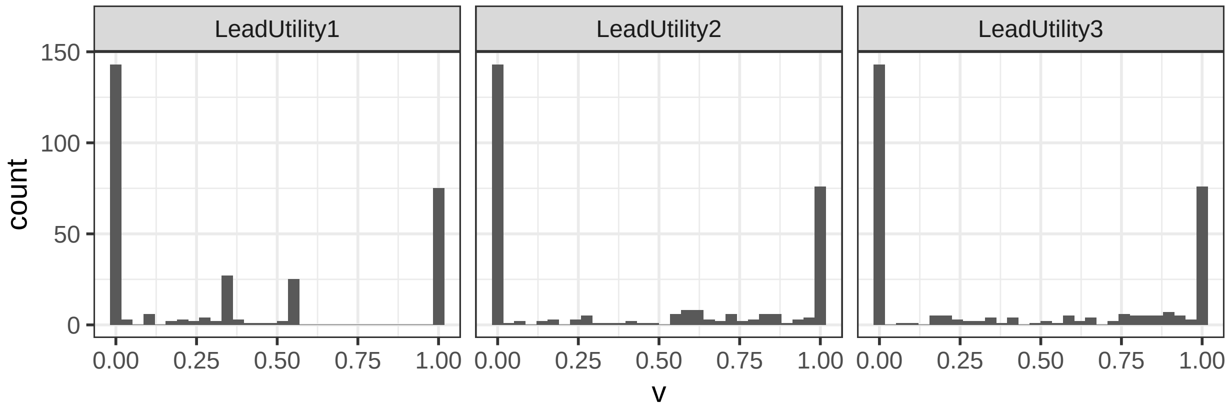
- The leads selected depend on the expected net revenue and the likelihood of acceptance. The former is given; the latter, inferred from the data. Three utility functions are defined. Utility1 only accounts for expected net revenue (labelled BidUtility here), while the latter two account for net revenue and likelihood of bid acceptance and discount BidUtility by the inferred amount.
  - utility 1 assumes we should bid on those leads with the highest expected payoff
  - utility 2 takes into account the likelihood of a bid being accepted given the bid amount using the data provided
  - utility 3 takes into account the likelihood of a bid being accepted given the bid amount naively assuming a linear relationship between bid amount and likelihood of acceptance

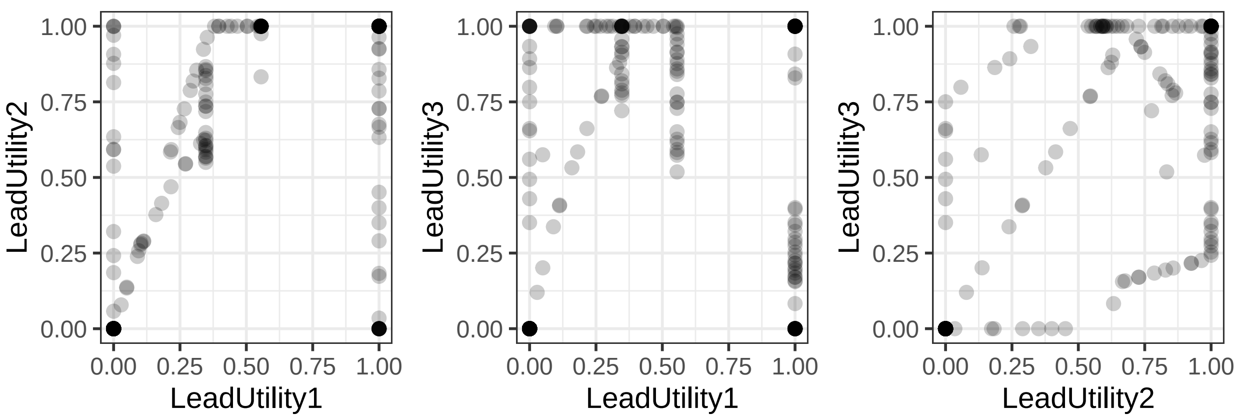
```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```

```
## Warning: Removed 12 rows containing non-finite values (stat_bin).
```




```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.

## Warning: Removed 12 rows containing non-finite values (stat_bin).

## Warning: Removed 4 rows containing missing values (geom_point).

## Warning: Removed 4 rows containing missing values (geom_point).

## Warning: Removed 4 rows containing missing values (geom_point).
```





- Observations
  - all three methods find utility in variously priced bids, but all choose 3
  - there is significant variability in which leads are chosen (see lower 3 facets; the darkest regions are 00 10 01 and 11)
  - all three produce extremes estimates (0 and 1) making it clear which leads we should NOT bid on but unclear which potential bids are most favorable
  - conclusion: these only seem useful for narrowing the search of what leads NOT to bid on but are extremely