Optimizing Player Strategy in Settlers of Catan

MGSC661 - Final Project Report

Student Name: Siraje Hatoum
Student No. 260509637

December 16, 2021

Contents

1 Introduction

1.1 Project Summary

In this project we will analyze a data set of 50 Catan games to identify patterns in the strategic decisions made by players and the impact of these choices on the final outcome. The aim of this analysis is to assess the existence of a set of common strategies amongst players as well as identify which choices have the most significant impact on player performance. The hope is that by identifying a set of strategies we can understand how the efficacy of these strategies is affected by events outside the player's control, such as dice rolls and opposing players' strategies.

1.2 Project Background

Catan is a strategy board game that has continuously grown in popularity since it's initial release in 1995. In addition to being a popular board game for casual fun amongst family and friends, it also touts a highly competitive community. Annual Catan tournaments are held worldwide at a regional, national, and international stage to crown the best of the best players as champions.

The base version of Catan is played by 4 players on a board consisting of 19 hexagonal tiles. Each tile has an associated resource and number (between 2 and 12). The placement of the resource tiles and their numbers is randomly selected at the beginning of each game. The object of the game is to be the first player to reach 10 or more points. Players accumulate points primarily by building settlements (1 point each) and cities (2 points each) on tile corners. At the beginning of each game players take turns to build a total of 2 settlements each. After this, players must expend resources to build more settlements and upgrade them to cities.

A player starts their turn by rolling a pair of 6-sided dice, which kicks-off the resource collection phase for all players. A player is allowed to collect resources if all of the following conditions are met:

1. The player has a settlement or city that is adjacent to the tile they are collecting from
2. The number associated with said tile equals the number rolled
3. The resource type collected matches the resource associated with said tile

There are a total of 5 resource types in the game: Sheep, Clay, Lumber, Wheat, and Ore. Different combinations of these resource types are used to build various structures yielding differing functional benefits. For example, upgrading from a settlement to a city requires 2 wheat and 3 ores and allows you to collect 2 resources from adjacent tiles instead of the 1 generated by a settlement.

In addition to collecting resources from the board, a player is allowed to barter for resources with other players on their turn. Trades between players can have any combination of resources exchanged as long as both players agree to the trade and have the agreed upon resources in hand. A player is also allowed to trade with the bank on their turn, this essentially allows the player to gain 1 of any resource by discarding 4 of the same resource type. A player also has the option to trade by port, if they have a port adjacent to any of their settlements or cities. These often work similarly to the bank but have a better rate while specifying the resource type to be discarded.

Aside from production and trade, resource management is also important. If a player holds more than 7 resources in hand when a 7 is rolled, that player is forced to discard half their resources (rounding up for odd numbers) as tribute.

The description above provided a high-level explanation of the main factors affecting the strategic decisions analysed in this project, for a comprehensive understanding of the rules please refer to Catan Game Rules and Almanac.

1.3 Project Motivation

As an avid Catan player, I have built my understanding of its intricate dynamics by playing over years and continuously reassessing my choices based on my experience and performance. This has allowed me to identify a few key considerations that I believe significantly improve a player's odds of winning. Due to the role chance plays in Catan, however, it can be difficult to assess the efficacy of strategies based on intuition. This project provides an opportunity to assess the validity of some of these beliefs by taking a statistical approach. This is especially interesting when evaluating efficacy based on events outside of a player's control.

2 Data Description

2.1 Data Source Analysis

The data used for this analysis was sourced from Kaggle. The data set contains 200 records i.e. 4 players over 50 games. Each record is uniquely identifiable by Player ID and Game ID. A single record contains observations on the total points accumulated by a player, the placement of their first 2 settlements, their resource production, trading behaviour, resource management, and the distribution of dice rolls throughout the game. Table 1 in Appendix A.1 provides a detailed data description table of all features in the initial data set.

As part of the initial data exploration the analysis focused on evaluating game performance (as measured by points) as a function of production volume, trading behaviour, resource management (as measured by tribute loss), and player order. Figure 1 shows a visual summary of these relationships.

From the graphs, we can see that production capacity and trade volume exhibit a positive trend with points accumulated. This is not surprising as production and trade are key mechanisms for generating the resources required to advance in the game. These data points are endgame stats, meaning they are collected after the game is completed and are not available at the start of a round. Unfortunately, they cannot be used to predict the outcome of a game that has not yet concluded. As such, in their current form, they have limited use in informing player strategy. Essentially, the data implies that in order to improve your chances of winning you should maximize your production and trade volume but provides no insight as to how.

On the other hand, the amount of resources lost as tribute shows no convincing link to points generated. This is somewhat unexpected, as poor resource management should negatively impact performance. The lack of a clear link may be attributable to the fact that we are analysing absolutes rather than proportion. Players with a high production rate are more likely to lose resources as tribute even if they effectively manage those resources. This relationship is made evident by the correlation matrix shown in Figure 2.

The distribution of points by player order, shown in Figure 1, suggests that the first player regularly performs worse than average while players who go second have a slight edge. This implies that initial placement of the first 2 settlements plays a role in determining player performance, as it is the only aspect of the game that is notably affected by player order. However, further analysis is required to understand what quantifiable difference, if any, exists between the strategies of players in different orders.

2.2 Feature Engineering Analysis

While the raw data set provides interesting preliminary insights, further data processing and feature engineering is required to achieve a strategically insightful model. The first feature engineering objective is to generate variables that provide better information regarding the strategic choices made

during initial settlement placement. For this purpose we created the following features based on the data set:

Expected Resource Gain
Based on the placement of the first two settlements the expected resource production per roll was calculated by resource type and overall. These features indicate the prioritization of production potential during the initial placement of settlements for specific resource types and overall resource production.

Tile Number Diversity
This feature is a simple unique count of all numbers associated with the tiles adjacent to the first 2 settlements. This feature measures the diversification of numbers in the initial set to minimize reliance on a small set of numbers for production.

Resource to Port Alignment
To account for the strategic inclusion of ports in the initial settlement placement, this variable was calculated as the product of the expected gain of the resource type that matches the port requirement. For example, if a player chose an ore port and their expected ore gain from the first 2 settlements was 0.1 per turn, then the alignment value is 0.1. In the case where a player used a general port, the resource type with the highest expected value was used to calculate alignment.

Figure 3 in Appendix A.2 shows the relationship between the expected resource production per round from the first 2 settlements and the points total. The results show a positive trend between overall expected return and total points accumulated. No single resource type exhibited a strong link to performance. This preliminary analysis implies that prioritizing any single resource type in a player's initial placements has no significant impact on game performance. Rather, players should focus on maximizing overall expected gain from their initial settlements.

Figure 4 in Appendix A.2 shows the distribution of initial port alignment, number diversity, and total points. No clear pattern emerges between points and port alignment or number diversity. This may be due to the lopsided distribution of both variables, with the majority of observations being on the top end or low end respectively. This suggests that if a pattern does exist, more observations are required to quantify it.

The next step of feature engineering focused on creating variables that provided deeper insights into player strategies around trading, production, and resource management.

Production and Trade Ratios
These feature explain what portion of total resources was acquired through production and trade respectively. They provide information on how much emphasis the player places on production as opposed to trading.

Trade Return
In order to better understand the trading strategy employed by a player, this feature provides the average gain to loss ratio of the player's trades.

Tribute Loss Ratio
To create a more meaningful measure of resource management, this feature provides the number of resources lost by tribute as a fraction of the total gained throughout the game.

Figure 5 in Appendix A.2 shows the relationship observed between game performance and these engineered features. Based on these plots no obvious link between any single feature and game performance is shown. In order to explore if the combination of these features can wield meaningful insights into game performance, we use clustering, tree-based, and classification methods in the next section.

3 Model Selection and Methodology

This section details the methodology used and models selected in our analysis. We begin by using the features identified above to assess the existence of common strategic combinations based on player choices. Next, the significance of these strategic patterns is evaluated with respect to game performance

and compared to other predictors. Finally, a classification model and regression model are developed and evaluated based on their ability to predict a game winner using the key features identified.

3.1 Identifying Strategy Clusters

Based on my previous experience playing Catan, I have noted that players often have different approaches to maximizing their likelihood of winning. Despite these variations, I believe that these approaches can likely be grouped into a small number of clusters representing high-level strategic patterns. The analysis performed at this stage relied on K-means clustering to evaluate the existence of distinguishable strategic patterns based solely on choices around initial placement, production reliance, trade behaviour, and resource management.

K-means was chosen as the clustering algorithm due to its widespread use in clustering applications and it's ease of tuning, attributed to its reliance on the number of clusters as the primary parameter. Due to its incorporation of both within cluster variance and cluster separation, the silhouette score was used to evaluate the validity of the clusters and choose the optimal number of clusters. Before running the model the features were standardized to avoid biasing the model based on predictor scale. Next multiple K-means models were run on the standardized data set with different cluster numbers. The average silhouette score of the clusters of each model was calculated and the model with the highest score was selected. The cluster labels were assigned to the non-standardized data set and the averages of the features were extracted for interpretation.

3.2 Evaluating Feature Importance

To better understand the significance of the identified clusters in determining game performance, their relative importance as predictive features was evaluated. This was done using a random forest model that processed the cluster labels as predictors, in addition to the previously engineered features, with total points accumulated as a target variable.

The random forest allowed for the calculation of the average drop in MSE in trees where a predictor wasn't used. Assessing this value as a measure of feature importance provided a ranking for the strategic clusters found in the previous stage against existing predictors. Additionally, this process identified predictors that reduced the accuracy of the model. These were labeled as poor predictors of performance and removed from the subset used in the next stage.

3.3 Predicting Player Success

After narrowing down the list of features based on importance, the attention turned to building a predictive model. Ideally, the model would have a high accuracy in determining whether a player will win the game based on their choices and its results could be interpreted to inform player strategy. At this point two modelling approaches were evaluated.

The first option involved creating a regression model to predict the number of points accumulated by a player, the winner would then be defined as the player with the highest number of predicted points. The initial step for this regression approach is to select an algorithm, and through tuning, minimize the error of the algorithm.

A Boosted Forest was selected due to it's strong predictive ability and tuning flexibility. A grid search method was used to determine the optimal number of trees, tree depth, and minimum observations per node for model parameters. When tuning the model, k-fold cross validation was performed to evaluate the MSE of the optimal hyper parameters. The results of the grid search can be found in Table 2 under Appendix B.1. Once the optimal hyper parameters were decided, winners were predicted from a random validation set of 5 games using the optimized model, trained on the remaining

45 games. This process was repeated 1000 times and the average accuracy of winner predictions was recorded to evaluate the predictive power of the model.

A major downside of this option is the loss of interpretation that occurs when using forest models. While we are able to compare the predicted performance of varying strategic choices, it's difficult to evaluate which changes would improve performance without rerunning the model. For this purpose we evaluated a Logistic Regression model, using the same predictors. For the target variable a binary feature was derived to indicate if the player won. The target was set to 1 if a player's score was greater than or equal to 10. As the game ends immediately after a player scores 10 or more points, this variable ensures that there is one winner per game. The primary advantage of the Logistic Regression model over the forest model, in this case, is the ability to interpret model coefficients to assess the impact of predictors on outcome.

Unlike the Boosted Forest model, Logistic Regression is less flexible in regards to tuning. The main choice available to the user is the likelihood threshold required to make a positive classification. Despite the utility of this choice, it presented a challenge due the lack of comparability with our previous approach. The method used to classify a game winner with the Boosted Forest prediction ensured that there is exactly one winner per game.

An alternative option was to classify all players with a predicted performance of greater than 10 points as winners. This method, however, would fall short in evaluating which player had the best strategy in the game, as anywhere between 0 to 4 players could be classified as winners per game. Hence the choice of classification method was made in order to incorporate the strategies of other players on the likelihood of winning. Using the Logistic Regression's classification output to determine whether a play would win or lose a game has the same shortfall, making it difficult to compare any accuracy metrics between the two approaches.

Another potential approach is to run the forest for classification rather than regression, using the win variable. This would certainly allow a more direct comparison between the predictions of the two models. In fact this was the initial modelling strategy, however it performed poorly. Specifically the classification models had a very high error rate when classifying winners, due to the categorical imbalance of the training data set. By design the data set has a 1:3 ratio for winning observations to losing observations. This meant that while the model achieved satisfactory accuracy overall, it performed poorly at identifying winners. Undersampling and oversampling options were considered to resolve the imbalance. Due to the limited size of the existing data set, undersampling methods were quickly dismissed. Oversampling methods, such as SMOTE, were considered and had potential but were discounted in favor of a regression analysis, to avoid unnecessary inferences on the data set.

Fortunately, the Logistic Regression model includes a functionality to predict classification probabilities rather than classification labels. These probability predictions were used to determine the game winners similarly to the previous model, i.e. the player with the highest probability of winning in a game was classified as the winner. A similar validation approach was taken to calculate the test accuracy of these predictions and compare to the performance of the Boosted Forest. Finally, the predictor coefficients were extracted for interpretation.

4 Results

This section will detail and analyse the output of the previously outlined modelling process with a focus on evaluating the validity of the strategy clusters found, understanding the feature importance ranking, and comparing model prediction accuracy between the two models.

4.1 Strategy Clustering

The K-Means model used to find strategic clusters implied the existence of two high-level strategic approaches based on the variables provided. Figure 6 in Appendix C.1 shows the change in silhouette score based on the number of clusters evaluated. The score is highest for 2 clusters at slightly above 0.15, implying that the generated clusters are not strongly separable and may not exist in truth. To interpret the meaning behind these clusters, we examine the predictor means for each cluster listed under Appendix C.1 in Table 3.

The data set contains 55 and 145 observations in Strategy cluster 1 and Strategy cluster 2 respectively, implying that Strategy 2 is a significantly more popular approach. To understand the functional difference between the two, we focus on the difference in Resource and Port Alignment. With an average alignment of 0, strategies in cluster 2 clearly do not incorporate ports in their initial settlement placement as opposed to strategies in cluster 1. Examining the variation between the two clusters for the remainder of predictors, we find that most differences can be explained by the choice of port alignment.

Tile Number Diversity is notably lower in cluster 1, this is due to the fact that ports do not have an assigned number and therefore cannot contribute to number diversity.

Expected Resource Gain is also lower for the majority of resource types and overall, this is because ports do not generate resources on roll. Therefore, players who choose a port over a resource tile will have less expected gain. The only observed exception to this rule is the expected ore gain, which is notably higher in cluster 1, implying that players who rely on ports for their initial settlements tend to prioritize ore production. This could be due to the ore requirement for building cities, a great way to overcome the reduced production capacity quickly, as it doubles the output of a settlement.

Production and Trade Ratios On average strategies in cluster 1 have a higher trade ratio and lower production ratio. This is not surprising as the trade ratio accounts for resources acquired through trades with players, banks, and ports. Players who prioritize port alignment early in the game will likely generate a higher proportion of resources by trading with ports and a lower proportion from resource tile production.

Trade Return is notably higher in cluster 1, this again is explained by compounding effects related to early prioritization of port alignment. As the port provides a better return on certain trades, the player is less likely to be forced to make poor return, 4 for 1 trades, through the bank. Players in cluster 1 are also likely to focus on maximizing return for a specific resource from other players, e.g. if I can exchange 2 wheat for any resource type then I will require a better return on trades where I give away wheat and a higher volume for any trades where I receive wheat, to justify the trade.

Tribute Loss Rate is only slightly higher in cluster 1, implying no significant impact on resource management. This is likely due to the fact that a players' risk tolerance is the key contributing factor in their resource management. Players with low risk tolerance are likely to trade or expend cards to ensure they are below the tribute threshold, while players with a higher risk tolerance may hold off on doing so in hopes of new resources that provide better alternatives on their next turn. This may imply that players who prefer ports in the early stage have a higher risk tolerance, but we lack of sufficient supporting evidence to assert this.

Based on this analysis and despite a low silhouette score, it appears that there is a strong logical argument for the existence of these two strategy clusters based on choices regarding initial placement, production reliance, trade behaviour, and resource management. The next step is to evaluate whether the strategy type chosen significantly affects game performance.

4.2 Feature Importance

To better understand the impact of the identified strategy clusters on performance, we begin by visualizing the discrepancy in points accumulated and win rate between them, as seen in Appendix C.2 Figure 7. For both metrics, it is clear that Strategy 2 averaged a higher performance than strategy 1, however the error bars indicate that the difference in averages is not statistically significant within a

90% confidence interval. In the absence of a larger sample set to improve confidence in these metrics, the focus shifted to determining the clusters' feature importance relative to the previously engineered features.

Figure 8 in Appendix C.2 shows the feature importance output by a Random Forest model, trained to predict the game points accumulated by a player. Based on the improvement to predictive power, the strategy cluster is ranked 7th of 12 predictors with an average improvement of approximately 1%. The most important feature is the overall expected resource gain with an improvement of 10%, more than double that of the next most important feature. The importance of the overall expected resource gain is not surprising, given the previously observed link between production volume and points accumulated. Interestingly, port alignment ranked second with an improvement of about 4%, this may be due to the link between port alignment and the other features as discussed previously.

On the lower end, tile number diversity, expected lumber return, and trade return rate all scored negative percentages, implying that their inclusion reduced the predictive power of the model and that they have no tangible impact on game performance. Particularly interesting is the inclusion of lumber in this mix, as it is the only resource type whose expected gain had a lower than 2% improvement contribution. At this stage it is unclear why this is the case, although it may be explained by a lower variance in the feature across players.

These results suggest that, while the strategy cluster may be a useful addition to a predicative model, it is not greater than the sum of its parts, wielding a significantly lower predictive importance than the predictors used to define it. The analysis also revealed some features that should be discounted from the predictors list, as they do not improve the predictive power of the model.

4.3 Success Predictions

Having narrowed down the list of useful predictors, and after optimizing the Boosted Forest hyper parameters for the data set, we are able to assess the predictive power of the regression model. Based on the cross validation performed during hyper parameter tuning the optimal model achieved an MSE of 4.6. Given the 10 point scale, an average error of roughly 2 points is large, indicating that the model is not a powerful predictor of points accumulated.

Despite this, the model may still have use when evaluating relative performance amongst players. Using the previously outlined validation methodology for predicting player outcome, the model achieved an accuracy of 34% when predicting winners, i.e. 34% of the predicted winners actually won the game. While this prediction rate is by no means impressive, it is a 9% improvement over the probability of randomly selecting a winner. This suggests that the Boosted Forest model does have some useful insight when comparing predicted player performance.

Next we evaluate the output of the Logistic Regression model and compare it's predictive power to the Boosted Forest. Table 4 in Appendix C.3 provides an overview of the model coefficients. The coefficient of Strategy cluster 2 seems to contradict the previous evidence by indicating that players who prioritize port alignment early in the game are more likely to win. However, as with the previous observations, the coefficient provided by the model does not appear to be statistically significant, as such no conclusion can be made at this time. Of the 9 predictors used only two are found to be statistically relevant, expected resource gain and production ratio. The relatively high positive coefficient implies that an increase in expected resource gain significantly improves a player's likelihood of winning.

Interestingly, production ratio had a negative coefficient, implying that in order to improve their chances of winning, players should diversify their sources of resource generation. Players can do this by increasing the volume and quality of their trades, to account for a larger percentage of resource generation. The production ratio coefficient may also be explained as an indicator of resource value, players that own resources that are scarce in the game are typically given more trades at a higher return rate. This theory could be tested by comparing a player's expected gain for each resource type

with other players', to evaluate the existence of a resource monopoly and the correlation of this feature with the production ratio.

Given the previously indicated feature importance, the low statistical significance of the majority of predictors is somewhat surprising. As these features have demonstrated value in predicting game performance in previous models, these results imply that the majority of features do no have a linear relationship with the log likelihood of winning a game. Indicating that a Logistic Regression model is not a good fit on the predictors without a non-linear transformation being applied. Before coming to a conclusion, the standardized validation test is performed and the accuracy of winning predictions is compared with the previous model.

The win prediction rate of the Logistic Regression is approximately 36%, preforming slightly better than the Boosted Forest. This combined with the insights generated through the coefficients and the relatively faster training rate make it the better choice for our final model.

5 Recommendations and Conclusion

Based on the analysis performed we can confidently conclude that player strategies built around initial placement, production reliance, and trade behaviour can significantly impact player outcome. In particular, our analysis found that a player can significantly improve their chances of winning by maximizing the expected resource production from their initially placed settlements. Our working assumption is that this approach will allow for faster growth in the early game, facilitating improved placement of subsequent settlements and increased rate of production overall.

Our analysis also revealed a key trade-off between maximizing initial expected resource gain and port alignment. In fact, our observations suggest that players' strategic approach is most easily differentiated by their decision to incorporate ports in their initial settlement placement. As a port must invariably take the place of a resource tile, it consistently reduces the resource production rate in the early game. However despite this trade off, our study did not reveal any significant differences in outcome between the two approaches. As such, we recommend careful evaluation of the larger context of the game before making a decision on port alignment in the early game.

Our results also indicated that players can generally benefit from diversifying their resource generation sources to lower their reliance on roll production in the long run. We recommend that players achieve this by increasing trade volume and trade return. To that end, we suggest taking advantage of ports where possible, as our observations have shown that their incorporation has a positive effect on trading behaviour, specifically trade return rate. We also suspect that diversification can be achieved by increasing player resource value, for which there are a few approaches available. These include monopolization of scarce resources and creation of artificial resource scarcity through the robber mechanism. Given its impact on our predictive model, we believe this should be a secondary decision after prioritizing expected resource gain in the early stages.

Interestingly, the analysis did not reveal any resource type as a statistically dominant force in affecting game outcome, however singled out early reliance on lumber generation as a particularly poor indicator of success. The same was found to be true for trade return rate and tile number diversity. On this basis, we do not recommend that players prioritize these strategic considerations over the ones discussed above.

In conclusion, despite a lower than anticipated predictive power, we found that a simple logistic regression, coupled with tree based and clustering methods, can wield insights into optimal game strategies. Interesting extensions to this problem include a sensitivity analysis on the predictive performance based on the probability of number rolls in a game and the incorporation of features that assess game context, such as scarcity measures for resource types by game. Major mechanisms that were not accounted for in this analysis include the generation and use of development cards as well as the incorporation of the longest road and largest army into player strategy. These mechanisms were

omitted due to the limitation of the available data set and would likely significantly elevate the insights and power of the predictive models evaluated.

A Data Description Details

A.1 Raw Data Exploration

Variable Name	Description
gameNum	Game identifier
player	Player starting position
points	How many points the player ended the game with
2	Number of times 2 was rolled during the game
3	Number of times 3 was rolled during the game
4	Number of times 4 was rolled during the game
5	Number of times 5 was rolled during the game
6	Number of times 6 was rolled during the game
7	Number of times 7 was rolled during the game
8	Number of times 8 was rolled during the game
9	Number of times 9 was rolled during the game
10	Number of times 10 was rolled during the game
11	Number of times 11 was rolled during the game
12	Number of times 12 was rolled during the game
settlement1	3 sets of paired columns indicating the resource type and number associated with the 3 tile adjacent to the player's first settlement
settlement2	3 sets of paired columns indicating the resource type and number associated with the 3 tile adjacent to the player's second settlement
production	Resources gained from settlements and cities
tradeGain	Resources gained from peer and bank trades
robberCardsGain	Resources gained from stealing with the robber, plus cards gained with monopoly cards
totalGain	Sum of all resources gained
tradeLoss	Resources lost from peer and bank trades
robberCardsLoss	Resources lost from robbers, knights, and other players' monopoly cards
tribute	Resources lost when player had to discard on a 7 roll
totalLoss	Sum of all resources lost
totalAvailable	Net gain of resources

Table 1: Description of raw data acquired from Kaggle

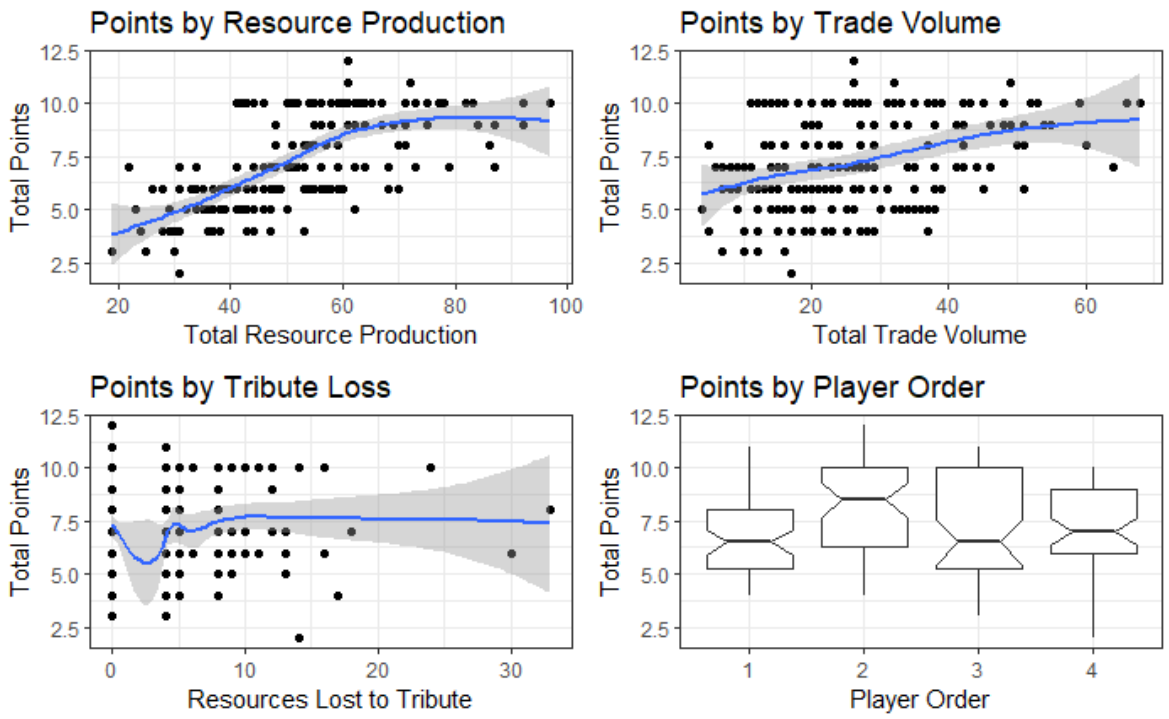

Figure 1: Visualization of game performance and factors included in raw data set

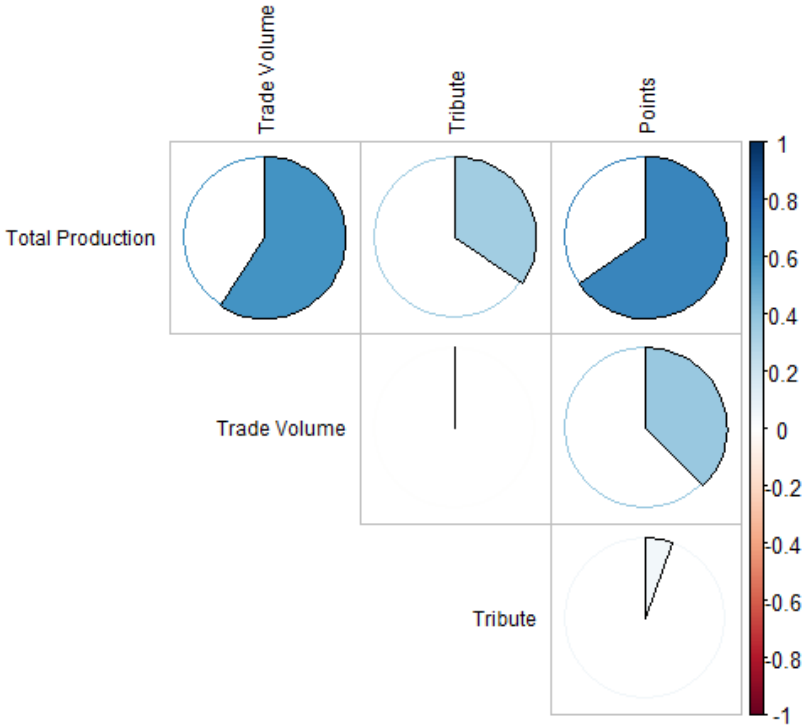

Figure 2: Correlation matrix of key features from raw data set

A.2 Engineered Feature Exploration

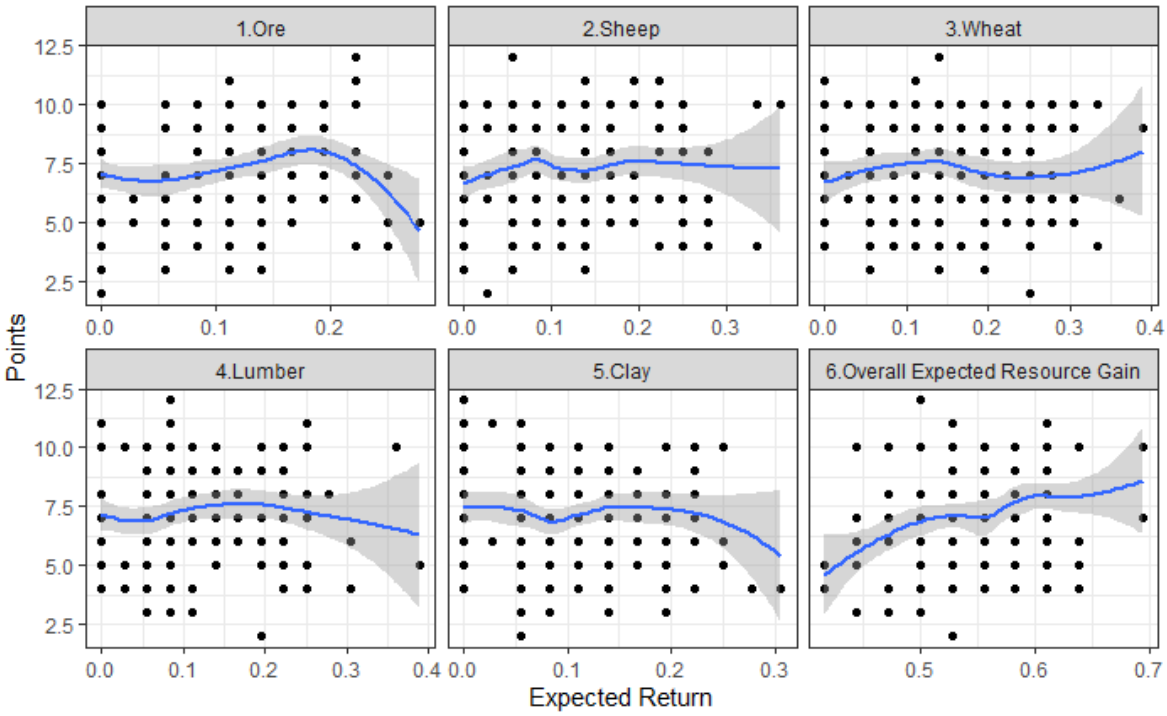

Figure 3: Visualization of game performance as a function of expected gain per resource type

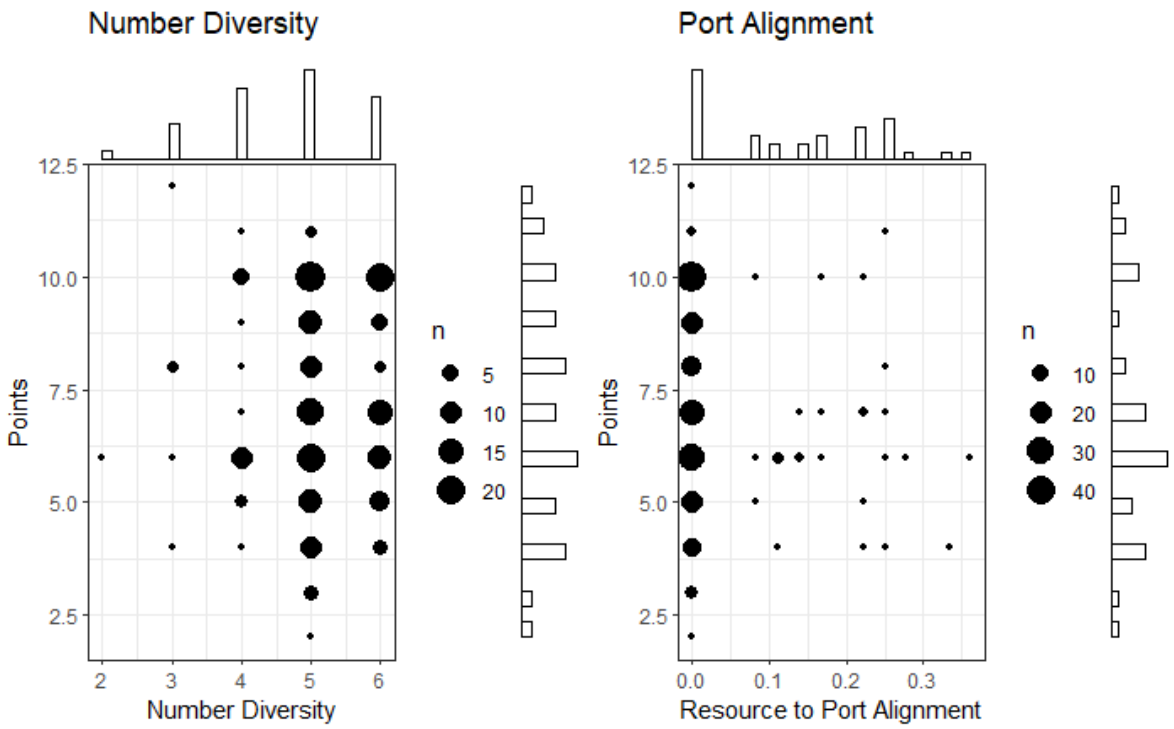

Figure 4: Visualization of points distribution by number diversity and port alignment

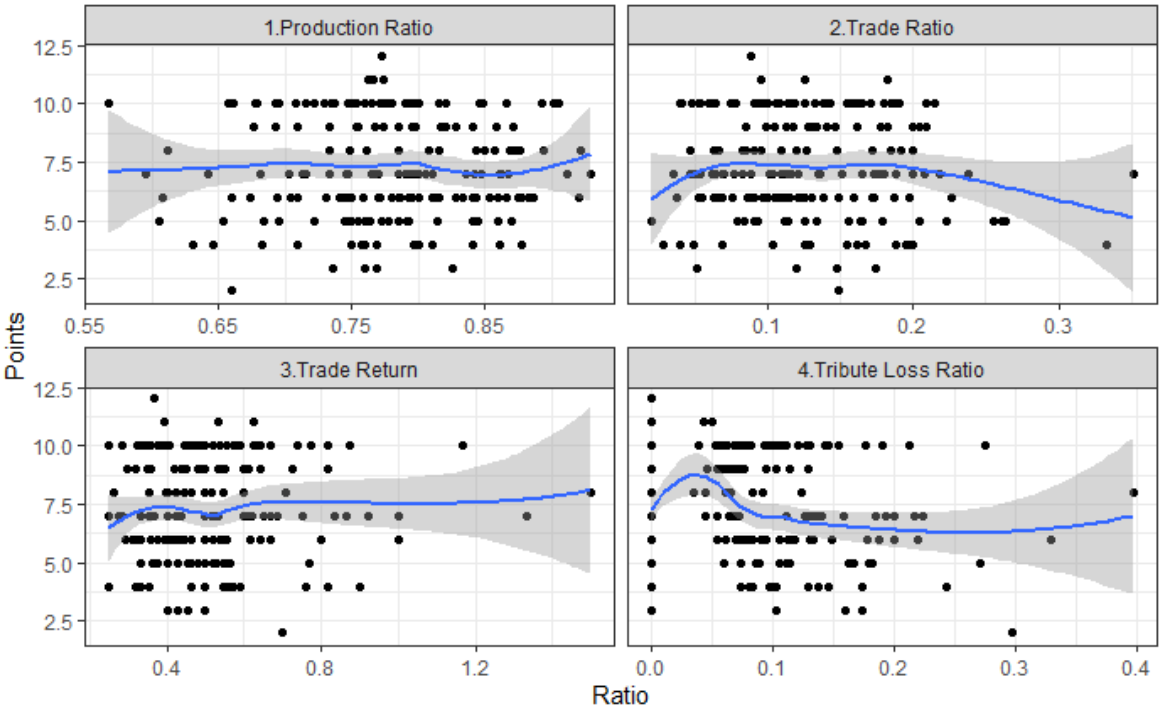

Figure 5: Visualization of game performance against the feature engineered ratios

B Model Selection and Methodology

B.1 Predicting Player Success

	Trees	Tree Depth	Minimum Observations per Node	Average MSE
1	100	3	5	4.608
2	100	2	3	4.674
3	100	4	4	4.783
4	100	2	4	4.791
5	100	2	2	4.794
6	200	4	2	4.802
7	200	2	5	4.876
8	100	4	3	4.910
9	500	2	4	4.943
10	100	4	5	4.952

Table 2: The hyper parameter configurations that returned the 10 lowest MSE

C Results

C.1 Strategy Clustering

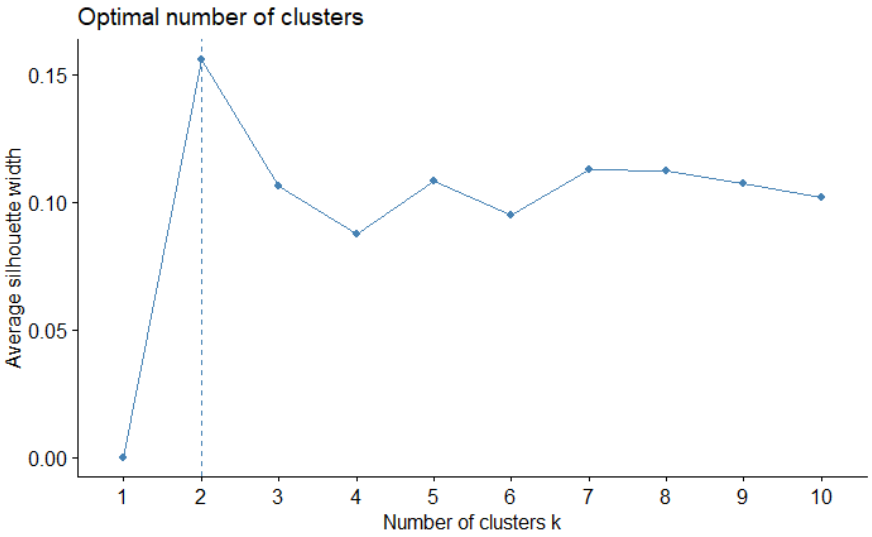

Figure 6: Silhouette score of by cluster number

Table 3: Predictor means of identified strategic clusters

	Strategy 1	Strategy 2
Tile Number Diversity	4.473	5.400
Expected Ore	0.122	0.092
Expected Wheat	0.121	0.134
Expected Sheep	0.087	0.119
Expected Clay	0.089	0.105
Expected Lumber	0.079	0.127
Expected Resource Gain	0.497	0.578
Resource and Port Alignment	0.09	0.00
Production Ratio	0.756	0.785
Trade Ratio	0.143	0.122
Trade Return	0.565	0.488
Tribute Loss Rate	0.079	0.073
Frequency	55	145

C.2 Feature Importance

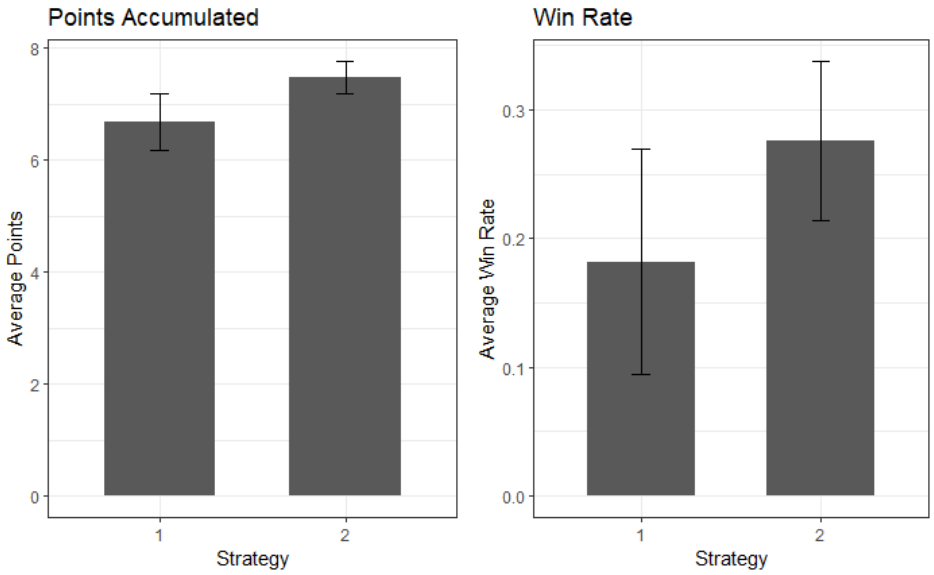

Figure 7: Visualizations showing the difference in game performance observed between the identified strategy clusters

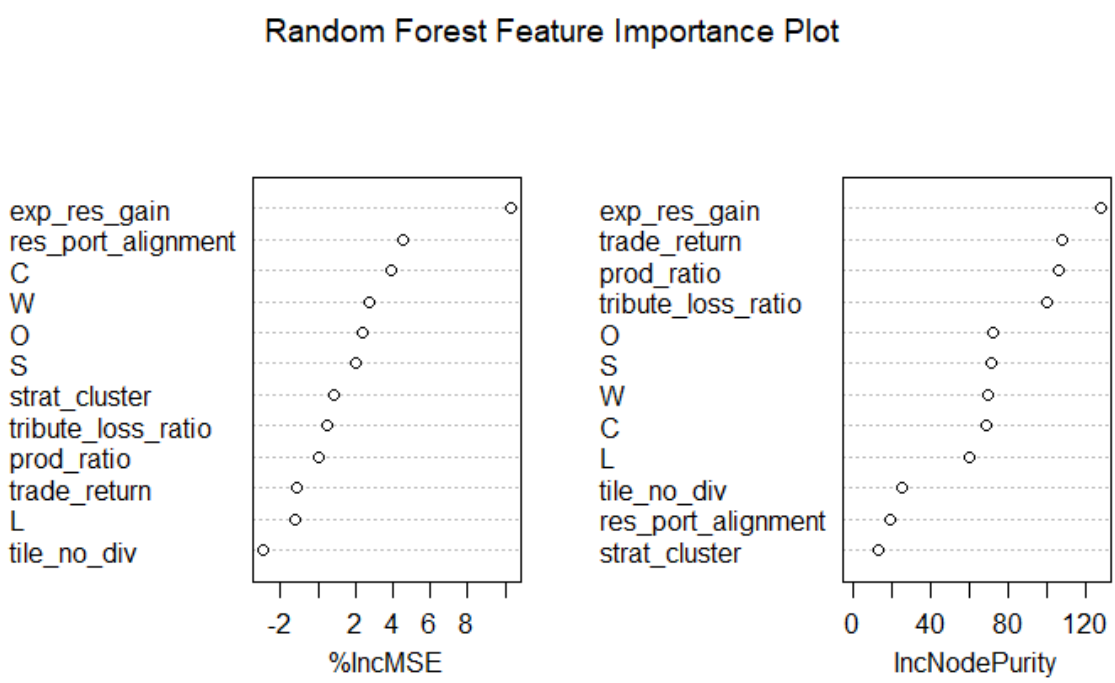

Figure 8: Feature importance plot generated by a random forest model

C.3 Success Predictions

Table 4: Summary of Logistic Regression Model

	Dependent variable:
	win
Expected Ore Gain	2.129
	(2.937)
Expected Sheep Gain	0.050
	(2.686)
Expected Wheat Gain	−1.786
	(2.608)
Expected Clay Gain	−0.038
	(3.267)
Expected Resource Gain	12.100**
	(4.876)
Resource and Port Alignment	−1.746
	(3.761)
Production Ratio	−4.945*
	(2.587)
Tribute Loss Rate	−1.247
	(2.383)
Strategy Cluster 2	−0.321
	(0.618)
Constant	−3.713
	(3.036)
Observations	200
Log Likelihood	−103.831
Akaike Inf. Crit.	227.662
Note:	*p<0.1; **p<0.05; ***p<0.01