Highway Amenities Prediction

Team 1

Business Case

Problem Statement:

- Identify the need to develop correct highway amenities for interstate drivers, by reducing waste and building of unnecessary amenities.

Solution Offered

- Employ predictive analytics to interpret drivers' preferences from coupon response data and help figure out what kinds of services drivers like most, by constructing best options that make trips better and increase profitability.

Methodology and Work Flow

Data Review

Data Review

Dataset:

id: unique number giving to the driver doing the survey
destination: No Urgent Place, Home, Work
passanger: Alone, Friend(s), Kid(s), Partner (who are the passengers in the car)
weather: Sunny, Rainy, Snowy
temperature:55, 80, 30
time: 2PM, 10AM, 6PM, 7AM, 10PM
coupon: Restaurant(<$20), Coffee House, Carry out & Take away, Bar, Restaurant($20-$50)
expiration: 1d, 2h (the coupon expires in 1 day or in 2 hours)
gender: Female, Male
age: 21, 46, 26, 31, 41, 50plus, 36, below21
education: Some college - no degree, Bachelors degree, Associates degree,..
occupation: Building & Grounds Cleaning & Maintenance, Farming Fishing & Forestry
.
.
.
Y: 1, 0 (whether the coupon is accepted)

Highway operation dataset

10K Records

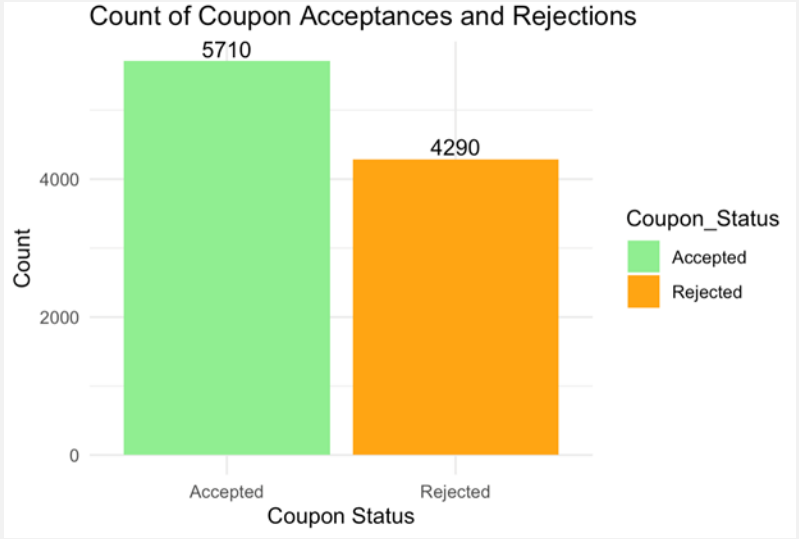
25 Columns

1 – Outcome -> Y

24 – Predictors

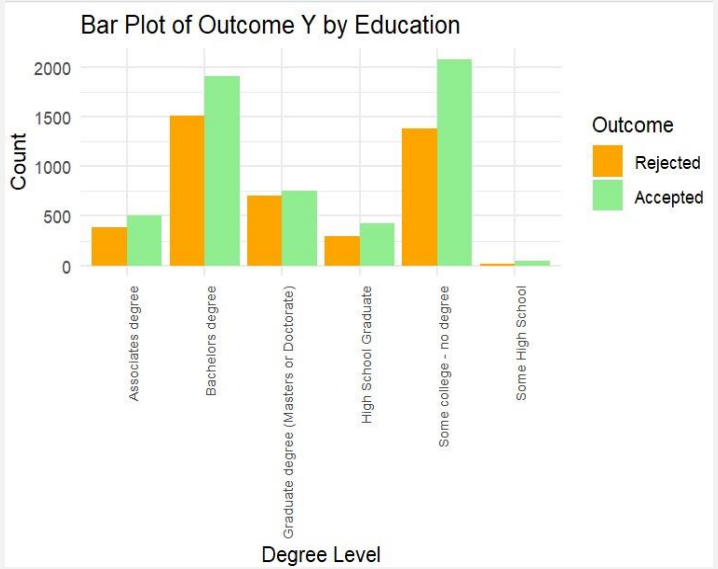
Result – Prediction -> Y

EDA



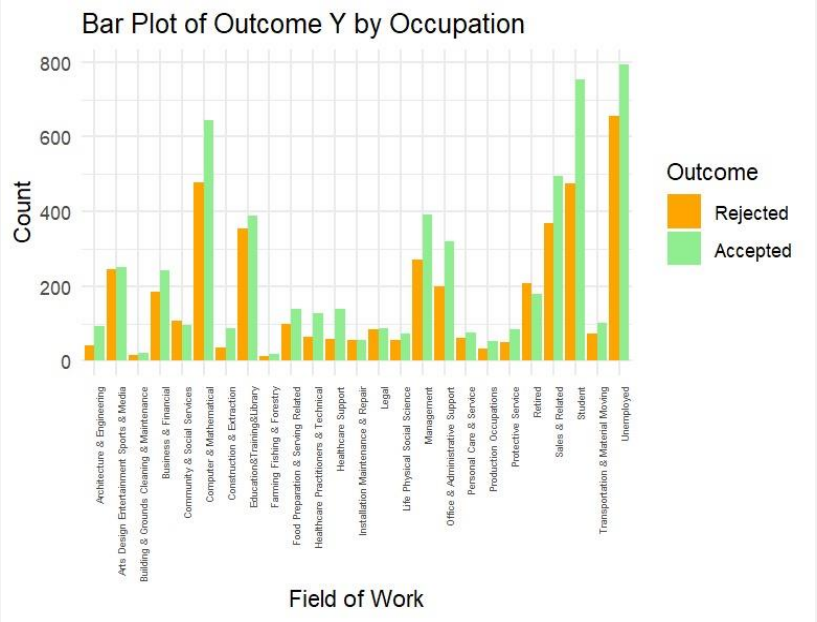
- 5,710 coupons accepted versus 4,290 rejected.
- More than 50% acceptance rate suggests effectiveness.

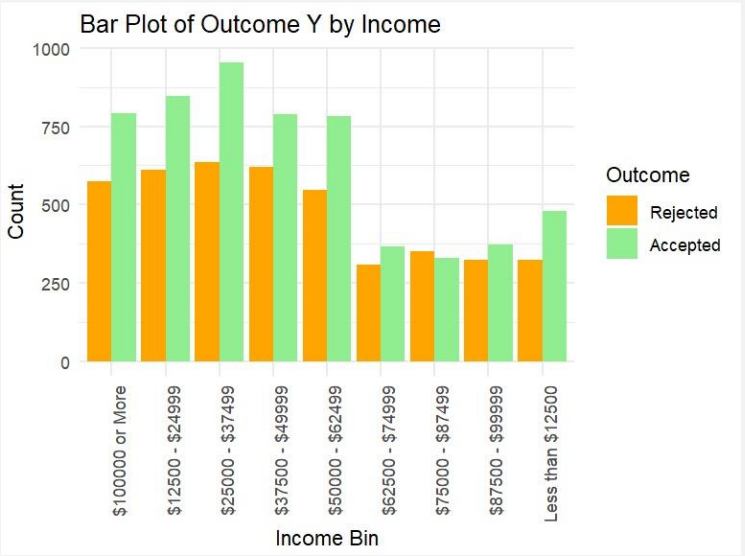
EDA-I



- Highest coupon acceptance rates observed among individuals with "some college - no degree."
- Bachelor's Degree Holders are Second-highest acceptance rate of coupons.
- High school degree is not important

- Occupation group possibly related to 'Unemployed' shows highest acceptance, nearing 800.
- High acceptance occupations could be targeted more aggressively.
- Occupation emerges as a significant factor influencing outcomes.

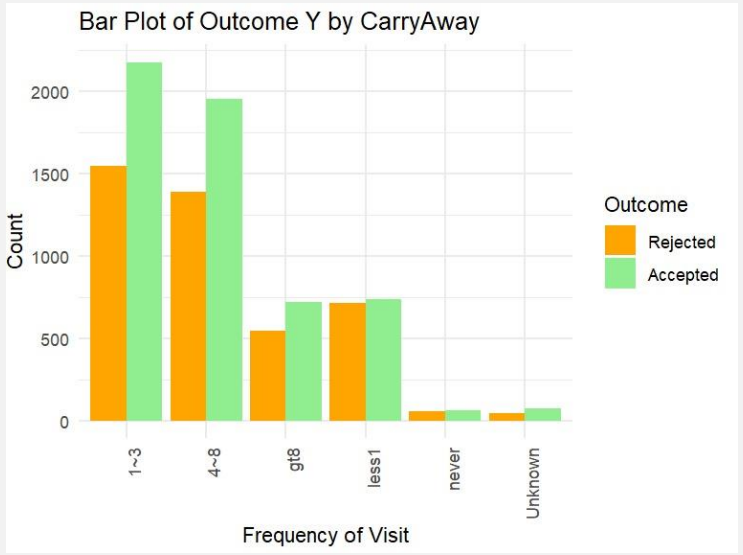


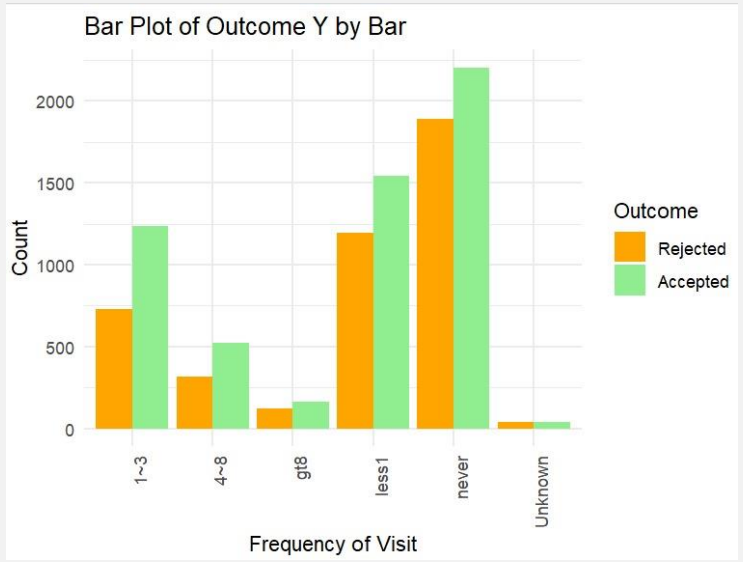


• Correlation observed between income levels and acceptance/rejection of Outcome Y.
• Acceptance counts increase with income, peaking at "$25,000 - $37,499" bracket, nearing 1,000 counts.

EDA-II

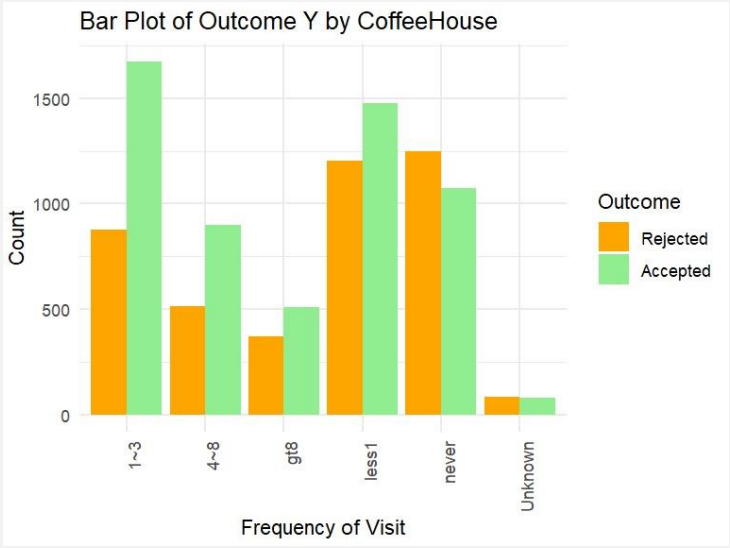
• ”1~3" and "4~8" categories show significantly higher counts for both acceptance and rejection.
• Less frequent interaction with "CarryAway" linked to lower acceptance and rejection counts.
• 1~3 frequent users may be more receptive.



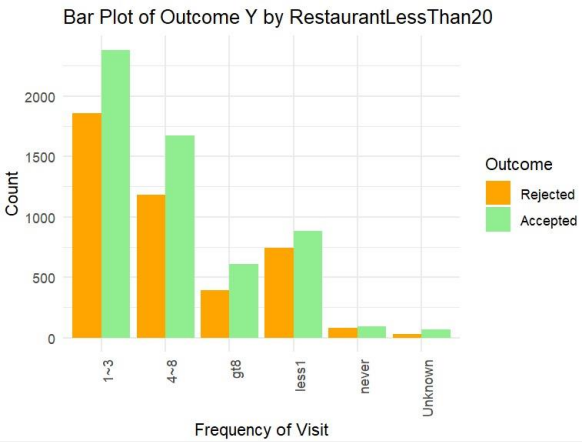


• ”never" group has highest count for both accepted and rejected outcomes.
•"1-3" group shows very low counts for both acceptance and rejection, expected due to lack of decision-making independence.
• " never" could be primary target for marketing strategies or interventions related to Outcome Y, given their high engagement.

• ”1~3” users shows the highest count to accepted
•"Never" and "Unknown" categories show lowest counts for both acceptance and rejection.

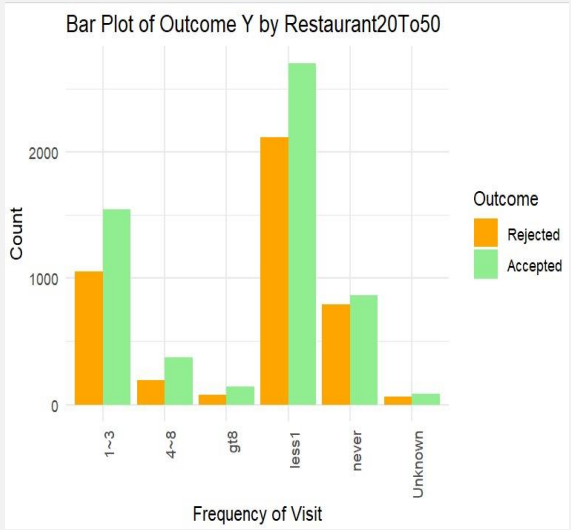


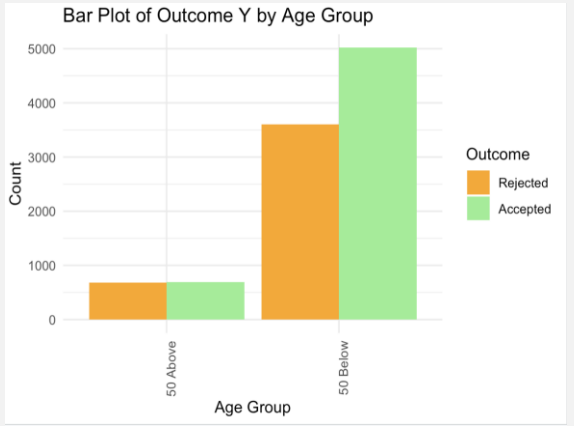
EDA-II



• Frequency category "1-3 times" shows significantly high acceptance count, suggesting strong engagement.
• Increase in frequency of visits correlates with decline in both acceptance and rejection counts.
• "Never" and "Unknown" categories exhibit very low counts for both acceptance and rejection.

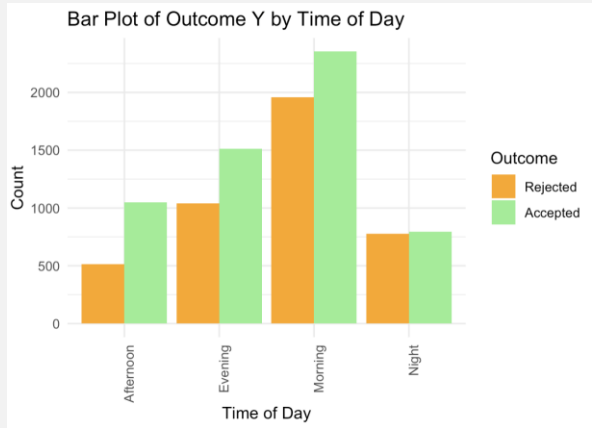
• “less1 times" frequency category shows highest acceptance count, indicating strong engagement with Outcome Y.



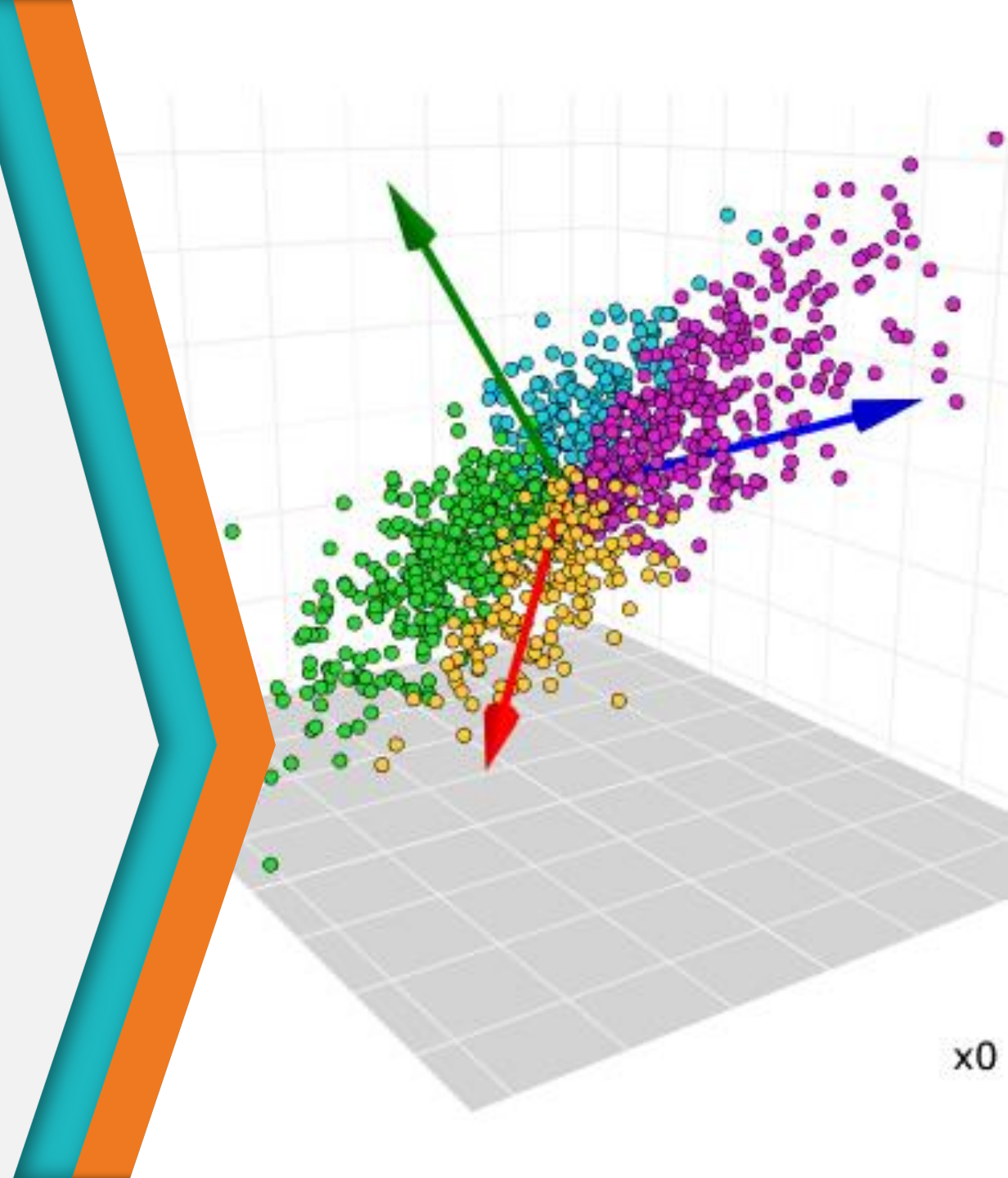


• "Above 50" age group exhibits substantially higher acceptance count and very few rejections.
• "Above 50" age group shows overwhelming preference for acceptance, indicating strong resonance with Outcome Y.

• Morning time slot shows highest number of acceptances, exceeding 2,000 counts.
• Both Afternoon and Night show more rejections than acceptances, indicating less favorability or relevance during these times.



Feature Engineering



Feature Engineering

1. Converted characters to factors.
2. Created a data frame consisting of columns of interest.
3. Filled blank data with an Unknown string for CarryAway, Bar, CoffeeHouse, Restaurant20To50, RestaurantLessThan20
4. Divided Age Group into 2 parts 50 Above and 50 Below.
5. Converted Time into 4 categories: Morning, Afternoon, Evening, and Night.

Feature Engineering

Before (highway_data)
- Existing Columns

After (highway_data)
- Selected Columns

Time

7AM, 10 AM — **Morning**

2 PM — **Afternoon**

6 PM — **Evening**

10 PM — **Night**

Time of Day

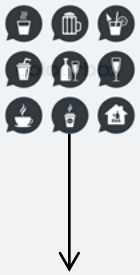
AGE

50 Above

50 Below

50 Above – 1378
50 Below – 8622

Bar, Coffee, Car away, Restaurant columns

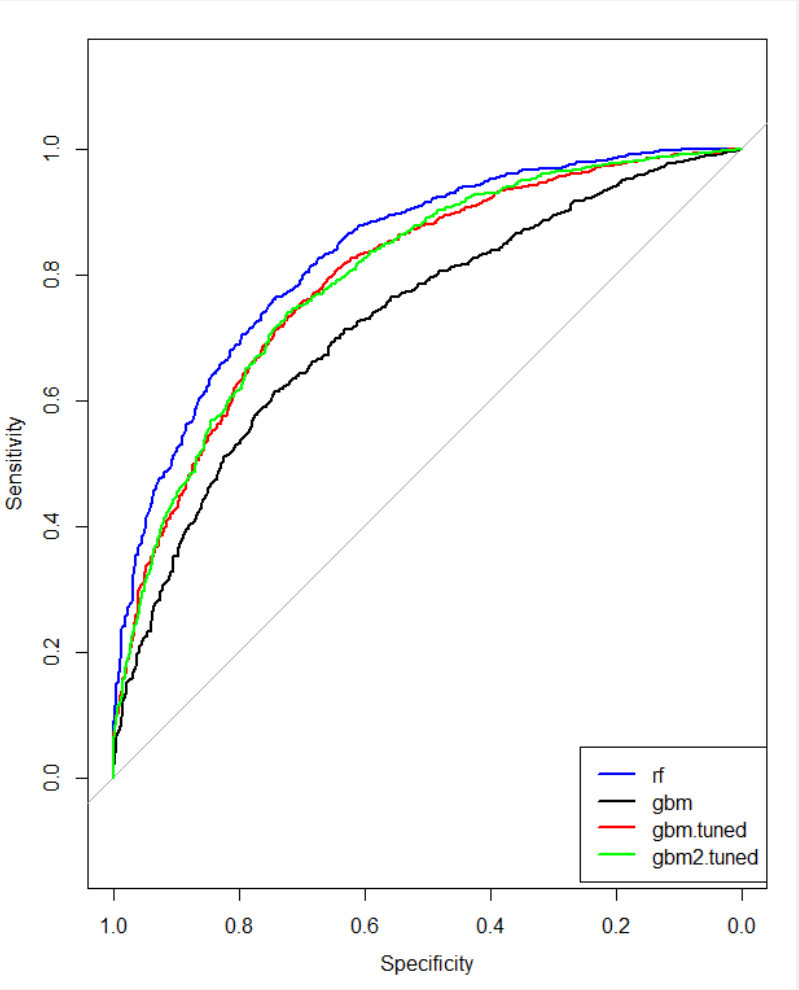


Added “Unknown” to blank columns

Model Review, Comparison & Interpretation

Model Review

Sr No	Model Name	Accuracy	F1 score
1.	Logistic Regression	55.51%	64.43%
2.	Random Forest	76.06%	80.21%
3.	Gradient Boosting	66.65%	72.54%
4.	GBM Tuned_1 (cross-validation = 20)	73.20%	75.66%
5.	GBM Tuned_2 (cross-validation = 20, repeated 5 times)	71.80%	74.15%



Selected Model and Feature of Importance

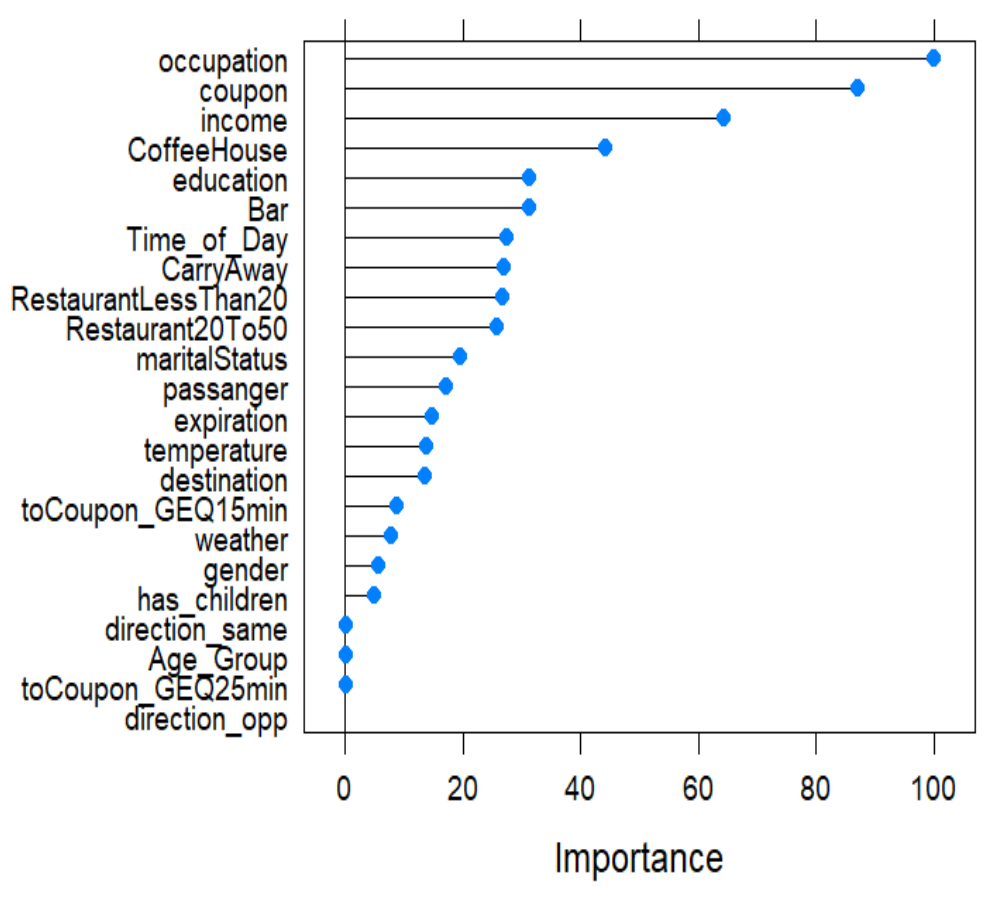
Selected Model **Random Forest (rF)** **Accuracy = 76.06%** **F1 Score = 80.21%** **80:20 Split**

Amenities:
- Coffee House
- Bar
- Carry Away
- Restaurant <20$
- Restaurant $20-$50

Important Features:
- Occupation
- Coupon
- Income
- Education
- Time of Day

Top Predictors:
- Occupation
- Income
- Education

Insights:
- Economic and social status play a role in amenity preference
- Preferences vary with the time of day, hinting at potential peak hours for specific amenities.



Implications & Recommendations

Implications and Recommendations - I

Important Variable	Implications	Recommendations
Occupation	The feature occupation shows the highest importance, suggesting that a driver's job significantly influences whether they accept the coupon and hence their amenity preferences	Establish partnerships with affordable chains or local vendors that can provide cost-effective options. Offer special discounts or loyalty programs for repeat customers to encourage ongoing patronage.
Coupon	The coupon and is also highly influential in this dataset, which can show the preference for the amenity	Since coffee houses are popular, providing a premium coffee experience could be a significant draw. That will be a good advice that building more coffeehouse with a range of options from standard brews to specialty drinks
income	Showing that financial considerations affect amenity preferences	Collaborate with popular mid-range amenity chains or support local eateries that can provide quality product at prices that align with middle-income budgets.

Implications and Recommendations - II

Important Variable	Implications	Recommendations
Education	Education is the fourth important factor influence.	Most people are around bachelor degree and less, we can make coupon rule more straightforward and directly.
Time of day	The acceptance of a coupon varies with the time when the driver is traveling.	Consider the operational hours of the amenities. Based on the bar chart, we recommend the operational time is from 7am to 6pm

Thank you


R
N
P
D
Memedroid
car won't start
until it gets coffee


DRIVER: "WHAT AM I SUPPOSED TO DO WITH THIS SPEEDING TICKET?"
OFFICER: "KEEP IT, WHEN YOU COLLECT FOUR OF THEM, YOU GET A BICYCLE."