1. Introduction and Motivation

When brainstorming ideas for potential final projects, we tried to look at things that linked us together, either through culture or common interest. Through discussion over time, we realised that we all had a common interest in the sport of cricket. We found ourselves talking a lot about the evolution of cricket as a game over time, from a slow, boring format that took place over five days to two hour games with explosiveness that is entertaining to watch. The Indian Premier League is a case in point. It is an Indian cricket league that specialises in the shortest time format of the game where teams can build rosters of players from all over the world. The IPL has seen tremendous growth since its inception in 2008, with a brand value of $6.8B in 2019.

As we discussed the league more, we realised that it was far from perfect. The key problem that we saw was how the IPL scheduled matches. At many points in the past couple of years, the scheduling of the tournament has led to teams having large disparities between their total number of games played. As such, a few teams can gain asymmetric information on their competitors, thus raising the risk of unfair competition.

We quickly realised that using data to optimize the IPL's scheduling system through our learnings in the class would be a great way to explore a topic we were collectively passionate about and interested in. This topic also had much greater potential in that we could incorporate even more layers of complexity such as the revenue per match and the pricing of tickets to make our initial scheduling problem more interesting.

2. Initial Scheduling Model

Overview and Potential Challenges

The IPL can be best understood as a round-robin tournament with 8 teams, where each team plays exactly 2 games with every other team, resulting in 56 games that are spread over a maximum of 56 days. Our scheduling model was designed to consider two main objectives - to minimize imbalance and thus, unfairness from the tournament and to further factor in player fatigue, viewership limits, among other constraints.

Model

Our decision variables took the format:

$$X_{i,j,k} = 1 \text{ } if\ team\ i\ plays\ team\ j\ on\ day\ k\ with\ i\ being\ at\ home$$

Implementing this for the 8 team tournament resulted in a total of 4096 decision variables, rendering the problem infeasible. Thus, we decided to simplify our tournament to consider 4 teams under the same arrangement, which resulted in a total of 192 binary decision variables, 16 for each day. The decision variables for Day 1 are reflected in Figure 2.1

	Day 1					
	Away					
Home		MI	CSK	RCB	DC	Total
	MI	0	0	0	0	0
	CSK	0	0	0	0	0
	RCB	0	0	0	0	0
	DC	0	0	0	0	0
	Total	0	0	0	0	

Figure 2.1: Decision variables for Day 1 Matches

The first 4 constraints are reflected in the following table. For ease of notation, let's assume that $X_{i,k}$ is the total number of games played by team i on day k, and further let $C_{i,k}$ be the total number of games played by team i until day k

Table 2.1: Constraint descriptions for scheduling problem

Constraint	Description	Mathematical Representation
Limit per day	No team plays more than one game on a given day. Note that this further ensures that $X_{i,i,k} = 0$ for all i and k since a team playing with itself would count as 2 games based on how we computed total games for each team by summing the rows and columns including the cell representing $X_{i,i,k}$ in both sums	$\forall k,\ X_{i,k} = 1$
Consecutive games	No team plays 3 matches consecutively	$\forall i,\ \forall k \in [1,\ 10]\ X_{i,k} + X_{i,k+1} + X_{i,k+2} \leq 2$
Home & Away	Each team plays 1 home and 1 away game with every other team	$\forall i,\ \forall j \neq i,\ \sum_{k=1}^{12} X_{i,j,k} = 1$
Day	No more than 2 matches are played in a given day	$\forall k,\ \sum_{i=1}^{4} \sum_{j=1}^{4} X_{i,j,k} \leq 2$
Schedule Imbalance	Maximum difference in total games played on a given day is 1	$\forall k,\ \forall i,\ MAX(C_{i,k}) - MIN(C_{i,k}) \leq 1$

However, determining maximum and minimum cumulative game counts for each team on a given day introduced non-linearity into the model due to the MAX and MIN functions. Thus, we decided to add 24 additional decision variables, with 12 representing maximum cumulative game counts on each day and the other 12 representing minimum cumulative game counts on each day.

To ensure that these variables represent the respective maximum and minimum cumulative counts, we added the following constraints:

$$\text{Let } MAX_k \text{ and } MIN_k \text{ be decision variables representing maximum and minimum cumulative game count on day } k$$

$$\forall k,\ \forall i,\ MAX_k \geq C_{i,k}\ ;\ \forall k,\ \forall i,\ MIN_k \leq C_{i,k}\ ;\ MAX_k - MIN_k \leq 1$$

The decision variables and constraints for days 1-6 are reflected below in Figure 2.2.

Team	Cumulative games played by day					
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
MI	0	0	0	0	0	0
CSK	0	0	0	0	0	0
RCB	0	0	0	0	0	0
DC	0	0	0	0	0	0
Maximum cumulative count	0	0	0	0	0	0
Mininimum cumulative count	0	0	0	0	0	0
Maximum Difference	0	0	0	0	0	0
Is less than	<=	<=	<=	<=	<=	<=
One	1	1	1	1	1	1

Figure 2.2: Decision variable & Constraint formatting to handle imbalance for Days 1-6

The primary objective of the model was to solve the scheduling. Thus, we decided to use a proxy objective function which was the sum of all 192 binary decision variables. This would yield a feasible schedule given that by the nature of the 3rd constraint, the total number of games for each team combination is fixed.

Ultimately, we got a model which was larger than the capacity of a traditional solver. Thus, we utilized open solver to obtain a solution.

Overview of Results

Ultimately, the model yielded a feasible schedule for the IPL, reflected in Figure 2.3.

Day	Match 1		Match 2	
	Home	Away	Home	Away
1				
2				
3				
4	DC	MI	CSK	RCB
5	CSK	MI		
6	DC	RCB		
7				
8	MI	RCB	CSK	DC
9	DC	CSK	RCB	MI
10				
11	MI	CSK	RCB	DC
12	MI	DC	RCB	CSK

Figure 2.3: Schedule generated through discrete optimisation

Despite this, the model suffered from a few limitations. Firstly, we simplified the tournament to 4 teams instead of 8. Additionally, the lack of an appropriate objective function hindered our ability to consider projected revenues as a part of scheduling. Going forward, we decided to address the second limitation as the next step since despite the smaller optimisation, the model did yield a reasonable blueprint for scheduling.

3. Revenue Forecasting

Overview and Potential Challenges

The scheduling attempt previously was just an initial step to understand if the optimization scheduling was possible. It used the bare minimum constraints that the league requires to operate and ignored a huge factor that influences the IPL - money. In order to delve into revenue maximization it was important to find the revenues for each combination of IPL games. Due to the lack of data, we decided that the best way to approach the problem would be to simulate the revenues earned for each game to accompany our analysis. These revenues found would end up being the coefficients for each game variable in the objective function.

Data & Assumptions

We found data on the stadium capacities along with prices for 8 different ticket classes. Since we couldn't find data on the number of seats allocated to each ticket class, we used data for 1 stadium and determined ratios of seats allocated to each ticket class to the total stadium capacity and used that as an approximation for seat distribution by section for all stadiums. We utilised online sources to determine that the average demand of the IPL game to be 35000 (refer to the bibliography). We further assumed variance to be 2500^2 for an IPL game.

Below is the data found on the stadiums and relevant capacities along with the proportion of seats dedicated to each seating segment. The price for a ticket in each section is shown below.

Stadium Capacities	
Stadium	Capacity
Wankhede Stadium (Mumbai)	33108
Chinnaswamy Stadium (Bangalore)	35000
MA Chidambaram Stadium (Chennai)	50000
Feroz Shah Kotla Stadium (Delhi)	41820

Ticketing Information		
Class	Price (₹)	Proportion
Block C1,D1,F1,G1,H1,K1.	400	0.18
Block B1,D,E,F1,G,H,J,L1	500	0.41
Block F	900	0.02
Block C & K	1000	0.17
Block L	1800	0.09
Block B	2100	0.09
Block CLUBHOUSE UPPER	3000	0.03
Block CLUBHOUSE LOWER	9000	0.02

Figure 3.1: Stadium capacities and ticket prices and proportions by class

We further assumed distributions for ticket demands and rainfall,which are reflected below.

Table 3.1: Distribution assumptions for revenue forecasting model

Assumption Name	Description	Assumption Statement	Source/Reasoning
Demand For Tickets Y_{ijk}	Y_{ijk} is the demand for tickets of match team i (home) versus j (away) in stadium seating section k.	$Y_{ijk} \sim N(\mu, \sigma^2)$	μ is proportion of tickets of class k multiplied by the average demand (35000) and σ^2 is computed to be the proportion squared times the overall variance (2500^2) since the demand for each class is effectively a linear function of the overall demand
Demand For Box Seat Tickets B_{ij}	Every game has a set number of premium box Suites that can be purchased for exorbitant prices.	Triangular(a,b, c)	Each stadium has a different number of box suites. Historically, they tend to be half full with rare events of being empty. Hence this is a triangular distribution where a is 0, b is half the stadium's box seats and c is 75% of the total number of box seats.
Rainfall Distribution R_{ij}	Every Game has a chance of being affected by rain which affects refund policy.	Gamma(α, β)	Based on the research paper we analysed, we determined that rainfall can best be modelled as a gamma distribution. We took the sample mean and variances from a research paper (appendix) and manipulated it into alpha and beta (alpha = E[X]^2/Var[X] and beta = E[X]/Var[X])

Model

Below is the logic behind the model for every single game that is to be played:

1. For any game, for each seating class, **demand** is forecasted using a normal distribution that is explained in Assumption 1.
2. Based on demand for each ticket class, **sales** for that ticket class are calculated using the maximum between demand and capacity.

$$Min\ (Early\ Demand,\ Stadium\ Capacity\ *\ Booking\ Limit\ Ratio\ *\ Proportion\ of\ Class\ (Upper/Middle/Lower))$$

3. **Revenue** is determined by the price of tickets multiplied by sales

$$Revenue\ =\ SUMPRODUCT(Prices\ for\ Different\ Classes,\ Simulated\ Sales\ for\ Different\ Classes)$$

4. Based on the rainfall distribution in assumption 3, **loss of revenue** for the match is determined by if the total rainfall is greater than given thresholds. If the Rainfall for a game is greater than 8, the match is delayed and a portion of revenues are refunded. Distributions for rainfall were derived from data regarding the different alpha and beta parameters for each city.
5. This is the **Operating Revenue** for the game (Forecasted Value)

For all the games that are played, each of their operating profits are calculated using the logic above. This fulfills the previous problem of being unable to find revenue data for the season.

Findings & Recommendations

Based on the Simulation, we were able to find the following matrix as the revenues for each combination of game to be played:

		Away			
	Team	MI	CSK	RCB	DC
Home	MI	0.00	22973186.44	24609297.31	24593874.39
	CSK	31782304.00	0.00	31705964.00	31630602.90
	RCB	33125732.96	33182556.07	0.00	33212654.27
	DC	31712742.13	31825496.74	31752325.81	0.00

Figure 3.2: Simulated revenues for each game

Now, using these values as objective coefficients, we have an understanding of which games at which stadiums pull in the most revenue. With the original scheduling model, the objective function of total operating revenue sums to INR 362,106,737.02

Revamping Schedule by Removing Constraint

Recently, due to COVID the IPL has discussed dropping the 1 home and 1 away constraint for every team. For example, for the MI vs. CSK game, the removed constraint would allow both games to be played in Mumbai if it brings more revenue to the table. To investigate this phenomenon, we utilised open solver along with the newly defined objective function to solve the optimisation problem, yielding the schedule reflected in the figure below and ticket revenues INR 389,682,972.34 i.e. 8% higher than the initial solution of 362,106,737.02, indicating that the IPL can boost revenues by opting for stadiums that have a favourable mix of seating and weather conditions.

Day	Match 1		Match 2	
	Home	Away	Home	Away
1				
2				
3				
4	DC	MI	CSK	RCB
5	CSK	MI		
6	DC	RCB		
7				
8	MI	RCB	CSK	DC
9	DC	CSK	RCB	MI
10				
11	MI	CSK	RCB	DC
12	MI	DC	RCB	CSK

Day	Match 1		Match 2	
	Home	Away	Home	Away
1	RCB	DC		
2				
3	CSK	MI		
4				
5	DC	MI		
6	RCB	CSK		
7	CSK	MI	RCB	DC
8	DC	MI		
9	RCB	CSK		
10	RCB	MI		
11	DC	CSK	RCB	MI
12	DC	CSK		

Figure 3.3: Schedules before (left) and after (right) factoring revenue optimisation & schedule relaxation

4. Ticket Pricing

Overview and Potential Challenges

Pricing tickets at a constant rate, as our revenue forecasting model assumes, may lead the IPL to lose out on more revenue. A constant pricing strategy fails to capitalise on customers' higher willingness to pay closer to the game. It doesn't take into account that on average, most IPL games are either sold out or close to selling out each stadium. The black market thrives in this scenario by pricing tickets at large multiples to official prices (at times even 3-4x). Thus, we were interested in determining how much more the IPL can make if they differentiate selling periods into "early bird" and "late price", (building upon the solution of a class homework problem) and to further identify the optimal strategy to maximize revenue.

Data & Assumptions

We assumed that tickets would be split into 2 classes - early bird and late, and early bird tickets would be sold until a configurable booking limit and further that late tickets would be priced at a multiple to the early tickets. We grouped the 8 classes into Lower, Middle, and Upper and took average prices of specific seats in that class (reflected in Figure 4.1). We assumed that the booking limit and price multiple associated with late tickets would be identical for each ticket class. Demands were assumed to be normally distributed similar to the revenue forecasting simulation and where mean and variance were computed using proportions as in the revenue forecasting model. The goal would be to find the best early booking limit (cap) configuration along with the best multiplier for the late pricing to maximize revenue.

Demand Distributions - Normal			
Class	Mean	Variance	S.D
Lower	21000	2250000	1500
Middle	12250	765625	875
Upper	1750	15625	125

Class	Early Bird Price (₹)
Lower	600
Middle	1600
Upper	6000

Conversion Probability	70%
Decline per multiplier step - Lower	30.0%
Decline per multiplier step - Middle	20.0%
Decline per multiplier step - Upper	15.0%
Conversion - Lower	10.0%
Conversion - Middle	30.0%
Conversion - Upper	40.0%

Figure 4.1: New Demand Distributions and Conversion Probabilities

Since we lacked data on demand at specific times, we simplified the model to consider only customers who demand tickets in the "early bird range" and decided to investigate if the IPL could benefit from setting an early ticket limit and a different price for late tickets. We modelled the "late demand" as a binomial distribution with the number of trials as the number of individuals that were unable to obtain a ticket in the early bird round. We assumed that the probability of purchasing for people across all classes would be 70%, which is a conservative estimate given the popularity of the IPL. We decided to factor the drop in this probability based on a customer's willingness to pay. For this, we assumed that the buyers in different classes are disjoint and those in the upper class have the highest willingness to pay. We modelled the decrease based on the multiplier and assumed that the drop in purchase probability every time the multiplier increases by 1 is 30%, 20%, and 15% for the lower, middle, and upper classes respectively:

$$Max\,(0,\ 70\% - Decline\,(Upper/Middle/Lower) * (Multiplier - 1)$$

Model

Our model has 36 streams of income, coming from each of the Home-Away match pairings separated by Lower, Middle, and Upper class. Below is the logic for one such revenue stream:

1. For any game and each seating class, **demand** is forecasted using a normal distribution that is explained in Assumption 1.
2. Based on demand for each ticket class, **sales** for that ticket class is calculated using the minimum between demand and capacity. This time, capacity is also capped by **the booking limit.**

$$Min\ (Early\ Demand,\ Stadium\ Capacity\ *\ Booking\ Limit\ Ratio\ *\ Proportion\ of\ Class\ (Upper/Middle/Lower))$$

3. Those that could not buy a ticket become the number of trials for the binomial distribution, modeled by the "Rejected" column. Late Demand is then modeled using Crystal Ball as a binomial distribution. Late Sales is calculated as the following:

$$Min(Late\ Demand,\ Stadium\ Capacity\ *\ Booking\ Limit\ Ratio\ *\ Proportion\ of\ Class\ (Upper/Middle/Lower)\ -\ Early\ Sales)$$

   Since late demand cannot go beyond the stadium capacity, we find the remaining capacity from the early sales and find the min.

4. Early Revenue and Late Revenue are determined via a sumproduct of the corresponding prices.

$$Early\ Revenue\ =\ SUMPRODUCT(Early\ Bird\ Prices\ for\ Different\ Classes,\ Sales\ for\ Different\ Classes)$$

$$Late\ Revenue\ =\ SUMPRODUCT(Late\ Prices\ for\ Different\ Classes,\ Sales\ for\ Different\ Classes)$$

5. Based on the rainfall distribution (modeled via gamma distribution), **loss of revenue** for the match is determined by if the total rainfall is greater than given thresholds.
6. **Final Revenue** (Forecasted Value) is calculated by summing all of the revenue streams.

Findings and Recommendations

	Multiplier (1.00)	Multiplier (1.50)	Multiplier (2.00)	Multiplier (2.50)	Multiplier (3.00)
Booking Limit (0.00)	387,480,660.52	494,780,250.95	544,452,801.46	536,499,164.24	470,919,147.58
Booking Limit (0.20)	425,348,902.02	508,180,832.30	546,526,843.63	540,387,296.72	489,762,641.61
Booking Limit (0.40)	463,214,510.96	521,583,323.01	548,599,814.54	544,273,329.43	508,604,591.96
Booking Limit (0.60)	500,850,673.16	534,963,106.79	550,672,929.19	548,160,290.69	527,437,715.39
Booking Limit (0.80)	530,569,209.32	545,297,289.66	552,213,991.71	551,398,159.06	543,163,981.85
Booking Limit (1.00)	540,700,155.08	540,712,749.99	540,720,460.14	540,722,985.78	540,720,858.77

Figure 4.2: Decision table reflected impact on ticket revenue from changes in booking limit and multiplier

Our model determined that the optimal booking limit for early bird tickets was .8, and the multiplier is set at 2.00. Thus, by selling a maximum of 80% of tickets in the early round, and selling the remaining at twice the original price, the IPL could earn a revenue of INR 552,213,991.71 which is approximately 53% higher than the revenue obtained in the original revenue forecast and further, approximately 42% higher than the revenue generated with the revamped optimisation model. Looking further into this strategy could further be beneficial for the league as it looks to perhaps recover the lost ticket revenue during COVID-19.

Limitations

We faced difficulties in determining the exact price elasticity of demand for buyers purchasing different classes of tickets. Additionally, we weren't able to find demand based at different times before the match which would have allowed us to model demands for each ticket class more effectively. Our model assumes a price elasticity of demand that followed the linear formula above, but the exact intercepts/coefficients will be much more important and difficult to find in reality,

5. Conclusion & Evaluation

In conclusion, we were able to accomplish three main objectives:

1. We created a schedule for a competition with 4 teams purely through discrete optimization. This incorporated existing constraints and regulations used by the IPL , and improved on it by adding in new constraints that we thought would benefit the league.
2. We further built on this scheduling problem through the usage of Crystal Ball in adding in match revenues. We were enthusiastic about getting an answer that maximised the IPL's revenue, and were particularly proud of the fact that we simulated each match's revenue as we did not have match attendance data on hand.
3. We explored the possibility of manipulating the IPL's ticket pricing strategy through an early bird and late price and were able to show that there was an optimal booking limit that they could put in place to gain more revenue.

There are definitely many things that we could improve and extend on this project. Firstly, we assumed that the league consisted of four teams to simplify the number of decision variables we had. We had to use Open Solver for this mini-league due to the number of constraints we had. Given more time and a more powerful optimizing engine, we would ideally like to conduct our trials on a full sized competition of eight teams.

Secondly, while we were able to simulate individual match revenues, data on match attendance and the number of tickets from different price categories solved would help verify our conclusion and improve the accuracy of our model. We could accomplish this by reaching out to the IPL and individual teams and ask them if they were willing to share such data, otherwise, we will have to wait until such data is made publicly available.

Finally, we worked on this model assuming that the current global pandemic had not occurred. This year's IPL was played in empty/low capacity stadiums in the UAE and not in India. Thus a lot of revenue was driven by TV viewership and advertisements, and some of the constraints that we had, such as the home and away game constraint, were made redundant. Assuming that next year's edition of the competition would be conducted like this, we could investigate the nature of this new format and try to find areas that can be optimized.

Appendix - Bibliography

(1) https://t20slam.com/ipl-2020-tickets-price-list-online-booking-guide/

(2) https://www.sportekz.com/list/popular-t20-cricket-leagues/

(3) https://en.wikipedia.org/wiki/List_of_attendance_figures_at_domestic_professional_sports_leagues

(4) https://www.sportskeeda.com/cricket/icc-t20-world-cup-2016-eden-gardens-seating-arrangement-india-pakistan-clash

(5) https://arxiv.org/pdf/1708.03144.pdf

(6) http://www.ipublishing.co.in/ijcserarticles/twelve/articles/volsix/EIJCSE6011.pdf