Final Project: A Second NFL Team in Chicago?

Eric Crnkovich & Jon Sax

Northwestern University

MSDS 457: Sports Management Analytics

May 30th, 2023

Executive Summary

With the rising popularity of the National Football League (NFL), the question has been proposed if the NFL should expand to include additional teams. Currently, there are 32 teams in the NFL, with each team being based in a different metropolitan area with the exception of Los Angeles and New York, which have two teams each. This report aims to verify the location of Chicago as the residence for the 33rd team and construct a projected roster optimized by performance and subject to various constraints.

The roster was constructed following closely to the NFL expansion draft rules set in 2002 when the Houston Texans were added to the League. The 53-Man roster has been filled using the following criteria: Multiple players at each position to account for potential injuries, selecting players who have a certain threshold of NFL experience, and maintaining a roster salary below the NFL salary cap of $224.8 million.

It was also necessary to verify the residence of this new team as Chicago. To accomplish this, a model was created that took into consideration many factors about a city and its surrounding metropolitan area such as population, public transit rank, and income per capita. As a result, Chicago was the city selected to be the residence of the NFL's 33rd team.

The findings from this study can be used to better analyze and make decisions regarding the addition of a professional sports team to a city. A constrained optimized approach to creating a roster is an ideal approach as long as the constraints are accurate and abide by current league rules. Different factors and metrics regarding any individual city can be analyzed and interpreted as input for a model to determine the best candidate city for a professional sports team.

Introduction

The National Football League (NFL) is a professional sports league based in the United States with an estimated value of $150 billion (BetMGM 2023). Currently, the League is composed of 32 teams with each team residing in its own metropolitan area with the exception of Los Angeles and New York which are home to two teams each. Chicago, the third largest city in the country behind the two aforementioned cities, is only home to one NFL team, the Chicago Bears. It is reasonable to assume that if the NFL were to add a 33rd team, Chicago would be an excellent candidate city to be its home.

This report first aims to construct a projected 53-man roster subject to various constraints while also optimizing for player performance. Next, an attempt to verify Chicago as the best choice to be the host city for this new team is completed using a self-created scoring model. Lastly, an expected ticketing and promotions strategy is detailed using information about the Chicago Bears gathered from *Ticketmaster.* This strategy then flows into predicting the expected attendance and performance of this new team.

The strategies and methodologies highlighted in this report can be generalized to any professional sports league seeking to create an expansion team. Of course, some aspects will need to be tailored to fit the league's requirements, but the overall concepts will remain consistent. More teams mean more money, and it appears inevitable that the dominant professional sports leagues in the United States (and around the world) will continue to expand and generate more revenue than ever before.

Literature Review

The most recent expansion team incorporated into the NFL was the Houston Texans in 2002 (Pierson 1999). As a result of this new team, the NFL had a new total of 32 teams and the divisions needed to be realigned to satisfy conference and divisional constraints. In order for an expansion team to be approved, other NFL team owners must vote on the creation of the new team, with a majority vote required for the new team to be approved. This new team in Houston also needed to place a $700 million bid prior to the vote by the other NFL team owners.

The question of what city should be chosen to host the newest expansion team for a professional sports organization is something that has been meticulously researched in the past. It has been expressed that Toronto may be an ideal choice to be the next city to host an NFL expansion team, as the other major professional sports leagues in the United States, the NBA, MLB, MLS, and NHL have all gone international with having at least one team not based in the United States (Perkins 2005). One reason behind Toronto being a prime candidate for an NFL expansion team was due to its proven success for professional sports franchises, i.e. the Toronto Raptors in the NBA, the Toronto Blue Jays in the MLB, Toronto FC in the MLS, and the Toronto Maple Leaf's in the NHL.

Projected Roster Methodology

Every NFL roster is composed of 53 individuals. How a team creates their roster and fills the various positions is up to the GM of each team. For example, one GM may elect to have 3 QB's on their roster, while another GM will opt for 2 QB's in order to roster another RB. For the scope of this project the following roster constraints were used: 2 Quarterbacks (QB), 4 Running Backs (RB), 6 Wide Receivers (WR), 3 Tight Ends (TE), 3 Centers (C), 3 Tackles (T/LT/RT), 3 Guards (G), 9 Defensive Linemen (DE/DT), 7 Linebackers (LB/ILB/OLB), 10 Defensive Backs (CB/S/FS/SS), 1 Kicker (K), 1 Punter (P), and 1 Long Snapper (LS).

First, data for 1,000 NFL players was collected from *Spotrac* that contained the players name, current team, position, and average salary. This data was then filtered by team and the five highest players per team were characterized as "protected" meaning they cannot be drafted by the expansion team. This is one of the key rules of the NFL expansion draft, with another being the expansion team cannot draft more than two players from a single team. Since insight is not able to be gathered regarding the opinions and potential actions of each team's GM regarding which players to protect, it was assumed that players with the top five highest average salaries would be protected (or top four plus QB). The protected players were then dropped from the dataset, leaving only players that the expansion team is allowed to draft.

To optimize a roster by performance, different metrics were obtained by position from *Pro Football Reference* for each individual player, see Table 1 below. Positions in football have very different jobs, this makes performance comparisons across positions difficult. For example, how would one compare the performance of a Wide Receiver and that of a Center? A generalized performance metric. This answer to this question and the formulation of this performance metric are explained later in this section.

Position	**Performance Metric Used**
Quarterback	QBR (Quarterback Rating)
Running Back	Yards per Touch
Wide Receiver	Yards per Touch
Tight End	Yards per Touch
Center	# of Snaps
Guard	# of Snaps
Tackle	# of Snaps
Defensive Line	Tackles per Snap
Linebacker	Tackles per Snap
Defensive Back	Tackles per Snap
Kicker	Made FG %

Punter	% Punts Inside 20-Yd Line
Long Snapper	Games Started

Table 1: Performance Metrics Used by Position

Next, players were first subset by position and then by team so that the aforementioned constraints involving the number of players per position and number of players drafted from any individual team can be fulfilled. After filtering, additional columns in the dataset were added to be used as a constraint matrix when solving the optimization model, see Figure 1 below.

player	team	position	average_sal	snaps	tackles	tackles_per_	QB	RB	WR	TE	C	TA	G	DL	LB	DB	K	P	LS
Matt Milano	BUF	OLB	14165000	4283	458	0.10693439	0	0	0	0	0	0	0	0	1	0	0	0	0
Randy Grego	DEN	OLB	14000000	1664	97	0.05829327	0	0	0	0	0	0	0	0	1	0	0	0	0
Preston Smit	GB	OLB	13000000	6063	363	0.05987135	0	0	0	0	0	0	0	0	1	0	0	0	0
Jerome Bake	MIA	ILB	12500000	4611	509	0.1103882	0	0	0	0	0	0	0	0	1	0	0	0	0
De'Vondre C	GB	ILB	10000000	5884	704	0.1196465	0	0	0	0	0	0	0	0	1	0	0	0	0
Bobby Okere	NYG	LB	10000000	3198	420	0.13133208	0	0	0	0	0	0	0	0	1	0	0	0	0
Uchenna Nw	SEA	OLB	9527500	2669	198	0.07418509	0	0	0	0	0	0	0	0	1	0	0	0	0
Travon Walk	JAC	OLB	9343155	788	49	0.06218274	0	0	0	0	0	0	0	0	1	0	0	0	0
Demario Dav	NO	ILB	9000000	10175	1136	0.11164619	0	0	0	0	0	0	0	0	1	0	0	0	0
Dre Greenlav	SF	LB	8200000	2386	326	0.13663034	0	0	0	0	0	0	0	0	1	0	0	0	0
Kayvon Thibc	NYG	OLB	7834760	740	49	0.06621622	0	0	0	0	0	0	0	0	1	0	0	0	0
Devin White	TB	ILB	7328954	3972	483	0.12160121	0	0	0	0	0	0	0	0	1	0	0	0	0
Kaden Elliss	ATL	OLB	7166667	826	101	0.12227603	0	0	0	0	0	0	0	0	1	0	0	0	0
Arden Key	TEN	OLB	7000000	2106	98	0.04653371	0	0	0	0	0	0	0	0	1	0	0	0	0
Germaine Pr	CIN	LB	6750000	2538	355	0.13987392	0	0	0	0	0	0	0	0	1	0	0	0	0
Eric Kendrick	LAC	ILB	6625000	7301	919	0.12587317	0	0	0	0	0	0	0	0	1	0	0	0	0
T.J. Edwards	CHI	LB	6500000	2328	389	0.16709622	0	0	0	0	0	0	0	0	1	0	0	0	0
Shaq Thomp	CAR	OLB	6300000	6017	709	0.11783281	0	0	0	0	0	0	0	0	1	0	0	0	0
Alex Anzalon	DET	ILB	6100000	3123	326	0.10438681	0	0	0	0	0	0	0	0	1	0	0	0	0
Alex Singletc	DEN	LB	6000000	2240	425	0.18973214	0	0	0	0	0	0	0	0	1	0	0	0	0
Quincy Willi	NYJ	LB	6000000	2254	275	0.12200532	0	0	0	0	0	0	0	0	1	0	0	0	0
Cole Holcom	PIT	LB	6000000	2740	388	0.14160584	0	0	0	0	0	0	0	0	1	0	0	0	0
Josh Allen	JAC	OLB	5685660	2693	185	0.06869662	0	0	0	0	0	0	0	0	1	0	0	0	0
Tyus Bowser	BAL	OLB	5500000	2443	152	0.06221858	0	0	0	0	0	0	0	0	1	0	0	0	0
Josey Jewell	DEN	ILB	5500000	2592	345	0.13310185	0	0	0	0	0	0	0	0	1	0	0	0	0
David Long	MIA	LB	5500000	1819	230	0.1264431	0	0	0	0	0	0	0	0	1	0	0	0	0
Bobby Wagn	SEA	ILB	5500000	10698	1523	0.14236306	0	0	0	0	0	0	0	0	1	0	0	0	0
Isaiah Simm	ARI	ILB	5166014	2277	258	0.11330698	0	0	0	0	0	0	0	0	1	0	0	0	0
Kyzir White	ARI	OLB	5000000	2878	388	0.13481584	0	0	0	0	0	0	0	0	1	0	0	0	0
Christian Kirl	HOU	ILB	5000000	6491	778	0.11985827	0	0	0	0	0	0	0	0	1	0	0	0	0
Jordan Hicks	MIN	ILB	5000000	6540	767	0.11727829	0	0	0	0	0	0	0	0	1	0	0	0	0
Anthony Nels	TB	OLB	5000000	1466	94	0.06412005	0	0	0	0	0	0	0	0	1	0	0	0	0
Azeez Al-Sha	TEN	LB	5000000	1523	199	0.13066316	0	0	0	0	0	0	0	0	1	0	0	0	0
Lorenzo Cart	ATL	OLB	4500000	2924	211	0.07216142	0	0	0	0	0	0	0	0	1	0	0	0	0
Frankie Luvu	CAR	OLB	4500000	1958	213	0.10878447	0	0	0	0	0	0	0	0	1	0	0	0	0

ATL	BAL	BUF	CAR	CHI	CIN	CLE	DAL	DEN	DET	GB	HOU	IND	JAC	KC	LAC	LAR	LV	MIA	MIN	NE
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1: Example Dataset for LB's

Datasets for each position were created in Microsoft Excel and then imported into R for further analysis and completion of the projected roster. The performance metrics for each position were then analyzed using boxplots (Figure 2) to determine the distribution of this variable and to recognize any potential outliers.

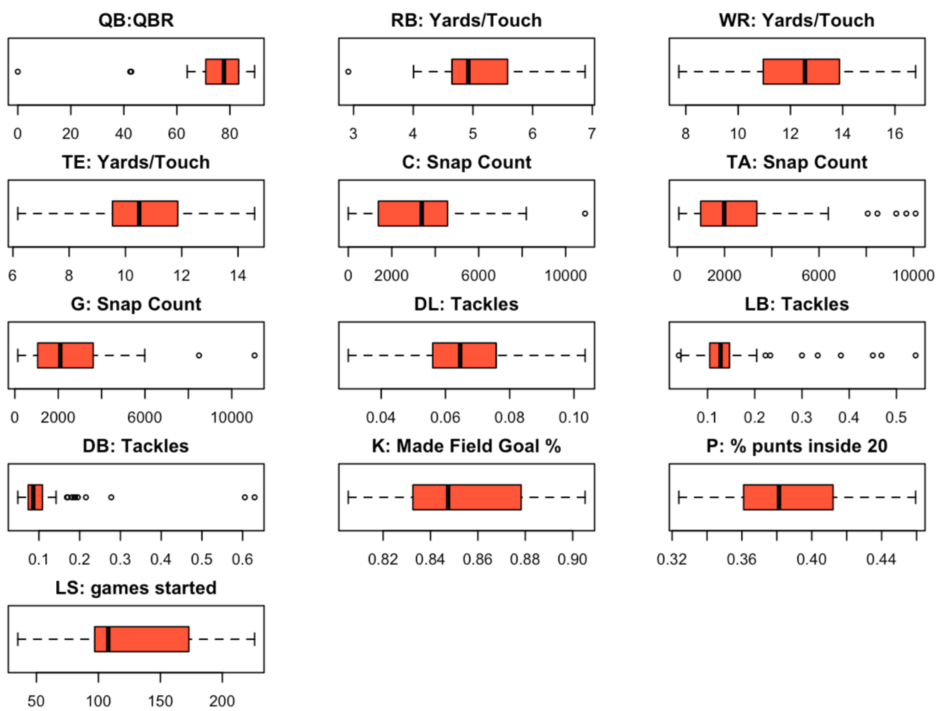


Figure 2: Performance Metric Distribution by Position

Next, we dropped all players at C, G, T, DL, LB, and DB below the 10th percentile in snap count to account for their lack of in game experience at these positions and the fact that they are outliers. After these players were dropped, the generalized performance metric for each position was created. This was accomplished by centering the performance metric for each position from Table 1. This allows for a common scale across all positions and can be utilized to measure the relative performance between positions.

Finishing up with data cleaning, the last step was to only select players in the 75th percentile or better at their respective position. Again, this was done so that only players with a significant amount of NFL in game experience can be considered for the potential roster.

Results and Recommendations

In order to solve this problem and create a projected roster, R's lpSolve package was utilized. The constraint matrix was defined and created taking into account positional constraints, team constraints, and also the salary cap constraint where the combined salary cannot exceed the NFL's salary cap of $224.8 million. The solution and final projected roster can be found below in Figure 3.

```
## # A tibble: 53 × 5
##    player                    team  position average_salary performance[,1]
##    <chr>                     <chr> <chr>             <dbl>           <dbl>
##  1 Jameis Winston            NO    QB              4000000           0.756
##  2 Nick Mullens              MIN   QB              2000000           0.782
##  3 Breece Hall               NYJ   RB              2253694           2.47
##  4 Travis Homer              CHI   RB              2000000           2.35
##  5 James Cook                BUF   RB              1458014           1.58
##  6 Rashaad Penny             PHI   RB              1350000           1.07
##  7 Marquez Valdes-Scantling KC    WR             10000000           2.18
##  8 Christian Watson          GB    WR              2310258           1.07
##  9 Marquise Goodwin          CLE   WR              1700000           1.44
## 10 George Pickens            PIT   WR              1688045           1.38
## 11 Van Jefferson             LAR   WR              1402784           1.22
## 12 Dyami Brown               WAS   WR              1236000           2.28
## 13 Kyle Pitts                ATL   TE              8227624           2.20
## 14 Jelani Woods              IND   TE              1343118           1.10
## 15 O.J. Howard               LV    TE              1232500           2.31
## 16 Jason Kelce               PHI   C              14250000           2.82
## 17 Ryan Jensen               TB    C              13000000           0.965
## 18 Corey Linsley             LAC   C              12500000           1.74
## 19 Duane Brown               NYJ   LT             10000000           2.83
## 20 Donovan Smith             KC    LT              9000000           2.18
## 21 Kelvin Beachum            ARI   LT              2575000           2.67
## 22 Kevin Zeitler             BAL   G               7500000           3.92
## 23 Andrew Norwell            WAS   G               5000000           2.69
## 24 Graham Glasgow            DET   G               2750000           1.38
## 25 DaVon Hamilton            JAC   DT             11500000           1.79
## 26 Sebastian Joseph          LAC   DE              8000000           2.44
## 27 Calais Campbell           ATL   DE              7000000           2.40
## 28 A'Shawn Robinson          NYG   DE              5000000           2.14
## 29 Khalen Saunders           NO    DT              4100000           1.47
## 30 Christian Wilkins         MIA   DE              3859775           2.10
## 31 Mike Purcell              DEN   DT              3833333           1.50
## 32 Lawrence Guy              NE    DT              2875000           2.40
## 33 Boye Mafe                 SEA   DE              2140563           2.23
## 34 T.J. Edwards              CHI   LB              6500000           1.22
```

```
## 35 Alex Singleton          DEN   LB            6000000          1.80
## 36 E.J. Speed              IND   LB            4000000          2.64
## 37 Dennis Gardeck          ARI   OLB           3333333          1.67
## 38 Oren Burks              SF    ILB           2500000          2.16
## 39 Chad Muma               JAC   LB            1356527          1.15
## 40 Matthew Adams           CLE   OLB           1232500          1.03
## 41 George Odum             SF    S             3166667          2.85
## 42 Siran Neal              BUF   S             3000000          3.82
## 43 Sam Franklin            CAR   S             2627000          1.70
## 44 Tavierre Thomas         HOU   CB            2250000          1.70
## 45 Jalen Pitre             HOU   S             2238610          1.63
## 46 Cody Davis              NE    FS            2200000          2.93
## 47 Miles Killebrew         PIT   S             2000000          3.58
## 48 Johnathan Ford          GB    S             1500000          1.57
## 49 Justin Bethel           MIA   CB            1317500          3.30
## 50 Jaquan Johnson          LV    S             1232500          1.67
## 51 Justin Tucker           BAL   K             6000000          1.89
## 52 Michael Dickson         SEA   P             3675000          1.11
## 53 J.J. Jansen             CAR   LS            1317500          1.77
```

Figure 3: Projected Roster Using Constrained Optimized Approach

The above roster follows all of the constraints previously mentioned. Each position has the required number of players, no more than two players have been drafted from the same team, and the total combined average salary is less than the NFL's salary cap of $224.8 million. In fact, the average salary of this projected roster is only slightly under the salary cap, coming in at $224.5 million. The generalized performance metric can be seen on the far right in Figure 3 and is the variable that was being maximized. The optimal total value of this variable subject to the data and the constraints is 105.0473. This number is difficult to interpret, but we believe it will make more sense if it can be compared to actual rosters for the other teams in the League. Then, we will be able to determine if the projected roster can immediately compete with the top teams in the League, or if it will take several years' worth of trades and draft picks to accomplish this feat.

City Selection

Today, there are only two metropolitan statistical areas (MSAs) boasting two NFL teams: New York-Newark-Jersey City and Los Angeles-Long Beach-Anaheim. The Chicago-Naperville-Elgin MSA is the third largest in the United States (US Census Bureau estimate 2022). The fourth largest, Dallas-Fort Worth-Arlington, has 1.5 million less people than the Chicagoland metropolitan area (2022). This is significant, as it represents the largest population difference between consecutive ranks in the top 10, excluding NY and LA (2022).

But metropolitan statistical area is not the only item NFL executives are considering when deciding where to build the league's 33rd team. This is an immense decision that involves a multitude of civic, political, geographical, legal, and economic factors (Bruggink & Schiz 2008). That is why we have selected 10 hypothetical candidate cities as potential hosts of the NFL's 33rd team and built a holistic scoring system to evaluate them based on 10 key components of what makes an area an appropriate location for an NFL team. Below is a list and brief explanation of the factors contributing to our scoring model:

Population (metropolitan statistical area): simply put, the larger the metropolitan area of a candidate city, the larger the potential fan base. In order to obtain accurate numbers for each of the 10 candidate cities, we used the most up-to-date comprehensive population data available from the United States Census Bureau, which is from 2021. It is important to note that we selected 2021 (instead of simply using the 2020 census) due to the significant urban population changes that occurred within the last three years resulting from the COVID-19 pandemic. 2021 provides a more accurate, up-to-date figure than the '20 census (US Census Bureau 2021).

Population (city proper): correlated with metro population, and also important to the analysis, as a higher city population not only means a higher potential fan base, but larger infrastructural integration opportunities with the expansion team (US Census Bureau 2021).

Population (city proper) % change (from 2010 census to 2021): observing population changes in the last 11 years will help forecast growth/decline of future pools of potential fans in the decades ahead (US Census Bureau 2010, 2021).

Income per capita (city proper, in dollars): cities with a stronger economy and higher income levels will score higher than those with weaker figures (US Census Bureau 2021).

Number of Fortune 500 companies headquartered in the metropolitan area: for the same economic reasons as income per capita, the higher the number of corporate headquarters, the larger the tax base, fan base, etc. (Fortune 2022).

Number of other professional sports teams already representing the city: we are striving to avoid overrepresentation by including this metric (as negatively impacting candidate cities) in our analysis (manually calculated).

Distance from the next closest city with an NFL team (in miles): like the number of other professional teams, we include this metric to provide and highlight an additional potential contributing factor to overrepresentation (manually calculated).

Unemployment rate ranking - February 2023 (metropolitan), not seasonally adjusted: cities with the highest rankings in this category had the lowest rates of unemployment (US Bureau of Labor Statistics - February 2023).

Remote work ranking (city proper): cities with the highest rankings in this category meant they were the most favorable areas to have remote jobs. We include this metric in our analysis for two reasons: (1) to recognize the rising emphasis on remote work permittance in white collar jobs and (2) items positively affecting the remote work score include safety, affordability, connectivity, and amenities – all components that will positively affect candidate city viability (Lawnstarter 2022).

Public transit rank (city proper): the goal attendance of each home game is over 60,000+ people, regardless of whether the stadium is within the city limits or not. Candidate cities with high public transit scores reduce congestion and pollution. Therefore, this is an important 21st-century consideration (WalletHub 2019).

Scoring Methodology

The data in Table 2 (see below) is compiled in many different units: number of people, companies, dollars, miles, ranking, etc. We want all of these factors to contribute equally to each candidate city's final score, regardless of their original scale or range. In order to create an overall analysis that makes sense, we loaded these results into a Pandas data frame in Python and used the Scikit-learn library's MinMax scaler class to preprocess the data by column. The resulting min-max normalized values all exist between 0 and 1, with the candidate city boasting the best score in a particular category as 1.0, and the candidate city with the worst score in said category as 0.0. Each other city's value proportionally reflects their score relative to the other nine cities in that category.

Candidate City	Metro Pop '21	City Pop '21	% Growth (11y)	Distance	Income per Capita	Fortune 500	Other Teams	Unemployment Rate Rank	Remote Work Rank	Public Transit Rank
Chicago	1.000000	1.000000	0.305991	0.000000	0.529786	1.000000	0.000000	0.027237	0.698413	0.868132
San Antonio	0.170684	0.501337	0.655640	0.382330	0.000000	0.057143	1.000000	0.225681	1.000000	0.659341
Oklahoma City	0.031410	0.195205	1.000000	0.580583	0.183357	0.057143	1.000000	0.894942	0.825397	0.076923
St. Louis	0.124371	0.037191	0.000000	0.471845	0.669254	0.228571	0.666667	0.844358	0.857143	0.109890
Sacramento	0.117048	0.130029	0.774568	0.185243	0.362432	0.000000	1.000000	0.000000	0.285714	0.340659
Salt Lake City	0.000000	0.000000	0.586545	1.000000	0.540188	0.000000	0.666667	1.000000	0.888889	0.000000
San Diego	0.251142	0.473196	0.517158	0.229126	0.715383	0.057143	1.000000	0.311284	0.000000	0.901099
Oakland	0.413360	0.093485	0.717614	0.009709	0.855811	0.000000	1.000000	0.470817	0.373016	0.967033
San Jose	0.074310	0.313697	0.453260	0.093204	1.000000	0.228571	0.666667	0.521401	0.166667	0.945055
Portland	0.159665	0.176551	0.672674	0.337864	0.748550	0.057143	0.666667	0.373541	0.968254	1.000000

Table 2: The 10 candidate expansion cities with their corresponding 10 topics for consideration

From here, we added a critical final transformation. There were four categories that had an adverse effect on a candidate city's campaign: Other Teams, Unemployment Rate Rank, Remote Work Rank, and Public Transit Rank. In other words, the higher the value, the lower the expansion market attractiveness of the city in that category. For example, before min-max normalization, the "Unemployment Rank" column listed each city's ranking among United States cities based on their unemployment rate. The city in our 10-candidate field with the lowest unemployment rate, Salt Lake City, ranked 51st nationwide, so the pre-normalized value was 51. The city with the worst unemployment rank in our field was Sacramento, at 308th nationwide, so Sacramento's pre-normalized value in this category was 308. In order to reverse the order of these values, we inverted their normalized scores. By applying the complement transformation, we converted the data to a scale where higher values represent stronger or better performance (instead of worse). The normalized data frame below now has all accurate inputs and is ready for analysis.

Candidate City	Metro Pop '21	City Pop '21	% Growth (11y)	Distance	Income per Capita	Fortune 500	Other Teams	Unemployment Rate Rank	Remote Work Rank	Public Transit Rank
Chicago	1.000000	1.000000	0.305991	0.000000	0.529786	1.000000	0.000000	0.027237	0.698413	0.868132
San Antonio	0.170684	0.501337	0.655640	0.382330	0.000000	0.057143	1.000000	0.225681	1.000000	0.659341
Oklahoma City	0.031410	0.195205	1.000000	0.580583	0.183357	0.057143	1.000000	0.894942	0.825397	0.076923
St. Louis	0.124371	0.037191	0.000000	0.471845	0.669254	0.228571	0.666667	0.844358	0.857143	0.109890
Sacramento	0.117048	0.130029	0.774568	0.185243	0.362432	0.000000	1.000000	0.000000	0.285714	0.340659
Salt Lake City	0.000000	0.000000	0.586545	1.000000	0.540188	0.000000	0.666667	1.000000	0.888889	0.000000
San Diego	0.251142	0.473196	0.517158	0.229126	0.715383	0.057143	1.000000	0.311284	0.000000	0.901099
Oakland	0.413360	0.093485	0.717614	0.009709	0.855811	0.000000	1.000000	0.470817	0.373016	0.967033
San Jose	0.074310	0.313697	0.453260	0.093204	1.000000	0.228571	0.666667	0.521401	0.166667	0.945055
Portland	0.159665	0.176551	0.672674	0.337864	0.748550	0.057143	0.666667	0.373541	0.968254	1.000000

Table 3: Expansion city candidates and their corresponding considerations, after min-max normalization.

Results and Recommendations

Once the min-max normalized scores across all categories were averaged, Chicago narrowly edged out Portland as the most-attractive city to host the NFL's 33rd team.

Rank	Candidate City	Final Score
1	Chicago	0.54
2	Portland	0.52
3	Oakland	0.49
4	Oklahoma City	0.48
5	Salt Lake City	0.47
6	San Antonio	0.47
7	San Jose	0.45
8	San Diego	0.45
9	St. Louis	0.40
10	Sacramento	0.32

Table 4: Final candidate city rankings

Chicago boasted the highest ranking in four of the 10 categories. No other city had more than two 1.0 scores. Still, after factoring in the city's weakest factors – several existing professional sports teams, troublesome unemployment rank, low population growth 2010-2021 – the Windy City's final score was within seven percentage points of half of the other candidate cities.

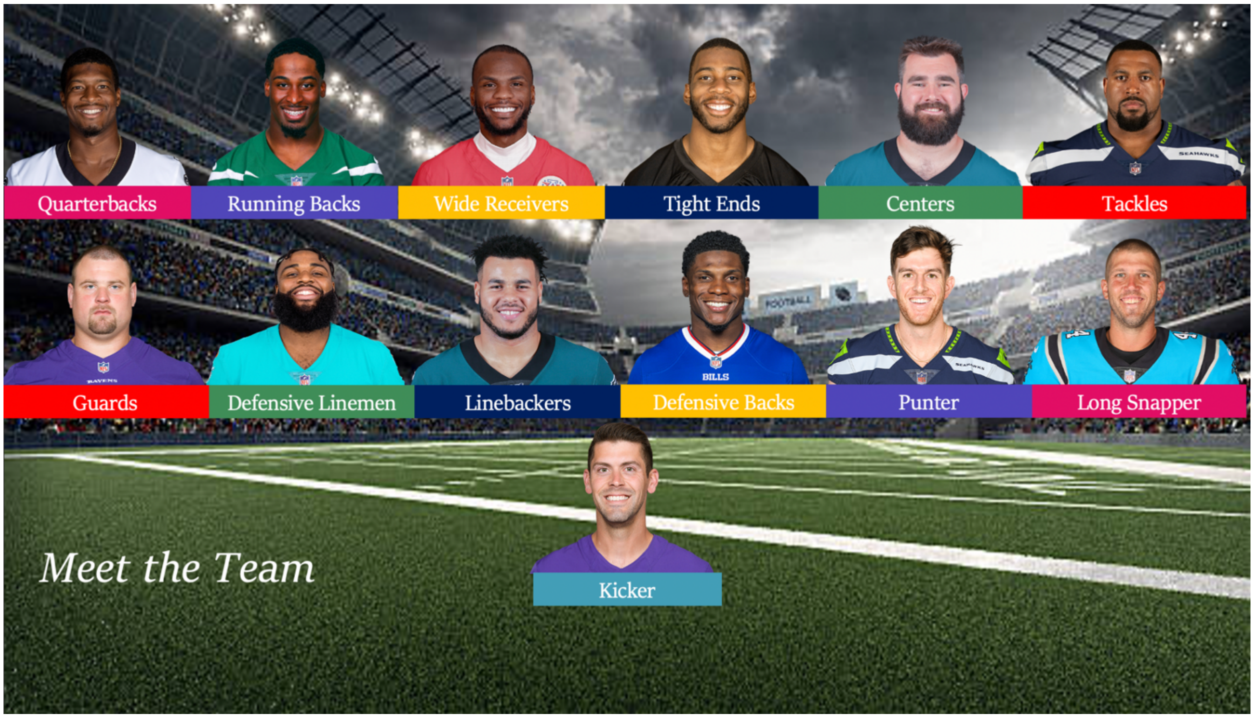
We already listed the reasons we included each of the 10 categories into our scoring model. But why did we select these 10 particular cities?

Oakland, St. Louis, and San Diego were included because all of these markets previously had NFL teams, or are already in the process of vying for the next one (Elgazzar 2023). In addition, Portland, Sacramento, San Jose, San Antonio, Salt Lake City and Oklahoma City either have or have had NBA or NHL teams, and additionally appeared in several articles discussing potential expansion cities (2023). Having a minimum of one existing professional sports team implies that there is already a large enough sports fan base in the region to potentially add that city to the field.

Why are there no international cities in our model? After all, as American football continues to disseminate around that globe, some may consider adding a second NFL team in Chicago to be redundant.

Fortunately, this question has already been answered by the NFL: some of the most important considerations in reinforcing international growth do not involve adding additional teams, but rather, creating direct access to international markets with existing teams, in what the league is calling International Home Marketing Areas (IHMA). Instead of adding an expansion team abroad, the NFL is already taking a more holistic approach to international awareness by “granting 18 clubs direct access to international markets…to drive NFL fan growth and evolve their global brands through marketing, fan engagement, and commercialization” (Elgazzar 2023). Therefore, our 10-city field vying for the NFL’s 33rd team were all within the United States.

Website Concept



What makes Chicago the best candidate for the NFL's 33rd team?

This is an immense decision that involves a multitude of civic, political, geographical, legal, and economic factors.
That is why we have selected 10 hypothetical candidate cities as potential hosts of the NFL's 33rd team, and built a holistic scoring system to evaluate them based on 10 key components (above). Click here for the results.

RESULTS

Chicago, as the United States' third largest metropolitan area, narrowly edged out all other candidate cities.

Rank	Candidate City	Final Score
1	Chicago	0.54
2	Portland	0.52
3	Oakland	0.49
4	Oklahoma City	0.48
5	Salt Lake City	0.47
6	San Antonio	0.47
7	San Jose	0.45
8	San Diego	0.45
9	St. Louis	0.40
10	Sacramento	0.32

Table 3. Final candidate city rankings

Appendix I

R Code for Constructing a Projected Roster Using a Constrained Optimized Approach

```
#import necessary libraries
library(lpSolve)
```

```
library(readxl)
```

```
final_qb_data <- read_excel("final_qb_data.xlsx")
final_rb_data <- read_excel("final_rb_data.xlsx")
final_wr_data <- read_excel("final_wr_data.xlsx")
final_te_data <- read_excel("final_te_data.xlsx")
final_c_data <- read_excel("final_c_data.xlsx")
final_ta_data <- read_excel("final_ta_data.xlsx")
final_g_data <- read_excel("final_g_data.xlsx")
final_dl_data <- read_excel("final_dl_data.xlsx")
final_lb_data <- read_excel("final_lb_data.xlsx")
final_db_data <- read_excel("final_db_data.xlsx")
final_k_data <- read_excel("final_k_data.xlsx")
final_p_data <- read_excel("final_p_data.xlsx")
final_ls_data <- read_excel("final_ls_data.xlsx")
```

```
#Boxplots for performance metrics to determine outliers
par(mar=c(2,2,2,2))
par(mfrow=c(5,3))
boxplot(final_qb_data$qbr,horizontal=TRUE,main="QB:QBR",col=c("tomato"))
boxplot(final_rb_data$yards_per_touch,horizontal=TRUE,main="RB: Yards/Touch",col=c("tomato"))
boxplot(final_wr_data$yards_per_touch,horizontal=TRUE,main="WR: Yards/Touch",col=c("tomato"))
boxplot(final_te_data$yards_per_touch,horizontal=TRUE,main="TE: Yards/Touch",col=c("tomato"))
boxplot(final_c_data$snap_counts,horizontal=TRUE,main="C: Snap Count",col=c("tomato"))
boxplot(final_ta_data$snap_counts,horizontal=TRUE,main="TA: Snap Count",col=c("tomato"))
boxplot(final_g_data$snap_counts,horizontal=TRUE,main="G: Snap Count",col=c("tomato"))
boxplot(final_dl_data$tackles_per_snap,horizontal=TRUE,main="DL: Tackles",col=c("tomato"))
boxplot(final_lb_data$tackles_per_snap,horizontal=TRUE,main="LB: Tackles",col=c("tomato"))
boxplot(final_db_data$tackles_per_snap,horizontal=TRUE,main="DB: Tackles",col=c("tomato"))
boxplot(final_k_data$made_fg_percentage,horizontal=TRUE,main="K: Made Field Goal %",col=c("tomato"))
boxplot(final_p_data$percent_punts_inside_20,horizontal=TRUE,main="P: % punts inside 20",col=c("tomato"))
boxplot(final_ls_data$games_started,horizontal=TRUE,main="LS: games started",col=c("tomato"))
```

```r
#Only choose C, G, TA, DL, LB, and DB with snap counts greater than the 10th percentile
final_c_data=final_c_data[final_c_data$snap_counts>=quantile(final_c_data$snap_counts,.1),]
final_g_data=final_g_data[final_g_data$snap_counts>=quantile(final_g_data$snap_counts,.1),]
final_ta_data=final_ta_data[final_ta_data$snap_counts>=quantile(final_ta_data$snap_counts,.1),]
final_dl_data=final_dl_data[final_dl_data$snaps>=quantile(final_dl_data$snaps,.1),]
final_lb_data=final_lb_data[final_lb_data$snaps>=quantile(final_lb_data$snaps,.1),]
final_db_data=final_db_data[final_db_data$snaps>=quantile(final_db_data$snaps,.1),]
```

```r
#Create a standardized variable 'performance' that can be generalized to all positions
final_qb_data$performance=scale(final_qb_data$qbr)
final_rb_data$performance=scale(final_rb_data$yards_per_touch)
final_wr_data$performance=scale(final_wr_data$yards_per_touch)
final_te_data$performance=scale(final_te_data$yards_per_touch)
final_c_data$performance=scale(final_c_data$snap_counts)
final_ta_data$performance=scale(final_ta_data$snap_counts)
final_g_data$performance=scale(final_g_data$snap_counts)
final_dl_data$performance=scale(final_dl_data$tackles_per_snap)
final_lb_data$performance=scale(final_lb_data$tackles_per_snap)
final_db_data$performance=scale(final_db_data$tackles_per_snap)
final_k_data$performance=scale(final_k_data$made_fg_percentage)
final_p_data$performance=scale(final_p_data$percent_punts_inside_20)
final_ls_data$performance=scale(final_ls_data$games_started)
```

```r
#Since we want to draft players that are actually good and experienced
#we will subset the data to only include players that are above the 75th
#percentile in the performance metric for their respective position
```

```r
#Subset QB's above 75th percentile
q3qb=quantile(final_qb_data$performance,.75)
final_qb_data=subset(final_qb_data, final_qb_data$performance>q3qb)
```

```r
#Subset RB's above 75th percentile
q3rb=quantile(final_rb_data$performance,.75)
final_rb_data=subset(final_rb_data, final_rb_data$performance>q3rb)
```

```r
#Subset WR's above 75th percentile
q3wr=quantile(final_wr_data$performance,.75)
final_wr_data=subset(final_wr_data, final_wr_data$performance>q3wr)
```

```r
#Subset TE's above 75th percentile
q3te=quantile(final_te_data$performance,.75)
final_te_data=subset(final_te_data, final_te_data$performance>q3te)
```

```r
#Subset C's above 75th percentile
q3c=quantile(final_c_data$performance,.75)
final_c_data=subset(final_c_data, final_c_data$performance>q3c)
```

```r
#Subset TA's above 75th percentile
q3ta=quantile(final_ta_data$performance,.75)
final_ta_data=subset(final_ta_data, final_ta_data$performance>q3ta)
```

```r
#Subset G's above 75th percentile
q3g=quantile(final_g_data$performance,.75)
final_g_data=subset(final_g_data, final_g_data$performance>q3g)
```

```r
#Subset DL's above 75th percentile
q3dl=quantile(final_dl_data$performance,.75)
final_dl_data=subset(final_dl_data, final_dl_data$performance>q3dl)
```

```r
#Subset LB's above 75th percentile
q3lb=quantile(final_lb_data$performance,.75)
final_lb_data=subset(final_lb_data, final_lb_data$performance>q3lb)
```

```
#Subset DB's above 75th percentile
q3db=quantile(final_db_data$performance,.75)
final_db_data=subset(final_db_data, final_db_data$performance>q3db)
```

```
#Subset K's above 75th percentile
q3k=quantile(final_k_data$performance,.75)
final_k_data=subset(final_k_data, final_k_data$performance>q3k)
```

```
#Subset P's above 75th percentile
q3p=quantile(final_p_data$performance,.75)
final_p_data=subset(final_p_data, final_p_data$performance>q3p)
```

```
#Subset LS's above 75th percentile
q3ls=quantile(final_ls_data$performance,.75)
final_ls_data=subset(final_ls_data, final_ls_data$performance>q3ls)
```

```
#edit each dataset so they all have the same variables
```

```
final_qb_data=subset(final_qb_data,select=-c(qbr))
final_rb_data=subset(final_rb_data,select=-c(touches,yards_from_scrimmage,yards_per_touch))
final_wr_data=subset(final_wr_data,select=-c(touches,yards_from_scrimmage,yards_per_touch))
final_te_data=subset(final_te_data,select=-c(touches,yards_from_scrimmage,yards_per_touch))
final_c_data=subset(final_c_data,select=-c(snap_counts))
final_ta_data=subset(final_ta_data,select=-c(snap_counts))
final_g_data=subset(final_g_data,select=-c(snap_counts))
final_dl_data=subset(final_dl_data,select=-c(snaps,tackles,tackles_per_snap))
final_lb_data=subset(final_lb_data,select=-c(snaps,tackles,tackles_per_snap))
final_db_data=subset(final_db_data,select=-c(snaps,tackles,tackles_per_snap))
final_k_data=subset(final_k_data,select=-c(field_goals_attempted,fied_goals_made,made_fg_percentage))
final_p_data=subset(final_p_data,select=-c(punts,punts_inside_20,percent_punts_inside_20))
final_ls_data=subset(final_ls_data,select=-c(games_started))
```

```
#Combine all players into a single dataset
working_data=rbind(final_qb_data,final_rb_data,final_wr_data,final_te_data,final_c_data,final_ta_data,final_g_dat
a,final_dl_data,final_lb_data,final_db_data,final_k_data,final_p_data,final_ls_data)
```

```
#create constraint matrix
constraint_matrix1=as.matrix(rbind(working_data$QB,working_data$QB,working_data$RB,working_data$RB,working_data$W
R,working_data$WR,working_data$TE,working_data$TE,working_data$C,working_data$C,working_data$TA,working_data$TA,w
orking_data$G,working_data$G,working_data$DL,working_data$DL,working_data$LB,working_data$LB,working_data$DB,work
ing_data$DB,working_data$K,working_data$K,working_data$P,working_data$P,working_data$LS,working_data$LS,working_d
ata$ARI,working_data$ATL,working_data$BAL,working_data$BUF,working_data$CAR,working_data$CHI,working_data$CIN,wor
king_data$CLE,working_data$DAL,working_data$DEN,working_data$DET,working_data$GB,working_data$HOU,working_data$IN
D,working_data$JAC,working_data$KC,working_data$LAC,working_data$LAR,working_data$LV,working_data$MIA,working_dat
a$MIN,working_data$NE,working_data$NO,working_data$NYG,working_data$NYJ,working_data$PHI,working_data$PIT,working
_data$SEA,working_data$SF,working_data$TB,working_data$TEN,working_data$WAS,working_data$average_salary))

dimnames(constraint_matrix1)=list(c("2QBMax","2QBMin","4RBMax","4RBMin","6WRMax","6WRMin","3TEMax","3TEMin","3CMa
x","3CMin","3TMax","3TMin","3GMax","3GMin","9DLMax","9DLMin","7LBMax","7LBMin","10DBMax","10DBMin","1KMax","1KMi
n","1PMax","1PMin","1LSMax","1LSMin","2ARIMax","2ATLMax","2BALMax","2BUFMax","2CARMax","2CHIMax","2CINMax","2CLEM
ax","2DALMax","2DENMax","2DETMax","2GBMax","2HOUMax","2INDMax","2JACMax","2KCMax","2LACMax","2LARMax","2LVMax","2
MIAMax","2MINMax","2NEMax","2NOMax","2NYGMax","2NYJMax","2PHIMax","2PITMax","2SEAMax","2SFMax","2TBMax","2TENMax"
,"2WASMax","SalaryMax"),working_data$player)
```

```
knapsack_object <-
    lp(const.mat = constraint_matrix1,
       objective = working_data$performance,
       direction = "max",
       const.rhs = c(2,2,4,4,6,6,3,3,3,3,3,3,3,3,9,9,7,7,10,10,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,
2,2,2,2,2,2,2,2,2,2,2,2,2,224800000),
       const.dir = c("<=", ">=", "<=", ">=","<=", ">=", "<=", ">=","<=", ">=", "<=", ">=","<=", ">=", "<=", ">=",
"<=", ">=", "<=", ">=","<=", ">=", "<=", ">=","<=", ">=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","
<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=","<=")  ,
       int.vec = 1:190, all.bin = TRUE)
```

```
print(working_data[as.logical(knapsack_object$solution),c("player","team","position","average_salary","performanc
e")],n=53)
```

```
new_roster=subset(working_data[as.logical(knapsack_object$solution),c("player","position","average_salary","perfo
rmance")],n=53)
total_payroll=sum(new_roster$average_salary)
cat("Total Payroll: $",round(total_payroll/1000000,1),"million")
```

Appendix II

Python Code for Scoring Model to Select Candidate City

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_auc_score
from sklearn import preprocessing
import matplotlib.pylab as plt
from sklearn.preprocessing import MinMaxScaler, StandardScaler
```

```
X = pd.read_excel("Expansion City.xlsx")
```

```
X.info()
```

```
X = X.dropna(axis=1, how='all')
X
```

```
# Drop City column from the front of the dataframe
# so SKlearn can evaluate the whole df correctly

X2 = X.drop(['Candidate City'], axis=1)
```

```
X2
```

```
scaler = MinMaxScaler()
normalized_data = scaler.fit_transform(X2)
normalized_df = pd.DataFrame(normalized_data, columns=X2.columns)
```

```
normalized_df
```

```
# Re-add City column to the front of the dataframe now that columns are normalized

City = X['Candidate City'].to_numpy()
```

```
normalized_df.insert(0, 'Candidate City', City)
```

```
normalized_df
```

```
# Specify the columns to be added and inverted

columns_to_add = ["Metro Pop '21" , "City Pop '21", "% Growth (11y)", 'Distance', 'Income per Capita',
                  'Fortune 500']
columns_to_invert = ['Other Teams', 'Unemployment Rate Rank', 'Remote Work Rank', 'Public Transit Rank']
```

```
# Inverting the values of the selected columns
normalized_df[columns_to_invert] = 1 - normalized_df[columns_to_invert]

normalized_df
```

```
# Calculate the final score and add final score to the end of the dataframe

normalized_df['Final Score'] = (normalized_df[columns_to_add].sum(axis=1) +
                                normalized_df[columns_to_invert].sum(axis=1)) / 10
normalized_df
```

```
df_rankings = normalized_df.sort_values(by='Final Score', ascending=False)
df_rankings
```

```
final_rankings = ['Candidate City', 'Final Score']
new_df = pd.DataFrame(df_rankings, columns=final_rankings)
```

```
# Removing all of the inputs and creating a new dataframe with just the final score

new_df
```

```
# Adding rank and background gradient

new_df['Rank'] = new_df['Final Score'].rank(ascending=False)
new_df['Rank'] = new_df['Rank'].astype(str).apply(lambda x: x.replace('.0',''))
first_column = new_df.pop('Rank')
new_df.insert(0, 'Rank', first_column)

final = new_df.sort_values(by='Final Score', ascending = False)
rounder = {'Final Score':'{:,.2f}'}
final.style.format(rounder).background_gradient(cmap= 'Blues')
```

References

Abdalazem, Reem. “The NFL announces international expansion plans in eight countries.” Diario AS. 16 December 2021. https://en.as.com/en/2021/12/17/nfl/1639710690_232451.html.

BetMGM. 2023. “How Much is the NFL Worth?” *BetMGM*, April 4, 2023. https://sports.betmgm.com/en/blog/nfl/how-much-is-the-nfl-worth-bm06/#:~:text=Therefore%2C%20the%20NFL%20is%20likely,gains%20in%20their%20overall%20evaluations.

Brown, Matthew. 2022. “NFL Expansion: How a Massive Expansion Could Work.” *Gridiron Heroics*, October 15, 2022. https://gridironheroics.com/nfl-expansion/#:~:text=NFL%20Expansion%20Draft&text=Basically%20each%20team%20has%20a,came%20with%20their%20current%20salary.

Bureau of Labor Statistics. “Unemployment Rates for Metropolitan Areas, Not Seasonally Adjusted.” February 2023. U.S. Department of Labor. Accessed 24 May 2023. https://www.bls.gov/web/metro/laummtrk.html.

Bruggink, Thomas H. and Doug Schiz. “Location Model in the National Football League: Predicting Optimal Expansion and Relocation Sites.” The Sport Journal. Volume 24. 14 March 2008. https://thesportjournal.org/article/location-model-in-the-national-football-league-predicting-optimal-expansion-and-relocation-sites/.

Elgazzar, Kareem. “Which cities should get an NFL expansion team?” Yardbarker. Updated 24 May 2023. https://www.yardbarker.com/nfl/articles/which_cities_should_get_an_nfl_expansion_team/s1__37690346#slide_3.

Kay, Alex. 2022. “What Would an NFL Expansion Roster Look Like in 2022.” *Bleacher Report*, July 13, 2022. https://bleacherreport.com/articles/10041787-what-would-an-nfl-expansion-roster-look-like-in-2022#:~:text=When%20a%20player%20is%20picked,back%20its%20remaining%20three%20players.

McCann, Adam. “Cities with the Best & Worst Public Transportation.” WalletHub. 10 September 2019. https://wallethub.com/edu/cities-with-the-best-worst-public-transportation/65028.

Perkins, Dave. 2005. “Tagliabue revives idea of bringing NFL to Toronto; City 'a candidate' for expansion team League may play game here next season: MET Edition.” *Toronto Star*, February 5, 2005.

Pierson, Don. 1999. “$700 MILLION BUYS HOUSTON AN NFL EXPANSION TEAM 8-DIVISION FORMAT ALSO SLATED FOR 2002: [CHICAGO SPORTS FINAL Edition].” *Chicago Tribune*, October 7, 1999.

Pro Football Reference. 2023. “List of all the Pro Football Franchises.” Accessed May 10, 2023. https://www.pro-football-reference.com/teams/.

Spotrac. 2023. “NFL Salary Rankings.”. Accessed May 10, 2023. https://www.spotrac.com/nfl/rankings/average/.

U.S. Census Bureau. “City Population Estimates.” 2010, 2021. Accessed May 3-24, 2023. https://www.census.gov/.

U.S. Census Bureau. “Metropolitan statistical area population estimates.” 2021. Accessed May 3-24, 2023.
https://www.census.gov/.

U.S. Census Bureau. “Per capita income in past 12 months (in 2021 dollars).” 2021. Accessed May 3-24, 2023.
https://www.census.gov/.

“Visualize the Fortune 500.” Fortune.com. 2022. Accessed 24 May 2022. https://fortune.com/franchise-list-page/visualize-the-fortune-500-2022/.

“2023’s Best Cities for Remote Workers.” LawnStarter. 12 October 2022. https://lawnstarter.wpengine.com/blog/studies/best-worst-us-cities-for-remote-workers/.