**Task 1 – Association Rules**

1. Mailing all seven catalogs to each of the 50,000 customers would cost Exeter $1,050,000.

1. If the sample data is representative of the whole population of Exeter customers, it would cost $311,306.53 to send a catalog to customers who had previously bought from the same type of catalog.
2. The most helpful rule is customers who buy from the computers catalog are 2.689 times more likely to buy from the garden catalog. 75% of those who bought items from the computers catalog also bought from the garden catalog, so future computer catalog customers will likely be interested in the garden catalog.

Also, customers who bought from the housewares and jewelry catalogs are 2.095 times more likely to buy from the novelty gift catalog. About 47% of the customers who made purchases in both jewelry and housewares also made purchases in the novelty catalog.

Finally, customers who bought items in the novelty gift catalog were 1.737 times more likely to buy from the garden catalog than those who did not purchase novelty gifts. About 49% of novelty gift customers also bought from the garden catalog, so they should receive both novelty gift and garden catalogs in the future.

1. If a customer only bought items from the jewelry catalog and the company can only send them two new catalogs, they should also receive the novelty gift catalog and the housewares catalog. This is because customers are 1.552 times more likely to buy novelty gifts if they purchased jewelry and 35% of customers who buy jewelry also buy novelty gifts. In addition, customers are 1.403 times more likely to buy housewares if they bought jewelry than if they had not purchased jewelry, and 60% of customers who bought jewelry also buy housewares. There is also a strong correlation where customers who bought jewelry and housewares also buy novelty gifts.
   1. If a customer makes a purchase from the garden catalog, they should also receive the computers catalog. 50% of customers buying from the garden catalog also buy from the computers catalog, increasing the likelihood of buying a computer by 2.869 times.
   2. If a customer buys from the computers catalog, they should also receive the garden catalog. 75% of customers buying from the computer catalog also buy from the garden catalog, increasing the likelihood of buying from the garden catalog by 2.869 times.
   3. If a customer buys both jewelry and garden products, they should also receive the automotive catalog. 63% of customers buying both of these products also buy from the automotive catalog, increasing the likelihood of buying automotive products by 1.459 times.
   4. If a customer buys both automotive and garden products, they should also receive the jewelry catalog. 69% of customers buying both of these types of products also buy jewelry, increasing the likelihood of buying jewelry by 1.455 times.

**Task 2 – Clustering: NBA Basketball Players Case**

You are to cluster the players along each of the following *dimensions* separately: (a) effectiveness as a scorer; (b) effectiveness as a defensive player; (c) effectiveness as a utility player (which may include assists, less turnovers, number of games, etc.); and (d) overall performance that includes all three

1. Justify variables:
   1. For the effectiveness as a scorer, I used the following metrics:

* Offensive rebounds- this describes rebounds secured by a player after one of his/her teammates missed a basket. If a player secures an offensive rebound, he/she has the opportunity to try to make another basket.
* Total rebounds- the total rebounds corresponds with the total offensive and defensive rebounds a player managed to secure. When the player has the chance to get the ball, he/she has a greater likelihood of being able to make a play or another basket. The more the ball is in the possession of the individual, the more likely his/her team can score.
* Field goal percentage- field goal percentage is the percentage of all shots excluding free throws. This category includes three point shots, jump shots, lay ups, etc. A high field goal percentage equates to a high efficiency, which can ultimately determine if a team wins or loses.
* Three-point percentage- this percentage is an average of three-points scored. Considering three points are the highest amount of points an individual can make in one shot, it is imperative that an effective shooter would be able to maintain a high three point percentage to overall increase the points earned during a game.
* Free throw percentage- this percentage is an average of free throws completed. An effective scorer is also talented at free throw. Teams depend on these points when a player is fouled. An effective scorer is able to make his/her free throws to help his/her team.
* Minutes per game- the minutes a player plays during a given game is also important to how effective he/she is. This metric shows the durability of a player as well as their stamina. An effective scorer must be able to play for a specific length of time, yet still maintain high shooting statistics.
  1. For the effectiveness as a defensive player, I used the following metrics:
* Defensive rebounds – this corresponds with the ability for a player to recover a ball after his/her opponent shot and missed. A good defensive player that can get the ball back for his/her team can set up plays in his/her favor.
* Steals per game- the steals per game metric is important when measuring effective defense players. If an individual can have high steals, he/she can turnover the ball from the opposing team and deliver the ball to his/her team, which in turn can lead to a profitable play from his/her team.
* Blocks per game- like the steals per game, the blocks per game are an important metric for a defensive player. The more blocks a player completes, the less points an opposing team scores, which ultimately helps his/her team increase the point spread of the game.
* Minutes per game- like an offense player, a defensive player needs to have adequate durability to be effective during a game. The minutes per game metric measures the stamina of a defensive player.
  1. For the effectiveness as a utility player, I used the following metrics:
* Games played- To be a true utility player, you must always be available for your team, so games played effectively shows the players who go out and play every day for their team.
* Field goal percentage – One mark of a utility player is their efficiency on the court. Even if they do not score in bulk, a utility player should use their looks effectively.
* Assistes per game – An effective utility player should be able to create shots for their teammates.
* Turnovers per 100 minutes – Utility players, as well as all others, should be able to hold onto the ball. The best way to measure this is Turnovers per 100 minutes.
  1. For the overall performance, I used the following metrics:
* Games started- the best overall player would be the starter for his team. Thus, the amount of games started is a good indicator with how effective the player is on his/her respective team.
* Field goal percentage- like previously mentioned in an effective shooter, field goal percentage is the percentage of all shots excluding free throws or “live shots”. The higher the field goal percentage, the more points a team can have, which leads to the increased likelihood of a team winning.
* Total rebounds- as mentioned in effective shooter and defender, a player that can regain ball control and pass it to his/her team can ultimately enable the team to score more points.
* Assists per game- an effective player also knows when to utilize his/her team members. The ability to know when to shoot as well as when to set up his/her team to shoot will overall help in his/her team’s chance at winning.
* Turnovers per 100 minutes- this is a more accurate representation for the amount of turnovers per player. An effective player would have low turnovers, as it would inhibit his/her team from scoring points and would enable the opposing team to have the ability to score more.

1. Average values of the chosen variables for each cluster:

Offensive

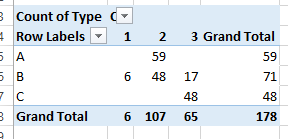
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster | Offensive Rebounds | Total Rebounds | Field Goal Percentage | Three-point percentage | Minutes per game |
| Cluster 1 | .29 | 1.29 | 41% | 30% | 7.46 |
| Cluster 2 | 1.56 | 6.14 | 47% | 32% | 34.3 |
| Cluster 3 | .94 | 3.73 | 47% | 28% | 22.1 |
| Cluster 4 | .96 | 4.62 | 48% | 29% | 27.88 |
| Cluster 5 | .86 | 3 | 47% | 25% | 16.4 |

Defensive

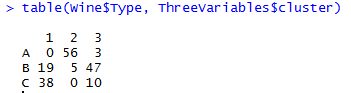
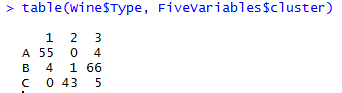
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Defensive rebounds | Steals per game | Blocks per game | Minutes per game |
| Cluster 1 | 2.92 | .76 | .4 | 24.22 |
| Cluster 2 | .99 | .19 | .15 | 7.46 |
| Cluster 3 | 3.56 | 1.02 | .58 | 29.87 |
| Cluster 4 | 4.99 | 1.16 | .79 | 34.95 |
| Cluster 5 | 2.31 | .52 | .44 | 16.9 |

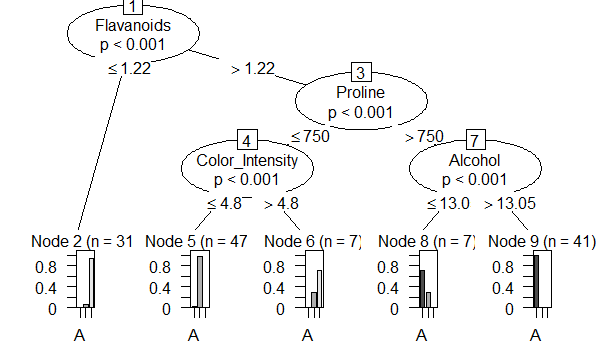
1. Further questions:
   1. Show the number of players in each group, as long as this number is greater than 1. The cluster numbers will go in order as Scorer, Defensive player, Utility, and Overall.
      1. 1,2,3,5: 12 players
      2. 1,2,4,5: 4 players
      3. 2,4,2,4: 3 players
      4. 2,4,2,3: 2 players
      5. 2,4,5,2: 5 players
      6. 3,1,2,4: 5 players
      7. 4,1,2,4: 3 players
      8. 4,3,2,4: 2 players
      9. 4,3,5,2: 4 players
      10. 5,5,1,1: 9 players
      11. 5,5,2,1: 5 players
      12. 5,5,3,5: 4 players
      13. 5,5,4,5: 3 players
   2. Examine the salaries of the players within each group. The average salary for each group is as follows. Clearly groups 4 and 5 have the highest salaries.
      1. $663,806
      2. $2,118,035
      3. $9,330,123
      4. $15,200,001
      5. $12,161,000
      6. $4,955,274
      7. $3,717,500
      8. $1,750,400
      9. $9,970,000
      10. $1,355,400
      11. $1,834,179
      12. $1,159,000
      13. $2,243,291
   3. The average salaries of each team are as follows
      1. Mavericks - $4,500,000
      2. Rockets - $3,500,000
      3. Grizzlies - $4,480,000
      4. Pelicans - $4,400,000
      5. Spurs - $4,000,000
   4. If you were a basketball team manager, select three players you would hire and justify
      1. Jeff Withey
         1. I chose this player because he fell in offensive player cluster 1, which seemingly specialized in a higher 3-point percentage and he is in Utility cluster 4, which have a high number of games played, a high field goal percentage, and a low number of turnovers. In addition, Withey has amongst the smallest contracts in the entire listing at $490,000
      2. Monta Ellis
         1. I chose Monta because he is amongst only 2 players that fall in defensive cluster 4, which specializes in high numbers of blocks and rebounds, and also in overall player cluster 3, which have a high number of games started, and the fewest amount of turnovers/100 min. His salary is also less than the other player which fills these two requirements at only $8,360,000
      3. Chandler Parsons
         1. I chose this player because he is in offensive category 2, which have very high 3-point percentages, defensive category 4, which have high minutes and rebounds per games. In addition, every other player that shares the same exact clusters as Chandler Parsons is paid at least $7,900,000, which Parsons is only paid $926,500, meaning that he is vastly underpaid for his skill level.

**Task 3- Wine**

1.) The following Pivot Table shows the accuracy of our clusters. In general, the clusters were relatively accurate. Each wine in Type A and C were clustered exactly to their type, but the Type B wines were distributed among the three clusters, so it was not as accurate.

2.) Looking under the spreadsheet “3 Variables”, we made a pivot table of all the variables and computed the values based on a percentage of Type A. This would show us how far away Type B and Type C are from Type A’s value. The variables with the highest percents would be those that cluster the wines the best. So we added the absolute value of the percent differences for Type C and Type B and found that Proline, Flavanoids, and Nonflavanoid Phenols cluster the wines the best.

3.) Same as #2, we picked the five variables Malic Acid, Proline, Flavanoids, Color Intensity, and Nonflavanoid Phenols by choosing the top 5 total percents away from type A. The five variable clusters were much more accurate than the three variable clusters, as evidenced in the tables to the right. The Types were much more distinguished with the five variable clusters. In total, there were 37 wines falsely categorized in the three variable data, whereas in the five variable data there were only 14 wines falsely categorized. Thus the more variables you have, the more accurate your clusters will be.

4.) Based on the spreadsheet “Classification”, we found that 86.67% of the predictions were accurate. This is a relatively high value, meaning that our Classification Tree is reasonably accurate.

Full Tree Rules: If Flavanoids are less than 1.22, then it is most likely Type C. If Flavanoids are greater than 1.22 and Proline is less than 750 and Color Intensity is less than 4.8, it is most likely Type B. If Flavanoids are greater than 1.22 and Proline is less than 750 and Color Intensity is greater than 4.8, it is most likely Type C but could also be Type B. If Flavanoids are greater than 1.22 and Proline is greater than 750 and Alcohol is less than 13, it is most likely Type A, but could also be Type B. If Flavanoids are greater than 1.22 and Proline is greater than 750 and Alcohol is greater than 13, it is most likely Type A.

5.) Based on our classification tree, all of the variables matched our top 5 variables except for alcohol. This means that Flavanoids, Proline, and Color Intensity were all variables that we considered in the earlier questions. Through R, the classification tree was able to find the most accurate variables whilst including outliers and extraneous data, whereas we used simple averages to find the most valuable variables. Obviously, our variables in #2 and #3 match, as we used the same calculations for both questions. The five variable data just had two additional variables.

A.) As shown in the “Clustering” spreadsheet, the net profit when the Types are randomly generated is $750. However, since the random generator continues to generate new numbers, this net profit will change each time we refresh the spreadsheet.

B.) When we used the clustering technique instead, our net profit was $1355. Obviously, this is a larger number because our more accurate predictions reduced the number of times we faced the $55 penalty.

C.)We then used the Classification method shown in the “Classification” Spreadsheet and scaled our net profit to an equal amount of bottles (178), rather than using the sample 45 bottles. In this, we found that the net profit was $4845.56. This is a much larger profit than both the random and the cluster profits, demonstrating that the classification technique is much more effective.

D.) In conclusion, we found that the Classification Technique provided a much more accurate representation of the Types than both the Clustering and Random Techniques. As a result, firms should rely on this data more than randomly generated samples and clustered samples.