Customer Acquisition

Yuecheng Fang
Kevin Gardner

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

Agenda

- Executive Summary
- Data Dictionary and ERD
- Code Flowchart
- Model Performance
- Prediction

Executive Summary

Pay more attention to those companies:

- Shorter duration of registrations
- Registered more recently
- Larger number of registrations

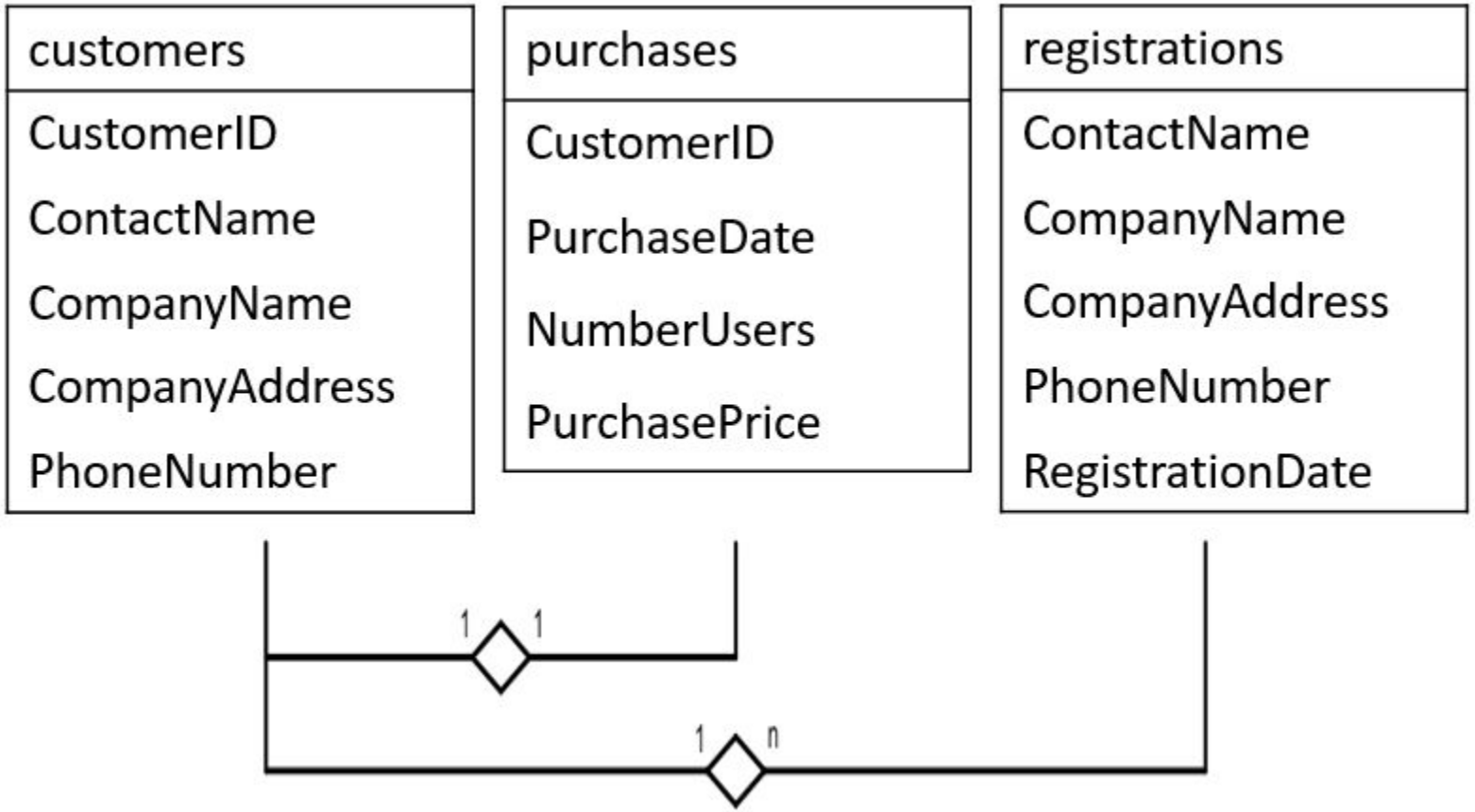
Data Dictionary

customers	
CustomerID	Customer ID
ContactName	Contact Name
CompanyName	Company Name
CompanyAddress	Where the company is located
PhoneNumber	Phone number for contact name

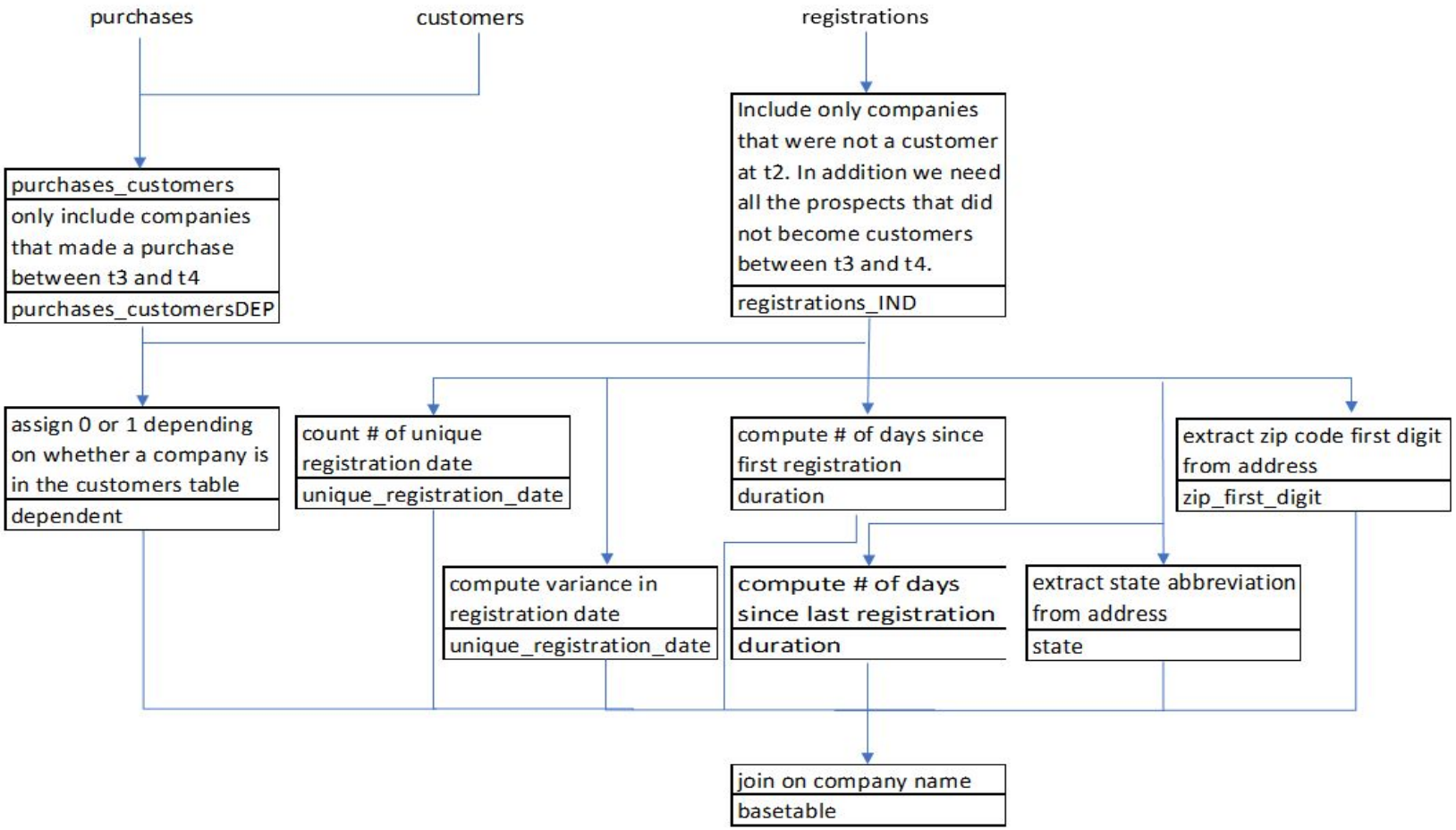
purchases	
CustomerID	Customer ID
PurchaseDate	Purchase date
NumberUsers	Number of users given access
PurchasePrice	Total price that was paid

registrations	
ContactName	Contact name
CompanyName	Company name
CompanyAddress	Where the company is located
PhoneNumber	Phone number for contact name
RegistrationDate	When the user registered

ERD


customers
CustomerID
ContactName
CompanyName
CompanyAddress
PhoneNumber
purchases
CustomerID
PurchaseDate
NumberUsers
PurchasePrice
registrations
ContactName
CompanyName
CompanyAddress
PhoneNumber
RegistrationDate
1
1
1
n

Flowchart


purchases
customers
registrations
purchases_customers
only include companies that made a purchase between t3 and t4
purchases_customersDEP
Include only companies that were not a customer at t2. In addition we need all the prospects that did not become customers between t3 and t4.
registrations_IND
assign 0 or 1 depending on whether a company is in the customers table
dependent
count # of unique registration date
unique_registration_date
compute # of days since first registration
duration
extract zip code first digit from address
zip_first_digit
compute variance in registration date
unique_registration_date
compute # of days since last registration
duration
extract state abbreviation from address
state
join on company name
basetable

Create Model

AUC	= 0.85
Top Decile Lift	= 2.18
Function Length	= 112 lines
Running Time	= 10.5 seconds

Positively Related Variables	**Negatively Related Variables**
Number of Registration Dates	Recency of Registration
Number of Registrations	Duration of Registrations
Variance of Registration Date	

Deploy Model

Function Length = 80 lines

Running Time = 3.7 seconds

Highest propensity leads:

1. Rainbow Polar
2. Seashell Consignment
3. Raisin Fire
4. Warrior Mania
5. Mango Videos
6. Mango Dolphin
7. Mushroom Leopard
8. Sunny Room
9. Myth Days
10. Mango Casting Call

Questions?

Appendix

Function calls

AUROC

Variable importance plot

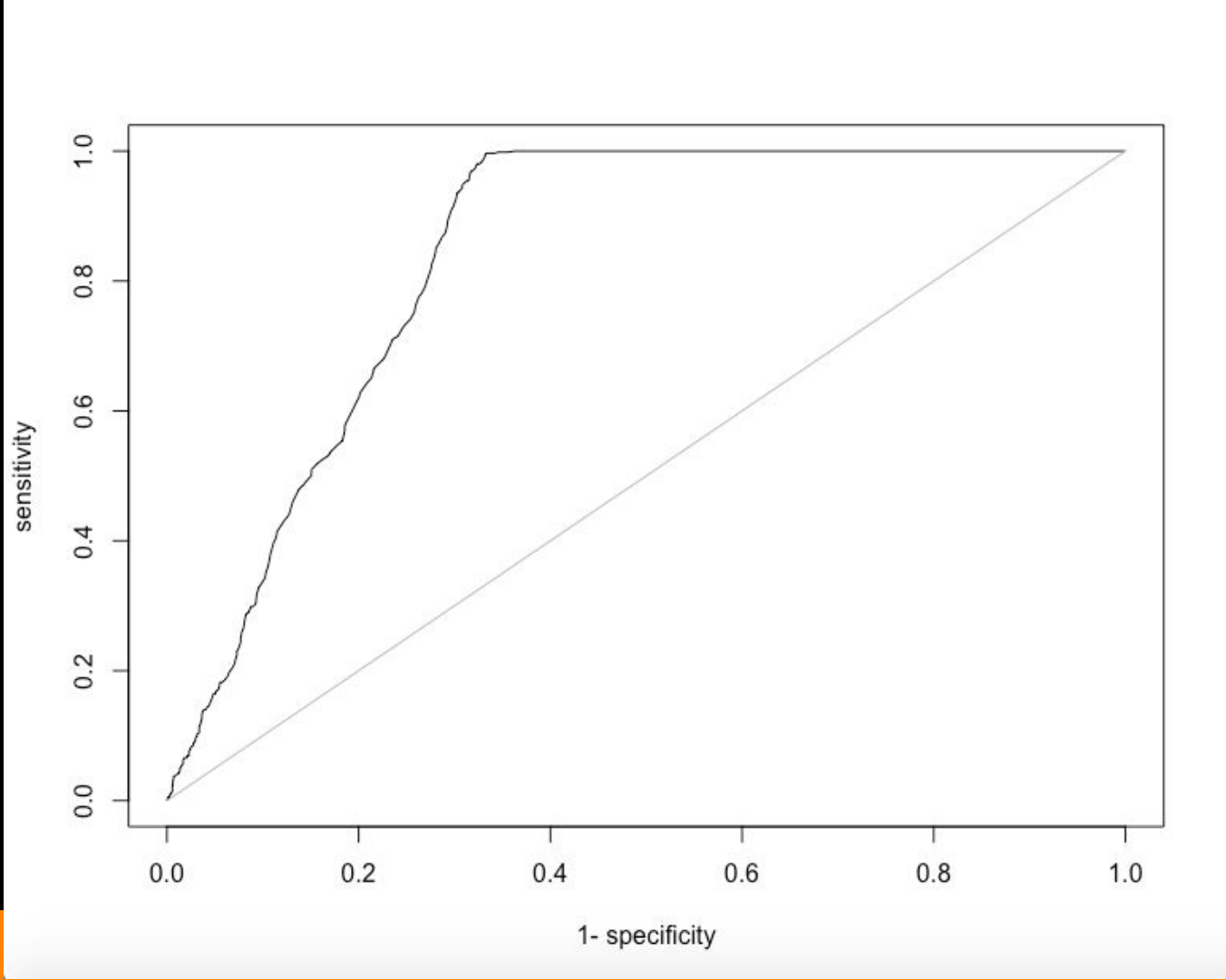
Partial dependence plots

Function Calls

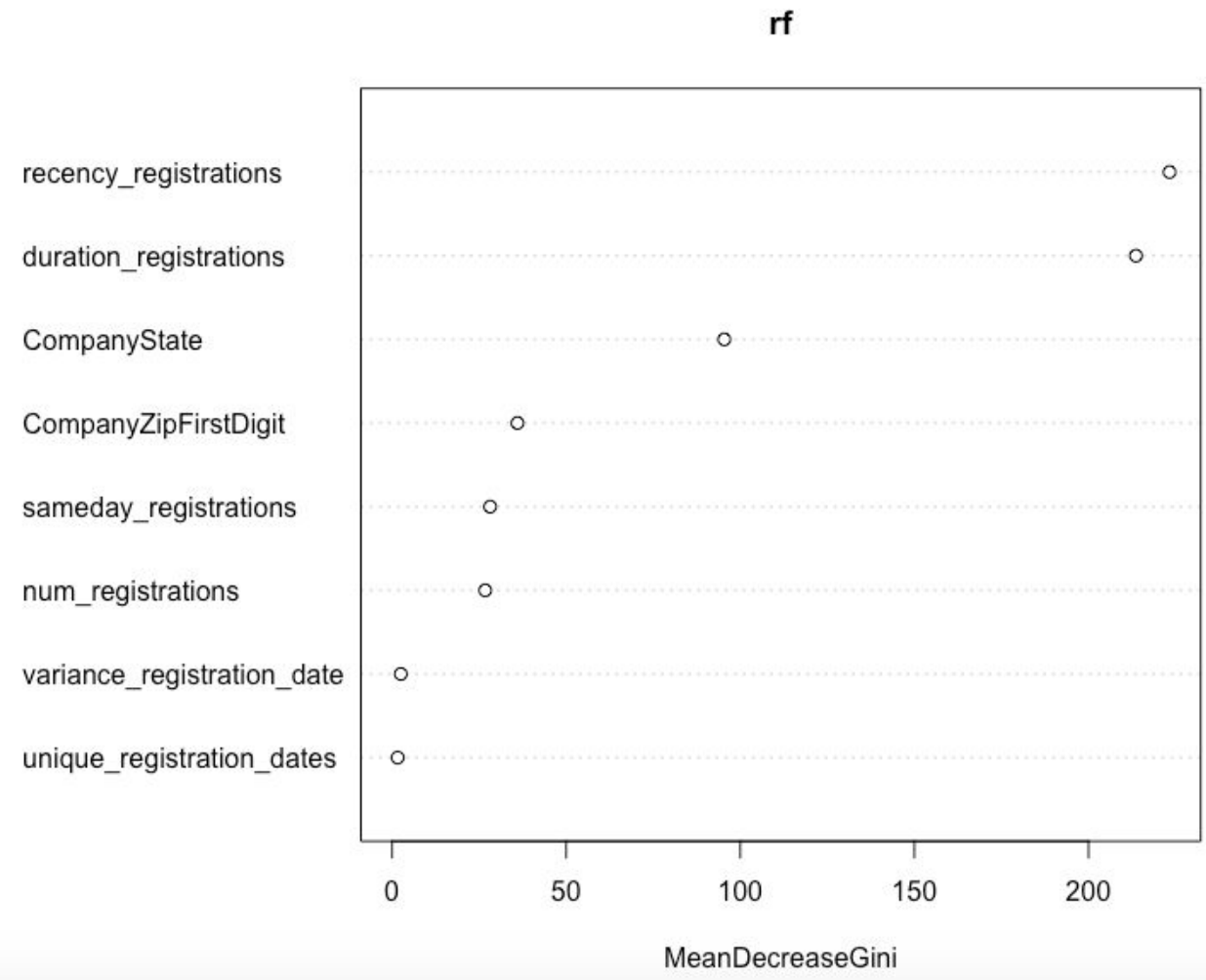
```
system.time(result <- create_model(start_ind="2014-05-31", # Earliest registration date
                                   end_ind="2015-07-30", # Operational period be 1 day
                                   start_dep="2015-07-31", # Dependent period be 32 days
                                   end_dep="2015-08-30" # The maximum purchase date
))
# Loading data..
# Creating basetable..
# Creating train and test set..
# Fitting random forest..
# user  system elapsed
# 5.863   0.142  10.485

system.time(result2 <- deploy_model(start_ind="2014-07-03",
                                    end_ind="2015-08-30",
                                    model=result$model))
# Loading data..
# Creating basetable..
# Scoring prospects..
# user  system elapsed
# 3.862   0.032   3.738
```

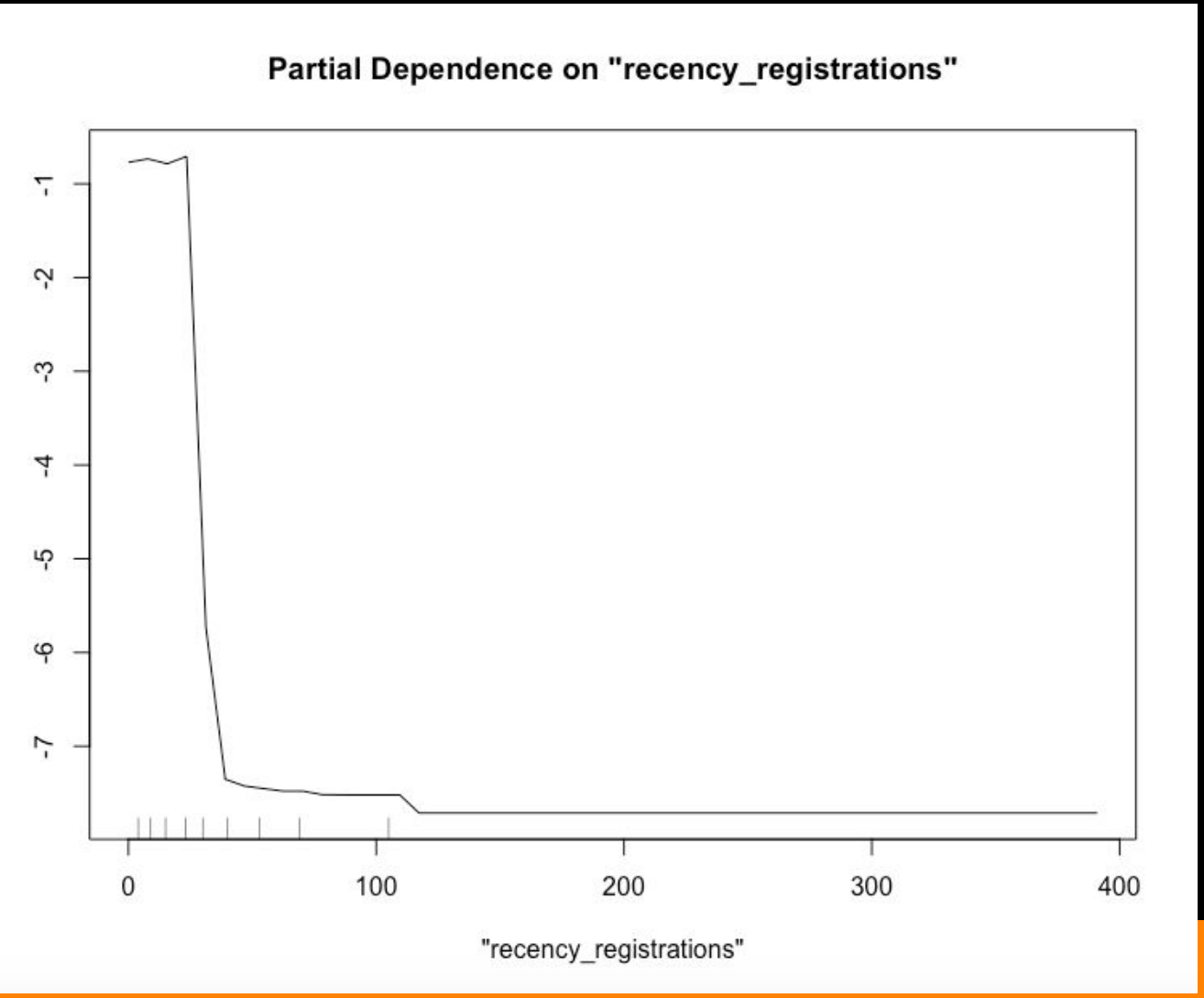
AUROC



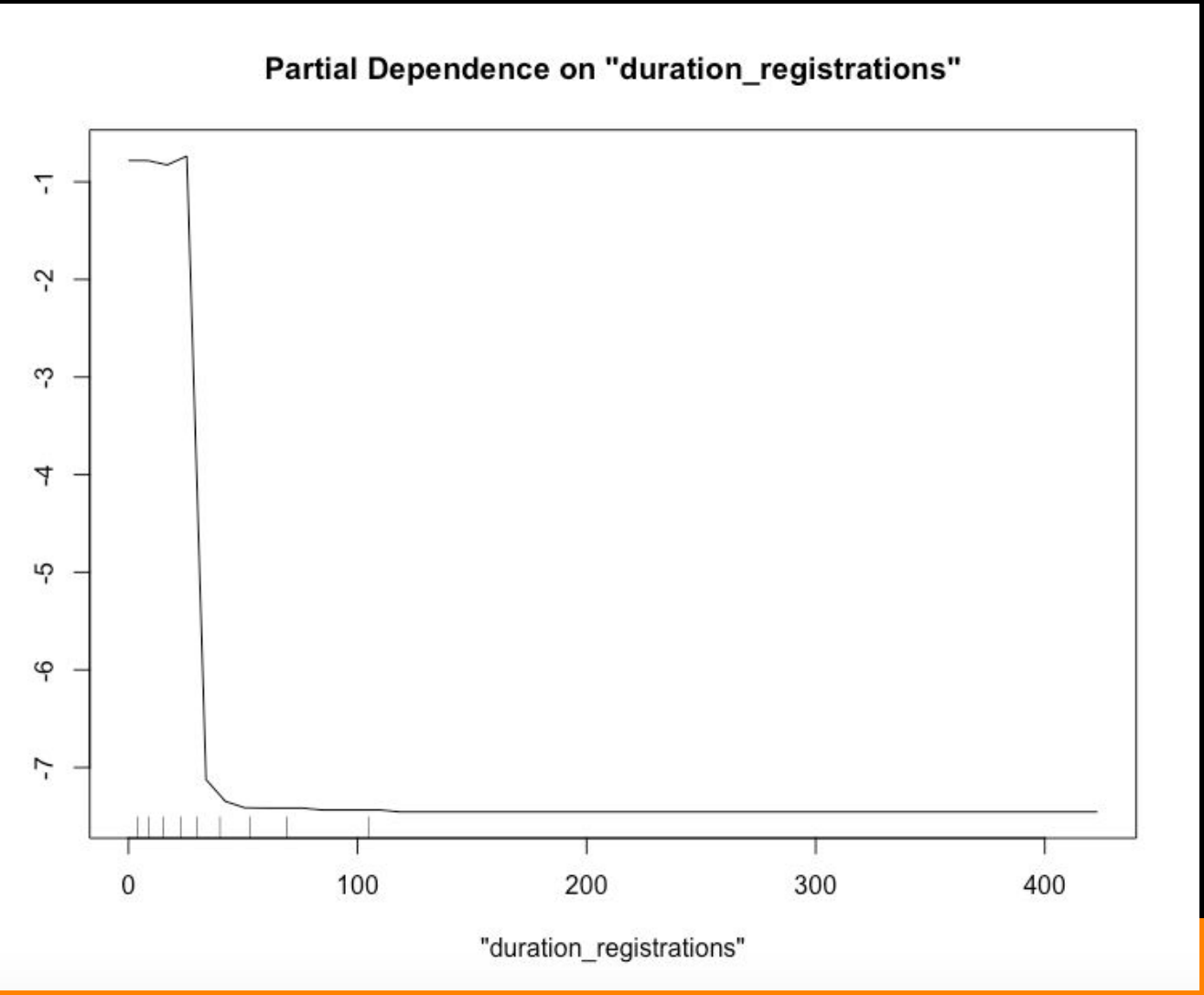
Variable Importance



Recency of registrations



Duration of registrations



Number of registrations