ABSTRACT

STRACK, MATTHEW GABRIEL. Team Assignment and Location Determination for the NCAA March Madness Tournament. (Under the direction of Brandon M. McConnell).

The National Collegiate Athletic Association (NCAA) March Madness tournament is a 68-team tournament played across eight locations in the first two rounds. Any team that travels more than 350 miles may fly to their game, which is an additional cost of 30–50k. The research objective is to schedule the tournament to reduce the number of teams flying to their respective game sites by selecting the best eight predetermined game locations while keeping the tournament structure intact. This paper investigates the use of two maximum covering models using Binary Integer Programming to choose the best eight predetermined game site locations and compare the number of teams flying to the actual tournament results for three seasons. Results show a significant reduction in the number of teams flying to their game sites for all three seasons providing considerable travel cost savings.

Team Assignment and Location Determination for the NCAA March Madness Tournament

by
Matthew Gabriel Strack

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Operations Research

Raleigh, North Carolina
2023

APPROVED BY:

Russell E. King

Kristin Thoney-Barletta

Brandon M. McConnell
Chair of Advisory Committee

BIOGRAPHY

I am currently an Army Officer getting my masters in Operations Research at NC State. I have served in the United States Army for 9 years that took me to Oklahoma, Tennessee, Texas, and North Carolina. In addition to those locations I deployed to Poland and Iraq. I started my military journey at the United States Military Academy in 2010 leaving my hometown in Champaign, Illinois.

ACKNOWLEDGEMENTS

Special thanks to Dr. Brandon McConnell for all his effort in helping me complete this paper. Also, thanks to the committee and other classmates who provided feedback throughout the process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

1.1 Background

The National Collegiate Athletic Association (NCAA) March Madness tournament, known as “March Madness," is a 68-team single-elimination tournament for division I college basketball. This multi-billion dollar tournament draws a significant amount of attention from the fans predicting the games resulting in millions of predicted brackets and thousands of workplace “bracket pools" where fans compete by forecasting tournament outcomes using a bracket as illustrated in Figure 1.1. Fans wagered roughly 8.5 billion U.S. dollars in the 2019 March Madness tournament Associate (2019). The amount of money and participation in this tournament shows the importance of this tournament. March Madness is a unique tournament in sports constructed in a way that is very different from other sports tournaments. March Madness has 68 teams in the tournament, which consist of 32 automatic bids and 36 at-large bids. Automatic bids are for the teams who win their respective conference tournament. The at-large bids are for teams selected by the committee who did not win their conference tournament. The committee conducts 4 steps to construct the first round matchups.

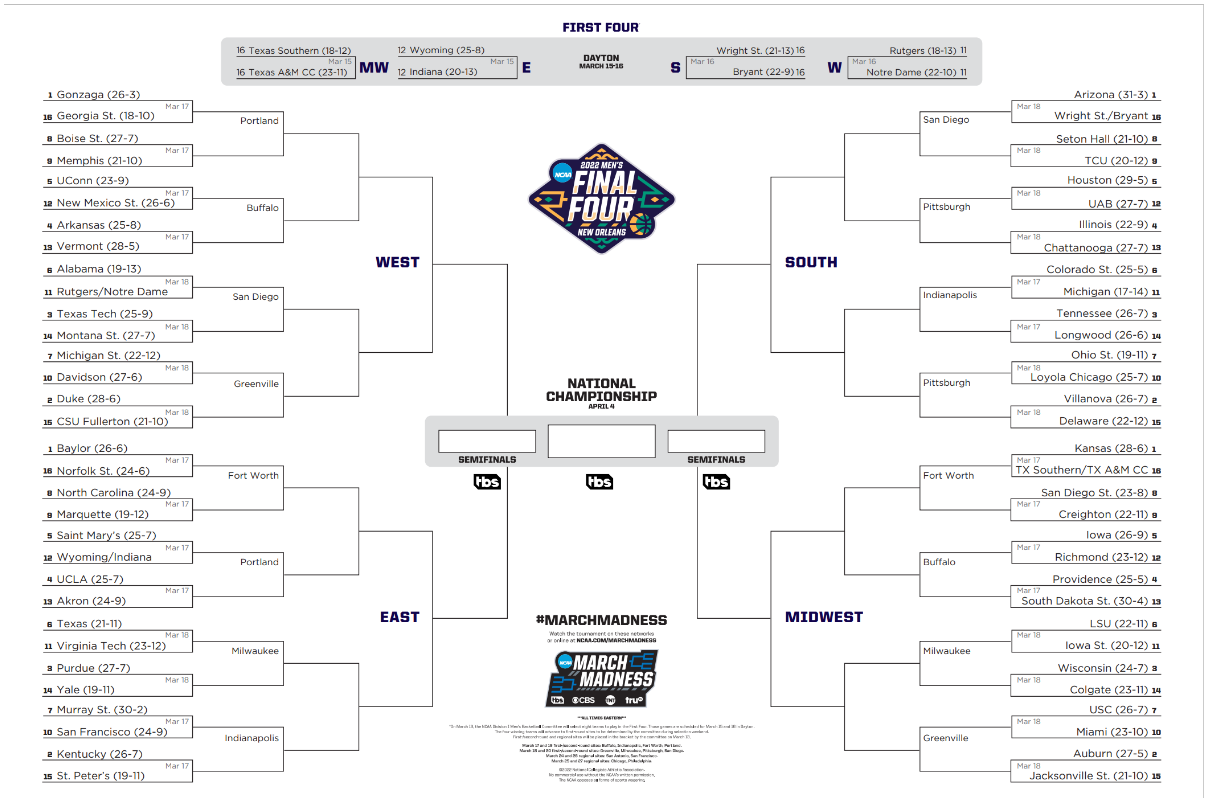

Figure 1.1: 2022 NCAA March Madness Bracket.

Step 1: Team Ranking. The first step the committee does after selecting the 68 teams is they rank order the teams from 1–68. The committee determines the ranking based on each team's resume: number of wins, quality wins, bad losses, conference ranking, and strength of schedule. The structure of the tournament is built into four regions. Each region has seeds from 1 through 16, which means each region has 16 teams. In each region, in the 1st round, the Number 1 seed plays the Number 16 seed, the Number 2 seed plays the Number 15 seed, the Number 3 seed plays the Number 14 seed, and so on, until the Number 8 seed plays the Number 9 seed. Figure 1.1 shows the tournament structure (Sager 2022). The tournament structure allows for 64 teams, but we have 68 teams. To solve this problem, the committee has eight teams play in four "play in" games. The four winners then get slotted into that 64-team field. The committee decides the eight teams who must play in these "play in" games based on the 1–68 team rank order list, commonly called the "s curve"(ncaa.com 2021). The lowest four automatic bid teams and the lowest four at-large bid teams on the s curve are selected to play in the “play in" games. The four play-in games are in Dayton, Ohio, each year.

Step 2: Assign Teams to Regions. The next step the committee does is slot the individual teams into a region with their respective seed. Figure 1.2 is an example of a region that the committee assigns. The figure is an example from the west region in the 2022 NCAA tournament. It shows the seed matchups. The first four teams in the s curve are all the Number 1 seeds, the following four teams (5–8) in the s curve are all the Number 2 seeds, and the following four teams (9–12) in the s curve are all 3 seeds. The committee does this procedure through all 16 seeds. The committee starts with the Number 1 seeds. They slot the Number 1 overall team in the s curve as a Number 1 seed to the closest region and location. They then slot the Number 2 overall team in the s curve as a Number 1 seed to the next closest region and closest location. They do the same process for the 3rd and 4th overall teams, which slots all the Number 1 seeds. The committee then places the No. 2 seeds in each region, placing the overall Number 8 team in the same region as the overall Number 1 team, the overall Number 7 seed in the same region as the overall Number 2 team, the overall Number 6 team in the same region as the overall Number 3 team, the overall Number 5 team in the same region as the overall Number 4 team. The logic is here is rewarding the Number 1 seeds by giving the best Number 1 seed the worst Number 2 seed. An important note is that the committee will uphold the principle of keeping the top four teams from the same conference in separate regions. The best four teams in a conference cannot be in the same region. The committee will then place the Number 3 seeds in their respective regions (overall Number 9 team with the Number 1 and 8, overall Number 10 team with the Number 2 and 7, overall Number 11 team with the Number 3 and 6, and overall number 12 seed with Number 4 and 5). The committee will maintain the principle of keeping the top four teams from the same conference in separate regions. The logic is that it gives the best number 2 seed the lowest-ranked Number 3 seed. The committee will then place the Number 4 seeds in their respective regions (overall No. 13 team with the Number 4, 5, and 12, overall Number 14 seed with the Number 3, 6, and 11, overall Number 15 seed with the Number 2, 7, and 10, and the overall Number 16 seed with the Number 1, 8 and 9). The logic is that the best overall 3 seed has the lowest 4 seed in their region. The committee will maintain the principle of keeping the top four teams from the same conference in separate regions. The committee uses the same logic to snake the rest of the seeds 5–16.

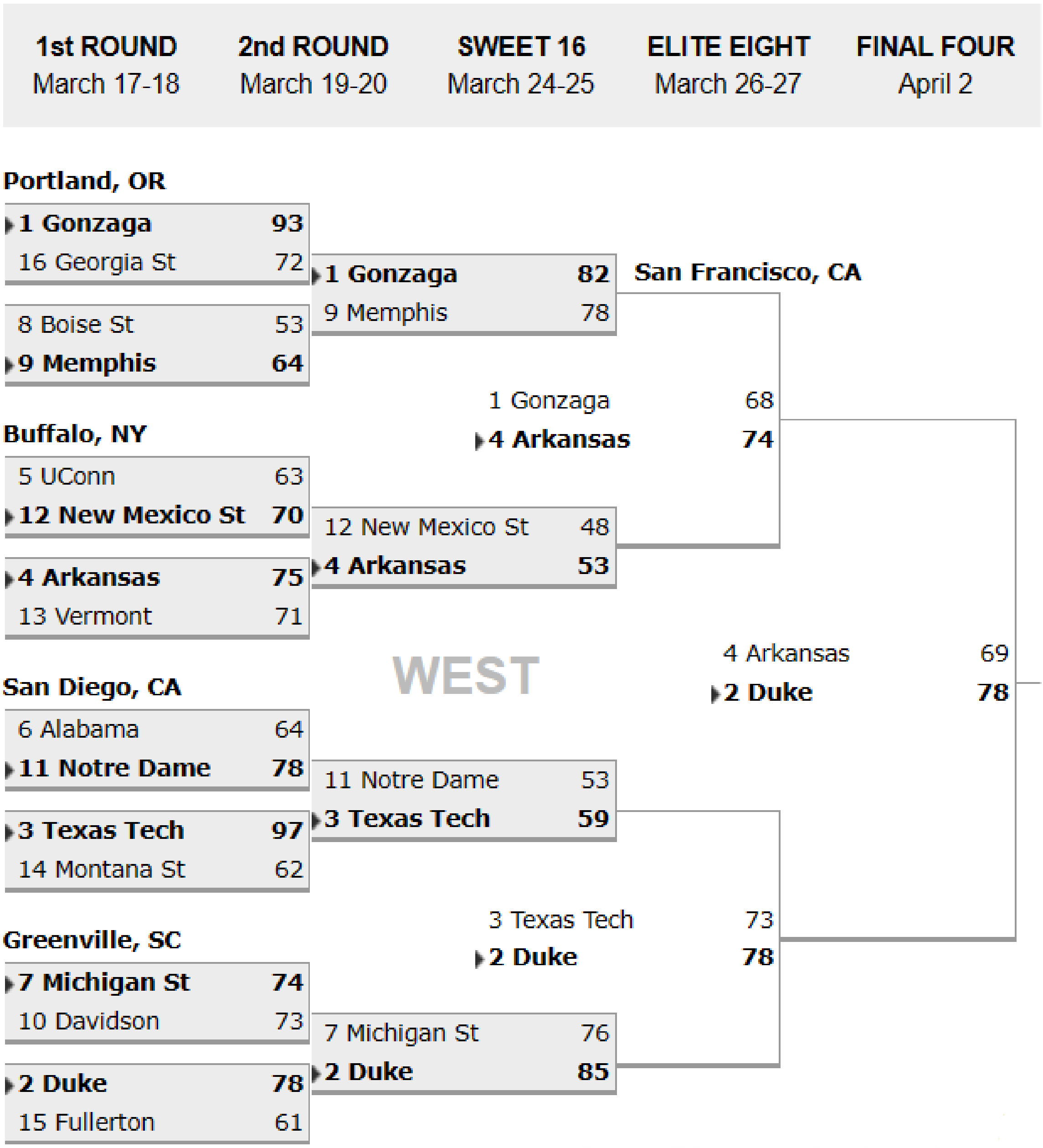

Figure 1.2: NCAA Region Example.

Step 3: Locations. After the committee assigns each team to its respective regions and seeds, the committee next decides on the locations of the games. There are 8 locations for the first-round games. However, they slot the teams in pods to ensure the second-round games are at the same location as their first-round game. A POD consists of teams playing against each other in the first round but also of the teams they may play in the second round. For example, the east region seeds 1,16,8,9 are in the same POD because the winner of the 1 vs. 16 will play the winner of the 8 vs. 9 game. Example of all the POD types are in table 1.1. Each location can have two pods or eight teams. The overall Number 1 team gets the closest location, so their POD 1,16,8,9 in that region will all play at that location. The overall Number 2 team gets the closest location to them, so their POD 2,15,7,10 all play there. The committee does this through the 16th overall team. Once a location has two pods assigned, it is complete, and no more teams can be assigned there and must choose the next closest location. Once the 16th overall team is assigned, all pods are assigned, and all locations are assigned. After this step, all teams have an assigned region, seed, and location.

Table 1.1: Pod Structure.

POD Number	POD Matchup
1	Number 1 vs Number 16
	Number 8 vs Number 9
2	Number 2 vs Number 15
	Number 7 vs Number 10
3	Number 3 vs Number 14
	Number 6 vs Number 11
4	Number 4 vs Number 13
	Number 5 vs Number 12

Step 4: Pairwise Comparisons. The committee's last step is conducting pairwise comparisons of each game and POD. They must ensure to satisfy all criteria; if not, they must switch teams. The committee uses geography as the driving force and will switch teams out if they have to travel too far or cross too many time zones (Palm 2016). There are no specific distance criteria or the number of time zones, so it is a subjective decision by the committee. To switch teams means to move teams within the same seed. For example, they may swap two different Number 2 seeds around to meet the criteria. In rare cases, they can move a team up or down a seed level to ensure they satisfy all criteria. There are

some objective criteria for each game. As mentioned earlier, the first four teams from the same conference must be in different regions. For example, if both Duke University and the University of North Carolina are in the top four of the Atlantic Coast Conference, they must be in different regions. Another criterion is that teams from the same conference cannot be in the same POD if they played twice during the regular season. The committee also tries to avoid non-conference rematches in the first round and avoid having teams in the same conference in the same POD. In extreme cases, the committee may move teams up or down one seed line or even two seed lines to satisfy all the bracketing principles. The committee goes through all the pairwise comparisons, and the field is complete once all principles are satisfied. This is the current scheduling process for March Madness.

1.2 Motivations

March Madness is one of the most popular events in the United States, with some years close to 10 billion dollars spent around the event. Roughly 700,000 people traveled to watch the games in person, while millions watched the games on TV. The championship game alone had 18.1 million viewers in 2022 (NCAA Staff 2022). Eight cities host the tournament's first two rounds and benefit from all the revenue generated by hotels, food, and transportation costs. With so much money and viewership on the line, the NCAA has treated this product as an asset. One problem that comes out of the tournament is the turnaround of releasing game locations and the actual games. The tournament game matchups and locations are released Sunday night, and the games are the following Thursday and Friday. The play-in games are on Tuesday and Wednesday, So the NCAA, teams, and fans have to coordinate all logistics as early as one day after the release of the tournament and with most having two days to figure it all out. The NCAA pays for team travel, so all schools must coordinate with the NCAA on transportation. The quick time makes logistics costs even more.

The NCAA uses a mileage policy to determine if a team will fly to their game or take a charter bus. For the NCAA March Madness tournament the policy is 350 miles, which means if a team has to travel 350 miles or farther the NCAA will pay for the team to fly (NCAA Staff 2023). There are multiple incentives to have a team close to game sites. First, a closer game will draw more fans to attend because they will have to pay less and use less time to get to a game. The tournament will generate more revenue for the sport and the local city hosting the game by having more fans. Second is the expense to fly a team vs.

drive a team. The average cost of a short turnaround private jet is approximately $40,000 compared to the average $1,500 to charter bus a team(evoJets 2023). This cost is an estimate and should not be used as a fact. It is based off average round trip cost for the allotted 75 personnel the NCAA authorizes to travel. There is a 2,566.67% increase in cost per team that flies instead of drives to their respective game site location. With roughly 48 hours for teams to book their travel it is much more expensive to fly than drive to their game site. This is motivation for the NCAA to reduce the number of teams having to fly to their game site location. Third, from the teams' perspective, the less travel, the better because traveling is the additional variable the NCAA must deal with while preparing to play a game.

1.3 Problem Statement and Open Questions

The NCAA March Madness tournament includes teams from all over the country, which makes distance a big concern. As the March Madness committee said, "Geography is king". Therefore, this paper will focus on the idea of minimizing the distance between teams and their game site location in the first round. This paper will not focus on future rounds because that requires probabilistic models of prediction and goes beyond what we are trying to accomplish in this paper. This paper will focus on two goals. The first goal is to minimize the total miles teams need to travel to their game site location. Decreasing the distance helps increase attendance while limiting travel hours for the teams to help facilitate better-quality games. The second goal is minimizing the number of teams flying because it costs the NCAA 2,566.67% more to fly than drive a team (evoJets 2023). This paper attempts to accomplish these goals by answering the following questions:

> Can we automate the NCAA March Madness tournament assignment process to minimize the total number of miles traveled while still meeting the March Madness principles?
>
> How much is the tournament solution impacted if we set a maximum distance travel limit?
>
> How do we minimize the total number of miles traveled and the number of teams flying while still meeting the March Madness principles?
>
> Does selecting our own eight game site locations reduce the number of miles traveled and the number of teams flying?

1.4 Structure of the Paper

The paper follows the structure of the chapters described. Chapter 2 states the current literature around sports scheduling and specifically the NCAA March Madness tournament. Chapter 3 explains the formulation the paper uses to minimize the total distance traveled in the preliminary round of the NCAA tournament. Chapter 4 describes how a penalty function is added to the model to reduce the number of teams flying to their respective game site. Chapter 5 uses two types of maximum covering model models to select 8 predetermined locations with the goal to reduce the number of teams flying to their game site. Chapter 6 is a brief summary of the results and the future work needed.

CHAPTER

2

LITERATURE REVIEW

This chapter discusses the literature and concepts in sports scheduling, tournament structure, location problems, and the NCAA March Madness tournament.

2.1 Sports Scheduling

Sports analytics is becoming increasingly popular as the industry continues to grow. A popular field within sports analytics is sports scheduling. Sports scheduling is involved worldwide, and a great example is the 2022 World Cup Final, where 32 countries played. Most countries have professional sporting teams that require complex scheduling. In the United States, the NFL, NBA, and MLB are big industries with complex scheduling problems. Henry and Holly Stephenson appeared in a documentary for MLB scheduling where they scheduled 30 teams for 162 games with many constraints like travel (Grantland Staff 2013). Durán (2021) provides a survey of sports scheduling survey paper tailored towards Latin American sports where multiple mathematical techniques were applied. Even e-sports is becoming famous worldwide, which more than 30 million people watch monthly. Dong et al. (2023) mentioned adopting different sports scheduling strategies in e-sports. Every tournament has different types of sports scheduling problems with a different set of constraints. There are two types of scheduling problems where one is the regular season, and the other is the tournament schedule. This paper focuses on the tournament schedule.

2.2 Tournament Structure

Scheduling a sports tournament depends on the type of tournament it is. There are three types of tournaments: Round Robin,Swiss-system, and knockout (Dong et al. 2023). Sziklai et al. (2022) measured the efficacy of the different types of tournaments to see which style was the best. Round robin is the most common type of tournament structure (Durán 2021). A round-robin is where each team plays all the other teams once. There are variants to this structure, like a double round robin where you play all the other teams twice, once at home and once away. The traveling tournament problem is an example of a round robin tournament structure (Easton et al. 2001).

The second type of tournament structure is the Swiss-system. The Swiss-system uses a fixed number of rounds that each team plays with the goal of matching teams with a similar ranking each game while trying to avoid rematches. After a fixed number of rounds, the team with the best score is the winner. The advantage to this style of structure is you do not play each time, which saves time. Csató (2021) discusses the current ranking issues seen in the Swiss-system. They use axioms to show the problems and how they effect who is the actual winner. Csató (2021) extended this process to robin tournaments as well. Dong et al. (2023) used integer programming to schedule a Swiss-system for the 2020 Honor of Kings World Champion Cup group stage. Another area that the Swiss-system tournament structure has a lot of research in is Chess. Ólafsson (1990) and Hartanto et al. (2016) use algorithms to create their pairwise comparisons. Their algorithm ranked teams in order to know who they should play. Csató (2013) used a pairwise comparison matrix to match the teams instead. A Swiss-system tournament purpose is to crown the best team the winner in an efficient way. However, not all tournaments are trying to crown the best winner. March Madness is so-named because people like to see upsets and Cinderella stories (Bender 2022; Lopresti 2022; Shen 2012; Anonymous 2022) so the idea of Swiss-tournament structures does not apply to all sports.

The third type of tournament is the knockout, a tournament structure where if you win, you advance to play the other winner until you are the last team left. This structure has variations, like double elimination, where you can lose twice before the team is eliminated. Both Hennessy and Glickman (2016) and Horen and Riezman (1985) focus on creating the best knockout tournament structure. They analyze the probability that the best team wins and use that as a factor in their design. This is the type of tournament structure for the NCAA March Madness tournament and what this paper uses. However, the format of the NCAA March Madness tournament is known, so this paper will not focus on how to design the structure but on what region and location to put each team in the tournament.

2.3 Location Problems

The location significantly impacts sports scheduling because of the traveling logistics and associated costs. Locations are either at the team's home location or a neutral site. The traveling tournament problem by Easton et al. (2001) uses integer and constraint programming to minimize the distance traveled while meeting the tournament restrictions. The Traveling Tournament Problem (TTP) is a double round robin, so each team plays each other home and away. The TTP has a similar objective to the objective in this paper to minimize distance, but in this problem, the locations are neutral sites. Melouk and Keskin (2012) focused on minimizing the distance with an integer program (IP) for the NCAA March Madness tournament. This strategy is similar to this paper that it is the same tournament, and the objective is to minimize the total distance traveled. However, this paper only utilizes the predetermined 8 locations given by the NCAA while this paper selects locations to minimize the total distance travel. Also, this paper wants to minimize the number of teams flying because of the high cost associated with flying.

Burke et al. (2004) used metaheuristics for neutral game sites to schedule a round robin tournament. This paper addresses a neutral site part but does not pick the neutral site locations to minimize distance. The neutral sites are known already. This paper picks the locations to minimize the number of teams flying. Currently no literature is found on picking the locations of game tournament. All literature is focused on already known locations. This is a different dynamic this paper focuses on.

2.4 NCAA March Madness

Most of the literature on the NCAA March Madness tournament focuses on prediction. There are two types of predictions. The first type predicts the teams to make the tournament, and the second deals with predicting the game's results. Fearnhead and Taylor (2010) solve this problem by concentrating on team strength of schedule and using historical data. Coleman and Lynch (2001) use probit analysis on team data to develop an equation to predict the teams selected for the committee. The second type of prediction is the outcome of the tournament. Kaggle's March Machine Learning Mania (Sonas et al. 2023) is a major event where people try to predict the results of March Madness through machine learning. As a result many machine learning programs are used to predict the tournament results. Kim et al. (2023) examined the use of different machine learning algorithms, including artificial neural network (ANN), k-nearest neighbors (kNN), support vector machine (SVM), logistic regression, and random forest (RF) that were used to predict the results of the NCAA March Madness tournament. Stekler and Klein (2012) employ statistical methods to predict games. Kaplan and Garstka (2001) use Markov models to predict winners. There are many prediction models, but this paper does not attempt to predict any results. This paper assumes that the tournament teams are already known and only focuses on the first game site locations for teams.

The closest NCAA March Madness tournament work to this paper comes from Melouk and Keskin (2012) who leverage integer programming to assign the tournament teams while minimizing the distance traveled. This paper uses this same concept but extends it to minimize the number of teams flying to their game site. This paper improves the formulation by changing some constraints, adding constraints, and modifying the objective. Also, they used the predetermined game sites from the NCAA, where this paper selects new game site locations to minimize distance. Smith et al. (2006) also attempt to minimize the distance traveled in the NCAA tournament with a mixed-integer program, utilizing the probability of tournament advancement to find pairings to lessen the total travel time. Their formulation does not follow the NCAA tournament's actual structure. For instance they fix the top 16 teams in the tournament to a specific a location. Also, they allow teams to move seed lines to improve the objective. This paper will keep all the tournament structures in place as how the NCAA tournament actually does them.

CHAPTER

3

METHODOLOGY

3.1 Approach

The NCAA March Madness Tournament starts with the four play-in games, and this paper assumes the winner of the game is known. Therefore the model accounts for 64 teams and treats the teams that lost in the play-in games as not a part of the tournament. There are two objectives in this paper. The first objective is to minimize the total distance traveled, and the second is to minimize the number of teams flying to their respective game site. The base model will minimize the distance traveled and the number of teams flying with some modifications. The model uses binary decision variables to decide which region and location each team will play. There are four regions in the NCAA tournament: the north, south, east, and west. The NCAA predetermines the eight-game site location before the season starts. The base model is the tournament's structure with all the constraints and assumptions to assign all the teams correctly.

3.2 Base Model

3.2.1 Assumptions

Similar to the assumptions made by Melouk and Keskin (2012), we assume

the 64 teams playing and their seeds are known,

the game sites are predetermined, and

the Number 1 seeds are preassigned to four separate locations closest to them, and

the distances between the team's home and game locations are known using the shortest path from the existing road network.

The shortest path calculation is done through the Matlog package (Kay 2023, 2016). This requires a distance matrix for all the teams to all possible locations. Matlog uses the ground distance multiplied by the circuity factor 1.2 to calculate the distance between two cities. For example, Matlog will calculate the team Illinois to all eight locations. Illinois is in Champaign so Champaign is used by Matlog to calculate the highway distance from Champaign to every location. This process is followed for every single team in the tournament.

Unlike Melouk and Keskin (2012), we do not assume each location has to have a different type of pods (e.g., two Number 1 seeds pods could not be at the same location). Their reasoning was to balance out the variety of games across the country. However, NCAA does not enforce this, as evidenced by Kansas and Baylor playing in the same game site location, Fort Worth, as Number 1 seeds in 2022.

The model only addresses the tournament's first two rounds. The reason is the first two rounds are at the same location for each team. Every other round uses different locations and requires predicting game results, which is outside the scope of this paper. The last assumption is that the s-curve is known and completed by the committee that selects and assigns all the teams in the NCAA March Madness tournament. This assumption is necessary to set the Number 1 seeds correctly.

3.2.2 Model Formulation

We formulate a binary integer program (BIP) to assign and schedule the 64 teams to their locations and regions.

Sets and indices

I Set of teams, $I = \{i : 1, \ldots, 64\}$

R Set of regions, $R = \{j : 1, \ldots, 4\}$

L Set of game sites (locations), $L = \{l : 1, \ldots, 8\}$

$l_1^*, \ldots, l_4^* \in L$ indicate game sites for the four Number 1 Seeds, respectively

C Set of conferences, $C = \{c : 1, \ldots, 32\}$

S Set of seeds, $S = \{s : 1, \ldots, 16\}$

T_s Set of teams with seed $s = 1, \ldots, 16$, for example,

$T_1 = \{1, \ldots, 4\}$ includes the four number one seeds;

$T_2 = \{5, \ldots, 8\}$ includes the four number two seeds; etc.

P Set of pods, $P = \{P_1, P_2, P_3, P_4\}$

P_1 Teams with seeds 1,8,9,16 that is $P_1 = T_1 \cup T_8 \cup T_9 \cup T_{16}$

P_2 Teams with seeds 2,7,10,15 that is $P_2 = T_2 \cup T_7 \cup T_{10} \cup T_{15}$

P_3 Teams with seeds 3,6,11,14 that is $P_3 = T_3 \cup T_6 \cup T_{11} \cup T_{14}$

P_4 Teams with seeds 4,5,12,13 that is $P_4 = T_4 \cup T_5 \cup T_{12} \cup T_{13}$

G_c Set of top four teams from conference $c \in C$

Parameters

D_{il} Distance between team location i and game site l, $i \in I$ and $l \in L$

K_{ic} K_{ic} =1 if team i is in conference c, and K_{ic} =0 otherwise, $i \in I$ and $c \in C$

Decision Variables

X_{ijl} X_{ijl} =1, if team i is assigned to region j at location l,

X_{ijl} =0, otherwise $i \in I$, $j \in R$, and $l \in L$

Y_{ij} Y_{ij} =1, if team i is assigned to region j

Y_{ij} =0, otherwise $i \in I$ and $j \in R$

Z_{il} Z_{il} =1, if team i is assigned to location l

Z_{il} =0, otherwise $i \in I$ and $l \in L$

Formulation

$$\text{minimize} \sum_{i \in I} \sum_{l \in L} D_{il} Z_{il} \tag{3.1}$$

Subject to:

Constraint 1: Team assignment

$$\sum_{j \in R} \sum_{l \in L} X_{ijl} = 1, \quad \forall i \in I$$

Constraint 2: Region assignment

$$\sum_{i \in I} \sum_{l \in L} X_{ijl} = 16, \quad \forall j \in R$$

Constraint 3: Site assignment

$$\sum_{i \in I} \sum_{j \in R} X_{ijl} = 8, \quad \forall l \in L$$

Constraint 4: Seed uniqueness

$$\sum_{l \in L} X_{ijl} = Y_{ij}, \quad \forall i \in I, \forall j \in R$$

$$\sum_{i \in T_s} Y_{ij} = 1, \quad \forall j \in R, \forall s \in S$$

Constraint 5: POD restrictions

$$\sum_{j \in R} X_{ijl} = Z_{il}, \quad \forall i \in I, \forall l \in L$$

$$\sum_{i \in T_s} Z_{ij} \leq 1, \quad \forall l \in L, \forall s \in S$$

$$\sum_{i \in P_k} X_{ijl} = 4 \sum_{i \in T_k} X_{ijl}, \quad \forall l \in L, \forall j \in J$$

Constraint 6: Conference Restrictions

$$\sum_{i \in P_k} K_{ic} = X_{ijl} \leq 1, \quad \forall k \in P, \forall c \in C, \forall j \in J, \forall l \in L$$

Constraint 7: Top 4 team restrictions

$$\sum_{i \in G_c} Y_{ij} \leq 1, \quad \forall j \in R, \forall c \in C$$

Constraint 8: Number 1 Seed location

$$\sum_{j \in J} X_{1jl_1^*} = 1,$$

$$\sum_{j \in J} X_{2jl_2^*} = 1,$$

$$\sum_{j \in J} X_{3jl_3^*} = 1,$$

$$\sum_{j \in J} X_{4jl_4^*} = 1,$$

Constraint 9: Binary constraint

$$X_{ijl}, \; Y_{ij}, \; Z_{il} \in \{0,1\}, \; \forall i \in I, \forall j \in J, \forall l \in L$$

The decision variables uniquely assign team i to region j at location l. The goal is to minimize the distance traveled between the game sites and the teams, which is accomplished by the objective function given in Equation 3.1. The NCAA March Madness tournament has specific rules and tournament structure guidelines that the model must follow while minimizing the objective function. The first three constraints set up the actual tournament structure, a 64-team tournament divided into four regions played at 8 locations. Constraint 1 assigns each team to one region and one location. Constraint 2 sets each region to exactly 16 teams. Constraint 3 sets each game site to 8 teams each. Constraint 4 sets seed uniqueness for each region. Each region has 16 teams and seeds from 1 to 16 in that region. Constraint 5 is what is called the POD structure in the model. pods enable the model to account for the NCAA's different restrictions and tournament structure. The most critical part of pods is that it forces the teams playing in the second round to play at the same game site. There are four different types of pods that consist of different seeds in each POD. Constraint 6 is a constraint set by the NCAA that states teams from the same conference cannot play each other in the tournament's first two rounds. What this means for this model is that teams from the same conference cannot be in the same POD as each other. To do this, constraint 6 limits the POD in a given region j and game site l to be teams from different conferences. Constraint 7 is an additional constraint set by the NCAA that states the top four teams from a conference cannot play in the same region as each other. Constraint 8 sets the Number 1 seeds to specific game site locations, which follows how the committee rewards teams for being the best four teams in the country. The order to slot Number 1 seeds comes from the committee's s-curve. Constraint 9 sets the decision variables to binary variables

The first change to Melouk and Keskin (2012) formulation is their home team constraint 8 is eliminated from this formulation. They assumed teams could not play in the city that they are from because that is an advantage. Now this assumption may be true, but it is not an actual restriction the NCAA committee formally has. Therefore, this paper allows a team to play in the same location they are from. Also, the formulation added a new constraint 8 that assigns all the number 1 seeds to a specified location. This follows the guidelines set by the NCAA March Madness Committee on how they assign teams to their location. This is also a reward to the number 1 seeds for their performance during the season. The other changes are updates to the sets and indices. The index for C (number of conferences) is updated from 31 to 32. This just an update to the NCAA adding more conferences. B_c set is updated from the top 3 to the top 4 in this paper. This set follows the rule that the top four teams from a conference cannot be in the same region as other teams. The NCAA changed the rules from the top 3 to the top 4.

CHAPTER

4

PENALTY METHOD

4.1 Approach

The base model in Chapter 3 minimizes the distance traveled from teams to their respective game site, saving transportation costs for both the NCAA and the fan base. However, the cost is significantly more when a team has to fly rather than drive. As mentioned earlier, the average cost of a short turnaround private jet is approximately $40,000 compared to the average $1,500 to charter bus a team (evoJets 2023). The base model's objective is to minimize distance but does not focus on getting teams below that 350-mile threshold to drive. Reducing the number of teams flying could save hundreds of thousands of dollars, so the model should focus on that. The second goal of this paper is to minimize the number of teams flying to their game. The NCAA has a specific policy for the NCAA March Madness tournament that if a team's game is farther than 350 miles, the team can fly to the game site (NCAA Staff 2023). Using this "threshold to fly" idea, this chapter introduces a penalty to the base model that motivates the model to avoid selecting a team to travel more than 350 miles if possible. The logic is limiting the number of teams having to fly utilizing this penalty method while still trying to minimize the distance traveled.

4.2 Formulation of Penalty Model

4.2.1 Assumptions

All assumptions and given information stated in the base model apply to the Penalty Model (see Section 3.2.1). We also assume there is no difference between teams flying to their games. All seeds are weighted equally in flying to their game site except for Number 1 seeds.

4.2.2 Penalty Model Formulation

Formulation. The formulation is the same as the base model with an additional decision variable, an added component to the objective function, and an added constraint. Below are the changes.

Parameters

τ — The "mileage threshold" or the minimum number of miles a team must travel to fly

N — A large value that penalizes the objective function in Equation (4.1)

M — A large value used in Constraint 10

Decision Variable Added

P_{il} — P_{il} =1, if team i is τ miles or more from location l,
P_{il} =0, otherwise for $i \in I$ and $l \in L$

Formulation Updates

Replace the objective function in Equation (3.1) with

$$\text{minimize} \sum_{i \in I} \sum_{l \in L} \left(D_{il} Z_{il} + N P_{il} \right) \tag{4.1}$$

and add a single constraint:

Added Constraint 10: Penalty Constraints

$$D_{il} Z_{il} \leq \tau + M P_{il}$$

The Penalty Model adds a decision variable that penalizes the objective function if the selected game site for that team is farther than τ miles. The NP_{il} term is added to the objective function in the equation to penalize the model for each team that has to travel farther than τ miles.

Constraint 10 is the added constraint the Penalty Model needs. It is a constraint that assigns the P_{il} a value of 1 if the distance the team has to travel to their game site is farther than τ miles due to the NCAA travel guidelines. The formulation is based off the big M technique.

The total number of teams flying is then given by $\sum_{i \in I} \sum_{l \in L} P_{il}$.

4.3 Analysis of Penalty Model

4.3.1 Penalty Value

After formulating the Penalty Model, the actual penalty value, N, needed to be determined. The penalty value must be big enough not to change the solution if parameters change in the model. To find the Penalty Model, a sensitivity analysis is required. Utilizing the 2022 season, penalty values were selected from 1,000 to 10,000. Only one season is utilized to get the magnitude of the penalty value needed. The number of teams flying in the optimal solution for each different penalty value is in Table 4.1. At a 5,000-mile penalty, the model converges on 34 penalties, and it does not matter how much larger the penalty is. To ensure a large enough penalty value, the model uses 10,000, doubling the 5,000 found. The reason not to go higher is that it makes the objective value significantly larger and less easy to comprehend.

Table 4.1: Penalty Value Analysis.

Penalty Value (N)	Number of Teams Flying
1000	36
2000	35
3000	35
4000	35
5000	34
10,000	34
100,000	34

4.3.2 Penalty Model Analysis

The first step in the penalty analysis is to use the penalty model for the 2022 season and 2019 season and comparing it to the base model and actual bracket. This paper did not use the 2020 or 2021 season because of COVID-19. NCAA canceled the 2020 tournament and decided to have the whole tournament in 2021 at one location. Comparing the results the Penalty Model reduced the number of teams flying from the base model by 18% and 12% for the 2022 and 2019 seasons respectively. That is rough savings of $320,000 and $240,000 in those years. The Penalty Model reduced the number of teams flying from the actual bracket by 27% and 21% for the 2022 and 2019 seasons. The Penalty Model successfully increases the number of teams driving to their game site. However, the overall distance traveled by every team did increase in the Penalty Model. Both Table 4.3 and Table 4.2 show that each model's average miles driven per team are roughly the same, so the flying miles increased for the Penalty Model. Figure 4.1 shows the different distance averages from the Base Model to the Penalty Model. The reason is that the actual travel distance is insignificant once a team is penalized for flying. That 10,000-mile penalty makes flying an extra 100 miles marginal. As a result, the Penalty Model makes the teams flying, fly farther, in order to allow other teams to drive. It does increase the cost to fly farther, but as seen earlier, it is far from the savings made by a team going from flying to driving. Increasing the mileage of the teams flying is not a significant time cost due to the speed of private jets. Overall, the Penalty Model worked in saving the NCAA money by reducing the number of teams flying.

Table 4.2: Penalty Model Results for 2019 Season.

Category	Base Model	Penalty Model	Actual Bracket
Distance	51,665	54,762	65,938
Number Teams Flying	50	44	56
Average Distance Per Teams Flying	979	1158	1145
Number Teams Driving	14	20	8
Average Distance Per Team Driving	191	189	221

Table 4.3: Penalty Model Results for 2022 Season.

Category	Base Model	Penalty Model	Actual Bracket
Distance	43,144	47,791	64,528
Number Teams Flying	44	36	49
Average Distance Per Teams Flying	876	1120	1245
Number Teams Driving	20	28	15
Average Distance Per Team Driving	230	230	233

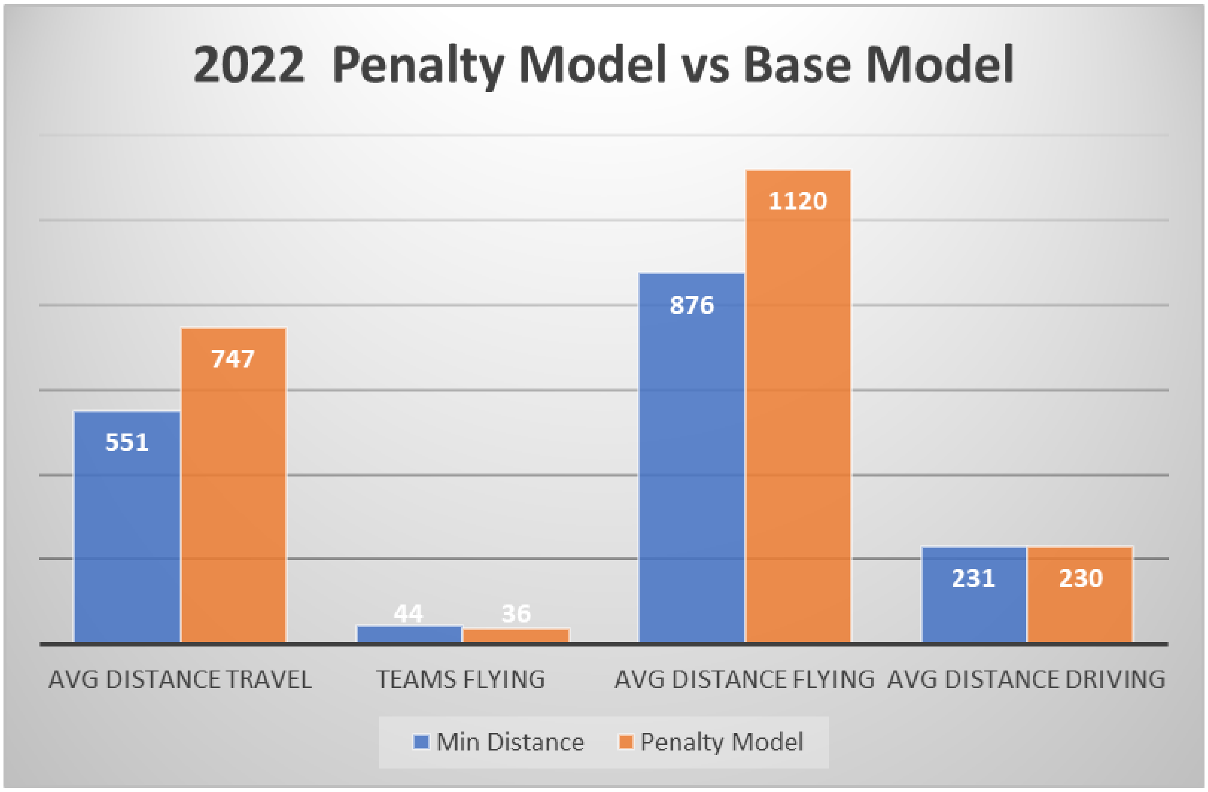

Figure 4.1: The overall distance traveled increased in Penalty model but that is due to flying distance which is not significant. Total number of teams flying is decreased in the Penalty Model.

4.4 Adjusting the 350 Mile Policy

After updating the base model with the penalty functions, we saw the number of teams flying decrease. However, the Penalty Model calculated these values based on the policy that teams farther than 350 miles had to fly. What if this changed? What if the new policy is 300 miles or 400 miles? Sensitivity analysis must be performed around the 350-mile threshold to answer these questions. Sensitivity analysis on the 300-600 mileage range was used because any shorter than 300 miles seemed unreasonable to fly to while driving over 600 miles would take more than one travel day. When the policy is changed to 300 miles, there is a 12% increase in the number of teams flying. If the NCAA makes the policy a shorter mileage policy, there will be a significant cost increase. Increasing the mileage policy by 50 miles only decreases the number of teams flying by 5%; while increasing the policy by 150 miles, there is almost a 28% decrease in the number of teams flying. That is a significant decrease if the NCAA increased the mileage policy. However, the NCAA March Madness tournament has special lower mileage policy than the other sports because of the traveling nature of the tournament. By increasing the policy the NCAA no longer abides in that. With a 50 mile increase only being a 5% increase it does not worth the NCAA increases the mileage policy. Increasing the policy by even 100 miles could result in teams driving two more hours. Figure 4.2 shows the different mileage policies.

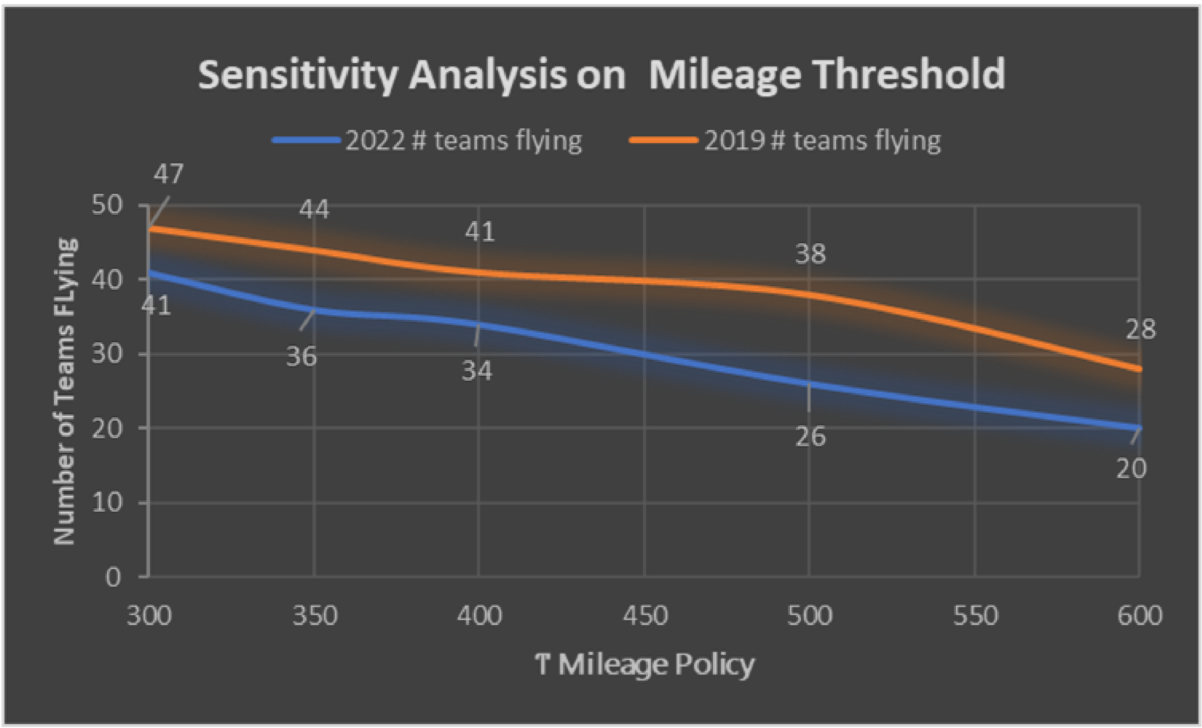

Figure 4.2: Policy Change.

In the different policies, we see the changes in the number of teams flying, but how close are other teams from that τ threshold? Figure 4.3 shows the 2022 season for three different τ values. Each histogram has a zero bin followed shortly after the τ value, which shows the Penalty Model does its best to make any team close to the threshold fly while making all the other teams fly farther. As a result, there are zero to a few teams above any given tau threshold. All the distances for 2019 and 2022 are in Appendix A.1

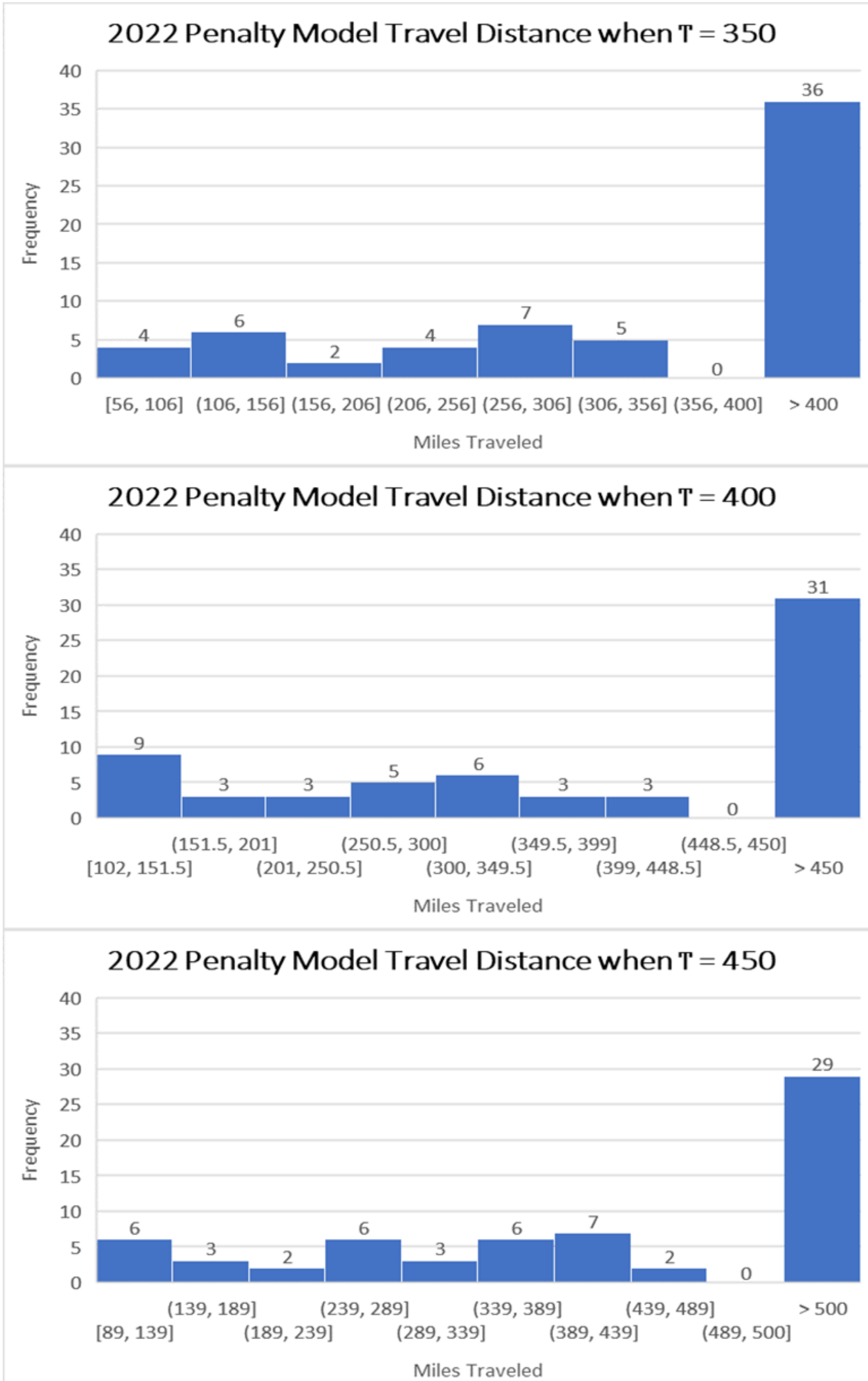

Figure 4.3: Histogram of the 2022 Season for all the Teams Traveling Distances for (top) $\tau = 350$, (middle) $\tau = 400$, and (bottom) $\tau = 450$.

The updated policy's objective function value decreased as the τ-mile threshold increased because fewer teams had to fly. With fewer teams flying, that 10,000-mile penalty decreases each time. Table 4.5 and Table 4.4 show the values. However, an interesting observation happened in the problem. When the penalty value is removed from the objective function value to get the actual total mileage traveled, it does not follow the same linear relationship as one would think. For example, the mileage policies for 350 and 400 miles increased the total distance from 47,791 to 49,359 miles while the 500-mile policy decreased to 43,750, the lowest among all policies. The reason it is not monotonic is because the τ-mileage threshold affects how many penalties the model can avoid. Sometimes the new τ threshold allows the model to reduce the number teams flying by a couple but it has to make other teams fly significant farther. As a result sometimes the total mileage traveled increases. Also, some of the different numbers is a result of the teams and locations that year for the tournament. The analysis used in the 2022 season was used for the 2019 season to validate these interesting observations. The same analysis for the 2019 season in Table 4.4 shows the same result. The objective function decreased, but the total travel distance did not decrease linearly. The total distance traveled increased from 54,535 to 54,762 from the 300 to 350 mileage policy. This helps concludes that the Penalty Model will make teams fly farther to reduce the total number of teams flying. By doing this the Penalty Model avoids the hefty penalty, but increases the total distance traveled. This makes sense, too, as the base model had a lower total distance traveled than the Penalty Model.

The final recommendation for the mileage policy is to change the policy to at least 450 miles, where the NCAA would save from $200,000 to $350,000. However, the NCAA has to weigh the savings with the knowledge of having teams drive that extra 100 miles.

Table 4.4: Mileage Policy Change 2019 Season.

τ	Teams Flying	Penalty Model Objective Function	Total Travel Distance
300	47	524,535	54,535
350	44	494,762	54,762
400	41	463,663	53,663
500	38	433,132	53,132
600	28	333,147	53,147

Table 4.5: Mileage Policy Change 2022 Season.

τ	Teams Flying	Penalty Model Objective Function	Total Travel Distance
300	41	456,457	44,645
350	36	407,791	47,791
400	34	389,359	49,359
500	26	303,750	43,750
600	20	245,095	45,095

CHAPTER

5

PREDETERMINED LOCATIONS

5.1 Approach

Another way to minimize the number of teams flying to their game site is to evaluate the actual game site locations. Models in Chapters 3–4 assume game site locations are known. The NCAA predetermines the locations before the year starts with site locations selected as far as four years (Staff 2023), but is there a way to pick better sites at the beginning of the year? The NCAA contracts with cities to host the games, so instead of trying to use every city in the United States or make specific criteria to choose cities, the model utilizes the last 60 cities that hosted NCAA March Madness games. The approach for this problem is to use a Maximum Covering Model to select the eight predetermined locations. We examine two different models for selecting eight game site locations that minimize the total number of teams flying to their game site location. The first model is a Maximum Covering Model formulation, while the second model tries to optimize the number of teams that have at least one game site location within the τ threshold. Both models are used to pick eight locations, and then those eight locations are used in the Penalty Model from Chapter 4 to see if they improve upon the solution from the NCAA selected sites.

5.2 Methodology of Maximum Covering Model

5.2.1 Assumptions

Assumptions and given information. The mileage policy in this problem is $\tau = 350$ miles. The other assumption is the NCAA continues to use the eight game site location tournament structure.

5.2.2 Maximum Covering Model Formulation

Formulation.

Sets and indices

I Set of teams, $I = \{i : 1, \ldots, e\}$
L Set of game site locations, $L = \{l : 1, \ldots, f\}$
P Set of game sites locations chosen, $P = \{l : X_l = 1\}$, $|P| = 8$

Parameters

a_{il} Distance between team location i and game site l, $i \in I$ and $l \in L$
τ The "mileage threshold" or the minimum number of miles a team must travel to fly

Decision Variables

x_l x_l =1, if location l is selected ,
x_l =0, otherwise, $l \in L$

Formulation

$$\text{maximize} \sum_{i \in I} \sum_{l \in L} a_{il} x_l \tag{5.1}$$

Subject to:

Constraint 1: Number of Game Sites

$$\sum_{l \in L} x_l = 8$$

Constraint 2: Binary Constraint

$$x_l \in \{0, 1\}, \forall l \in L$$

The Maximum Covering Model utilizes three sets—the first set I, which are the teams used to select the game site locations. The value e is the total number of teams used in this problem, which is 48. There are 48 teams rather than 64 (number of teams in the tournament) for this model because it will not try to predict the tournament but utilize a set of criteria that make the team "relevant." The list of teams are in Table B.2. The logic is that if the model can have locations within τ miles of all the relevant teams, this could give the best chance to minimize the number of teams flying in March. The goal is to beat the penalty model by picking better locations. Also, adding a non-relevant team that may not make the tournament could alter the model. Last, the teams seeded 13–16 (a total of 16 teams) are conference tournament champions from the small mid-major conferences, and it is almost a new team every year that wins the conference tournament. This means that it is usually new team every year and not reliable to use for choosing game site locations.

The criteria to determine if a team is relevant are the following:

- The team has at least 15 tournament appearances
- The team has appeared once in the tournament in the last four years
- The team has appeared at least four times in the tournament in the past ten years

The value f is 60 for this problem, which represents 60 potential game site locations the NCAA can use to host the tournament. The 60 game locations L in Table B.1 mentioned come from previously chosen sites from the NCAA. They are the last 60 game site locations for the NCAA tournament. p is the list of eight locations from the 60 total locations that will host the first two rounds of the NCAA tournament. The parameter a_{il} checks if the distance between any game site and team location is within that 350-mile distance. There are two decision variables in the formulation. The first decision variable X_l tracks which actual location the model chooses. The second decision variable is if a team location is covered by a game site location that is open. The objective function is to maximize the number of teams that have a site within 350 miles of their team. So out of the 48 teams, it is trying to get the most teams to have at least one site within 350 miles of their location. Constraint 1 limits the number of locations to eight. Constraint 2 is a coverage constraint that states for each team if a team has a game site location within 350 miles or not. Constraint 3 requires the decision variable to be binary.

5.2.3 Analysis of Maximum Covering Model

The Maximum Covering Model focuses on opportunity, which means maximizing the number of site locations within 350 miles of a team location. Two game site locations close to each other maybe be chosen because they are both within 350 miles of many teams. The model only cares if one game site already covers that team; the reason is to create as many opportunities for teams to have game site locations within that 350 miles. With the pod structure, a team is forced to play at the location the top team is assigned, so having more game sites within the τ creates more opportunities for the team to be within driving distance. The Maximum Covering Model chose cities in the Midwest and made a small circle around them. With most teams in a geographically smaller Midwest, the model is trying to give all those teams the most opportunities for game sites within 350 miles. Figure 5.1 displays the layout of the cities. The only problem just looking at the model is that many teams on the east coast, west coast, and south will not have a close location.

5.3 Methodology of Modified Maximum Covering Model

We modify the model from maximizing the number of locations within τ miles of the team game sites to maximizing the number of teams that have one game site location within τ miles.

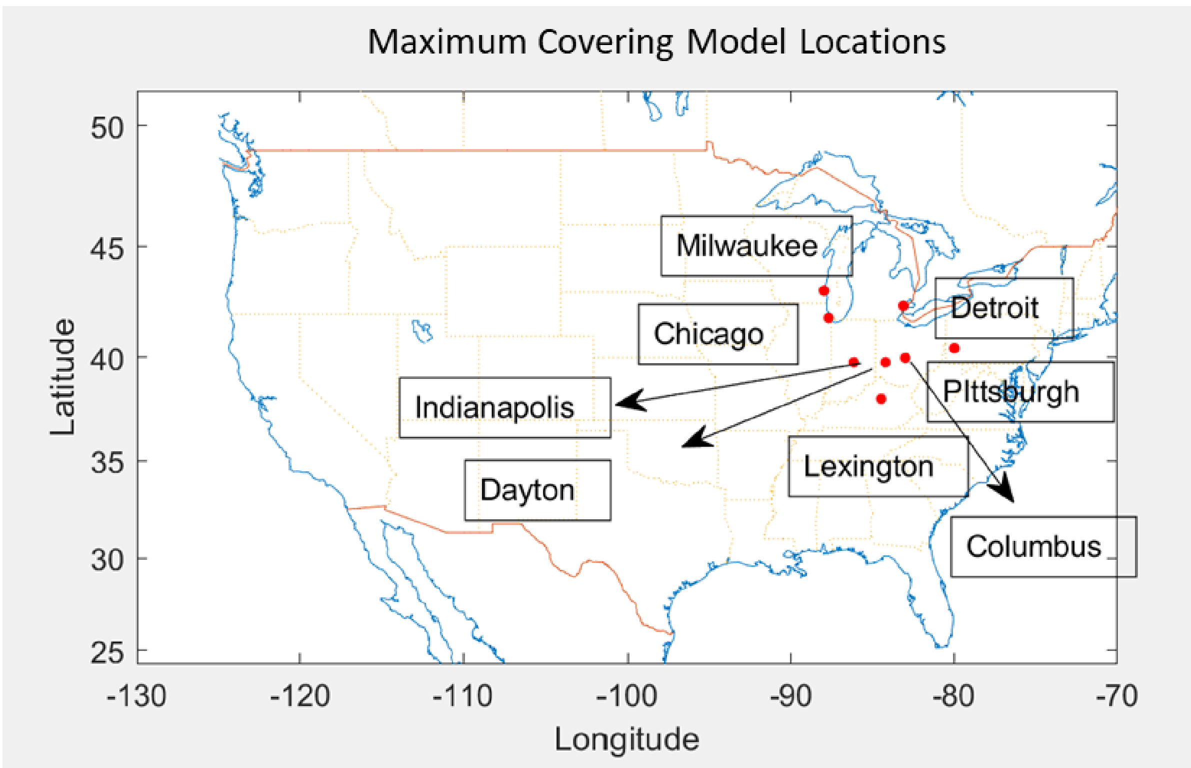

Figure 5.1: Maximum Covering Model Locations.

5.3.1 Assumptions

Assumptions and given information. The τ mileage policy for this problem is the 350-mile NCAA March Madness policy. Also, we will assume the same eight location tournament structure as the NCAA tournament.

5.3.2 Modified Maximum Covering Model Formulation

Formulation.

Sets and indices

I Set of teams, $I = \{i : 1, \ldots, 48\}$

L Set of game site locations, $L = \{l : 1, \ldots, 60\}$

P Set of game sites (locations), $P = \{l : 1, \ldots, 8\}$

Parameters

a_{il} Distance between team location i and game site l.
$a_{il} = 1$ if the distance $\leq \tau$ and 0, otherwise. $i \in I$, $l \in L$

τ The "mileage threshold" or the minimum number of miles a team must travel to fly

Decision Variables

X_l $X_l = 1$, if location l is chosen ,
$X_l = 0$, otherwise, $l \in L$

Z_i $Z_i = 1$, if team i is covered by a location,
$Z_i = 0$, otherwise, $i \in I$

Formulation

$$\text{maximize} \sum_{i \in I} Z_i \tag{5.2}$$

Subject to:

Constraint 1: Location Number

$$\sum_{l \in L} X_l = 8$$

Constraint 2: Coverage

$$Z_i \leq \sum_{l \in L} a_{il} X_l, \ \forall i \in I$$

Constraint 3: Binary

$$X_l, Z_i \in \{0, 1\}, \ \forall i \in I, \forall l \in L$$

The modified Maximum Covering Model is similar to the Maximum Covering Model formulation. It uses the same set, indices, and parameters while adding one decision variable. The decision variable X_l is the same but Z_i is needed for this model. The objective function changes to maximize the number of teams with a game site location within 350 miles of their location. Here a specific team having more than one location within 350 miles is not an advantage to the model. The constraints are the same, but *Constraint 2* from the Maximum Covering Model is added. These are the only differences to the modified Maximum Covering Model.

5.4 Analysis of modified Maximum Covering Model

The modified Maximum Covering Model picked eight-game site locations within 350 miles of the 44 team locations. The models covers 44 of the 48 possible teams, which is around (92%) of the total teams. Figure 5.2 maps the locations and where they are in the United States. The model chose two locations, Portland and San Diego, for the West Coast, with the rest on the eastern side of the United States. This layout makes sense because most teams are on the eastern side of the United States. The spread is throughout the United States in an attempt to have a location close to each team. The locations closely resemble

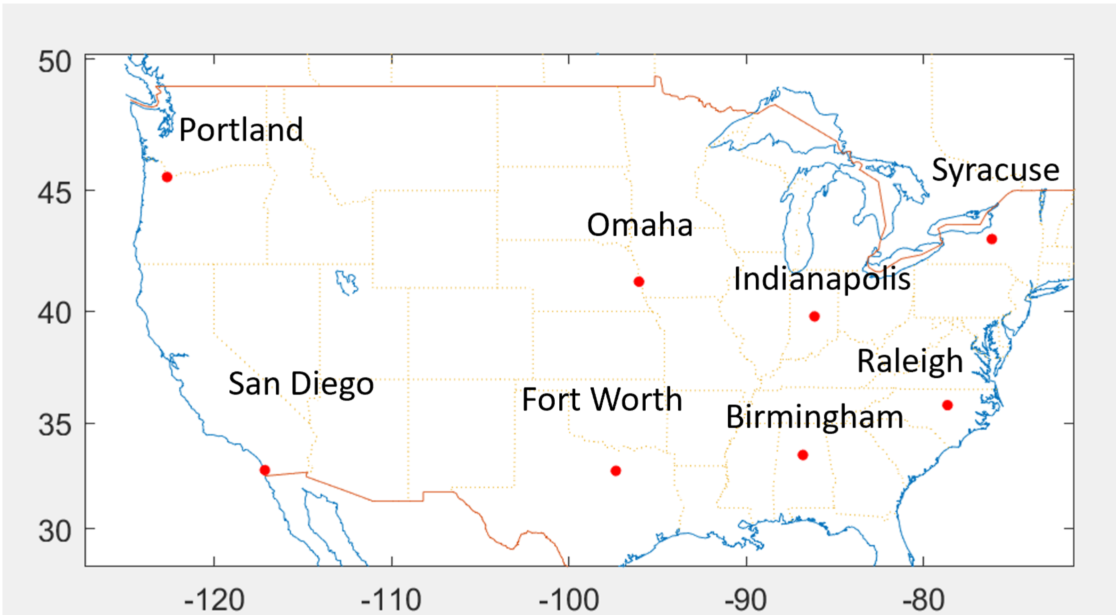

Figure 5.2: Maximum Covering Model Locations.

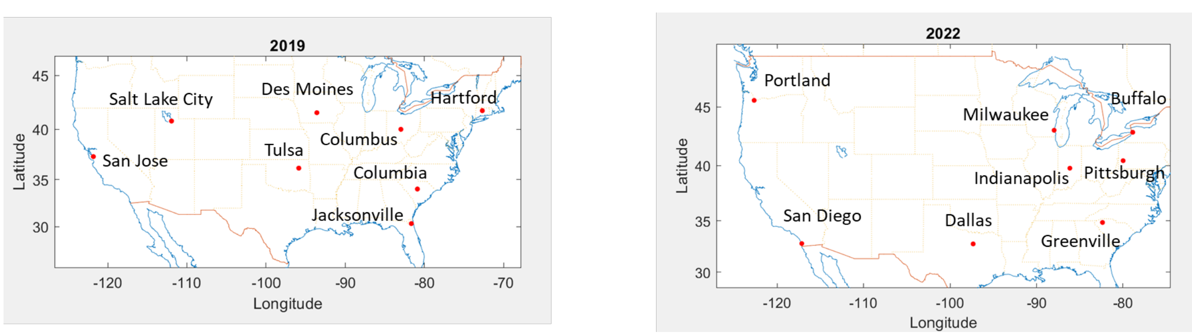

Figure 5.3: Actual game site locations from the NCAA in 2019 (left) and 2022 (right).

the spread of the NCAA's chosen locations. Figure 5.3 represents the actual locations the NCAA used for the March Madness tournament for the year 2019 and 2022. This demonstrates the NCAA does have a method to picking their locations and it is close to an optimal spread, but the analysis will show it is not the best selection of cities.

5.5 Result Comparisons

With these two additional models, there are now four models (Base model, Penalty model, Maximum Covering Model, and Modified Maximum Covering Model) to compare with the actual tournament logistics. The next step is to utilize the three models for the 2018, 2019, and 2022 seasons to see how they compare with the actual bracket regarding the total number of teams flying to their game site. The Penalty Model utilized the NCAA predetermined locations with its respective formulation. The Maximum Covering Model and Modified Maximum Covering Model utilize their determined locations and the Penalty Model formulation to calculate the total distance and number of teams flying. Figure 5.3 displays the results. The Maximum Covering Model has the most teams flying for all three seasons and is dominated by all the proposed models. The selected cities in the Maximum Covering Model are all in the Midwest or close by, which makes too many teams outside that region have to fly to their game. The modified Maximum Covering Model has the least total number of teams flying across all three years. The penalty model is a close second, with only three more teams flying. Comparing the actual results with the modified maximum coverage model, there is a significant difference in the number of teams flying. In 2018 there was a (30%) decrease in the total number of teams flying, a (32%) decrease in 2019, and a (26%) decrease in 2022.

If the NCAA used the modified Maximum Covering Model for their assignment of the NCAA March Madness tournament, they could save an average of $613,333 per tournament. The savings, again, is based on the assumption that the NCAA saves $40,000 per team that drives instead of flies. Figure 5.4 shows the savings for three different seasons based on how many fewer teams have to fly. The NCAA switching from a committee that uses pairwise comparisons to a model would save from $520,000 to $720,000 based on the historical data. The savings are significant yearly, and the model follows all the NCAA committee guidelines.

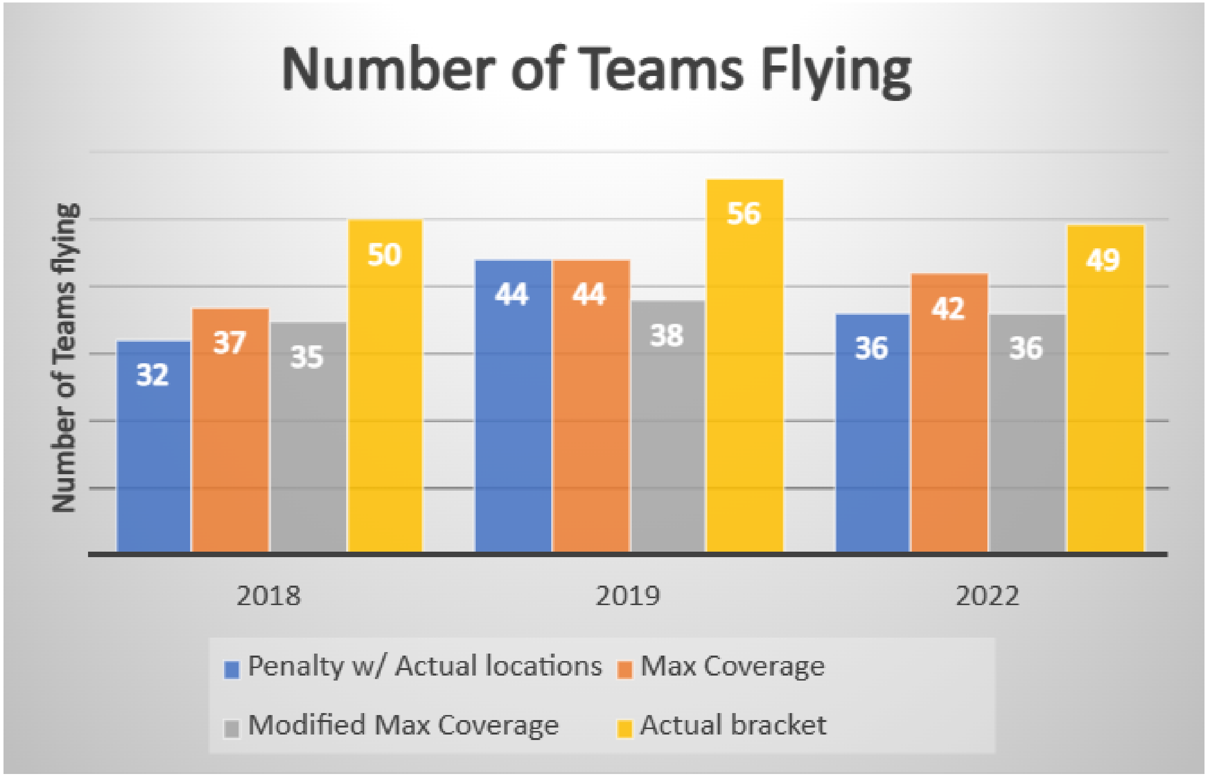

Figure 5.4: Teams Flying.

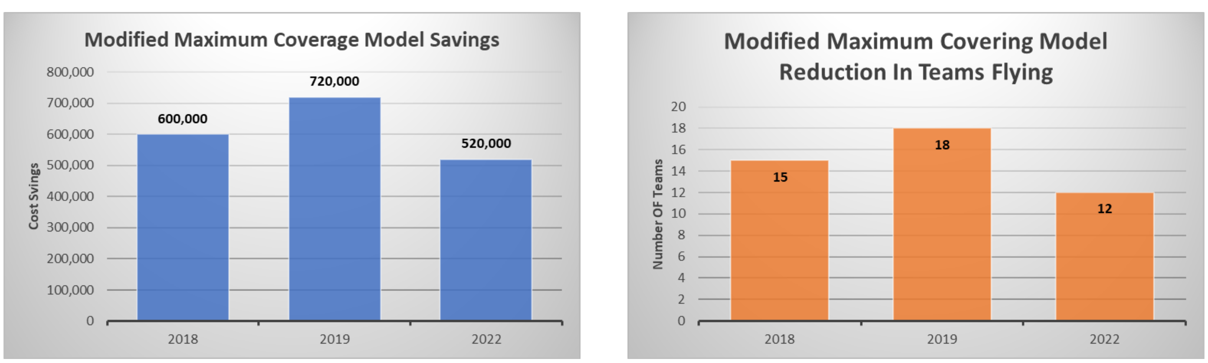

Figure 5.5: The Cost savings (left) and the reduction in number of teams flying (right) when the Modified Maximum Covering Model is used vs. the actual results.

After analyzing three seasons of data, we will test the models on the recent 2023 season to see if we obtain similar results. The results in Table 5.1 show identical results to our analysis. The Maximum Covering Model is dominated again by the Penalty and Modified Maximum Covering Model by having the most teams flying at 43. The Modified Maximum Covering Model has the lowest number of teams flying at 36 compared to the Penalty Model having 40. Also, using the Modified Maximum Covering Model instead of the actual bracket, there is an (11%) reduction in overall travel distance and (36%) fewer teams flying to their game site. The estimated savings is $800,000 in transportation costs. The results for 2023 support the analysis that using the Modified Maximum Covering Model saves at least $500,000 in transportation costs while reducing the number of teams flying by at least (25%).

Table 5.1: Results Comparison for 2023 Season.

Category	Penalty Model	Max Cov Model	Modified Max Cov Model	Actual
Distance	51,919	50,691	51,808	58,171
Number Teams Flying	40	43	36	56
Avg Dist Per Teams Flying	1177	1079	1212	1012
Avg Dist Per Team Driving	202	204	227	190

CHAPTER

6

CONCLUSIONS

6.1 Summary of Results

This paper aims to assign the teams in the NCAA March Madness tournament to reduce the number of teams that have to fly to their game site location while minimizing the total distance traveled. The paper utilized a Binary Integer Program to minimize the total distance while using a penalty to reduce the number of teams flying. We investigated the use of two Maximum Covering Models to select eight predetermined locations to also reduce the number of teams flying. The recommendation is to use the Penalty Model, which reduces the number of teams flying by 12% compared to the Base Model. The results showed the Maximum Covering Model to be dominated by the rest. However, the Modified Maximum Covering Model reduced the number of teams flying by 26–32% compared to actual recent tournament logistics. The recommendation is to use the Modified Maximum Covering Model to pick the eight locations and the Penalty Model to assign the teams to their game site location. By saving $40,000 per team that flies, a significant saving exists by using a model to assign teams to locations.

The NCAA could also look at changing its 350 mileage policy. If they increase the policy by 50 miles to 400, there is only around a 5% decrease in the number of teams flying, whereas an increased policy to 450 miles decreases the number of teams flying by around 16%. The extra mileage is not worth NCAA considering. The NCAA has a shorter mileage policy for the NCAA March Madness tournament because of all the potential traveling a team may have to do. Increasing more than 50 miles adds a significant amount of driving time while not resulting in significant savings. Therefore, the recommendation is to keep the policy as is.

6.2 Future Work

There are many different avenues for future work in this problem. The first is exploring a different objective in this paper. This paper focused on the number of teams flying, but future work could focus on cost. Cities have a different costs to host the games and different potential earnings levels. The logistics cost vs the cost of having the tournament in different cities could be considered. For example, does having the games in Las Vegas make significantly more money than in San Diego? The impact of the overall cost would determine where to put the game site locations. Also, more analysis on the cost impacts if the NCAA selected sixteen locations rather than eight. Sixteen locations keep the teams in the same location for the first two rounds and would reduce travel time. However, further analysis is needed to determine if the cost is worth expanding to more locations.

Another idea is using a probability model to help forecast the teams in the tournament instead of using historical data. Those teams then can be used to help select the eight game site locations. This paper used historical data to help predict possible NCAA tournament teams where a probability model could help prove a better set of teams. Another area is looking at all the locations chosen in the NCAA tournament. This paper focused on the first two rounds because they are at the same locations and have the most teams. However, future rounds have different locations, and location selection could matter. Probability models help predict the best locations to use in future rounds. There are a lot of different ideas out there to expand on the current problem.

The last future work idea is adding a time zone constraint to the model. Mentioned earlier in Chapter 1 the NCAA March Madness Committee cares about how many time zones a team travels. There are currently no specified restriction in the number of time zones a team can travel. However, a constraint that has a restriction like a team cannot travel three or more time zones can easily be added to the model. This additional constraint could enforce the Committee's principles in a quantitative way.

REFERENCES

Anonymous (2022). 4 Reasons Why NCAA Tournament is the Best. CollegeInsider.com. Accessed 6 March 2023, https://collegeinsider.com/4-reasons-why-ncaa-tournament-in-the-best.

Associate, A. G. (2019). 2019 March Madness Betting Estimates. https://www.americangaming.org/resources/2019-march-madness-betting-estimates/.

Bender, B. (2022). Top 20 most shocking first-round NCAA Tournament upsets. sportingnews.com. Accessed 6 March 2023, https://www.sportingnews.com/us/ncaa-basketball/news/ncaa-tournament-march-madness-top-20-biggest-first-round-upsets-2-15-seeds-richmond-cleveland-state-florida-gulf-coast/86s1szj9le8x1jhd2i14jr3od.

Burke, E., De Werra, D., Silva, J. L., and Raess, C. (2004). Applying heuristic methods to schedule sports competitions on multiple venues. In *Practice and Theory of Automated Timetabling V*, volume 5, pages 451–456, Pittsburgh, PA. 5th International Conference, PATAT 2004, August 18–20, http://www.cs.nott.ac.uk/~pszjds/research/files/jdls_patat2004_1.pdf.

Coleman, B. J. and Lynch, A. K. (2001). Identifying the NCAA tournament "dance card". *Interfaces*, 31(3):76–86.

Csató, L. (2013). Ranking by pairwise comparisons for swiss-system tournaments. *Central European Journal of Operations Research*, 21:783–803.

Csató, L. (2021). *Tournament design: How operations research can improve sports rules*. Springer Nature, Cham, Switzerland. https://doi.org/10.1007/978-3-030-59844-0.

Dong, Z.-L., Ribeiro, C. C., Xu, F., Zamora, A., Ma, Y., and Jing, K. (2023). Dynamic scheduling of e-sports tournaments. *Transportation Research Part E: Logistics and Transportation Review*, 169. https://doi.org/10.1016/j.tre.2022.102988.

Durán, G. (2021). Sports scheduling and other topics in sports analytics: a survey with special reference to Latin America. *Top*, 29(1):125–155.

Easton, K., Nemhauser, G., and Trick, M. (2001). The traveling tournament problem description and benchmarks. In *Principles and Practice of Constraint Programming—CP 2001: 7th International Conference, CP 2001 Paphos, Cyprus, November 26–December 1, 2001 Proceedings 7*, pages 580–584. Springer.

evoJets (2023). How much does it cost for private jets. https://www.evojets.com/charter-flight-costs-pricing basics.

Fearnhead, P. and Taylor, B. M. (2010). Calculating strength of schedule, and choosing teams for March Madness. *The American Statistician*, 64(2):108–115.

Grantland Staff (2013). 30 for 30 Shorts: 'The Schedule Makers'). https://grantland.com/features/our-latest-30-30-documentary-short-tells-story-how-husband-wife-ended-creating-schedule-mlb/.

Hartanto, J., Arifin, Y., and Hanafiah, N. (2016). Xiang qi tournament algorithm applying swiss pairing system. In *2016 International Conference on Information Management and Technology (ICIMTech)*, pages 253–256. IEEE.

Hennessy, J. and Glickman, M. (2016). Bayesian optimal design of fixed knockout tournament brackets. *Journal of Quantitative Analysis in Sports*, 12(1):1–15.

Horen, J. and Riezman, R. (1985). Comparing draws for single elimination tournaments. *Operations Research*, 33(2):249–262.

Kaplan, E. H. and Garstka, S. J. (2001). March Madness and the Office Pool. *Management Science*, 47(3):369–382. https://www.jstor.org/stable/2661505.

Kay, M. G. (2016). Matlog: Logistics Engineering Using MATLAB. *Journal of Engineering Sciences and Design*, 4(1):15–20.

Kay, M. G. (2023). Matlog: Logistics Engineering MATLAB Toolbox. https://mgkay.github.io/matlog/.

Kim, J., Magnusen, M., and Jeong, S. (2023). March Madness prediction: Different machine learning approaches with non-box score statistics. *Managerial and Decision Economics*. https://doi.org/10.1002/mde.3814.

Lopresti, M. (2022). We're tracking every upset in the NCAA men's tournament. NCAA.com. Accessed 6 March 2023, https://www.ncaa.com/live-updates/basketball-men/d1/2022-mens-march-madness-upset-tracker.

Melouk, S. H. and Keskin, B. B. (2012). Team assignments and scheduling for the NCAA basketball tournament. *The Journal of the Operational Research Society*, 63(5):620–630.

NCAA Staff (2022). 2022 DI men's basketball championship game sets single-game viewing records. https://www.ncaa.com/news/basketball-men/article/2022-04-05/2022-di-mens-basketball-championship-game-sets-single-game-viewing-records.

NCAA Staff (2023). Division I Championships Transportation and Per Diem Policies. https://ncaaorg.s3.amazonaws.com/championships/resources/travel/2022-23D1Champs_TravelPolicies.pdf.

ncaa.com (2021). The NCAA bracket S-curve, explained). https://www.ncaa.com/news/basketball-men/article/2021-03-14/ncaa-bracket-s-curve-explaineds.

Ólafsson, S. (1990). Weighted matching in chess tournaments. *Journal of the Operational Research Society*, 41(1):17–24.

Palm, J. (2016). Bracketology: Geography and balancing the bracket. https://www.cbssports.com/college-basketball/news/bracketology-iowa-state-falls-in-bracket-now-on-ncaa-tournament-bubble-after-loss-to-west-virginia/.

Sager, R. (2022). 2022 NCAA bracket: Scores, stats, records for March Madness men's tournament). https://www.ncaa.com/news/basketball-men/article/2022-07-12/2022-ncaa-bracket-mens-march-madness-scores-stats-records.

Shen, J. (2012). 10 Reasons March Madness Is the Best Postseason in America. BleacherReport.com. Accessed 6 March 2023, https://bleacherreport.com/articles/1118941-10-reasons-march-madness-is-the-best-postseason-in-america.

Smith, J., Fraticelli, B. M., and Rainwater, C. (2006). A bracket assignment problem for the National Collegiate Athletic Association Men's Basketball Tournament. *International Transactions in Operational Research*, 13(3):253–271. https://doi.org/10.1111/j.1475-3995.2006.00547.x.

Sonas, J., Maggie, and Cukierski, W. (2023). March Machine Learning Mania 2023. https://kaggle.com/competitions/march-machine-learning-mania-2023.

Staff, N. (2023). Future dates and sites. https://www.ncaa.com/mens-final-four/future-info.

Stekler, H. O. and Klein, A. (2012). Predicting the outcomes of NCAA basketball championship games. *Journal of Quantitative Analysis in Sports*, 8(1). https://doi.org/10.1515/1559-0410.1373.

Sziklai, B. R., Biró, P., and Csató, L. (2022). The efficacy of tournament designs. *Computers & Operations Research*, 144. https://doi.org/10.1016/j.cor.2022.105821.

APPENDICES

APPENDIX

A

TRAVELING DISTANCES

Table A.1: List of all the Traveling Distances by year for each τ.

Team Travel Distances for 2019 and 2022						
τ	350		400		450	
year	2019	2022	2019	2022	2019	2022
	0	56	0	88	41	89
	41	102	41	102	50	102
	50	103	50	103	73	104
	103	104	73	104	114	112
	114	110	114	110	123	133
	123	112	123	112	126	133
	126	117	126	117	153	142
	186	133	197	133	211	180
	197	155	211	138	219	194
	211	194	219	158	240	196
	219	201	224	194	251	247

	224	213	240	201	258	248
	240	237	258	213	331	267
	258	247	324	237	339	267
	262	248	331	248	351	284
	285	275	339	275	367	285
	324	284	351	284	370	295
	330	285	367	285	378	302
	331	290	370	290	382	335
	351	295	370	295	401	346
	402	301	378	301	402	349
	410	302	382	302	405	356
	478	335	402	335	410	362
	532	346	478	346	426	373
	533	346	532	349	478	376
	569	349	533	349	507	411
	592	349	537	356	508	414
	612	411	569	378	519	414
	662	417	612	395	532	416
	681	427	635	411	537	427
	703	495	681	416	566	429
	735	522	703	427	569	437
	787	525	735	522	592	447
	799	536	787	536	612	452
	809	553	800	546	661	522
	875	564	809	553	681	536
	891	570	875	564	787	564
	893	587	891	570	800	570
	912	671	893	587	875	587
	926	708	926	659	891	627
	938	711	940	671	940	661
	1018	723	962	708	1015	671
	1019	734	1015	723	1019	708
	1024	823	1018	728	1024	723

	1050	841	1019	734	1045	752
	1082	853	1024	823	1050	767
	1114	878	1045	878	1114	778
	1126	887	1050	886	1126	853
	1152	962	1114	933	1152	931
	1191	978	1126	962	1158	962
	1219	1015	1152	1066	1182	1015
	1484	1110	1395	1139	1254	1066
	1583	1139	1583	1215	1544	1137
	1755	1215	1755	1376	1755	1139
	1831	1376	1831	1618	1765	1376
	1879	1618	1879	1626	1831	1399
	2038	1678	1921	1647	1921	1618
	2176	1750	2038	1826	2091	1647
	2214	1956	2214	1956	2214	1956
	2308	1965	2308	2253	2308	2266
	2434	2266	2434	2266	2522	2315
	2522	2596	2522	2596	2677	2596
	2677	2803	2677	2803	2706	2803

APPENDIX

B

INPUT DATA FOR LOCATION MODELS

Table B.1: List of Possible Game Site Locations L.

Index	City	State
1	Birmingham	AL
2	Des Moines	IA
3	Orlando	FL
4	Sacramento	CA
5	Albany	NY
6	Columbus	OH
7	Denver	CO
8	Greensboro	NC
9	Las Vegas	NV
10	New York	NY
11	Kansas City	MO
12	Louisville	KY
13	Houston	TX

14	Buffalo	NY
15	Indianapolis	IN
16	Fort Worth	TX
17	Portland	OR
18	Greenville	SC
19	Milwaukee	WI
20	Pittsburgh	PA
21	San Diego	CA
22	Columbia	SC
23	San Jose	CA
24	Jacksonville	FL
25	Salt Lake City	UT
26	Hartford	CT
27	Tulsa	OK
28	Wichita	KS
29	Dallas	TX
30	Boise	ID
31	Charlotte	NC
32	Detroit	MI
33	Nashville	TN
34	Los Angeles	CA
35	Atlanta	GA
36	Boston	MA
37	Omaha	NE
38	Memphis	TN
39	Glendale	AZ
40	Providence	RI
41	Raleigh	NC
42	St. Louis	MO
43	Oklahoma City	OK
44	Spokane	WA
45	Anaheim	CA
46	Philadelphia	PA

47	Chicago	IL
48	Seattle	WA
49	Cleveland	OH
50	Syracuse	NY
51	San Antonio	TX
52	Auburn Hills	MI
53	Lexington	KY
54	Dayton	OH
55	Washington	DC
56	Albuquerque	NM
57	Tucson	AZ
58	Tampa	FL
59	New Orleans	LA
60	Newark	NJ

Table B.2: List of Teams used to choose locations.

Index	Team
1	Kentucky
2	North Carolina
3	Kansas
4	UCLA
5	Duke
6	Lousiville
7	Syracuse
8	Indiana
9	Villanova
10	Notre Dame
11	Texas
12	Michigan State
13	Arizona
14	Arkansas
15	Marquette
16	UConn

17	Cincinnati
18	Oklahoma State
19	Illinois
20	Purdue
21	Kansas State
22	Georgetown
23	Ohio State
24	BYU
25	West Viriginia
26	Oklahoma State
27	Iowa
28	Maryland
29	Michigan
30	Missouri
31	Wisconsin
32	Gonzaga
33	Tennesse
34	Virginia
35	Creighton
36	Houston
37	Alabama
38	Florida
39	Iowa State
40	Memphis
41	Providence
42	USC
43	Florida State
44	Texas Tech
45	VCU
46	Oregon
47	Baylor
48	Saint Mary's