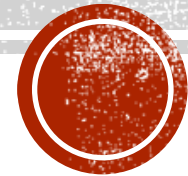


POTENTIAL CUSTOMERS ANALYSIS FOR NATIONAL CITY BANK



Jason Jia

CSCI-96 Data Mining for Business
July 2019 Harvard University

OUTLINE

1. Background/Introduction
2. Data Cleaning
3. EDA of the Current Marketing Results
4. Models
 - a. Compares 5 Techniques
 - b. Choose the Best Performance Model – Logistic Regression
5. Results & EDA of Predicted Top 100 Customers
6. Conclusion



BACKGROUND

🔗 Goal

- 🔗 Create a customer propensity model for a new product
- 🔗 Identify the next top 100 customers to reach.
- 🔗 Specifically run a line of credit against a household's used car

🔗 Data Resource

- ✂ 4000 historical data from previous marketing campaign



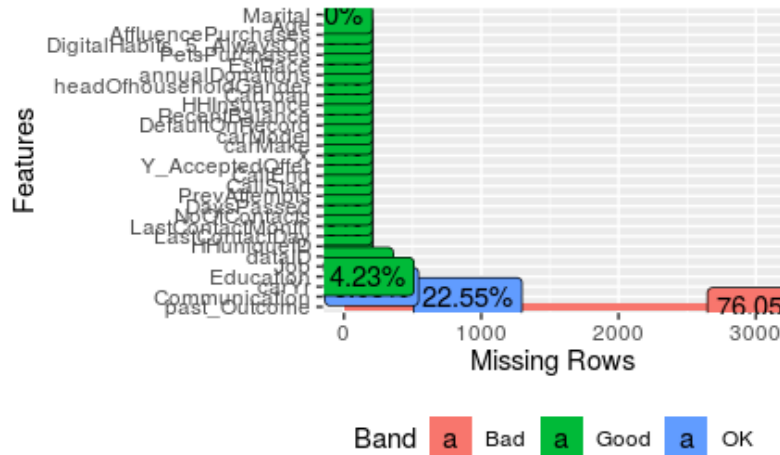
DATA JOINING AND CLEANING

Left-joined: “Current customer Marketing results” + “Household Axiom Data” + “Household credit data” + “Household vehicle data”

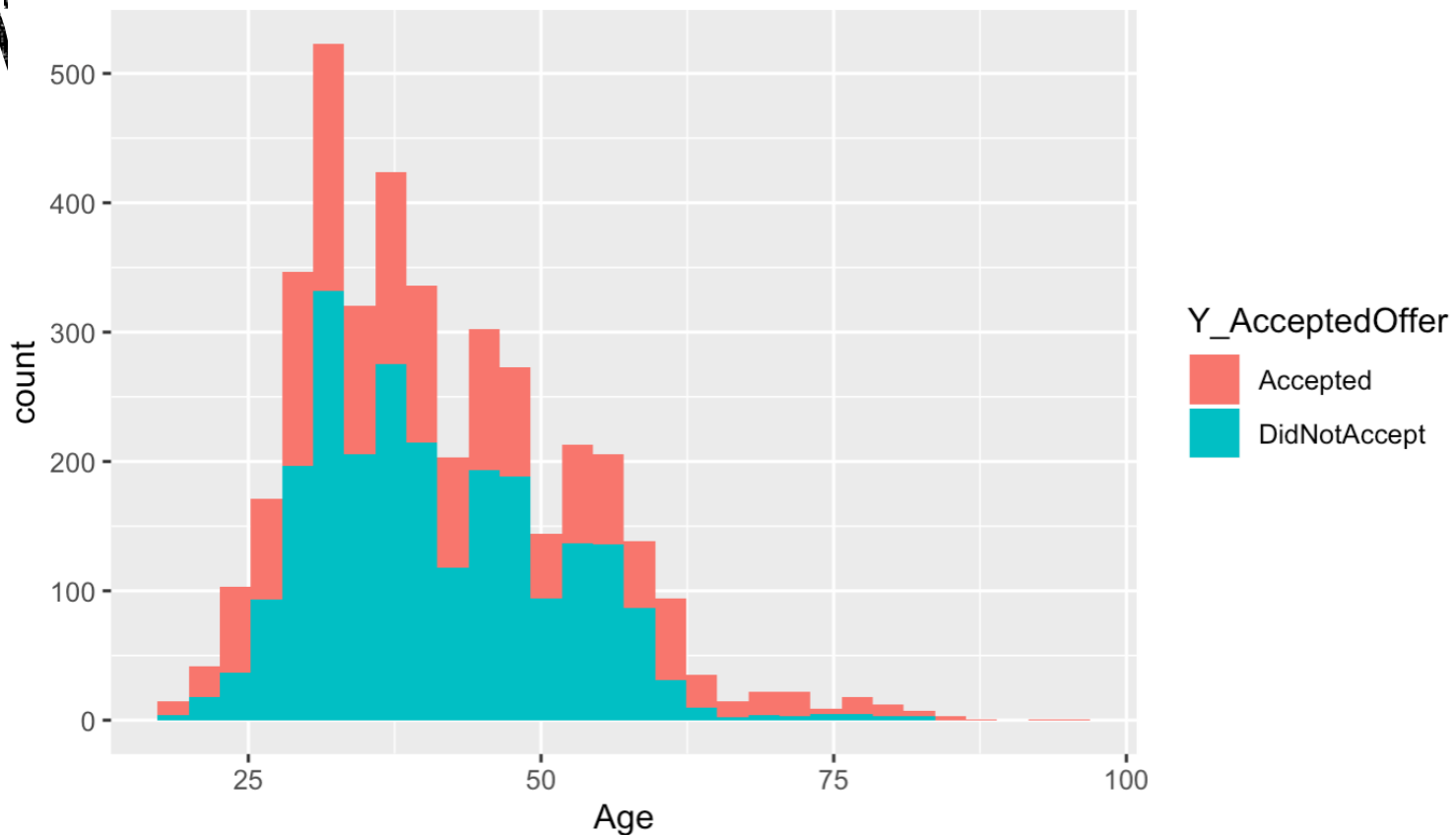
```
> names(clean.data)
```

[1] "X"	"LastContactDay"
[3] "LastContactMonth"	"NoOfContacts"
[5] "DaysPassed"	"PrevAttempts"
[7] "Y_AcceptedOffer"	"carMake"
[9] "carYr"	"headOfhouseholdGender"
[11] "annualDonations"	"EstRace"
[13] "PetsPurchases"	"DigitalHabits_5_AlwaysOn"
[15] "AffluencePurchases"	"Age"
[17] "Job"	"Marital"
[19] "Education"	"DefaultOnRecord"
[21] "RecentBalance"	"HHInsurance"
[23] "CarLoan"	

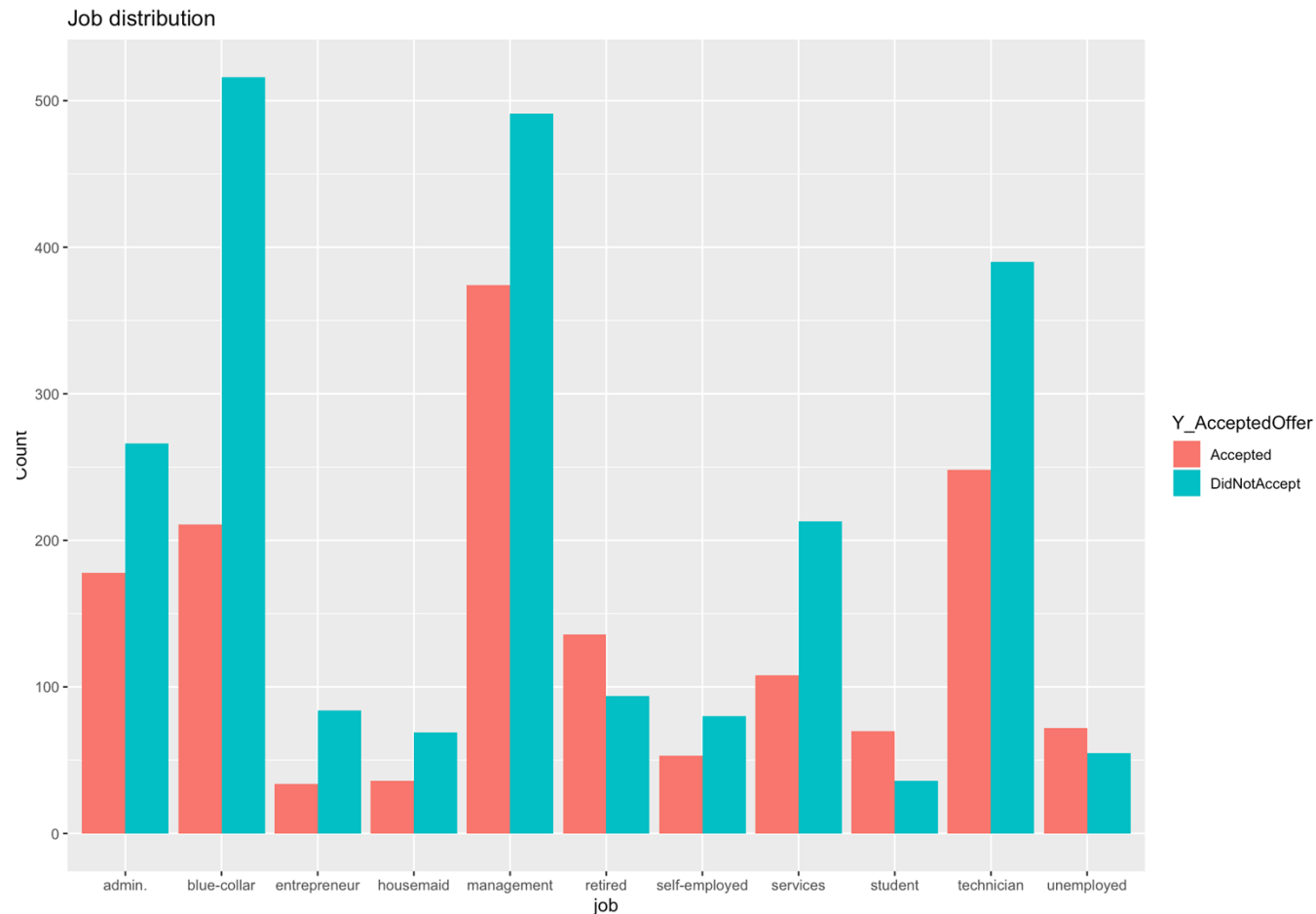
:-inc



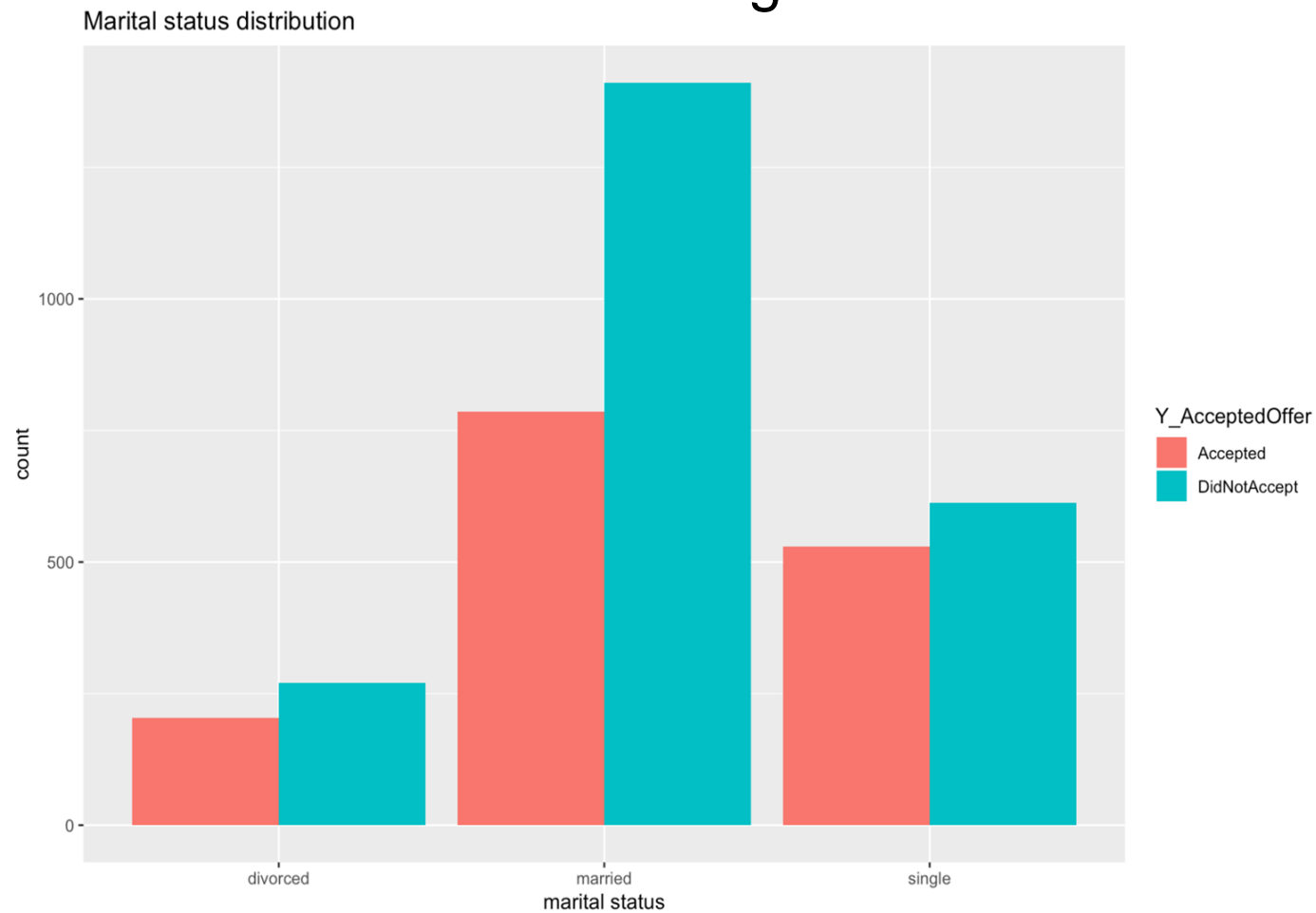
THE DEMOGRAPHICS OF OUR CURRENT MARKETIN



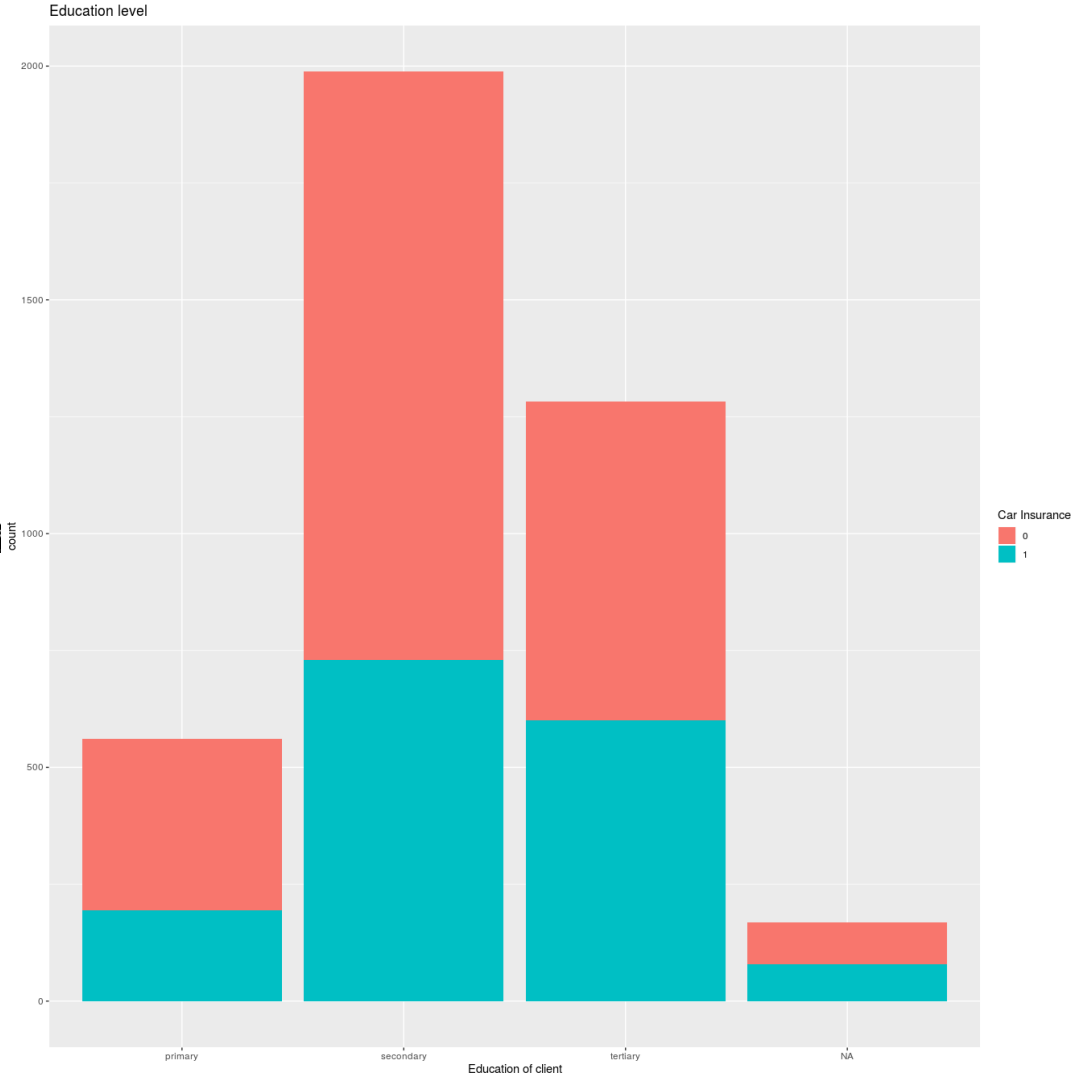
The Demographics of Our Current Marketing Results



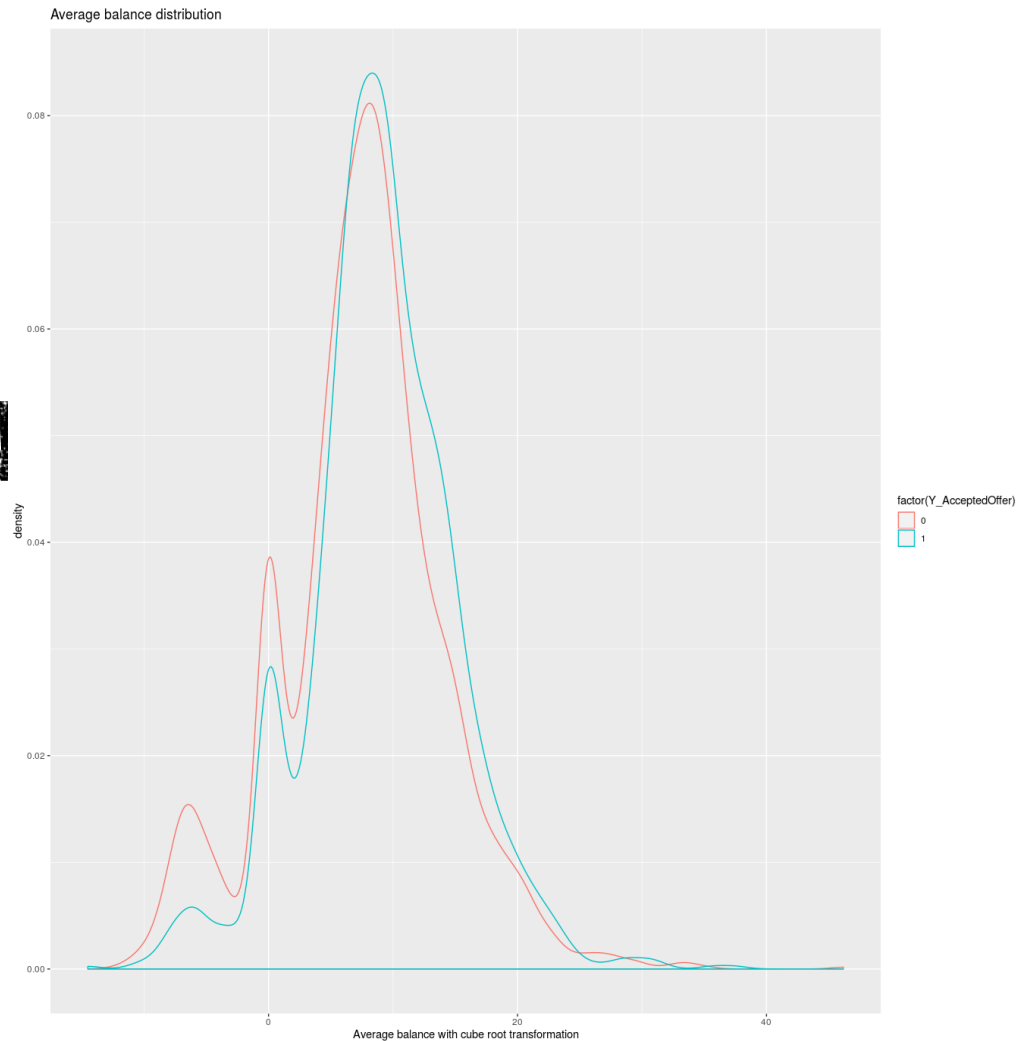
The Demographics of Our Current Marketing Results



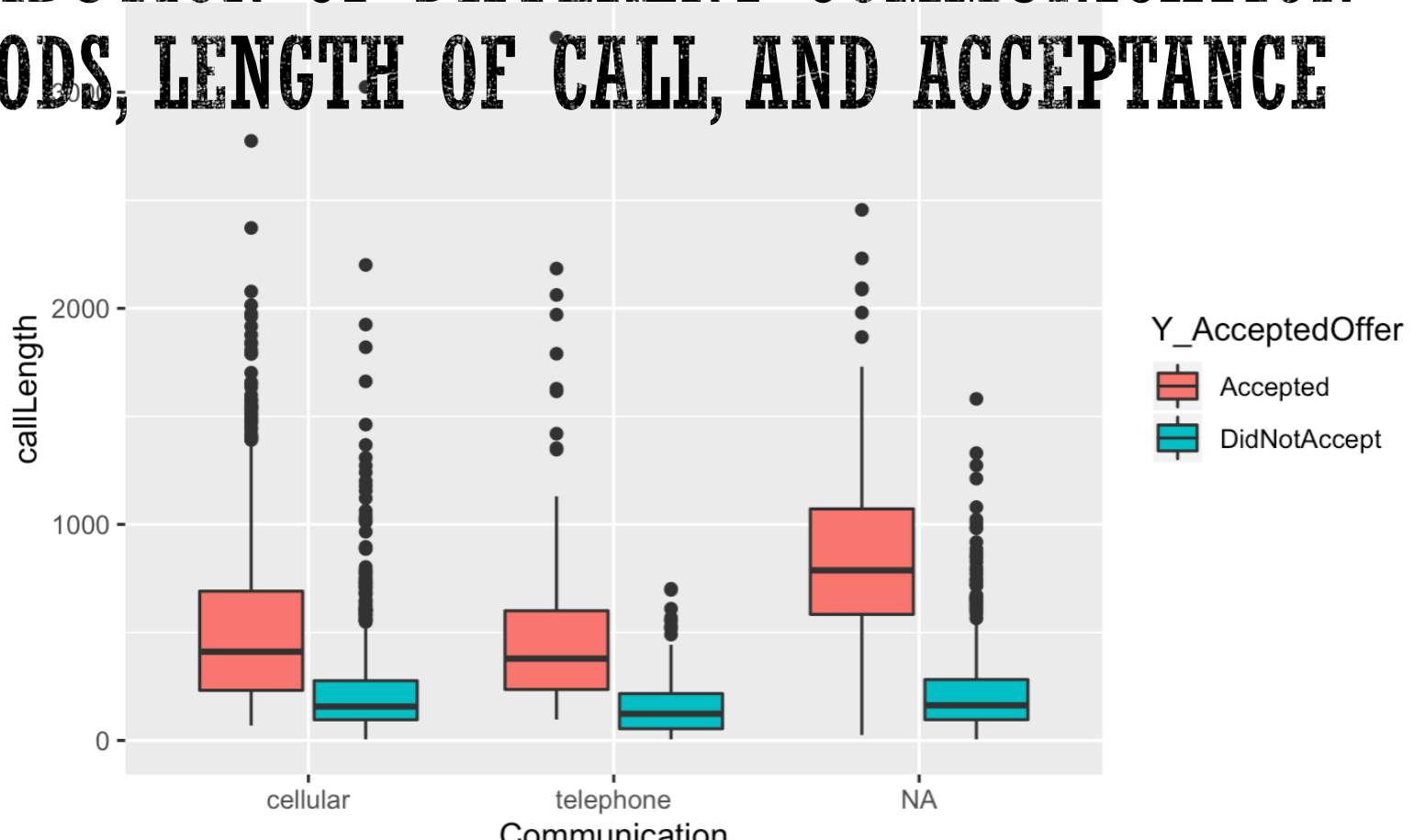
THE EDUCATION AND CAR LOAN STAT



THE AVERAGE BALANCE DISTRIBUTION

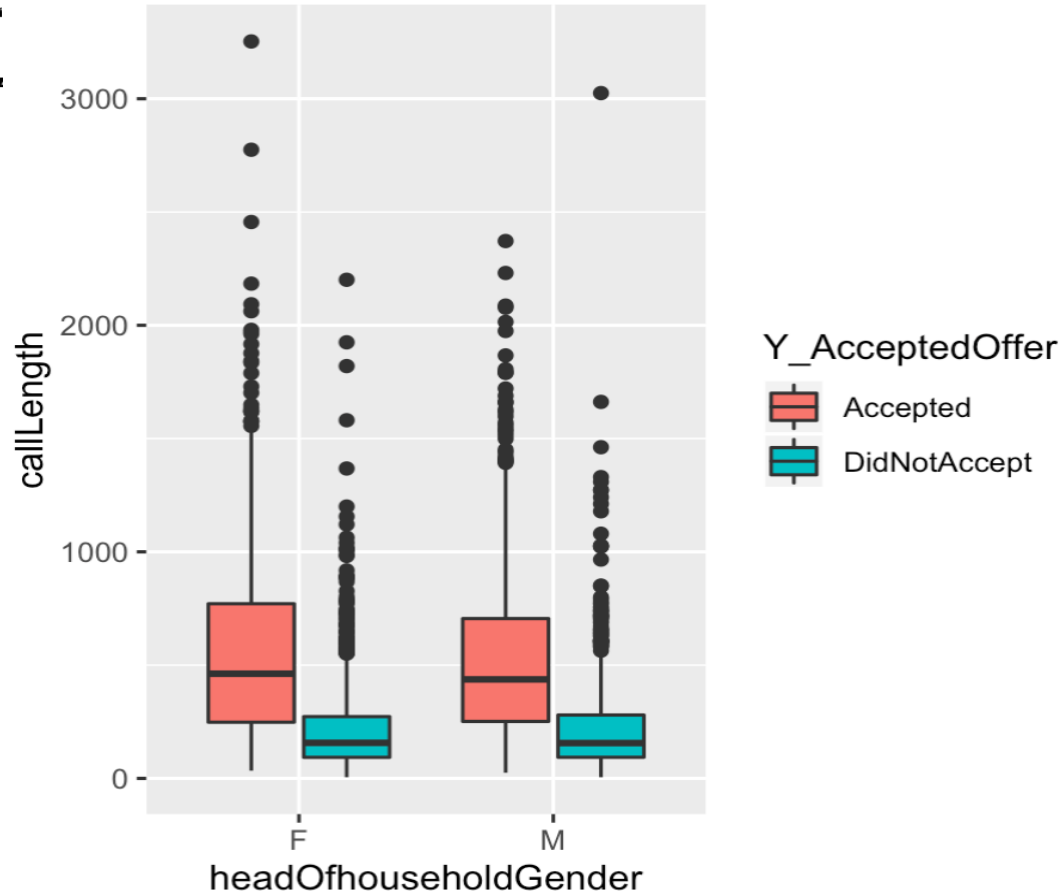


DISTRIBUTION OF DIFFERENT COMMUNICATION METHODS, LENGTH OF CALL, AND ACCEPTANCE



DISTRIBUTION OF DIFFERENT COMMUNICATION METHODS, I

ANCE

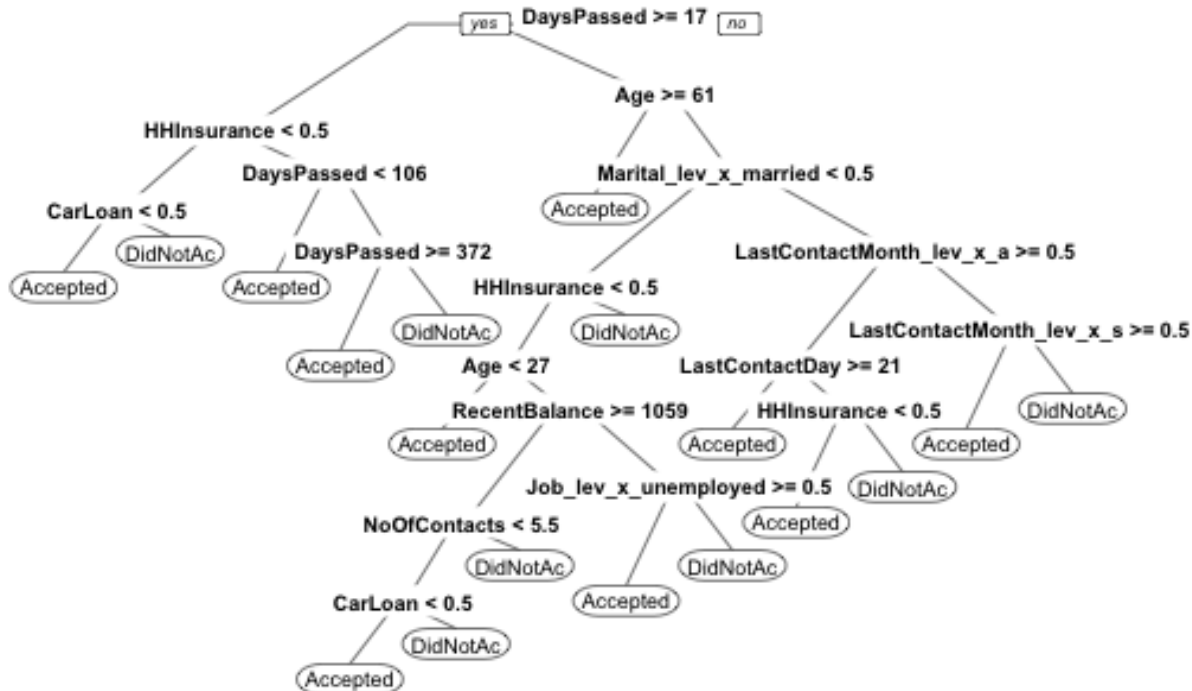


THE ACCURACY OF DIFFERENT MODELS

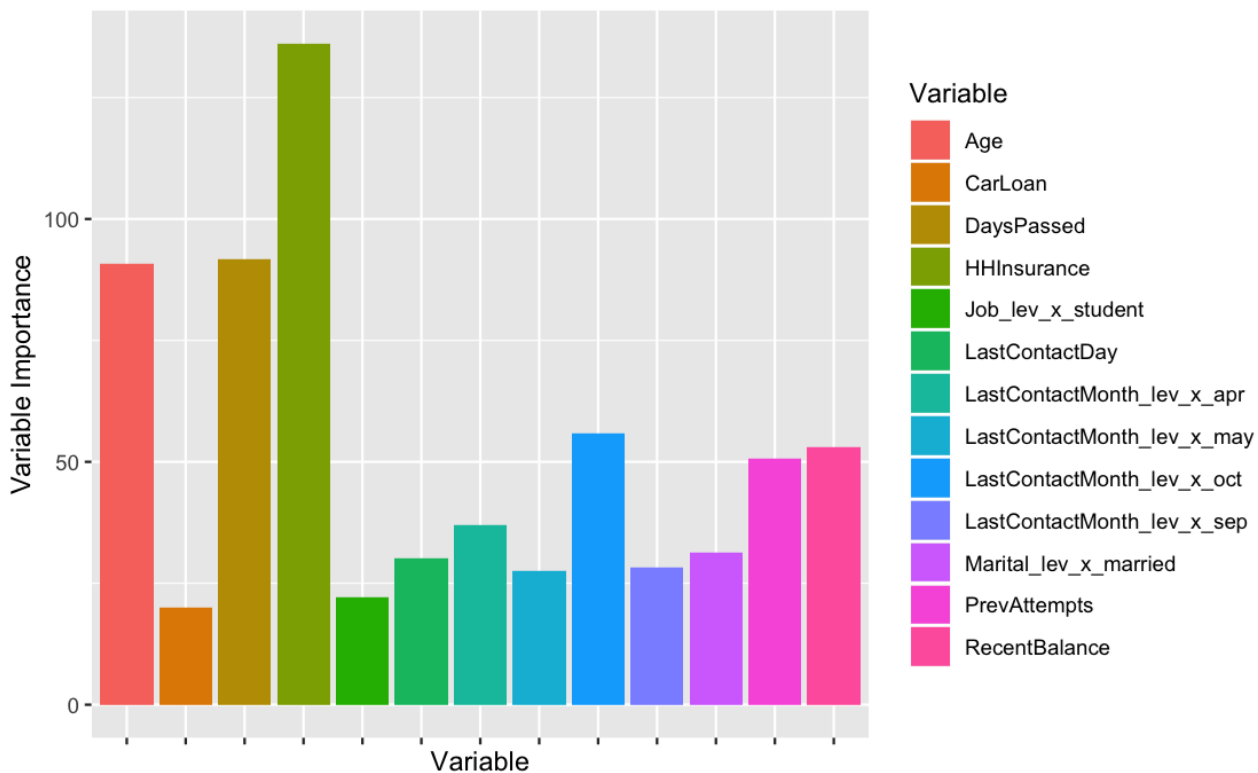
Model Type	Testing Accuracy
KNN	71%
Logistic	75%
Neural Net	62%
Classification Tree	69%
Random Forest	71%



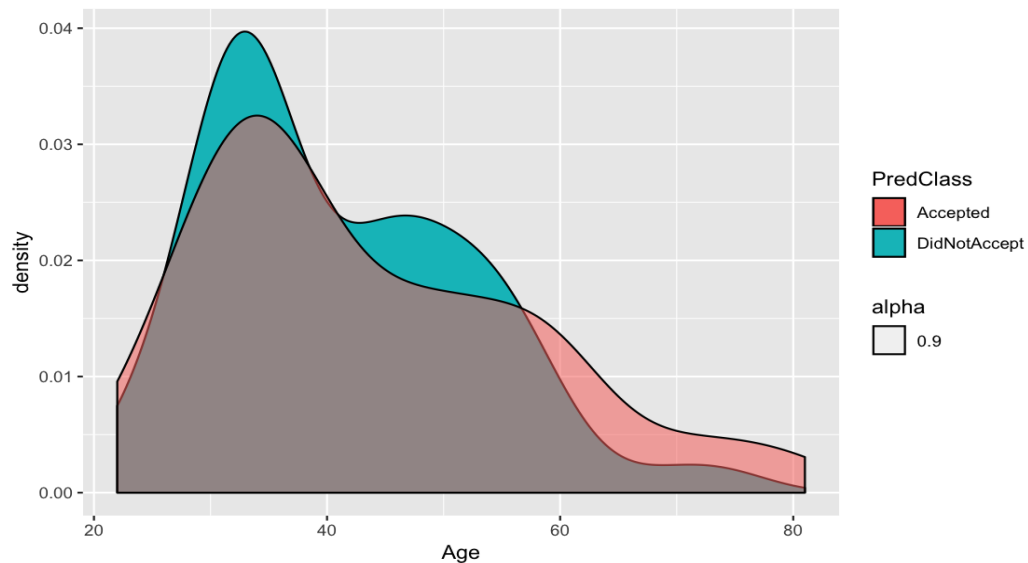
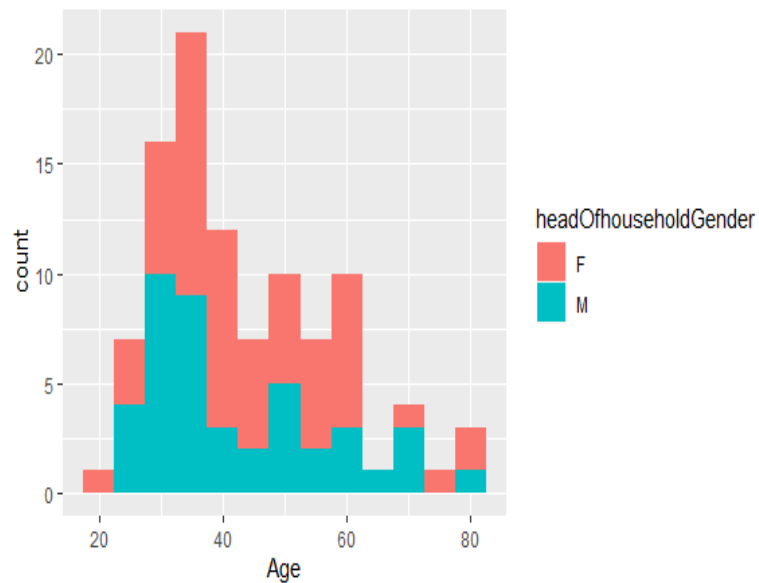
CLASSIFICATION TREE

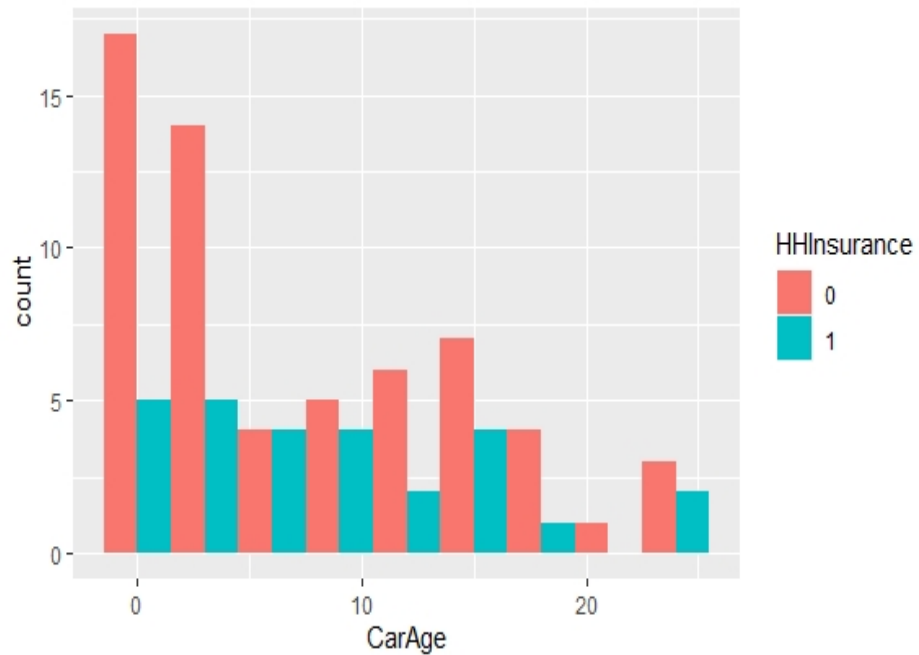
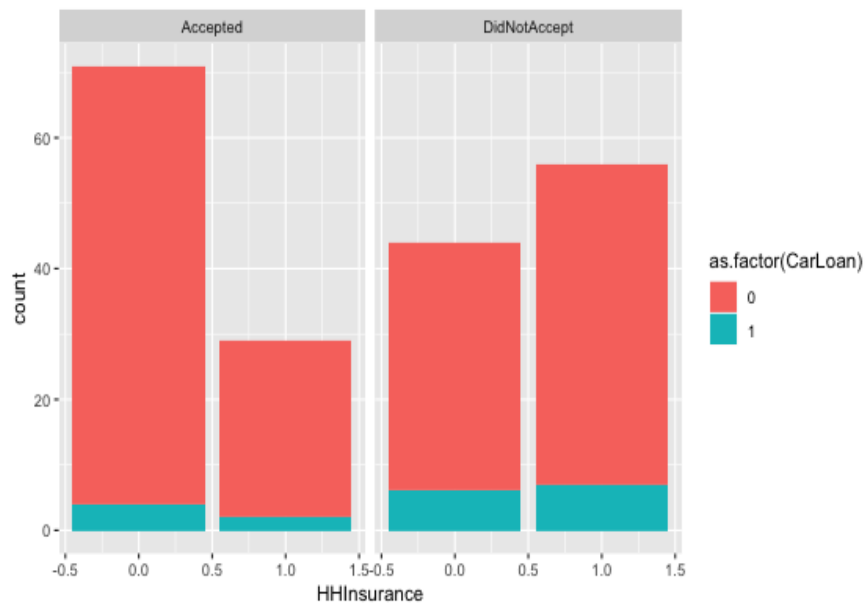


THE VARIABLE IMPORTANCE

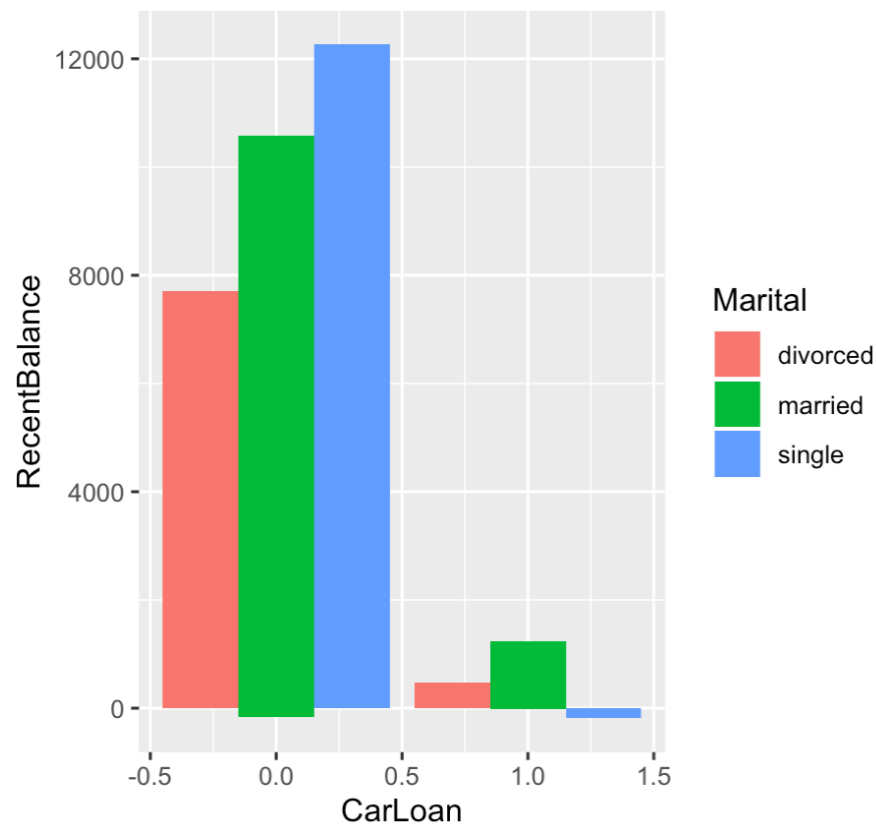


THE PREDICTED CUSTOMER DATA

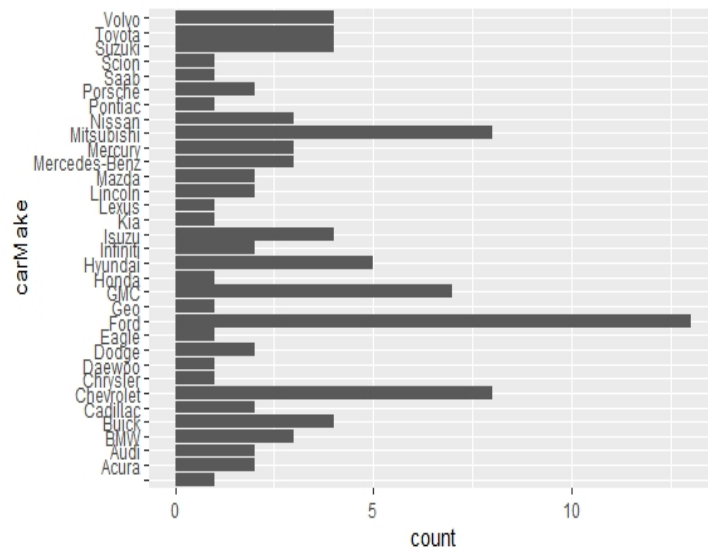
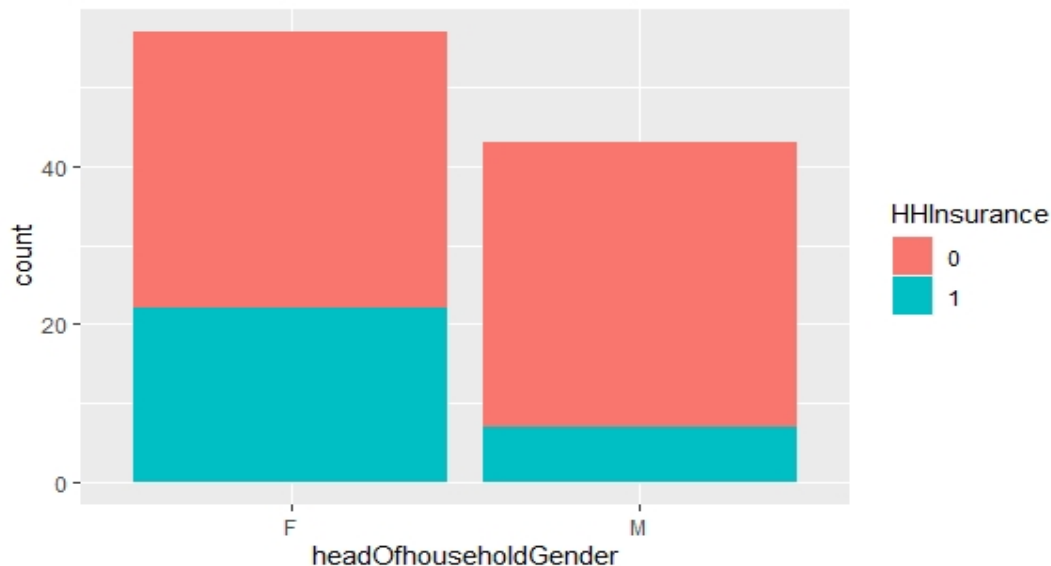




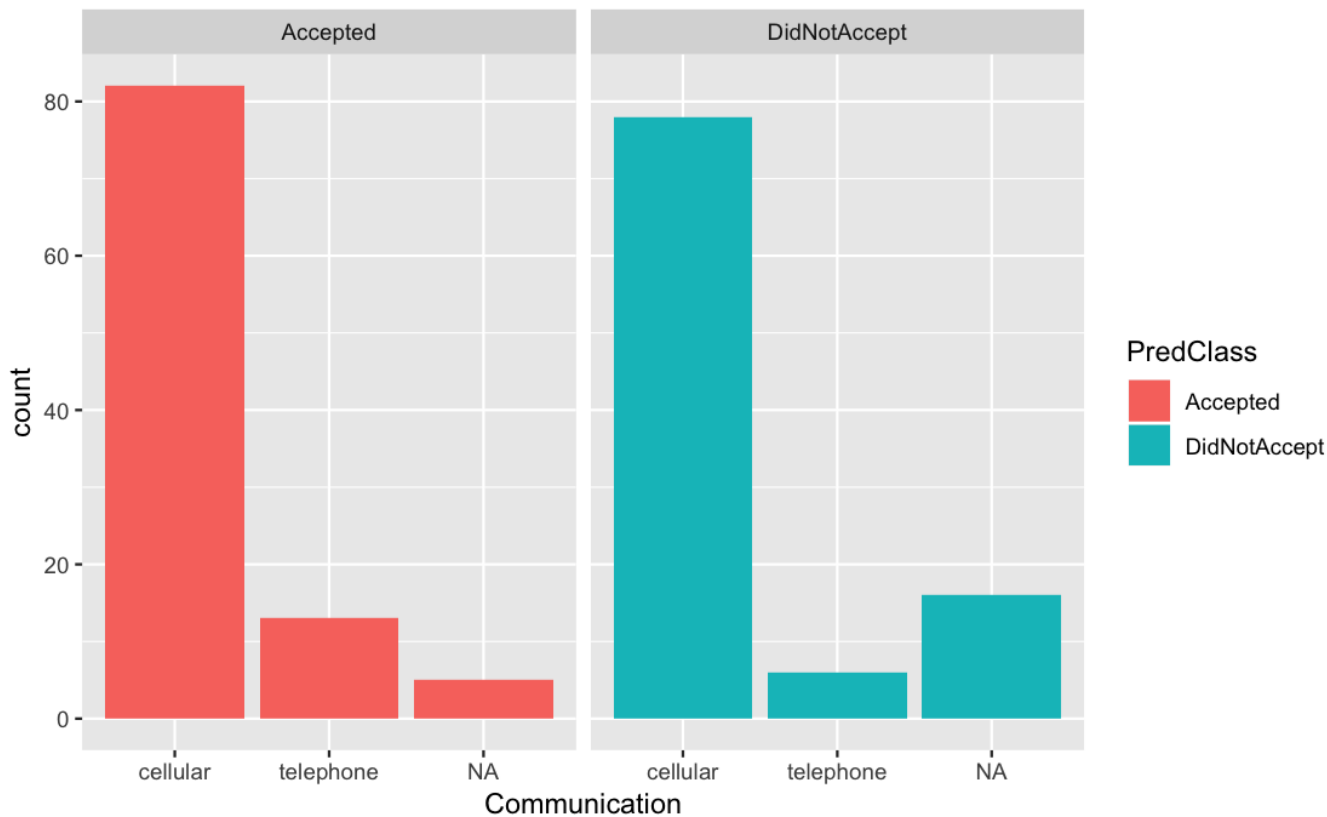
The difference between groups of
Different Marital, Car Loan,
and Recent Balance

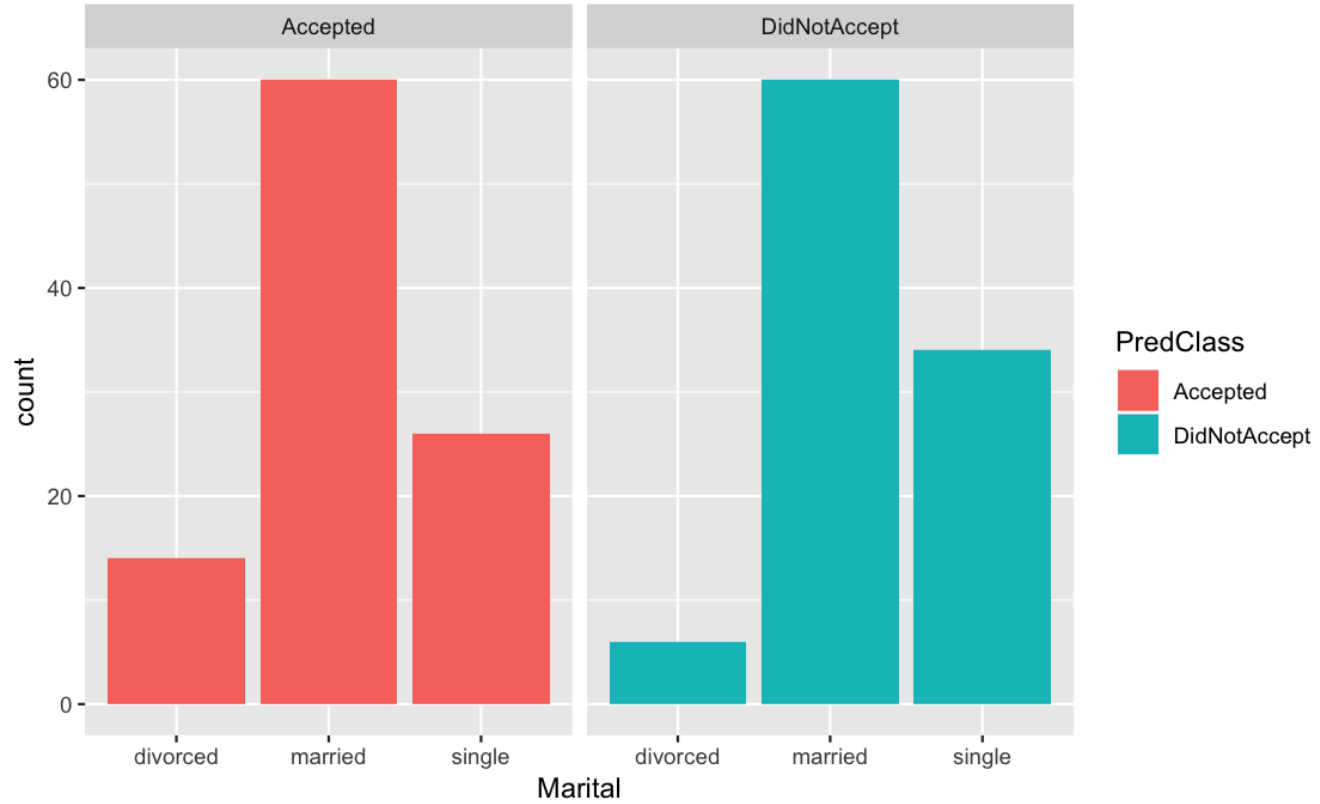


GENDER & INSURANCE AND CAR LOANS



BASED ON LOGISTIC MODEL PREDICTIONS





CONCLUSION

- 🔗 Demographics of the Current Marketing Results
- 🔗 Run 5 Models of Supervised Classification
- 🔗 Select the Best Model ---- Logistic Regression
- 🔗 Create EDA for the Next 100 Customers' Profile

