

SETTING THE BOARD

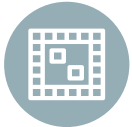


BACKGROUND



Catan

Build settlements and trade resources to win



The Dataset

Statistics from 50 online games

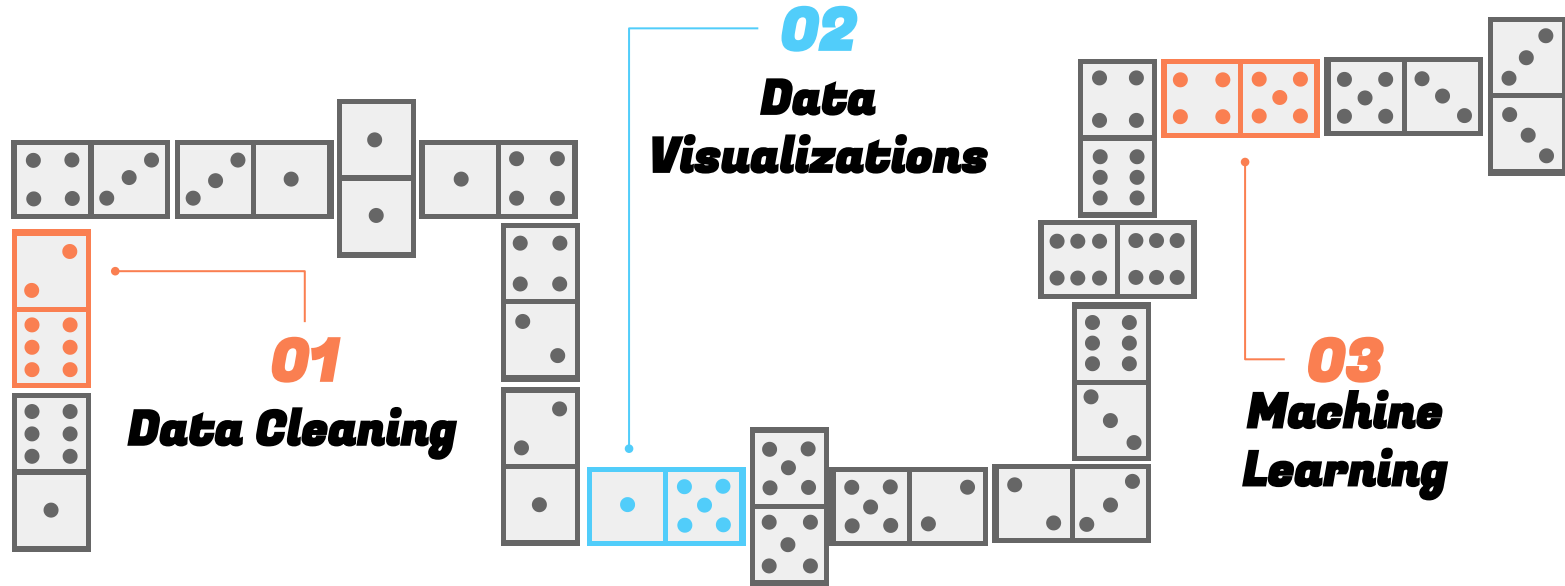


Our Goal

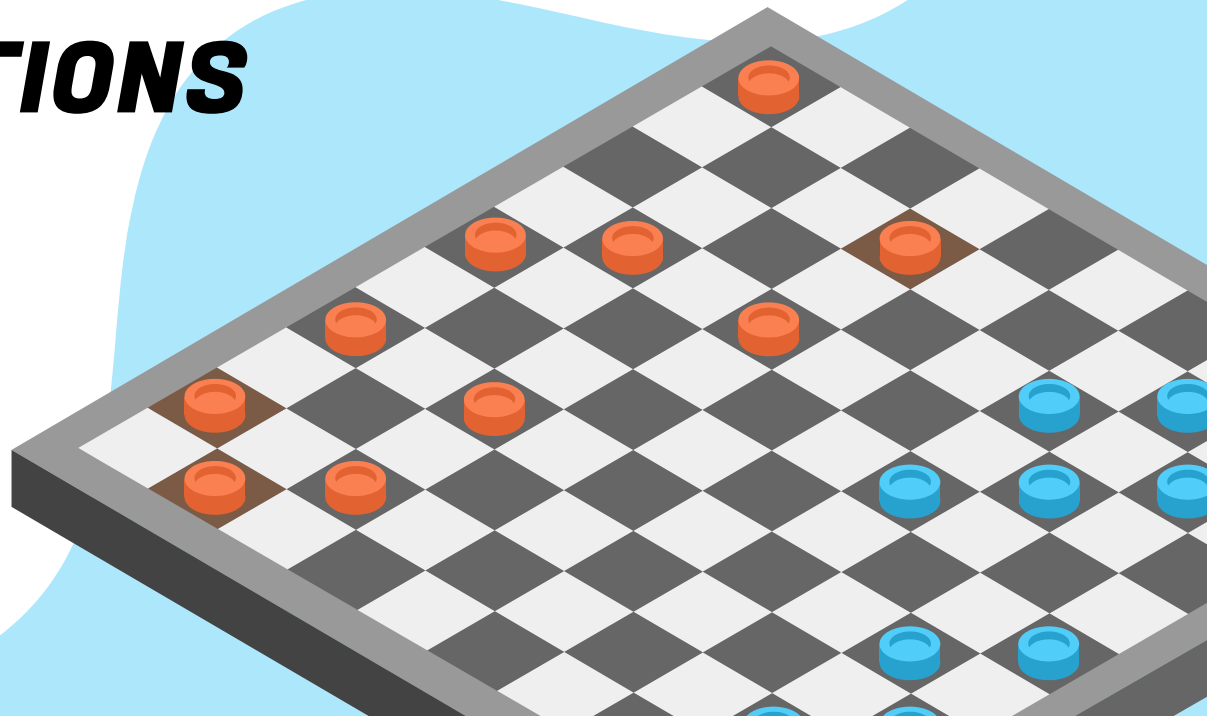
Determine what factors influence player outcome

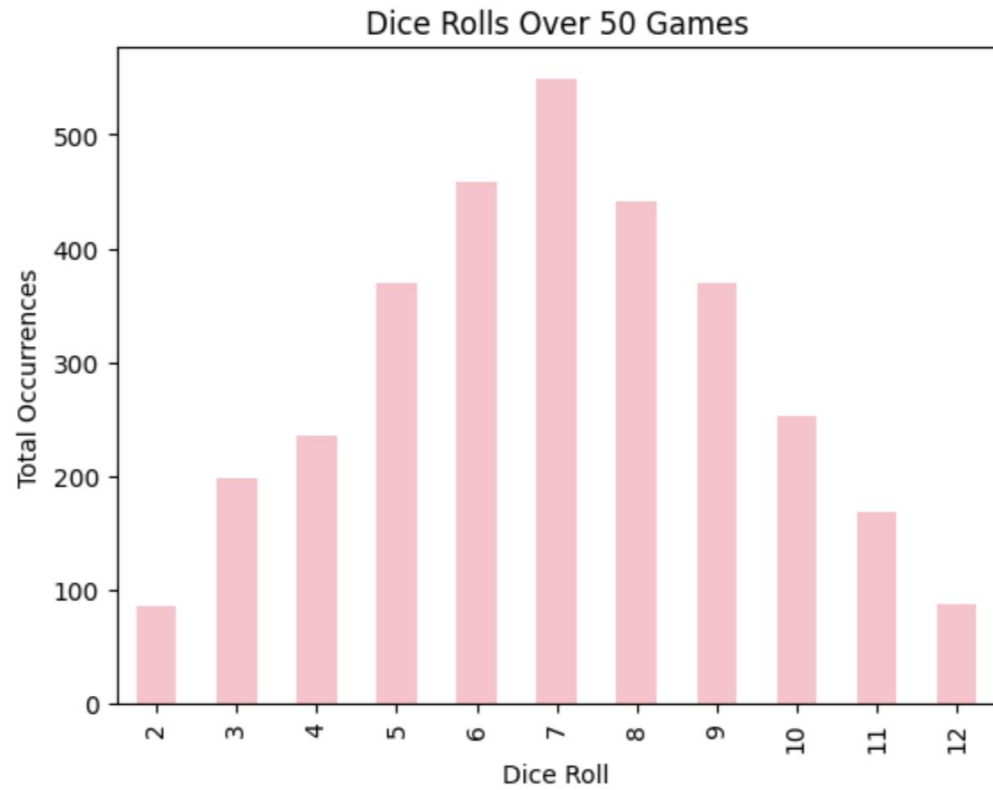


METHODOLOGY

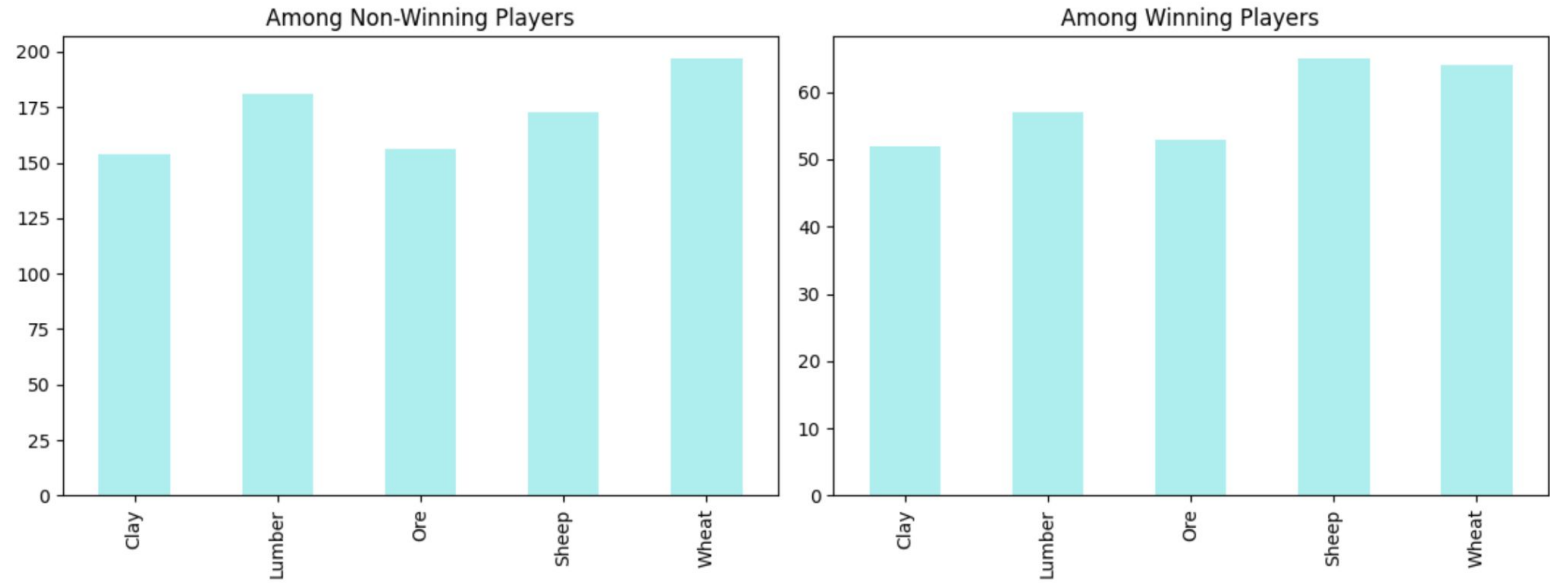


DATA VISUALIZATIONS



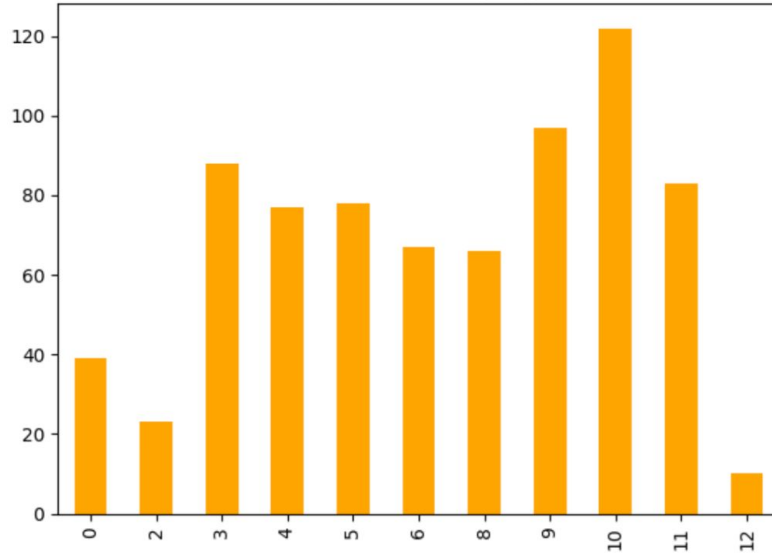


Frequency of Starting Resources

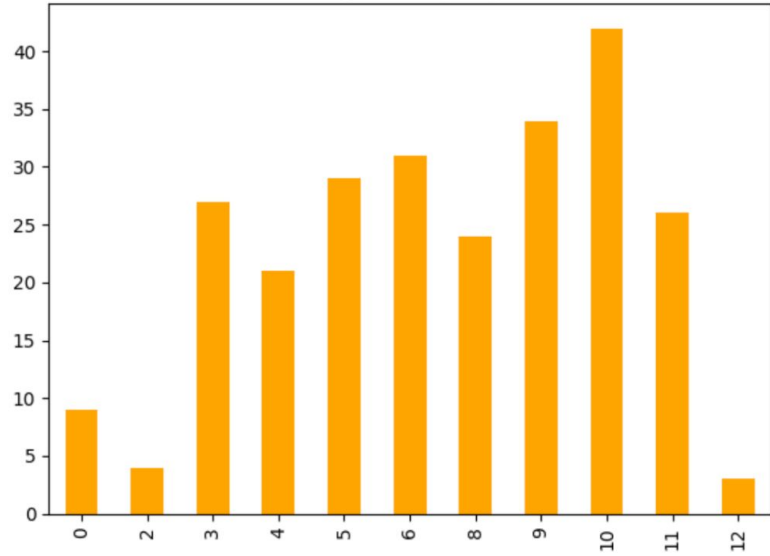


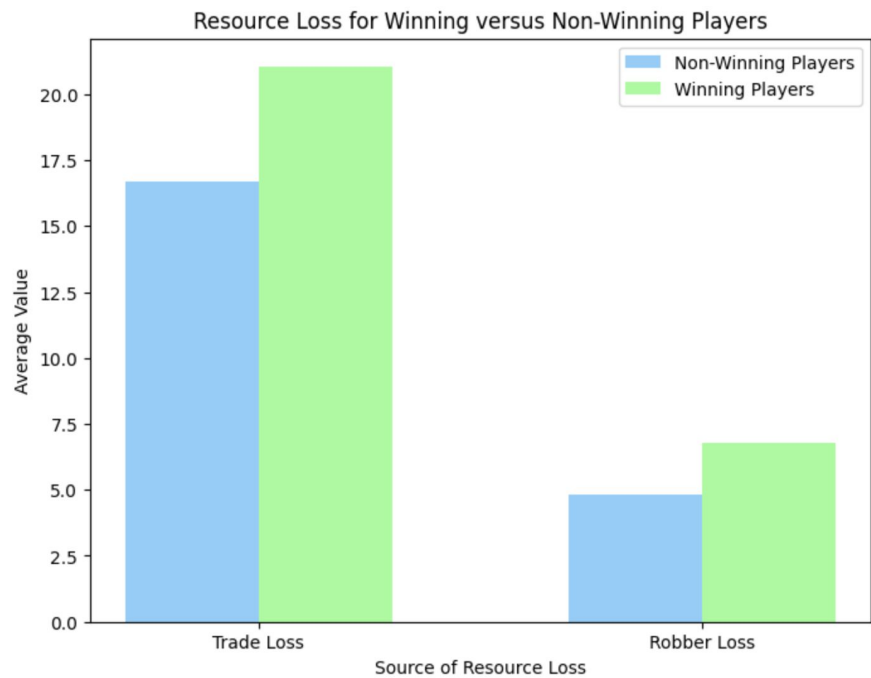
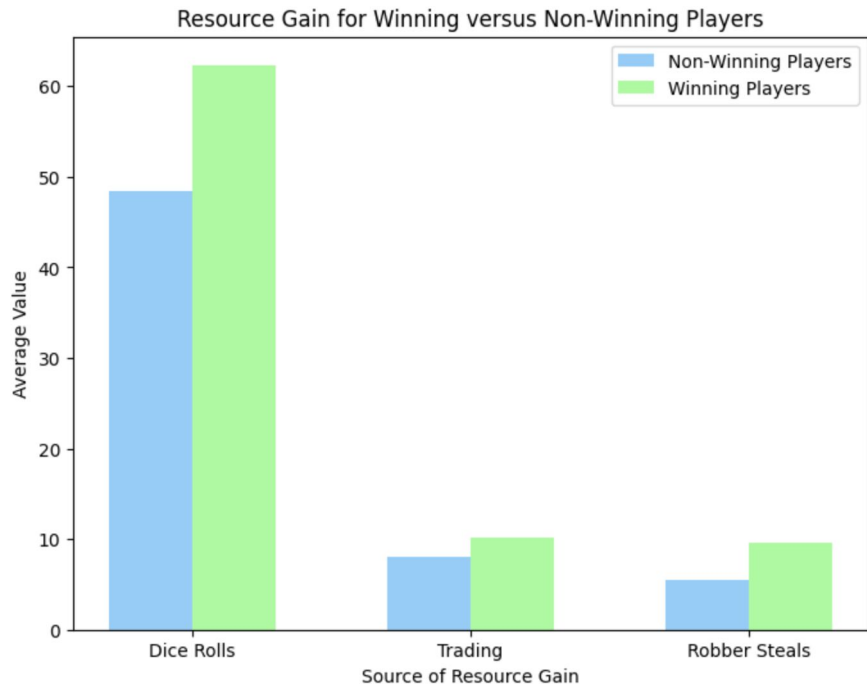
Frequency of Starting Number Placements

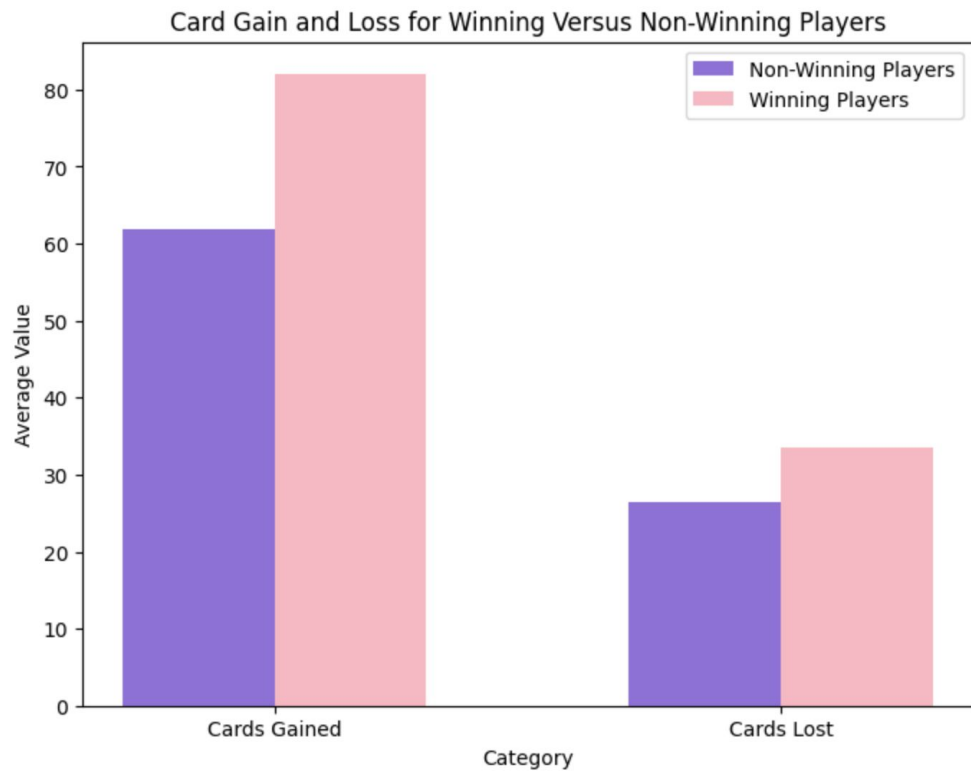
Among Non-Winning Players



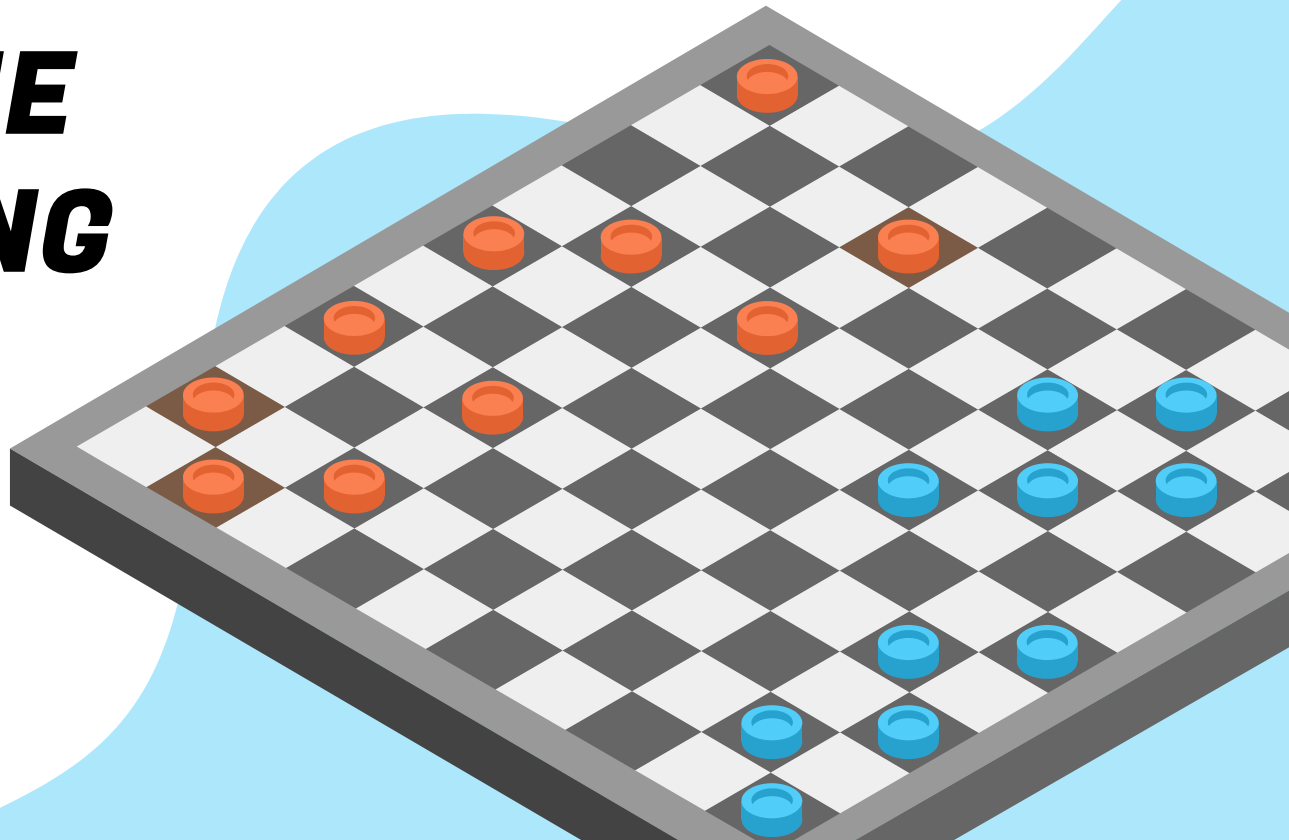
Among Winning Players







MACHINE LEARNING



PROCESS



Initial Model

KNN Classification with only starting resource counts



Next Iteration

KNN Classification with only starting resource counts using SMOTE oversampling



Next Iteration

KNN Classification with starting resource and number counts using SMOTE oversampling



Next Iteration

Logistic regression with starting resource and number counts using SMOTE oversampling



Next Iteration

Logistic regression with starting resource and number combinations using PCA

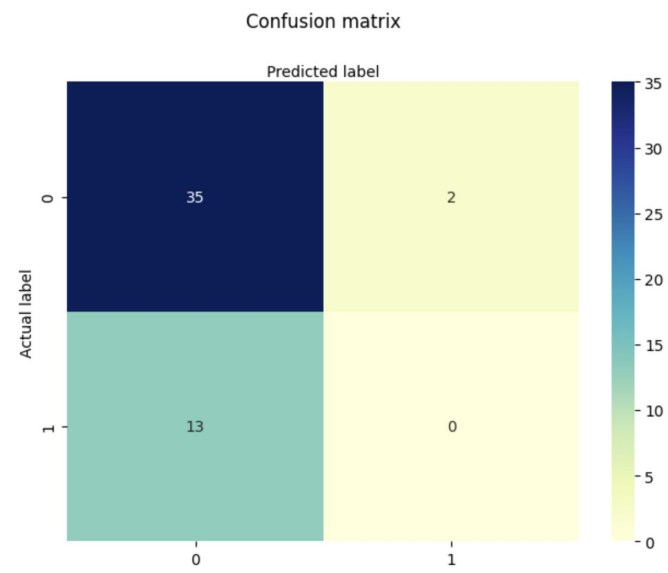


Analysis

Analysis of varying model performances and outcomes

Initial Model: KNN Classification with Only Starting Resource Counts

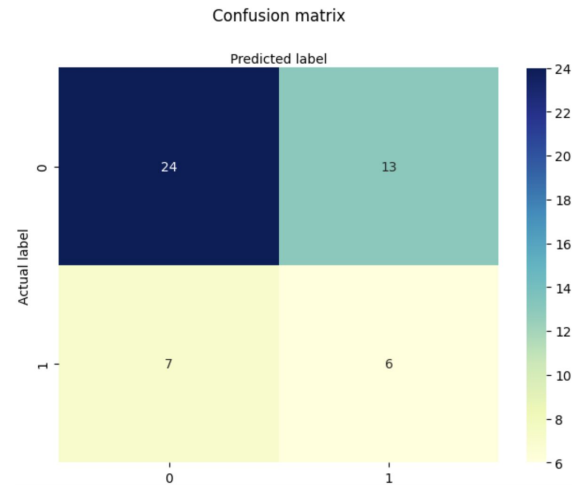
- ***K Value of 4 chosen***
- ***Decent accuracy, however, upon further analysis it became clear the model is unable to predict any wins at all***
 - ***The model is simply recommending the most common outcome, which is losses always***



Model	Accuracy	Precision	Recall	f1-Score
KNN Classification with Starting Resource Counts	Overall: 0.70	0: 0.73 1: 0.00	0: 0.95 1: 0.00	0: 0.82 1: 0.00

Next Iteration: KNN Classification with Only Starting Resource Counts with SMOTE Synthetic Oversampling

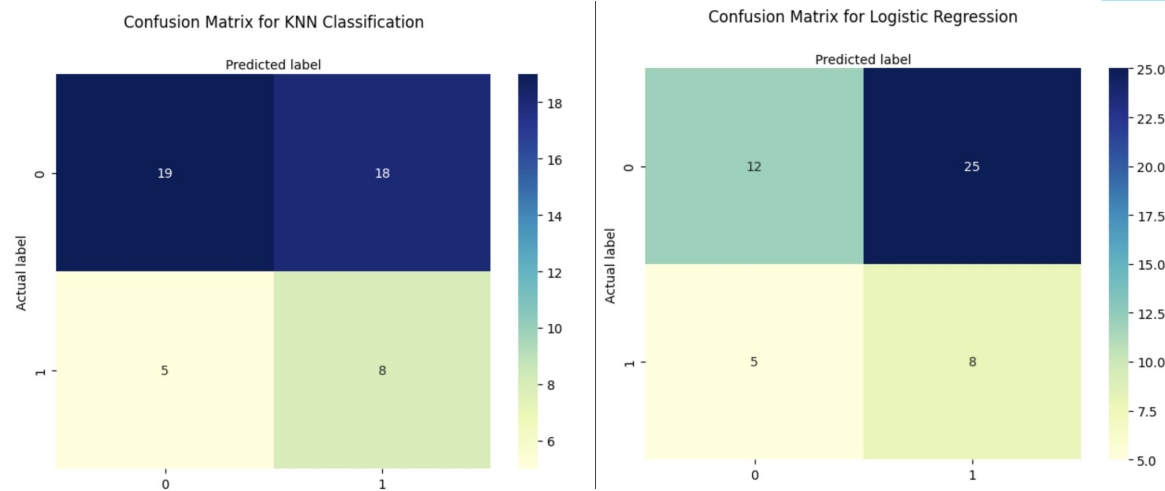
- *K Value of 5 chosen*
- *SMOTE oversampling results in an equal number of total wins (synthetic and organic) and total losses in the training data, with 113 of each in the now 226-row data set*
- *Improved precision and recall in model for wins, coupled with a decrease in accuracy*



Model	Accuracy	Precision	Recall	f1-Score
KNN Classification with Starting Resource Counts with SMOTE Synthetic Oversampling	Overall: 0.6	0: 0.77 1: 0.32	0: 0.65 1: 0.46	0: 0.71 1: 0.37

Next Iteration: KNN Classification and Logistic Regression with Starting Resource and Number Counts using SMOTE Oversampling

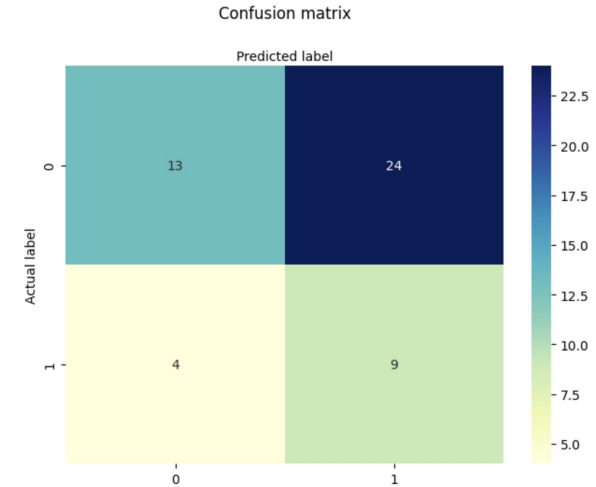
- ***K Value of 20 chosen***
- ***KNN model is perhaps the most effective of all iterations***
- ***Both utilize SMOTE synthetic oversampling to compensate for underrepresentation of the wins in the organic data***



Model	Accuracy	Precision	Recall	f1-Score
KNN Classification with Starting Resource and Number Counts with SMOTE Synthetic Oversampling	Overall: 0.54	0: 0.79 1: 0.31	0: 0.51 1: 0.62	0: 0.62 1: 0.41
Logistic Regression with Starting Resource and Number Counts with SMOTE Synthetic Oversampling	Overall: 0.40	0: 0.71 1: 0.24	0: 0.32 1: 0.62	0: 0.44 1: 0.35

Logistic Regression using Starting Placement Resource and Number Combinations, via PCA

- **No significant improvement in model performance**
- **Utilizes Principal Component Analysis (PCA) to**
- **Improved precision and recall in model for wins, coupled with a decrease in accuracy**



Model	Accuracy	Precision	Recall	f1-Score
Logistic Regression with Starting Resource and Number Combinations with PCA	Overall: 0.44	0: 0.76 1: 0.27	0: 0.35 1: 0.69	0: 0.48 1: 0.39

Comparing Model Outcomes

Model	Accuracy	Precision	Recall	f1-Score
KNN Classification with Starting Resource Counts	Overall: 0.70	0: 0.73 1: 0.00	0: 0.95 1: 0.00	0: 0.82 1: 0.00
KNN Classification with Starting Resource Counts with SMOTE Synthetic Oversampling	Overall: 0.6	0: 0.77 1: 0.32	0: 0.65 1: 0.46	0: 0.71 1: 0.37
KNN Classification with Starting Resource and Number Counts with SMOTE Synthetic Oversampling	Overall: 0.54	0: 0.79 1: 0.31	0: 0.51 1: 0.62	0: 0.62 1: 0.41
Logistic Regression with Starting Resource and Number Counts with SMOTE Synthetic Oversampling	Overall: 0.40	0: 0.71 1: 0.24	0: 0.32 1: 0.62	0: 0.44 1: 0.35
Logistic Regression with Starting Resource and Number Combinations with PCA	Overall: 0.44	0: 0.76 1: 0.27	0: 0.35 1: 0.69	0: 0.48 1: 0.39

FUTURE WORK



Neural Networks



***Game Outcome
Prediction***



Decision Tree



ANOVA Analysis

