



# Predicting In-Vehicle Coupon Acceptance

## Group 11

Rushika Badri Prasad

Trupti Shailendra Gandhi

Girish Sunkadakatte Chandrappa

# Problem Statement

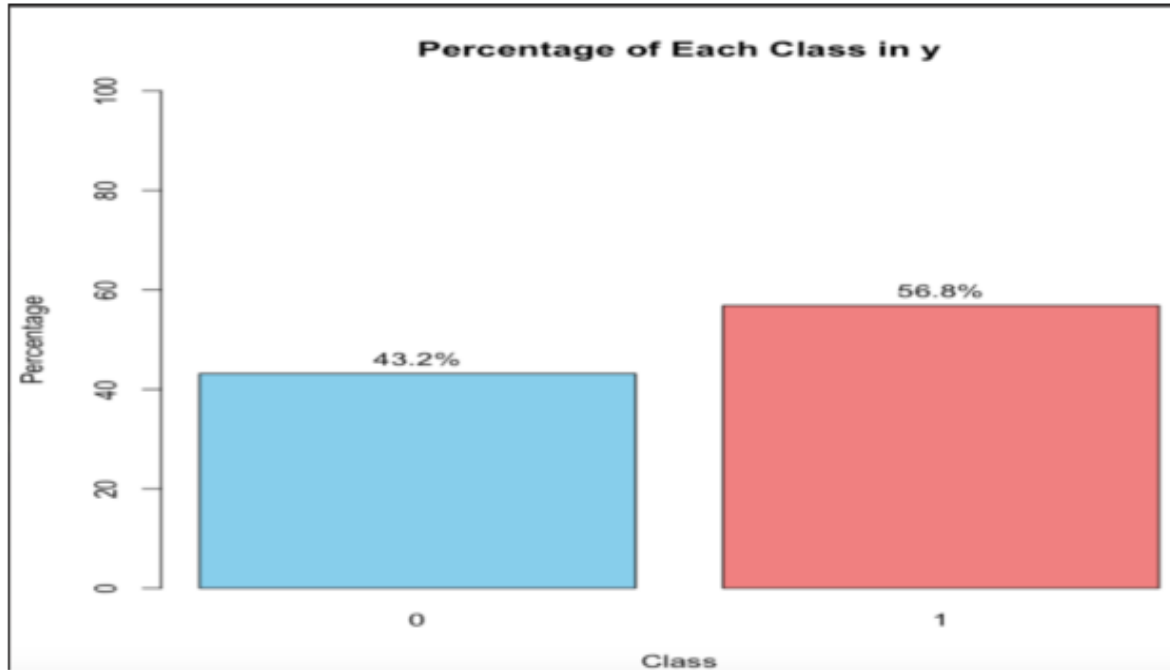
Predict whether a customer will accept an in-vehicle coupon based on demographic, contextual, and behavioral factors to help businesses optimize marketing strategies and improve engagement.

# Dataset Description

- The dataset comprises **12,684 observations and 26 features**, including both **categorical and numerical variables** from various domains, aimed at understanding the factors influencing coupon redemption behavior.

Category	Features
Demographic	Gender, Age, has_children, Education, Income, maritalStatus, Occupation
Behavioral	Bar, CoffeeHouse, CarryAway, RestaurantLessThan20, Restaurant20to50
Coupon Characteristics	Coupon, Expiration
Environmental Factors	Destination, Passenger, Time, Weather, Temperature
Proximity Information	toCouponGEF_5min, toCoupon_GEQ15min, toCoupon_GEQ25min
Navigational Features	direction_same,direction_opposite
Target Variable	Y

# Exploratory Data Analysis

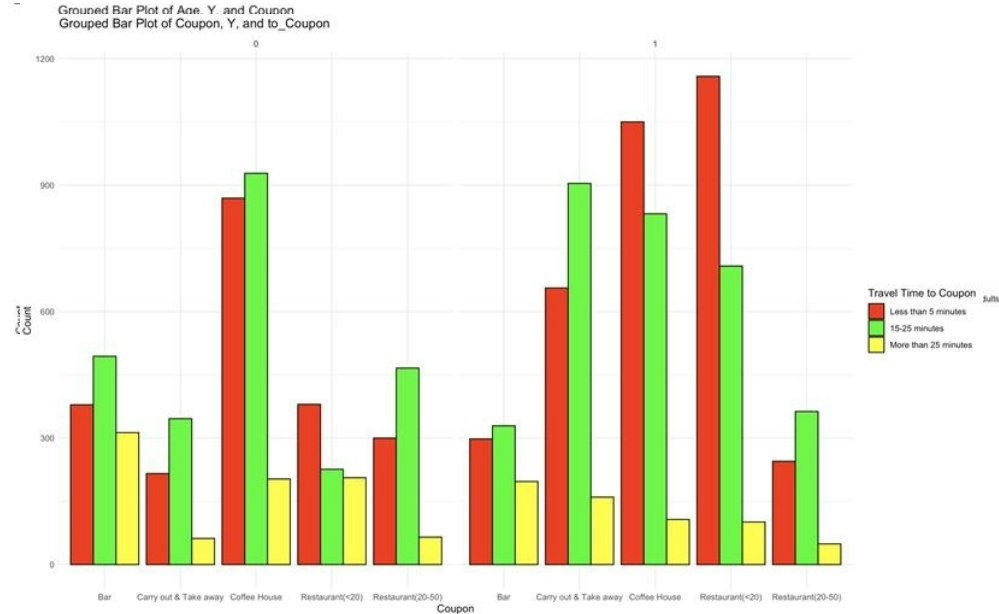


# Exploratory Data Analysis

- Coffee House coupons are the most frequently offered, with the highest acceptance rates for Carry Out & Take Away and Restaurant (<\$20) coupons, while Bar coupons have the lowest acceptance rate.
- Most users are young adults, highest acceptance rate is observed on both teenagers and young adults.
- Most common weather is “Sunny” and shows the highest coupon acceptance.
- Majority of users are from low and medium-income groups, which also have the highest coupon acceptance rates.
- Coupons are most accepted when users have destination 'No Urgent Place'.
- Target variable is balanced in nature.

# Exploratory Data Analysis

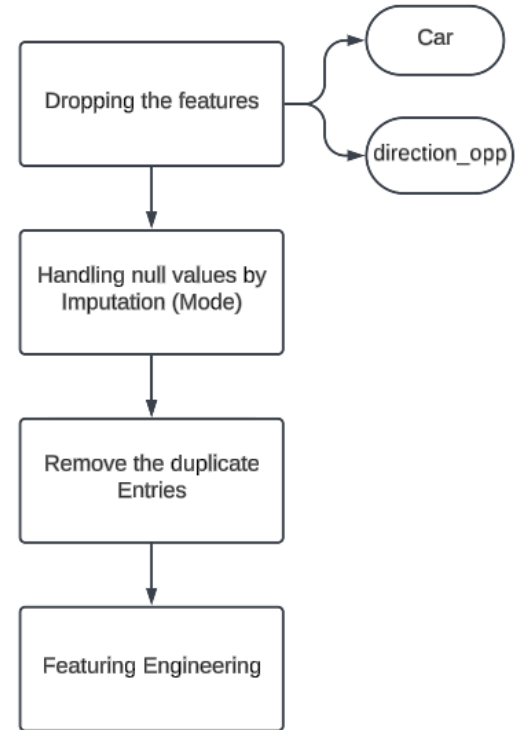
- ❑ Coupons are most accepted in sunny weather, with "Restaurant (<20)" and "Carry out & Take away" being accepted more frequently than rejected.
- ❑ Young adults have the highest coupon acceptance rates, while seniors and teenagers show significantly lower acceptance regardless of the coupon type.
- ❑ "Coffee House" and "Restaurant (<20)" coupons are most accepted with travel times under 5 minutes, while coupons requiring over 25 minutes of travel are rarely accepted.



# Data Preprocessing

- ❑ Missing values in feature car (99.14%).
- ❑ Redundancy in Data : direction\_opp
- ❑ Handling missing values by Mode Imputation.
- ❑ Feature Engineering.

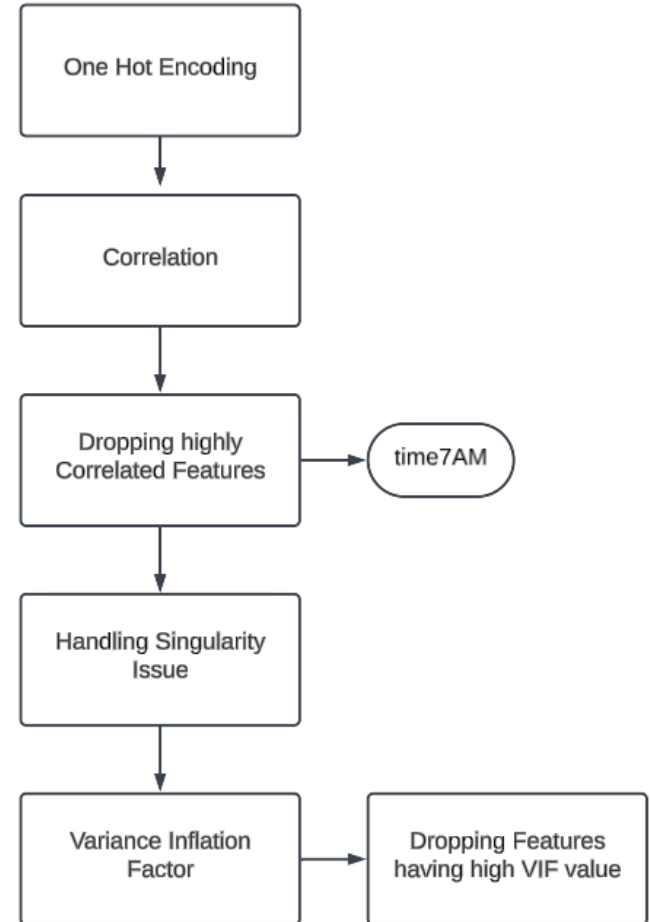
Features	Details
to_coupon_geq5min to_coupon_geq15min to_coupon_geq25min	to_coupon
age	Teenagers Young Adults Middle Aged Adult Seniors
Occupations	Into Different Categories
Income	Low_Income Medium_Income High_Income





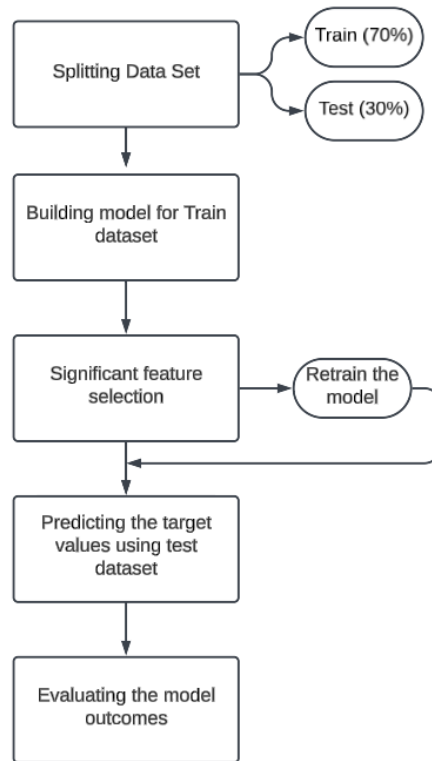
# Data Preprocessing

- ❑ Time7AM and destination\_passengerWork\_Alone is highly correlated.
- ❑ Observed singularity issue for destination\_passengerWork\_Alone.
- ❑ Feature Selection by VIF( $VIF < 0.5$ )



# Models and Evaluation

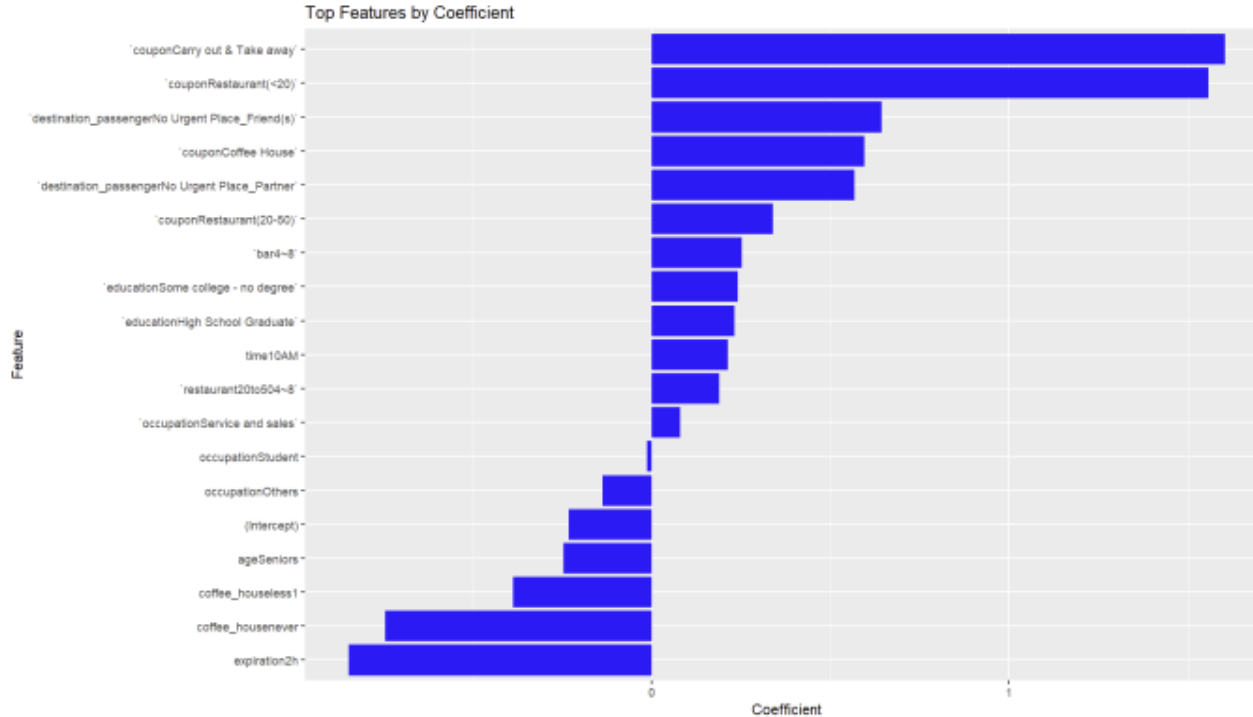
- ❑ Logistic regression demonstrates showing good generalization but low accuracy
- ❑ LDA has better sensitivity than Logistic regression, indicating possible overfit
- ❑ QDA shows a significant drop in overfitting, despite its slightly better



Models	Sample Accuracy
Regression	Training – 68.57% Test – 64%
LDA	Training – 69.29% Test – 62.33%
QDA	Training – 87.29% Test – 63.67%

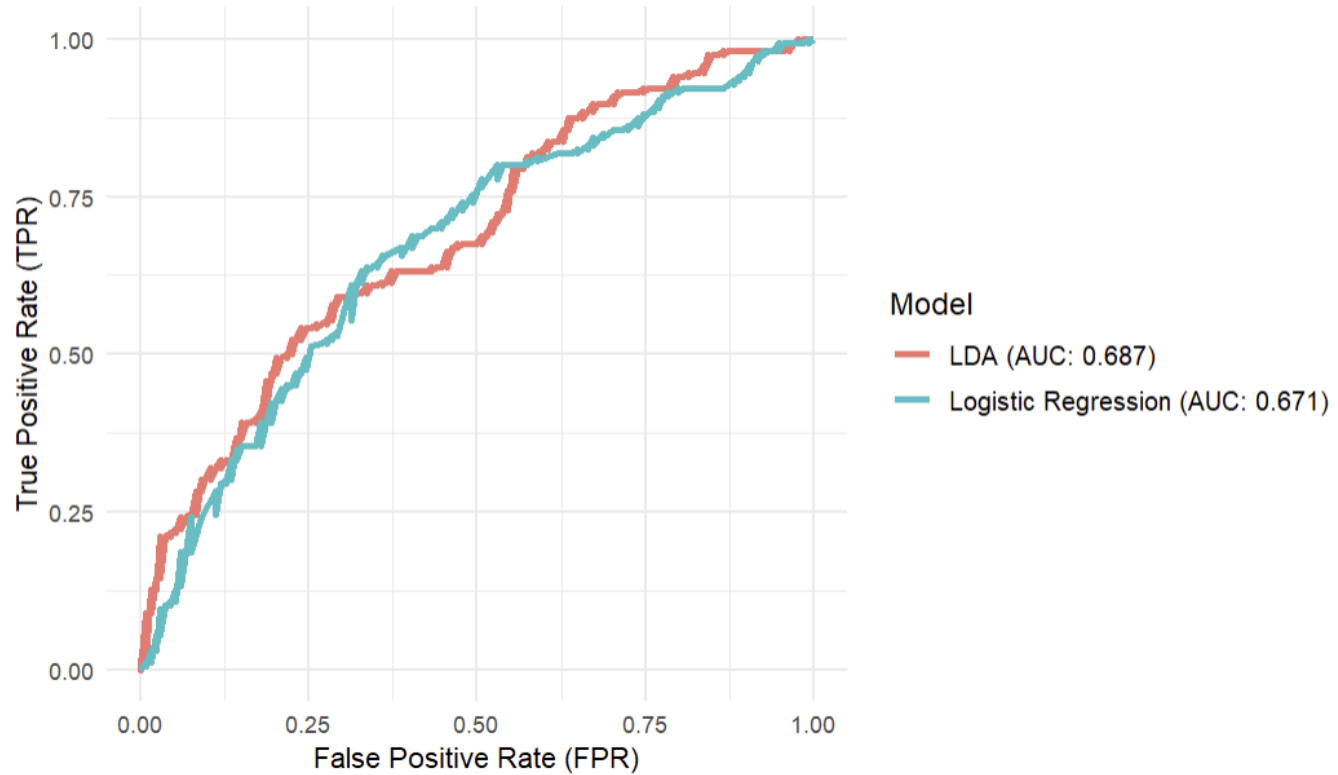
Specificity	Positive Prediction	Negative Prediction
73.49%	61.40%	65.59%
63.25%	57.34%	66.88%
63.25%	58.50%	68.63%

# Feature Importance



# ROC Curve

ROC Curves for Logistic Regression and LDA Models



# Summary and Conclusion

- ❑ Logistic Regression achieves the best balance with **68.57% training accuracy** and **64% test accuracy**, demonstrating strong generalization.
- ❑ LDA performs **better on sensitivity** (61.19%) but shows a drop in test accuracy (62.33%), suggesting potential overfitting.
- ❑ QDA has the highest training accuracy (87.29%) but struggles with test accuracy (63.67%) due to significant overfitting.
- ❑ **Logistic Regression** is preferred for its balanced and generalizable performance.

# Future Scope

- ❑ Explore Random Forest or XGBoost for improved accuracy.
- ❑ Extend to multi-class classification for coupon names.
- ❑ Enhance performance with advanced feature engineering/selection.
- ❑ Leverage larger datasets for better results.

Models	Dataset Accuracy	Sample Accuracy
Logistic Regression	Training – 68.49% Test – 67.47%	Training – 68.57% Test – 64%
LDA	Training – 69.09% Test – 67.31%	Training – 69.29% Test – 62.33%
QDA	Training – 73.18% Test – 70.07%	Training – 87.29% Test – 63.67%

# References

- Dataset: <https://archive.ics.uci.edu/static/public/603/in+vehicle+coupon+recommendation.zip>
- Research Paper: <https://jmlr.org/papers/volume18/16-003/16-003.pdf>
- <https://github.com/dikaaka/In-Vehicle-Coupon-Recommendation-Project/blob/main/STAGE%20/FINAL%20PROJECT%20-%20STAGE%20.pdf>
- <https://www.kaggle.com/code/maherabdellatif/invehicle-coupon-recommendation>

Thank



You