Course Project

Source

https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Leveraging-O.R.-to-Achieve-the-Competitive-Edge

Introduction

The ANFP decided to tackle the long-standing problem of how to schedule their soccer teams. From the abstract, it appears they were quite successful, but what kinds of issues do they address, what data is needed, and what do their models reveal about the underlying relationship between sporting events, the teams that participate, and the fans that attend?

While there are several peripheral concerns that can be modeled and analyzed, the focus is on three that are common to every sporting event: prediction and economics of ticket sales, the scheduling of matches during the season, and focused engagement and advertising efforts.

Scheduling

Before anything else can happen, teams have to be scheduled to play. Everything else is contingent of this: if there are no games, there is no reason for players to play, fans to cheer, and no money made for the ANFP (and any other league for that matter). At a minimum, at least one

team has to play at least one other team for everything else to matter.

Obviously, everyone knows that minimum is insufficient for a sustainable league, so how should teams be scheduled to create a profitable and exciting venture for the league, teams, players, and fans? We model it using Integer Programming.

Integer Programming is a linear optimization technique restricted to only integer values that will attempt to solve the following objective function: maximize profit, with profit expressed as revenue minus costs. Only integer values are considered as teams shouldn’t be able to play against a fraction of another team; it’s a bit unfair.

Simple, right? If you believe so, then you are likely already very wealthy. For the rest of us, there are several variables and constraints that we should consider to create a working model.

Let us start with the basics: we need to know the number of teams participating and where they are located. This is settled by the governing body before the start of each season. Each team has a venue they generally operate out of, but it’s possible there are alternate locations in case of special need. Beyond that, the ANFP and other leagues have rules in place that enforce certain restrictions.

What kinds of restrictions? The simplest are structural by nature: a minimum and maximum number of games a team must play in general, and the maximum and minimum number of times a team must play each other team. In addition, there might be additional rules governing

advanced structures like conferences and divisions; teams in one conference might only be allowed to play a team from another conference under special conditions. League rules might also dictate rules to protect player safety by specifying a minimum amount of time to exist between games. These rules, and more, can be obtained by contacting the ANFP or other governing body.

Apart from safety and legal mandates, the ANFP wants to maximize the revenue generated by each game. There are going to be times where it is predictable that certain matchups will yield a greater than average attendance; attendance data is generally obtainable from the ANFP and hosting venues. Existing rivalries might have a tradition of specific matchups occurring on a particular day or within a specific time-period. It’s important to research the histories of each team in the league to make sure important games are scheduled during the season, especially in cases where it might not be a given that two teams will play each other. Failure to do so might result in decreased short-term ticket sales and long-term interest by the fan bases, stifling potential secondary market opportunities like branding and merchandizing.

There are practical considerations as well. For example, If Team A is scheduled to host a major rivalry game with Team B, but Team A’s primary venue cannot support the expected number of fans, there can be a potentially avoidable drop in revenue resolved by making sure Team A’s larger alternate location is used by the scheduler.

Finally, costs must be taken into account once mandatory and practical considerations are dealt with. Ultimately, if costs exceed the projected

revenue, the entire league will fail as a result. The Alliance of American Football league is a recent example of projected costs quickly eclipsing projections, leading to an early termination.

For each venue, we need to minimize cost as part of the objective function; in order to maximize profit, we want to maximize revenue while also minimizing cost. To do this, it's advantageous to schedule games so that teams can travel the least number of times and the least distances. One such example is scheduling regional teams to play each other as a geographic cluster; then, several teams from a region can travel and play another entire region, avoiding the issue of traveling back and forth over long distances. In fact, these geographic clusters might naturally be organized into the conference and divisions I referenced earlier.

Advertisement costs can be utilized as well. Advertising likely has escalating costs at different levels of influence: national TV ads are far costlier than a simple social media campaign. Games can be scheduled to take advantage over multiple rivalry games happening around the same time to combine major advertising pushes, while allowing more local media to promote more mundane offerings.

This Integer Programming model is likely to be re-evaluated at the conclusion of each season; this allows the league to see the effects of league policy changes, advertising costs, venue fees and travel costs, etc. Scheduling for the upcoming season can add the previous season's data to the overall dataset for further refinement.

Ticket Sales

Ticket sales are a major factor used by the Integer Programming model to create the league schedule. Often, they are a principal factor for whether or not a particular franchise is allowed to exist; if a team cannot sell tickets, the owner might feel pressured to move the team or sell. Stability is important for overall league health, so we want to be able to predict ticket sales at different pricing models and methodologies.

Ticket sales include not just physical “person-in-seat” sales, but also television and streaming viewers. For example, a person can buy a digital ticket and be able to view a sporting event. However, we still want to know if we can sell physical tickets, box suites, season tickets, or other in-person services such as a ticket that allows entry into special restaurants with a great view of the game. Even someone watching for free might be generating revenue as a metric required for continued support by advertisers or investors; their view is essentially an indirect ticket that can be modeled.

So how can we determine the probability that a person will buy each different type of ticket? We can set up a logistic regression to model the likelihood between 0 and 1 that a person will buy each type of ticket for each team and/or venue. In fact, each type of ticket can be modeled individually, with the maximum of those probabilities being the likelihood that a person will buy any ticket. The total predicted number of each ticket is the probability of the ticket purchase for a given target population, and the total number of tickets is simply the sum of each of

these results.

But what is the target population?

Target Population

In simple terms, the target population are the people expected to purchase each type of ticket. Why is this important? I can effectively advertise to them if I know who they are.

For digital sales, the potential market is limited only by TV/Internet access and governmental restriction (national-only games, blackout rules). However, this doesn’t mean every person online or watching TV will buy a ticket. We can purchase data from companies like Google and Facebook who use their advanced analytics to pinpoint people that have recently searched “soccer tickets”, or the name of the league/team, or have bought associated merchandise or related sporting equipment. The more specific and frequent the searches, the more targeted we can be with advertising different ticket types. Age demographics might also correlate; perhaps newer fans are more comfortable with watching a match online versus going to a physical venue.

For local physical ticket sales, the majority of sales are likely to originate in a short radius from the venue. However, this doesn’t account for potential density where advertisement of one ticket might be far more effective than another. For example, advertisement of box suites and sponsorship deals will likely net more revenue if targeted at affluent business districts rather than middle-class suburbs. Areas with

more personal transportation might be more likely to buy season tickets than areas that rely on public transportation; or vice-versa.

We can use K-Means Clustering to group and focus advertisement efforts in an attempt to increase ticket sales of every variety.

Digital-based clusters can map analytics and associated browsing and purchasing behavior with each type of ticket purchased. In fact, online questionnaires from the team might offer more direct insight into what level of ticket or service was purchased as a result of online advertisement.

Geographic-based clusters can help pinpoint what local markets are being actively engaged or potentially ignored. Aside from the previously mentioned advertisement benefits, different communities can be modeled to help create the ticket sales models to determine different pricing structures and promotional opportunities.

The primary challenge is to be careful not to use potentially biasing data (income, location) in a manner that can be construed as discriminatory; clusters should not be used for predatory purposes or exploitation.

Conclusion

Each model is used to fuel future refinement of the others. The eventual goal is to better understand customer habits and desires, predict primary revenue streams, and schedule games to allow for maximum revenue and long-term interest growth while minimizing on-going and

emerging costs.