Researching Dynamic Pricing at AZ Alkmaar

Master of Science Business Analytics, Vrije Universiteit Amsterdam (36 ECTS)

Author: Sammy Junior Obeng Eshun

Student Email: s.j.o.eshun@student.vu.nl

Student Number: 2598332

Supervisors: Mr. Barend Verkerk, Dr. Ger Koole

Company: AZ Alkmaar

Abstract

Comparing the football industry found in Western Europe to other types of sports found in America, it doesn’t take long to conclude that our industry is lagging behind in innovation. This is shown by only a few teams publicly adopting data science to further enhance their first teams and club operations, until recently. AZ is one of the leading teams in the Netherlands when it comes to advance usage of data science. Under the guidance of Billy Beane, Robert Eenhoorn, formerly Marijn Beuker and Barend Verkerk, data science is now widely used to make data-driven decisions at AZ Alkmaar. Data analysis, forecasts, and a pricing policy are the main aspects of a dynamic pricing strategy.

This thesis provides an outline of how this can be applied within the setting of football matches at AZ Alkmaar. Elements like occupancy and costumer perception are all included in this analysis to make it feasible for the club. First, a literature research is done to provide some possibilities for the practical implementation of dynamic pricing at AZ Alkmaar. After this, a data analysis is done which provides useful insights on revenue potential in the stadium. A first attempt to obtain price elasticity follows the data analysis, this ultimately fails but show that price is negatively correlated with sales.

Acknowledgements

In front of you lies the thesis "Researching Dynamic Pricing at AZ Alkmaar". It functions as the initial body of research into the theme of dynamic pricing at AZ (read, AZ Alkmaar). It intends to serve as the basis for further research that AZ Alkmaar could use in order to deploy a strategy which dynamically changes prices of match tickets. It was written to fulfill the complete program of the Master of Science (MSc) Business Analytics at the Vrije Universiteit Amsterdam. I researched and wrote this thesis from May 2021 till January 2022.

Robert Eenhoorn and Barend Verkerk requested me to research dynamic pricing at AZ Alkmaar, a football club at the highest level in the Netherlands. I found it difficult during the thesis to discover one ultimate path which could produce dynamic pricing in the short term. I explored different routes which gave me much to ponder. Extensive research and guidance, is what made it possible to hand in a body of work that is useful to AZ. This will serve as a pillar for future research in pricing within AZ. I would like to thank both of these gentlemen for giving me the opportunity to research this challenging and interesting project at AZ Alkmaar. I could not have imagined a more enriching environment where I could write a Master thesis.

I would like to thank my supervisors for their unwavering commitment, guidance and support during these months. Their expertise and critical thinking were vital in successfully completing this thesis and internship. Their flexibility amidst challenging personal circumstances were of great help. I would further like to thank all those at AZ Alkmaar who helped me in any shape or form. This includes the ticketing team that provided me with the necessary data and information. As does it include my colleagues in the data science team who have sat with me and reviewed information with me; blessing me with their data science proficiency whenever necessary.

I would also like to thank my parents and siblings for their support from the moment I first spoke to Barend for this project, up until now. To make them proud has been a driving force throughout this process, and without it I would not have been able to complete it.

I hope you enjoy reading this thesis!

Junior Eshun

Amsterdam, January, 2022

Management Summary

The main relevant insights gained for AZ by this internship, are in the initial prices of match tickets. It was made clear that some of the seat sections that are similar in price, are not similar in desirability. Seats that were located at medium height were most preferred, likely because this gives the best view of the match. Lower seats were preferred to upper seats. It was found to be best to dissect the stadium rows in three categories; low, middle and high.

There were other asymmetries in the desirability of the seat sections. Comparable seat sections further away from the away section are preferred over those which are closer to the away section. Currently they share the same price and again, the data analysis points to changing the current setting. A raise in the more desired seat sections or discount in the less desired seat sections could boost revenue for AZ.

When it comes to the dynamic pricing at AZ, an attempt to obtain a proxy of the price elasticity failed. However, an overview of how to obtain price elasticity and demand functions has been discussed in the thesis. This is by taking some time to experiment with changing prices of tickets. Further avenues AZ Alkmaar could take regarding algorithms to use, with or without price elasticities, have also been provided. This thesis then serves as a base for further research on dynamic pricing at AZ Alkmaar.

Contents

1 - Introduction

Current setting at AZ

Currently at AZ there is a data science department which was established about three years ago. This consists of the Head of Data Science Barend Verkerk.

CONFIDENTIAL

Over the past few years, AZ has ended third, second/first, and fourth. When the size, status, history of the club, and the size of the city/region (including the competitors in this region) are considered, that is an excellent achievement. AZ is generally considered to be the fourth biggest club of Holland after Ajax, PSV, Feyenoord respectively. This is mainly because of the excellent vision and policy the club upholds. This is even after various dramatic events such as the collapse of DSB Bank and the recent collapse of AFAS stadium roof.

AZ's goal is to be able to compete with Ajax, PSV and Feyenoord and become a household name within Europe. AZ wants to win Eredivisie titles, KNVB Beker titles, and also to make it to the Champions' League group stage. Whilst AZ has all these goals, they wish to stick to the identity of the club. Which is promoting youth players from the stellar youth academy. The academy has frequently been voted the best in the Netherlands and thus one of the best in the world. An attacking brand of football is inherent to the club, and this is the style which the fans like to see. Impressive about AZ then, is that these goals aren't compromised by the fact that AZ has less funding than the clubs aforementioned.

AZ is privy to the knowledge that youth players are essential, hence why they have adopted a similar model to that of Dortmund and Ajax. When a club excels at developing young players to professional standards, major clubs are then enticed to come and bid for these players for a significant amount of money. While improving the performance of the first team, the money secured from the sale of football players can be of good value to the club. For example, it can be used to scout other talents, enhance the academy, improve the club departments and make balanced signings for the first team. This balance is key for the growth of AZ, that has voiced their goal to win a league title and be in the top 50 European clubs before 2025. Vision, strategy and excellence is needed to grow, because of the smaller budget.

One of the ways for AZ to excel, is through usage of data science. This thesis focuses specifically on the research and potential application of (dynamic) pricing. Businesses usually want to maximize profit, and they may do this by maximizing revenue, minimizing costs or both. For AZ when it comes to the cost of their product, AZ achieves this by having the best team possible for the minimum costs. This is done by using academy players and the scouting of excellent players. Some areas which the club cannot significantly minimize costs are: lighting, security, cleaning and stadium costs. These are some of the direct costs of a football match.

However, there is still some room for improvement when it comes to maximizing revenue. Match tickets and season tickets pricing policies could be improved, resulting in higher revenue. That was the task of this research and internship - specifically with respect to dynamic pricing. Ideally AZ would increase revenue of the regular match tickets, whilst keeping a full stadium.

The research question then would be: "How can dynamic pricing be deployed at AZ Alkmaar?"

Revenue management & Dynamic pricing

The word revenue management came up very early on in the internship. Coincidently, it was one of the topics that prof. Ger Koole was lecturing in the course Optimization of Business Projects. It has also appeared in much of the literature concerning dynamic pricing, where previous graduates of the Master of Science Business Analytics are now revenue management experts. There is some relation between revenue management and dynamic pricing which will be briefly explained.

Revenue management and dynamic pricing both use analytics to optimize prices. Revenue management predicts demand and can optimize attendance and price. This notion has been defined by Robert Cross as “selling the right product, to the right customer, at the right time, at the right price.” [Cross (1997)]

They both rest on the principle that inventory is perishable. This means that the product is available for a set amount of time, and after time it becomes useless. This can be translated into football tickets. A ticket has value until the gates of a stadium closes or when the ticket is scanned. Another principle that is necessary is that different customers are heterogenous in their willingness to pay for this product. In the case of AZ that would mean that in a seat section for a particular match, one customer is willing to pay X for a ticket, and the other customer is willing to pay Y. If all customers had the same willingness to pay, the price would just be fixed to that amount.

Moreover, it should be possible to divide the fans into segments with a limited arbitrage opportunity. If these conditions are met, and the costs of applying dynamic pricing is smaller the perceived increase in revenue, then dynamic pricing is a good option for a sports team.

[Narahari (2005)]

The main aim of revenue management is to maximize the profit through market demand forecasting. Revenue management uses analytics to predict where and when there will be demand. By analyzing past purchases and consumer behavior, it becomes possible to find the best prices to set a product for different times. Revenue management strategies work best when the products are sensitive to changes in price. But to have a successful strategy, it is optimal to know the customer, and to know what value the product holds. Revenue management is much more focused on understanding customers and forecasting demand. It also focuses on segmenting products and finding the best way to allocate products to customers depending on their willingness to pay. Additionally, it can be used to tailor fit different customers that have different behavior.

Dynamic pricing is not one specific pricing strategy. Instead, it is a way of pricing a product according to various rules where the focus is on price and adjusting them. While revenue management can include a dynamic pricing of seat inventory, a successful revenue management strategy can also be based on fixed pricing.

Dynamic pricing uses intelligent algorithms to calculate and adjust prices in real-time. In this way, it is possible to recalculate and optimize prices as often as needed to in order to maximize revenue and increase profit margins. This is the study of determining optimal selling prices of products or services, in a setting where prices can easily and frequently be adjusted. Furthermore, dynamic pricing with learning can be regarded as the combined application of two research fields. Statistical learning, specifically applied to the problem of estimating demand functions, and price optimization. [Den Boer (2015)]

The application of dynamic pricing in sports

The concept of dynamic pricing can be applied to various sectors. The dynamic changing of prices can increase revenue as demand is more precisely met. Albeit under a different name, it is applied in the hospitality industry, airline industry and entertainment industry. It is important to note that dynamic pricing in the USA sports industry is widely applied among the top sporting franchises in the MLB and NBA [Kemper et al. 2016]. One renowned external party that applies this strategy is Qcue and they have developed a tool that can be customized for various clubs.

In Europe there have been less documented success stories regarding the deployment of dynamic pricing in sports. This could have different reasons, of which one could be the difference in culture among the clients and supporters. Another reason could also be the franchises. America is well known for their advanced culture regarding sports and this has been going on for quite some time when the movie Moneyball is considered.

Sports and entertainment have such a huge presence in America. There is a huge market on that side of the world. Stadiums are much bigger, and prices are also higher. Americans have no issue purchasing tickets to enjoy the sports they love meaning huge sums of money are spent and invested in the sports industry. This is a different story to what is being written over here in Europe.

One of the problems with dynamic pricing is, that it can easily be perceived as unfair, which is a dangerous term in the football sector. It is not well known whether fans would accept paying higher fees than others, for similar seats. This seems to be the case in America, but it remains to be seen how this translates to football in Europe. For this reason, a thoroughly designed approach is necessary, as supporters don't just hold ticket value but also net value. They make use of concessions, purchase merchandise, and subscribe to social media, with the potential to keep coming. A naive maximization of the revenue that can be generated for one match can be harmful, as this may decrease revenue in the long-term.

Transparency with Dynamic Pricing

Various approaches can be made to increase the likelihood that supporters will accept dynamic pricing. According to [Choi et al. (2005) – Choi et al. (2004)], it is necessary to communicate the use of dynamic pricing to fans. This communication needs to be extensive and worded as a win-win for both parties. Typically, customers find it unfair when the price of a product increases whilst the cost for the seller doesn't. This means that only the seller has an advantage. However, when a customer's gain is communicated clearly, their perception changes to fair as both parties gain. One way this can be done is to rather make mention of price discounts rather than price premiums. "Through dynamic pricing it is possible to pay a discounted price". This is good for the customer, but also good for the seller as potentially more seats can be sold, and on top of that a higher revenue can be achieved [Kimes et al. II (2003)].

Other than this part of the communication, the customers should also know how the dynamic pricing strategy works. This need not to be in detail, but basic understanding should be achieved where customers know how to make use of the discount. When these strategies are combined, this improves the perceived fairness of dynamically priced tickets [Choi et al. (2004)]. When it comes to dynamic pricing, it is important to tell the story of customers being given the opportunity to influence their price, by achieving discounts on the price if certain behavior is performed. It is also good to be transparent and show what behaviors leads to which results.

It is also interesting to note how customers perceive the fairness of variable pricing which may vary over time. It was found that pricing which was originally seen as unfair, such as hotel rates, have now become much more accepted by society. Cultural elements can affect the acceptance of variable pricing and its perceived fairness. Many companies and different sectors now deploy dynamic pricing such as Uber. It was prognosed that in the years to come this acceptance will only rise. Now in 2022, this prognosis seems to have been realized [Kimes et al. (2003)].

To deal with these issues, in the sporting industry many at times franchises choose to make prices non-decreasing. The price can only increase over the booking horizon. This approach has several positive effects;

- This causes customers to book tickets earlier when they know of this strategy. This has the benefit of making demand prediction easier as the booking times can be predicted.
- This strategy can easily be framed as a win-win situation where the customer knows how to influence the price of their ticket, and hence is likelier to be perceived as fair.

This is an interesting approach, however [Jiaqi et al. (2019)] advocates against this as it is not efficient revenue wise. Far higher revenue can be generated when this strategy is abandoned. It is explained to be a case of being too careful and hence losing the gain of the dynamic pricing strategy.

Important to note is that in the context of the situation at AZ, revenue maximization is just one of the aspects. It is also preferable to fill the stadium, as this helps to achieve more sponsorships as advertisements would reach a bigger audience. Similarly, a full stadium increases the home advantage causing the home team to perform better [Kemper et al. 2016], this is of huge value to the first team and the potential trophies and prize money can be won from it.

Dynamic pricing and price elasticity

Dynamic Pricing rests on the principle that prices are elastic. When there is a change in price, there is also a change in the demand of that product. Optimal dynamic pricing would then be to change the prices in such a way, that the revenue (total sales * price), is maximized. For simplicity, total sales can be exchanged with demand. When there is an indication of how much the demand will be when a certain price change is evoked. The pricing strategy can then be validated through that price elasticity.

If prices aren’t elastic, prices can be increased until they reach elasticity. If this was the case, prices could just be raised higher and higher however, this is inconsistent with all literature regarding the price elasticity of football tickets.

The definition of price elasticity is then given to further discuss this concept; “a measure of the effect of a price change or a change in the quantity supplied on the demand for a product or service.”

In mathematical terms, it is expressed like this;

$$\text{Price Elasticity of Demand} = \frac{\text{\% Change in Quantity Demanded}}{\text{\% Change in Price}}$$

It is difficult to identify the price elasticity when there is no evidence of historic price changes of products. However, in this thesis a way will be proposed on how this can be achieved.

Summary of interviews with experts

By means of the network that AZ has, it was possible to speak to various big and experienced names in the dynamic pricing and revenue management sectors. Some of the most interesting insights gained during these online and physical meetings are discussed to further aid the research.

Qcue

There are several experts in the field of applying dynamic pricing to the sports industry. One of the leading parties is Qcue, which is an American company that applies dynamic pricing in leagues such as the MLB, NBA and NFL. Most of the franchises in the MLB are signed to their services. It became apparent that most clubs tend to delegate the task of dynamic pricing to the experts such as Qcue or Demand Analytics during our meeting. These parties supply the franchises with price recommendations based on a tool that they have developed. This tool contains a lot of parameters that can be adjusted to the various clubs. In that sense, the tool is quite general but can be fine-tuned to the needs of the specific clubs.

They use price rules and limits, and are capable to work with SRO which is the ticketing system that AZ currently uses. In their experience there has been very little backlash from fans when it comes to applying dynamic pricing to tickets. This could have to do with a different background of Americans which has been hinted at earlier in the introduction.

Thomas Duncan

Thomas Duncan is a data scientist at the Sacramento Kings, a basketball franchise in the NBA. Some of the interesting things picked up during this meeting where that he mentioned that the dynamic pricing model that they used, wasn't incredibly sophisticated. He mentioned that sophistication is nice but unnecessary.

Sacramento Kings had some personal challenges with a new arena, it was important to price in a way where the new arena would be filled. By achieving this, more sponsorships could be generated. Ultimately, they used about 260 price levels by extreme granularity which Qcue couldn't support, there would be massive overfitting. They were able to perform dynamic pricing without price elasticity or demand functions of themselves, but relied on what was happening in the secondary market. In quite a deterministic way, they then adjust the prices of the seats.

Important tips they gave, were to protect the season ticket holder value, this should always be the lowest average price per game and they should never be undercut. They found that some games never really sell-out no matter what, but after the new arena they were able to go on a sell-out streak from about 80 games.

Make a model which ranks games, incorporate information on which players on the opposing team will be able to play, check if there are other events in the region and that the best models can lead astray. It is important to find something simple that works.

Daniel Hopman

Daniel Hopman is a former Business Analytics student, who has worked at KLM and Majid Al Futtaim as a revenue management analyst. He wrote his PHD titled: The Real Theory and Practice of Revenue

Management. This eventually became a book which challenges the old and impractical implications of the general accepted theory behind revenue management.

Some of the interesting insights gained from this interview were considerations for a price that doesn't decrease, which makes it easier to predict demand as demand becomes more predictable when there is such a pattern. The good thing about this is that it becomes easier to sell to the supporters, that loyal supporters can buy their tickets earlier. His book also has references on researching regarding customer perception.

Rik van Leeuwen

Rik van Leeuwen is another former MSc Business Analytics student, who has worked with Ger Koole. Currently he is employed at Ireckonu, a company which specializes in hospitality. In his thesis, he looked at demand forecasting in hospitality using smoothed demand curves. Interesting about this method is that demand could be predicted without having the price as a factor by splines. It uses a method of looking at matches that look like the current trajectory of the match looked at, by linear or dynamic programming it is then possible to maximize the revenue.

Perry Hendriks and Convius

Perry Hendriks is the head of marketing at Almere City and has formerly worked for NAC Breda, FC Utrecht and in the NFL. He was a sent representative of Almere City's chairman; John Bes, renowned for his interest in Data Science. At Almere City they did a test with Dynamic Pricing and Real Time Pricing. The Real Time Pricing was what this thesis considers Dynamic Pricing and the Dynamic Pricing was some fixed pricing strategy. The former was unsuccessful at Almere City but successful elsewhere, whereas the latter was successful. The concepts will be explained here after. Both were performed by Convius, a party also specialized in the pricing of tickets of events and festivals.

It was possible to suggest a ticket price with some predetermined minimum price. However, as a supporter fills in their willingness to pay, it is tied to how much money the club would be able to spend if everyone would buy a seat for this price and where their expected league position would be as a result of that. This feature made it a game-effect, in which a greater attendance and revenue was achieved as fans were willing to pay more for the good of the first team. This shows that there is some sense of price elasticity where for different prices, a great attendance was reached.

In this experiment, there were different factors that would influence the prices of tickets. Some of those factors were the location of the seat, the weather, traffic on the website, the opponent and the number of tickets left for sale. This pilot was manual, but in terms of achievement rather poor which had to do with some external factors other than this experiment. Interesting to note however was that there was no backlash over this experiment whatsoever.

In summary, the greatest insights gained in these interviews were what would be the end of the horizon, the application tool created by Qcue. Is it worthwhile for AZ to deploy dynamic pricing internally, or is paying an external company more effective? Dynamic Pricing is certainly possible in football, Convius has shown that it is even possible in the Netherlands, there are various directions to take in terms of the mathematical model or even branch that is used, but its lucrative nature is evident.

The framework of this thesis is created based on a literature research towards dynamic pricing in Section 2, speaking of the practical directions. Then follows an extensive data analysis in Section 3, which includes a description of the data available for working towards a dynamic pricing model at AZ

Alkmaar. A logistic regression model is presented in Section 4, followed by the conclusion and discussion in respectively sections 5 and 6.

2 – Literature Research Towards Dynamic Pricing

Obtaining demand functions.

Currently the bottle neck is that demand functions aren't present, nor are there accurate or reliable indications for price elasticities at AZ. There is no historic data where AZ has experimented with price changes, so price elasticities or demand functions are not straight-forward to obtain. There are some price elasticities calculated in the literature. However some of these are very old or club/league-specific. The results are also quite varying across different papers. AZ wishes not to copy these estimates and build models upon them. So an important questions to ask are: Are demand functions and price elasticities necessary and if they are, how can they be obtained? If they aren't, is it still possible to perform a dynamic pricing strategy? These questions are answered in this section.

One suggestion to try and obtain price elasticity, is to run a logistic regression model, which will be explained in section 4. This regression ultimately doesn't yield the needed outcome, but improvement and alterations may lead to a more useful outcome. The insights of the regression model will be provided in that section.

A way to find demand functions and price elasticities, would be to experiment with ticket prices for a significant period of time. A suggestion would be to change prices for two seasons (to be able to make comparisons), and observing what happens to the demand or sales. In the McKinsey Quarterly [Baker et al. (2001)], a flagship business strategy publication, firms were recommended to research and deploy dynamic pricing by initially doing price experiments. When small adjustments in prices are made, firms can find the demand functions for their customers. There are many advanced theoretical models in literature, but in practice, making predictions of these demand functions proves to be rather difficult on historic sales. Implementing pricing strategies is then barely feasible, that's why McKinsey recommends to start with some experiments. Here at AZ, demand may be measured by some web measure of clicks, or how many people reach some stage in the ticket purchase process. Easier would be to look at the amount of tickets sold. When price changes are made, demand functions may then be estimated through tracking of the sales made, this would then lead to obtaining demand functions and price elasticities can be obtained. Important in this approach is to be able to obtain some benchmark. Because even if prices remain the same, the demand changes over time. So some knowledge of the amount of tickets that would be expected to sell, must be known. This may be determined by some regression of factors, where the number of tickets sold is the explained variable by time, opponent, weather and so forth. Discrepancies between the demand after price changes and the predicted demand might then give some evidence of the effect achieved by the price changes.

Another approach would be to create some demand functions initially, and to validate them with the experts. The idea would be to use them to optimize prices over time and then to generate new and more accurate demand functions. However this approach is not free of problems. The demand functions in the initial basis could be unboundedly inaccurate, which makes it difficult to make projections for profitability.

It is also vital to have a well-defined way of changing prices, and changing prices significantly. It wouldn't cost AZ Alkmaar much to change prices by few euro's. This would make it possible to deduct some indicators for the price elasticity and even demand functions for some range of the data. However long term, the idea is also to cause the prices to change to a further extend if that is the clever thing to do. Such price changes may be damaging when not applied properly in an experimental phase. For that reason, even during the experimenting phase there needs to be some mechanism that can reliably push the prices towards some direction. Price elasticity can then direct how much into the direction that price change must be pushed.

For that reason it is advisable to have some correcting model that determines whether the applied price change is good or bad, and give some measure of how good or bad the price change is. The method developed in [DiMicco et al. (2001)] and earlier discussed, is then used as a suggestion. From AZ Alkmaar there is a clear demand for a mechanism that gives feedback and shows whether price changes are going into the right direction. Some of the model options will be discussed now.

Dynamic pricing with demand functions

In [Kemper et al. 2016] a framework is presented that shows how to design a dynamic pricing model in the setting of a highest division football team (Bayern Munich). Based on other read literature, this seems to be a very good way of deploying dynamic pricing when demand functions are present. It is consistent with the expertise of Dr. Prof. Ger Koole on the usage of dynamic programming with demand functions.

So if demand functions were present, it would be advised that dynamic programming is the direction that AZ should take. This would be a dynamic pricing strategy that could be automated and run on itself and give price recommendations. The dynamic pricing developed for Bayern Munich in this paper, is based on probabilities that a ticket will be sold, and increasing them over the time that they are not sold by decreasing the price. There are some limitations with the approach suggested in the paper, but ideas to help those are provided as well. In the long-term at AZ, the modelling becomes more complex when demand functions are estimated for all the seat sections, and all the different locations. On top of that, the usage of factors that may influence demand such as win/loss ratio, weather and opponent ranking amongst others, would further make such a model more complicated. However, one need not to make the dynamic pricing application thus sophisticated that everything is included in the model. It is possible to model certain seat sections or areas in the stadium, and translate the general trajectory to the remaining seat sections or areas.

They estimate eight demand functions based on eight different price categories in the setting at Bayern Munich. For these categories, they use secondary market data in order to use a logistic model that generates a demand probability for a given price level. This is currently not feasible for AZ with the data present but AZ could generate this data by doing experiments. In [Cain et al. (2020)], another way to calculate the probability that a ticket will be sold is shown which doesn't only comprise of secondary market data. This is another avenue that AZ could take. The experimental model used a logistic regression model as a probabilistic classifier that then fed into an optimization model. The logistic regression predictive model supplied the probability that a ticket would be sold on the primary market by factoring in specific traits associated with the ticket, which is exactly one of the limitations that were present in designing the dynamic pricing model for Bayern Munich.

Ultimately, a dynamic programming approach is then applied to the demand functions in order to calculate prices. The dynamic programming used for Bayern Munich, applies the Bellman – equation to then optimize prices using beta demand functions and iso-capacity lines.

It is good to note and the paper doesn't fail to stress that, there is extra-ordinary potential at Bayern Munich regardless of the application of dynamic pricing. Most of their matches are sold out, and prices have been very low in comparison to demand. More people would be willing to watch a match, and a lot of people have higher willingness to pay than the current ticket prices. In the likely deployment of dynamic pricing at AZ Alkmaar, probably once a week, a bit more often just before the gameday, someone of the ticketing department would adapt some changes based on some recommendations generated by the algorithm in the initial phase of the booking horizon. There would have had to be some research done on the exact modelling and whether the modelling would be granular enough, yet not too complex. It probably wouldn't be worth it to build such a complex model that needs huge manpower and hence huge resources. AZ looks to improve revenue but its revenue isn't as high as some of the top sport franchises in Europe, let alone in the United States. There would need to be significant revenue improvement, with little manpower. It shouldn't take an entire team of the best available data scientist to deploy this, and preferably more margin should be achieved for the risks that are present. All this is to say that if one professional with a salary of 50.000 euro's yearly and under, is able to implement this within a year, it would be a good direction to go in. Especially if said professional contains the skills to then automize this and move on to other of the data science dealings at AZ Alkmaar.

Dynamic pricing without demand functions

Without the knowledge of price elasticities and demand functions, based on the literature it is still possible to apply dynamic pricing. These strategies use other mechanisms in order to dynamically price items and generate significantly more revenue than fixed pricing mechanisms. Some of these will be discussed in this segment or are referred to from earlier mentions in this section.

One of such frameworks is developed in [DiMicco et al. (2001)]. This framework is a dynamic pricing algorithm that adapts price changes based on information on the performance of the current sales. They give an approach that makes no assumptions or attempts to segment the buyer population into sub-groups.

To further explain the approach suggested in the article, they develop and present the Learning Curve Simulator, a platform for running dynamic pricing algorithms in simulated markets. By analyzing different pricing strategies under various conditions, they show how a seller can take advantage of changing demand levels to earn more revenue and optionally sell more inventory. Which aligns with the desire at AZ Alkmaar. The conditions are; a finite market with a finite time horizon, seller inventory and a buyer population with fluctuations in demand. Taking advantage in this case means that prices are changed based on these fluctuations in demand. The researchers expect dynamic pricing to be common in markets with a finite time horizon over time, as seemingly this is worthwhile.

Opposing to some beliefs that in dynamic pricing it is necessary to segment the customers, [Biller et al. (2000)] make clear that this isn't the case. In revenue management, the techniques applied rely on being able to make advanced assumptions and predictions about the behavior of the customers. However Gallego and van Ryzin discussed that revenue management needed some ideas of dynamic pricing in order to further dynamically adjust prices[Gallego et al. (1997)]. The main difference being that in dynamic pricing, pricing is determined in response to consumer demand. Several algorithms exist that use historical data to predict the behavior of customers [McGill et al. (1999)], however it is always somewhat probable that using historic data to make decisions on the future might go wrong. For AZ in talks it has been apparent that there is a strong desire to have some algorithm that has incorporated a mechanism to be able to adjust when there is a risk of being wrong.

In the research it was concluded that under varying market conditions, a fixed-price strategy performs unexpectedly well. Another good option is to have some pre-set ordered prices that the seller may change between based on the demand. Practically this might be more feasible than actual daily re-optimization.

They argue that when the demand curve is known, a best fixed price can be found which almost maximizes revenue but in reality it is improbable to know the correct demand curve. [Gallego et al. (1997)] show in the automative industry that a theoretical model could be designed for a marketplace with unknown changing demand levels. They show that as long as demand changes, there is a dynamic pricing strategy that performs better than a fixed one.

The actual dynamic pricing performed in this tool makes use of two adaptive strategies. These strategies make no assumptions about behavior of the customers, but just track changes in sales corresponding to the adjustments in price. These observations then lead to the adaptive strategy that incrementally implement price changes. These two adaptive strategies are called; the Goal-Directed and Derivative-Following strategies [DiMicco et al. (2001)].

The Goal-Directed (GD) strategy is called goal-directed as it has the goal of directing the sales in a way that the entire inventory should only be sold on the last day of the market. It aims to sell tickets to the highest paying buyers on every day during the time horizon. It lowers prices when sales are low and raises them when they are high. By some mathematical techniques, this strategy is able to make very extreme price changes in the last days before the event, generating high revenue. It performs well when there is a lot of variation in willingness to pay and when sales are not very important in the first period of the booking horizon. This is deemed likely to be the case at AZ Alkmaar [DiMicco et al. (2001) ; Morris et al. (2001)]

The Derivative-Following (DF) strategy adjusts prices by basing the price change on the revenue earned before, as result of the price change made the period before. If yesterday's price change generated more revenue per good than the previous day, then the strategy makes a similar change in price with its magnitude depending on the success. If not, the strategy suggest an opposing price change. Here the seller aims to sell at the highest price that generates sales. Opposed to the Goal-Directed strategy, this strategy performs best in the first days in the booking horizon. It might be interesting to use a combination of these algorithms to maximize revenues.

Another dynamic pricing framework that was designed was conducted in [Jiaqi Xu (2019)]. This paper is mainly useful because it was actually designed in the context of sports and doesn't use demand functions in order to apply the dynamic pricing. It uses dynamic programming which optimizes on a demand model. This demand model mainly consists of logistic regression models, similar to those performed in this thesis. This demand model stochastically captures demand, ticket quantity and seat section choice. This demand model uses relevant factors that drive sports ticket sales. The modelers were keen to allow prices to decrease, and with the use of the model insights they achieved a potential revenue improvement of 17.2% through daily price re-optimization. This performance is similar to a strategy which would actually know the demand and could use hindsight to set an optimal fixed price.

Another paper is provided, just to give some overview of the options present in further research. Titled, Optimal Dynamic Pricing for Sports [Won (2008)]

Internal or External Deployment?

One interesting thing of this dynamic pricing project at AZ Alkmaar is whether the advice would be to deploy this internally or externally. There are parties who are specialized in the application of dynamic pricing in the sports industries and they have built sufficient name for themselves to suggest that profits are being made with them. However for larger clubs, that can take more capacity and hence sell more tickets, possibly for even higher prices, it is far more lucrative to apply dynamic pricing. It was briefly mentioned that QCue asks for 20.000 dollar on a yearly basis to apply dynamic pricing, this is excluding the time and manpower it takes to streamline the business between AZ and their data scientists. It would have been useful to have an idea of the additional profit they achieve, and whether this would outweigh the costs for AZ. For AZ, it would probably take an experienced data scientist about a year to build a good model internally. This would cost 50.000 euro's, from then on, it would take very little time to keep updating it, with all parties involved I would say a measly 5000 euro's on year basis (4 hours a week, incl. meeting time with Barend and the ticketing team).

One thing I am certain of, is that it will be used at some point at AZ Alkmaar and most of the bigger clubs in Europe. [Nufer et al. (2013)] came to the same conclusion based on an analysis of European football leagues. Along with the experience of dynamic pricing in America, they said that it can be only a matter of time before dynamic pricing will be applied by major football clubs in Europe. [Drayer et al.(2012)] presented a general overview on dynamic pricing in sports from the viewpoint of managerial aspects. The authors reviewed whether dynamic pricing can be applied in sports based on the criteria instituted in [Kimes (1989)] and [Kimes et al. (1998)]. Their conclusion is that the sports industry is a feasible one to deploy dynamic pricing strategies, with significant revenue potential. Back then it was argued that much more research would have to be done in terms of which algorithms should be used, currently much of that research has been performed with already successful algorithms and deployments published.

In 2016 there was a research conducted in which a total of 72 managers and executives from the four major North American professional sports leagues were surveyed. The sample consists of respondents from various sports such as the National Football League (NFL), Major League Baseball (MLB), National Basketball Association (NBA), National Hockey League (NHL), and Major League Soccer (MLS) [Bouchet et al. (2016)].

It was found that updating of ticket prices in the sports industries happened far less often compared to passenger air travel and in the hospitality industry. These industries have used dynamic pricing since the early 80's opposed to the 10's of this millennium. It appeared that 70% of the responders update their prices on weekly/monthly basis. 25% percent re-optimizes prices daily and less than 5% optimizes prices multiple times per day. Data showed that teams that automated prices in-house, more frequently optimizes prices compared to those that outsources to third-parties like Qcue. When it comes to automating dynamic pricing decisions, over 50% that responded shared that there is no automation, 25% outsource automation to a third party and 20% automate inhouse. There is clearly a lack of automation and especially in-house. This was considered an area of improvement, but it also shows that it is not a necessity to be able to automate prices daily in-house. The paper doesn't show whether more teams outsourced dynamic pricing, or performed it inhouse, but there were 15 teams automating via a third party, and 12 teams automating inhouse. Also interesting is the type of data these franchises used. Over 90% of the respondents uses inventory/supply data, 88% primary sales data (from the club) and 86% secondary sales data are used. This is promising for AZ as it again shows that secondary sales data is not completely necessary, although very useful.

In conclusion, when sports franchises use a dynamic pricing strategy correctly, they are almost certain to gain higher ticket revenues. It is said not to be a case of whether the major teams will deploy this, but rather when, and will it be internally or externally. It shows that both options are used heavily, and that re-optimization is rarely automated especially daily, which matches the manpower that would be likely available at AZ. Kemper argues that fans would also have a benefit in average ticket prices that may be lower, especially in the context of AZ and increased attendance. He concludes that in the event a club decides against increasing prices to the optimal level, even fractionally reaching the potential revenue by dynamic pricing would boost revenue in comparison to a fixed pricing. [Kemper et al. (2016)]

3 - Data Analysis

Available Data

There are three data sets available for the analysis during. One dataframe contains all the various events. An event contains this kind of information: An Eredivisie match played against Ajax in the AFAS Stadion, containing the date of the match, a consistent opponent name in Ajax. The events extracted from this dataframe are only the Eredivisie-matches played at home from season 2016/2017 till season 2018/2019.

All the individual transactions are available in another dataframe that contains an unique ID, the type of transaction (regular ticket or season ticket), row number, seat number, the seat section, the name of the item purchased etc. This information can be linked to the dataframe containing all the events .

Then there is also information on the behavior of season ticket holders. They have access to watch every Eredivisie home match by virtue of having a season ticket. However, they don't always actually make use of this and skip various matches. There is data to determine whether a season ticket was scanned at a particular match. This information will be used in this thesis.

The dataframe with the events starts with about 800 events including test events, parking tickets and normal matches. After the transformations it comes back to 51 events (17 x 3 home matches).

The dataframe with all the transactions reduces from just under 1.2 million transactions to 0.5 million transactions.

The dataframe that has all the scanned tickets, consists of about 0.4 million individual transactions that denote the specific chair for which a ticket was scanned.

Revenue analysis

Here is a plot of the revenue of regular tickets of all the matches. The peaks all belong to Ajax, PSV, and Feyenoord.

CONFIDENTIAL

Figure 1 - The revenue made by the sale of regular tickets over all 51 matches during the three seasons.

CONFIDENTIAL

Table 1 - The revenue by the sales of regular tickets, for each season separately.

These are the revenues per season for the regular tickets when it comes to the Eredivisie. There are no transactions available for the pre-season matches, KNVB Beker and European matches but a rough consideration would tilt these towards an extra CONFIDENTIAL euro's a season.

Before analysis is even performed, if a 10% increase would be achieved by dynamic pricing and smarter fixed pricing, that is about CONFIDENTIAL euro's extra on average for AZ. Which is a substantial amount, deducting the costs of maintaining the algorithms would still be a process worthwhile considering.

Some of the analysis and insights gained could also further optimize the revenue gained through season tickets. For the past 3 seasons, this was roughly the revenue achieved by season tickets in the Eredivisie.

CONFIDENTIAL

Table 2 - The revenue by the sales of season tickets over each season.

If a 5% increase could be made by smarter selling of season tickets by some of the pricing insights that will be gained in this section, that would drive about CONFIDENTIAL extra euro's roughly every season. It is now interesting to look at the insights that could be gained.

Location of the seats at AZ

The AFAS Stadion has undergone a transformation in the recent year. After the roof collapsed back in 2019, luckily with no injuries, the club has taken the lockdown as an opportunity to build a new roof and increase the capacity of the stadium.

Based on the data available containing matches from Eredivisie seasons 16/17 – 17/18 – 18/19, some analysis can be performed in this old setting. One of the more interesting things is to observe which seats in the stadium are most valued or are most picked. These seats may be priced differently in the future as the price currently might not reflect the desirability and thus the value of that seat. This information can be used for the initial pricing of seats, and may even result in certain price rules or price relationships where seats in the middle of the stadium should be priced higher than seats that are closer to the pitch for example.

An analysis is performed in this report regarding these locations in the stadium and also on a more aggregated level, that is, per seat section.

Current overview

Figure 2 - The stadium overview including the sections and stands.

Figure 3 - The stadium 3D overview, with the business seats and clear overview of the length of the sections.

These are maps of the AFAS stadium in which the various seat sections are visible. Certain sections are excluded of this analysis as these sections are either away sections, sections where people on the business side attend, or sections that will be changed; in short the entire Victorie-Tribune and section K and L (Away-fans) are not included in the data analysis.

The various locations are specified in this way;

The row numbers are in ascending order from bottom to top, meaning that the top row is the row with the highest number, and the first row starts with row number 1. Following, are the categories that are defined, in order to have variation but not too much granularity that makes drawing conclusions difficult.

- First Row (the first row of every seat section)
- Low (classified as rows from the height of the first row till the eighth row based on section J)
- Middle (classified as the ninth rows up until the twenty-second row based on section J)
- High (the rest of the rows)

Clubcard

For some of the matches it is mandatory to have a Clubcard, this is because these matches are considered higher risk for various reasons. In these matches, only fans that have a Clubcard and hence are registered can buy a ticket for such a match. The additional restriction is that they can only buy one extra ticket besides their own. These are limiting factors on the sales of tickets and this can also be seen in the data for these matches.

For these reasons some matches don't sell all tickets because of this restriction.

Selling Out

Even before looking at the popular locations within seat sections, it is of interest to know which sections are more popular than others and tend to sell out frequently. Another question to be answered is; how often is the entire stadium sold out? This information is important for determining a potential dynamic pricing model. Scarcity and overflow of tickets may determine the complexity of such models when relevant to the model of choice. If seat sections frequently sell out and it is observable that other sections then sell out significantly faster, it means that there is some relation between seat sections and these can't be modelled individually in such a dynamic pricing model. However if it's the case that customers only want to sit in certain sections for numerous reasons, it is then important to price those sections according to the willingness to pay of those customers specifically.

When determining which seat sections are sold out for all the fifty-one matches present in the dataset, the following problems arise. A section is rarely fully filled. There are often singular seats that can be found in between chairs as match-goers tend to come in groups of 2 or more. For this reason, often whilst there is some demand for the match and chairs are available, an event must be judged to be sold out as realistically no further purchases could be made.

This problem is accounted or by concluding that a section has sold out if 92.5% or 95% in a section are filled. [Di Domizio et al. (2015)]

Another issue in determining how often a match is sold out, is the problem that sometimes sections are purposely not filled because of safety. For this reason it is not easy to judge whether the demand trumps the supply. Often at times, it is also the case that matches that do sell out, sell out at the latest moment. Hence why a simple rule that incorporates inactivity of sales through time doesn't work.

It is visible on AZ's YouTube page that certain seat sections are purposely not filled during matches vs Ajax, PSV, Feyenoord and some other opponents, mainly the seat sections next to the away section. For this reason, some filtering will be done in order to look at the matches with stable and equal conditions.

One might argue that the matches versus Ajax, PSV and Feyenoord are essentially sold out. More interesting is it then to look at the other 42 matches in the dataset, where conditions are more stable and clear.

Some matches are tenser than other matches, this can also be seen in the behavior of the away fans. For this reason, the sections next to the away section are unpopular. The hard core fans of AZ sit opposite to the away section. The analysis that will follow this introduction, seems to hint that people like to watch a calm match of football without too much disturbance of hooligans in the away section. For this reason, section J, M and sometimes N are a bit less popular. Even stronger than that is the fact that they are sometimes purposely kept empty in order to avoid trouble with away fans. It is time-consuming to track for all matches whether this was the case, so sections J and M will also separately be considered in tracking whether these other 42 matches are sold out and their popularity. Section H is the section for family and friends of the players and staff, it is good to take note of this during the analysis.

Potentially, for a quite similar reason, sections U and V are considerably less popular than other sections as well. They are seated next to the AZ hard core fans, and also when looking at the location of those sections in the stadium, it is clearly one of the lesser favorite sections. It is not always clear whether these sections are intentionally kept empty.

Summary of results

Here is an overview of the occupancy of each seat section (area in table) over all the 51 matches. Match 50 (the last match, as the first match is match 0) is shown in the figure to give an example of the data preceding these final three columns.

All the areas are indexed and for all the seat sections the occupancy is calculated for the game. This occupancy is area-wise summed in 'Total Sold Out' next to match 50. The next column is named Total Matches, which stands for 51, the total amount of matches. Sold Out Frequency is then the former divided by the latter. This gives an indication of the most sold areas in the stadium.

Area	Area Capacity	8021	8032	Total Sold Out	Total Matches	Sold Out Frequency
H	349	0,84	0,99	44,52	51	0,87
J	1746	0,64	0,87	42,95	51	0,84
M	498	0,18	0,4	31,5	51	0,62
N	617	0,75	0,97	43,45	51	0,85
O	819	0,85	0,96	45,47	51	0,89
P	886	0,88	0,99	46,46	51	0,91
Q	562	0,96	1	49,11	51	0,96
R	886	0,92	1	47,39	51	0,93
S	819	0,94	0,99	47,47	51	0,93
T	614	0,93	0,99	47,51	51	0,93
U	535	0,83	0,99	44,03	51	0,86
V	541	0,74	0,95	39,74	51	0,78
W	461	0,93	1	42,75	51	0,84
X1	463	0,94	1	45,4	51	0,89
X2	432	0,84	0,99	45,12	51	0,88
X3	462	0,82	0,99	44,19	51	0,87
Y	406	0,88	0,97	43,67	51	0,86
Z	349	0,93	0,99	47,9	51	0,94
Overall	11445	0,82	0,95	44,37	51	0,87
Matches	11445	Heracles	PSV			

Table 3 - The seat sections and the summed ratio of their occupancy. Seat Section Q is the most popular and M is the least popular.

Area	Area Capacity	8021	8032	Total Sold Out	Total Matches	Sold Out Frequency
H	349	0	1	18	51	0,35
J	1746	0	0	20	51	0,39
M	498	0	0	10	51	0,20
N	617	0	1	24	51	0,47
O	819	0	1	25	51	0,49
P	886	0	1	30	51	0,59
Q	562	1	1	48	51	0,94
R	886	1	1	36	51	0,71
S	819	1	1	40	51	0,78
T	614	1	1	38	51	0,75
U	535	0	1	15	51	0,29
V	541	0	1	10	51	0,20
W	461	1	1	21	51	0,41
X1	463	1	1	26	51	0,51
X2	432	0	1	21	51	0,41
X3	462	0	1	15	51	0,29
Y	406	0	1	14	51	0,27
Z	349	1	1	44	51	0,86
Overall	11445	0,39	0,89	25,28	51	0,50
Matches	11445	Heracles	PSV			

Table 4 – The binary evaluation of whether a seat section is sold out.

In the appendix there are figures of bar charts and a link to interact with these bar charts. These bar charts show the overall score and the score for when only binary selling out is considered. Binary overviews of the popularity are available in the appendix, in section C. It is more interesting to know how often a section sells out, and the tables are sufficient for that.

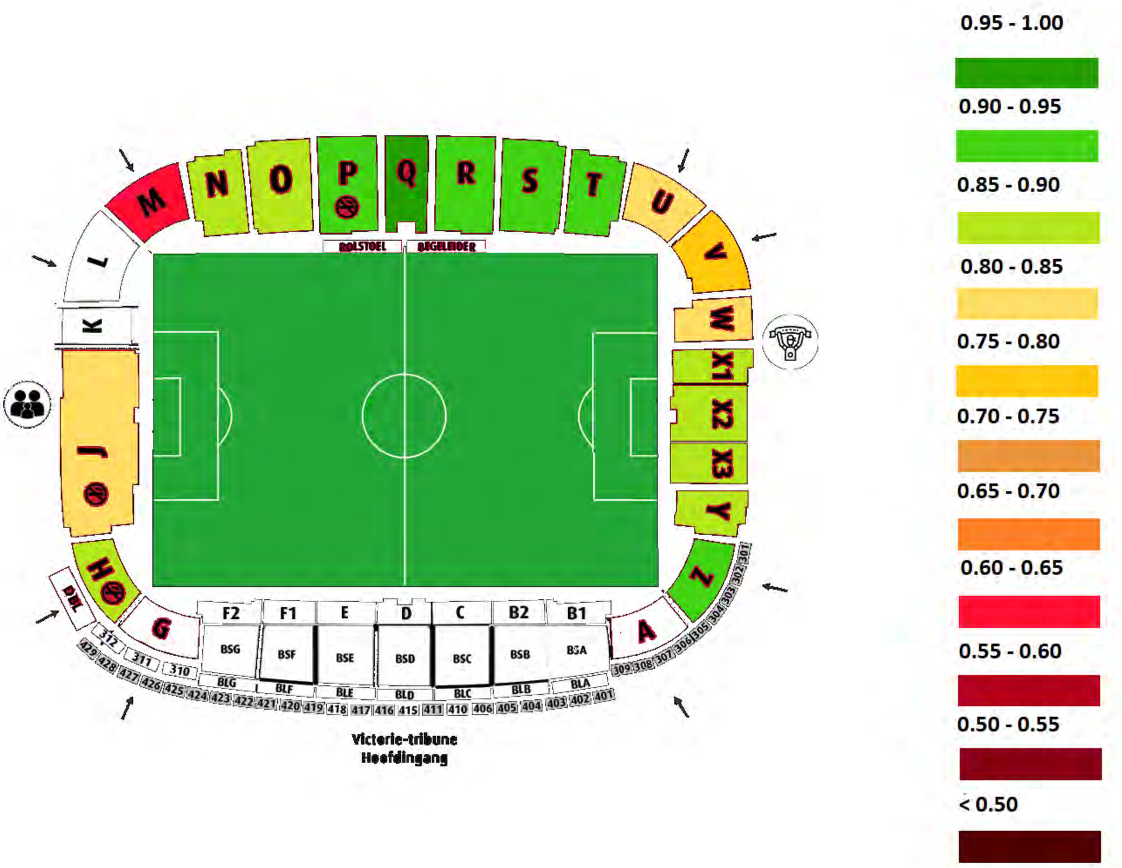


Figure 4 – Overview of the popularity of the seat sections in the stadium, with a legend (all matches).

This is the figure when we take into consideration the raw load percentages of all the areas. Areas M, V, W, J, U are clearly the worst performing but also have the problem that these areas are often censored. Section Q is clearly still the most popular section, and M the least. The sections Z, X1 and the entire Molenaar stand are easily the most selling-out sections.

For all matches, plots have been made of the ticket sales over time, containing the ticket sales of the various area's and the capacities. All these plots can be found in the Appendix and their titles start with; 'Cumulative with capacities' in the folder; 'Alle Wedstrijden'. This gives insight to the numbers that AZ is dealing with. The cumulative sales are split either in area's or in season tickets and regular tickets.

Splitting the matches

The matches are split in three different groups. The top matches are treated differently. These matches are said to be the standard sold out matches but for safety reasons some sections are purposely held empty. It seems to mainly happen for the sections J and M. Some cumulative plots are graphed in the Appendix as well, in order to show the graphs for these matches.

There was a very successful campaign with the supermarket Vomar Voordeelmarkt, one of AZ's main sponsors. During this campaign, loyal customers had the opportunity to buy tickets for Eredivisie games with huge discounts. This attracted a lot of people who had never visited the stadium before. These people however don't tend to sit where the hard core fans sit, so other sections that are usually unpopular, become very popular in this setting. The interesting thing is that for that matter, the hard core sections are less filled than the others which is caused by this promotion. That leaves us with the matches that aren't against Ajax, PSV, or Feyenoord and weren't matches during the Vomar promotion. This is a small part of matches that should have regular circumstances. However singular matches may still have restraints for various reasons.

Top Matches

-Percentage

	Area	rea Capaci	1148	1150	1152	4184	4219	4260	8014	7995	8032	otal Sold O	otal Match	Out Frequency
0	H	349	0,99	0,97	0,97	0,90	0,98	0,98	0,87	1,00	0,99	8,64	9	0,96
1	J	1746	0,94	0,76	0,84	0,52	0,91	0,97	0,50	0,97	0,87	7,28	9	0,81
2	M	498	0,69	0,30	0,61	0,08	0,65	0,82	0,09	0,81	0,40	4,45	9	0,49
3	N	617	0,99	0,87	0,99	0,75	0,98	0,99	0,65	1,00	0,97	8,17	9	0,91
4	O	819	0,98	0,97	0,99	0,87	0,98	1,00	0,84	1,00	0,96	8,58	9	0,95
5	P	886	0,98	0,96	0,99	0,92	0,98	1,00	0,86	1,00	0,99	8,68	9	0,96
6	Q	562	1,00	0,99	0,99	0,99	0,99	1,00	0,95	1,00	1,00	8,91	9	0,99
7	R	886	0,99	0,96	0,98	0,96	0,99	1,00	0,91	1,00	1,00	8,78	9	0,98
8	S	819	0,99	0,97	0,98	0,97	0,99	0,99	0,95	1,00	0,99	8,82	9	0,98
9	T	614	0,98	0,97	0,96	0,97	0,98	0,99	0,92	1,00	0,99	8,77	9	0,97
10	U	535	0,96	0,93	0,95	0,84	0,99	0,99	0,81	1,00	0,99	8,46	9	0,94
11	V	541	0,97	0,75	0,90	0,69	0,99	0,98	0,78	1,00	0,95	8,02	9	0,89
12	W	461	0,89	0,81	0,84	0,86	0,95	0,94	0,95	1,00	1,00	8,24	9	0,92
13	X1	463	0,99	0,91	0,93	0,94	0,95	0,96	0,98	1,00	1,00	8,67	9	0,96
14	X2	432	0,97	0,91	0,90	0,89	1,00	0,98	0,91	0,99	0,99	8,54	9	0,95
15	X3	462	0,98	0,94	0,92	0,92	1,00	0,98	0,87	1,00	0,99	8,58	9	0,95
16	Y	406	0,97	0,92	0,91	0,86	1,00	0,98	0,87	1,00	0,97	8,49	9	0,94
17	Z	349	0,97	0,99	0,97	0,97	1,00	0,99	0,97	1,00	0,99	8,87	9	0,99
18	Overall	11445	0,96	0,88	0,92	0,83	0,96	0,97	0,82	0,99	0,95	8,27	51	0,92
19	Matches	11445	Ajax	Feyenoord	PSV	Feyenoord	Ajax	PSV	Feyenoord	Ajax	PSV	8,27	51	0,92

Table 5 - The binary occupancy of the matches.

It is clear that for the top matches, very high loads are achieved. It is safe to assume and considered with the ticketing department that these matches are always sold out. For the safety of the fans, sections J, M, N, V and W seemed to be restraint purposely for some matches.

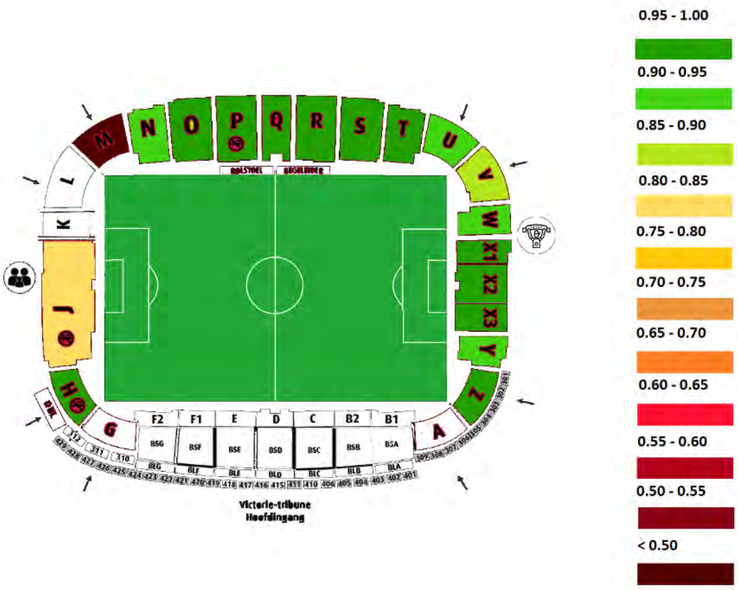


Figure 5 - The percentage wise occupancy for each seat section.

-Binary

When looking at the results for binary classifying whether a section is sold out, clearly section M is never sold out and sections J, H, W, V are often not sold out. For all the other sections, they are basically always sold out. Some had some club card restrictions through which it wasn‘t possible to get tickets for friends without a club card in the young adult section.

	Area	rea Capaci	1148	1150	1152	4184	4219	4260	8014	7995	8032	otal Sold O	otal Match	Out Frequency
0	H	349	1	1	1	0	1	1	0	1	1	7	9	0,78
1	J	1746	1	0	0	0	1	1	0	1	0	4	9	0,44
2	M	498	0	0	0	0	0	0	0	0	0	0	9	0,00
3	N	617	1	0	1	0	1	1	0	1	1	6	9	0,67
4	O	819	1	1	1	0	1	1	0	1	1	7	9	0,78
5	P	886	1	1	1	1	1	1	0	1	1	8	9	0,89
6	Q	562	1	1	1	1	1	1	1	1	1	9	9	1,00
7	R	886	1	1	1	1	1	1	1	1	1	9	9	1,00
8	S	819	1	1	1	1	1	1	1	1	1	9	9	1,00
9	T	614	1	1	1	1	1	1	1	1	1	9	9	1,00
10	U	535	1	1	1	0	1	1	0	1	1	7	9	0,78
11	V	541	1	0	0	0	1	1	0	1	1	5	9	0,56
12	W	461	0	0	0	0	1	1	1	1	1	5	9	0,56
13	X1	463	1	1	1	1	1	1	1	1	1	9	9	1,00
14	X2	432	1	1	0	0	1	1	1	1	1	7	9	0,78
15	X3	462	1	1	1	1	1	1	0	1	1	8	9	0,89
16	Y	406	1	1	1	0	1	1	0	1	1	7	9	0,78
17	Z	349	1	1	1	1	1	1	1	1	1	9	9	1,00
18	Overall	11445	0,89	0,72	0,72	0,44	0,94	0,94	0,44	0,94	0,89	6,94	51	0,77
19	Matches	11445	Ajax	Feyenoord	PSV	Feyenoord	Ajax	PSV	Feyenoord	Ajax	PSV	6,94	51	0,77

Table 6 - The binary occupancy of the matches.

Vomar Matches

-Percentage

	Area	rea Capaci	4256	4257	4258	4259	4261	4262	8026	8039	8029	8035	7998	8045	8001	8010	8004	8042	8007	8021	tal Sold O	tal Match	Out Frequency
0	H	349	0,94	0,76	0,87	0,97	0,97	1,00	0,77	0,81	0,90	0,97	0,95	0,86	0,87	0,92	0,86	0,85	0,94	0,84	16,04	18	0,89
1	J	1746	0,97	0,80	0,93	0,98	0,96	0,97	0,69	0,75	0,96	0,97	0,83	0,89	0,88	0,94	0,74	0,82	0,80	0,64	15,53	18	0,86
2	M	498	0,97	0,78	0,32	0,89	0,79	0,99	0,49	0,99	0,40	0,99	0,37	0,69	0,36	0,89	0,34	0,96	0,65	0,18	12,04	18	0,67
3	N	617	0,99	0,55	0,88	0,99	1,00	1,00	0,61	0,99	0,91	0,99	0,96	0,95	0,64	0,88	0,71	0,88	0,79	0,75	15,48	18	0,86
4	O	819	0,81	0,76	0,92	0,98	0,98	0,99	0,83	0,90	0,99	0,98	0,97	0,95	0,93	0,93	0,91	0,87	0,89	0,85	16,46	18	0,91
5	P	886	0,86	0,84	0,92	0,99	1,00	1,00	0,87	0,91	0,97	0,99	0,92	0,95	0,92	0,94	0,93	0,89	0,91	0,88	16,68	18	0,93
6	Q	562	0,94	0,94	0,99	0,99	1,00	1,00	0,94	0,96	0,99	0,99	0,98	0,97	0,97	0,97	0,97	0,97	0,97	0,96	17,52	18	0,97
7	R	886	0,90	0,86	0,96	0,99	0,99	0,99	0,91	0,93	0,99	0,98	0,95	0,96	0,94	0,95	0,97	0,94	0,94	0,92	17,06	18	0,95
8	S	819	0,90	0,89	0,95	0,98	0,99	0,98	0,91	0,94	0,98	0,98	0,94	0,96	0,92	0,96	0,94	0,94	0,95	0,94	17,05	18	0,95
9	T	614	0,93	0,88	0,95	0,98	0,99	0,99	0,89	0,96	0,98	0,99	0,96	0,97	0,94	0,94	0,95	0,94	0,95	0,93	17,12	18	0,95
10	U	535	0,77	0,77	0,81	0,99	0,98	0,99	0,79	0,82	0,93	0,99	0,89	0,94	0,88	0,95	0,88	0,82	0,83	0,83	15,85	18	0,88
11	V	541	0,68	0,62	0,72	0,99	0,96	0,99	0,76	0,78	0,83	0,98	0,86	0,82	0,79	0,81	0,81	0,79	0,80	0,74	14,71	18	0,82
12	W	461	0,80	0,77	0,80	0,98	0,97	1,00	0,91	0,95	0,89	0,92	0,94	0,92	0,90	0,90	0,95	0,93	0,95	0,93	16,40	18	0,91
13	X1	463	0,83	0,85	0,83	0,98	1,00	1,00	0,99	0,92	0,92	0,91	0,95	0,93	0,88	0,93	0,94	0,92	0,94	0,94	16,64	18	0,92
14	X2	432	0,86	0,84	0,89	0,99	0,97	1,00	0,85	0,93	0,94	0,98	0,94	0,95	0,89	0,91	0,91	0,90	0,96	0,84	16,54	18	0,92
15	X3	462	0,78	0,79	0,84	0,99	0,97	0,99	0,85	0,85	0,88	0,99	0,95	0,86	0,83	0,92	0,85	0,87	0,89	0,82	15,94	18	0,89
16	Y	406	0,81	0,78	0,81	0,98	0,97	0,99	0,82	0,88	0,89	0,98	0,92	0,89	0,88	0,88	0,89	0,89	0,85	0,88	15,98	18	0,89
17	Z	349	0,90	0,91	0,92	0,98	0,97	0,98	0,91	0,93	0,97	0,98	0,96	0,95	0,94	0,95	0,95	0,98	0,94	0,93	17,03	18	0,95
18	Overall	11445	0,87	0,80	0,85	0,98	0,97	0,99	0,82	0,90	0,91	0,98	0,90	0,91	0,85	0,92	0,86	0,90	0,89	0,82	16,12	51	0,90
19	Matches	11445	Roda JC	VVV-Venl	Sparta Ro	FC Gronin	Vitesse	PEC Zwoll	NAC Bred	Vitesse	PEC Zwoll	sc Heeren	De Graafs	Willem II	Excelsior	FC Utrech	FC Emmer	VVV-Venl	FC Gronin	Heracles /	16,12	51	0,90

Table 7 - The occupancy of the matches.

For the successful Vomar promotion, even matches versus smaller opponents achieved high overall sell outs or loads. Especially for the sections M when not restrained, N, when not restrained, O, P, Q, R, S, and T. For the smaller matches in the Vomar Promotion, sections U, V, W, X1, X2, X3 and Y were still often worse performing.

Here it can be seen that four matches had an average load of over 0.95.

Figure 6 - The percentage wise occupancy for each seat section.

-Binary

When it is looked at from the binary perspective, there are a few matches in which all or nearly all sections were sold out.

	Area	rea Capaci	4256	4257	4258	4259	4261	4262	8026	8039	8029	8035	7998	8045	8001	8010	8004	8042	8007	8021	tal Sold O	tal Match	Out Frequency
0	H	349	1	0	0	1	1	1	0	0	1	1	1	0	0	1	0	0	1	0	9	18	0,50
1	J	1746	1	0	1	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	8	18	0,44
2	M	498	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0	0	5	18	0,28
3	N	617	1	0	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	9	18	0,50
4	O	819	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	11	18	0,61
5	P	886	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	13	18	0,72
6	Q	562	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18	18	1,00
7	R	886	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	18	0,89
8	S	819	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17	18	0,94
9	T	614	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	16	18	0,89
10	U	535	0	0	0	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0	7	18	0,39
11	V	541	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	4	18	0,22
12	W	461	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	14	18	0,78
13	X1	463	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	14	18	0,78
14	X2	432	0	0	0	1	1	1	0	1	1	1	1	1	0	1	1	0	1	0	11	18	0,61
15	X3	462	0	0	0	1	1	1	0	0	0	1	1	0	0	1	0	0	0	0	6	18	0,33
16	Y	406	0	0	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	5	18	0,28
17	Z	349	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17	18	0,94
18	Overall	11445	0,39	0,11	0,44	0,94	0,94	1,00	0,33	0,61	0,72	1,00	0,78	0,67	0,44	0,78	0,56	0,44	0,56	0,39	11,11	51,00	0,62
19	Matches	11445	Roda JC	VVV-Venl	Sparta Ro	FC Gronin	Vitesse	PEC Zwoll	NAC Bred	Vitesse	PEC Zwoll	sc Heeren	De Graafs	Willem II	Excelsior	FC Utrech	FC Emmer	VVV-Venl	FC Gronin	Heracles /	11,11	51	0,62

Table 8 - The binary occupancy of the matches.

Interesting is to see again the discrepancy between areas that are symmetric in view of the half way line. It is visible that U is a lot more popular than M, S more popular than O (which is also the seat section for the old hard core fans), T is more popular than N. Seat sections J are far less popular than the seat sections at the Van Der Ben Tribune.

Remaining Matches

-Percentage

Area	rea Capaci	1151	1154	1155	1157	1158	1159	1160	1161	4078	4123	4127	4194	4216	4217	4218	4255	8017	7992	tal Sold O	tal Match	Out Frequen
H	349	0,96	0,87	0,80	0,82	0,87	0,82	0,81	0,85	0,75	0,83	0,71	0,94	0,80	0,72	0,77	0,87	0,79	0,82	19,81	24	0,83
J	1746	0,97	0,77	0,91	0,85	0,90	0,92	0,90	0,82	0,71	0,56	0,75	0,88	0,97	0,69	0,85	0,85	0,77	0,80	20,16	24	0,84
M	498	0,99	0,85	0,26	0,71	0,78	0,93	0,99	0,55	0,25	0,11	0,13	0,71	0,83	0,79	0,10	0,71	0,62	0,49	15,02	24	0,63
N	617	0,98	0,86	0,69	0,94	0,93	0,73	0,87	0,91	0,85	0,58	0,54	0,98	0,97	0,80	0,60	0,98	0,71	0,76	19,78	24	0,82
O	819	0,95	0,92	0,85	0,82	0,90	0,87	0,92	0,90	0,78	0,74	0,76	0,76	0,80	0,79	0,80	0,85	0,80	0,89	20,44	24	0,85
P	886	0,91	0,95	0,90	0,86	0,91	0,91	0,92	0,90	0,83	0,80	0,80	0,80	0,84	0,85	0,88	0,90	0,85	0,89	21,08	24	0,88
Q	562	0,97	0,98	0,96	0,92	0,97	0,95	0,98	0,98	0,90	0,94	0,90	0,88	0,92	0,93	0,96	0,98	0,93	0,95	22,71	24	0,95
R	886	0,91	0,95	0,90	0,88	0,92	0,90	0,92	0,92	0,87	0,86	0,87	0,84	0,87	0,89	0,89	0,93	0,87	0,91	21,51	24	0,90
S	819	0,92	0,92	0,89	0,87	0,92	0,90	0,92	0,92	0,87	0,86	0,86	0,84	0,89	0,86	0,91	0,91	0,91	0,93	21,60	24	0,90
T	614	0,93	0,92	0,87	0,86	0,91	0,89	0,92	0,89	0,87	0,89	0,86	0,88	0,87	0,89	0,90	0,94	0,92	0,93	21,63	24	0,90
U	535	0,91	0,87	0,84	0,84	0,87	0,84	0,87	0,85	0,75	0,73	0,75	0,73	0,74	0,77	0,75	0,77	0,82	0,84	19,70	24	0,82
V	541	0,94	0,74	0,68	0,66	0,68	0,69	0,70	0,69	0,60	0,57	0,65	0,84	0,62	0,60	0,67	0,67	0,73	0,74	16,98	24	0,71
W	461	0,76	0,74	0,70	0,72	0,75	0,72	0,74	0,75	0,70	0,67	0,74	0,71	0,73	0,77	0,79	0,79	0,90	0,99	18,09	24	0,75
X1	463	0,83	0,82	0,81	0,80	0,81	0,78	0,79	0,84	0,84	0,76	0,80	0,78	0,83	0,89	0,87	0,87	0,90	0,94	20,08	24	0,84
X2	432	0,94	0,83	0,81	0,79	0,82	0,78	0,82	0,81	0,82	0,82	0,83	0,79	0,85	0,86	0,88	0,85	0,90	0,95	20,05	24	0,84
X3	462	0,93	0,81	0,82	0,81	0,85	0,82	0,84	0,84	0,78	0,77	0,76	0,73	0,81	0,77	0,79	0,78	0,81	0,89	19,67	24	0,82
Y	406	0,95	0,80	0,83	0,79	0,81	0,81	0,84	0,79	0,71	0,71	0,70	0,69	0,71	0,74	0,73	0,80	0,86	0,89	19,21	24	0,80
Z	349	0,97	0,92	0,93	0,92	0,91	0,91	0,96	0,92	0,89	0,86	0,86	0,86	0,91	0,88	0,94	0,89	0,93	0,95	22,00	24	0,92
Overall	11445	0,93	0,86	0,80	0,82	0,86	0,84	0,87	0,84	0,76	0,73	0,74	0,81	0,83	0,81	0,78	0,85	0,83	0,86	19,97	51	0,83
Matches	11445	Sparta Rot	PEC Zwolle	Excelsior	ADO Den H	FC Groning	Roda JC	FC Twente	FC Utrecht	ADO Den H	NAC Breda	Excelsior	FC Utrecht	Willem II	FC Twente	Heracles A	sc Heeren	Fortuna Sit	ADO Den H	19,97	51	0,83

Table 9 - The occupancy for each seat section.

For the remaining matches, it can be seen that lower overall loads are achieved with a few outliers. The match vs Sparta Rotterdam was the best watched match, in which in Section W there was a far lower overall load. No match seemed to be sold out however.

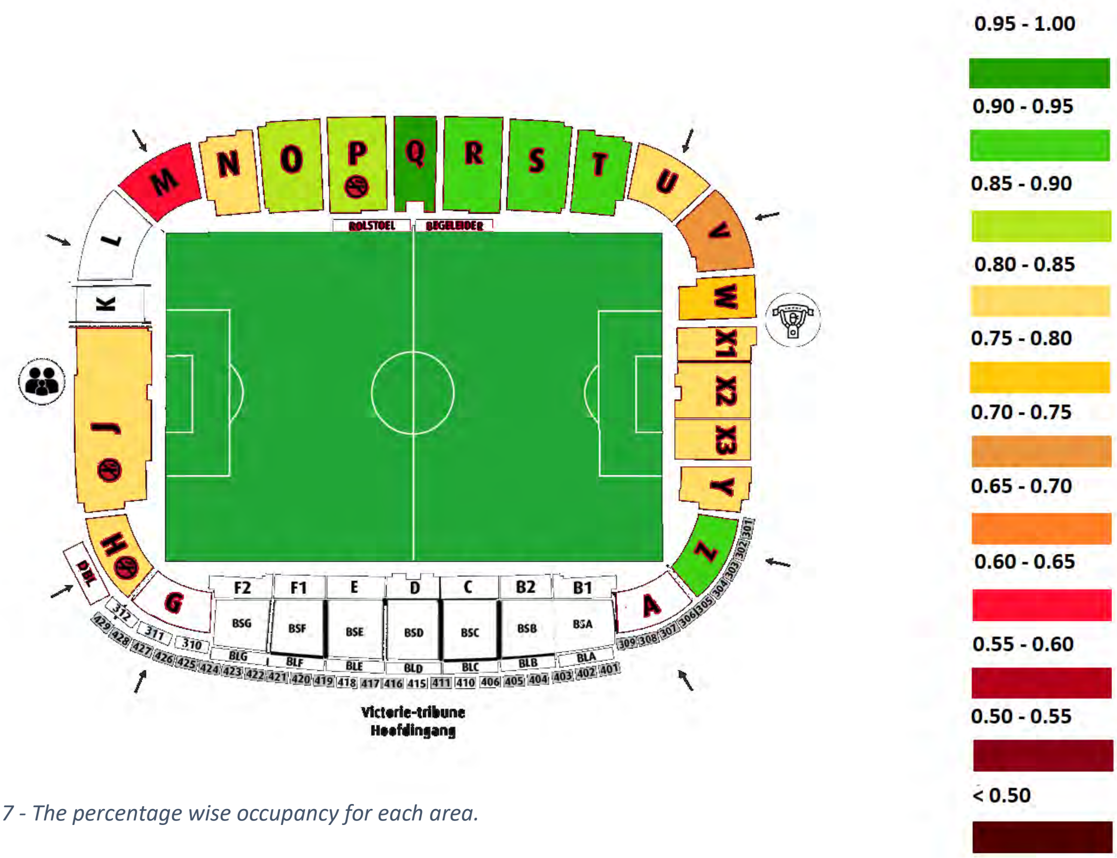


Figure 7 - The percentage wise occupancy for each area.

-Binary

For the binary case, it is clear that no match had a significant amount of sold out sections. Mainly Sections Q and Z were often sold out as often the case. The only match in contention for that would be the match against Sparta Rotterdam. Here about 90% of the sections were sold out and the only sections not sold out were the sections where the AZ die-hard supporters usually sit. It might be fair to say that match was sold out for the standard profile of a fan.

Area	rea Capaci	1159	1160	1161	4078	4123	4127	4194	4216	4217	4218	4255	8017	7992	tal Sold O	tal Match	Out Freque
H	349	0	0	0	0	0	0	1	0	0	0	0	0	0	2	24	0,08
J	1746	1	1	0	0	0	0	0	1	0	0	0	0	0	8	24	0,33
M	498	1	1	0	0	0	0	0	0	0	0	0	0	0	5	24	0,21
N	617	0	0	1	0	0	0	1	1	0	0	1	0	0	9	24	0,38
O	819	0	1	1	0	0	0	0	0	0	0	0	0	0	7	24	0,29
P	886	1	1	1	0	0	0	0	0	0	0	0	0	0	9	24	0,38
Q	562	1	1	1	0	1	0	0	1	1	1	1	1	1	21	24	0,88
R	886	1	1	1	0	0	0	0	0	0	0	1	0	1	11	24	0,46
S	819	0	1	1	0	0	0	0	0	0	1	1	1	1	14	24	0,58
T	614	0	1	0	0	0	0	0	0	0	1	1	1	1	13	24	0,54
U	535	0	0	0	0	0	0	0	0	0	0	0	0	0	1	24	0,04
V	541	0	0	0	0	0	0	0	0	0	0	0	0	0	1	24	0,04
W	461	0	0	0	0	0	0	0	0	0	0	0	1	1	2	24	0,08
X1	463	0	0	0	0	0	0	0	0	0	0	0	1	1	3	24	0,13
X2	432	0	0	0	0	0	0	0	0	0	0	0	1	1	3	24	0,13
X3	462	0	0	0	0	0	0	0	0	0	0	0	0	0	1	24	0,04
Y	406	0	0	0	0	0	0	0	0	0	0	0	0	0	2	24	0,08
Z	349	1	1	1	0	0	0	0	1	0	1	0	1	1	18	24	0,75
Overall	11445	0,33	0,50	0,39	0,00	0,06	0,00	0,11	0,22	0,06	0,22	0,28	0,39	0,44	7,22	51	0,30
Matches	11445	Roda JC	FC Twente	FC Utrecht	ADO Den H	NAC Breda	Excelsior	FC Utrecht	Willem II	FC Twente	Heracles A	sc Heeren	Fortuna Sit	ADO Den H	7,22	51	0,30

Table 10 - The binary occupancy for each area.

Very similar conclusions are found in terms of symmetry-wise comparisons. Normally, the seat sections are rarely sold out. These bars could be of use when looking at setting the prices also for season tickets. The most popular area's are made evident in it.

Overall

The symmetric pairs are listed here, making visual the effect of seat sections and their closeness to the away section. The 'winning values' are highlighted in yellow.

Section Side 1 (closer to away section)	Section Side 2 (further from away section)	Popularity Side 1 Preference; High to Low	Popularity Side 2 Preference; High to Low	Speed Side 1 PF; Low to High	Speed Side 2 PF; Low to High
H	Z	0.87	0.94	0.96	0.85
J	X2 & X3 & Y	0.84	0.89 & 0.87 & 0.86	0.98	0.97 & 0.98 & 0.98
W	X1	0.84	0.89	0.98	0.98
M	V & U	0.62	0.77 & 0.84	0.98	0.99 & 0.98
N	T	0.85	0.93	0.96	0.96
O	S	0.89	0.93	0.95	0.95
P	R	0.91	0.93	0.94	0.91

Table 11 - The scores for the popularity of the sections in terms of occupancy and speed. Sections are compared to symmetric comparable sections.

The Speed measure is measured by keeping track the number of days it took for an area to sell out, this is then divided by the total amount of time. The comparison between section H and section Z is not fair. Section H mainly consists of free tickets for family and friends. This explains why section Z is more popular and also sells out sooner than section H.

Section J sells out similarly to the opposing sections, also in terms of speed. This is interesting as very different types of people visit those sections. Section J is rather for families whereas sections X2, X3 and Y are more for the hard core fans and young adults. In its turn, section X1 is more popular than section W, it seems to be the main section for the hard core fans. Section W tends to get fuller when section X1 is already full.

Clearly sections V and U are considerably more popular than the then section M despite they are symmetrical in location. The influence of the away section can be easily identified. However the truncation of seat sales is also incorporated in this comparison. A further analysis should be done of the popularity when there is no interference.

Section T, S, R are more popular than respectively N, O, P. This can easily be explained as the latter sections are closer to the away sections. However it may well be that some of the sections had censored ticket sales as well by interference from the ticketing staff. This needs to be identified in order to be able to draw sound conclusions.

When the pairwise dualities are moved away from, the following conclusions are made based on these results. Here the average ratio of the occupancy of tickets in an area is divided by the total capacity of that area. There are a few interesting insights to find here especially when we incorporate background knowledge. Section Q is #1, this is predicted by the ticketing department as well. This section is exactly on the same line as the half-way line, This is deemed the most preferable section to watch a football match. Other than that, section Z scores surprisingly high. This section is very comparable to section H which performs 6% worse. This might have to do with the fact that section is relatively closer to the away-fans, which seems to be a recurring factor for the popularity of the sections.

T, S, R, P, O are all next to each other and have a spread of 4% which is to be expected, section N is considerably lower with 85% compared to their mean of 91%, this is likely because of its proximity to the away fans section. We see that section X1 far outperforms section W, which is surprising as according to the knowledge present, these sections should be similar. Section X2 and X3 which are next to X1 perform reasonably well. It might be that they have a synergy ongoing whereas Section V and W have a negative synergy.

Section U also scores considerably better than Section V whilst they are neighbors, there seems to be something going on at sections W and X1. It might be that the view is a bit better at section U, and section X1 has very good atmosphere and Sections X2 and X3 bounce of that synergy. What is very relevant is the prices as well of course. It could be that certain sections are prices wrongly for their quality or attractivity. Here we find the results when we incorporate speed of selling out similarly as we have prior in this part of the report. It would be interesting to determine a way of penalizing a section that doesn't sell out and penalize them by the amount of seats that are still left free. Comparing them in a platonic manner, we need to stress that the variance between the values is very different for both methods and hence the rank isn't easily transferable from one method to another. Deeper analysis would have to be performed.

One thing that is noteworthy, is that the rank of sections X1, X2 and X3 is much lower in terms of speed compared to the occupancy of the section. This might mean that these sections get filled very last-minute, fans buying tickets late during the booking horizon. Furthermore we see that it is opposite for section N. Which sometimes gets sold out very quickly, this section might be very popular during certain matches and not so much during others. It would be interesting to find out for which matches this holds. Other than that, the ranks of the sections remain fairly consistent in both methods. It is clear that without factoring in price as a factor, the sections on the Molenaar-Tribune are most popular. These are also the most expensive sections coincidentally. MUVJWX1X2X3 are clearly less popular.

Location Popularity

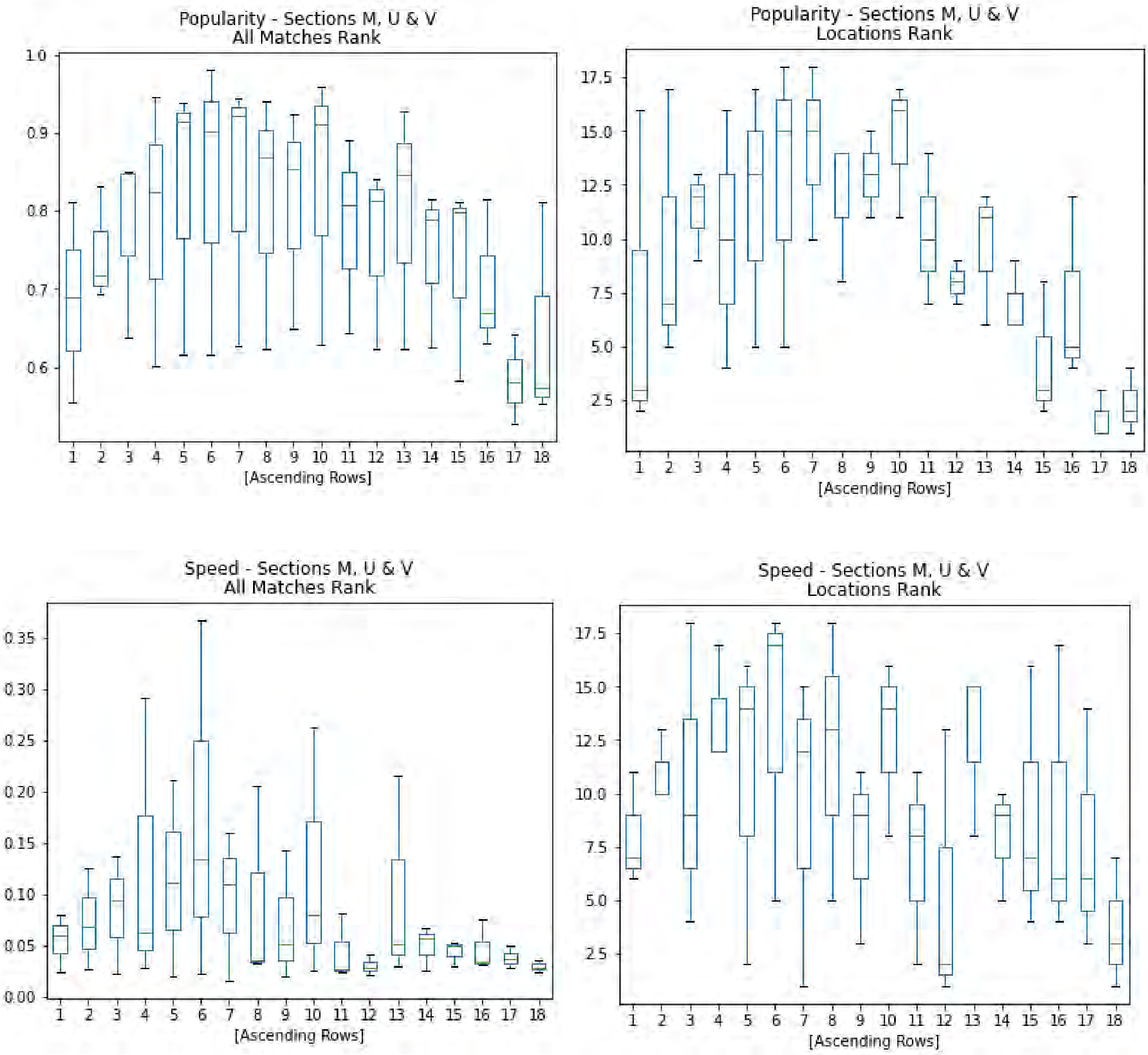


Figure 8 - Boxplots showing the popularity in seat sections M, U, V. Measured in occupancy and speed.

When looking at the sections M, U, V it becomes apparent that the higher rows are least popular. The most popular rows are the first 10 rows. These are the seat sections in the corners and not close to the ground. It is about the same height as the middle seats of other seat sections. Later than that there is a drop in popularity. It would be wise to incorporate this information in the segmentation of the stadium.

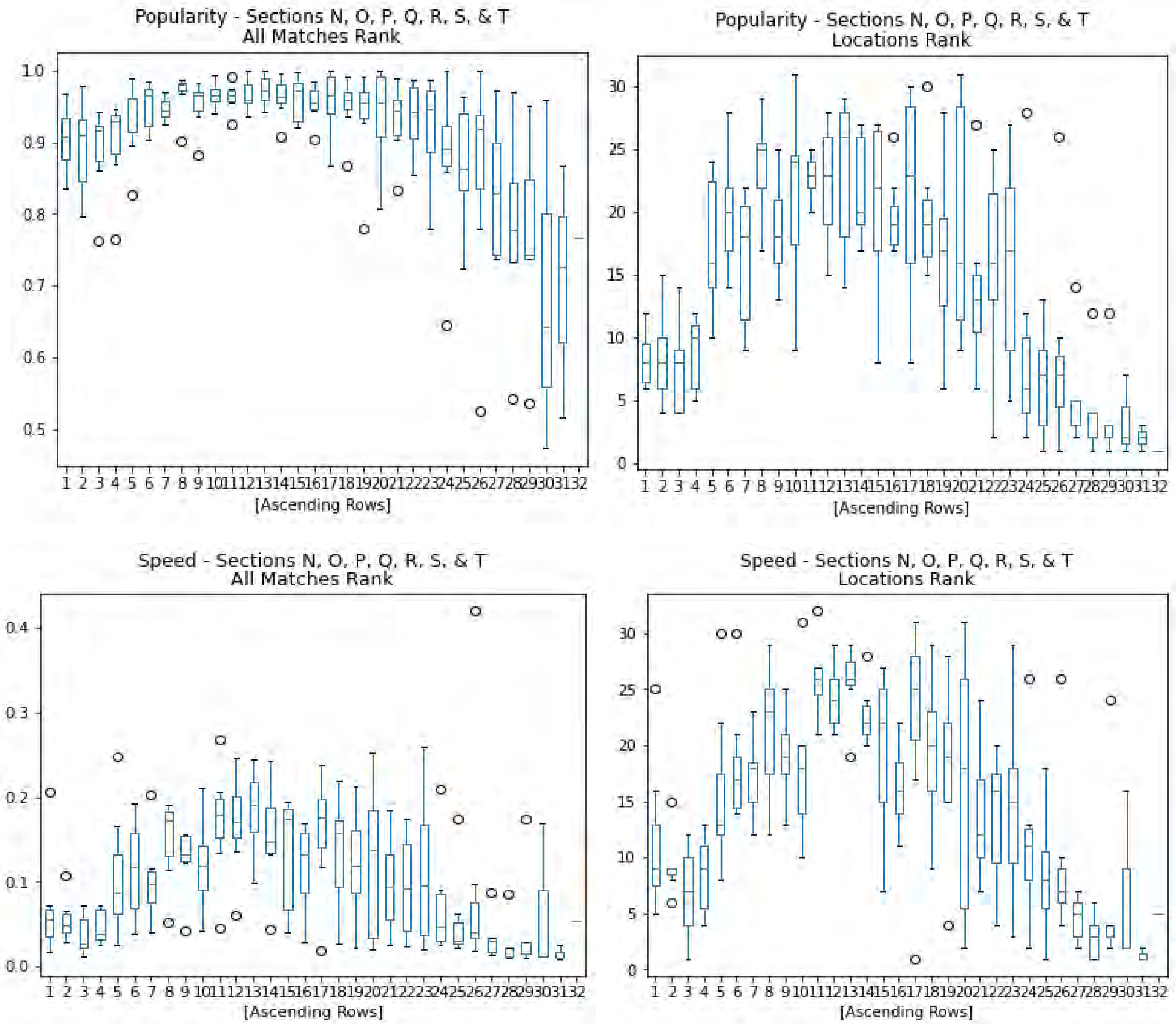


Figure 9 - Boxplots showing the popularity in seat sections N, O, P, Q, R, S and T. Measured in occupancy and speed.

For the seat sections on the Molenaar-Tribune, it is apparent that the middle of the seat section is most preferred. It has the highest occupancy and significantly higher than the high seat sections. The lower seat sections are still quite preferable as they are in the middle of the stadium and provide a good view on the stadium. It fits the general idea that the stadium locations should be classified in low, medium and high locations.

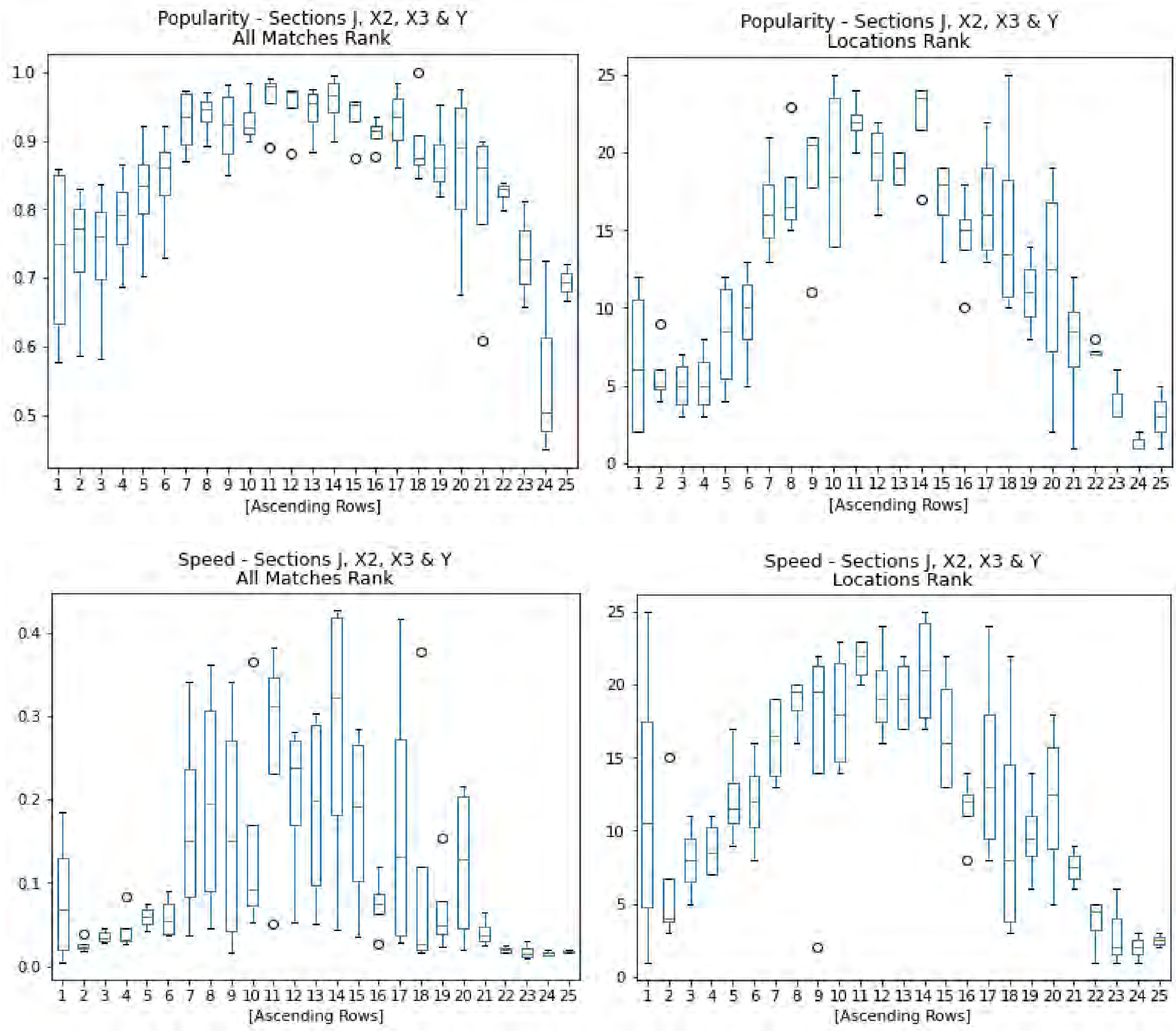


Figure 10 - Boxplots showing the popularity in seat sections J, X2, X3 and Y. Measured in occupancy and speed.

For the sections that are on the opposing ends, the Alkmaarder Hout-tribune and the Van Der Ben-tribune, it is seen that the seats in the middle are again preferred over the lower and higher seats in the old stadium setting. The results are consistent for both the speed in which the different rows are solds, as for the amount of tickets are sold. This further enhances the view that the stadium locations should be split in three.

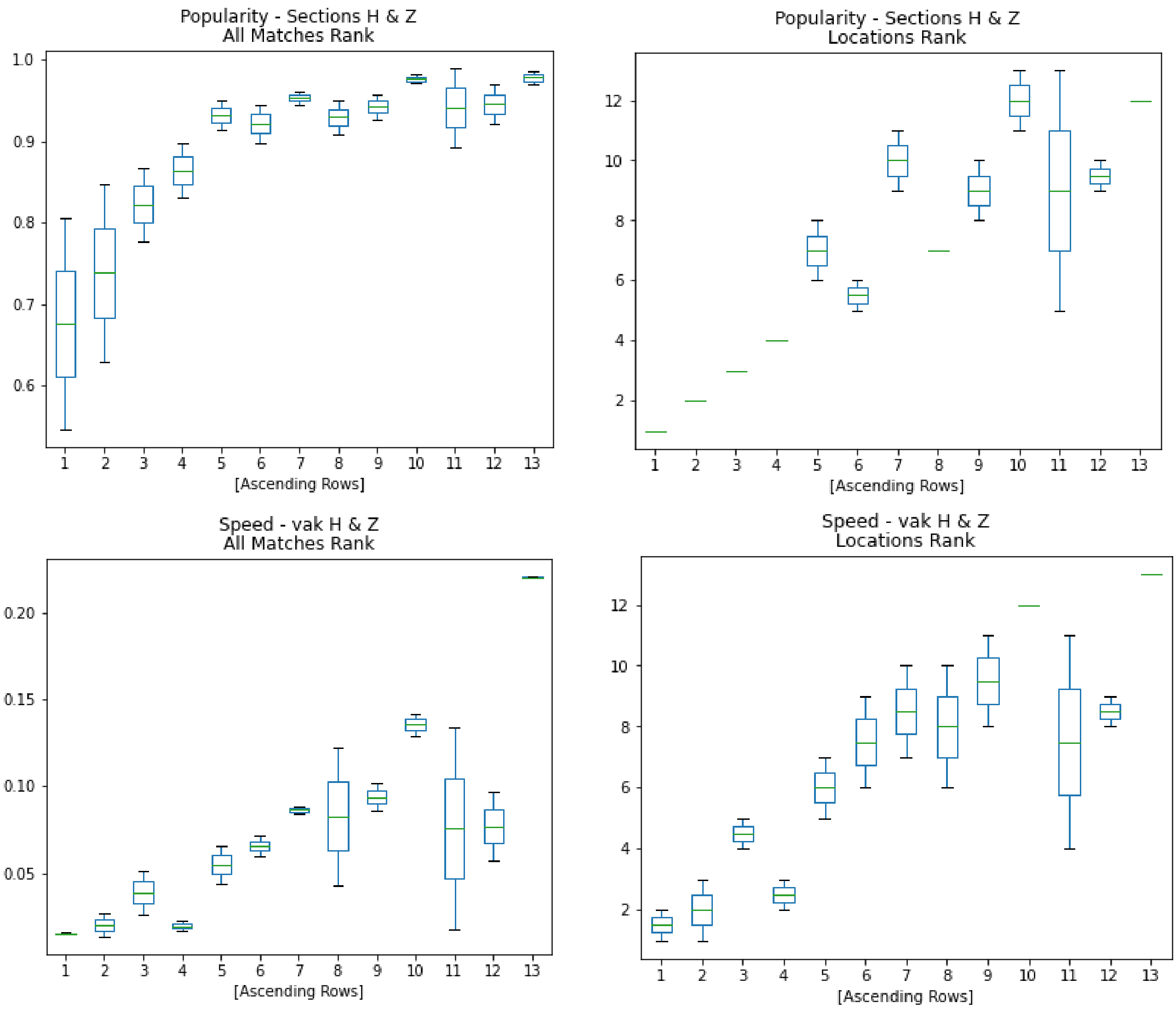


Figure 11 - Boxplots showing the popularity in seat sections H and Z. Measured in occupancy and speed.

The seat sections H & Z are in the corners at the Victorie-Tribune. These sections are just in front of the skyboxes and give the feel of a sky-box experience. There used to be heaters as well at these seat sections. This makes it that the higher seats are preferred to the lower seats. The higher one goes, the more comfort one enjoys. That makes these seat sections special and the current situation should be incorporated in any decision making regarding these seat sections.

Capturing overflow between Sections

Potential dynamic pricing algorithms will be discussed in a later section. However to consider which model will be used for the dynamic re-evaluation of prices, it is important to know to which degree there is scarcity. Are matches or sections often sold out? This is an important aspect of some dynamic pricing algorithms. This question was answered in the previous section, in the coming section, the presence or absence of overflow will be discussed.

The reason why it is important to capture overflow of demand is because of the reason that in some models, if prices would be priced dynamically on section-level, sections are then seen as variables or states. If these sections are co-dependent, e.g. if section A is full, demand for tickets in section A moves to section B, then these models become more complex as the states or variables have to be multi-dimensional. It is best to at least determine this before that model is attempted, the complexity makes it a model not viable to implement in the remaining period of this project.

Through some data analysis it is possible to calculate on a daily basis when a seat section is sold out. A seat section is set to be sold out if about a certain percentage of the seat section is sold out. Mostly 95% will be used, when this percentage is reached there are mostly individual empty seats scattered through the seat section.

Below is a print of what is calculated, all the seat sections have capacities and an aggregate is recorded for each day. This will then determine whether a seat section is sold out.

df_sold_out - DataFrame

Index	Area	ea Capaci	'OfSelling	Count	)ccupancy	Indicator
0	H	349	204	345	0.9885	1
1	J	1746		1517	0.8688	0
2	M	498		199	0.3996	0
3	N	617	238	598	0.9692	1
4	O	819	195	787	0.9609	1
5	P	886	192	874	0.9865	1
6	Q	562	69	562	1	1
7	R	886	169	882	0.9955	1
8	S	819	157	808	0.9866	1
9	T	614	181	607	0.9886	1
10	U	535	207	531	0.9925	1
11	V	541	252	514	0.9501	1
12	W	461	210	461	1	1
13	X1	463	217	463	1	1
14	X2	432	243	429	0.9931	1
15	X3	462	224	459	0.9935	1
16	Y	406	206	392	0.9655	1
17	Z	349	42	345	0.9885	1

Table 12 - Table of the areas, their capacities, the amount of tickets sold at one point, and their occupancy.

First of all, what is then calculated is the weight of tickets sold of each seat section as a ratio of all tickets sold that day. If a seat section gets full, its ratio is thought to decrease as less tickets will be sold in that seat section relative to the total sales. With this, the difference between each day can be calculated, this can also be averaged out (smoothed) over a given window so that changes in the values can be met. The interesting thing is to note if there are significant changes in these ratios when seat sections get sold out. Tables that show what happen can be found in the appendix – section B. this is depicted in the next figures and is also made graphically for a match.

Area	level_1	0
Y	W	0.07143
M	Y	0.05542
Y	M	0.04343
M	U	0.03399
N	S	0.03075
M	V	0.02367
W	W	0.01852
J	H	0.01813
W	X	0.01642
V	N	0.01477
O	Q	0.01424
W	Y	0.01342
V	O	0.01203
U	T	0.01135
U	S	0.01134
N	P	0.01073
M	Z	0.0107
X1	J	0.01068
W	M	0.01058
U	N	0.01005

Table 13 - This table shows the highest correlation values.

This figure shows that when seat section Y gets sold out, the ticket sale in seat section W increases by 7%. These statistics are still indefinite because the daily random change exceeds these values.

It would be interesting to note whether there is some mathematical framework to find in order to determine statistical significances.

Figure 12 - This is a plot of the cumulative tickets sold for a specific match against Ajax, these include the season tickets which are assigned to the seat sections.

The cumulative ticket sales can also be made graphically for each match. These plots can also be accessed through a Zip File that is linked in the Appendix, which is interactive. After looking at these plots some overflow is captured, but it is very limited. Mainly in X2 and X1 it is visible that X2 sells more when sales in X1 drop. This might mean that there is a preference for X1 and then X2. These seat sections are when the hard core fans sit and they are exactly next to each other.

4 – Logistic Regression

Analysis of ticket attractivity

In dynamic pricing it is often important to find the price elasticity of the match tickets in order to determine how much the price should change in order to maximize the expected revenue. The price elasticity is difficult to be determined because of the lack of variation of tickets under similar circumstances. No two opponents are the same, and every match is different to the other. However by attempting to group matches together that have used different prices, it was thought during the internship that a proxy for the price elasticity could be found. In hindsight, it would have been more useful to go for demand functions, but the model will still be explained in this section. This model could be improved or altered in order to retrieve more information.

In the literature there are also various indications given for price elasticity in football. Generally, demand of tickets in the football industry are said to have low price elasticity meaning that football clubs could raise their revenues by increasing ticket prices.

Given the situation of the available data at AZ Alkmaar, an attempt will be made in order to capture the price elasticity.

Data available

- Dataframe containing;
    - Transaction, transaction number, event name, event data, section, row number, seat number
- Dataframe containing;
    - Events, EventDate, The Competition, The Opponent
- Dataframe containing;
    - 51 dataframes with realized ticket transactions, scan of a transaction on the matchday

These data frames can be merged into a data frame with every transaction containing the various features as; transaction number, event name, event date, section, row number, seat number, the opponents name and competition. For the season ticket holders, a separate data frame can be made with indications on whether the season ticket holder was present (1 if present, 0 if absent). This is also described in the beginning of the Data Analysis. Other than these features, regressors like the Day of the Week, the Status of the opponent, The Ranking of AZ and the Opponent are incorporated in the data frames. In this manner, it is easy to include the regressors in a logistic regression that scores the influence that these regressors have on the probability of a season ticket holder showing up.

Grouping matches together

There is data available of 3 seasons of Eredivisie matches, the seasons 16/17, 17/18 and 18/19. This data contains all the home games in the Eredivisie over that time, both the transactions data and the data of the season ticketholders and whether they did or did not scan a ticket. Season ticketholders have purchased a season ticket at the start of the season and have therefore paid an initial and final fee. This ticket gives the purchaser the access to visit all the home games in the domestic competition. Typically, season ticketholders show different patterns of behavior for different type of games and additional factors. They decide whether to consummate their right to visit the match based on their personal preferences, it seems that the average AZ-season ticket holder would rather visit a match vs Ajax or PSV than Excelsior let's say. It is easily observable that different groups of matches have different numbers of absentees. This can be used as a way to rank matches based on their attractiveness, without incorporating price as a factor. Each match receives an attractivity score based on the average level of season tickets that show up for the match. This is predicted by certain features to matches that could have some predictive value. This is incorporated into a logistic regression and the results of will be discussed, but first some of these regressors are described in the next table.

Regressor Name	Description of the regressor
DayInWeek	This is a regressor that captures the type of day in the week, whether the match was played on a Weekday, Saturday or Sunday
Under5Degrees	This is a regressor that indicates whether the temperature was below 5 degrees.
Status	This is a regressor with manual determined statuses of clubs regardless of the current ranking of a team in the league
Speelronde	This is a regressor that captures the #th home match in the Eredivisie. It gives a measure of time.
Standing AZ	This is a regressor that gives the current ranking of AZ
Standing Opponent	This is a regressor that gives the current ranking of the opponent of AZ
Standing Difference	This is a regressor that gives the difference in the current ranking of AZ and their opponent
Within Seven Days	This is a regressor that indicates whether in 7 days before or 7 days after the match, there has been a (top) match played at home in the AFAS Stadion
ClubkaartVerplichting	This is an indicator for whether clubcards are compulsory for ticket purchase
Ajax	This is an indicator for whether the match is played against Ajax
Feyenoord	This is an indicator for whether the match is played against Feyenoord
PSV	This is an indicator for whether the match is played against PSV
Season	This is an indicator for which season the match is played in

Table 14 - The regressors and their description.

Before looking at the regression, the individual regressors will be looked at individually in order to see how they correlate with the behavior of the season ticket holders and the amount of tickets sold.

Data Analysis on the regressors

DayInWeek

When a boxplot is made of the amount of tickets sold on Saturday's, Sunday's and Weekday's we see that Week Day's are least popular with the lowest ceiling and lowest average. This is in alignment with what would be expected.

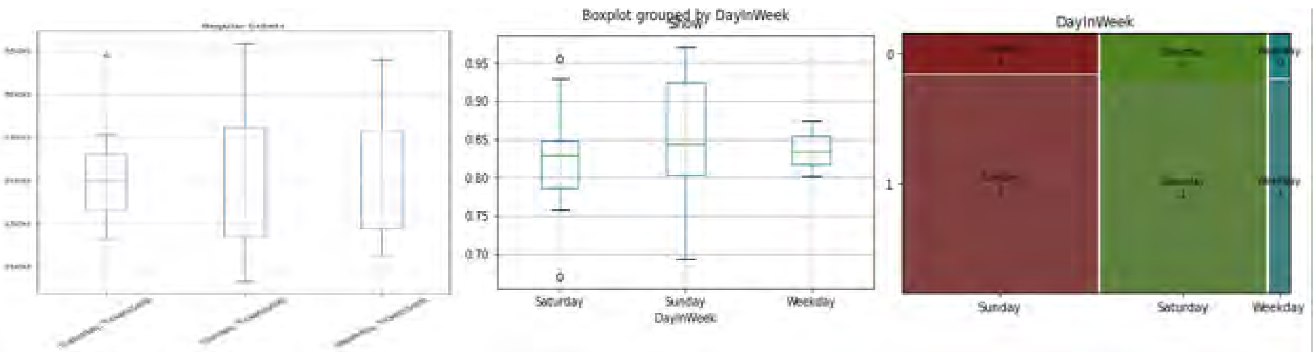


Figure 13 - Two Boxplots of the regressor DayInWeek pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

Important to note however is that there were only 3 matches played on weekdays, this is also easily observable in the 3rd plot which contains the ratio of the data available. Saturdays seem to sell more, but the ceiling is also lower. When it comes to the season ticket holders, Sunday seems more favorable for them to go to the match. Saturday then even seems to do it less well than Weekdays.

Here is a depiction of all the matches and the day during the week. An indication can be directly made which days perform well in this model.

	Venue	AZ	Tegenstander	EventDate	DayInWeek	TicketsSold
43	AFAS Stadion	AZ	Vitesse	2018-04-18	Weekday	3400
15	AFAS Stadion	AZ	FC Twente	2017-11-24	Weekday	1753
26	AFAS Stadion	AZ	Heracles Almelo	2019-04-23	Weekday	1134

These are the only 3 matches during the week in the 3 seasons, the Vitesse match which was a Vomar Action, massively skews the results and the results are therefore neglected.

Under5Degrees

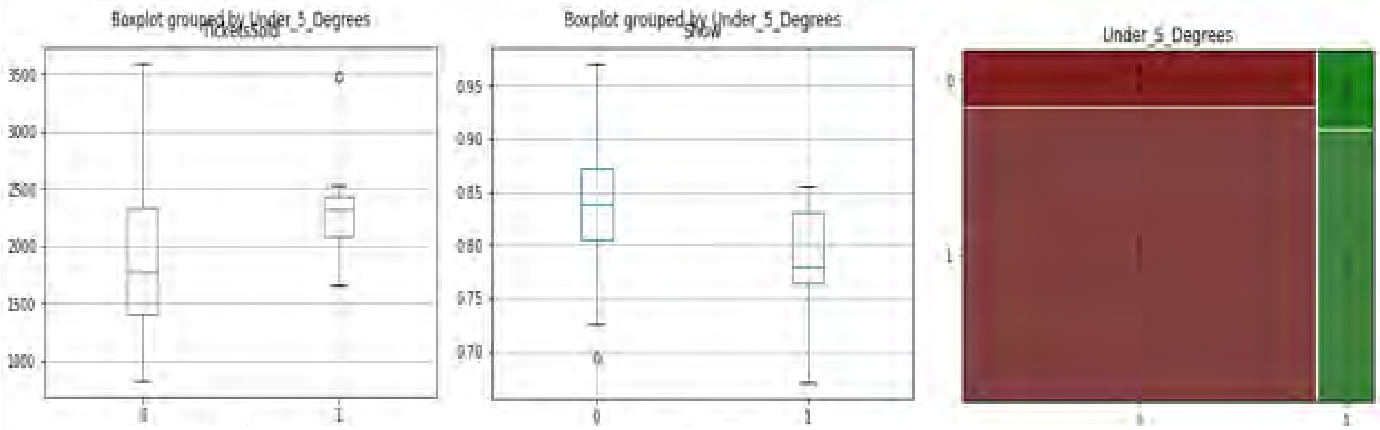


Figure 14 - Two Boxplots of the regressor DayInWeek pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

Out of the 51 matches, 7 matches were below 5 degrees. It is visible that less season ticket holders show up for the colder matches, but there were still many tickets sold for these matches. Long term it would be wise to do a longer analysis on which weather circumstances impacts season ticket holders. It might not be the average daily temperature but rather the average temperature of feeling during the day or evening. 5 degrees might not be the optimal cut-off point.

Status

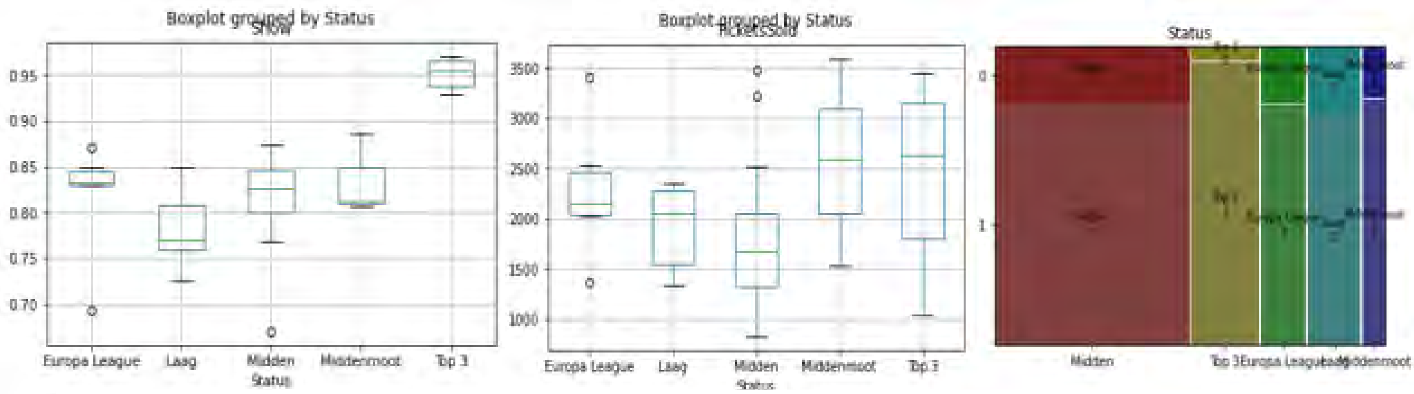


Figure 15 - Two Boxplots of the regressor Status pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

When it comes to the status' of clubs, it is easily observable that the traditional Top 3 live their own lives. Europa League, Midden, Middenmoot are all very comparable to one another and then the Lower ranked teams are visibly less popular again. In terms of status it then seems best to divide all the opponents in 3 different sections. In Top 3 teams, Average Teams, and then Low classed teams.

Speelronde

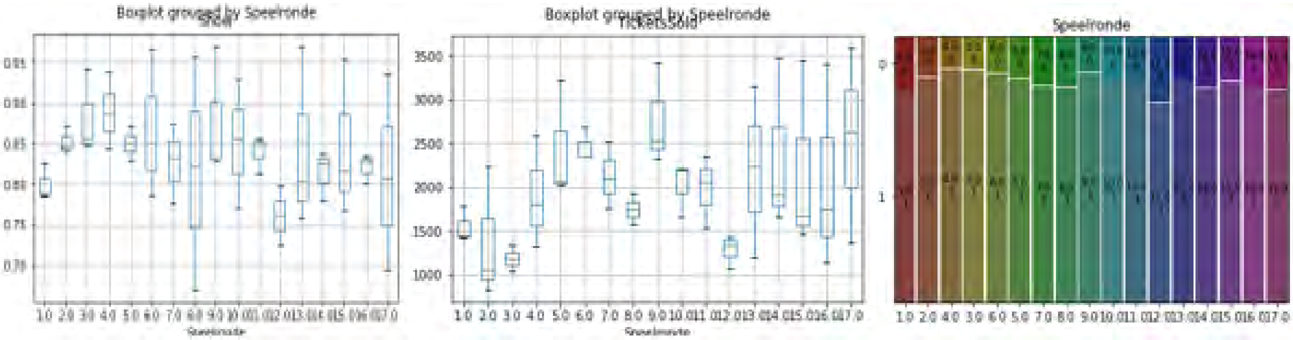


Figure 16 - Two Boxplots of the regressor Speelronde pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

The gameweek is included as a factor that incorporates time and season. Earlier in the season, the trophies aren't dealt, but rather later in the season. What can be observed from this, is that the first few weeks are a bit less popular, but as the season goes on, it becomes a little bit more popular over time. Which may drop again depending on probably whether AZ is in contention for the top matches.

Standing AZ

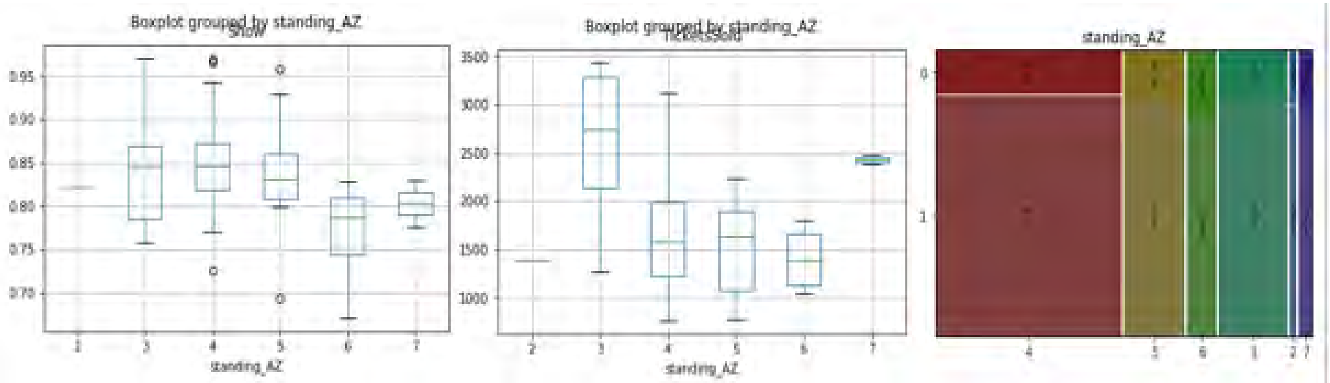


Figure 17 - Two Boxplots of the regressor StandingAZ pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

It seems like that the standing of AZ is not the best indicator of the popularity of match tickets. Only when AZ is ranked 6th or 7th is it observable that season ticket holders are less likely to show up, however still for those singular matches, many tickets may be sold.

These rankings may also be categorized to;

- 1-3 | Top – 3
- 4-6 | Europees
- 7 – 15 | Middenmoot
- 16-18 | Degradatie

In which the first few matches of the season rankings are manually determined, by looking at the average position of AZ and the opponent in previous seasons.

Standing Opponent

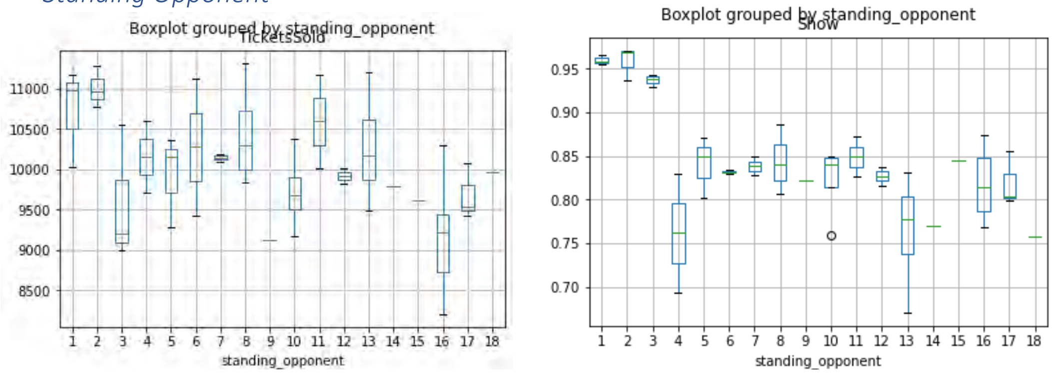


Figure 18 - Two Boxplots of the regressor Standing Opponent pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

These plots show that the opponents that occupy first, second and third place are clearly most popular with the most tickets sold or season tickets that are present for the matches. Other than that, there is a big middle section and only about the bottom 5 teams are considered less popular.

Standing Difference

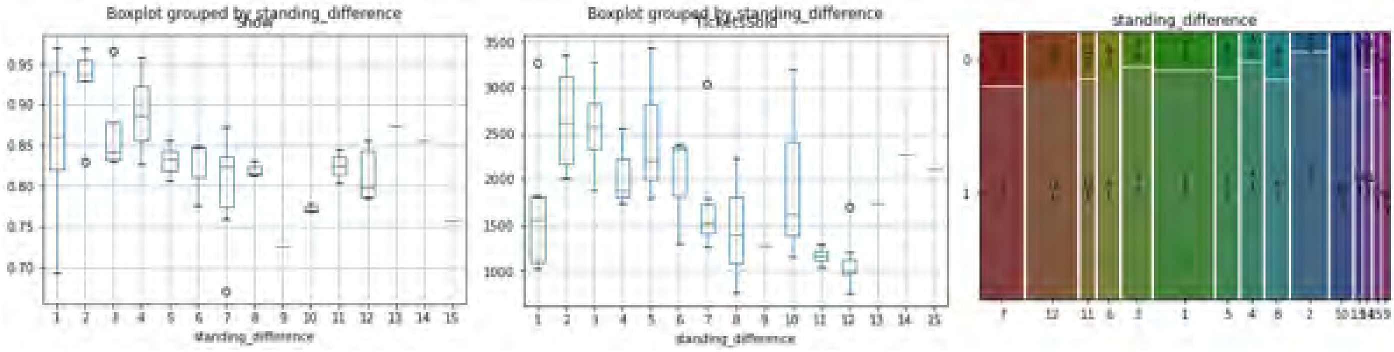


Figure 19 - Two Boxplots of the regressor Standing Difference pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

When it comes to the difference in standing of AZ ad the opponent, it is visible that for opponents which stand closer to AZ, more tickets are sold and season ticket holders tend to show up more often. These numerical levels can also be categorized in minimal difference (smaller or equal than 2), a small difference (between 3 and 5) and a big difference (bigger than 6) in ranking.

One of the issues with this metric is that this metric particularly becomes useful at a latter stage in the season because then rankings are a better reflector of the actual standings of those teams. The idea is that when AZ is playing a tight match, this should be more attractive because there is more at stake during such matches and they are deemed to be of higher quality as well

Within Seven Days

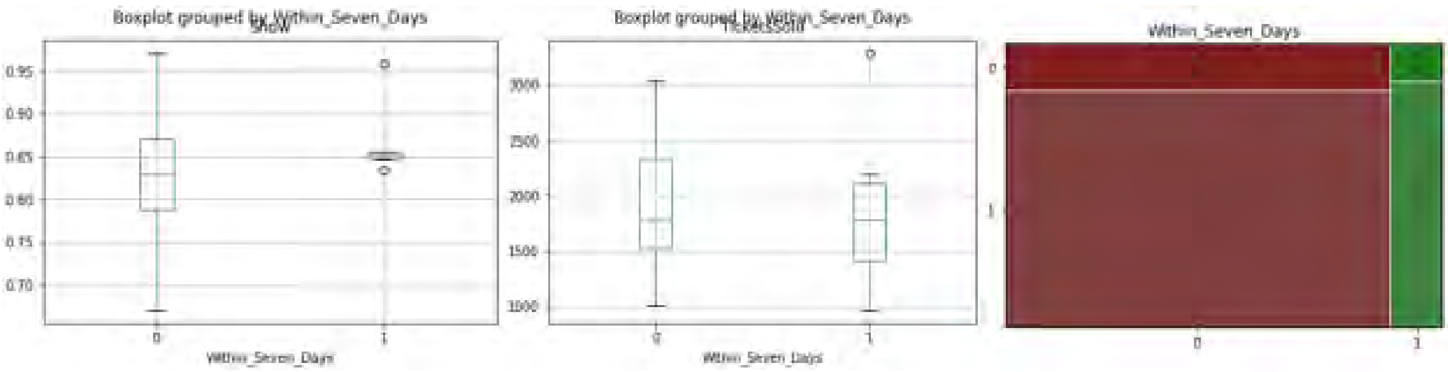


Figure 20 - Two Boxplots of the regressor Within Seven Days pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was

From the 3rd plot it is clear it doesn't happen very frequently that AZ plays 2 big home matches within 7 days to each other. When this happens, there seems to be very little observed effect to the number of tickets sold or the variable presence of the season ticket holders. This occurred for some of the more enticing opponents such as Feyenoord, Vitesse, sc Heerenveen making it difficult to determine the causality. The full list shows in the appendix.

Season

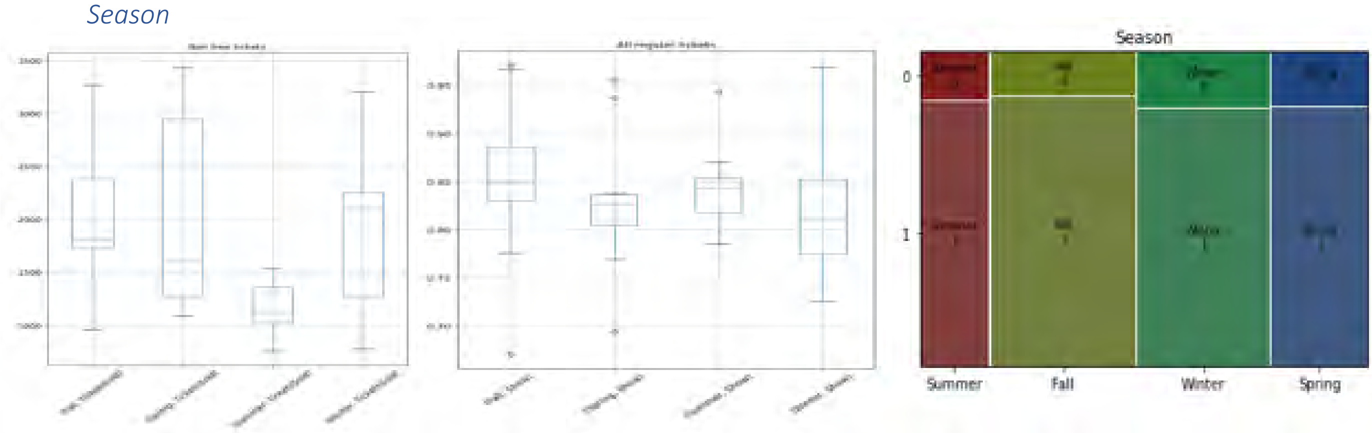


Figure 21 - Two Boxplots of the regressor Season pertaining to respectively regular tickets and to the degree of season ticket holders showing up. Right is a mozaic plot that shows the proportions of the regressors and if a season ticket was present (1) or not (0).

For the season ticket holders, it seems that spring and winter are least popular where fall an summer are most popular. When it comes to the tickets sold, Summer is really unpopular. This could be explained by people not yet having returned from holidays, the teams not having reached forms yet and the short span in which tickets could be bought for these matches. Similar to the game week, the season is intended to incorporate the temperature and stage in the season. It is interesting to see whether playing against the Top-3 in Summer might be a bit less interesting than playing against a team you're competing against for European football late in Spring.

Ajax

Ajax matches sell a lot of tickets, it is the biggest club in the Netherlands and the most enticing match for AZ fans. This match has to be a regressor on it's own and the data easily shows that.

Feyenoord & PSV

For Feyenoord and PSV it is the case that they are also the part of the traditional top -3 clubs in the Netherlands. However regression wise it becomes problematic to include them as regressors and a categorical level for the standing of the opponent as well. The effect for these teams is lesser than Ajax, so it is avoided to include them as regressors for the regression.

Price

It is also useful to have an idea of the prices of match tickets. Therefore histograms have been made for the Prices, Price Types, and the combination of them both. For this reason a combination of the price and price types of regular nonfree tickets for the matches against PSV and Go Ahead Eagles are shown in a histogram.

CONFIDENTIAL

Figure 22 - Shows from low to high, the different price types and frequency for a match vs PSV.

CONFIDENTIAL

Figure 23 - Prices and price types for a match against Go Ahead Eagles.

When a close look is taken at the prices it is observable that the most expensive ticket for AZ – PSV is 40 euro, whereas the most expensive ticket for AZ vs Go Ahead Eagles is 30 euro's. Other than that we see relatively far higher ticket prices in terms of count for PSV as that Histogram is tilted to the right by about 5 – 10 euro's with considerably more bins. This gives an indication for the prices that are of use in this problem.

Next the correlations of the regressors with the response variables are calculated.

Within 7 Days	Season Fall	Season Spring	Season Summer	Season Winter	Standing AZ Europees	Standing AZ Top-3	Standing Opponent Top-3	Standing Opponent Europees	Standing Opponent Middenmoot
1.51	1.71	2.11	1.54	2.17	1.58	1.57	inf	4.63	4.30
Show	Standing difference - Big	Standing difference - Small	Standing difference - Minimal	DayInWeek - Saturday	DayInWeek - Sunday	DayInWeek - Weekday	Ajax	PSV	Feyenoord
4.72	4.65	4.07	5.84	2.77	2.90	1.74	inf	inf	inf

Table 15 - The correlation coefficients for all the regressors.

Easily observable is that none of the variables have a strong correlation with the response variable which is whether a season ticket holder showed up for the match. The strongest correlation is when a match is played against an opponent who is standing in the top 3 or if it's Ajax, other than that the correlations are quite low so it is difficult to directly say that when this variable goes up, the response variable goes up or down as well.

Within 7 Days	Season Fall	Season Spring	Season Summer	Season Winter	Standing AZ Europees	Standing AZ Top-3	Standing Opponent Top-3	Standing Opponent Europees	Standing Opponent Middenmoot
0.03	0.07	-0.06	-0.00	-0.02	-0.03	0.03	0.16	-0.06	-0.07
Standing Opponent Degradatie	Standing difference - Big	Standing difference - Small	Standing difference - Minimal	DayInWeek - Saturday	DayInWeek - Sunday	DayInWeek - Weekday	Ajax	PSV	Feyenoord
-0.05	-0.10	0.07	0.04	-0.02	0.02	0.01	0.11	0.09	0.06

Table 16 - The variance inflation factor for all the regressor.

However, the combination of various factors could do well to explain whether an individual will show up for a certain match or not. In that sense, this is an interesting thing to research but not necessary. Interesting is to look at various correlations as this is a correlation with binary regressors. In that sense it could also be interesting to look at Phi association or Cramer's coefficient.

Phi association is useful as it gets rid of negative associations, so it measured the association in an absolute and undirected manner. However in this case, there seems to be that all the regressors have no or a negligible direct relation with the response again. Based on these results, none of the regressors are excluded of the model [Goodman et al. (1979)].

Within 7 Days	Season	Standing AZ	Standing Opponent	Standing Difference	Day In Week	Ajax
0.03	-0.04	0.02	-0.00	0.10	0.04	0.09

Table 17 – The phi association for all the regressors.

Another consideration to be made is the Variance Inflation Factor, where each regressor should have a variance information factor which is beneath 5. If not, it means that they are too much correlated with other regressors. Easily observable is that the standing of the opponent being top 3 has a coefficient of infinity, as do the regressors Ajax, Feyenoord, PSV. This is the case because they are the only teams to have occupied the top-3 positions especially when the standings during the first few matches are generated deterministically.

Within 7 Days	Season Fall	Season Spring	Season Summer	Season Winter	Standing AZ Europees	Standing AZ Top-3	Standing Opponent Top-3	Standing Opponent Europees	Standing Opponent Middenmoot
1.50	1.75	2.03	1.54	2.09	1.58	1.55	4.29	4.07	4.24
Show	Standing difference - Big	Standing difference - Small	Standing difference - Minimal	DayInWeek - Saturday	DayInWeek - Sunday	DayInWeek - Weekday	Ajax		
4.72	4.45	4.07	4.55	2.77	2.90	1.74	1.68		

Table 18 - The variance inflation factor for a subset of the regressors.

For this reason, the regressors PSV and Feyenoord are removed and Top-3 and Ajax are kept in order to have reasonable Variance Information Factors that ensure the sensibility and validity of the model.

Regression Models

Transaction regression

When all the transformations on the dataset are performed, the data set is left with about 500.000 transactions over 51 matches. This is the first regression, a logistic regression to determine whether a season ticket holder is absent or present for the match. This logistic regression uses a train set of 80% and test set of 20%. The trainset and test set are stratified, such that the various levels are both present in the train set and in the test set. Under these circumstances, the following results were yielded.

The choice for a logistic regression was made because of the binary nature of the explained variable, however the model that follows here after is of different nature. Then a look is taken at the number of tickets sold, which serves as a proxy for the demand, as a function of the attractivity of the match and its price in order to capture the relationship between the (average) price of a match and the number of tickets sold (demand). In this model, the choice is made to use a Poisson Regression with a Binomial Link function. Other options could be made such as Machine Learning Models, Binomial Tree's, OLS-Regression, however, in the context of the situation at AZ it is important to keep the methods explainable. In that way, it is more likely that they will be used despite possibly achieving lower accuracy and performance.

Model

The idea is to retrieve whether a season ticket was actualized and calculate the proportion of season ticket holders present, including regressors that could have predictive value for this issue.

Presence;	1	If the season ticket holder is present
	0	If the season ticket holder is absent

Presence here serves as a proxy for the attractivity of the match, which is then a function of the regressors with predictive value. That is then modelled in the following way.

Attractivity = Intercept +β_1* Season +β_2* Standing AZ + β_3* Status + β_4* Standing Opponent+ β_5* Day In Week + β_6* Match Within 7 Days + β_7* Speelronde + β_8* Ajax + β_9* Clubkaartverplichting

The backward elimination technique was used to reduce the regressor set to the regressors that have significance. Only the regressors with a significance level below 0.05 are used.

Results and Discussion

```
In [18]: regression_one = logistic_regression(df_show_transactions, expression)
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                   Show   No. Observations:               318982
Model:                            GLM   Df Residuals:                   318967
Model Family:                Binomial   Df Model:                           14
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:            -1.3653e+05
Date:                Tue, 19 Oct 2021   Deviance:                   2.7306e+05
Time:                        13:45:34   Pearson chi2:                 3.21e+05
No. Iterations:                     6
Covariance Type:            nonrobust
```

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	2.8874	0.044	65.784	0.000	2.801	2.973
DayInWeek[T.Sunday]	0.0653	0.013	5.009	0.000	0.040	0.091
DayInWeek[T.Weekday]	0.1800	0.023	7.738	0.000	0.134	0.226
Status[T.Laag]	-0.2273	0.029	-7.783	0.000	-0.285	-0.170
Status[T.Midden]	0.1156	0.023	5.052	0.000	0.071	0.160
Status[T.Top 3]	1.3269	0.029	45.123	0.000	1.269	1.385
Season[T.Spring]	1.1536	0.043	26.865	0.000	1.069	1.238
Season[T.Summer]	-0.6981	0.028	-24.912	0.000	-0.753	-0.643
Season[T.Winter]	0.7801	0.026	30.398	0.000	0.730	0.830
Under_5_Degrees	-0.1709	0.017	-10.354	0.000	-0.203	-0.139
Speelronde	-0.1445	0.004	-33.999	0.000	-0.153	-0.136
standing_AZ	-0.0850	0.005	-16.491	0.000	-0.095	-0.075
standing_opponent	-0.0138	0.002	-7.324	0.000	-0.018	-0.010
Ajax	0.6651	0.046	14.519	0.000	0.575	0.755
ClubkaartVerplichting	-0.0859	0.014	-6.181	0.000	-0.113	-0.059

Table 19 - Plot of the coefficients of the logistic regression.

Below the predicted occupancy for 10 different matches are found, graphically the predicted counts quite closely follow the actual counts and when looking at a scatter plot of the Fitted Values against the Actual values or residuals the patterns are promising in terms of drawing conclusions on normality of the errors.

The area under the curve is a performance measurement for classification problems. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. In this case, the higher the AUC, the better the model is at distinguishing season ticket holders that show up for a match and season ticket holders that don't.

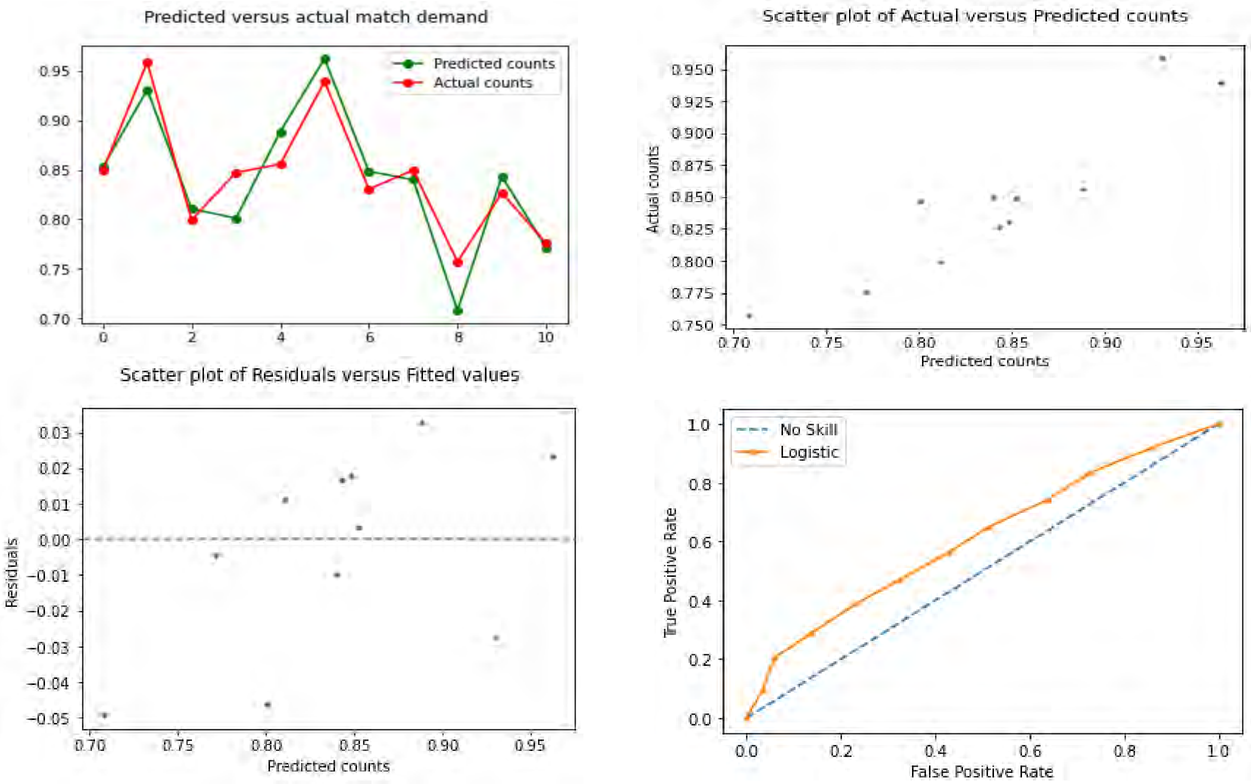


Figure 24 - The performance of the logistic regression with respect to accuracy and normality.

Some measures are calculated in order to validate the model and its performance. This validation makes the model usable to statistically significant degree or not.

A classification model that is unintelligent, would guess the false positives and true positives right a half of the times. This logistic regression model, raises that to 61% of the time which is a slight improvement but not strong. Strong improvements are considered to be from 70% [Hosmer Jr et al. (2015)].

The Jarque Bera and Shapiro statistics both have p-values greater than 0.05 and that means that their null hypothesis are both not rejected. Their null hypothesis is that the errors are normally distributed, so normality of errors is not rejected in this case. This is important for the validation of the model and shows that the logistic regression model doesn't fit the model well enough to make conclusions on. A high R^2 is found in 85%, this means that 85% of the variation in the occupancies of the matches are explained by the logistic regression model. This is paired with a MAPE of under 3% which is very high. It means that the values that the model predicts are very accurate [(Lewis 1982)].

Measure	**Value**
No Skill: ROC AUC	0.500
Logistic: ROC AUC	0.610
Jarque Bera (statistic, pvalue)	0.99 – 0.61
Shapiro (statistic, pvalue)	0.92 – 0.31
R^2	0.85
Mean Absolute Percentage Error	2.66

Table 20 - Statistic Measures for performance, accuracy and normality.

Price regression

The first part of the regression was to reduce all matches to one metric, and that is their attractivity without bringing price into the equation. That was possible, but the validity of the regression model was insufficient. Here the second regression is performed, in order to get some idea of the influence of price on the amount of tickets sold. In this second regression, matches are reduced to their attractivity score.

Model & Method

This is the model in the form of an equation;

$$\text{The Amount of Tickets Sold} = \text{Intercept} + \beta_1 * \text{Attractivity} + \beta_2 * \text{Price} \qquad (2)$$

This equation is calculated by performing a Poisson regression because the amount of tickets sold for every match is count data. Ajax, PSV and Feyenoord are omitted in this regression as they automatically sell much tickets, but also correlate with the highest prices.

This is still a linear regression, left with 42 matches with data on the amount of tickets sold, they have an attractivity score based on the "Transaction Regression", and the average price of all the sold regular tickets is used as a measure of their different prices. In total there are 42 observations, each observation containing three values. One value that is explained, by two regressors.

Results and Discussion

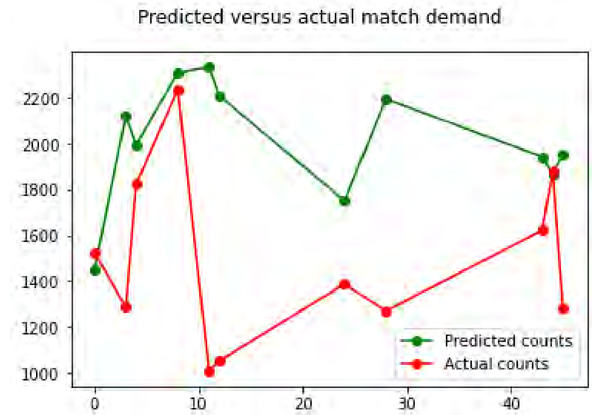
```
In [47]: price_regression(df_scored_transactions, expression_price_regression)
Training data set length=31
Testing data set length=11
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:           Tickets_Sold   No. Observations:                   31
Model:                            GLM   Df Residuals:                       28
Model Family:                 Poisson   Df Model:                            2
Link Function:                    log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -3586.7
Date:                Mon, 13 Dec 2021   Deviance:                       6886.2
Time:                        07:54:01   Pearson chi2:                 6.79e+03
No. Iterations:                     4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.1688      0.090     91.040      0.000       7.993       8.345
Price         -0.0899      0.002    -45.756      0.000      -0.094      -0.086
Score          0.9626      0.110      8.717      0.000       0.746       1.179
==============================================================================
```

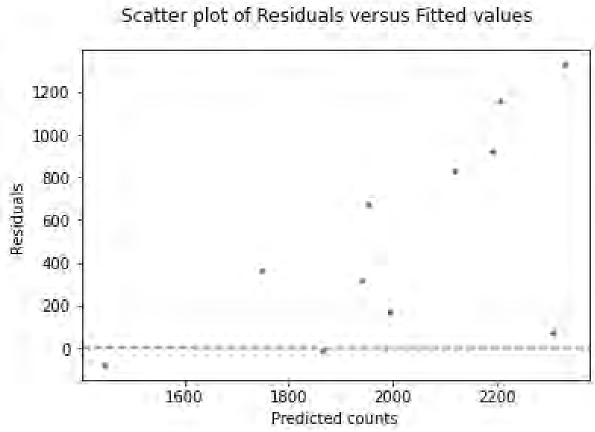
Table 21 - Plot of the regression of the amount of tickets sold, with the regressors Price and Score.

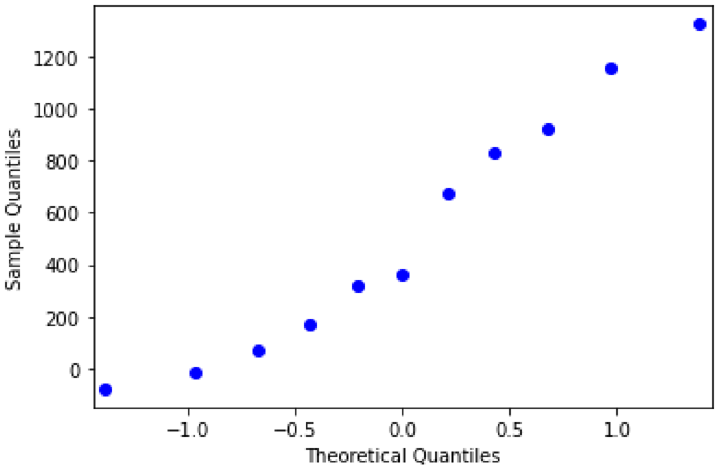
The regression coefficients intuitively make sense. Price has a negative coefficient of -0.0899 meaning that an increase of one in average price goes with 9% decrease in the amount of tickets sold. For the Scores, it means that an increase in one for the Score goes together with a 96% increase in the amount of tickets sold. Note that the scores are between 0.72 and 0.89 and prices between 12 euro's and 22 euro's.

Intuitively, it makes sense that price has a negative effect on the amount of tickets sold and that attractivity has a positive effect on the tickets sold. This confirms the hypothesis. It would be wise to segregate again however between the top matches, vomar matches, and remaining matches to try and extract the exact relationship.

Important things to note however, are the lack of observations. 42 observations is very small and it seems not worthwhile to show results on a regression that contain 9, 18 and 24 observations.







Measure	**Value**
Jarque Bera (statistic, pvalue)	0.37 – 0.83
Shapiro (statistic, pvalue)	0.95 – 0.75
R^2	0.08
Mean Absolute Percentage Error	28

Figure 25 – Figures and tables showing the accuracy and normality of the regression.

The accuracy of the regression is reasonably poor, which is to be expected with such few observations. However it looks like the errors are quite normally distributed based on the QQ-plot.

Statistic tests on normality such as the Jarque Bera test or Shapiro test are not valid for such small samples, though shown. So normality is difficult to be validated, the R^2 is very low meaning that only 8% of the variation in tickets sold is explained by the regressors and the Mean Absolute Percentage Error is also very high. These are poor results for a model, and meaningful conclusions can't be drawn from it. For that reason, a different attempt is made in the next regression.

Seat-based Regression

A potential solution for the problem described in the previous section is doing a regression on seat level. A presupposition is, that each match the same seats are all available for sale. And for each match, it could be identified whether a specific seat has been bought. This seat could be allocated all the features earlier described in this section, on transactional level. However, a price for this seat could also be determined. Certainly, for the seats that were sold, it is known what the buyer paid for that seat. However suggestions could also be made for the prices of seats that weren't sold by looking at the seat section in which that seat is located, and making an estimate based on all the price types of the seats that were sold in that section. This seems like a fair way in order to give all seats a price.

In that manner, the influence of price on whether a seat will be bought can directly be determined on transactional level, this would be a regression over hundreds of thousands of transactions, which is much better for the performance of these models compared to the previous model which ultimately had to be grouped towards 42 observations. This approach is therefore analyzed in this section, in order to yield an indication of the price elasticity.

One of the challenges is the fact that tickets are often given freely, the question is whether to include these tickets in the analysis as it might skew the regression performance. Both the inclusion and exclusion of these type of tickets will be presented in this section.

Model

The modelling is the same as the modelling in the previous section, only here there will be three considerations made in order to perform the logistic regression on seat-level.

In the first setting, all tickets are considered including season tickets. In that setting, season tickets will use the data that is already provided whereby a season ticket chair is between 5 euro's and 7 euro's (this is the average price of a match, the cost of a season ticket divided by all the home matches). These seats will then not be filled with a proportion of the regular ticket types sold. A problem with this approach is that cheap tickets are always purchased, and more expensive tickets (unsold seats that were simulated by regular sold tickets) are not purchased. These seats are a lot more expensive, so it biases the price elasticity on a match level.

In the second setting, only regular non-free tickets will be considered and all the other tickets are then not considered. Possible problem is that ends up with a lot of empty seats, and this may bias the seats that were actually occupied by season tickets.

In the third setting, only the seats that are not occupied by season tickets are considered. The remaining tickets are either sold to a regular ticket holder or are unsold. For the unsold ticket, the data is distributed in proportion to the data of the sold tickets.

For the robustness of this model, the top matches will not be considered nor will matches with necessary clubcards for ticket sale. These matches greatly influence the chairs that aren't bought and hence bias the price elasticity. The same regressors as for the transactional model are used, but because of the set-up of this model, the seat section and location are also included as regressors. If they are omitted, the model can't be evaluated in terms of how well it can predict the number of seats that will be sold in a given seat section.

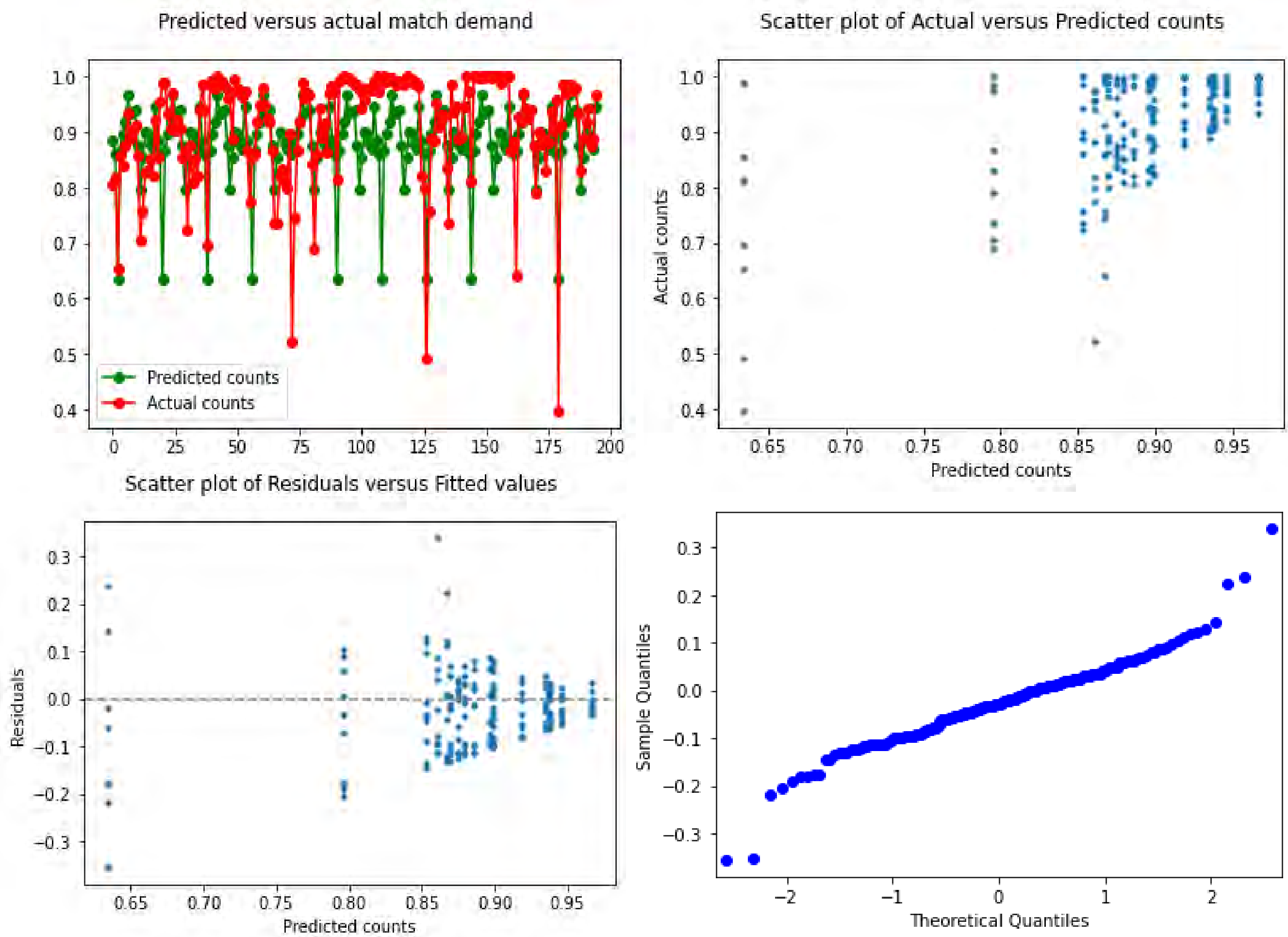
First setting – Results and Discussion

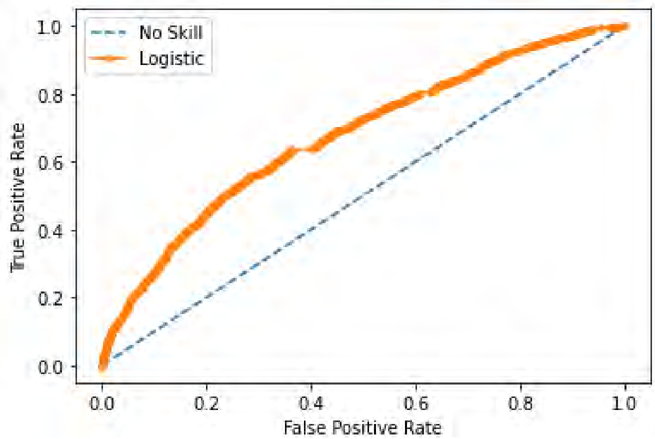
In the first setting season tickets are considered. However when season tickets are used, every match, the same seats will always be registered as bought for a low price. In that sense, the predictor will predict largely the same amount of tickets sold for every match and will find that for the more popular matches, which are more expensive, a higher price correlates with a higher demand, resulting in a positive price coefficient. For this reason, it is better to not include season tickets. The model coefficients follow now.

```
In [102]: regression_four = logistic_regression(df_all_seats_per_match,
expression_All)
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:              Purchased   No. Observations:               403498
Model:                            GLM   Df Residuals:                   403465
Model Family:                Binomial   Df Model:                           32
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:            -1.5242e+05
Date:                Mon, 13 Dec 2021   Deviance:                   3.0483e+05
Time:                        16:21:36   Pearson chi2:                 4.04e+05
No. Iterations:                     6
Covariance Type:            nonrobust
==============================================================================
------------------------------------------------------------------------------
Intercept                2.1219      0.060     35.223      0.000       2.004       2.240
Season[T.Spring]         0.0629      0.016      3.907      0.000       0.031       0.095
Season[T.Summer]        -0.3936      0.014    -28.706      0.000      -0.421      -0.367
Season[T.Winter]         0.1719      0.014     12.308      0.000       0.145       0.199
Area[T.J]               -0.2459      0.027     -9.025      0.000      -0.299      -0.193
Area[T.M]               -1.5578      0.029    -52.924      0.000      -1.615      -1.500
Area[T.N]               -0.2119      0.031     -6.865      0.000      -0.272      -0.151
Area[T.O]                0.0993      0.031      3.215      0.001       0.039       0.160
Area[T.P]                0.3416      0.032     10.732      0.000       0.279       0.404
Area[T.Q]                1.2839      0.044     29.018      0.000       1.197       1.371
Area[T.R]                0.5906      0.033     17.857      0.000       0.526       0.655
Area[T.S]                0.6315      0.034     18.786      0.000       0.566       0.697
Area[T.T]                0.6641      0.036     18.682      0.000       0.594       0.734
Area[T.U]               -0.1186      0.032     -3.753      0.000      -0.181      -0.057
Area[T.V]               -0.7216      0.030    -24.286      0.000      -0.780      -0.663
Area[T.W]               -0.3028      0.032     -9.367      0.000      -0.366      -0.239
Area[T.X1]               0.1395      0.035      3.974      0.000       0.071       0.208
Area[T.X2]               0.1111      0.036      3.125      0.002       0.041       0.181
Area[T.X3]              -0.0596      0.034     -1.753      0.080      -0.126       0.007
Area[T.Y]               -0.1565      0.034     -4.607      0.000      -0.223      -0.090
Area[T.Z]                0.8198      0.043     19.027      0.000       0.735       0.904
Location[T.High]         0.0842      0.047      1.792      0.073      -0.008       0.176
Location[T.Middle]       0.0003      0.046      0.007      0.994      -0.090       0.090
DayInWeek[T.Sunday]     -0.0381      0.012     -3.157      0.002      -0.062      -0.014
DayInWeek[T.Weekday]    -0.1809      0.022     -8.243      0.000      -0.224      -0.138
Status[T.Laag]          -0.3175      0.029    -11.115      0.000      -0.373      -0.262
Status[T.Midden]        -0.2772      0.022    -12.533      0.000      -0.321      -0.234
Status[T.Top 3]          0.0600      0.026      2.274      0.023       0.008       0.112
standing_AZ              0.0098      0.006      1.627      0.104      -0.002       0.021
standing_opponent        0.0198      0.006      3.561      0.000       0.009       0.031
standing_difference     -0.0442      0.005     -8.106      0.000      -0.055      -0.033
Price                    0.0086      0.001      9.600      0.000       0.007       0.010
Within_Seven_Days       -0.0372      0.015     -2.444      0.015      -0.067      -0.007
==============================================================================
```

Table 22 - Graph showing the regression coefficients..

Price here has a slightly positive coefficient which is in contrast to the findings found in the previous section. However looking at the model challenges earlier mentioned, it could be that indeed matches that sell more tickets, are more often more expensive matches. An idea might be to regress every match separately and aggregate the results.





Measure	**Value**
No skill: ROC AUC	0.500
Logistic: ROC AUC	0.678
Jarque Bera (statistic, pvalue)	81.90 – 0.0
Shapiro (statistic, pvalue)	0.96 – 0.0
R^2	0.27
Mean Absolute Percentage Error	8

Figure 26 - Figures showing the accuracy and the normality of the regression.

With the high amount of observations, normality is rejected for both the Jarque Bera and Shapiro test, the qq-plot affirms this as well. The Mean Absolute Percentage Error and the ROC AUC give confidence as well of a well performing model where the plot of residuals vs fitted values show that the model is often reasonably accurately finding how many people bought a ticket in a seat section. What makes it difficult is that the values are within a certain window for a majority of the data, so it is bound to not go beyond this window.

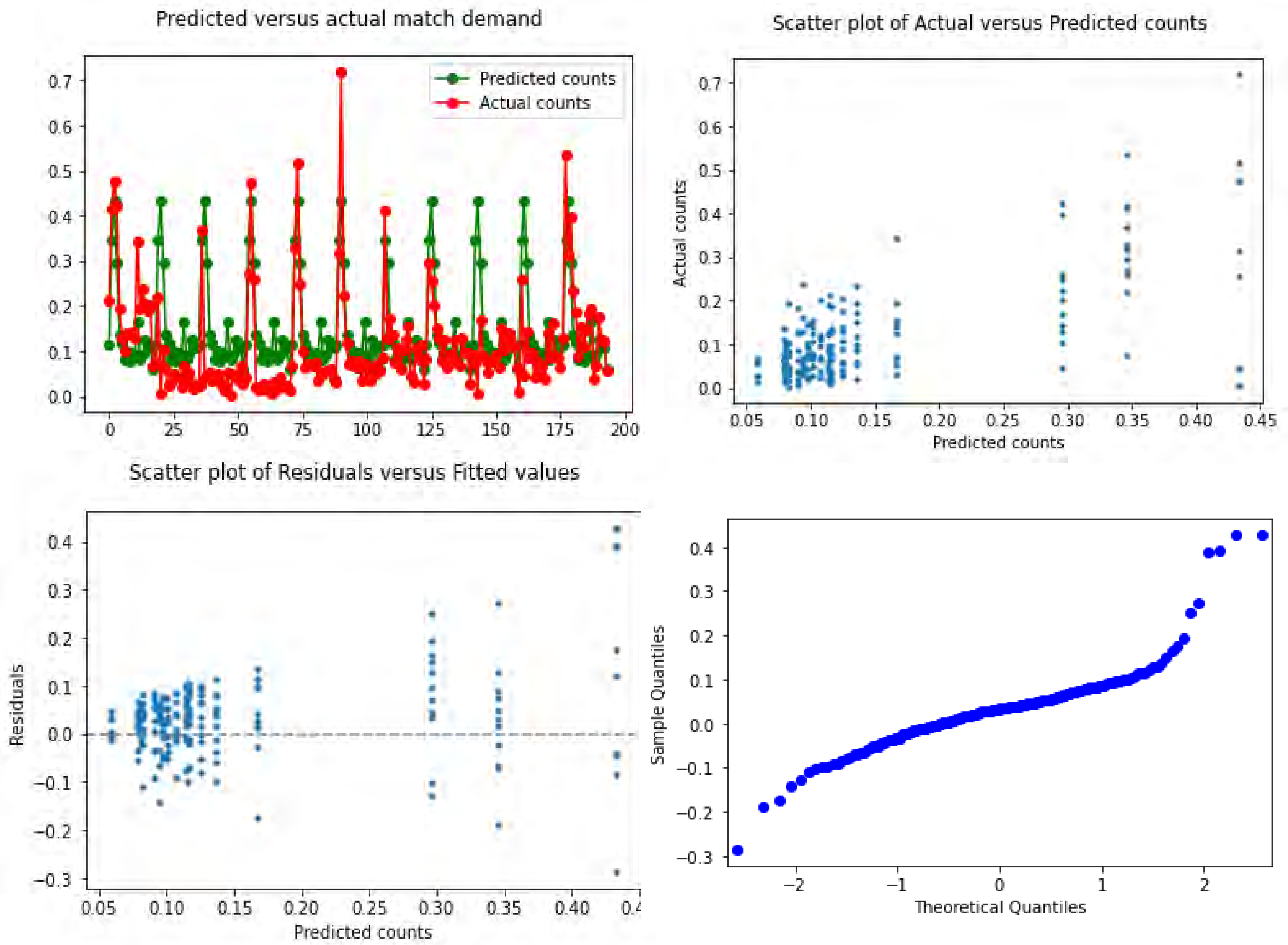
Second setting – Results and Discussion

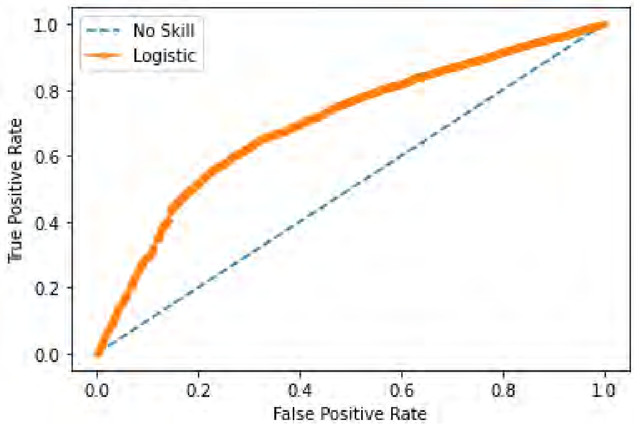
In the second setting, only regular tickets with non-zero prices are considered. This leaves a lot of seats unsold however, the impact of this on the regression is not clear. It might diminish the actual influence of price on whether a ticket is sold.

```
In [104]: regression_four = logistic_regression(nonregular, expression_All)
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:              Purchased   No. Observations:               358203
Model:                            GLM   Df Residuals:                   358170
Model Family:                Binomial   Df Model:                           32
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:            -1.4503e+05
Date:                Mon, 13 Dec 2021   Deviance:                   2.9006e+05
Time:                        16:33:11   Pearson chi2:                 3.62e+05
No. Iterations:                     6
Covariance Type:            nonrobust
==============================================================================================
                                coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Intercept                    -0.7030      0.053    -13.364      0.000      -0.806      -0.600
Season[T.Spring]              0.0599      0.014      4.378      0.000       0.033       0.087
Season[T.Summer]             -0.5906      0.018    -33.629      0.000      -0.625      -0.556
Season[T.Winter]              0.3493      0.014     25.864      0.000       0.323       0.376
Area[T.J]                     1.2147      0.029     42.527      0.000       1.159       1.271
Area[T.M]                     1.5558      0.032     49.248      0.000       1.494       1.618
Area[T.N]                     1.1115      0.031     35.732      0.000       1.051       1.172
Area[T.O]                     0.1951      0.034      5.791      0.000       0.129       0.261
Area[T.P]                     0.0421      0.035      1.219      0.223      -0.026       0.110
Area[T.Q]                    -0.3247      0.040     -8.194      0.000      -0.402      -0.247
Area[T.R]                    -0.1337      0.038     -3.526      0.000      -0.208      -0.059
Area[T.S]                    -0.3991      0.037    -10.681      0.000      -0.472      -0.326
Area[T.T]                    -0.2138      0.036     -5.977      0.000      -0.284      -0.144
Area[T.U]                    -0.3319      0.036     -9.119      0.000      -0.403      -0.261
Area[T.V]                     0.3272      0.033      9.955      0.000       0.263       0.392
Area[T.W]                    -0.3500      0.039     -8.973      0.000      -0.426      -0.274
Area[T.X1]                   -0.1882      0.040     -4.702      0.000      -0.267      -0.110
Area[T.X2]                   -0.0947      0.040     -2.355      0.019      -0.173      -0.016
Area[T.X3]                   -0.0916      0.039     -2.319      0.020      -0.169      -0.014
Area[T.Y]                    -0.1665      0.038     -4.366      0.000      -0.241      -0.092
Area[T.Z]                    -0.7850      0.045    -17.287      0.000      -0.874      -0.696
Location[T.High]              0.0897      0.038      2.350      0.019       0.015       0.164
Location[T.Middle]           -0.0624      0.037     -1.691      0.091      -0.135       0.010
DayInWeek[T.Sunday]           0.0516      0.011      4.740      0.000       0.030       0.073
DayInWeek[T.Weekday]         -0.0237      0.021     -1.125      0.261      -0.065       0.018
Status[T.Laag]               -0.4990      0.026    -19.100      0.000      -0.550      -0.448
Status[T.Midden]             -0.1254      0.019     -6.472      0.000      -0.163      -0.087
Status[T.Top 3]               0.4708      0.023     20.413      0.000       0.426       0.516
standing_AZ                  -0.1155      0.005    -21.863      0.000      -0.126      -0.105
standing_opponent             0.0140      0.005      2.803      0.005       0.004       0.024
standing_difference          -0.0393      0.005     -8.048      0.000      -0.049      -0.030
Price                        -0.0335      0.001    -43.997      0.000      -0.035      -0.032
Within_Seven_Days            -0.0773      0.015     -5.240      0.000      -0.106      -0.048
==============================================================================================
```

Table 23 - Plot showing the regression coefficients.

The train set consists of 358000 observations, a 80% train set is used with a 20% test set. In this second setting, Price has a significant negative coefficient which is consistent with the results in the previous model, the coefficient is significantly smaller however. It has gone from about -8% to -3%.





Measure	**Value**
No skill: ROC AUC	0.500
Logistic: ROC AUC	0.701
Jarque Bera (statistic, pvalue)	394 – 0.0
Shapiro (statistic, pvalue)	0.86 – 0.0
R^2	0.43
Mean Absolute Percentage Error	43

Figure 27 - Figures showing the accuracy and normality of the regression.

The model accuracy in first and fifth graph show reasonable results, Jarque Bera and Shapiro reject the normality of errors. The Mean Absolute Percentage Error is very poor, the model doesn't do well in predicting how many people will fill an area at all times. This makes sense as only regular tickets are considered, the range of the actual values is much greater with much variance. The way the model is shaped, this is more difficult to detect for the model.

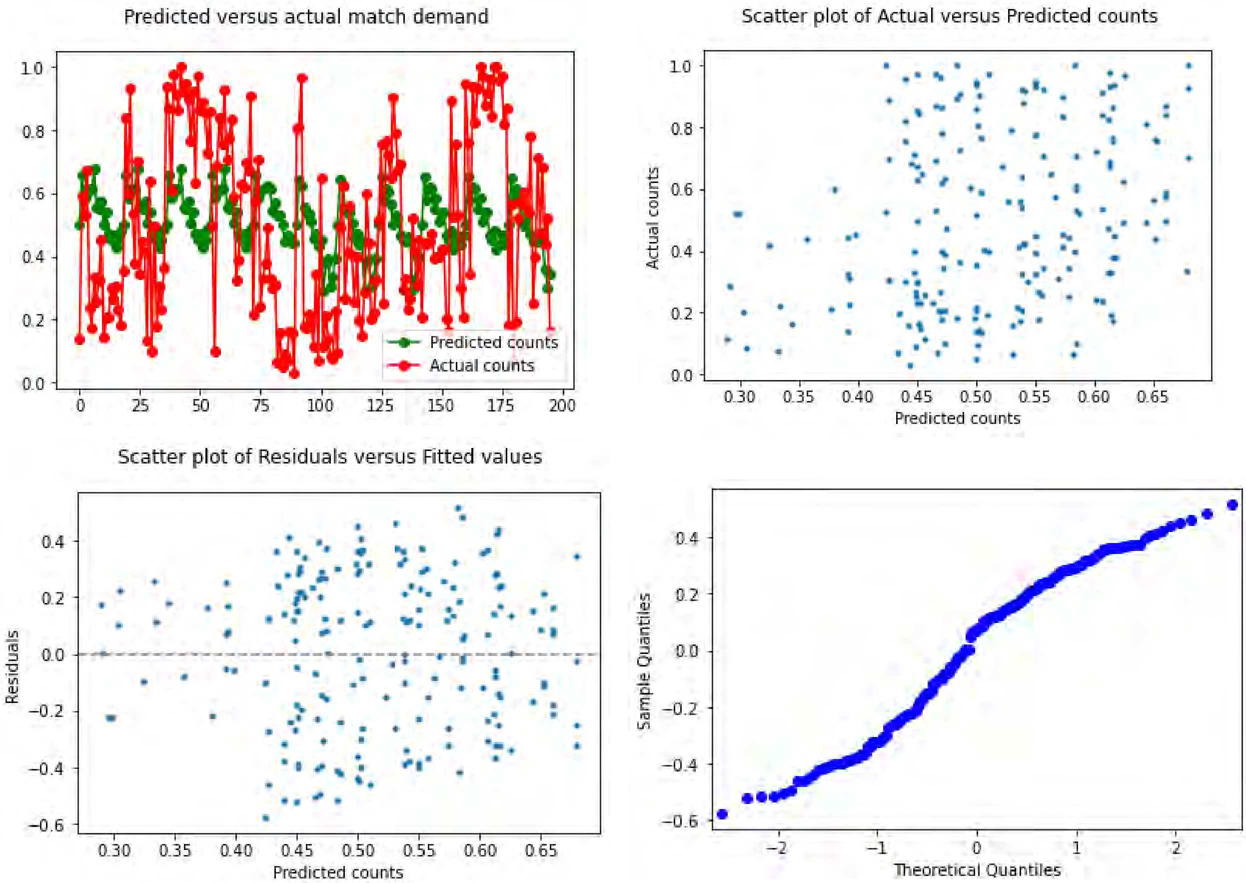
Third setting – Results and Discussion

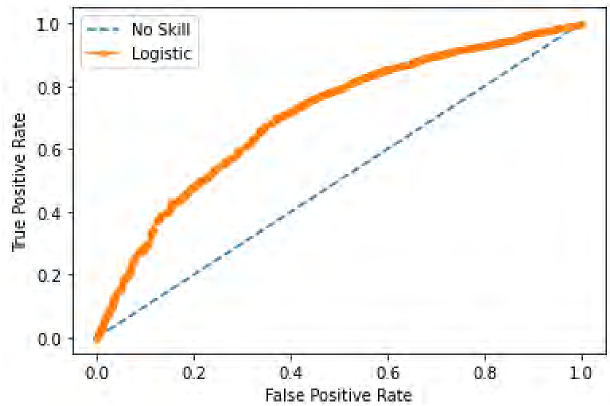
In the third setting, the seats occupied by the season tickets are removed. The season tickets remain fairly constant but each iteration the season tickets are considered once again. Then the transactions and the prices are filled in for the actual prices. Chairs that eventually weren't booked receive a sample of the ticket price of the regular sold tickets in that section. This will be used for the logistic regression.

```
In [105]: regression_four = logistic_regression(df_remaining_seats_per_match, expression_
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:              Purchased   No. Observations:               110969
Model:                            GLM   Df Residuals:                   110936
Model Family:                Binomial   Df Model:                           32
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -68426.
Date:                Mon, 13 Dec 2021   Deviance:                   1.3685e+05
Time:                        16:38:23   Pearson chi2:                 1.12e+05
No. Iterations:                     4
Covariance Type:            nonrobust
=================================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept          2.1207      0.087     24.339      0.000       1.950       2.292
Season[T.Spring]   0.2374      0.024     10.086      0.000       0.191       0.283
Season[T.Summer]  -1.0413      0.021    -49.658      0.000      -1.082      -1.000
Season[T.Winter]   0.1349      0.019      7.104      0.000       0.098       0.172
Area[T.J]          0.4402      0.040     11.136      0.000       0.363       0.518
Area[T.M]         -0.2548      0.042     -6.140      0.000      -0.336      -0.173
Area[T.N]          0.4876      0.043     11.291      0.000       0.403       0.572
Area[T.O]          0.4424      0.047      9.387      0.000       0.350       0.535
Area[T.P]          0.5047      0.049     10.292      0.000       0.409       0.601
Area[T.Q]          1.1538      0.063     18.304      0.000       1.030       1.277
Area[T.R]          0.7235      0.056     12.877      0.000       0.613       0.834
Area[T.S]          0.4422      0.054      8.228      0.000       0.337       0.548
Area[T.T]          0.5381      0.052     10.340      0.000       0.436       0.640
Area[T.U]         -0.3265      0.049     -6.682      0.000      -0.422      -0.231
Area[T.V]         -0.3279      0.044     -7.399      0.000      -0.415      -0.241
Area[T.W]         -0.5136      0.052     -9.867      0.000      -0.616      -0.412
Area[T.X2]        -0.0005      0.057     -0.008      0.993      -0.112       0.111
Area[T.X3]        -0.2324      0.054     -4.291      0.000      -0.339      -0.126
Area[T.Y]         -0.2579      0.051     -5.032      0.000      -0.358      -0.157
Area[T.Z]          0.1571      0.066      2.376      0.018       0.027       0.287
Location[T.High]   0.1942      0.054      3.593      0.000       0.088       0.300
Location[T.Middle] -0.1413     0.052     -2.707      0.007      -0.244      -0.039
DayInWeek[T.Sunday]  0.0356    0.017      2.121      0.034       0.003       0.068
DayInWeek[T.Weekday] -0.2483   0.041     -6.119      0.000      -0.328      -0.169
Status[T.Laag]    -0.3068      0.056     -5.441      0.000      -0.417      -0.196
Status[T.Midden]   0.0475      0.046      1.039      0.299      -0.042       0.137
Status[T.Top 3]    0.4225      0.039     10.933      0.000       0.347       0.498
standing_AZ       -0.0197      0.013     -1.575      0.115      -0.044       0.005
standing_opponent -0.1852      0.012    -15.678      0.000      -0.208      -0.162
standing_difference 0.1319     0.012     10.893      0.000       0.108       0.156
Price             -0.0553      0.001    -49.903      0.000      -0.057      -0.053
Within_Seven_Days  0.0330      0.025      1.327      0.184      -0.016       0.082
=================================================================================
```

Table 24 - Plot that shows the regression coefficients.

Because of the model set-up, this train set uses less observations because all seats that are taken up by season ticket seats that are omitted. Some of the regressors are insignificant but what is being seen here again a negative price coefficient which is greater in magnitude. So this does seem to resemble what was found in the previous section that it can be made visible that price has a negative influence on the sales of match tickets.





Measure	Value
No skill: ROC AUC	0.500
Logistic: ROC AUC	0.708
Jarque Bera (statistic, pvalue)	11.53 – 0.0
Shapiro (statistic, pvalue)	0.96 – 0.0
R^2	0.08
Mean Absolute Percentage Error	47

Figure 28 - Figures that show the accuracy and normality of the regression.

In this setting, normality of errors is again rejected. The accuracy of the model in terms of predicting how many tickets will be sold is again quite poor, the prediction is staying very central of the range of values the observed value assumes. Little percentage of the variation is explained by the regressors, but Price does really help to identify the true positives.

5 – Conclusion

The main conclusion is that the models performed, lack the statistical power to make affirmative conclusions regarding price elasticity. In the first place, what they measure is also not exactly the price elasticity but rather what influence price has on the amount of tickets sold or whether a ticket is sold. What does an unit change in price have as effect on the amount of tickets sold over all the tickets. But it doesn't provide us a demand function, not a reliable one.

Based on the different models and settings, it does seem that price has a negative influence. Especially when the set-up is better streamlined to point the effect of price on the amount of tickets sold, price seems to have a negative coefficient ranging from about 3% to 8%. Meaning that when the price increases by 1 euro, the response decreases by about 3 to 8% on average over all the average prices used. But what exactly the negative influence is, is still not clear as this depends on the model set-up, and still there is no history of price changes made in order to find a demand curve and hence the price elasticities.

It is concluded that this information is currently inconclusive to use in further dynamic pricing models.

Other than that, little may be said about the regressors as their signs change based on the set-up of the regression. Which makes it hard to make affirmative statements about their values. A discussion on the regression model will be given in the next section.

The conclusion on the entire research will be given nonetheless. AZ can already improve the revenue by making changes in the initial prices of tickets. It is worthwhile to perform a further analysis into the desirability of seat sections and the various locations when things stabilize regarding the covid-19 crisis. This has to be done because of the new roof and all the alterations that come with it, this information can also be used for the pricing of season tickets. When this is done, the stadium will also be segmented which is then ready made for the ultimate use of dynamic pricing.

To get there, the advise would be for AZ to experiment for two seasons with minor price changes. AZ should look into changing prices on the basis of the height of the seats, the varying desirability in seat sections, and the fluctuating demand. These three different pricing mechanisms will have an effect upon one another, as the price changes will work in three dimensions. For this reason the advice is to raise/drop prices in small steps initially. Steps of 3%, 5% and 10% if the results warrant it. Starting too big leads to the problem of risking major losses, without knowing the price changes that can lead to favorable results. It is also advisable to group similar matches together, in order to have more observations for comparisons. This will result in demand functions, that can be used to deploy a dynamic programming model that gives price recommendations. This model doesn't need to run on all the segments that have been determined, but the price recommendations could be translated to these segments as well. It is often stressed that is important to find a policy that suits the capacity and situation of the club. The literature review and interviews with the experts show that it isn't necessary to run a completely perfect dynamic pricing model with perfect demand functions and perfect automation. A simple but good model can already improve revenue to a level that makes dynamic pricing impossible to ignore in the long-term.

6 - Discussion

One of the issues when it comes to the two-piece transaction regression, is that now the transactions are grouped to match-level meaning that the regression is based on 51 matches that have a certain amount of tickets sold, attractivity score and price level. This is a very low number of observations to yield some statistical significance. In these 51 matches there are significant differences in the context surrounding the matches. 9 of these matches are versus Ajax Amsterdam, PSV Eindhoven, Feyenoord Rotterdam, a dozen of these matches were during the Vomar sale which attracted lots of people for low prices. The main problem is that this makes that the number of observations are very few to perform a regression on. Especially also when the context of these matches is taken into place. There are then at least 2 very different types of matches in that dataset, matches during the Vomar sale and matches that weren't. Above that, there are the very low-placed teams, and the higher placed teams. The level of granularity in these different type of matches, make regressions very tedious if the influence of one regressor wants to be extracted. Because of this, generally during these regressions: meaningful conclusions were difficult to form, and significant ones even harder.

For the seat regression models, the mean absolute percentage errors were very poor. By that, the accuracy of the model was insufficient. Because of the set-up of this model, there might be some biases towards what the dominating category is in that set-up. The normality of errors was rather good, this is helped by the big amount of observations. Also it seems that by the third model, the influence of price is much better identified. A further avenue that could be taken is, to regress on less aggregated data again, because of the many observations available. For instance, it would be interesting to again dissect the top matches, matches during the Vomar action, make groups based on status, or even perform the regression per match. One could also look to improve the regressors. Most of these regressors were constructed during this internship. Finding a better regressor that captures whether it was freezing cold in the stadium. A regressor that indicates whether AZ still has much to play for at that stage in this season. In this regression set-up, at best, a more significant coefficient for price would be found with a nice area under the curve, good predictions and good evaluation measures. It would never yield demand functions but only general information over all the transactions. It rather finds the correlation between differences in price and the amount of tickets sold. This isn't a one to one translation of demand functions which capture much more information. However such a price coefficient could be generalized and used for calculations, it doesn't seem to be the best way of doing things.

In terms of the data analysis, the popularity of the location in the seat sections seems fairly obviously split up in three sections. Low, Middle and High. It would be wise to reconsider this situation in the new situation at AZ, and definitely considering the matches and seat sections that are purposely held empty. When it comes to the most preferred seat sections, the same considerations should be made. It is however necessary to also look at the percentage of season ticket holders in those seat sections. Caution is necessary when changing the prices of season tickets, and it is also important to know the population of those sitting in the seat section. Section S is the seat section where the 'hard core of yesterday' are seated. It is probably not wise, to significantly raise prices here. For the overflow of match tickets to other seat sections, it seems that there is some overlap between X1 and X2. However, it is advisable to only further look into this when a model is ultimately chosen that needs this information. Based on the literature, and conversations had with experts: this information is not always necessary to apply dynamic pricing.

Appendix

Section A

Google Drive with all the files (https://drive.google.com/drive/folders/1gaJ_OYQ82DMEeJ6RBg0bs-GUflbKpbtx?usp=sharing)

All the bar charts and sales graph can be found here.

Sold Out Sales

Bar Chart All Matches – Binary

Bar Chart All Matches – Percentage

Bar Chart Top Matches – Binary

Bar Chart Top Matches – Percentage

Bar Chart Vomar Matches – Binary

Bar Chart Vomar Matches – Percentage

Bar Chart Remaining Matches – Binary

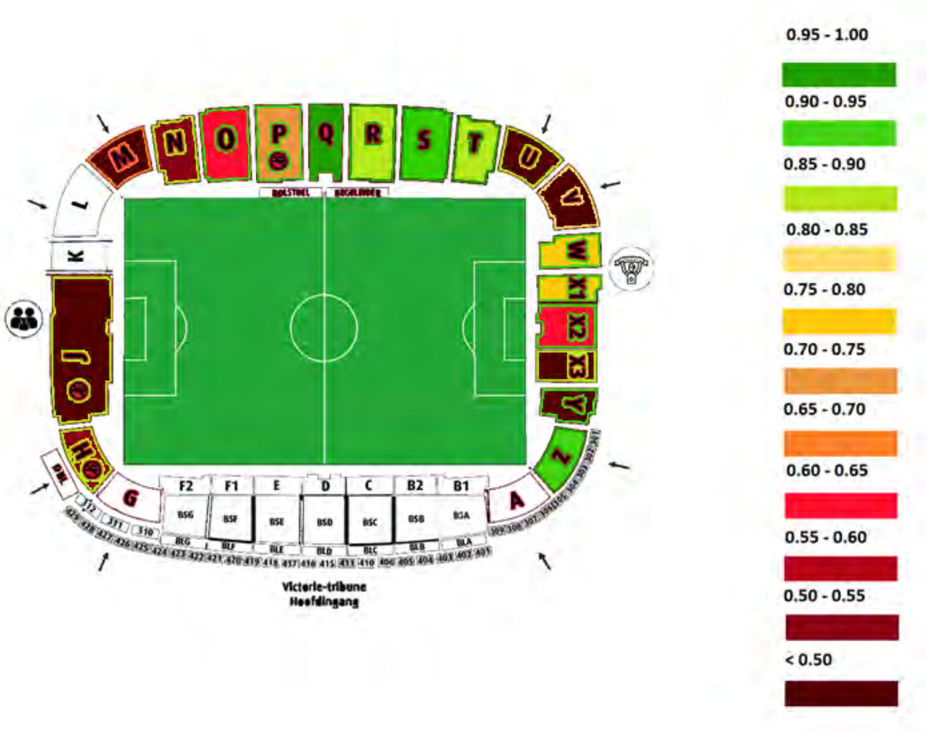
Bar Chart Remaining Matches – Percentage

Section C

All Matches

Top Matches

Vomar Matches



Remaining Matches

Section D

df_area_part_rolling - DataFrame

Index	Day 66	Day 67	Day 68	Day 69	Day 70	Day 71	Day 72	Day 73	Day 74	Day 75	Day 76	Day 77	Day 78	Day 79	
0	3.965e-18	0.03175	0.03175	0.03175	0.03175	0.03175	0.03175	0.03175	3.965e-18	3.965e-18	3.965e-18	3.965e-18	0.08571	0.08571	0.08
1	0.4143	0.4143	0.3	0.2643	0.3357	0.3357	0.2429	0.1	0.1	0.07143	0.07143	9.516e-17	9.516e-17	0.04762	0.04
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0.007143	0.007143	0.007143	0.007143	0.007143	0.007143	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	0.05714	0.1524	0.15
4	0.05	0.007143	0.007143	0.007143	0.007143	0.007143	0.02381	0.08095	0.08095	0.08095	0.1762	0.1762	0.1762	0.1524	0.09
5	0.1071	0.1548	0.1548	0.04762	0.04762	0.04762	0.04762	0.04762	0	0.1224	0.1224	0.2653	0.2653	0.2653	0.26
6	0	0	0	0.04082	0.04082	0.04082	0.06463	0.06463	0.06463	0.06463	0.02381	0.02381	0.02381	0	0
7	0.07143	0.03175	0.03175	0.05215	0.05215	0.05215	0.09977	0.09977	0.2109	0.2109	0.1905	0.1905	0.1905	0.1429	0.14
8	4.956e-17	4.956e-17	0.04286	0.04286	0.04286	0.04286	0.04286	0.04286	0.04286	4.758e-17	4.758e-17	4.758e-17	4.758e-17	4.758e-17	4.75
9	0	0	0.02857	0.02857	0.02857	0.02857	0.02857	0.02857	0.02857	0	0	0	0	0	0
10	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	0.02041	0.02041	0.02041	0.02041	0.02041	0.02
11	0.02143	0.03889	0.03889	0.0593	0.0593	0.1164	0.1093	0.1093	0.07755	0.07755	0.05714	0.05714	0	0	0
12	0.1214	0.1071	0.1214	0.1827	0.2541	0.1469	0.1469	0.2041	0.2041	0.1898	0.1286	0.05714	0.05714	0.05714	0
13	0	0	0.02857	0.02857	0.02857	0.1143	0.1143	0.1429	0.1429	0.1143	0.1619	0.1619	0.07619	0.07619	0.04
14	0.1429	0.1429	0.1429	0.1429	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.172e-17	3.17
15	0.03571	0.03571	0.03571	0.03571	0.03571	0	0	0	0	0	0	0	0	0	0.14
16	0.02857	0.02857	0.02857	0.02857	0.02857	0.02857	0.04762	0.04762	0.04762	0.04762	0.04762	0.04762	0.04762	0	0
17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19e-17	1.19
19					['Q']										

Table 1 – Table showing the ratios smoothened over 7 days on a daily basis. The last row indicates whether a seat section was sold out that given day.

total_list - DataFrame

Index	Day 33	Day 39	Day 47	Day 48	Day 51	Day 47	Day 51	Day 54	Day 63	Day 68	Day 75	Day 80	Day 101	Day 103	Day 124	Day 131	Day 135	Day 137	Day 138	Day 141
0	-0.005291	0	-0.001444	-0.004762	0.001249	0.01905	0	0.009524	0	0	0	0	0	0	-0.01348	0	-0.004082	0	0.001116	0
1	-0.02579	0.004699	0.05256	0.02288	0.01892	0.04762	0.003968	-0.004762	0.09524	-0.007519	0	0.05714	-0.03061	0.04233	-0.003032	-0.009354	0.04184	0.02997	0.0194	0.02185
2	0	0	0.004892	0.002323	0.005917	0	0	0	0	0	0	0	0	0	0.01786	-0.01105	0.025	0.01719	0.008822	0.0159
3	-0.005291	-0.005952	-0.007735	-0.002439	0.004075	-0.03571	0	0	0	0.01099	-0.01099	0	0.02041	0.02721	-0.01348	0.0102	0.002551	-0.01603	-0.0102	-0.01242
4	0.06614	0.02256	0.003448	0.002323	-0.006548	-0.01786	0	0	-0.09524	-0.01504	0	-0.02597	0	0.01058	-0.01887	0	0.003571	0.0012	0.003348	0
5	-0.005291	-0.02005	-0.02525	-0.009524	-0.003861	0.01071	-0.01058	-0.01429	0	0	0.007143	-0.02597	-0.01429	0	0.03571	-0.02551	-0.01224	0.002401	-0.0102	-0.01242
6	0	0.01128	0	0	0.006762	0	-0.01587	0	0	0.01446	-0.02198	0	0.02041	-0.08163	0	0	0	-0.009217	0	0.001232
7	-0.01587	0	0.01174	-0.003717	-0.004302	0	0.01786	0	0	-0.04916	0.003297	-0.02208	0	0	-0.01887	0	-0.02602	0.0012	0.003348	0.001232
8	0	0.004699	-0.002423	0.006969	-0.00728	-0.01786	-0.01587	0	0	0.00347	0.003297	0	0	-0.04082	0	0	0.003571	0.003601	0.001116	-0.003748
9	0	-0.01848	-0.01407	0.005807	-0.01754	-0.01786	-0.01058	0	0	0	0.01429	0	0	0	0	0	0	0.0012	0.008929	0.002463
10	0	-0.002193	-0.01845	-0.007201	0	0	0.01786	0	0	-0.01504	0.04286	0	0	0	0	0	0	0	0.002232	0.003695
11	0	0.007519	-0.01361	-0.02985	-0.003936	0	-0.02116	0	0	0.02198	-0.007692	0	0	0.01058	-0.002695	0.02721	-0.02653	-0.005228	-0.02163	-0.01194
12	-0.005291	-0.00282	-0.0004659	0.003484	0.005917	-0.01786	0.008929	0.05714	-0.04762	-0.01504	0.007143	-0.02597	-0.02857	0	0.008929	-0.008929	-0.004082	0.007203	0.007812	0.01847
13	-0.02116	0.007519	0.00834	-0.004878	-0.0005551	-0.008333	0.0003307	-0.009524	0.04762	0.03297	-0.004396	0	0	0	0.02139	-0.01998	0.001531	0.00639	-0.002338	-0.009959
14	0	-0.0119	0.002935	0.006969	-0.004264	0.009524	-0.01058	-0.009524	0	0	0	0	0	0.01058	0	0.01361	-0.01224	-0.04147	-0.01137	-0.008085
15	0	0.01504	-0.003401	0.003484	0.004189	0	0.01786	0	0	0.00694	-0.02198	0	0.04082	0.01058	0	0.003401	0.007143	0.003601	0.007759	0.004926
16	0.01786	0	0.002935	0.006969	0.003381	0.02857	0.01786	-0.02857	0	0	0	0	-0.02857	0.01058	-0.01078	0.02041	0	-0.002014	-0.01036	-0.01119
17	0	-0.0119	0	0.001161	-0.002132	0	0	0	0	0.01099	-0.01099	0.04286	0.02041	0	-0.002695	0	0	0	0.002232	0
18	['Z']	['Q']	['S']	['R']	['T']	['S']	['T']	['Q']	['Z']	['P']	['R']	['U']	['H']	['O']	['N']	['X2']	['X1' 'Y']	['X3']	['V']	['J']

Table 2 – Table showing every day that some seat section got sold out, based on all the matches, this can be used to summarize all the data.

Area	H	J	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
H	0.003873	0.0002161	-0.000332	0.001652	-0.008291	0.007674	0.0003411	-0.002095	-0.0004794	0.001423	0.001234	-0.003278	0.001083	0.0005307	-0.003233	-0.0004235
J	9.541e-05	0.004116	0.01626	-0.01594	0.003385	-0.002981	-0.01406	0.004441	0.008737	-0.02034	-0.001295	0.01467	-0.01286	-0.009608	0.02118	-0.0005786
M	0.00297	-0.004284	-0.001204	0.001848	-0.007862	0.001443	0.0007527	0.002357	0.003059	-0.00334	0.00666	0.0036	-0.021	0.001607	-0.003973	0.0002306
N	-0.001597	-0.007082	0.01438	-0.0002526	0.001971	-0.001675	-0.004043	0.0008598	-0.0003448	0.004935	-0.003481	-0.01227	0.002261	-0.007363	-0.006811	0.0005386
O	-0.01386	0.008809	0.005653	0.003742	-0.01071	-0.002842	0.01702	0.00415	-0.006025	0.003228	-0.008153	0.001357	-0.000937	-0.002304	-0.005548	-0.01789
P	-0.002557	-0.003352	-0.01377	0.0008477	0.00439	-0.003481	-0.001419	0.003363	-0.0001416	-0.007037	-0.005576	-0.00139	0.001006	-0.001715	-0.005431	0.03025
Q	-0.0002746	-0.002882	-0.001671	-0.0006238	-0.002501	0.001099	-0.002074	-0.008253	-0.001425	-0.001428	-0.001885	7.429e-05	-0.001182	0.001478	-0.0002606	0.002357
R	-0.004455	-0.0006031	-0.0185	-0.003874	0.001437	-0.003247	0.001684	-0.001579	0.0007279	-0.006585	-0.008864	0.001218	0.002426	-0.001891	-0.001848	-0.004779
S	8.243e-06	-0.004031	-0.003004	-0.006607	-0.003919	-0.002102	0.0004314	-0.005647	0.0001451	0.006407	0.0007526	0.0001059	0.007186	-0.0004546	-0.001298	-0.001785
T	-0.001752	0.004025	-0.001278	0.002076	0.001502	-0.005315	-0.001043	-0.001033	0.00174	0.0008497	-0.0002421	0.0007073	-0.01187	-0.000974	0.004377	0.001397
U	0.002762	-0.001516	0.002865	0.003974	0.005124	0.001594	0.004763	0.002069	-0.005543	0.001933	-0.001668	-0.003752	0.005928	0.001625	0.00209	-0.003574
V	0.005318	-0.000957	-0.004742	0.003688	-0.0002434	0.004232	0.01124	0.003481	-0.003229	0.003967	0.001741	-0.006348	0.007664	0.001452	-0.005897	-0.0105
W	-0.001555	0.002035	0.006616	0.0008285	-0.000474	-0.002844	0.004886	0.0004289	0.001147	0.0022	0.001597	0.005676	0.002806	0.006128	0.003213	-0.005376
X1	0.001	0.001943	0.002197	0.005126	0.001029	0.004715	-0.007919	0.0006539	-0.0006138	0.002838	0.006651	0.002349	0.008771	0.005922	-0.0003107	-0.004001
X2	-0.002646	-0.001541	-0.000284	0.000806	0.006488	9.549e-05	-0.005299	-0.000165	0.004173	0.002748	0.001344	0.001213	0.001772	0.001332	0.0009324	-0.002098
X3	0.009356	0.003808	-0.006418	0.002531	0.005035	0.00259	0.0001105	-0.003872	-0.004482	0.0006854	0.006974	0.0008433	0.003008	0.001282	0.003175	0.02065
Y	0.00283	0.0005222	0.003241	-0.001382	0.007358	0.0008307	-0.003814	-0.0003897	0.002076	0.005063	0.001684	-0.002301	0.00558	0.003247	-0.0006857	-0.0006987
Z	0.0004801	0.0007761	0	0.001565	-0.003715	0.0002145	-0.001559	0.001229	0.0004783	0.00245	0.002527	-0.002473	-0.001644	-0.0002944	0.0003289	-0.003727

Table 3 - This matrix is then a summary of what happens to ratios when a seat section (column-wise) get's sold out.

References

Baker, W.L., E. Lin, M.V. Marn and C.C. Zawada. (2001). "Getting Prices Right on the Web." The McKinsey Quarterly 2.

Biller, S., L.M.A. Chan, D. Simchi-Levi and J. Swann. (2000). "Dynamic Pricing and the Direct-to Customer Model in the Automotive Industry." GM Research & Development Center.

Bouchet, A., Troilo, M. and Walkup, B.R., 2016. Dynamic pricing usage in sports for revenue management. *Managerial Finance*.

Cain, B., Saporoschetz, N. and Ginting, T., 2020. Out of the Box: A Dynamic Pricing Model for Professional Sports Teams. *The Journal of Purdue Undergraduate Research*, *10*(1), p.8.

Den Boer, A.V., 2015. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, *20*(1), pp.1-18.

DiMicco, J.M., A. Greenwald and P. Maes. (2001). "Dynamic Pricing Strategies under a Finite Time Horizon." In Proceedings of the ACM Conference on Electronic Commerce (EC '01), Tampa, FL.Morris, J.,

DiMicco, J.M., Maes, P. and Greenwald, A., 2003. Learning curve: A simulation-based approach to dynamic pricing. *Electronic Commerce Research*, *3*(3), pp.245-276.

Gallego, G. and G. van Ryzin. (1997). "A Multiproduct Dynamic Pricing Problem and Its Applications to Network Yield Management." Operations Research 45(1), 24–41.

Goodman, L.A. and Kruskal, W.H., 1979. Measures of association for cross classifications. *Measures of association for cross classifications*, pp.2-34.

Jiaqi Xu, J., Fader, P.S. and Veeraraghavan, S., 2019. Designing and evaluating dynamic pricing policies for major league baseball tickets. *Manufacturing & Service Operations Management*, *21*(1), pp.121-138.

Kemper, C. and Breuer, C., 2016. How efficient is dynamic pricing for sport events? Designing a dynamic pricing model for Bayern Munich. *International Journal of Sport Finance*, *11*(1), pp.4-15.

Kimes, S. E. (1989). The basics of yield management. Cornell Hotel and Restaurant Administration Quarterly, 30(3), 14-19.

Kimes, S. E., Chase, R. B., Choi, S., Lee, P. Y. & Ngonzi, E. N. (1998). Restaurant revenue management: Applying yield management to the restaurant industry. Cornell Hotel and Restaurant Administration Quarterly, 39(3), 32-39.

McGill, J.I. and G. van Ryzin. (1999). "Revenue Management: Research Overview and Prospects." Transportation Science 33(2), 233–256.

Morris, J., P. Maes and A. Greenwald. (2001). "Learning Curve: Analysis of an Agent Pricing Strategy Under Varying Conditions." In Proceedings of the International Conference on Artificial Intelligence (ICAI), Las Vegas, NV.

Narahari, Y., Raju, C.V.L., Ravikumar, K. and Shah, S., 2005. Dynamic pricing models for electronic business. *Sadhana*, *30*(2), pp.231-256.

Nufer, G., & Fischer, J. (2013). Ticket pricing in European football - Analysis and implications. Sport and Art, 1, 49-60.

P. Maes and A. Greenwald. (2001). “Learning Curve: Analysis of an Agent Pricing Strategy Under Varying Conditions.” In Proceedings of the International Conference on Artificial Intelligence (ICAI), Las Vegas, NV

R.G. Cross. Revenue Management. Crown Business, New York, NY, 1997.

Shapiro, S. and Drayer, J. (2012), “A new age of demand based pricing. An examination of dynamic ticket pricing and secondary market prices on Major League Baseball”, Journal of Sport Management, Vol. 26 No. 6, pp. 532-546.

Sheryl E Kimes and Jochen Wirtz. Has revenue management become acceptable? findings from an international study on the perceived fairness of rate fences. Journal of Service Research, 6(2):125–135, 2003.

Sheryl E Kimes and Jochen Wirtz. Perceived fairness of revenue management in the us golf industry. Journal of Revenue and Pricing Management, 1(4):332–344, 2003.

Sunmee Choi and Anna S Mattila. Hotel revenue management and its impact on customers’ perceptions of fairness. Journal of Revenue and pricing Management, 2(4):303–314, 2004.

Sunmee Choi and Anna S Mattila. Impact of information on customer fairness perceptions of hotel revenue management. Cornell Hotel and Restaurant Administration Quarterly, 46(4):444– 451, 2005.

Won, D.C. and Lee, Y.H., 2008. Optimal dynamic pricing for sports games with habitual attendance. *Managerial and Decision Economics*, *29*(8), pp.639-655.