DIRECT MARKETING

BELLEVUE UNIVERSITY

FALL 2020

TORREY CAPOBIANCO

CONRAD IBANEZ

EDRIS SAFARI

BACKGROUND

- Direct marketing is a business sector that can leverage predictive analytics to determine the type of customer to send their advertisements to
- Direct marketing is costly
- Past data on customers can be used to see who will act on an advertisement
- Predicting the type of customer can shift the company’s marketing resources to those that will spend vs. those that will not spend money
- Prevent wasted marketing resources on customers who will not act on the mailer

PROBLEM STATEMENT

➢Can we predict how much a customer will spend from direct mailing?

- The company uses direct mailing catalogs to send to customers
- From data collected from customers, can this information be used to determine the type of spender a new customer will be?
- The data insights will be used to expand their market regions to those that are predicted to spend more

DATA

- The data was retrieved from Kaggle
- 1,000 samples
- 10 total variables
- Characteristics of customers & households

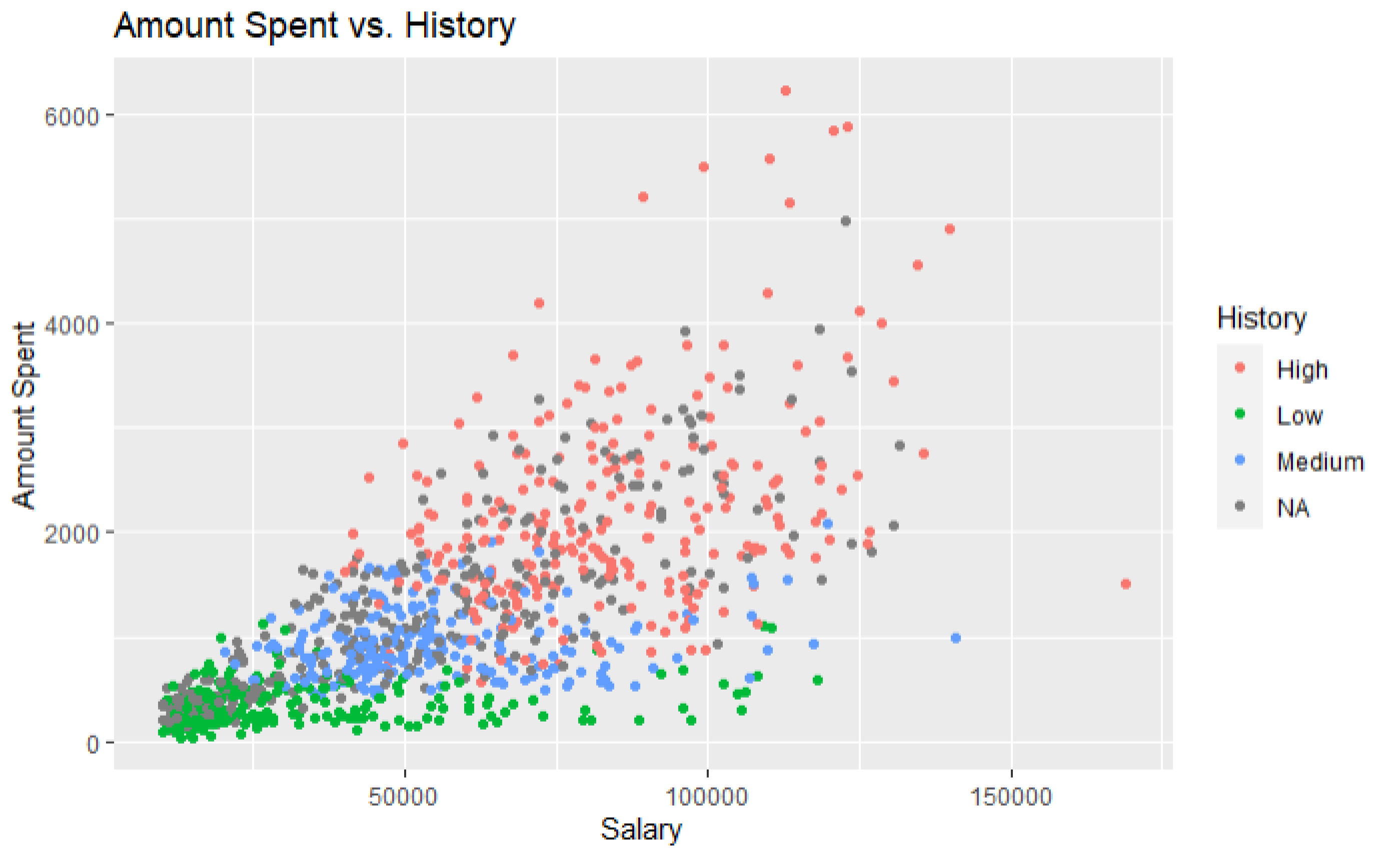
Variables

- Age
- Gender
- OwnHome
- Married
- Location
- Salary
- Children
- History
- Catalogs
- Amount Spent

DATA PREPARATION

Missing Values

- History had 303 missing values, over ¼ of the data
- Could not derive the NA from other data
- Dropped History from data frame



Normalize

- For one model, continuous variables were normalized to be on the same scale

DATA PREPARATION

Feature Creation

- Classification Spending Label
  - Categorize into low, medium, high spenders
  - Based on the quantiles of amount spent

Class	Range	Amount
Low	Below Median Quantile	≤ $962.00
Medium	Between Median and 3rd	
High	Above 3rd Quantile	> $1688.50

Dummy Variables

- Explode categorical variables into dummy variables
- Dropped 1 from each to avoid correlation

Age	Gender	OwnHome
Old	Male	Own
Young	Female	Rent
Middle	Female	Own

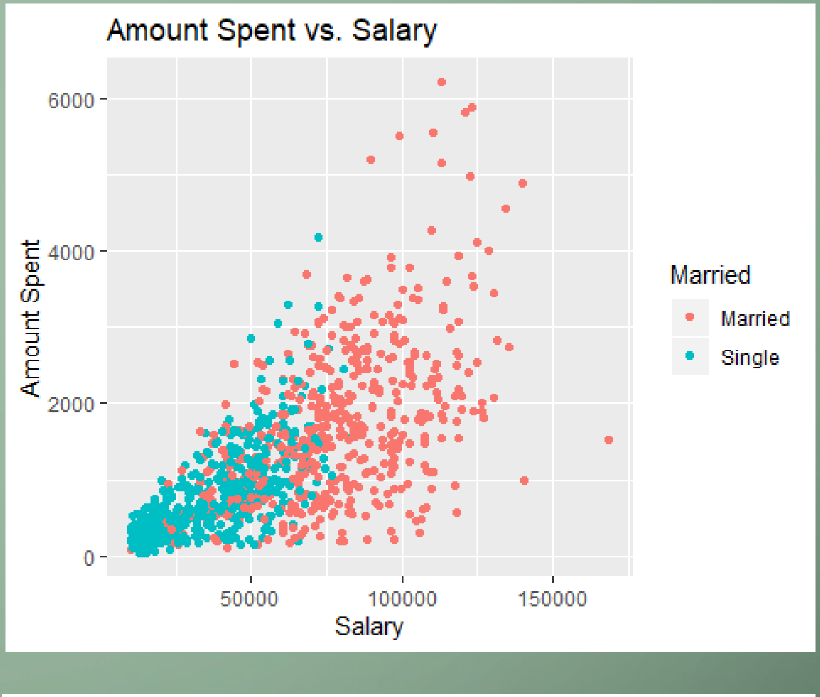
↓

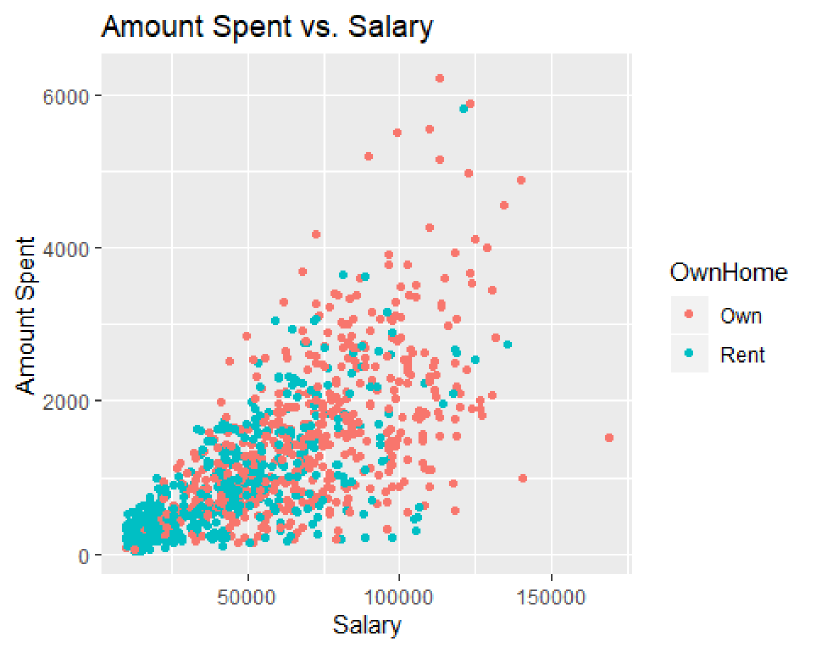
Age_Old	Age_Young	Gender_Male	OwnHome_ Own
1	0	1	1
0	1	0	0
0	0	0	1

EXPLORATORY DATA ANALYSIS

Salary has a positive correlation with amount spent

- Married households tend to make more, thus spending more
- Those that have a higher salary also tend to own a home vs. renting one. They also tend to spend more





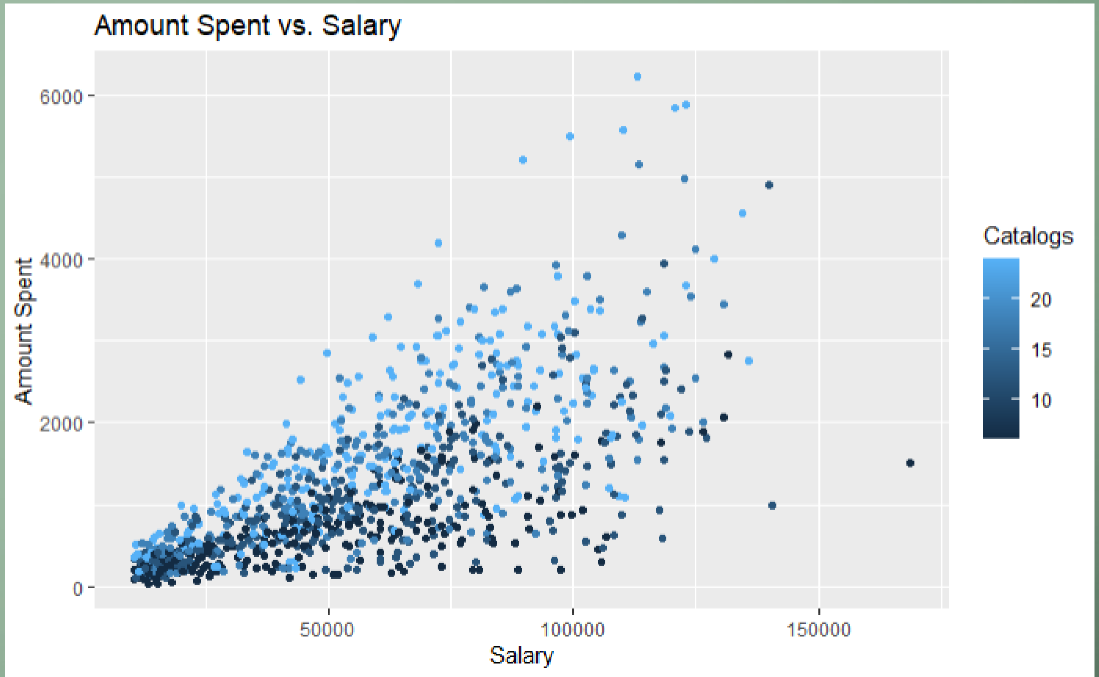
EXPLORATORY DATA ANALYSIS

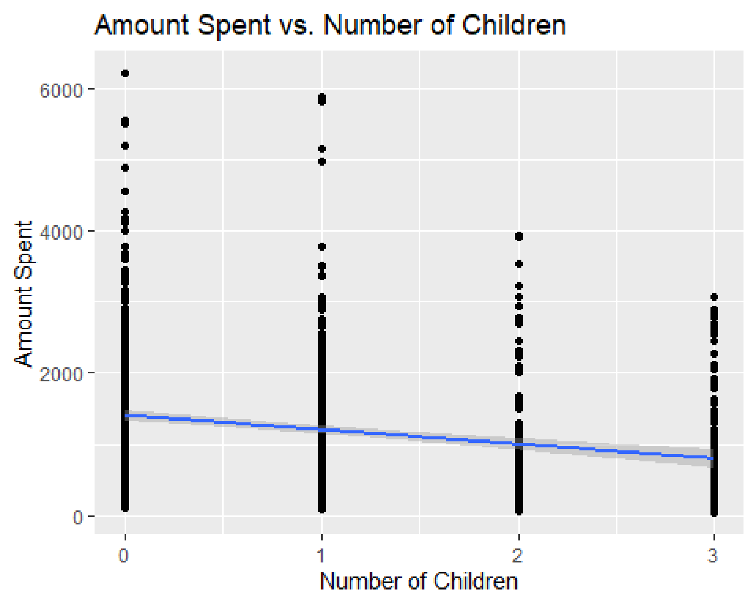
Catalogs

- The number of catalogs received seems to not determine how much they spend based on their salary
- Unsure of what effect catalogs would have for a new customer
  - Tested removing catalogs and did not change the model outcome

Number of Children

- As children increase, spending starts to decrease



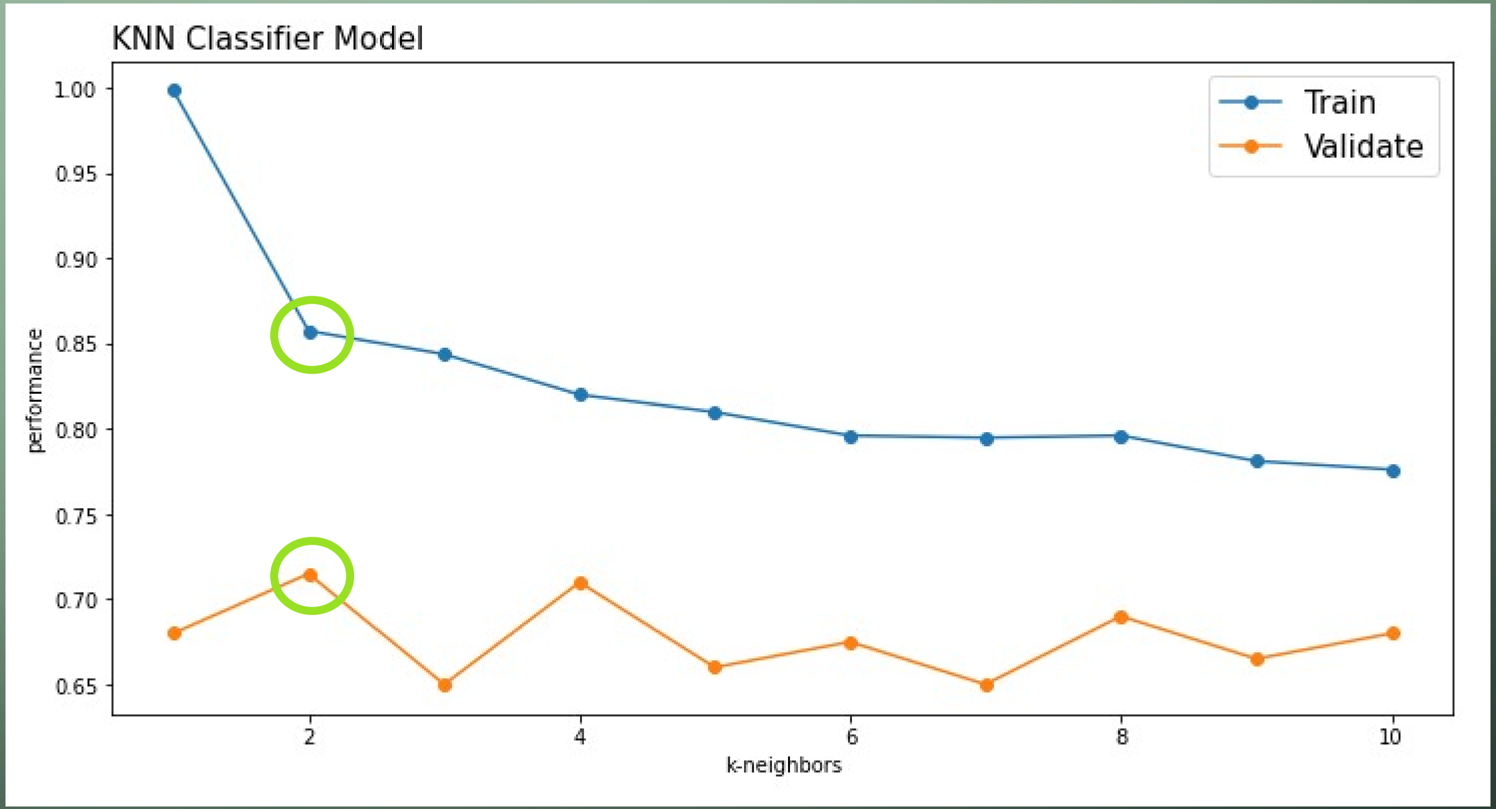


MODELING

- We present four different predictive models
  - K-Nearest Neighbors
  - Decision Tree
  - Linear Regression
  - Logistic Regression
- Training & Testing Sets: 80/20 split

K-NEAREST NEIGHBORS

- Classification Model
- Predicts type of spender by its neighbor's class
- 2 Nearest Neighbors for optimal prediction

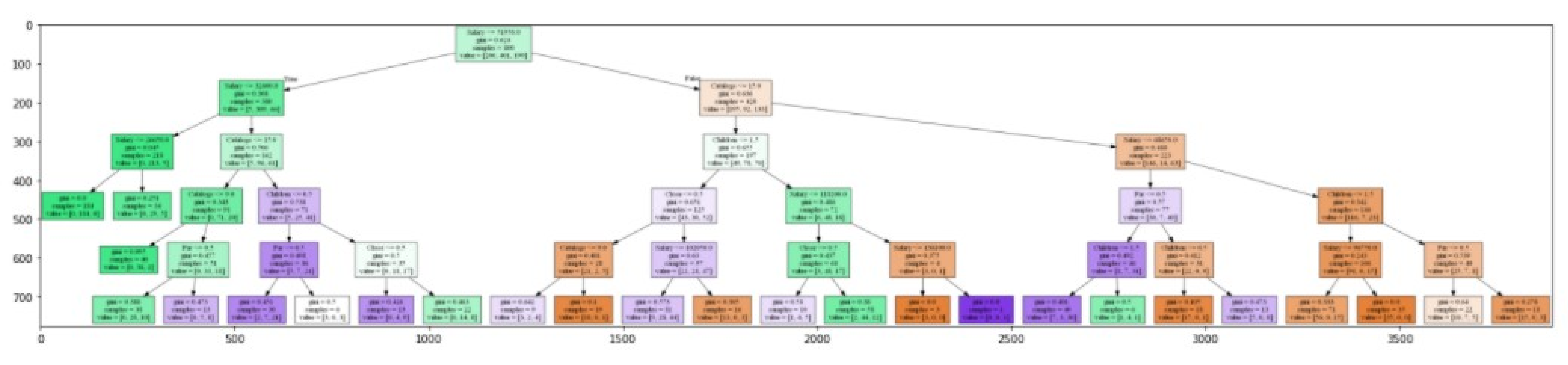


Train: 85.8%

Test: 71.5%

DECISION TREE

- Classification Model
- Predicts type of spender by splitting on decision nodes
- At depth 5, test and train have close accuracy

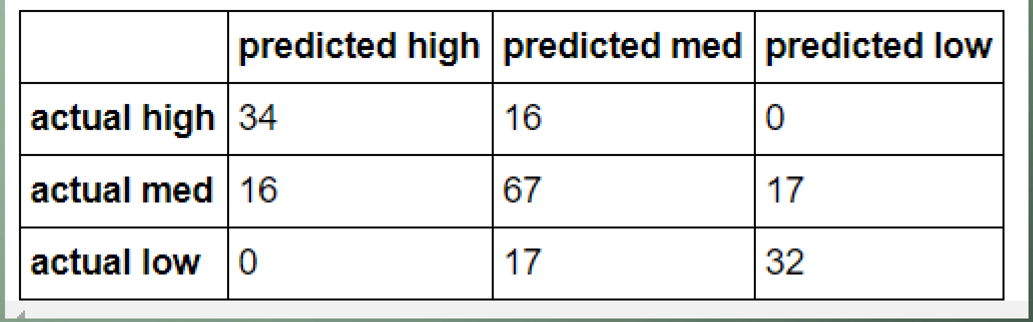


Train: 80.1%

Test: 81%

LINEAR REGRESSION

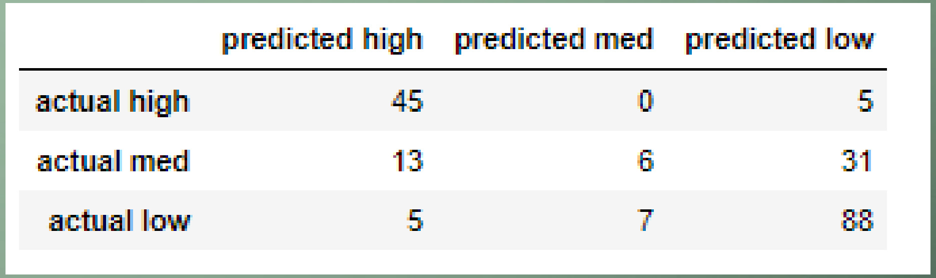
- Predicts actual dollar spent
- PCA removed all but one feature
- Train accuracy of 47%
- Test accuracy of 66.5 %
- Confusion Matrix after classifying predicted values to low, med, high



	predicted high	predicted med	predicted low
actual high	34	16	0
actual med	16	67	17
actual low	0	17	32

LOGISTIC REGRESSION

- Predicts the probability of a categorical dependent variable
- Train accuracy of 69.5%
- Test accuracy of 67.3 %



	predicted high	predicted med	predicted low
actual high	45	0	5
actual med	13	6	31
actual low	5	7	88

ACCURACY

Accuracy Score:
Proportion of predictions that the model classified correctly

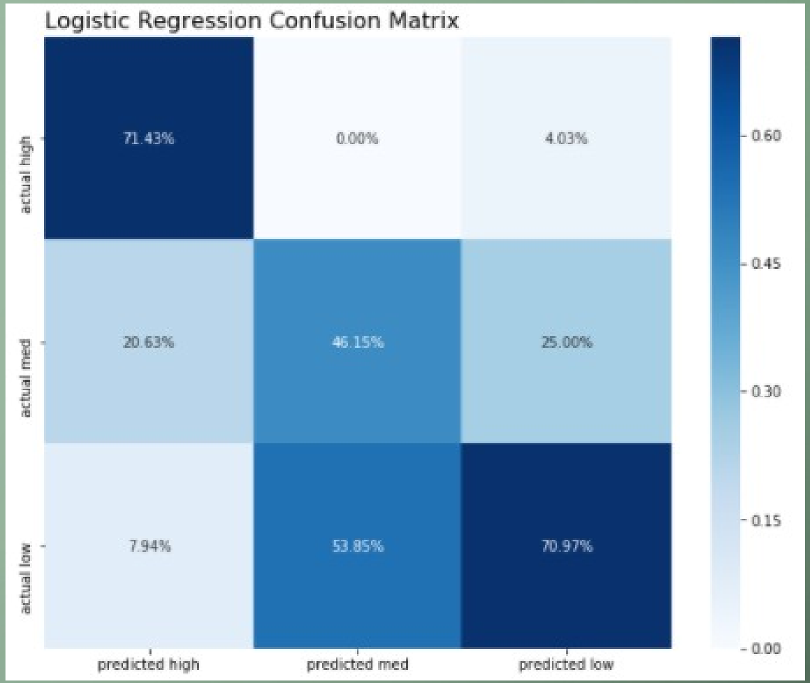
Highest Accuracy: Decision Tre

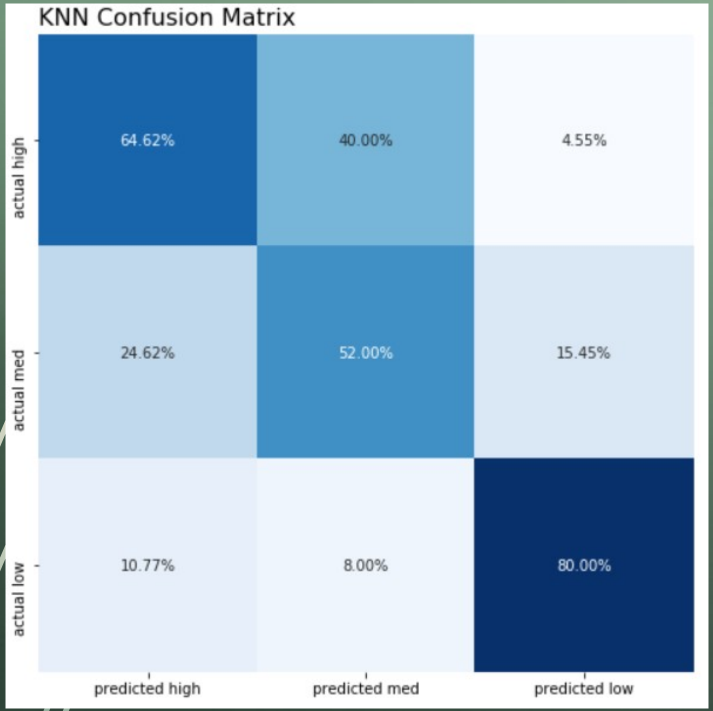
K-Nearest Neighbors	Decision Tree	Linear Regression	Logistic Regression
71.5%	81%	66.5%	67.3%

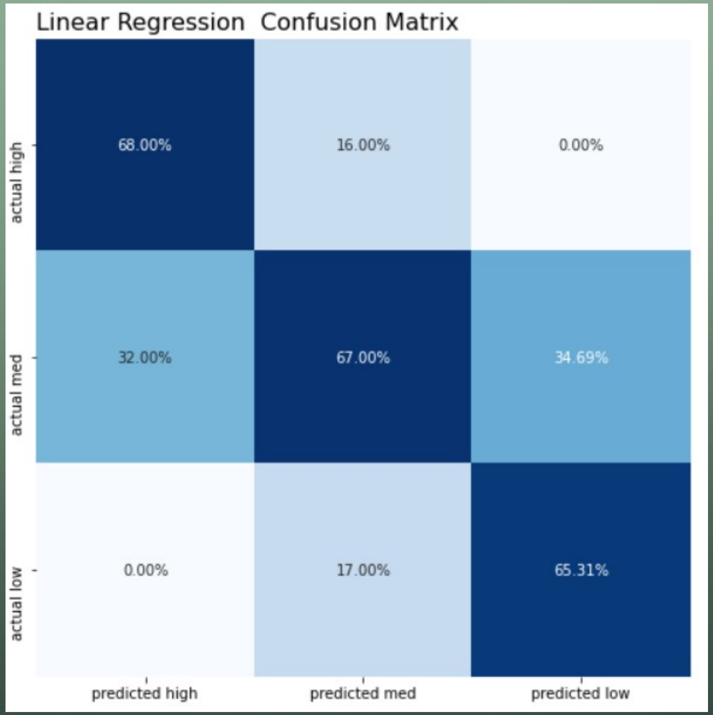
CONFUSION MATRIX

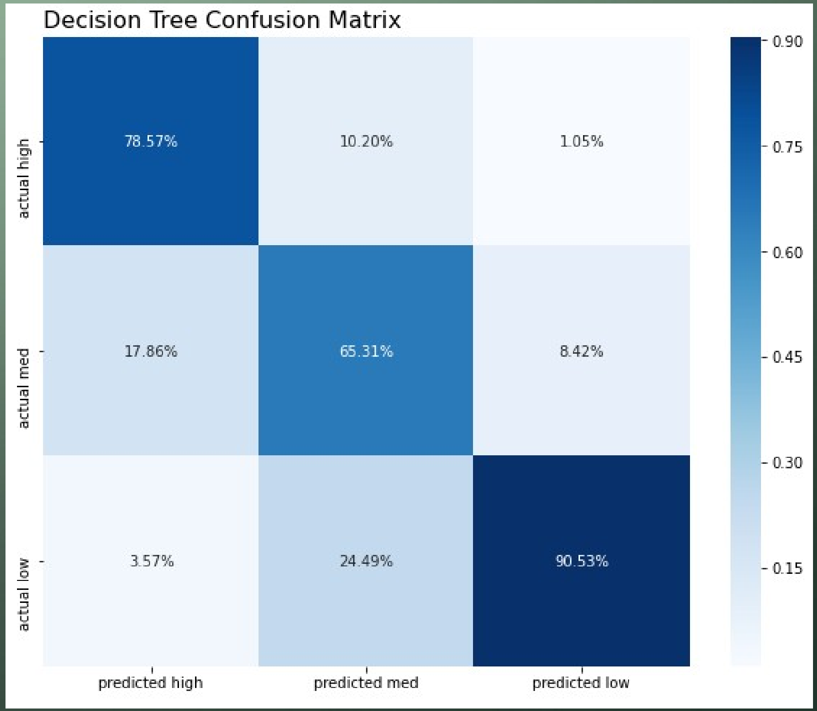
Decision tree has the best classification scores for high and low spenders

Additional analysis to gain additional insights can be done for these results


Logistic Regression Confusion Matrix


KNN Confusion Matrix


Linear Regression Confusion Matrix


Decision Tree Confusion Matrix

CLOSING

- Opportunities for improvement:
  - To achieve higher accuracy we recommend the company collect additional customer data
  - Collect data on those that did not make a purchase when they received a catalog