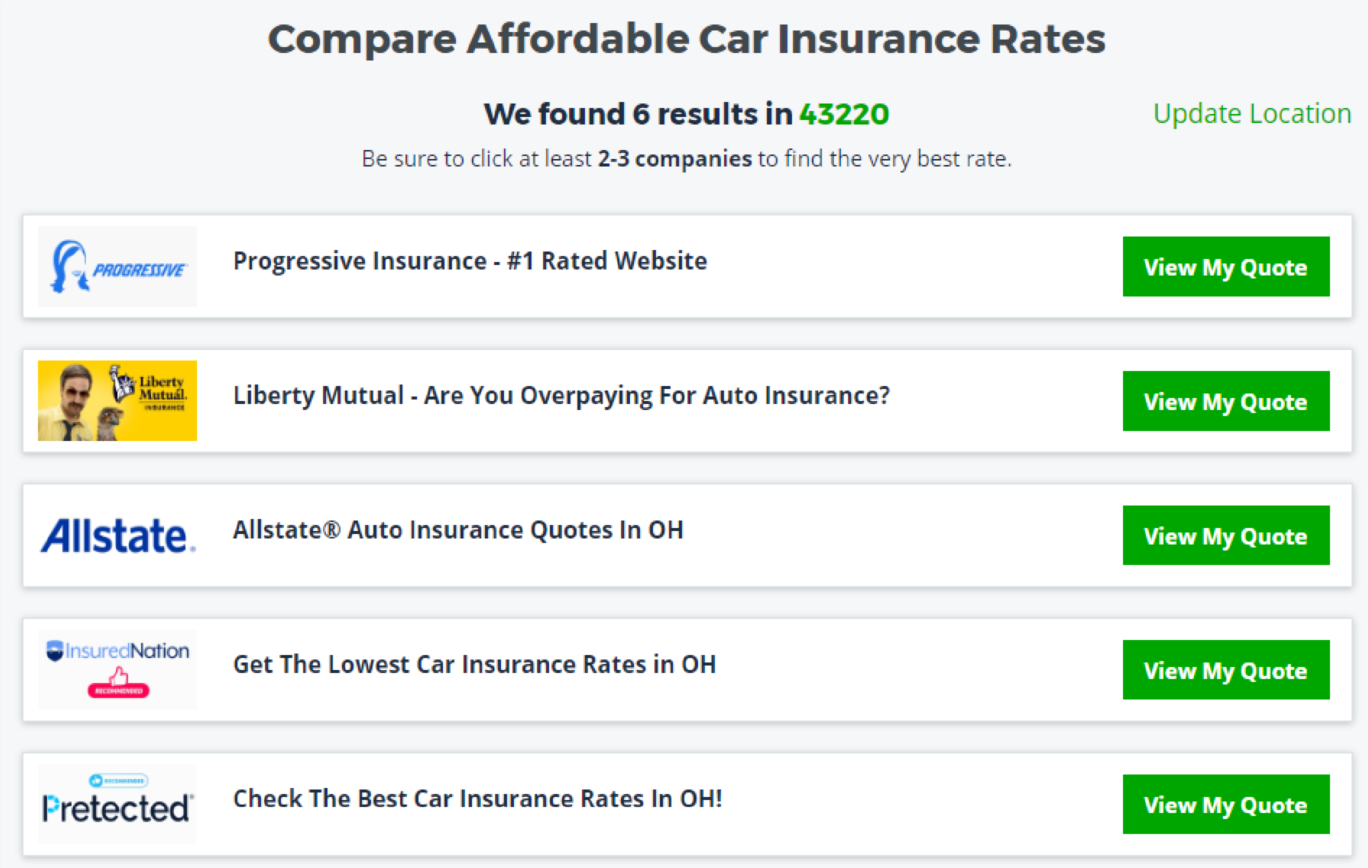
# TEAM 12: “THE ROOTS” ROOT INSURANCE CHALLENGE

By: MinSub Kim, Simon Li, Thomas Mussmann, Sri Ravindran, Angelo Taranto, and Shuo Xu.

# VERTICAL SEARCH WEBSITES FOR CAR INSURANCE




- The website lists ads based on the consumer's characteristics.
- The order of ads is determined by an online auction: higher bid, higher **rank**
- The companies can set different bid prices for different **types** of customers.
- The companies only pay if their ad is **clicked**.
- Our company currently bids \$10 per customer and gets 545 **sales**.
- <u>Goal</u>: Find the best bidding strategy which optimizes the **cost** per **sale**, with at least 400 **sales** for every 10,000 ads.

# OUR APPROACH

**Data Analysis**

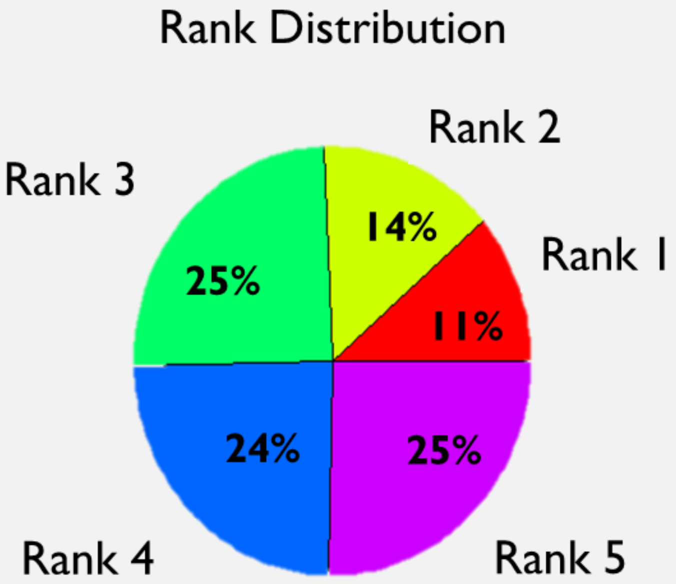
**Set Model**

**Optimization**

# DATA DESCRIPTION

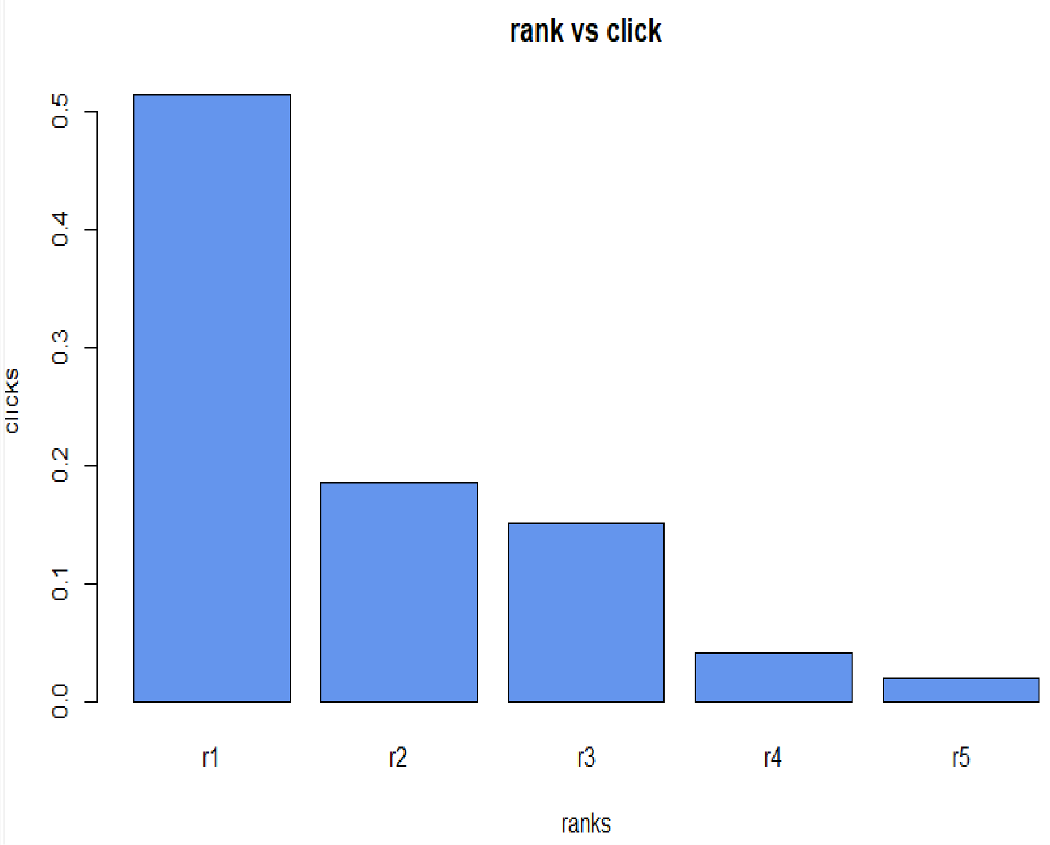
- Dataset contains 10,000 observation
- Consumer Characteristics
  - Are Currently Insured?
  - How many drivers in the policy: 1 or 2
  - How many cars they have: 1 or 2
  - Is the consumer single or married
- Given these characteristics, we group these consumers into 16 types: one for each combination
- Our ranking given our bid of \$10: {1,2,3,4,5}

$$2^4 = 16$$



# DATA ANALYSIS: P(CLICK)

## Plot



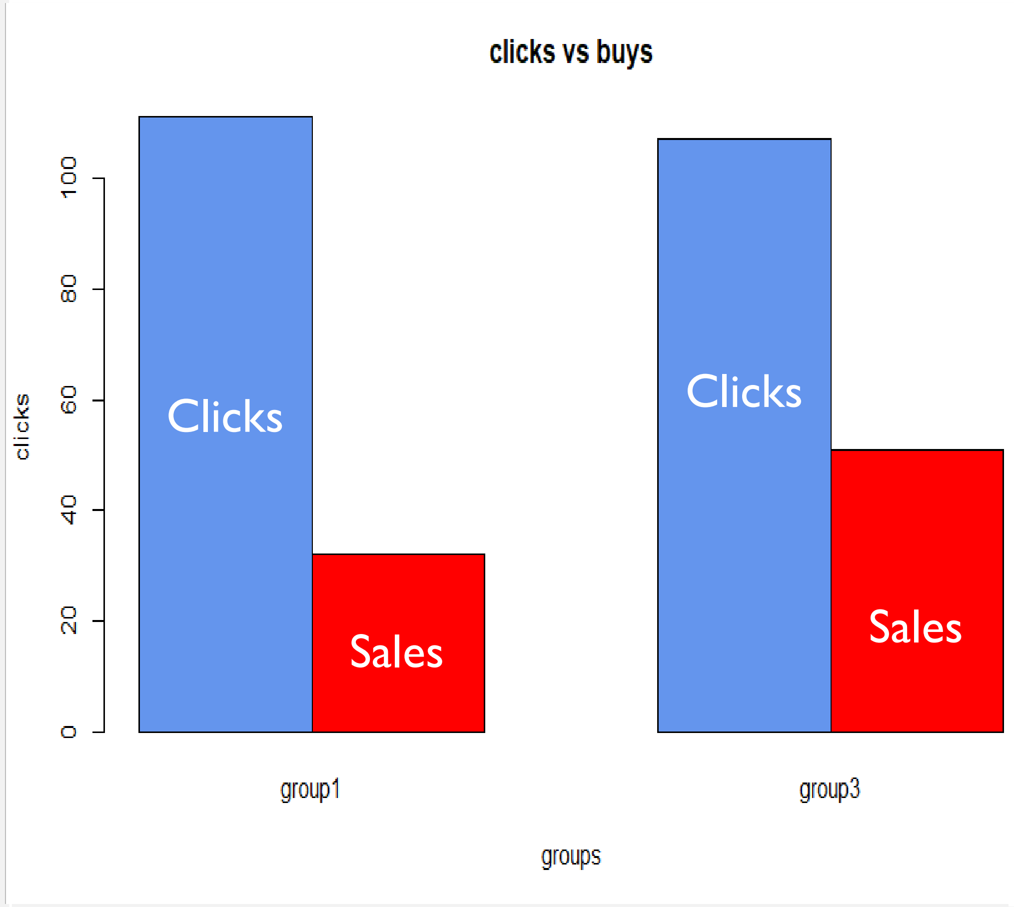

## Description

Higher rank ads → Higher number of clicks.
LOGIT test found type had insignificant effect on clicks.

| ranks | Prob.clicks |
|---|---|
| 1 | 0.51 |
| 2 | 0.19 |
| 3 | 0.15 |
| 4 | 0.04 |
| 5 | 0.02 |

# DATA ANALYSIS: P(SALE|CLICK)

## Plot




## Description

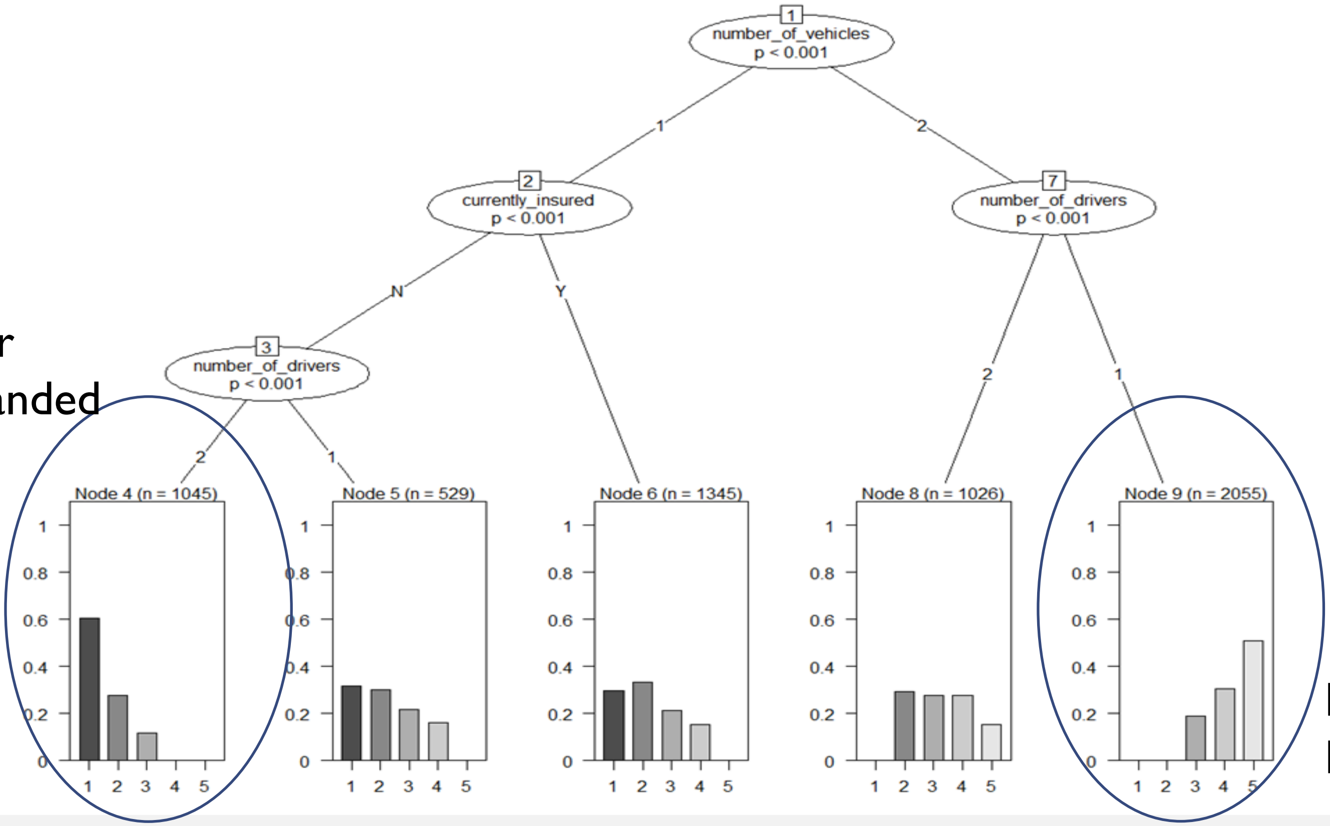
Policy sold primarily depends on the type of customers, not only the clicks.
Group1: {veh:2, driv:1, Ins:Y, Marital:M}
Group3: {veh:2, driv:1, Ins:N, Marital:M}

| Group | Clicks | Policy.sold |
|---|---|---|
| 1 | 111 | 32 |
| 3 | 107 | 51 |

# DATA ANALYSIS: P(RANK|TYPE)

The tree diagram provides useful information about the ad's rank for the different type of customers for cost $10.

Lower Demanded

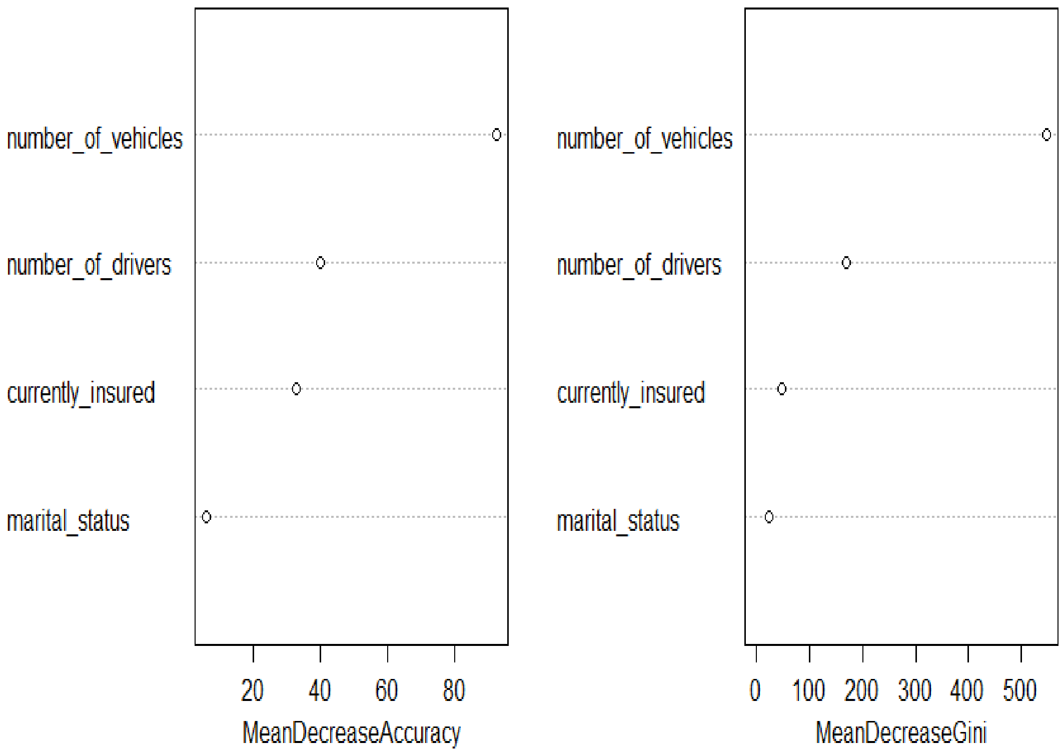



Higher Demanded

| Vehicles | Drivers | Insured | Most Likely Rank |
|---|---|---|---|
| **2** | 1 | Y/N | 5 |
| **1** | 2 | N | 1 |

# DATA ANALYSIS: RANDOM FOREST

Variable Importance Plot




Description

The variable importance plot of random forest suggests the following order of characteristics (from highest to lowest) to predict ranking.

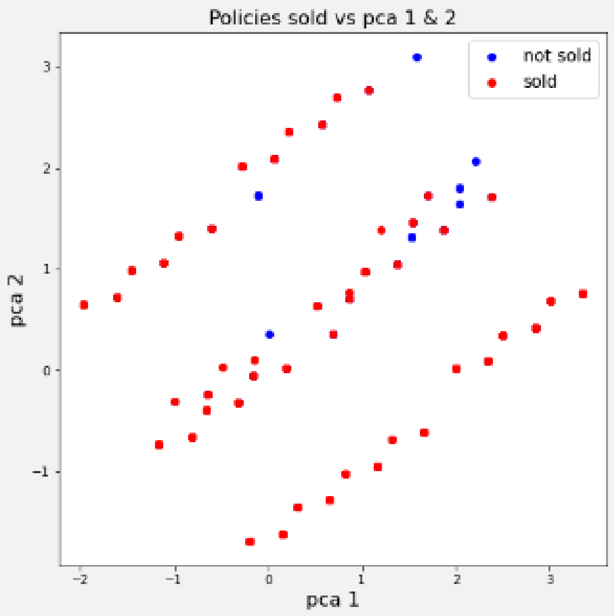
| Importance | Characteristic |
|---|---|
| 1. | Number of Vehicles |
| 2. | Number of Drivers |
| 3. | Currently Insured |
| 4. | Married |

# DATA ANALYSIS: OTHER ALGORITHMS

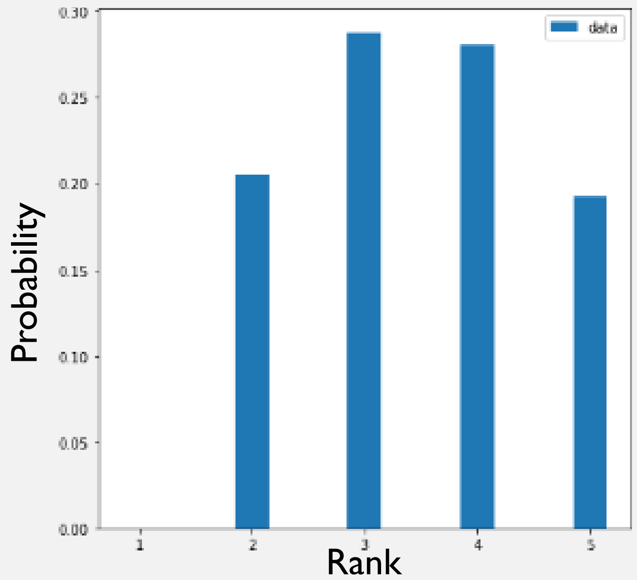
- Supervised learning:

| | Policy sold, given ad clicked and type | Ad clicked, given rank | Rank given type |
|---|---|---|---|
| **Best model** | Linear Discriminant Analysis | Random Forest Classifier | Support Vector Machine |
| **Accuracy** | 61% | 73% | 42% |
| **Precision** | 61% | 74% | 52% |
| **Recall** | 61% | 73% | 42% |

- Unsupervised learning:
  - PCA: We have few enough variables already!
  - Clustering: Also ineffective
- Low performance of SVM caused us to try to predict a *distribution* of rank given type, rather than a single rank. Population vs. rank distribution was best described by a sum of uniform distributions – about 20% for the outer ranks and about 30% for the inner ranks. Then what we predict is the center of this distribution, among the ranks.



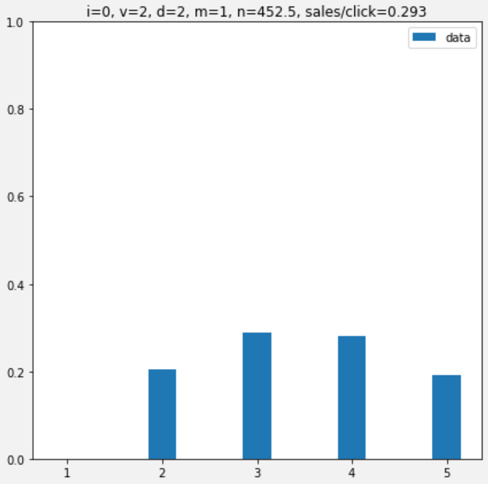

Probability Distribution of Rank Given Type



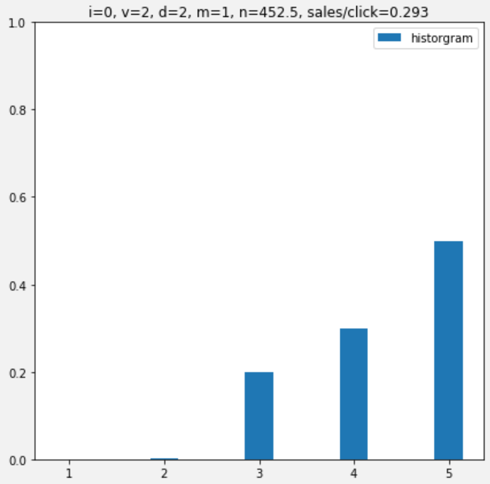
# DATA ANALYSIS: APPROXIMATE BID EFFECTS

Since our company only bids 10$ for every customer, we make 3 assumption to allow us to vary bids

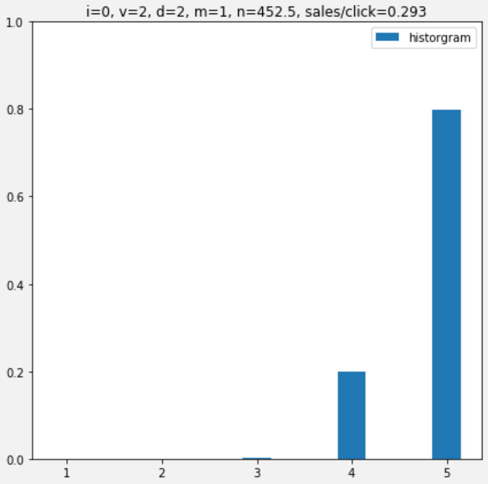
1) bid of $1 gives all rank 5
2) bids shift only the center of the distribution
3) change in bid per shift is linear



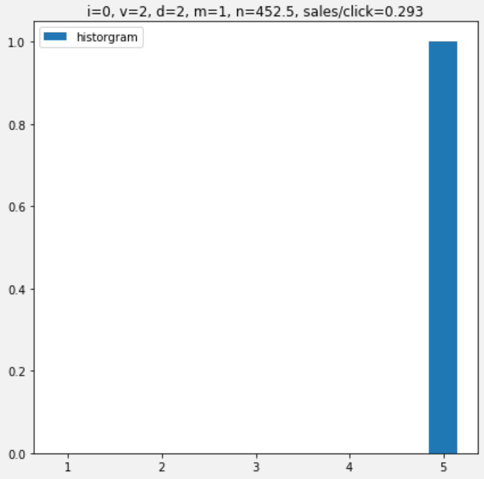

No shift
Bid: $10




1 shift down




2 shifts down




3 shifts down
Bid: $1

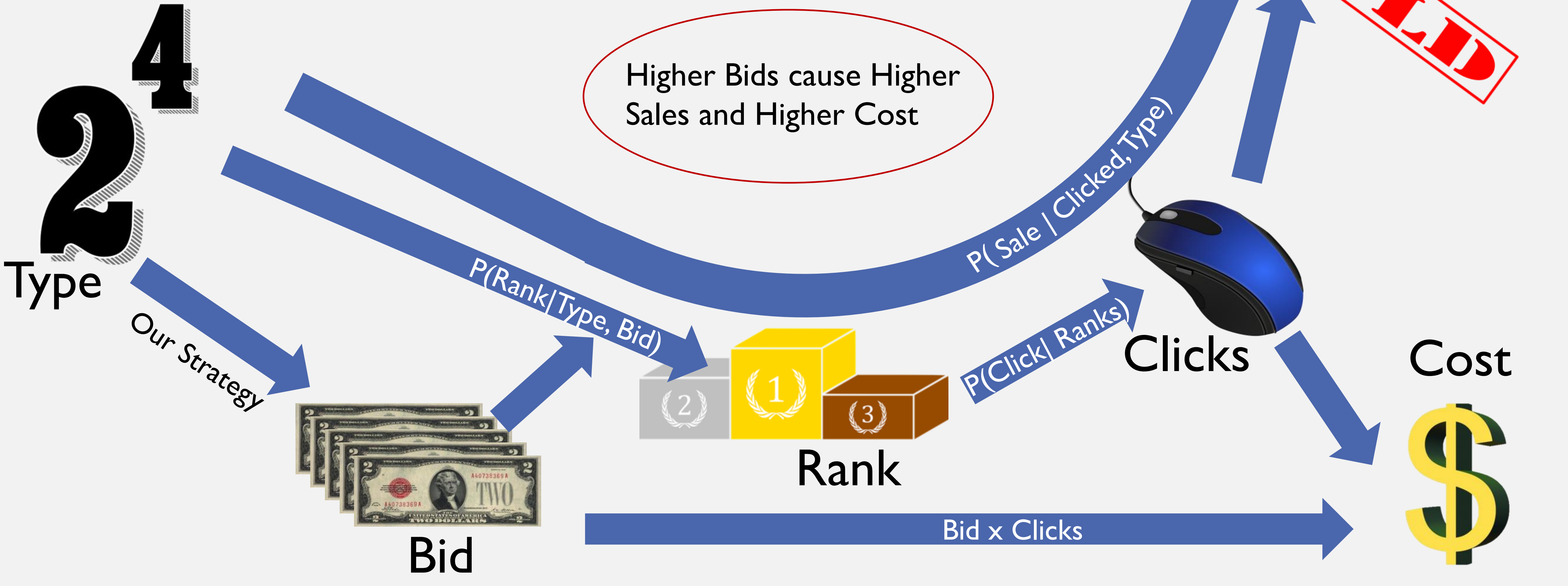
For this type: change_in_bid/shift =($10-$1)/3 shifts=$3/shift

# OUR MODEL

Sales= $\sum_{type} P(type) \sum_{rank} P(rank|type, bid) P(click|rank) P(sale|click, type)$

Cost= $\sum_{type} P(type) \sum_{rank} P(rank|type, bid) P(click|rank) bid(type)$

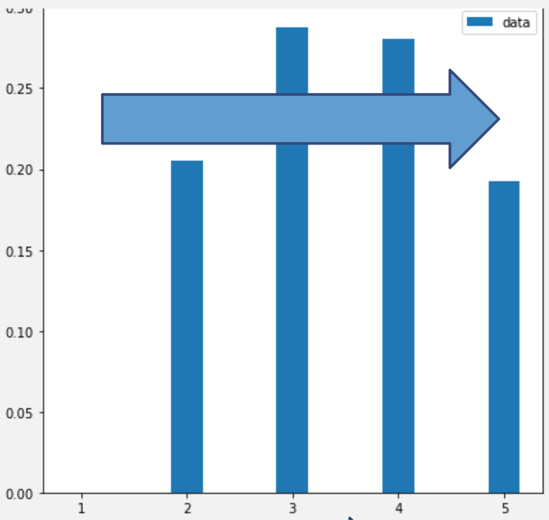
Our Strategy



# HOW WE BID

Bid decrease per shift is $3.00

Sales/Click=.26




13 sales
$460

7 sales
$171

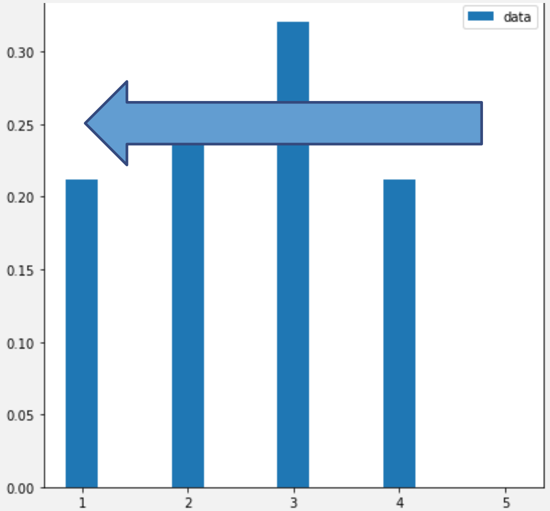
Combining both shifts gives

$137 Saved

We should bid higher for types which have high sales/click

Bid increase per shift is $2.25

Sales/Click=.59




13 sales
$308

8 sales
$156

## What if it isn’t this obvious which way we should shift?

- Wait until all the “low hanging fruit” is gone
- Test a fractional cost movement for all 16 types
- Choose the movement with the lowest cost/sale

# OPTIMIZATION & RESULTS

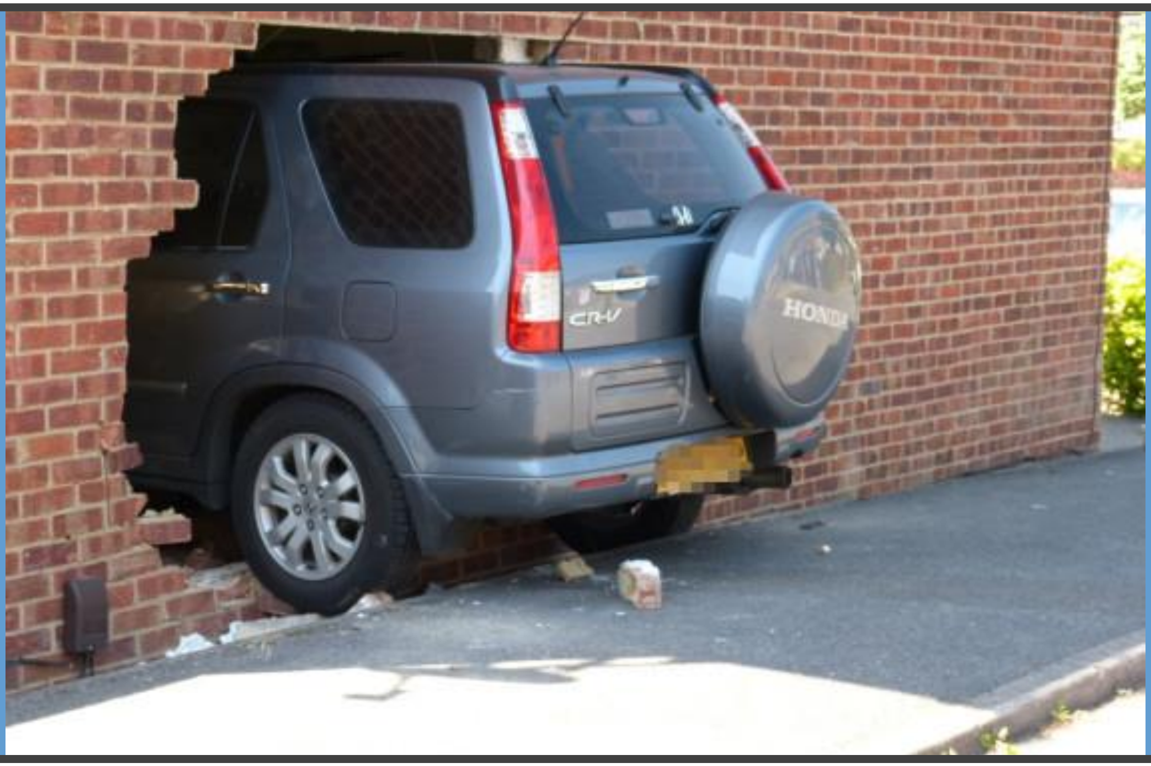
- Goal: Optimize the **cost** per **sale**, with at least 400 **sales** for every 10,000 ads
- We use a modified gradient descent algorithm with multiple initiation points to achieve an optimal bid for each type (all converged to the same optimum)
- We expect cost per sale to be consistent across types when optimized


| | Original | Big Savings | Same Sales | Big Sales |
|---|---|---|---|---|
| Sales | 545 | 400 | 545 | **590** |
| Cost | $13,740 | **$7,620 ($6,120 saved)** | $12,101 ($1,639 saved) | $13,740 |
| $/Sale | $25.21 | $19.05 | $22.20 | $23.27 |

# POTENTIAL ADDITIONS TO MODEL

- Need to know: The change in the probability of the ranking due to marginal change in bid.
- With longevity/profit information we can improve the model
- Create versions which optimize sales more so than cost effectiveness, and vice versa for bullish vs. bearish company strategy.
- Model other companies' bidding strategies to stay competitive
- One driver, non married are poorly represented in our data

# QUESTIONS?




## TEAM 12: "THE ROOTS"

MINSUB KIM, SIMON LI, THOMAS MUSSMANN, SRI RAVINDRAN,
ANGELO TARANTO, AND SHUO XU.