CUSTOMER SATISFACTION AT CARMAX

STAT 331-130: TEAM 4

MICHAEL CASTAGNA, KELSIE MILLER, ANH NGUYEN,
TUNG NGUYEN, & SHUYI WU

MOTIVATION

- A study from Cox Automotive Car Buyer Journey based on an online survey of 2,976 U.S. consumers finds that 74% of used vehicle buyers in 2021 stated they were highly satisfied with their overall purchasing experience with the dealership or retailer
- We are interested in learning about what factors influence customer satisfaction ratings and how it can be used to improve a company's performance

https://www.coxautoinc.com/news/2021-car-buyer-journey-study/#:~:text=In%20the%20study%2C%2078%25%20of,the%20vehicle%20selection%20in%202021.

WHAT IS NET PROMOTER SCORE (NPS)?

- NPS measures customer perception based on how likely it is that they would recommend a product or service to someone they know
- Why is NPS important?
  - A good indicator of future growth
  - Understand what is driving the customer experience or satisfaction
  - Help businesses to improve their products and services, gauge their performance, and channel their customer service efforts to drive growth

https://www.salesforce.com/uk/learning-centre/customer-service/calculate-net-promoter-score/#:~:text=Net%20promoter%20score%20helps%20businesses,revenue%20through%20referrals%20and%20upsells
https://www.netpromotersystem.com/about/

WHAT IS NET PROMOTER SCORE (NPS)? CONT.

- The rating ranges from 0 (not at all likely to recommend) to 10 (extremely likely to recommend)
  - Higher score → more satisfied, more likely to generate value for the business
  - Lower score → dissatisfied, potential to switch business and cause damage to reputation
- Companies that use NPS
  - BestBuy, PayPal, SAP, Verizon, KPMG, J&J, GE, Delta Airline, etc.

https://www.netpromotersystem.com/about/companies-that-use-net-promoter/

CARMAX

- Largest retailer of used cars in the U.S.
- Competitors: AutoNation, Carvana, Penske Automotive Group
- Primary Objective:
  - Utilize data analysis and statistical methods to optimize CarMax's operational efficiency

KEY STUDY QUESTIONS

- What are the main factors that drive an increase or decrease in NPS?
- Where should CarMax focus its efforts to enhance its current business model for customer relationship management?

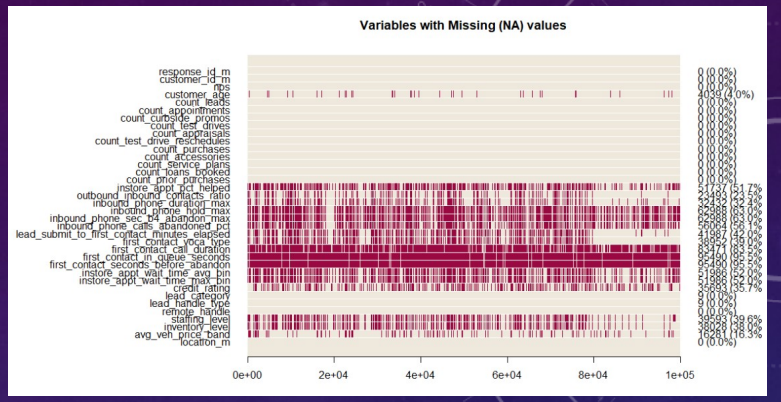
CARMAX DATA SET

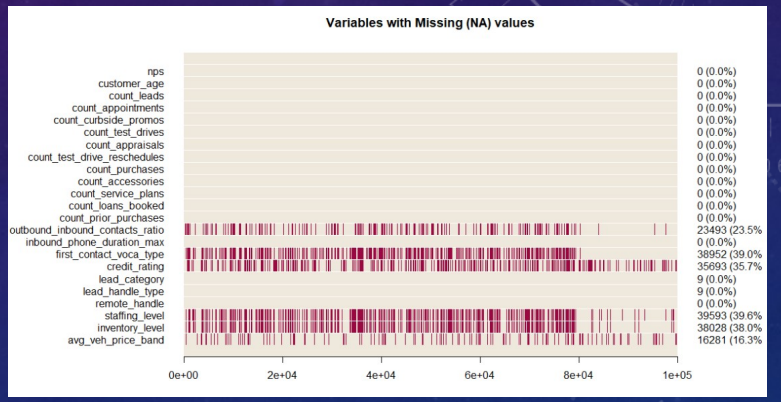
- Sample of 100,000 NPS survey results from customers who recently purchased a car from CarMax
  - Each row or record represents one survey response
  - Variables include: 36 variables

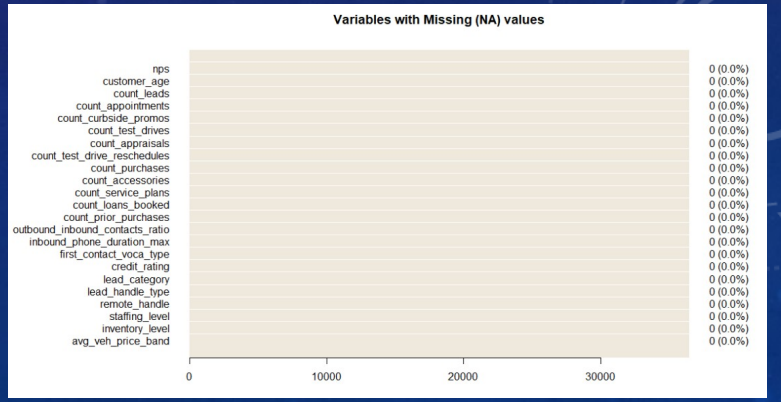
- response_id_m
- customer_id_m
- nps
- customer_age
- count_leads
- count_appointments
- count_curbside_promos
- count_test_drives
- count_appraisals
- count_test_drive_reschedules
- count_purchases
- count_accessories
- count_service_plans
- count_loans_booked
- count_prior_purchases
- instore_appt_pct_helped
- outbound_inbound_contacts_ratio
- inbound_phone_duration_max
- inbound_phone_hold_max
- inbound_phone_sec_b4_abandon_max
- inbound_phone_calls_abandoned_pct
- lead_submit_to_first_contact_minutes_elapsed
- first_contact_voca_type
- first_contact_call_duration
- first_contact_in_queue_seconds
- first_contact_seconds_before_abandon
- instore_appt_wait_time_avg_bin
- instore_appt_wait_time_max_bin
- credit_rating
- lead_category
- lead_handle_type
- remote_handle
- staffing_level
- inventory_level
- avg_veh_price_band
- location_m

CLEANSING TASKS

- Use “lapply” function to convert categorical data to factors
- Use “plotMiss” function to visualize missing values in the data set
- Use the “select” function to select only those columns in the data set which have less than 40% missing values
- Use the “preProcess” function to impute missing values for numerical variables with the median
- Use the “na.omit” function to remove any remaining missing values
- Use the “duplicated” function to check for and remove any duplicate values

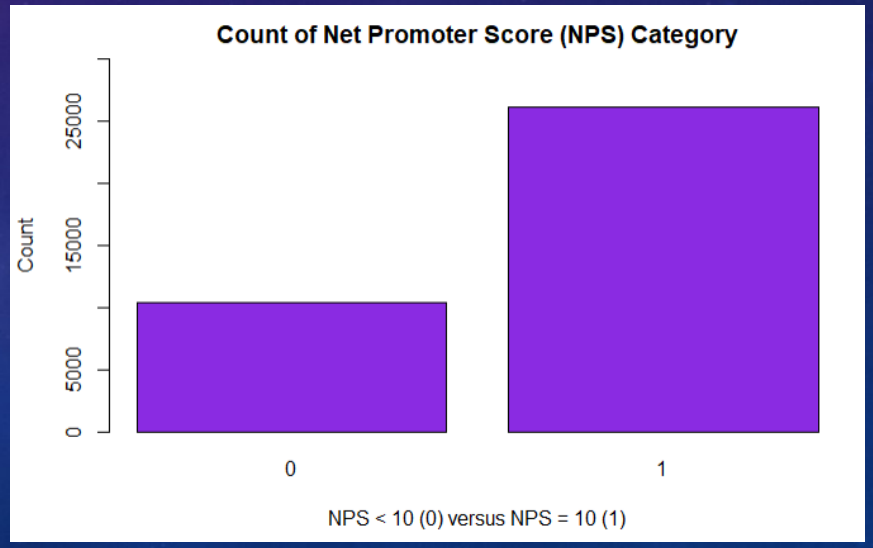






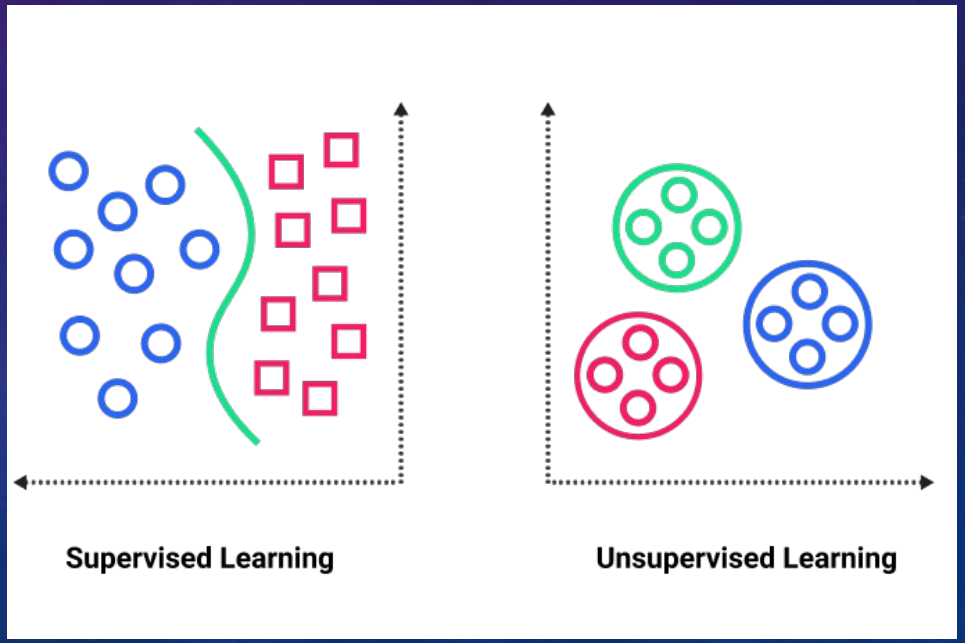
OUTCOME VARIABLE: NPS CATEGORY

- CarMax maintains an average NPS of 9.15 (considered relatively high)
- Use the “mutate” function to create a new binary dependent variable “nps_cat”
  - 1 – indicates that CarMax has earned NPS of 10
  - 0 – indicates that CarMax has earned NPS < 10
- Dimensions of data pre- and post-cleansing:
  - 100,000 observations → 36,446 observations
  - 36 variables → 24 variables



EXPLORATION TASKS

- Supervised Learning Algorithms:
  - Logistic Regression
- Unsupervised Learning Algorithms:
  - Hierarchical Clustering Analysis



MODEL DEVELOPMENT: LOGISTIC REGRESSION

- Model the relationship between count of appointments, curbside promos, test drives, prior purchases, and NPS category, where appointments, curbside promos, test drives, and prior purchases are the independent variables to predict the probability of NPS category as the dependent variable
- LR Equation:
  - log = 1.10982 - 0.13040 *(appointments) - 0.42429 *(curbside promos) - 0.04062*(test drives) + 0.04717*(prior purchases)

```
Call:
glm(formula = nps_cat ~ count_appointments + count_curbside_promos +
    count_test_drives + count_prior_purchases, family = binomial(link = "logit"),
    data = train_carmax)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8807  -1.4976   0.7985   0.8125   1.5309

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              1.10982    0.02894  38.355  < 2e-16 ***
count_appointments      -0.13040    0.01645  -7.929 2.21e-15 ***
count_curbside_promos   -0.42429    0.05879  -7.217 5.33e-13 ***
count_test_drives       -0.04062    0.01896  -2.142  0.03219 *
count_prior_purchases    0.04717    0.01439   3.278  0.00104 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

MODEL VALIDATION: LOGISTIC REGRESSION

- Evaluate how well the LR model developed using train_CarMax will perform in an independent data set in terms of predicting the probability of having NPS of 10, given we know the count of appointments, curbside promos, test drives, and prior purchases
  - Create predicted values of “nps_cat” (binary dependent variable) in testing data: test_CarMax
    - pr_lr_Testing <- predict(lr1, newdata= test_CarMax, type= "response")
  - Using “ifelse” function, convert the predicted values of the dependent variable to be recoded as "1" if probability > 0.5, or as "0" otherwise
    - pr_lr_coded <- ifelse(pr_lr_Testing > 0.5, 1, 0)
  - Create the Confusion Matrix
    - table(test_CarMax$nps_cat, pr_lr_coded)

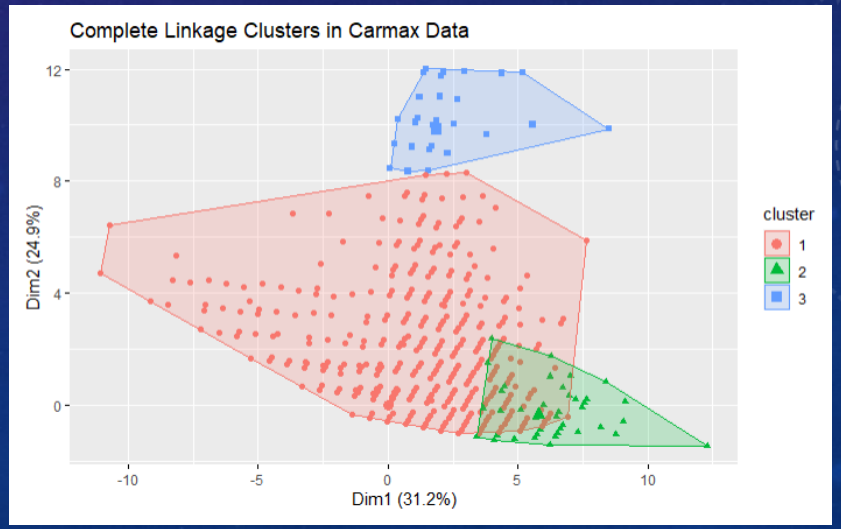
CONFUSION MATRIX INTERPRETATION

- Precision → Given the model predicted CarMax receives NPS of 10, how many times did CarMax actually receive NPS of 10?
  - 10440/(4099+10440)*100 = 71.81%
- Recall → Given CarMax receives NPS of 10, how many times did the model predict it correctly?
  - 10440/(19+10440)*100 = 99.81%
- Note: There is a trade-off in the metrics we choose to maximize
  - As we increase precision, we decrease recall and vice-versa

		LR model predicted NPS of 10	
		No (0)	Yes (1)
Actual NPS is 10	No (0)	21	4099
	Yes (1)	19	10440

HIERARCHICAL CLUSTERING ANALYSIS

- Goal:
  - Cluster customers based on their count of appointments, promotional offers redeemed, test drives, and prior purchases (numerical)
  - Complete Linkage: creates clusters using the maximum distance between an item and a cluster
- Result:
  - Model identified 3 clusters (red, green, blue – see below) using these explanatory variables → each dot is an observation (customer)



CLUSTER MEANS INTERPRETATION

- CarMax should focus its efforts on Groups 2 & 3
- Group 2
  - Develop an outreach/promotional program to encourage customers to not only make appointments and schedule test drives but also purchase a vehicle
- Group 3
  - Because this group of customers is already purchasing the most, the marketing strategy should focus on retention and continuous satisfaction to maintain this level of purchase behavior

Group.1 <int>	count_appointments <dbl>	count_curbside_promos <dbl>	count_test_drives <dbl>	count_prior_purchases <dbl>
1	1.321604	0.03323338	0.6007318	0.4863408
2	7.057692	0.00000000	2.8653846	0.7692308
3	1.377778	0.00000000	0.7555556	11.7333333

SUGGESTED TARGET SEGMENTS

- Millennials and Gen X
- 37% and 33% of the sample population respectively
- Both generations made prior purchases with CarMax

Why?

- The largest population in the US
- Bought more new cars than any other age group, entirely online
- Have the potential to become consistent CarMax customers

INSIGHTS & RECOMMENDATIONS

Build brand loyalty through first-time buyer and promotional programs for customers who have poor and fair credit score and book a loan

Deliver a positive customer experience at each stage of the customer journey (leads and appointments)

Suggest the appropriate vehicle price range for a particular generation

Invest in informative and interactive contents across a variety of social media and mobile devices

FUTURE ANALYSIS NEEDED

- Leverage multiple data sources to retrieve additional data contexts
- Research other customer loyalty KPIs (i.e., customer retention rate, churn rate, CLV, etc.)
- Further trend modeling on the drivers of NPS score over time

THANK YOU!

Q&A

WORKS CITED

- Dessel, Van Gert. "Net Promoter Score (NPS) – use, application and pitfalls." *CheckMarket*, June 2011, https://www.checkmarket.com/blog/net-promoter-score/
- Joyce, Gemma. "The Pros and Cons of NPS." *Brandwatch,* June 14th 2019, https://www.brandwatch.com/blog/nps-pros-cons/
- Roldós, Inés. "NPS Analysis: 6 Ways to Analyze Your NPS". *MonkeyLearn,* March 4th 2020*,* https://monkeylearn.com/blog/nps-analysis/