If You Build It, Will Soccer Fans Come?
Evaluating the Factors that Make Markets Viable for Major League Soccer

by
Eric Roden

Abstract

Major League Soccer (MLS) is growing quickly by improving its existing franchises, and also by expanding to new markets, adding new franchises. MLS has the potential to see significant gains from this expansion into new markets; however, as with every business decision, there are risks. There is always the potential that a new franchise may fail if placed in the wrong city. This is a concern with expansion in any professional sports league, but particularly so in one as young as Major League Soccer. Given the sizable risk, it is surprising that relatively little research exists on the evaluation of potential markets for new sports franchises. This study fills this void by analyzing how a variety of factors that characterize a geographic market – i.e., population, community attachment, the presence of other professional sports franchises, household entertainment spending, and education level (at a metropolitan statistical area level) – affect MLS attendance. Results indicate a positive relationship between population, entertainment spending, and education level with the success of MLS franchises while a negative relationship exists between the presence of other professional sports franchises and community attachment with franchise success. Subsequently, the observed relationships are used to predict which geographic markets would be most suitable for future Major League Soccer expansion in America, with Atlanta, San Francisco, Miami, and Phoenix appearing most attractive.

Key words: Major League Soccer, new market entry, market viability, professional sports expansion

I. Introduction

Soccer is growing in America; participation in the game has risen steadily in recent years. US Youth Soccer reported an increase of over 200,000 registered players from 2013 to 2014[1], and the top professional league in the country, Major League Soccer (MLS), has seen its per game average attendance grow the last three years running, increasing by almost 2,500 last year alone. This recent surge has indicates that the MLS is now the fastest growing American professional sport[2]. MLS attendance has seen a steady rise since its founding in 1993(Figure 1) and MLS commissioner Don Garber is trying to ensure that the league captures as much of the American market as possible by continuing to expand the brand and improving profitability.

Figure 1: MLS Average Single Game Attendance Over Time
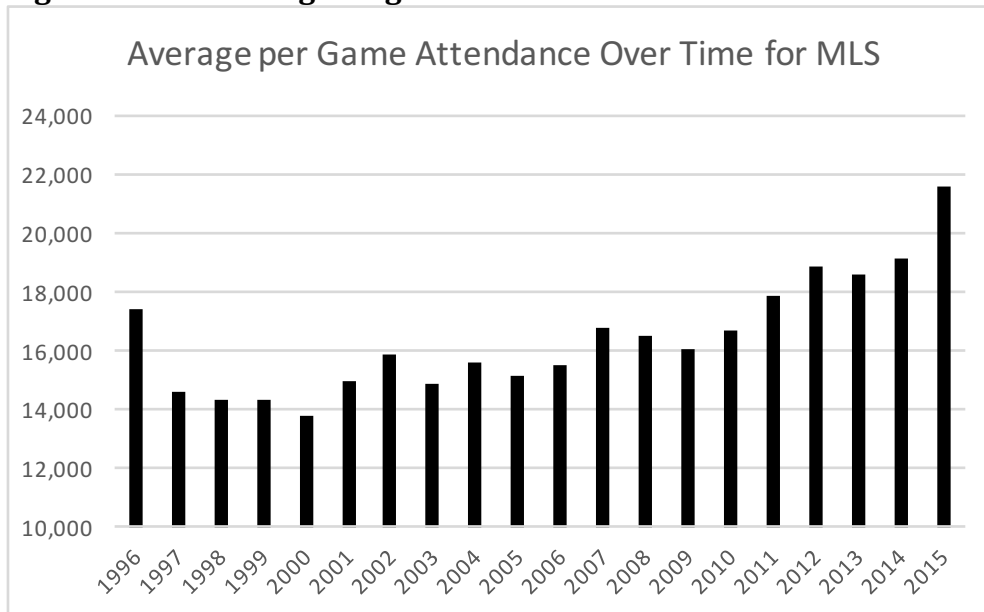
MLS attendance has seen gradual growth over the last decade following a sharp drop in attendance in its second season. Note: Attendance starts at 10,000, not at 0.

---

[1] http://www.usyouthsoccer.org/media_kit/keystatistics/
[2] https://en.wikipedia.org/wiki/Major_League_Soccer_attendance#MLS_attendance_vs._other_North_American_professional_sports_leagues

Since its first season, 20 years ago, the MLS has repeatedly expanded into new markets. Starting with only 10 teams in 1993, the league grew to 20 teams in 2015. The league plans to expand to 24 teams by 2018, with new franchises already announced for Atlanta, Minneapolis/St. Paul, and Miami. MLS's path to this point has been bumpy, as several franchises have folded over the course of the league's history. For example, after six years in the league and relative success on the field, the Tampa Bay Mutiny, a founding member of the MLS, folded in 2001; the Miami Fusion dropped from the league that same year, after only four years of play, the last of which saw them finish in first place at the end of the regular season; and in 2014, Los Angeles based Chivas USA folded, despite the comparative success of cross-town rivals the LA Galaxy, who maintain one of the largest fan bases in MLS. Although a lack of on field success may have played a part in the downfall of some of these franchises, a driving factor was simple lack of attendance.

In contrast, successful franchises, such as Seattle Sounders FC, Sporting Kansas City, and the Portland Timbers, routinely pack their stadiums and appear stable enough to last years into the future, despite being located in smaller markets. This inconsistency begs the question: what factors are most important in determining the viability of a market for Major League Soccer franchises, and what potential markets, presently lacking teams, would be most viable for expansion?

In order to determine what factors make a market viable, and ultimately which markets would make the best expansion locations, it is perhaps useful to examine which factors associate with high attendance in existing MLS markets. Moreover, if we make the assumption that a model of attendance in current MLS metro areas could be extended to predict attendance at new sites, the results of this analysis might be used to inform the

selection of franchise expansion locations in the future. It is likely that some of the best MLS

markets are already playing host to a team; however, the selection process thus far has

been mostly based on metropolitan population size and the presence of a willing

ownership group. Comparing Panel A to Panel B in Figure 2 shows, attendance depends on

a complex array of factors, beyond simple metro population. Young franchises, such as

Orlando City FC, Seattle Sounders, and Portland Timbers, who rank among the most

attended MLS franchises, despite all being initially overlooked, lend credence to the idea

that good markets likely remain untapped.

Figure 2: Comparing Population to Attendance in Existing MLS Locations
The two figures compare the metropolitan area population of all current US cities with a MLS franchise to their 2015 MLS attendance. Comparing the two maps shows that although there may be some relationship between population and attendance, it is far from the only factor at play.

Panel A: 2014 Metropolitan Statistical Area Populations for MLS Cities

2014 Metropolitan Statistical Area Population

Panel B: 2015 Attendance for MLS Cities

MLS Attendance 2015

Note: New York City has two teams; however, for readability, only the attendance for New York City FC is shown, the larger of the two in 2015. NYC FC averaged 29,016, while the New York Red Bulls averaged 19,657. Graphic created using Tableau public.

To develop a model of attendance, general demographic and consumer information for cities that currently host MLS teams was analyzed. Specifically, the impact that community attachment, the availability of substitute entertainment options (i.e., other sports franchises), education level, population, and psychological need for aesthetics have on MLS attendance was investigated. This analysis was performed via multiple regression, assessing the relationship between fan attendance and location factors. Further analysis also examined the above-described factors ability to predict which markets have a franchise (i.e. the presence of an MLS franchise). The development of the second model allows for comparison between using attendance as a dependent variable and using just the presence of a team in the market. Subsequent analysis assessed which markets make the best fit for future MLS expansion with consideration given to both models.

Although there exists a large amount of research that studies the factors that motivate consumers to attend sporting events, and a smaller body of research that has touched on professional sports expansion, this study is novel in that it provides what is to my knowledge a first attempt to identify and compare possible new markets for the creation of soccer franchises, based on an assessment of the common features of cities that presently host successful franchises. The findings of this study therefore have the potential to inform the location choices for new MLS franchises. This would impact local and national soccer enthusiasts by providing improving their ability to make a case for receiving an MLS franchise. Moreover, in turn, it would benefit those local economies, through the jobs associated with the creation of any new franchise.

More specifically, this study addressed five hypotheses and utilized the results to evaluate the predicted success of a hypothetical new MLS franchise in various potential markets. The five hypotheses examined are as follows:

- **Hypothesis 1: Community Attachment -** *A positive relationship exists between community attachment,* as measured by the percentage of the population who was born in the city in which they currently live, *and MLS franchise success.*

- **Hypothesis 2***: **Total Sports** - A negative relationship exists between the number of other professional sports teams in a market and MLS franchise success.*

- **Hypothesis 3***: **Need for Aesthetics** - A positive relationship exists between the need for aesthetics and MLS franchise success.*

- ***Hypothesis 4****: **Population*** *- A positive relationship exists between population size and MLS franchise success.*

- ***Hypothesis 5****: **Knowledge*** *- A positive relationship exists between desire to acquire knowledge and MLS franchise attendance.*

The two regression models generated subsequently both suggest that Atlanta, San Francisco, Phoenix, and Miami hold the greatest potential as top expansion locations.

II. Literature Review

To determine which factors have the greatest impact on attendance in MLS markets and in predicting the presence of an MLS franchise, the existing literature on sports consumption and motivation was consulted, resulting in several insights. Most of the existing literature focused on the impact various factors have on certain psychological outcomes, such as attachment and motivation, in a specific population. Additionally, the psychological studies typically focused on a single market and a single sports team. Finally, the literature around studies that specifically focused on understanding professional sports expansion through economic modeling failed to account for the degree of success of existing franchises, instead focusing on just the presence or lack of a franchise in a metro area.

The design of this study addresses both of the issues mentioned above. First, the study design examines data from multiple markets and multiple teams to develop a clearer picture of the breadth of applicability of the psychological findings. Second, this study compares the results of models based on a measure of success (attendance) to those based on the presence of a franchise.

i. The one market, one sport approach

       As mentioned above, one recurring trend through the literature was that researchers typically limited their studies to one market, and the fans of one specific franchise. For example, James, Kolbe and Trail (2002) studied specifically individual supporters of a specific new Major League Baseball franchise. Kim and Trail (2010) used an even more specific approach, focusing their research on the fans in attendance at a single WNBA game. Funk, Beaton and Alexandris (2012) took the broadest approach in their study, which examined consumers of various Australian rugby and football teams. This paper, however, still treated the sample as one population and did not look to examine the differences between the two sports' fans or the various markets studied. Keeping the studies to a limited number of markets and franchises likely made it easier to identify factors that were affecting the population because it held socio-cultural and geographic factors relatively constant. Limiting the studies breadth of markets and franchises and focusing on the psychological outcomes compared to actual attendance also had its weaknesses. The literature left a gap in determining what findings can be generalized to other populations and also did not allow for cross-market comparisons.

       Additionally, the literature typically took a more psychological approach to sports consumption behavior rather than directly focusing on attendance. Uhlman and Trail (2012), for example, focused on variables' correlation with motivation rather than actual attendance and James et al. (2002) focused primarily on levels of psychological involvement. Kim and Trail (2010) focused on the causes of variation in attendance and used psychological variables examined in other research, such as attachment to the team,

implying that increased motivation ultimately leads to increased consumption. Additionally, focusing on motivation and similarly abstract concepts provided a basis for determining the factors that were likely to be the most significant in determining what makes a good MLS market.

ii. Similar factors influence sports consumption across studies

Another trend in the existing research was the emergence of several factors which influence sports consumption, and thus attendance, in independent studies with independent populations. For example, Funk et al (2010) and Uhlman and Trail (2012) both determined that the fans' sense of esteem and superiority influenced consumption. Uhlman and Trail also found that consumption correlated with various forms of attachment, consistent with Kim and Trail (2010). The literature review uncovered further similarities which informed the hypotheses of this research. I discuss those similarities in coordination with the corresponding hypotheses below. Finding these consistencies allowed for the evaluation of markets based on these shared factors.

Overall, these findings provided the basis for the rest of this research, but they also presented some problems and illustrated areas where the current literature fell short. With the exception of Kim and Trail (2010), the majority of prior research focused on factors that are positively correlated with consumption. In contrast, Kim and Trail discuss the constraints that limit individuals' ability to consume sports, such as disposable income. These are also important factors to consider because they clearly impact attendance.

Additionally, factors beyond basic demographic considerations impacted individual's motivation to consume sports. Funk et al (2012) and Uhlman and Trail (2012),

found that factors related to the franchises' accomplishments also played an important part in determining consumption habits – i.e., the success of the franchise had a significant impact on attendance. Kim and Trail (2011) supported this by finding that a team's success acted as a motivator to attendance and on-field failure acted as a constraint, limiting attendance. In fact, the study conducted by Kim and Trail, they also evaluated various factors impact on attendance and found that lack of success accounted for 10% of the 34% of variance their model explained. The impact ranked only behind team attachment in their study. A team's success is not an inherent feature of a market in which a team is located, but rather a function of management and operations. Therefore, it was important to control for this in order to compare the viability of markets, which is independent of the impact of management decisions once the franchise entered the market. For example, if a team is managed exceptionally well and consistently places near the top of the league then they would have a higher attendance than if the same team were performing poorly, and that variation has little to do with the viability of the market but rather the management of the franchise.

iii. Neglecting franchise success to focus on just the presence of a franchise

The final insight that emerged from the literature, and perhaps the most important, showed that the few studies that focused on professional sport expansion used the presence of a team to indicate a good market, not the success of those teams. Bruggink & Zamparelli (1999), Baehr (2006), and Davis (2006) all took this approach in modeling professional baseball, professional soccer, and minor league baseball respectively. Each model used the presence of a franchise as the dependent variable and conducted

multivariable regressions. Brugginck & Zamparelli and Baehr used the results to see which geographies were most similar to those with franchises already. They used that result to determine ratings for new franchise locations. This provided a foundation of proven variables useful in predicting the location of franchises and also gave insight into the best way to statistically analyze this type of problem.

Focusing on the presence of teams in markets and using this to generate results implies an assumption that the existing franchises were all placed in desirable markets and that the same decision mechanism used for placing these franchises should be used for placing future franchises. The approach fails to learn from previous decisions. Additionally, by focusing on just the presence of a team, all teams are valued equally. For example, in soccer, as mentioned, the Seattle Sounders consistently have the highest yearly average attendance and thus cities more similar to Seattle would be desirable for expansion. However, using the above described approach, the characteristics of Seattle would be diluted by the noise created by cities like Denver, home of the Colorado Rapids, one of the least successful franchises in MLS. In short, MLS wants more teams like Seattle, yet by focusing on what predicts team presence, cities like Denver would be viewed as no more or less ideal than cities like Seattle because analysis failed to evaluate actual indications of team performance.

iv. Summary

Considering these themes and their shortcomings, it was apparent that this paper served as a bridge that connected and integrated the findings of prior, isolated works. This paper holistically examined a number of factors that have repeatedly been observed to play

a significant role in determining consumption and attendance in the literature, across a variety of studies and markets. By considering these factors jointly, and exploring their effects across a number of geographic markets, it became possible to reliably identify the characteristics that make an area particularly suitable for MLS expansion. Applying this approach to American professional soccer, the present study expanded upon the "very little research [which] has been done on this topic of the location of expansion teams in professional sports" (Bruggink & Zamparelli, 1999, p. 50).

III. Methodology

i. Formulation of Hypotheses

To effectively address the research question, a number of hypotheses were evaluated via multivariable linear regression in R. Each of the hypotheses focused on measuring a market's potential, rather than the ability of management to maximize that potential, conditional on entry. Data regarding the factors needed to evaluate each hypothesis came entirely from the US Census and US Consumer Expenditure Survey. Accordingly, factors under the control of team management, such as a team's on-field success, were incorporated merely as control variables; these variables were not the focus of the analysis because they did not determine the inherent viability of the market, and could not be used to inform decisions about where to locate a new franchise in the future. Data regarding team specific variables came from Wikipedia and was verified through the MLS's website, team websites, and other online resources.

Each hypothesis focuses on franchise success as the dependent variable; however, the factors discussed in the hypotheses were also examined in the context of their ability to

predict the presence of a franchise (**presence model**) in a market and the resulting

regression model provided a benchmark to which the regression based on attendance

(**attendance model**) could be compared.

Hypothesis 1: Community Attachment *- A positive relationship exists between community attachment and MLS franchise success.*

The first factor considered was community attachment. I proposed that metro areas

with a higher sense of community attachment would be more successful overall. This

hypothesis draws from research by Uhlman & Trail, which found that Seattle Sounders fans

cited city pride and attachment as a more significant factor in their loyalty to the team than

their actual passion for the game of soccer, or even the specific team. This theory was

derived from previous work done by Anderson and Stone (1981), Funk et al (2001), and

Mahony, Nakazawa,, Funk, James, & Gladden (2002).

Hypothesis 2: Total Sports - *A negative relationship exists between the number of other professional sports teams in a market and MLS franchise success.*

Next, I considered the presence of alternative entertainment options, which

intuitively possessed a negative impact on MLS franchise success. Kim & Trail (2010) found

this to be the case when they examined factors that affect attendance at a women'

professional basketball game. This held weight particularly for MLS because both MLS and

women's professional basketball are relatively young professional sports, which are quite

likely to be eclipsed by larger leagues in the same market, specifically teams from the big

12

four leagues of North American sports: the MLB, NBA, NFL, and NHL. The presence of teams from other sports leagues was expected to detract from interest in an MLS franchise because, in the eyes of general sports fans, MLS is a substitute, and potentially even an inferior good.

Additional support for this hypothesis was provided in the work of Funk et al., which noted that some of the appeal of attending major sporting events was attributed to the novelty of watching sports in person. The more opportunities to watch professional sports in a particular geographic market, the less attention an MLS franchise receives. There existed some literature which argued that having additional professional sports franchises established a city as a "sports city." Proponents of this line of thinking argued that supporting teams becomes part of the local identity, leading it to be a "major league" city (Bruggink & Zamparelli 1999). This may hold merit for larger leagues; however, for a young league, like MLS, it seemed more likely that a franchise would be overshadowed by other professional franchises.

Hypothesis 3: Need for Aesthetics - *A positive relationship exists between the need for aesthetics and MLS franchise success.*

A study conducted by Funk et al. (2012) indicated that there was a positive correlation between consumers need for aesthetic value and their consumption of sports. Aesthetic value, according to Funk et al. "represents performance as the inherent desire for aesthetic representation and movement, desire for excellence, beauty, and creativity of athletic performance, fantasy and subjective experiences that are pleasurable" (359). In

their research it was determined that the higher the need for aesthetics an individual had, the more likely they were to consume sports.

Hypothesis 4: Population - *A positive relationship exists between population size and MLS franchise success.*

Another important factor to considered was the general population of a metro area. By having a larger population in the metro area, there are more potential attendees, and an overall higher number of potential customers. In previous professional sports expansion models such as the one put forward by Bruggink & Zamparelli, population size was a top predictor of the presence of a professional sports franchise. Population was also a top predictor of MLS franchise locations according to Baehr (2006). Although these previous models only sought to predict the presence of a team, not a team's success, it was still important to evaluate sheer population in this work, because it was quite likely to predict MLS franchise attendance, and thus success. Additionally, population was important to include in a predictive model because, at some point, it does likely become a limiting factor. That is, in some markets, the population is simply too small for an MLS franchise to be economically viable. In fact, population alone serves as a reasonable predictor of franchise presence as illustrated in Figure 3.

Figure 3: Presence of an MLS franchise by population for all metro areas in the US

Panel A: Population Distribution of MSAs with an MLS Franchise

Population of MLS MSAs
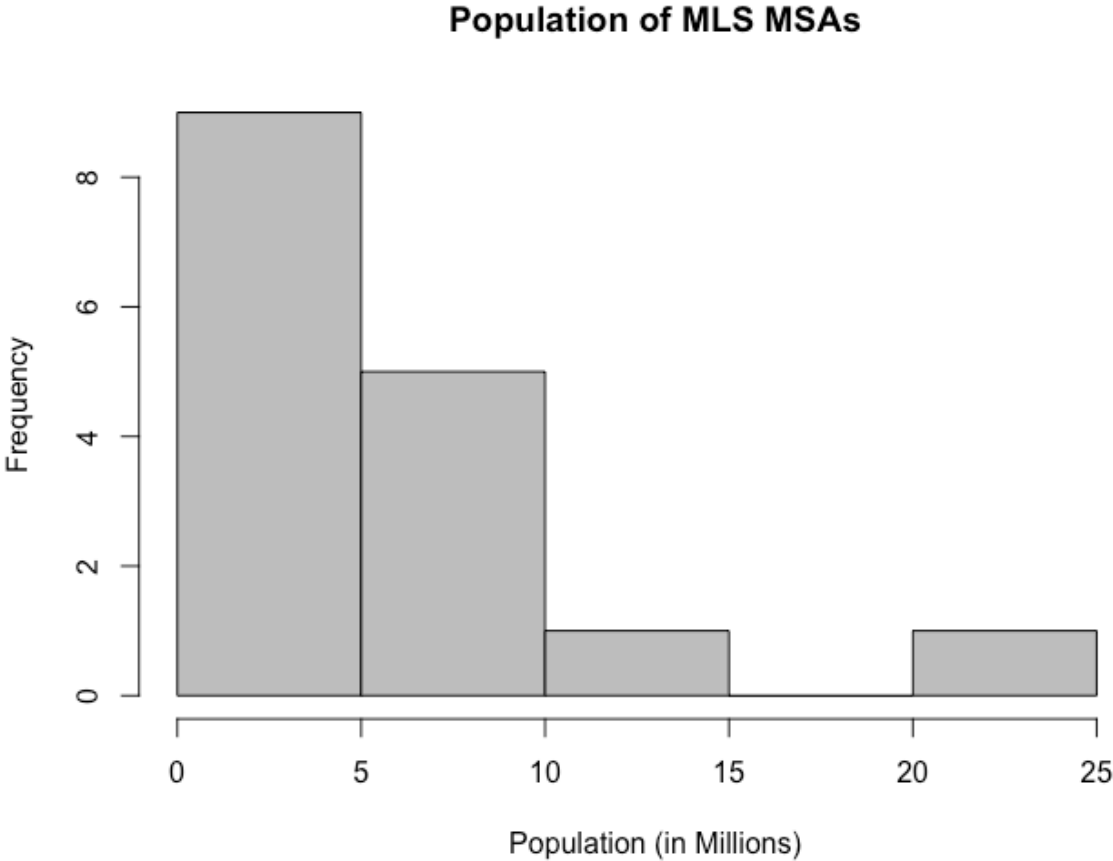
Panel B: Population Distribution of MSAs with out MLS franchises

Population of Non-MLS MSAs

Population (in Millions)
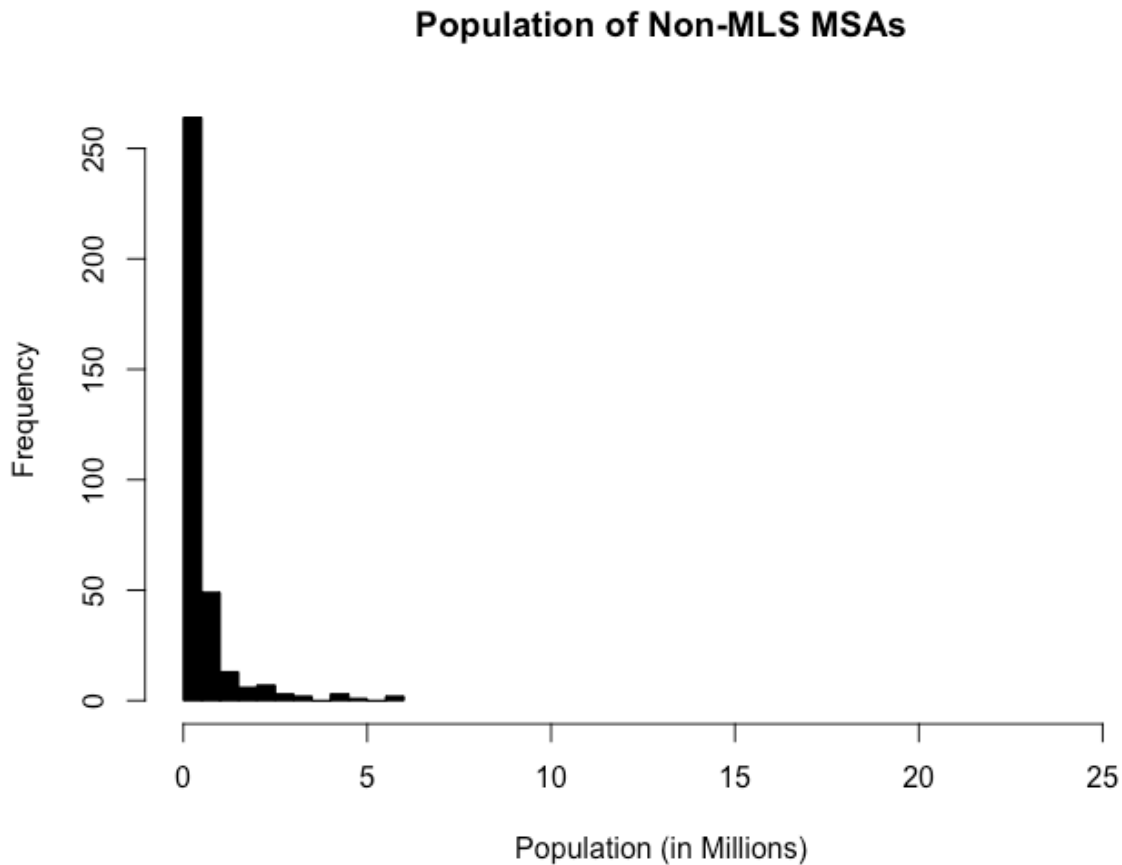
As the cities get larger, you become more likely to find an MLS franchise located in that city.

Note: The Y-axis changes from Panel A to Panel B due to the higher overall volume of MSAs with out franchises.

Hypothesis 5: Knowledge - *A positive relationship exists between desire to acquire knowledge and MLS franchise attendance.*

The final factor evaluated was the local population's desire to acquire knowledge. This variable had not been included in other professional sports franchise prediction models; rather, it was drawn from fan motivation theory. James and Trail cited the

acquisition of knowledge as a factor that played an integral role in fan behavior. Fans often

viewed sports as a hobby to be studied, indicating a higher desire to acquire knowledge

about their team. This often led to support of a specific team in the form of TV viewership

and, importantly for this study, game attendance.

ii. Data Collection

Due to the scope and nature of the research question, primary data collection was

waived. Fortunately, a number of secondary data sources were readily available that

addressed the research question. A number of proxies were identified, operationalizing the

constructs discussed in the hypothesis development section. First, average single game

attendance was identified as the most suitable proxy for franchise success and collected

from worldfootball.net[3]. Smaller leagues such as the MLS and WNBA tend to rely more

heavily on attendance as a source of revenue while larger leagues rely on TV contracts

(Kim and Trail, 2010). As mentioned in the literature review, the use of some measure of

success as a dependent variable to evaluate multiple hypotheses sets this research apart

from previous studies. The majority of previous sports league expansion models assume

that existing franchise locations were optimal and thus that future expansion sites should

be as similar as possible to those existing sites. By focusing on attendance, and comparing it

to the presence of a team, this study identified population-related factors that associated

with the success of MLS franchises, following their creation.

---

[3]2014 table but each year from 2005-2014 was collected.
http://www.worldfootball.net/attendance/usa-major-league-soccer-2014/

Proxies for community attachment, need for aesthetics, and desire for knowledge were also needed. Comprehensive community attachment measures for major metropolitan areas across America were not readily available and would necessitate extensive data collection, which fell outside the scope of this project. However, the US Census, the main source of data regarding the factors examined in this study, incorporates data on the number of individuals that live in the same state where they were born[4]. This measure was used as a proxy for community attachment. It stood to reason that individuals who were born in a specific state, and ultimately decided to continue living in that same geography, likely developed a degree of attachment to their community.

Similar to community attachment, collecting primary data on the desire for knowledge (as James and Trail did for their initial study) in all metro areas across the US would have entailed a great deal of time and effort, and was thus deemed out of scope for this paper. However, a suitable proxy existed in form of educational attainment[5]. Thus, rather than employ a specific metric of desire to acquire knowledge, the analysis drew upon a measure of the percentage of the population with at least a bachelor's degree. This data was available on a Metropolitan Statistical Area (MSA) level from the US Census. Using this as a proxy for the desire for knowledge followed the logic that if an individual was motivated to complete a bachelor's degree, then they have some desire to acquire knowledge. Certainly factors other than the desire for knowledge motivate (and limit)

---

[4] 2014 table as an example for repeatability.
http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_1 4_5YR_B06001&prodType=table
[5] 2014 table as an example for repeatability.
http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_1 4_1YR_S1501&prodType=table

individuals to pursue higher education, but overall a higher percentage of individuals with bachelor's degrees indicates a higher percentage of individuals seeking out knowledge.

Entertainment spending also required the use of a proxy and thus utilized household entertainment spending. This was a suitable, and even potentially advantageous proxy for a number of reasons. First, it not only represented a consumer's desire to enjoy aesthetics, but also their ability and willingness to pay for those desires. This contrasted previous studies which examined the impact of household income on the presence of sports franchises, an approach that failed to account for how income is spent. Household entertainment spending was collected from the US Consumer Expenditure Survey[6].

MLS point totals for each season were collected from MLS.com[7]. Data about the presence of other 'big four' North American professional sports franchises in each market was drawn from Wikipedia[8].

With respect to sample selection, there were effectively two types of markets that were considered. The first was the sample of markets that currently host MLS teams. All data was collected on these markets (MSAs / franchises) for the period between 2005 and 2014; as far back as census data was readily available. This period saw a fairly consistent set of teams, only one team folded and five were added in the US. Montreal, Toronto, and Vancouver also launched over this period but are excluded from the dataset. Data on those franchises located in Canada (Montreal Impact, Vancouver Whitecaps, and Toronto FC) were excluded from the study because of data inconsistency issues (e.g., the Canadian

---

[6] http://www.bls.gov/cex/csxmsa.htm
[7] 2014 table as an example for repeatability.
http://www.mlssoccer.com/standings/mls/2014/
[8] https://en.wikipedia.org/wiki/U.S._cities_with_teams_from_four_major_league_sports

government records data in different formats, etc.). The second set of markets that needed to be considered were MSAs that did not have MLS franchises – i.e., potential expansion locations. MSAs were excluded from the attendance model projections if their population was less than one million people, because this was the approximate size of the smallest metro area that currently hosts an MLS franchise; the Salt Lake City metro area and training of the model occurred on a dataset where all MSAs possessed relatively large populations so using the model on a dissimilar dataset proved problematic.

Beyond the key independent variables of interest, a number of control variables were also considered. Through research and intuition, it was determined that three factors must be accounted for in any analysis of game attendance: team on-field success, franchise novelty, and league growth. These factors only apply to the attendance model and not the presence model as they depend on the presence of a team in a specific city and were held constant in the presence model. In other words, if control variables entered the presence model, they would hold high predictive value but not actually provide meaningful insights because only markets with a team had values for those factors.

Previous research, such as the work by Kim et al. and that by Parrish (2013), supports the notion that winning impacts attendance greatly. Team performance was therefore measured as a franchise's points per game. Average points per game was used, rather than sheer number of wins, or the win-loss ratio, or total points accumulated over a season. This was because many MLS games end in a tie. Final rankings are determined by total points each year, with a win resulting in three points, a tie resulting in one point, and a loss resulting in no points. Because the number of games played by each franchise varied over time, with the introduction of new teams, raw point totals provided an inconsistent

measure of performance. In contrast, average points per game was a stable, consistent

measure, over time. The literature also revealed that the novelty of a franchise impacted

support for a team, so a franchise age variable was also included (Baade, 1994). Finally,

dummy variables indicating years were also incorporated in the analysis, to capture larger

temporal trends, such as changes in the league's popularity overall. Table 1 contains a

summary of each hypothesis and the data used to evaluate it.

Table 1: Hypotheses and Measures

Hypotheses by Relationship with Attendance			
Label	Hypothesis	Expected Relationship with Attendance	Support
Population	The population of metropolitan statistical area has a positive relationship with attendance	+	The more people in an area, the more potential soccer fans. This has been shown to be true in previous studies.
Sports	The total number of MLB, NBA, NFL, and NHL teams located in a metro area will have a negative relationship with attendance.	-	As MLS is viewed as a lower quality product than these leagues, they act as competitors in the market.
Entertainment	Household entertainment spending has a positive relationship with attendance.	+	Individuals need for aesthetics drives them to sports and entertainment spending shows how much they are willing and able to spend on that need.
Resident	The percentage of a metro population that are currently residents in the state they were born in will have a postive relationship with attendance.	+	Community pride has been shown to be a strong factor in driving soccer attendance. By choosing to live in the state an individual was born in it indicates some level of pride in that community, serving as a suitable proxy.
Education	The percentage of a metro population with at least a bachelor's degree has a positive relationship with attednace.	+	The desire to aquire knowledge has been shown to be a factor that motivates sports attendance. Having the desire to seek higher education can serve as a proxy for this variable in a metro population.

Notes: This table gives each hypothesis a label, states the hypothesis in terms of the independent variables (not the original factors from the literature), states the expected relationship, and provides a brief description of the supporting argument.

iii. Analysis

Once the data was gathered, statistical analysis was conducted in the form of three

multiple linear regression models. The first model was an analysis of how the control

variables impacted attendance. The control variables in this instance included points per

game (on-field success), dummy variables capturing each year, to control for time effects, as well as a franchise age variable.

The formula for this initial statistical model was as follows:

$$Attendance = \beta_0 + \beta_1 Points\ Per\ Game + \beta_2 Franchise\ Age + \beta_3 2014\ Dummy$$
$$+ \beta_3 2013\ Dummy \ldots + \beta_{12} 2006\ Dummy$$

After gaining a better understanding on the impact that winning, novelty, and yearly changes[9] had, the full model expanded on the baseline model, to include all independent variables.

$$Attendance = \beta_0 + \beta_1\ Points\ Per\ Game + \beta_2 Franchise\ Age + \beta_3\ 2014\ Dummy$$
$$+ \beta_4\ 2013\ Dummy + \cdots + \beta_{12}\ 2006\ Dummy + \beta_{13}\ Population$$
$$+ \beta_{14}\ Entertainment\ Spending - \beta_{15}\ Total\ Sports\ Franchises$$
$$+ \beta_{16}\ Resident\ in\ State\ of\ Birth + \beta_{17}\ Education$$

Each of the betas being tested was evaluated based on p-values and was held to a significance level of five percent. The sign of each independent variable was also adjusted to align with the anticipated relationship based on the stated hypotheses (i.e. an expected negative relationship generates a negative coefficient). The only variable this impacted was β_{14} which pertained to the presence of other professional sports teams which was hypothesized to have a negative relationship with attendance. Once the full attendance model was evaluated, it provided coefficients that were used to determine which potential expansion sites would be best for MLS franchises. Both this and the control model were trained on the sample containing existing MLS franchises.

---

[9] The impact of franchise age squared was also considered but ultimately removed as the relationship between success and age appeared to be linear.

Finally, I generated a third model based on the presence of an MLS franchise. This third model relied on the complete set of MSAs for a training data set and the resulting model predicted the likelihood of each MSA having an MLS franchise (presence of a franchise). The presence model removed the control variables from the attendance model because they did not impact the likelihood of a franchise existing, but rather the success of that franchise. Also, all predictions for the presence model utilized 2014 data leading to the removal of the year dummy variables. This resulted in the following model:

$$Presence = \beta_0 + \beta_1\ Population + \beta_1\ Entertainment\ Spending$$

$$- \beta_2\ Total\ Sports\ Franchises + \beta_3\ Resident\ in\ State\ of\ Birth$$

$$+ \beta_4\ Education$$

This presence model originally generated curious results, with many smaller cities receiving higher scores than MSAs that intuitively seemed to better fit the expectations for an MLS expansion location. Upon further inspection, the presence model generated a negative coefficient for population (Table 2) as a result of the covariance between population and total sports (Table 3).

Table 2: Original Presence Model Regression Results

Presence Model Results				
	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	-0.0464227	0.06449926	-0.71974	0.47215206
Population (in Millions)	-0.043720	5.39E-09	-8.1180081	7.60E-15
Total Sports in Market	0.2326853	0.00874925	26.5948983	5.70E-87
Entertainment Spend (in Thousands)	0.035291	2.12E-05	1.6664277	0.09649803
Resident in State of Birth (as a %)	-0.000787	0.00033862	-2.3240674	0.02067788
Education (as a %)	0.00363041	0.00131315	2.76465859	0.00599153

Note: Population had a negative coefficient in the original model resulting in many larger

cities lacking other professional sports being considered poor locations for MLS franchises.

Table 3: Correlation of Original Presence Model

	Team Present	Population	Total Sports	Entertainment Spend	Resident in State of Birth	Education
Correlation Table for Presence with Total Sports Included						
Team Present		0.64053047	0.87461293	0.075903706	-0.140945464	0.264655839
Population	0.640530466		0.84081065	0.09488649	-0.138284209	0.29603859
Total Sports	0.87461293	0.84081065		0.050876625	-0.114987324	0.264474259
Entertainment Spend	0.075903706	0.09488649	0.05087662		-0.159400258	0.042516869
Resident in State of Birth	-0.140945464	-0.1382842	-0.1149873	-0.159400258		0.054656968
Education	0.264655839	0.29603859	0.26447426	0.042516869	0.054656968	

The attendance model also exhibited this behavior but due to it only considering

MSAs with populations over one million the impact was less pronounced. As a result, total

sports was ultimately removed from the presence model leaving the model below as the

complete presence model.

$$Presence = \beta_0 + \beta_1\ Population + \beta_1\ Entertainment\ Spending$$
$$+ \beta_3\ Resident\ in\ State\ of\ Birth + \beta_4\ Education$$

iv. Assumptions

The results of the attendance model and the control model were compared to

evaluate which metro areas provide the most opportunity for MLS expansion. In order to

do this, the theoretical new franchises in the models were assumed to have 1.39 points per

game (the MLS average in 2014), and were assumed to be a new franchise that year,

meaning their franchise age was 0.

IV. Results

Once all data collection was completed, it was analyzed using the R programing language. Specifically, the three multiple linear regressions described in the methodology section were created to address what factors were important in determining the viability of an MLS market as measured by attendance and by presence. The coefficients from the full (control variables and independent variables) attendance regression model were then used to project 2014 attendance for potential MLS expansion locations, and the independent variables from the presence model were used likewise. This section details the results of those analyses.

i. Evaluation of Hypotheses

Table 4: Summary of Hypothesis Tests

Full Model Results			
	Coefficient	Significance	Supported?
Population (In Millions)	850	**	Supported
Total Sports	-1347	*	Supported
Entertainment Spending (in Thousands)	3849	*	Supported
Resident In State of Birth (as a %)	-145	***	Not Supported
Education (as a %)	277	*	Supported
	R-Squared	0.527	

*Table 4 shows the results of a multiple linear regression ran on all control variables and all independent variables. "***" indicates a significance <.001, "**" < .01, "*" < .05, and " " > .05.*

The five hypotheses posed by this study were evaluated by creating a multivariable linear regression model with attendance as the dependent variable. Findings indicated that *a **positive** relationship existed between **population** size and **MLS franchise success**.* Results also indicated a statistically significant relationship between attendance and population. For each additional one million people in a metro area, a team saw approximately an 850 person increase in per game attendance. However, for larger cities, counter to this impact, a

negative relationship existed between the **number of other professional sports teams** *in a market and* **MLS franchise success**. Essentially, although larger populations tend to lead to more support, they also tend to attract more competition in the form of other sports franchises, each of which reduced projected attendance by approximately 1,300 per game.

The results further indicated that *a* **positive** *relationship existed between* **spending on entertainment** *and* **MLS franchise success**. Metro areas that spent more on entertainment saw attendance rise approximately 4 people per every additional dollar spent on entertainment on average. Support also indicated *a* **positive** *relationship between* **educated share of the population** *and* **MLS franchise success**. MSAs where a greater percentage of individuals completed a least a bachelor's degrees tended to better support MLS franchises.

Counter to expectations, testing revealed *a* **negative** *relationship exists between* **community attachment** *and* **MLS franchise success**. Essentially the results indicated that the higher percentage of people living in their place of birth, the less they supported an MLS franchise.

As a point of comparison, each hypothesis included in the presence model (population, need for aesthetics, desire to acquire knowledge, and community attachment) resulted in a coefficient with the same sign, signifying a comparable relationship with presence. However, for the presence model only population and desire for knowledge (as educational attainment) resulted in statistically significant findings (Appendix A). Comparing the results of the control model, the attendance model, and the presence model yielded Table 5 below which compares the resulting beta values (coefficients) and significances of each factor considered in each model. Table 8 also provides the R^2 value

that resulted for each model with the attendance model holding the highest value followed

by the presence model and lastly the control model.

Table 5: Comparison table of results of the control, attendance, and presence models

	Control Model		Attendance Model		Presence Model	
	Coefficient	Significance	Coefficient	Significance	Coefficient	Significance
(Intercept)	12285.1781	***	8467.41721		-0.0090	
Population (in Millions)	N/A	N/A	849.767306	**	0.0744	***
Total Sports	N/A	N/A	-1347.2231	*	Removed due to covariance	
Entertainment Spending (in Thousands)	N/A	N/A	3848.95421	*	0.0044	
Resident in State of Birth (as a %)	N/A	N/A	-144.53152	***	-0.0009	
Education (as a %)	N/A	N/A	277.19112	*	0.0047	*
Points Per Game	5161.31013	***	5186.62129	***	N/A	N/A
Franchise Age	-578.57757	***	-607.17037	***	N/A	N/A
year14	8117.7748	***	6311.80969	**	N/A	N/A
year13	6772.47222	***	5673.34644	**	N/A	N/A
year12	5828.4982	**	4748.15461	*	N/A	N/A
year11	4472.41037	*	3544.04721	*	N/A	N/A
year10	3095.75881		2718.77904		N/A	N/A
year09	2491.42101		1285.71087		N/A	N/A
year08	2835.32497		1106.49422		N/A	N/A
year07	2523.05977		1721.35158		N/A	N/A
year06	508.881575		1044.25686		N/A	N/A
R-Squared	0.371		0.527		0.420	

*Notes:. "***" indicates a significance <.001, "**" < .01, "*" < .05, and " " > .05.*

ii. Control model results

The table below (Table 5) outlines the generic results of the control model which

included dummy variables for each year in the data set, the age of the franchise in that year,

and the points per game associated with the franchise for that year. For the full results, see

Appendix B.

The control model had an R^2 value of .3711, indicating that it explained

approximately 37% of the variance in attendance. This provided a baseline to which the full

attendance model is compared.

Table 6: Summary of Control Model Results

Control Model Results	Coefficient	Significance
Points Per Game	5161	***
Franchise Age	-579	***
2014 Dummy	8118	***
2013 Dummy	6772	***
2012 Dummy	5828	**
2011 Dummy	4472	*
2010 Dummy	3096	
2009 Dummy	2491	
2008 Dummy	2835	
2007 Dummy	2523	
2006 Dummy	509	
	R-Squared	0.3711

*Notes: Table 2 shows that points per game (on-field success), franchise age (novelty), and some years, did have a significant impact on attendance. "***" indicates a significance <.001, "**" < .01, "*" < .05, and " " > .05.*

The control model also showed that several of the control variables were significant, including points per game and franchise age. It also showed that MLS has seen year over year growth with each year having a positive relationship with attendance to varying degrees of significance. These results supported previous research which indicated that novelty and franchise on-field success have an impact on sports consumption and fan motivation. This model also generated predicted attendance values for 2014 for existing MLS locations (Table 6) where were compared to actual attendance numbers as a check of the accuracy of the model.

Table 7: Control Model Predicted Attendance vs Actual Attendance for 2014

MSA	Predicted Attendance	Actual Attendance	Difference	Difference %
Seattle-Tacoma-Bellevue, WA Metro Area	27,213	43,734	(16,521)	-38%
New York-Newark-Jersey City, NY-NJ-PA Metro Area	17,576	19,421	(1,845)	-10%
Boston-Cambridge-Newton, MA-NH Metro Area	18,350	16,681	1,669	10%
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area	24,489	17,631	6,858	39%
Columbus, OH Metro Area	17,885	16,881	1,004	6%
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area	18,969	17,030	1,939	11%
Dallas-Fort Worth-Arlington, TX Metro Area	18,195	16,816	1,379	8%
Chicago-Naperville-Elgin, IL-IN-WI Metro Area	16,617	16,076	541	3%
Denver-Aurora-Lakewood, CO Metro Area	14,840	15,082	(242)	-2%
Los Angeles-Long Beach-Anaheim, CA Metro Area	19,227	21,258	(2,031)	-10%
Portland-Vancouver-Hillsboro, OR-WA Metro Area	26,100	20,806	5,294	25%
San Jose-Sunnyvale-Santa Clara, CA Metro Area	14,531	14,947	(416)	-3%
Houston-The Woodlands-Sugar Land, TX Metro Area	21,710	20,117	1,593	8%
Salt Lake City, UT Metro Area	23,712	20,351	3,361	17%
Kansas City, MO-KS Metro Area	17,421	20,003	(2,582)	-13%

iii. Full model results

The full regression model built off the control. The full attendance model overall had an R^2 value of .527, indicating that it explained an additional 15% of variation in attendance compared to the control model. For the full results, including exact coefficients and significance values, as well as the results of the control variables in the full model see Appendix C. As a check of accuracy, the resulting model generated predicted attendance values for the 2014 season which were compared to the actual attendance for 2014 as shown in Table 7.

Table 8: Full Model Predicted Attendance Vs Actual Attendance for 2014

MSA	Predicted Attendance	Actual Attendance	Difference	Difference %
Seattle-Tacoma-Bellevue, WA Metro Area	29,954	43,734	(13,780)	-31.5%
Portland-Vancouver-Hillsboro, OR-WA Metro Area	28,082	20,806	7,276	35.0%
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area	22,408	17,631	4,777	27.1%
Los Angeles-Long Beach-Anaheim, CA Metro Area	22,150	21,258	892	4.2%
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area	21,458	17,030	4,428	26.0%
Houston-The Woodlands-Sugar Land, TX Metro Area	21,162	20,117	1,045	5.2%
New York-Newark-Jersey City, NY-NJ-PA Metro Area	21,004	19,421	1,583	8.2%
Salt Lake City, UT Metro Area	20,893	20,351	542	2.7%
San Jose-Sunnyvale-Santa Clara, CA Metro Area	17,330	14,947	2,383	15.9%
Boston-Cambridge-Newton, MA-NH Metro Area	16,730	16,681	49	0.3%
Chicago-Naperville-Elgin, IL-IN-WI Metro Area	16,109	16,076	33	0.2%
Dallas-Fort Worth-Arlington, TX Metro Area	15,910	16,816	(906)	-5.4%
Kansas City, MO-KS Metro Area	15,637	20,003	(4,366)	-21.8%
Columbus, OH Metro Area	14,545	16,881	(2,336)	-13.8%
Denver-Aurora-Lakewood, CO Metro Area	13,462	15,082	(1,620)	-10.7%

iv. Projected 2014 attendance and presence for potential expansion locations

Once coefficient variables were determined through the full linear regression model, those coefficients were used to estimate attendance values for potential expansion locations for MLS franchises. As previously mentioned, potential MLS expansion locations were limited to US metropolitan statistical areas with populations over one million people. The results are shown in Appendix D and the projected attendance values in the table represent the expected attendance for an expansion franchise, so for metro areas that already have MLS franchises, this represents a second franchise. As a point of reference, the average MLS attendance for 2014 was 19,148. Utilizing the coefficients generated by the presence model, predicted presence values were also generated for all MSAs in the US, these results are shown in Appendix E. To compare the results of the models, Figure 4 shows a scatter plot of the projected presence model results against the projected attendance model results. These results are limited to populations over one million but the full results can be seen in Appendix F.

Figure 4: Projected Presence vs Projected Attendance for MSAs with Populations over one million
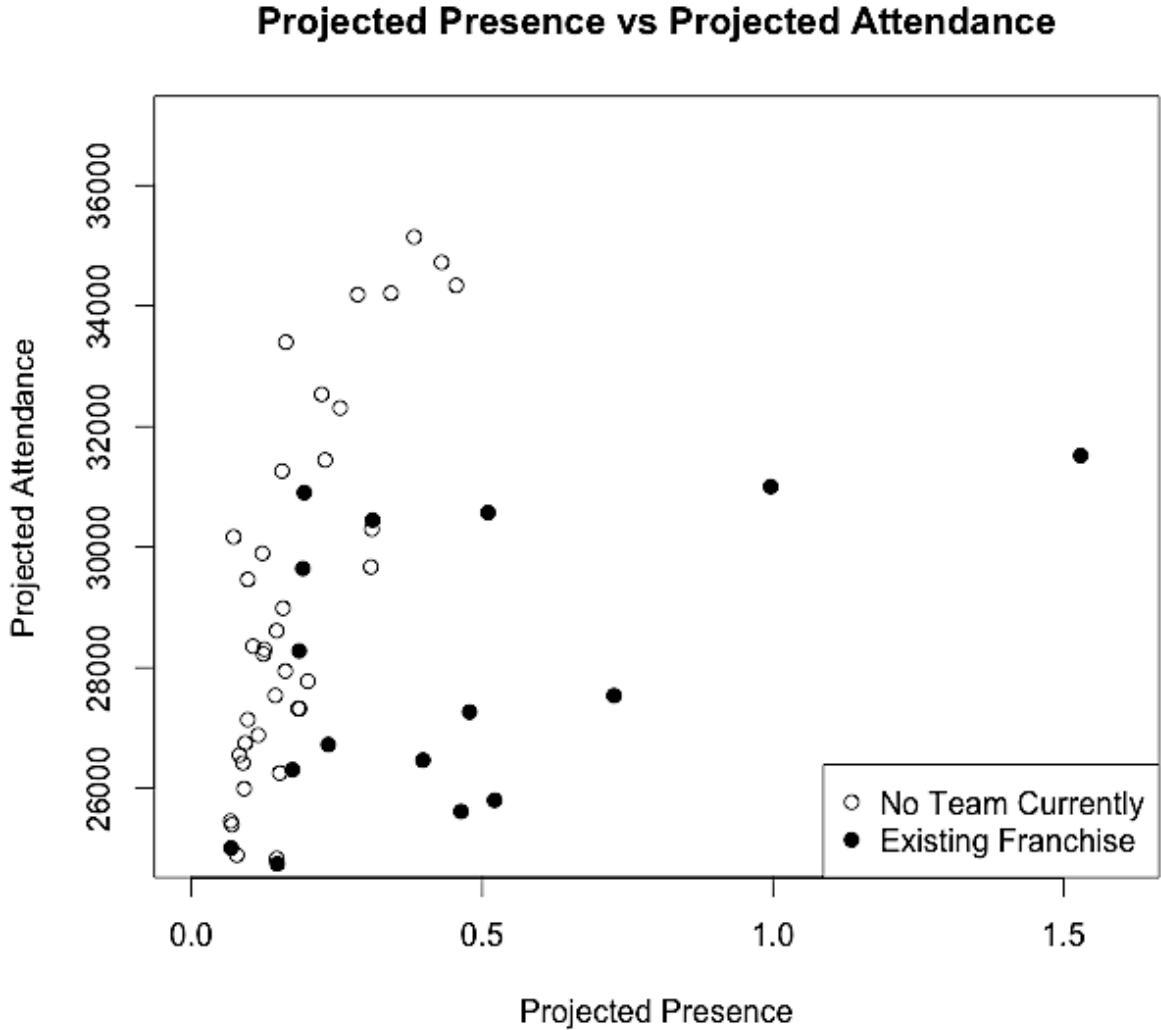

Projected Presence vs Projected Attendance

Figure 4 illustrates the difference between the two models. Notably, the top ten results from the attendance model place above locations with franchises, and all ten of these MSAs scored below a .5 in the presence model, but still rank among the top locations based on that model as well.

V. Discussion

This section will discuss interesting insights, new questions for research, and opportunities for improving upon this study.

i. Attendance Model

Considering the comparison to the actual attendance data shown in Table 7, the attendance model provides a fairly reliable predictive tool. Seattle created the largest discrepancy between predicted and actual as expected due to its status as an outlier compared to typical MLS attendance. That said, each of the models created possessed strengths and weaknesses. As intended, the attendance model provided an insight into the potential success of a new franchise, not just the ability to say where a team should or should not exist. The correlation between population and total sports may distort the model some (Table 8), but this did not have as large of an impact on the attendance model as on the presence model and I felt that despite the covariance, both were important factors to evaluate.

Table 9: Full model independent variable correlation table

Independent Variable Correlation Matrix					
Variables	Population	Entertainment	Resident	Education	Sports
Population	1.00	-0.05	0.15	0.32	0.96
Entertainment	-0.05	1.00	0.04	-0.14	-0.09
Resident	0.15	0.04	1.00	-0.06	0.13
Education	0.32	-0.14	-0.06	1.00	0.42
Sports	0.96	-0.09	0.13	0.42	1.00

Note: Total sports and population were highly correlated but ultimately were both included in the full attendance model.

Also, due to being trained on a data set made up of only large cities (those with MLS franchises already), the attendance model generated relatively a small coefficient for population. This tended to bias the model to be more forgiving towards smaller cities in comparison to the presence model and because of this, had I not filtered out populations smaller than one million, Boulder, Colorado would have been the top location for MLS expansion projecting nearly 37,000 in average attendance despite the MSA only having a population of approximately 350,000.

Further, the majority of projected expansion locations have higher projected attendances than current MLS team. This resulted partially due to each of these franchises being new and not feeling the negative impact of franchise age, but also indicates that the model likely tended to overestimate attendances. The negative impact of age may have resulted from the impact of novelty but it also may have resulted early MLS franchise locations not being wisely selected or managed. In 2014 four of the five least attended franchises were founding members of the league. The fifth of those franchises, Chicago, is also among the older teams in the league, joining two years after the league's founding. This has serious implications for MLS and for struggling franchises as it provided evidence suggesting some franchises should either be relocated or folded in favor of better locations, or atleast newer franchises should MLS continue to struggle financially. This raises the question of did franchise age and decreasing novelty really drive down MLS attendance, as suggested by the model, or were the original franchises just poorly chosen?

ii. Presence Model

The presence model in this study served as a benchmark to which the attendance model could be compared. Counter to the attendance model's bias towards small cities, the presence model reflects bias towards generally larger cities because it was estimated on a data set containing all MSAs where population varied more dramatically. This model appeared to more closely reflect the approach MLS has taken, with eight of the top ten predicted markets already hosting franchises and with Miami and Atlanta about to make it ten out of ten.

iii. Control variables play a big part in attendance

One insight gained from the results showed that the control variables examined explained a fair amount of the variation in attendance in MLS. This is interesting because it means that with all reasonable markets (here defined as metro areas with populations over one million) there is a certain baseline attendance you would expect to see. It also means that a fair amount of the franchise's success comes down to effective management which was most clearly represented by the large impact that winning had on MLS attendance. Referring back to Table 8, Kansas City embodied this fact. For years Sporting Kansas City (previously the Kansas City Wizards) struggled to attract fans but in recent years the team's success (including a championship in 2013) pulls in fans leading to one of the top attendance marks in MLS. The Philadelphia Union showed the opposite impact as their market seemed poised for success but years of a poor on field performance has likely driven down attendance. Understanding the impact of multiple years of on-field success or failure provides an avenue for future studies to explore. Although other factors played an

important role in predicting attendance, these findings also lend some credit to the famous

quote from Al Davis, former Oakland Raiders owner and General Manager; "Just win, baby!"

iv. Models agree MLS is now making solid expansion decisions

Despite having no formally expressed, research based method for evaluating MLS

expansion locations, the league seems to have a good understanding of what makes a

market attractive. MLS franchises have been announced for Atlanta, Minneapolis-St. Paul,

and Miami which represented three of the top five projected cities for MLS attendance, and

three of the top eight cities for predicted presence in MSAs without franchises.

Additionally, these results also provide Atlanta FC, Minnesota United, and the future

Miami franchise a rough estimate of how many fans to expect in their first year and could

be used to determine ideal capacities for stadiums. In Minnesota United's case, they

recently announced plans for a 20,000 seat stadium but the model suggests they could fill a

stadium significantly larger, although projected attendance from this model appears

somewhat inflated.

v. Model could be expanded to include more variables for higher accuracy

With both the control and independent variables, there certainly existed more

variables that could be measured and included in the model and this would have driven up

the R^2 value (perhaps just artificially based on the mechanics of how linear regression

works) to explain more of the variance in attendance. This list of variables was deliberately

a shorter list with an emphasis on variables with support in existing literature to avoid

artificially inflating the R^2 value, but the list of variables one could explore in building

similar models is almost unlimited. Another reason this list of variables was used is that there was little correlation between variables other than population and number of professional sports teams in that metro area. Other factors that were initially included but eventually removed from consideration due to lack of support in literature or high correlation included average age, ethnic makeup, stadium age, and household income to name a few.

vi. Residency in state of birth a poor proxy for community attachment

Based on the results, one area that should be revisited in future work is how to evaluate community attachment because residency in state of birth seemed to do a poor job of measuring that sentiment. Community attachment was a factor that came through the literature as being one of the most strongly influential factors in determining MLS fan motivation, so to see it negatively correlated with MLS attendance in the results seemed unlikely. It is possible that the underlying assumption was incorrect and in fact individuals who choose to live in a state different from where they were born have a greater sense of attachment because they chose that community despite having roots in another location. Whatever the cause may be, the result seemed to be not that community attachment negatively impacts MLS attendance, but rather that community attachment was not accurately measured.

VI. Conclusion

This section serves to reinforce the main takeaways of the research, illustrate the limitations, synthesize the new knowledge while continuing opening up new areas for exploration.

i. Summary of key takeaways

The research and results used in this paper have several strengths worth noting with important implications going forward. First, this study used publicly available data and existing research research but applied it in a new way to help expand the literature around sports expansion modeling and, in a broader sense, the way businesses thing about entering a new market.

Additionally, the study provided a framework that it could be applied to other professional sports leagues, and other businesses, looking for ways to determine ideal expansion locations. It would be particularly useful to smaller, younger professional leagues, or franchises, that may be viewed as inferior in the eyes of the general public when compared to more established, traditional brands like the NFL, MLB, NHL, and NBA in soccer's case. Additionally, the validity and accuracy of this study would be easy to test as all information used was publically available and requires no primary data collection. That said, the quality of the study would be greatly enhanced with wide scale primary data collection. Additionally, the hypotheses proposed are relatively generalizable and require little if any expertise in the field of sports which means more individuals will be able to utilize this research in their own work, even if it is in another field.

Also, the use of multiple linear regression allowed for the analysis of multiple variables at once, while providing information about how those variables work together as

well. Examining simply one variable in a simple linear regression, or even utilizing multiple simple linear regressions would be an overly narrow approach and would be less efficient at minimizing the noise in the data. By putting all of the variables into one model, it informed the model about how those variables worked together to build a more complete picture of how to model attendance for MLS.

Further, it is important to keep in mind that although this research provided a framework for future research, each expansion question will have its own nuances and its own important variables. What is true for MLS may not be true for selecting McDonalds expansion locations, and may not even be true for selecting other sports expansion locations. Due to this, it remains important to always start with the literature to understand what may be most significant in determining the success of expansion locations in a specific industry, then use the available data to determine the degree and significance of the effect.

ii. Study limitations

The methodology that was used in this study also had many limitations. First, proxies had to be used to evaluate some of the variables being studied due to limitations in the data available. Specifically, the proxy used for community attachment has little grounds in the literature, but community attachment was a new idea worth testing and this was the most viable option with the data available. There were also some limitations to the entertainment spending data. The Consumer Expenditure survey did not collect data on all metro areas but rather on a select few from each of their four predefined regions. These metro areas provided varied slightly year to year but for example, the 2014 metro areas were Chicago, Detroit, Minneapolis - St. Paul, Cleveland, New York, Philadelphia, Boston,

Washington D.C., Baltimore, Atlanta, Miami, Dallas – Fort Worth, Houston, Los Angeles, San Francisco, San Diego, Seattle, and Phoenix. For cities not listed specifically, their regional average was used. As is evident by the list provided, most MLS metro areas were provided, but this dataset was a larger limitation on projecting expansion cities and likely generated regional biases based on the need to use regional averages.

Additionally, the study only accounted for MLS franchises based in America. This is a limitation both on the original model's accuracy and on the ability to select the best possible cities for expansion. The original model was limited by not being able to include several years of data for Toronto FC, Vancouver Whitecaps, and Montreal Impact. Additionally, metro areas such as Edmonton, which has strong support for a lower division soccer franchise (Edmonton FC), were left out due to the data inconsistencies Canadian teams would introduce.

Further, the lack of transparency in MLS, and professional sports overall, with regards to financial reporting, provided a limiting factor. Attendance was used as a dependent variable to measure the success of a franchise however attendance does not necessarily mean profitability for a team. Factors such as TV deals, merchandise, ticket prices, and others all contribute to the overall profitability and sustainability of a franchise. For most larger professional sports leagues, Forbes also provides yearly valuations along with revenue and profitability reports, however these were only available for three years for MLS at the time of this study, which provided limitations of its own. Instead, raw attendance data was used and for smaller sports leagues like the MLS, this represents a more reasonable approach to take because the majority of their revenue does come from attendance (Kim & Trail, 2010). That said, MLS now operates under a new TV deal through

their collective bargaining agreement that will dramatically increase the percentage of revenue coming from TV, thus this approach may need to be altered in future studies, but did not impact this study as it took effect beginning in the 2015 season.

Finally, this study attempted to address the limitations of previous work that had studied the presence of a franchise by using some measure of success. By making this decision, it introduced a new limitation into the study. Because attendance became the dependent variable, the sample studied for the attendance model consisted of only locations that had MLS franchises. It is possible these locations had something about them that made them systematically better locations and that factor may not have been captured by the attendance model. This at least partially manifests itself in the need to filter MSAs with low populations because the model was trained on data containing only high populations (thus limiting the variance of that factor). In and effort to address this, a third model was created that utilized the same variables as the presence model (eg no control variables and removing total sports), but used attendance as the dependent variable, and was trained on the complete set of MSAs. Due to the high number of MSAs having no attendance at all, the predicted attendance values as a whole were quite low however this provided an additional ranking system for evaluating the quality of expansion locations. Interestingly enough this model also generated Miami, Atlanta, San Francisco, and Phoenix as the most appealing expansion locations, with Minneapolis rating highly as well (Table 9). This only further strengthened the claim other models made that also project these as top locations.

Table 10: Attendance Model Trained on All MSAs

Geography	Predicted Attendance
New York-Newark-Jersey City, NY-NJ-PA Metro Area	29091
Los Angeles-Long Beach-Anaheim, CA Metro Area	19089
Chicago-Naperville-Elgin, IL-IN-WI Metro Area	13874
Dallas-Fort Worth-Arlington, TX Metro Area	9735
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area	9459
Houston-The Woodlands-Sugar Land, TX Metro Area	9083
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area	8842
Miami-Fort Lauderdale-West Palm Beach, FL Metro Area	**8369**
Atlanta-Sandy Springs-Roswell, GA Metro Area	**8364**
Boston-Cambridge-Newton, MA-NH Metro Area	7506
San Francisco-Oakland-Hayward, CA Metro Area	**7455**
Phoenix-Mesa-Scottsdale, AZ Metro Area	**6596**
Detroit-Warren-Dearborn, MI Metro Area	**6184**
Minneapolis-St. Paul-Bloomington, MN-WI Metro Area	**6100**
Seattle-Tacoma-Bellevue, WA Metro Area	6073
Riverside-San Bernardino-Ontario, CA Metro Area	**6063**
San Diego-Carlsbad, CA Metro Area	**5002**
Baltimore-Columbia-Towson, MD Metro Area	**4720**
Denver-Aurora-Lakewood, CO Metro Area	4602
Tampa-St. Petersburg-Clearwater, FL Metro Area	**4098**
San Jose-Sunnyvale-Santa Clara, CA Metro Area	3856
St. Louis, MO-IL Metro Area	**3851**
Portland-Vancouver-Hillsboro, OR-WA Metro Area	3750
Pittsburgh, PA Metro Area	**3661**
Kansas City, MO-KS Metro Area	3294
Austin-Round Rock, TX Metro Area	**3274**
Orlando-Kissimmee-Sanford, FL Metro Area	3245
Sacramento--Roseville--Arden-Arcade, CA Metro Area	**3174**
Cincinnati, OH-KY-IN Metro Area	**3122**
Las Vegas-Henderson-Paradise, NV Metro Area	**3079**
Charlotte-Concord-Gastonia, NC-SC Metro Area	**2934**
Columbus, OH Metro Area	2871
Indianapolis-Carmel-Anderson, IN Metro Area	**2791**
San Antonio-New Braunfels, TX Metro Area	**2730**
Nashville-Davidson--Murfreesboro--Franklin, TN Metro Area	2623

Note: MSAs that do not currently have MLS franchises are bolded as potential expansion locations.

iii. Closing thoughts

Although the study did have a number of limitations, this study also contributed to

the existing literature in a meaningful way. Some research had been done into evaluating

expansion locations for professional sports leagues, but in comparison to other sports management research questions, the research base was relatively thin. By taking another look at this question with new data, new research variables, and a slightly altered approach, this research continued to push forward the understanding of what makes some cities support professional sports, and others ignore them. This research also validated recent decisions made by Major League Soccer while leading to questions about older expansion locations. It also opened the floor examine what other variables not included in this model might impact MLS attendance. Finally, this research can serve soccer fans located in markets without MLS teams as something they can point to as they try to persuade potential team owners, and the league itself, to bring the beautiful game to their city.

Appendix A: Presence Model Results

Presence Model Excluding Total Sports				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.01	0.11087635	-0.0815468	0.9350523
Population (in Millions)	0.07	5.24E-09	14.2108562	1.03E-36
Entertainment Spend (in Thousands)	0.004399	3.64E-05	0.12097516	0.90377801
Resident in State of Birth (as a %)	-0.000867	0.00058221	-1.4896958	0.13717736
Education (as a %)	0.004738	0.00225675	2.09957585	0.03645992

Appendix B: Control Model Results

Control Model Results				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12285.1781	2737.61893	4.48754133	1.66E-05
Points Per Game	5161.31013	1492.43734	3.45830942	0.00075296
Franchise Age	-578.57757	87.0831112	-6.6439699	9.44E-10
year14	8117.7748	2021.17975	4.0163547	0.00010344
year13	6772.47222	2000.86581	3.38477083	0.00096314
year12	5828.4982	1984.61146	2.936846	0.00397596
year11	4472.41037	1973.62843	2.2660853	0.02523814
year10	3095.75881	1999.12511	1.54855682	0.12412162
year09	2491.42101	2038.08136	1.22243452	0.22393895
year08	2835.32497	2070.72886	1.36924009	0.17348148
year07	2523.05977	2107.01061	1.19745945	0.23348756
year06	508.881575	2107.59383	0.24145144	0.80961734

Appendix C: Full Model Results

Full Model Results				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8467.41721	6347.72114	1.33393024	0.18486246
Population (in Millions)	849.767306	263.764059	3.22169484	0.00165771
Total Sports	-1347.2231	575.294938	-2.3417956	0.020912
Entertainment Spending (in Thousands)	3848.95421	1.54287719	2.49466013	0.01402813
Resident in State of Birth (as a %)	-144.53152	41.4587349	-3.4861537	0.00069472
Education (as a %)	277.19112	136.117281	2.03641388	0.04400582
Points Per Game	5186.62129	1349.50407	3.84335358	0.00019935
Franchise Age	-607.17037	89.3712301	-6.7938012	5.09E-10
year14	6311.80969	1917.3759	3.29189998	0.0013221
year13	5673.34644	1834.28937	3.09293971	0.00248777
year12	4748.15461	1818.34152	2.61125567	0.01022334
year11	3544.04721	1798.48372	1.97057509	0.05117543
year10	2718.77904	1784.95762	1.52316167	0.13046346
year09	1285.71087	1866.2406	0.68893093	0.49225434
year08	1106.49422	1945.5725	0.56872423	0.57065229
year07	1721.35158	1904.79519	0.90369379	0.36804693
year06	1044.25686	1880.65347	0.55526278	0.57979344

Appendix D: Projected Attendance For Top Candidate MSAs

Geography	2014 Projected Attendance
San Francisco-Oakland-Hayward, CA Metro Area	34989
Atlanta-Sandy Springs-Roswell, GA Metro Area	34567
Miami-Fort Lauderdale-West Palm Beach, FL Metro Area	34185
Phoenix-Mesa-Scottsdale, AZ Metro Area	34057
Minneapolis-St. Paul-Bloomington, MN-WI Metro Area	34033
Las Vegas-Henderson-Paradise, NV Metro Area	33244
Baltimore-Columbia-Towson, MD Metro Area	32380
San Diego-Carlsbad, CA Metro Area	32153
New York-Newark-Jersey City, NY-NJ-PA Metro Area	31363
Tampa-St. Petersburg-Clearwater, FL Metro Area	31289
Austin-Round Rock, TX Metro Area	31106
Los Angeles-Long Beach-Anaheim, CA Metro Area	30849
San Jose-Sunnyvale-Santa Clara, CA Metro Area	30748
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area	30416
Seattle-Tacoma-Bellevue, WA Metro Area	30293
Riverside-San Bernardino-Ontario, CA Metro Area	30142
Tucson, AZ Metro Area	30013
Raleigh, NC Metro Area	29742
Detroit-Warren-Dearborn, MI Metro Area	29516
Portland-Vancouver-Hillsboro, OR-WA Metro Area	29489
Hartford-West Hartford-East Hartford, CT Metro Area	29307
Sacramento--Roseville--Arden-Arcade, CA Metro Area	28832
Nashville-Davidson--Murfreesboro--Franklin, TN Metro Area	28457
Jacksonville, FL Metro Area	28199
Virginia Beach-Norfolk-Newport News, VA-NC Metro Area	28150
Orlando-Kissimmee-Sanford, FL Metro Area	28123
Providence-Warwick, RI-MA Metro Area	28069
Cincinnati, OH-KY-IN Metro Area	27789
St. Louis, MO-IL Metro Area	27619
Indianapolis-Carmel-Anderson, IN Metro Area	27383
Chicago-Naperville-Elgin, IL-IN-WI Metro Area	27380

Note: MSAs that do not currently have MLS franchises are bolded as potential expansion locations.

Appendix E: Projected Presence Values for top MSAs in 2014

Geography	Prediction Value
New York-Newark-Jersey City, NY-NJ-PA Metro Area	1.53
Los Angeles-Long Beach-Anaheim, CA Metro Area	1.00
Chicago-Naperville-Elgin, IL-IN-WI Metro Area	0.73
Dallas-Fort Worth-Arlington, TX Metro Area	0.52
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Are	0.51
Houston-The Woodlands-Sugar Land, TX Metro Area	0.48
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Are.	0.46
Miami-Fort Lauderdale-West Palm Beach, FL Metro Area	**0.46**
Atlanta-Sandy Springs-Roswell, GA Metro Area	**0.43**
Boston-Cambridge-Newton, MA-NH Metro Area	0.40
San Francisco-Oakland-Hayward, CA Metro Area	**0.38**
Phoenix-Mesa-Scottsdale, AZ Metro Area	**0.34**
Seattle-Tacoma-Bellevue, WA Metro Area	0.31
Riverside-San Bernardino-Ontario, CA Metro Area	**0.31**
Detroit-Warren-Dearborn, MI Metro Area	**0.31**
Minneapolis-St. Paul-Bloomington, MN-WI Metro Area	**0.29**
San Diego-Carlsbad, CA Metro Area	**0.26**
Denver-Aurora-Lakewood, CO Metro Area	0.24
Tampa-St. Petersburg-Clearwater, FL Metro Area	**0.23**
Baltimore-Columbia-Towson, MD Metro Area	**0.22**
St. Louis, MO-IL Metro Area	**0.20**
San Jose-Sunnyvale-Santa Clara, CA Metro Area	0.19
Portland-Vancouver-Hillsboro, OR-WA Metro Area	0.19
Pittsburgh, PA Metro Area	**0.19**
Orlando-Kissimmee-Sanford, FL Metro Area	0.19
Charlotte-Concord-Gastonia, NC-SC Metro Area	**0.18**
Kansas City, MO-KS Metro Area	0.17
Las Vegas-Henderson-Paradise, NV Metro Area	**0.16**
Cincinnati, OH-KY-IN Metro Area	**0.16**
Sacramento--Roseville--Arden-Arcade, CA Metro Area	**0.16**
Austin-Round Rock, TX Metro Area	**0.16**
San Antonio-New Braunfels, TX Metro Area	**0.15**
Columbus, OH Metro Area	0.15

Note: MSAs that do not currently have MLS franchises are bolded as potential expansion locations. Salt Lake City, home of Real Salt Lake (a current MLS franchise) is not shown in this table because it ranked so low. Its predicted presence value was approximately .07.

Appendix F: Projected Presence vs Projected Attendance for all MSAs

Projected Presence vs Projected Attendance

Works Cited

Anderson, D. F., & Stone, G. P. (1981). Sport: A search for community. Sociology of sport: Diverse perspectives, 164-172.

Baade, R. a. (1994). Stadiums, Professional Sports , Economic Development: Assessing the Reality. *Policy Study*, 62(4), 1–39.

Baehr, M. J. (2006). An Analysis of Expansion and Relocation Sites for Major League Soccer. *International Journal of Sport Managment.*, *7*(January), 1–12.

Bruggink, T. H., & Zamparelli, J. M. (1999). *Sports Economics: Current Research*. (J. Fizel, E. Gustafson, & L. Hadley, Eds.). Westport, CT: Praeger Publishers.

Gitter, S. R., & Rhoads, T. A. (2010). Determinants of Minor League Baseball Attendance. *Journal of Sports Economics*, *11*(6), 614–628. http://doi.org/10.1177/1527002509359758

Kim, Y. K., & Trail, G. (2010). Constraints and motivators: A new model to explain sport consumer behavior. *Journal of Sport Management*, *24*, 190–210. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-77955114752&partnerID=40&md5=b6c376860b51883bbf5ced7ee28f569f

Baehr, M. J. (2006). An Analysis of Expansion and Relocation Sites for Major League Soccer. *International Journal of Sport Managment.*, *7*(January), 1–12.

Bruggink, T. H., & Zamparelli, J. M. (1999). *Sports Economics: Current Research*. (J. Fizel, E. Gustafson, & L. Hadley, Eds.). Westport, CT: Praeger Publishers.

Davis, M. C. (n.d.). *Income and the Locations of AAA Minor League Baseball Teams*. University of Missouri-Rolla.

Gitter, S. R., & Rhoads, T. A. (2010). Determinants of Minor League Baseball Attendance. *Journal of Sports Economics*, *11*(6), 614–628. http://doi.org/10.1177/1527002509359758

Funk, D. C., Beaton, A., & Alexandris, K. (2012). Sport consumer motivation: Autonomy and control orientations that regulate fan behaviours. *Sport Management Review*, *15*, 355–367. http://doi.org/10.1016/j.smr.2011.11.001

James, J. D., Kolbe, R. H., & Trail, G. T. (2002). Psychological Connection to a New Sport Team : Building or Maintaining the Consumer Base ? *Sport Marketing Quarterly*, *11*(4), 215–226.

Kim, Y. K., & Trail, G. (2010). Constraints and motivators: A new model to explain sport consumer behavior. *Journal of Sport Management*, *24*, 190–210. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-77955114752&partnerID=40&md5=b6c376860b51883bbf5ced7ee28f569f

Mahony, D. F., Nakazawa, M., Funk, D. C., James, J. D., & Gladden, J. M. (2002). Motivational factors influencing the behaviour of J. League spectators. Sport Management Review, 5(1), 1-24.

Parrish, C. (2013). Soccer specific stadiums and designated players: Exploring the Major League Soccer attendance assumption. International Journal of Sport Management Recreation and Tourism, 12, 57-70.

Uhlman, B. T., & Trail, G. T. (2012). Consumer Behavior An Analysis of the Motivators of Seattle Sounders FC Season Ticket Holders : A Case Study. *Sport Marketing Quarterly*, *21*(4), 243–252.