National City Bank Case

Isaiah New isaiahnew@icloud.com

Group Project

Introduction

- Data Set: The training data contain 4,000 customers of National City Bank whose marketing department had previously contacted to sell a credit product (used vehicle line of credit).
- Supplemental third-party data on current and prospective customers were also available.

Goal

- Identify the 100 most propense users to accept a used car credit line
- Share some insights about our current customer base

We ran multiple models over the data to come up with the most accurate outcome.

Cleaning & Analysis

Obtain the dataset ...

```
> read.csv('CurrentCustomerMktg
Results.csv', na.strings=c(""))

> joinData <- left_join (currentData, … )
```

- Data obtained from 4,000 National City bank customers included 30 variables

```
joinData    4000 obs. of 30 variables
  dataID : int 1 2 3 4 5 6 7 8 9 10 ...
  HHuniqueID : chr "HHd4d0af8c72" "HH8d3e87c164" "HHdd53ef1db6" "HH6fa0de6516" ...
  Communication : chr "telephone" NA "cellular" "cellular" ...
  LastContactDay : int 28 26 3 11 3 22 17 12 18 12 ...
  LastContactMonth : chr "jan" "may" "jun" "may" ...
  NoOfContacts : int 2 5 1 2 1 1 1 4 1 2 ...
  DaysPassed : int -1 -1 119 -1 -1 109 -1 -1 -1 -1 ...
  PrevAttempts : int 0 0 1 0 0 1 0 0 0 0 ...
  past_Outcome : chr NA NA "failure" NA ...
  CallStart : chr "13:45:20" "14:49:03" "16:30:24" "12:06:43" ...
  CallEnd : chr "13:46:30" "14:52:08" "16:36:04" "12:20:22" ...
  Y_AcceptedOffer : chr "DidNotAccept" "DidNotAccept" "Accepted" "Accepted" ...
  X : chr "Accepted" "Accepted" "Accepted" "Accepted" ...
  carMake : chr "Mitsubishi" "Hyundai" "Ford" "Buick" ...
  carModel : chr "Eclipse" "Elantra" "Explorer Sport" "Regal" ...
  carYr : int 2014 2009 2011 2008 2008 2013 2019 2017 1985 2018 ...
  DefaultOnRecord : int 0 0 0 0 0 0 0 0 0 0 ...
  RecentBalance : int 1218 1156 637 373 2694 1625 1000 538 187 3 ...
  HHInsurance : int 1 1 1 1 0 0 1 1 1 1 ...
  CarLoan : int 0 0 0 0 0 0 0 0 0 1 ...
  headOfhouseholdGender : chr "F" "M" "F" "F" ...
  annualDonations : chr NA NA NA NA ...
  EstRace : chr "Fijian" NA NA NA ...
  PetsPurchases : logi TRUE TRUE FALSE TRUE TRUE TRUE ...
```

… and capture missing values

What about missing values?

```
> joinData[joinData=="NA"]=NA

> joinData[joinData==""]=NA

> joinData[joinData=="<NA>"]=NA
```

Remove X attribute, adds no quantitative or qualitative value to our data

Deal with NA, "" and <NA>

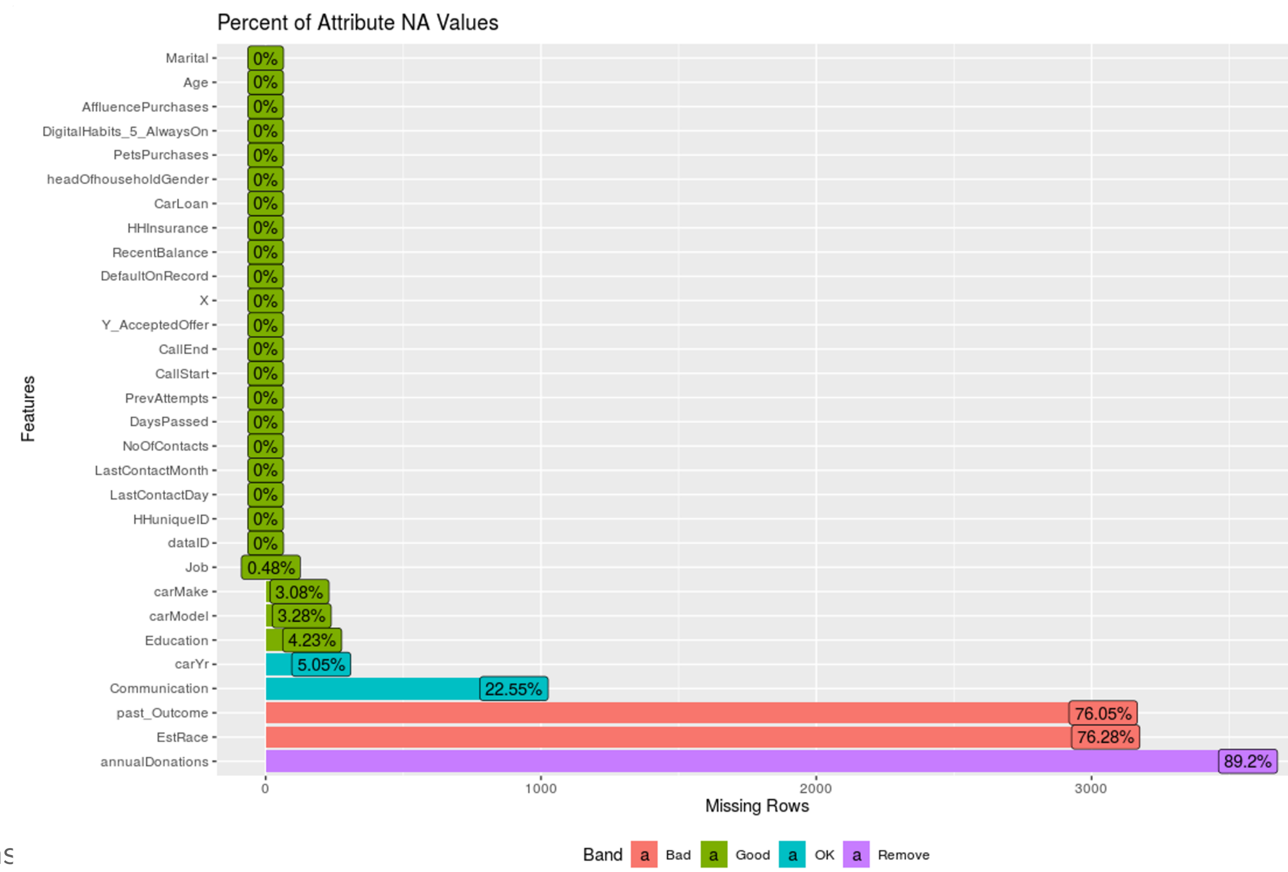
```
joinData                          4000 obs. of 30 variables
  dataID : int 1 2 3 4 5 6 7 8 9 10 ...
  HHuniqueID : chr "HHd4d0af8c72" "HH8d3e87c164" "HHdd53ef1db6" "HH6fa0de6516" ...
  Communication : chr "telephone" NA "cellular" "cellular" ...
  LastContactDay : int 28 26 3 11 3 22 17 12 18 12 ...
  LastContactMonth : chr "jan" "may" "jun" "may" ...
  NoOfContacts : int 2 5 1 2 1 1 1 4 1 2 ...
  DaysPassed : int -1 -1 119 -1 -1 109 -1 -1 -1 -1 ...
  PrevAttempts : int 0 0 1 0 0 1 0 0 0 0 ...
  past_Outcome : chr NA NA "failure" NA ...
  CallStart : chr "13:45:20" "14:49:03" "16:30:24" "12:06:43" ...
  CallEnd : chr "13:46:30" "14:52:08" "16:36:04" "12:20:22" ...
  Y_AcceptedOffer : chr "DidNotAccept" "DidNotAccept" "Accepted" "Accepted" ...
  X : chr "Accepted" "Accepted" "Accepted" "Accepted" ...
  carMake : chr "Mitsubishi" "Hyundai" "Ford" "Buick" ...
  carModel : chr "Eclipse" "Elantra" "Explorer Sport" "Regal" ...
  carYr : int 2014 2009 2011 2008 2008 2013 2019 2017 1985 2018 ...
  DefaultOnRecord : int 0 0 0 0 0 0 0 0 0 0 ...
  RecentBalance : int 1218 1156 637 373 2694 1625 1000 538 187 3 ...
  HHInsurance : int 1 1 1 1 0 0 1 1 1 1 ...
  CarLoan : int 0 0 0 0 0 0 0 0 0 1 ...
  headOfhouseholdGender : chr "F" "M" "F" "F" ...
  annualDonations : chr NA NA NA NA ...
  EstRace : chr "Fijian" NA NA NA ...
  PetsPurchases : logi TRUE TRUE FALSE TRUE TRUE TRUE ...
```

Missing Data

A number of attributes from the amalgamated data were missing substantial values or altogether irrelevant. To improve the predictive ability of our model, we decided to remove those attributes from the data set. These include:

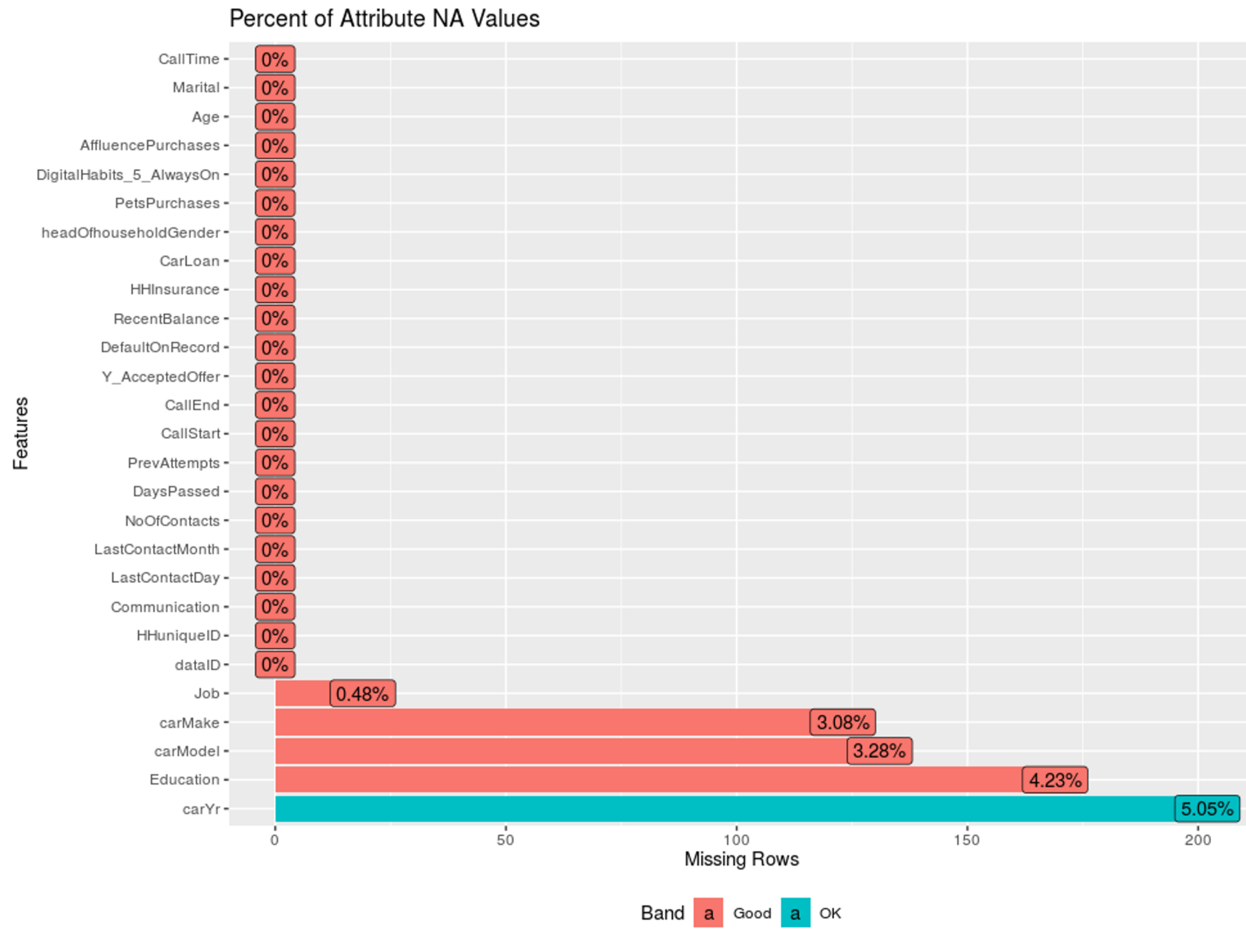
- X
- past_Outcome
- EstRace
- annualDonations

For Communication, we imputed "cellular" for missing values, which was more than 99% of existing responses.



Missing Data

After cleaning data, the percent of attributes NA values dropped dramatically.



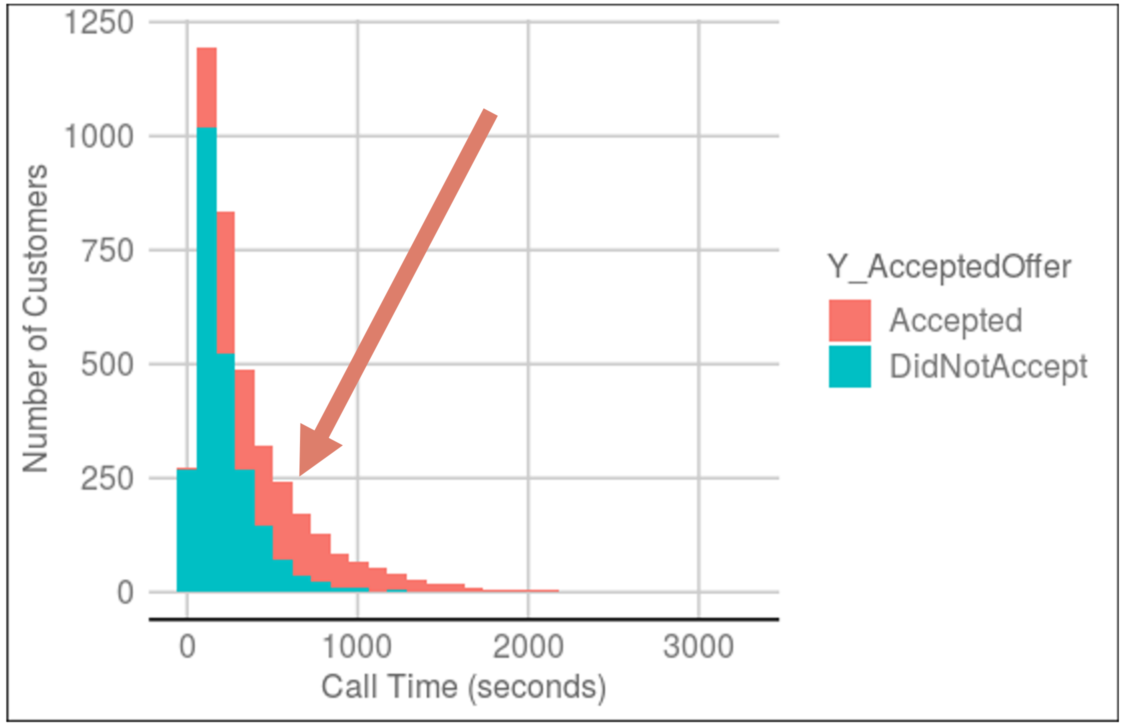
New Data

What value-added insights can we glean from our existing data?

CallTime <- callend - callstart

joinData$**CallTime** <- CallTime

Longer calls are correlated with a greater likelihood of accepting offer



Customer Marketing Insights

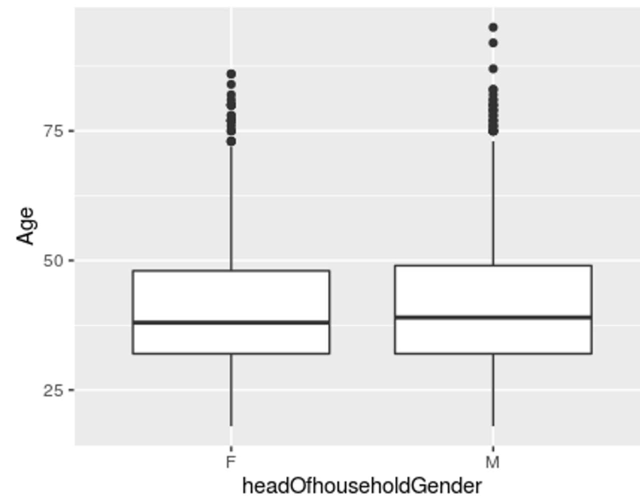
Females v Males

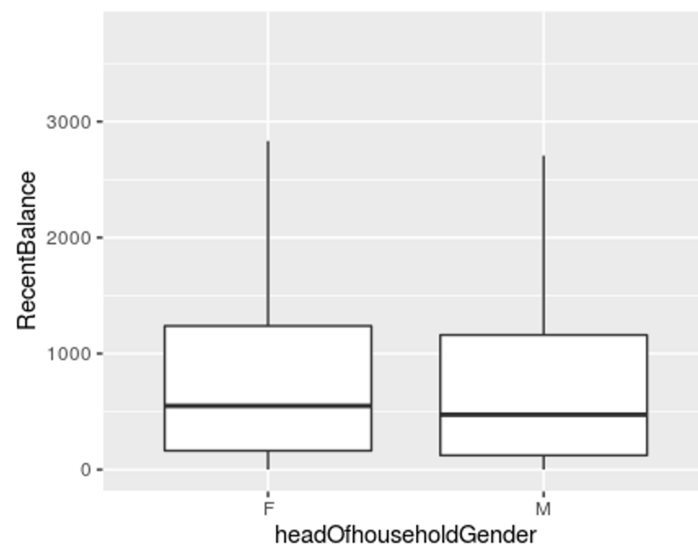
Median age for both males and females are about the same, with a few older outliers for men.

Female heads of household have slightly higher bank balances than males.

The median female has $58 more dollars in the bank than the median male.

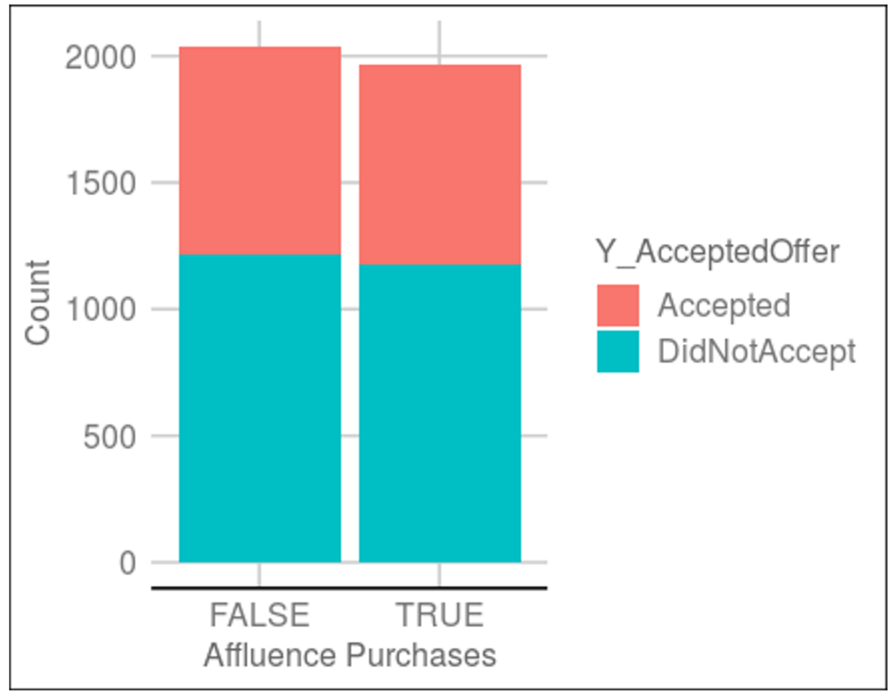
```
 headOfhouseholdGender medianbalance
  <chr>                         <int>
1 F                               581
2 M                               523
```





Females v Males

Affluence purchases are no more likely than other customers to accept the offer.



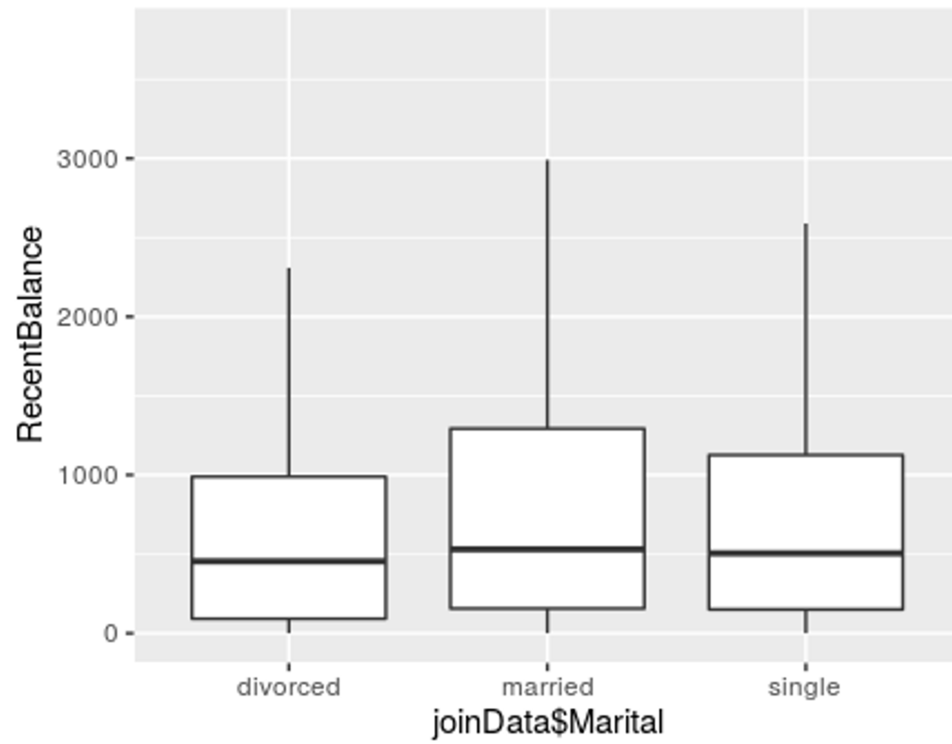
Divorcees

Divorcees have lower bank balances

Divorcees have *$138 less in the bank* than Married individuals.

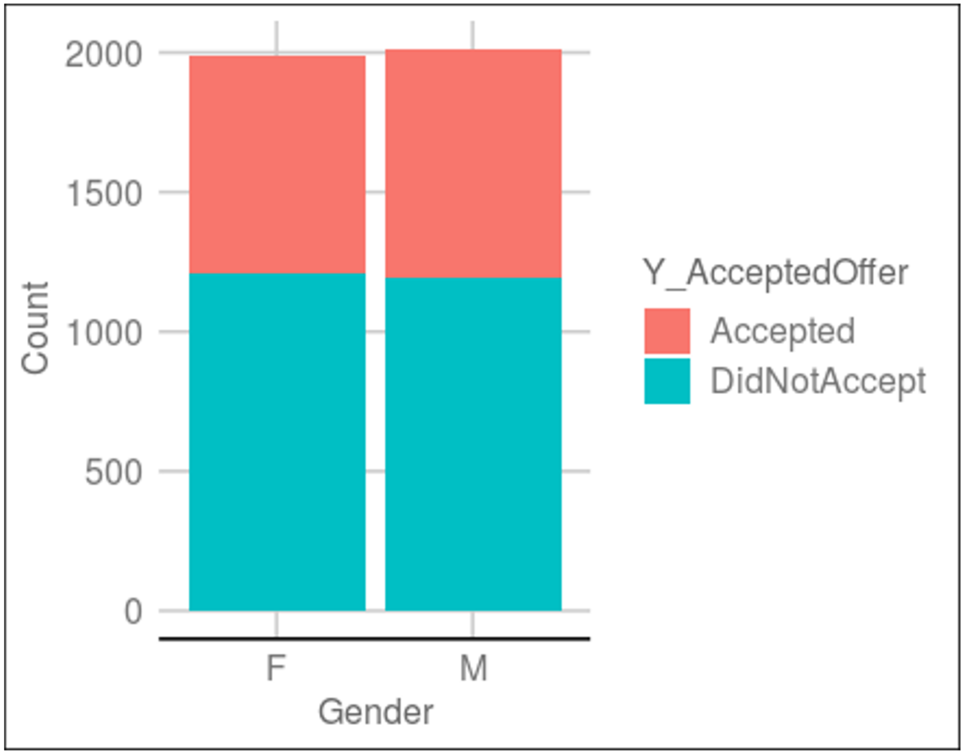
That's a 23% lower bank balance.

```
Marital   medianbalance
  <chr>            <dbl>
1 divorced          458
2 married           596.
3 single            533
```



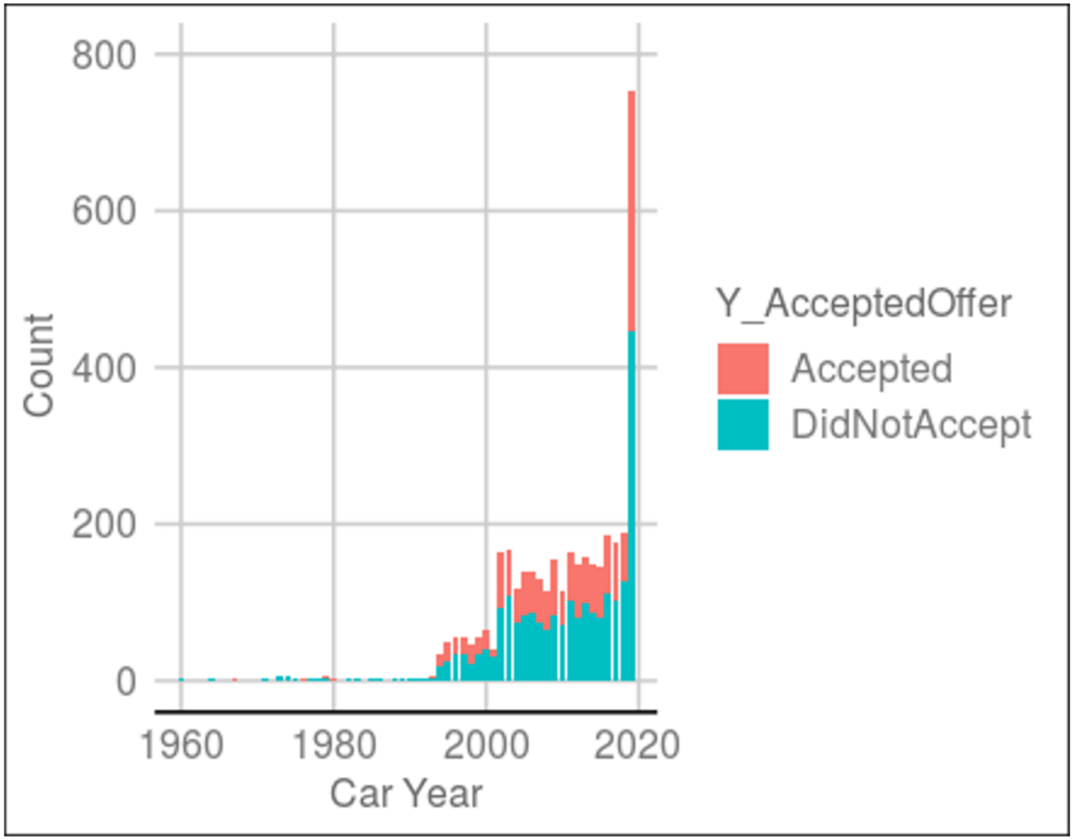
Accepting the Offer

Gender plays no role in acceptance of credit offer: Men and Women accepted the offer in equal proportion



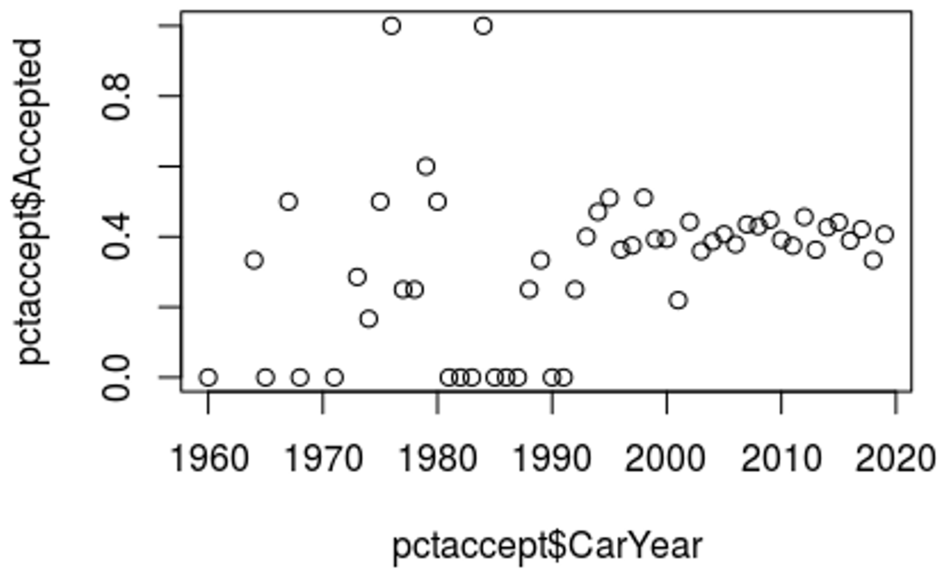
Accepting the Offer

2019 models more likely to accept offer



Accepting the Offer

Data show percent acceptances level off at about 40% of offers tendered for vehicle models made between 1995-2019



Targeting our users

Length of call by gender is equal at 232 seconds

I.e., **Median call length was 3 minutes 52 seconds**

joinData %>% group_by(headOfhouseholdGender) %>% summarise(mediancalltime = median(CallTime))

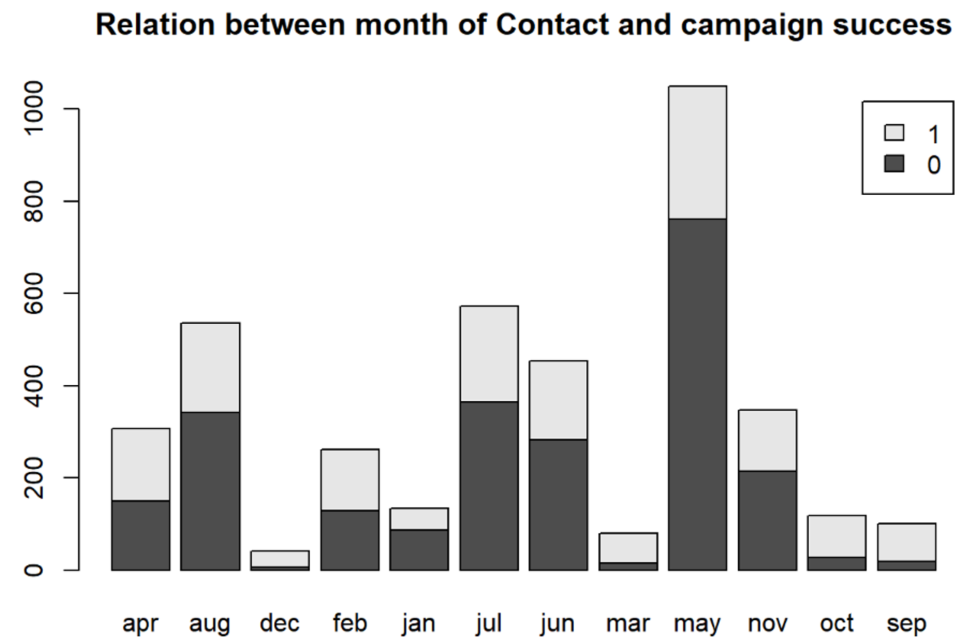
```
headOfhouseholdGender mediancalltime
  <chr>                 <drtn>
1 F                     232 secs
2 M                     232 secs
```

Targeting our users

Rate conversion by month provides insights on when to contact users and for how long to target the campaign

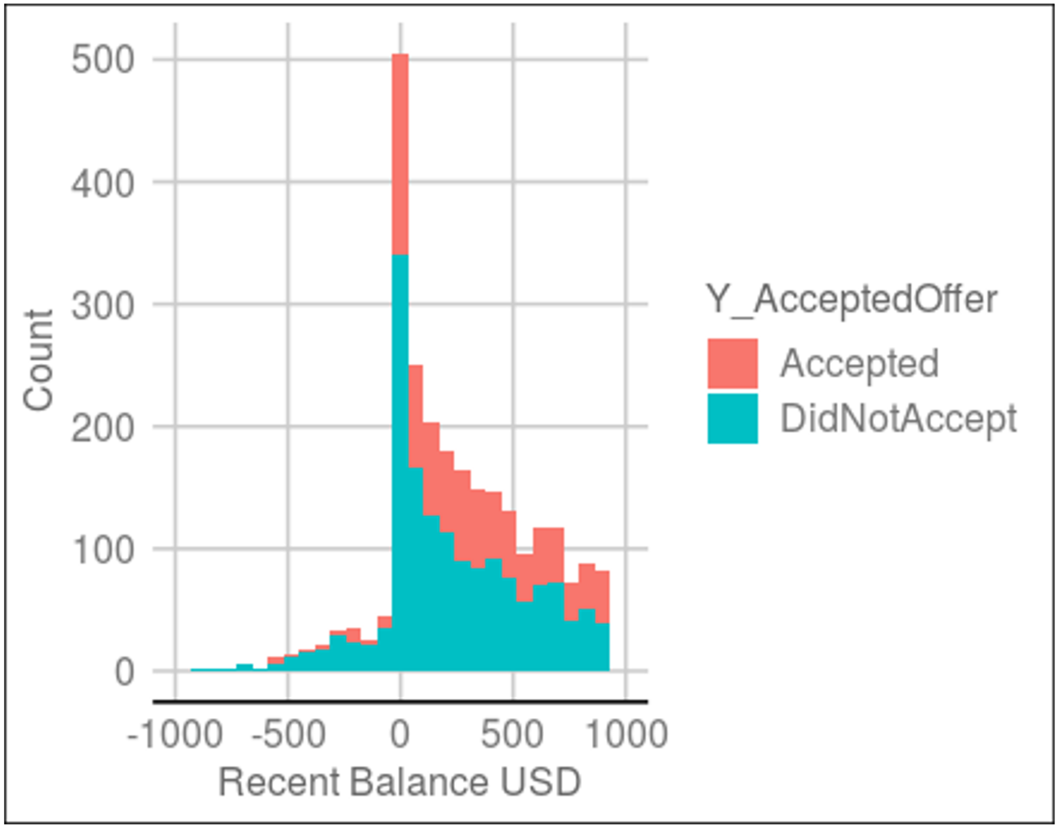
Length of call by gender is equal at 232 seconds

I.e., **Median call length was 3 minutes 52 seconds**



Accepting the Offer

Customers with negative bank balances are less likely to accept offer



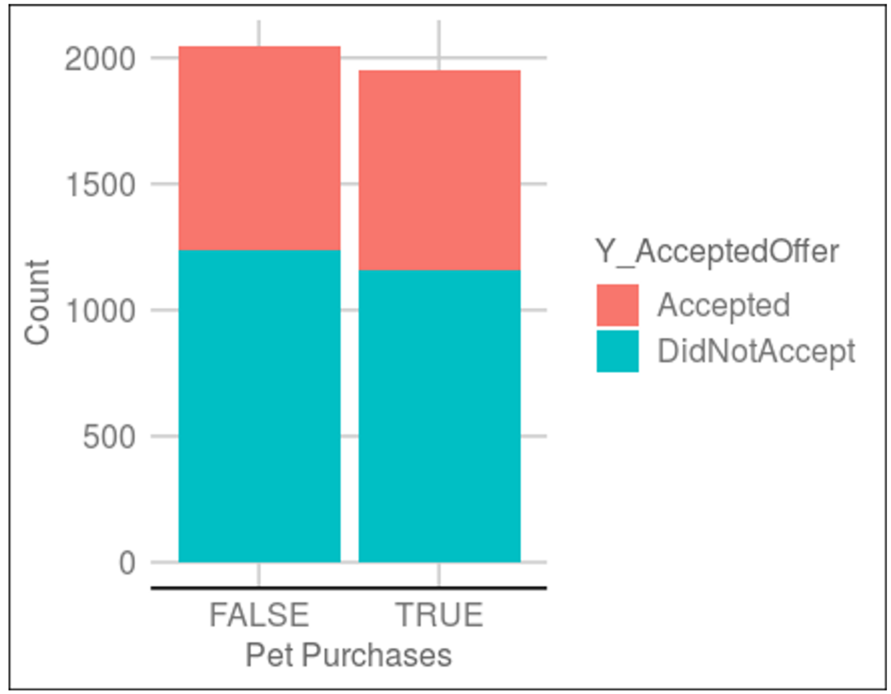
Pet Purchasers

Pet purchasers have more purchasing power.

They hold **$79 more dollars in the bank** than non-pet purchasers

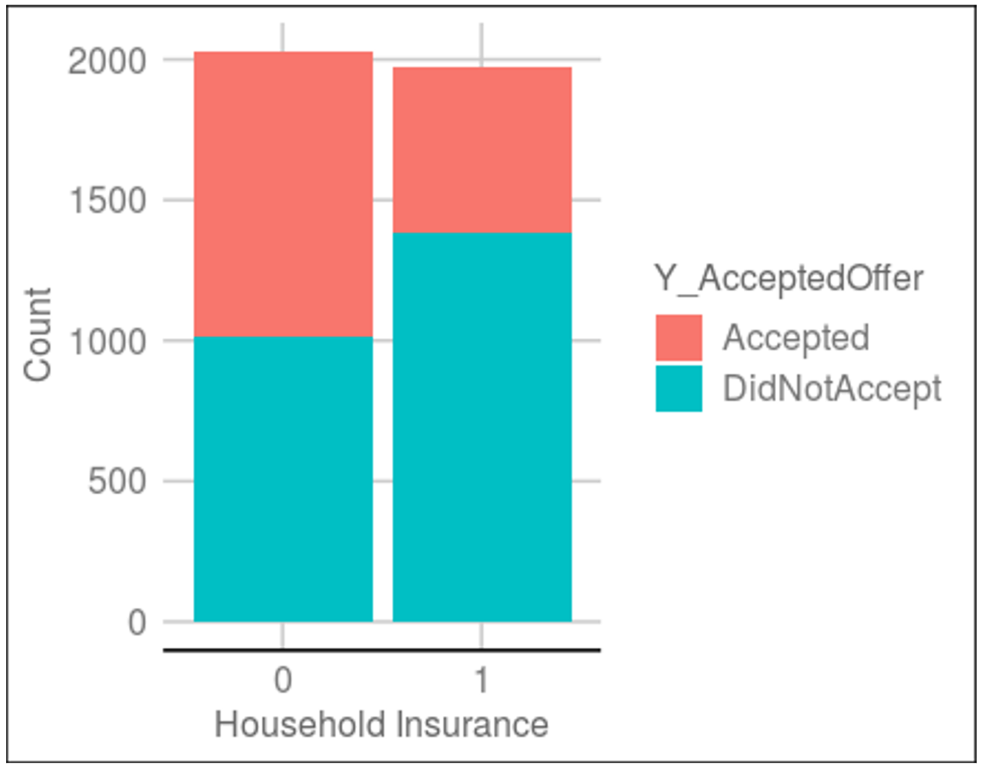
```
PetsPurchases   medianrecentbalance
  <lgl>                       <int>
1 FALSE                         514
2 TRUE                          593
```

But NOT more likely to accept the credit product.



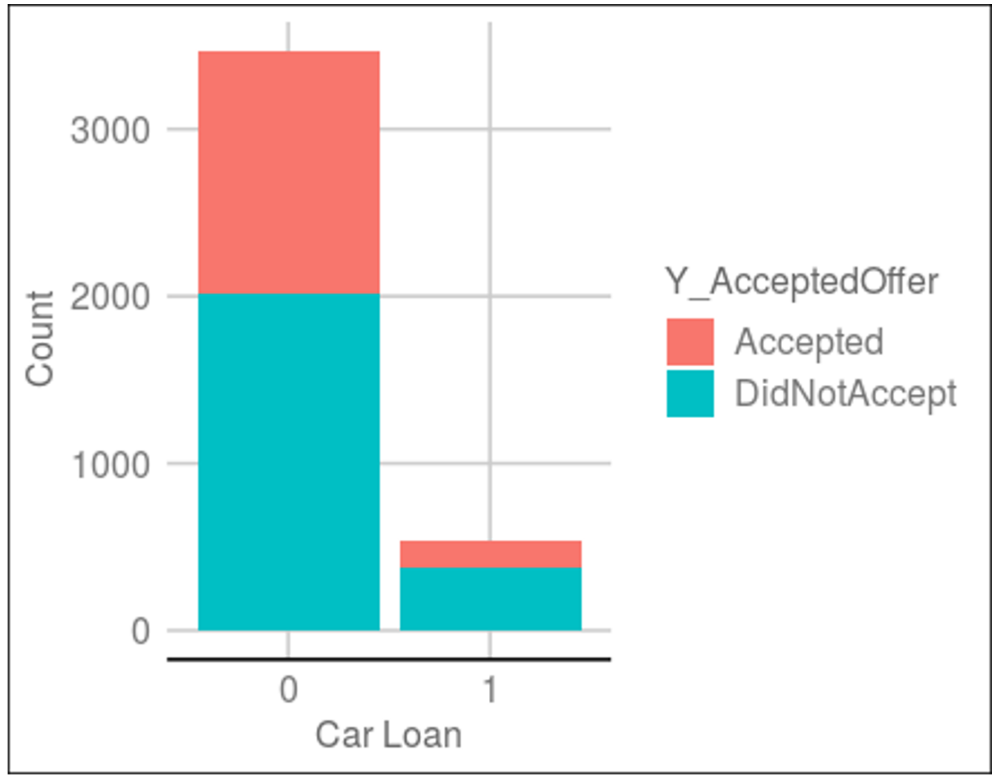
Household Insurance

Customers **WITHOUT household insurance** with the bank were more likely to accept offer.



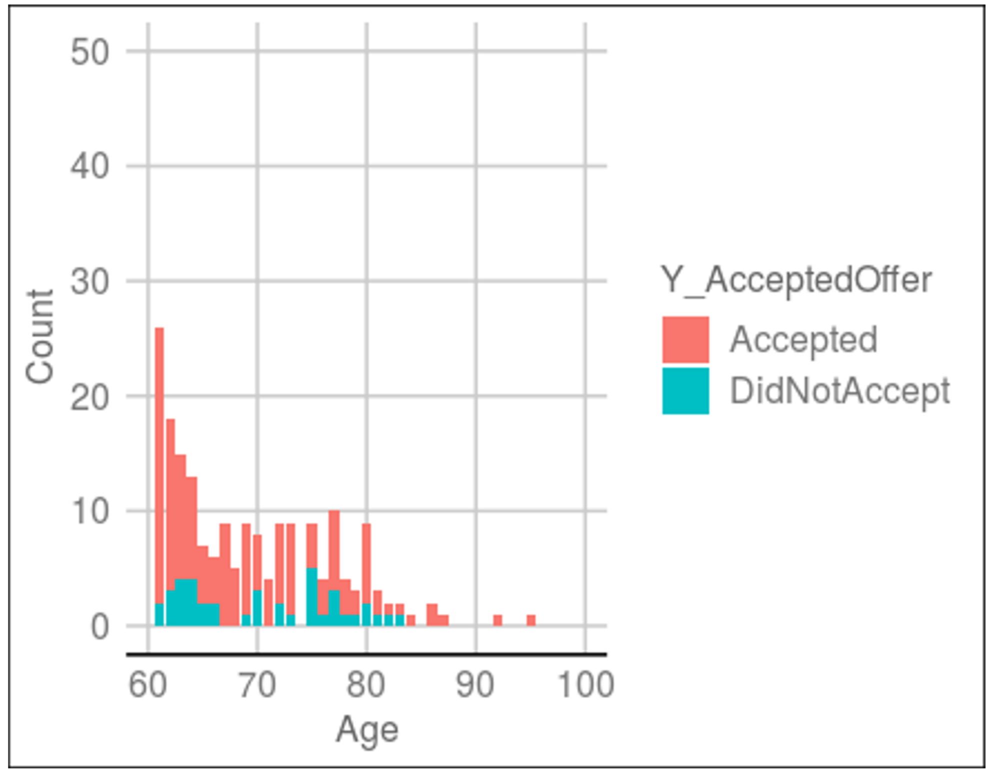
Car Loan

Customers WITHOUT a car loan were **13% more likely** to accept offer.



Old age

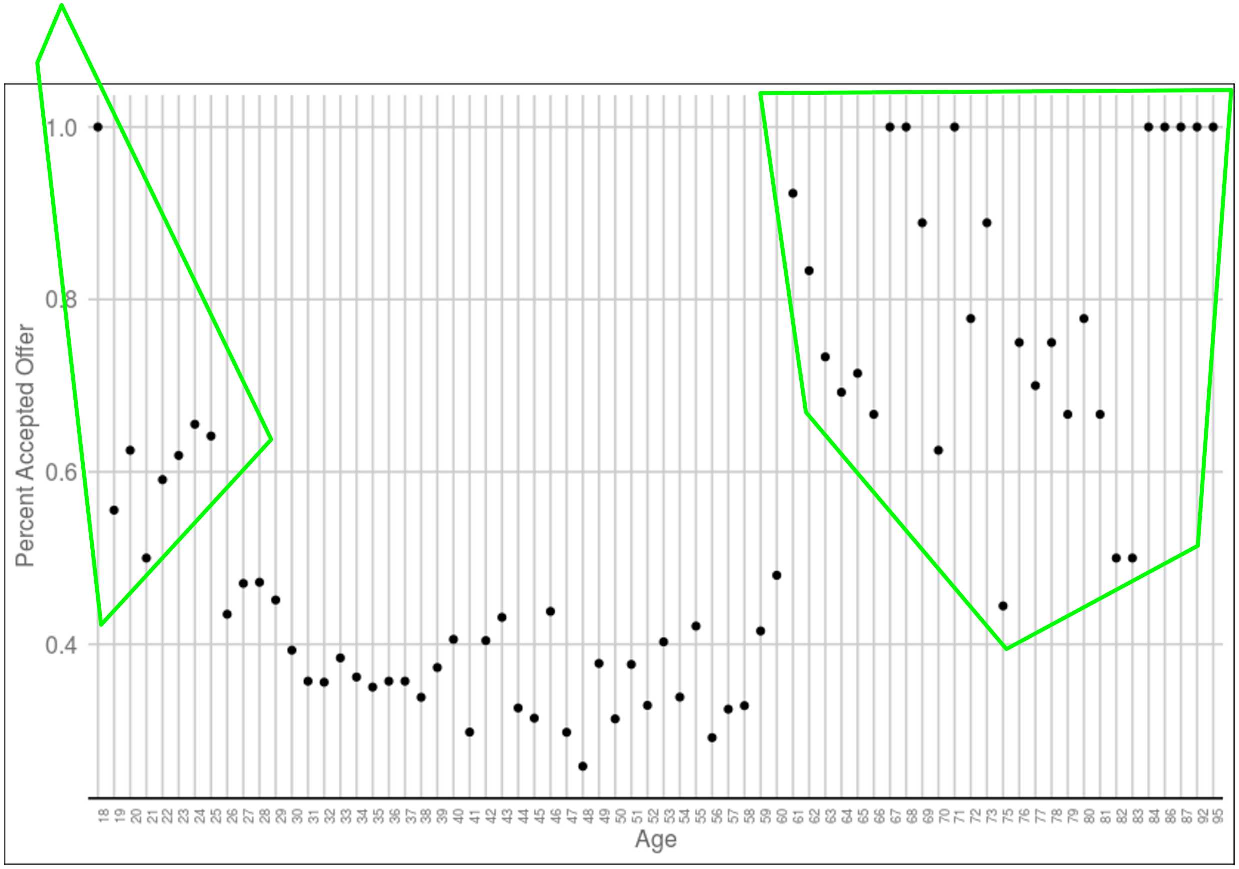
Customers aged 61 and older were more likely than all other age groups to accept offer...



Old age

But don’t count out the youngin’s

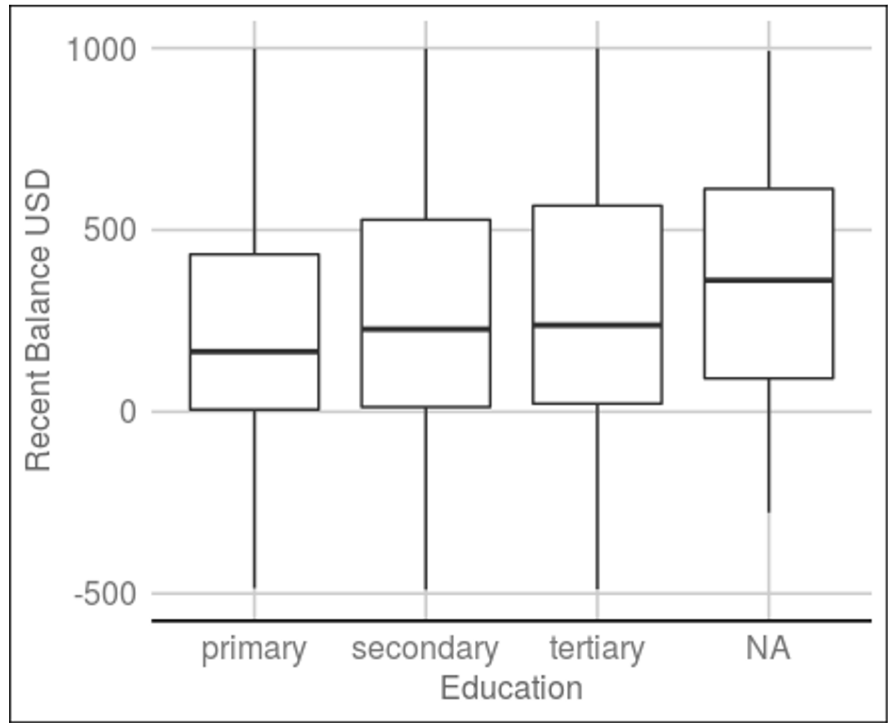
Customers ages 18-26 are also statistically more likely than not, to accept the product offer.



Education

Customers with higher Educations are more likely to have higher recent balances

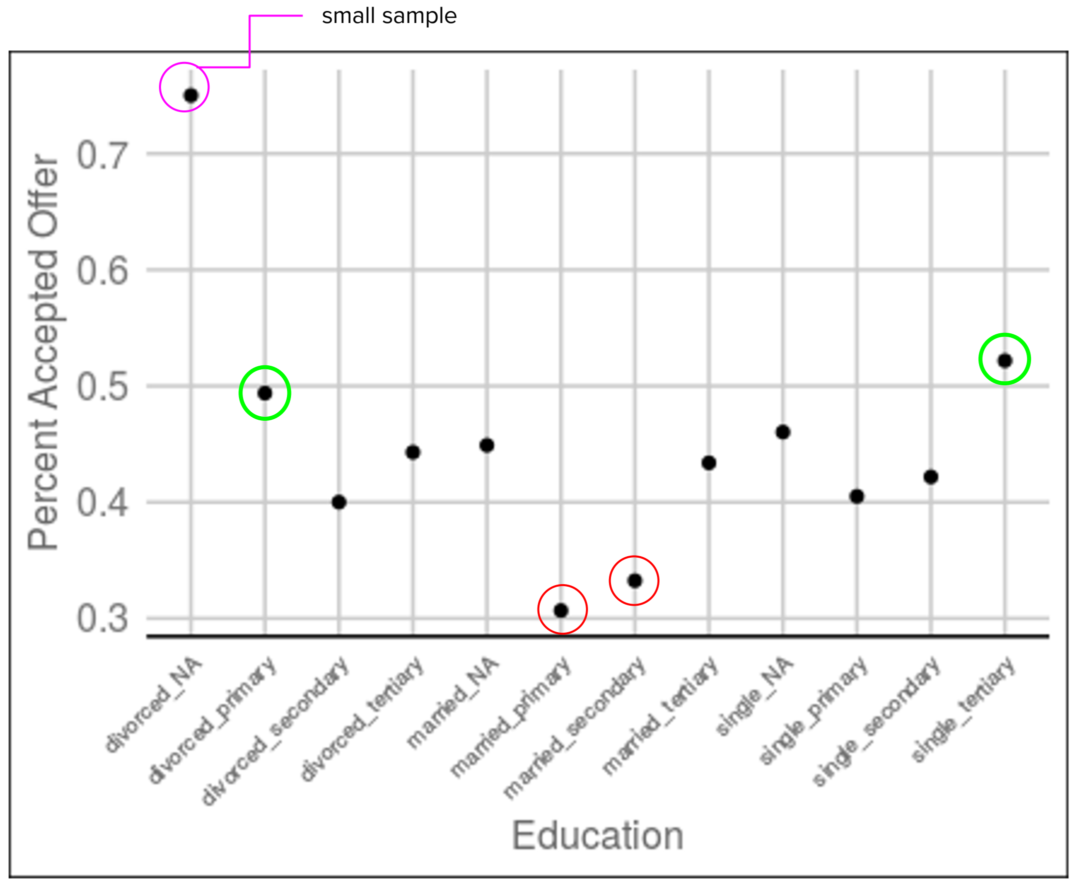
Customers with tertiary educations have median recent balances **$211 higher** than customers with only a primary education.



Marital + Education

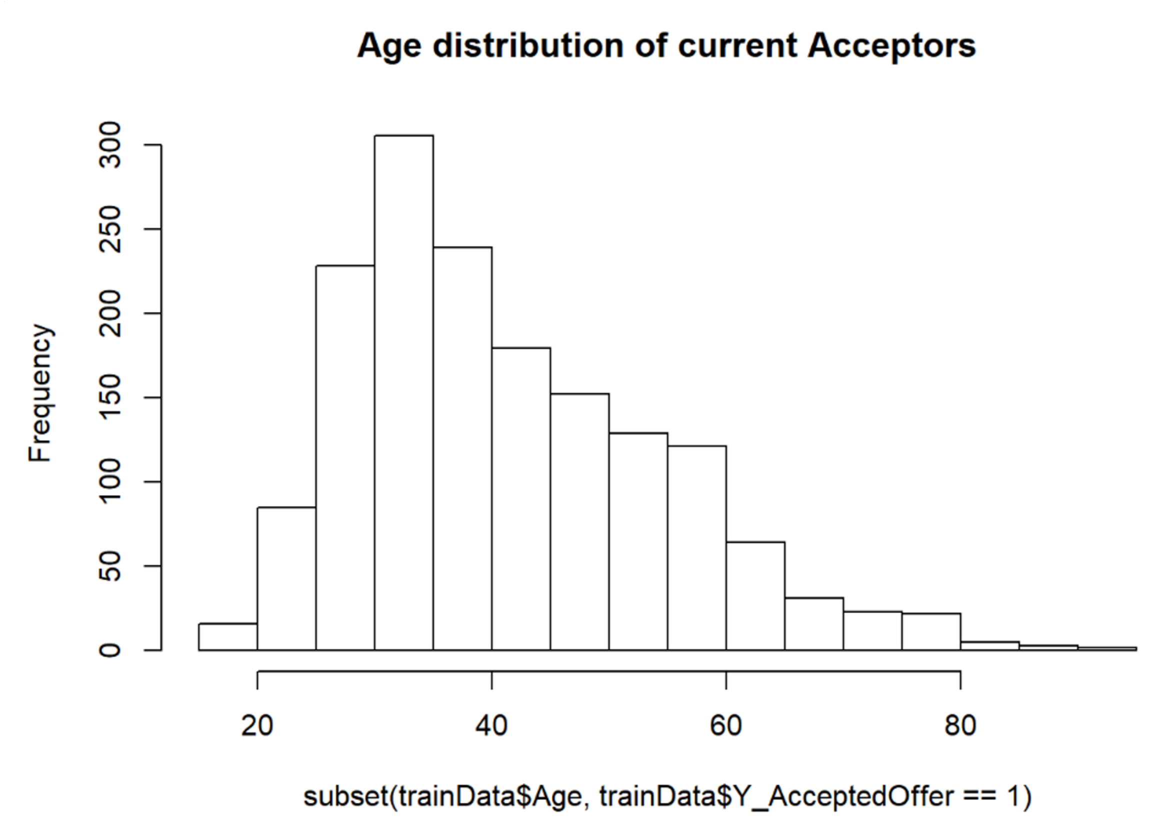
Divorced customers with primary educations & **single customers with tertiary educations** more likely to accept offer.

Married customers with primary or secondary educations **least likely to accept.**

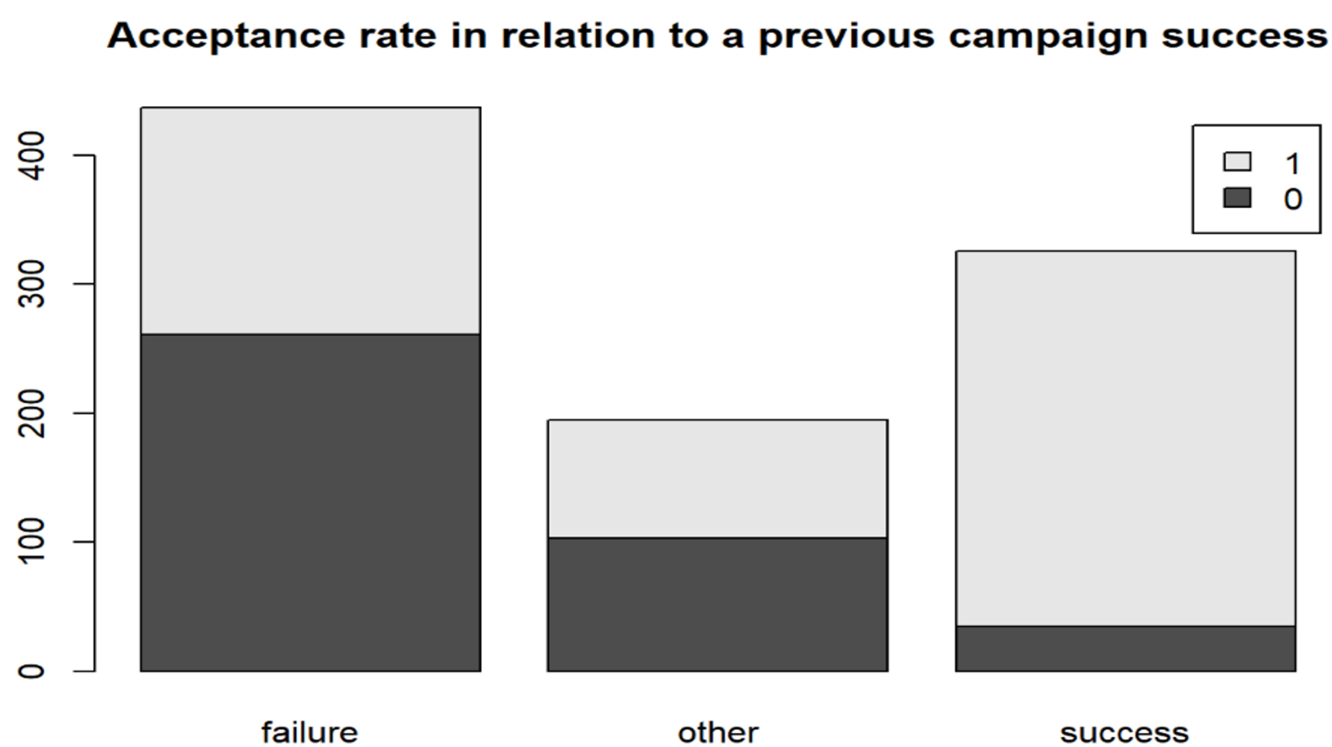


Age

The majority of the users that have accepted a loan range in ages 20 to 40



Previous Offer Acceptance



Summary of the current customers

- ➔ *Age 25-40*
- ➔ *Evenly spread among males and females*
- ➔ *Predominantly managers or technicians*
  (*retired have as well a good conversion rate ->opportunity)*
- ➔ *Single or married*
  (divorced customers having a much lower conversion rate)
- ➔ *Prefer to be contacted via mobile*
- ➔ *The customers that had already accepted an offer in the past are much more likely to accept an offer in this campaign.*

Other cleaning and enrichment steps

- After EDA, we removed the variables Call End, Call Start and duration of the call from the prospective data to match inquiry vectors used in both the training and scoring data sets.

- We dropped the car model, and the last contacted day. (Not informative enough)

Model

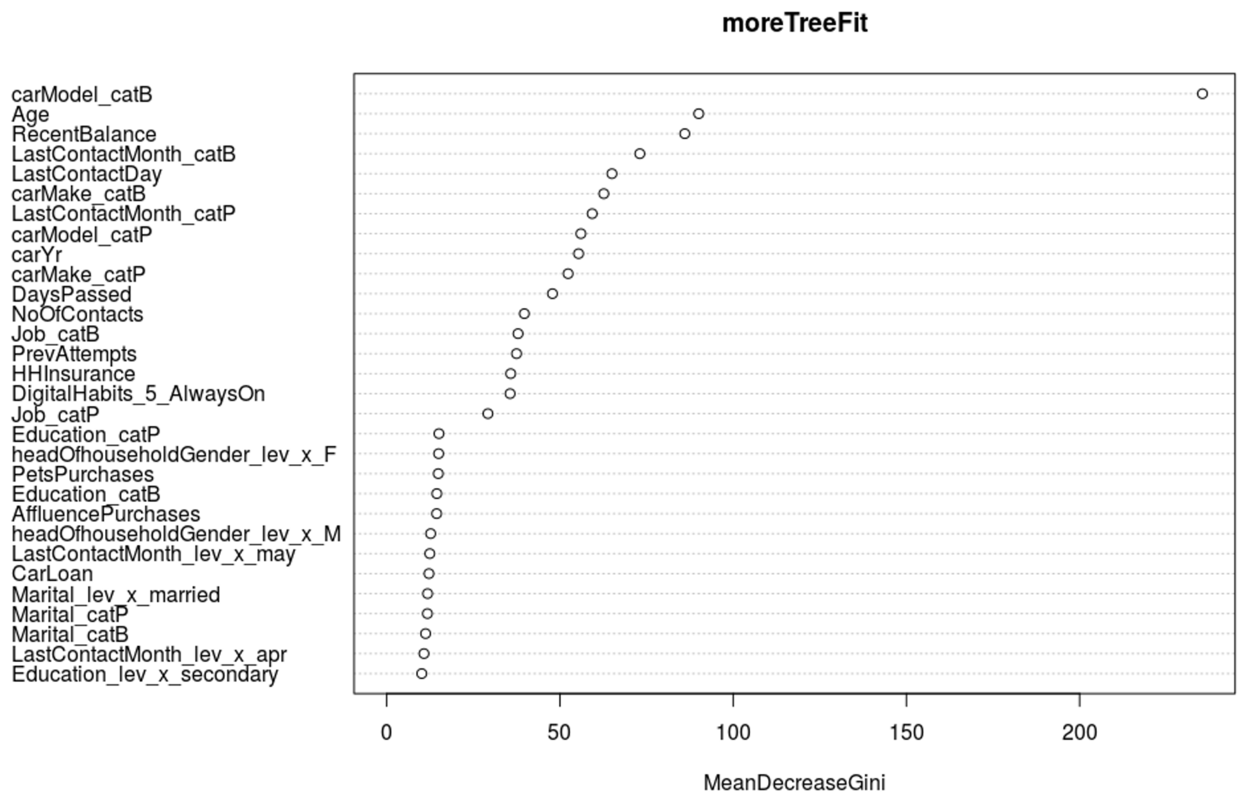
83%

Our model accurately predicts top 100 customers having an average 83% likelihood of accepting the financial product offer

We ran four models - KNN, Log Reg, Random Forest, and Decision Tree. We compared the accuracies across all the models. We chose and optimized a Random Forest model to predict the most likely prospective respondents.

Our Model

Found **carModel**, **Age** and **RecentBalance** among the most important predictive variables in accepting the product offer.



Your Top Prospective Customers

Our model finds these 100 prospects most likely to accept the offer.

```
> top100 <- head(successclasses, 100)
> top100
  [1] "HH7f6d4585ad" "HHdb87dc6e7a" "HH43c8759373" "HH1a33226c3d" "HH5ed73b551f" "HHc47c129a68"
  [7] "HH27aa9c576a" "HH24950c7841" "HH6edbc8212a" "HH446184e6cf" "HH7e5d9ebe4d" "HH7057883c68"
 [13] "HH43154ad8fd" "HH0793198f3e" "HH7f0f16440c" "HH09cdde08cf" "HHf1eed2cadb" "HH723000ae17"
 [19] "HH59847f671d" "HHe787f56b64" "HHf0b25e587f" "HH3fbb640624" "HHf134d2ac30" "HHa790aa13fe"
 [25] "HH07d8cade5d" "HHe8c454dfab" "HH161bc4c4c0" "HHad613bc2eb" "HHccc4bcb00c" "HHc78254e402"
 [31] "HHfbaf4748cd" "HHd5915c4d00" "HHe5252221ce" "HHa069b24baa" "HH38b558270f" "HHad9f6632a3"
 [37] "HH4f58ca91f4" "HH566a73acc0" "HH6c08f5b786" "HH6a5835f323" "HH2f94c20bf2" "HHccb6c83f3d"
 [43] "HH5a97ec205d" "HHf70acd4c5e" "HH7b64e60645" "HHc9b1d48c10" "HHdd7ab8dc96" "HHdde8e2bbbd"
 [49] "HH205b964fef" "HH6bc5e432d5" "HH7ada840fcc" "HHc7707daa69" "HH49df785161" "HHd60991b43c"
 [55] "HHee3f05a8af" "HH8e3c45eb7e" "HH621e6c0fdb" "HH1e772cc201" "HH7b3f008f2f" "HH69d287d315"
 [61] "HHa3aa29b7f2" "HH138070ca75" "HHcd2530dc33" "HHcb19d64c50" "HH5fbe40c17c" "HH84f50a3671"
 [67] "HH314fde17b5" "HH840291d606" "HH67cc96e24f" "HH3ec7dbbe37" "HH1d05a7ffb3" "HH59b703c8ce"
 [73] "HH9f3f799917" "HH714f3c26a2" "HH32827939bb" "HH6d48cca377" "HH971108bf95" "HH18cd0c68fd"
 [79] "HHc5c1770498" "HH495acba15a" "HHad15a796e1" "HHebaa51c47f" "HHc2d068f82f" "HHb395536c6c"
 [85] "HHa1727a4bf4" "HH09a83de3b5" "HHaa66459333" "HHc2ebfbbe83" "HH8cacb60619" "HHe742e0f52c"
 [91] "HH7beacbd3d2" "HHb2eec2c397" "HHca84cc79a6" "HH3ec1aac9f2" "HH22fd5cb8f6" "HH0e899bce96"
 [97] "HH011ae3bd0a" "HH39b55ca312" "HHee6f4eedcd" "HH5d7fa4c802"
```

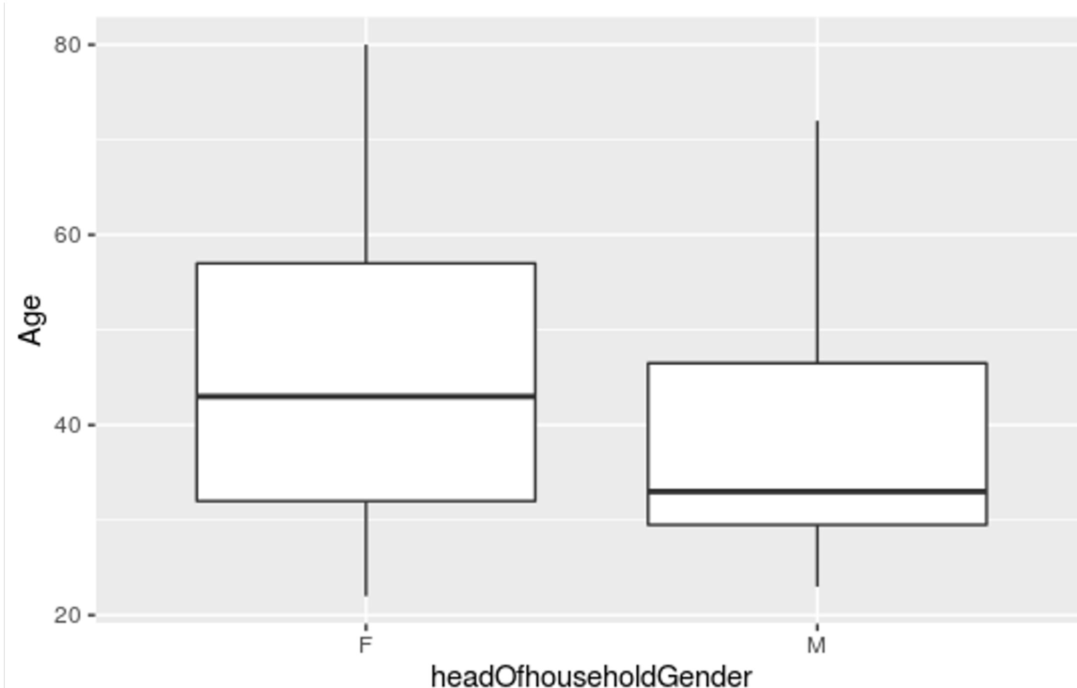
Prospective Targets - Insights

Your Top Prospective Customers

Median Age

F: 43

M: 33



Your Top Prospective Customers

*Top prospective customers have **24% higher recent balances** than current customers who accepted the line of credit.*

```
joinData %>% group_by(Y_AcceptedOffer) %>%
summarise(medianbalance = median(RecentBalance))
```

```
Y_AcceptedOffer medianbalance
  <fct>              <dbl>
1 Accepted            700
2 DidNotAccept        457
```

```
top100joined %>% group_by(headOfhouseholdGender) %>%
summarise(medianbalance = median(RecentBalance))
```

```
headOfhouseholdGender medianbalance
  <fct>                      <int>
1 F                           916
2 M                           926
```

Conclusion

Your prospective customers can be targeted by many factors. They are:

- Are more likely to be female
  - Potential for campaigns targeting middle-aged females
- Are more likely to be married
  - Potential for marketing campaigns targeting married couples with recent model cars.
- More likely to have higher bank balances
  - Targeting wealthier individuals, who may be more familiar with lines of credit products.