PREDICTIVE ACCURACY OF DEFAULT OF CREDIT CARD CLIENTS

Collin Guidry

EXECUTIVE SUMMARY: PREDICT DEFAULT OF PAYMENTS

Payment Default Data	• Dataset of interest: credit card payment defaulters
Models Tested	• KNN • K-means Clustering • Neural Network
Best Model	• Neural Network • Has highest accuracy

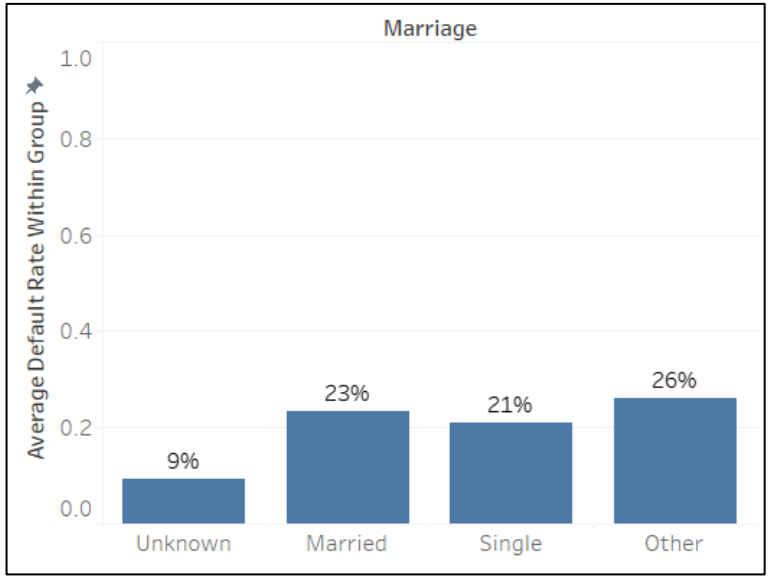
DATA EXPLORATION

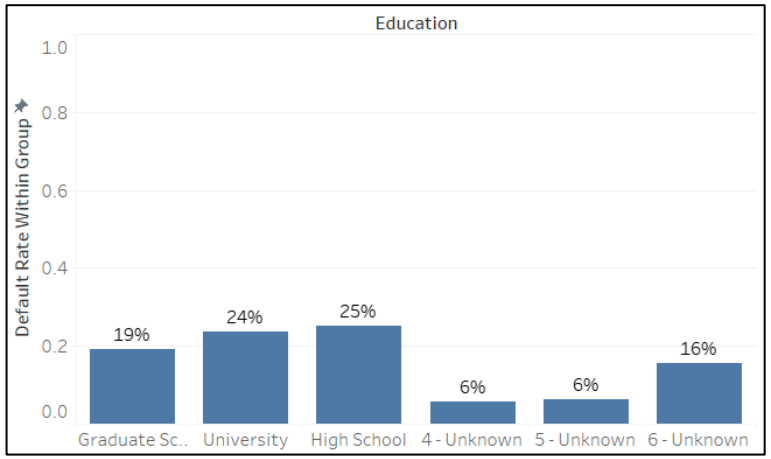
- This data includes a series of historical credit card payments for bank customers and whether each customer defaulted on his/her most recent payment.
- The amounts due and amount paid for previous months is shown, as well as demographic information such as age, gender, marital status, and education

Average Default Rate by Demographics
Marital Status
- No particular marital status type indicates a higher rate of payment default.

Education Level
- Higher education individuals are generally less likely to default on payments

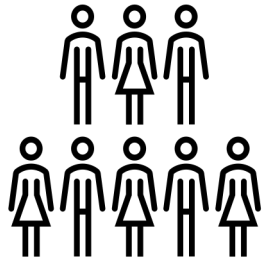
There are two unknown education types which have very low default rates





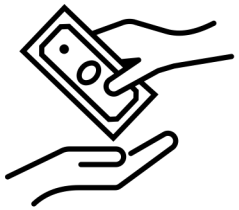
SAMPLE DATA BREAKDOWN

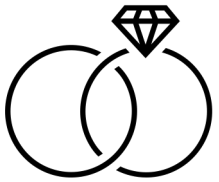
30,000 CUSTOMERS



MOSTLY FEMALE

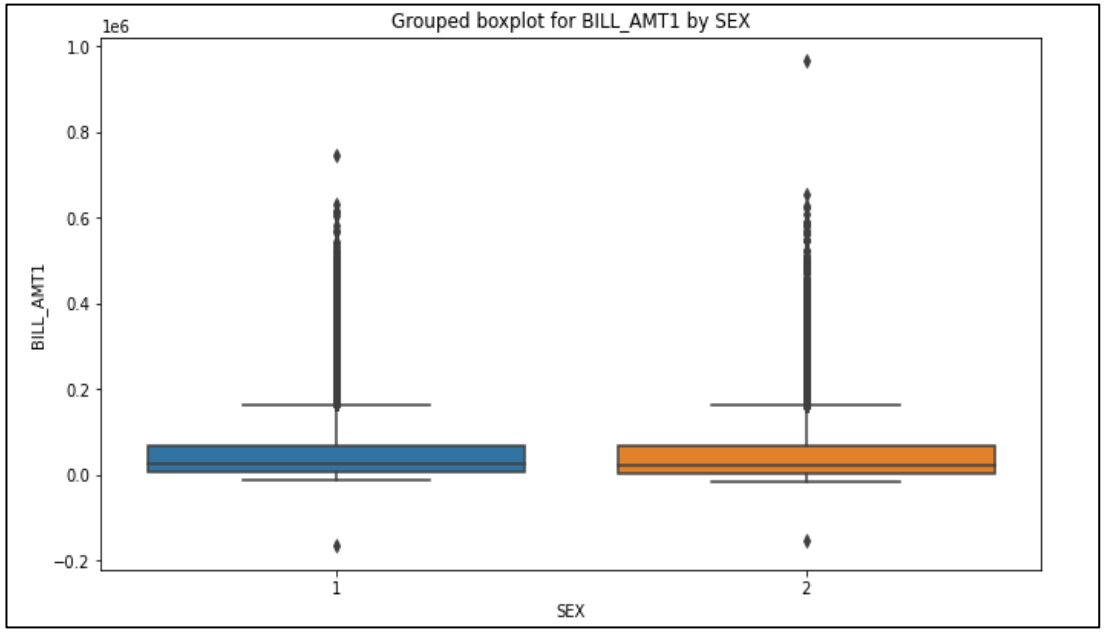
MOSTLY FEMALE DEFAULTERS

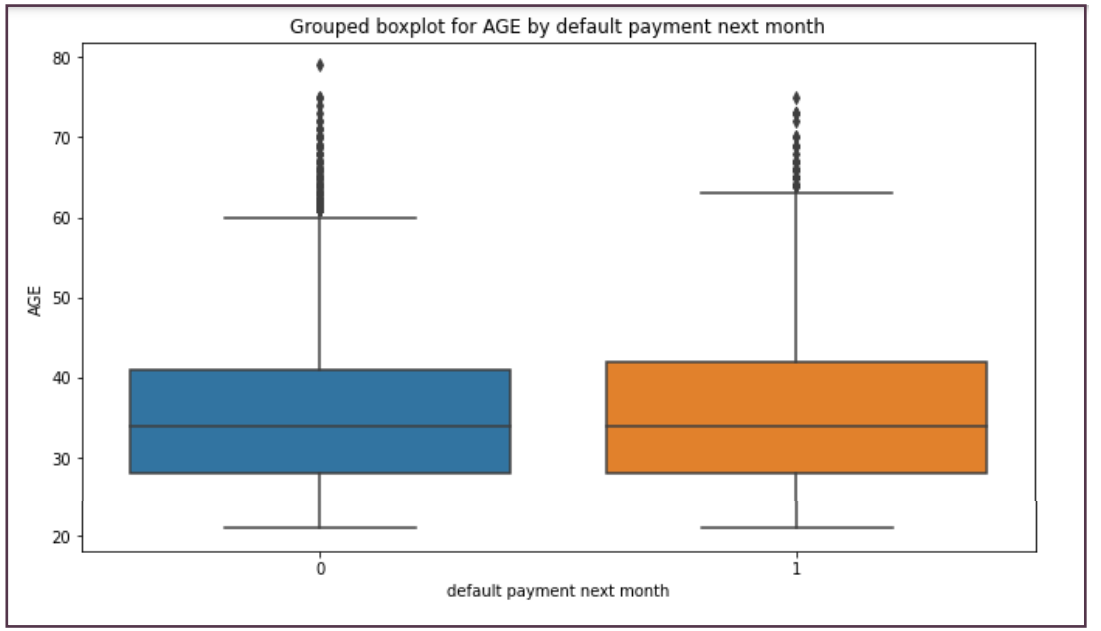




4 DISTINCT MARRIAGE VALUES

DATA DISTRIBUTIONS





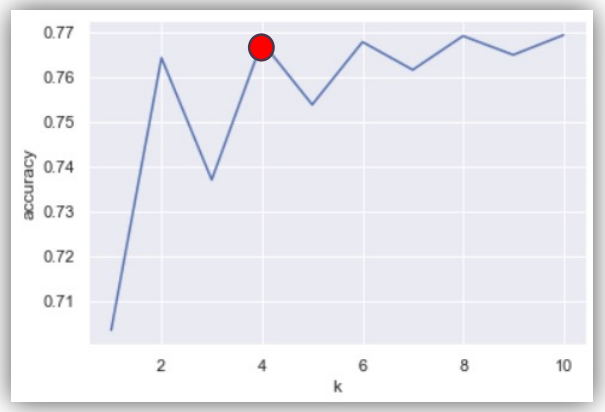
- Distribution of the amount due per month is relatively equal among males and females

- The bulk of customers are between approximately 30-40 years of age regardless of default status

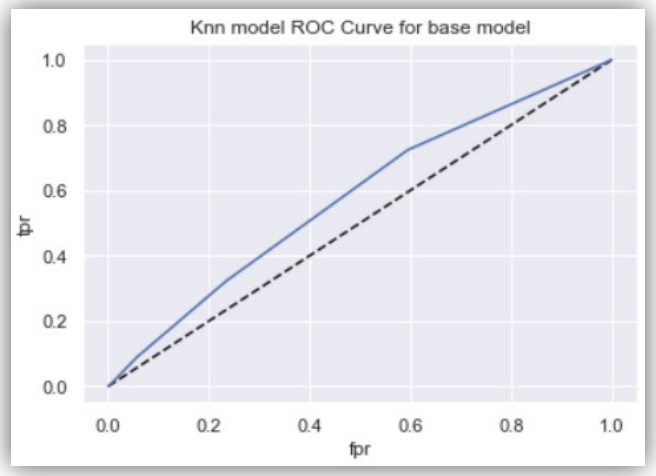
KNN MODEL

Value of k chosen: 4

(Accuracy does not significantly increase after 4 *see Appendix A for additional information*)



ACCURACY



ROC

Measure	Value
Accuracy	77%
Misclassification	23%
True Positive	11%
False Positive	4%
Specificity	96%
Precision	44%
Prevalence	23%

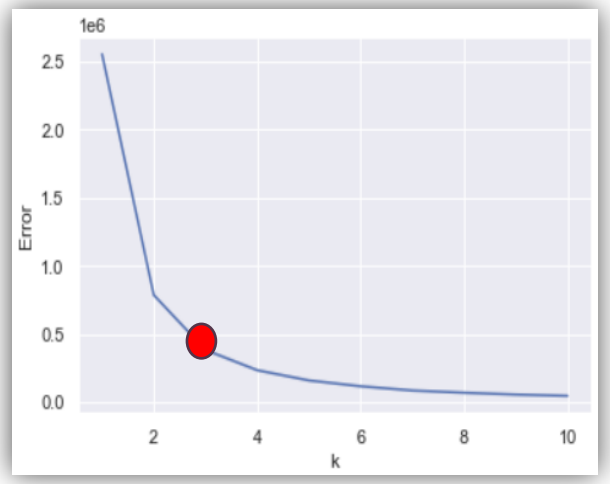
K-MEANS CLUSTERING KNN MODEL

K-means Clustering

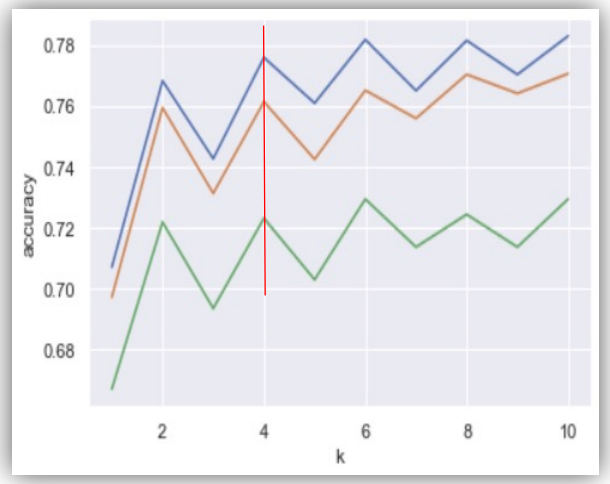
- Variables used: Age and Default Payment
- Number of segments chosen: 3
- The last significant decrease in error occurs at k = 3

KNN Model

- Value of k chosen: 4
- Low value of k chosen without sacrificing accuracy
- Weighted Average Accuracy of all models: **76%**



SEGMENT ACCURACY



SEGMENT KNN

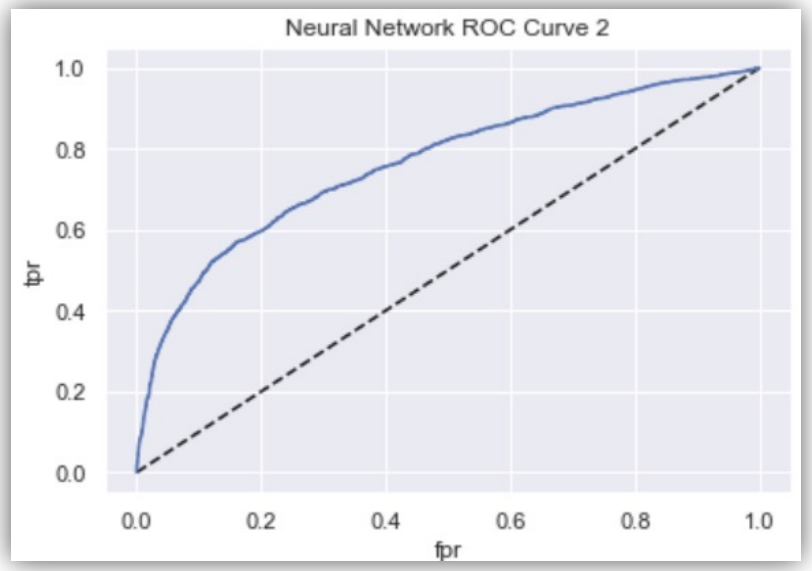
Segment Age	Accuracy	Misclassification	True Positive	False Positive	Specificity	Precision	Prevalence
21-32	76%	24%	9%	4%	96%	37%	23%
33-44	78%	22%	11%	4%	96%	40%	21%
45-79	72%	28%	9%	6%	94%	35%	26%

See Appendix A for all segments' ROC curves

NEURAL NETWORK MODEL

- Layer 1: 12 Nodes
- Layer 2: 8 Nodes
- Layer 3: 1 Node

Measure	Value
Accuracy	81%
Misclassification	19%
True Positive	35%
False Positive	5%
Specificity	95%
Precision	65%
Prevalence	23%



ROC

MODEL COMPARISON

- KNN model
  - Accuracy: 76%
- KNN models of separate segments
  - Accuracy: 76%
  - Using a KNN model for each age group does not outperform a single model
- Neural Network
  - Accuracy: 81%
  - Due to the pattern-based nature of the payment data, the neural network's ability to "memorize" allows it the highest level of accuracy.
- Cross validation used for all models such that over-fitting is not a concern

CONCLUSION

BEST MODEL: NEURAL NETWORK

- Highest accuracy rate
- Highest true positive
- Highest area under ROC curve
- Recognizes patterns

THANK YOU

APPENDIX A

- ROC for all 3 segments' KNN models