Overview

This report is sought to better understand how people feel about MLS. How does MLS attract fans in the States? Since not all respondents to the survey are MLS fans, we arranged different questions for three groups: MLS fans, other sports fans, and non-sports fans. MLS fans and other sports fans were asked what sports they watch and why they enjoy sports games. Other sports fans and non-sports fans were asked what barriers prevent them from following the MLS and what might be the potential reason that would lead them to the MLS fandom in the future. Also, MLS fans were asked what they thought the MLS could've been better.

Exploratory Research

Our exploratory research discovered that soccer is swelling due to immigration from soccer-loving nations and a few decades of supporting youth soccer leagues. It helps that soccer is already the world's most popular sport. On the other hand, the MLS has successfully attracted a younger audience, particularly millennials, partly due to the game's fast-paced nature and less ad saturation compared to other American sports. The league's diverse player base also resonates with the nation's immigrant population, making it more inclusive. Soccer's growing popularity among the youth and its status as a global sport contribute to its increasing fanbase in the U.S. Last but not least, Lionel Messi's move to Inter Miami CF has significantly boosted the league's visibility. While the increase in average attendance was modest, there was a dramatic rise in MLS season pass subscriptions and ticket prices, showcasing the massive influence of the Messi effect.

The results of the exploratory research informed the confirmatory research by suggesting that future surveys should target potential audiences more directly to gather evidence for effective strategies. It also indicated that while star players can attract immediate attention, long-term growth may require a focus on cultivating a younger and more diverse fanbase and considering the rarity of acquiring players with Messi's level of influence.

Confirmatory Research

Survey Goal

Our exploratory data mainly reflects present circumstances, so we would like to build forecasting models or surveys to predict long-term fan behavior and preference trends. Moreover, because of Messi's exceptional influence, caution should be exercised in generalizing findings based on his impact alone. Hence, the questions aim to assess the excitement level of sports games, the barriers to becoming a sports fan, the following of MLS, fan loyalty, viewing habits, and potential improvements to enhance the fan experience. Additionally, demographic information is collected to analyze the data within different population segments.

Survey Result

Our confirmatory research surveyed 173 participants; all were asked in an online and anonymous survey. The total number of responses we use is 169 because we dropped 4 responses due to inappropriate answers.

There are 66 MLS fans, 66 other sports fans, and 37 non-sports fans. Since “MLS” is in the survey title, the proportion of the fans and non-fans should not be considered. Especially for non-sports fans' responses should be more biased. Hence, we used a two-way chi-squared test to determine how age, gender, and annual household income differentiate between MLS and other sports fans.

Our two-way chi-squared test results indicated that age, gender, and income do not affect sports fans’ preferences between MLS and other sports significantly. However, the bar chart shows that for people aged 18-34, MLS fans are 60% more than other sports fans, while respondents at other ages have a 50-50 chance of becoming one of the two groups. This corresponds to the fact that MLS has successfully attracted a younger audience. Also, for those whose annual income is below $49,999, MLS fans are 50% less than other sports fans, while respondents at other ages have a 50-50 chance of becoming one of the two groups.

In addition, our goal is to improve the MLS’s promotion, so we’ll figure out MLS fans' preferences and other sports fans' preferences by cluster analysis.

Segmentation

MLS Fans

“Matches_Excitement,” “Supporting_Favorite_Team,” “Specific_Players,” “Community_Camaraderie,” and “Sports_Related_Content_n_News_Engagement” are our variables to split the clusters. By performing a hierarchical cluster analysis, we decided to identify three groups due to the resulting dendrogram. We then ran a K-means cluster analysis with three clusters on the dataset and obtained a set of cluster centers describing the segments. There are three types of MLS fans: highly engaged, moderately engaged, and low engaged. Of our 66 MLS fans, there are 38 highly engaged fans, 13 moderated fans, and 15 low-engaged fans.

On a scale from 1 to 5, the excitement level of the aspects of sports games and events, which are the five variables, to highly engaged fans are rated 5. Moderately engaged fans value the excitement of the matches, the support of their favorite teams, and specific players more, having a rating of 4 on those factors and a rating of 3 on community camaraderie and sports-related content and news engagement. Lastly, low-engaged fans followed sports-related content and news with a rating of 3 and had other variables rated 2.

Highly engaged fans are enthusiastic in all aspects of the game, so they might be interested in every piece of the MLS. For moderately engaged fans, we think match highlights and team-related and player-related stats or merchandise might draw their attention the most. As for low-engaged fans, they might be more interested in some breaking news, interviews with the athletes, and insightful podcasts. Note that our proportion of the fanbase might be biased due to the title name (highly engaged fans might be more interested in answering the survey); the league should implement its promotion strategy after clarifying what percentage each cluster shares.

Other Sports Fans

Not only “Matches_Excitement,” “Supporting_Favorite_Team,” “Specific_Players,” “Community_Camaraderie,” and “Sports_Related_Content_n_News_Engagement,” but there are also “No_MLS_Unfamiliar_with_rules_of_soccer” and “No_MLS_have_other_preferred _sports” as variables to split the clusters. Three types of other sports fans are Selective Sports Followers, Casual Sports Watchers, and Neutral Sports Fans. Of our 66 other sports fans, there are 23 Selective Sports Followers, 9 Casual Sports Watchers, and 35 Neutral Sports Fans.

Selective Sports Followers rated most categories highly, including “No_MLS_Unfamiliar_with_rules_of_soccer” and “No_MLS_have_other_preferred _sports.” Those might be loyal to their favorite sports and not interested in engaging in a new fanbase.

Casual Sports Watchers have the lowest scores across all categories; however, this might indicate that they would have a higher possibility of following the MLS if there is an opportunity for them to watch the MLS matches since they do not think that there are significant barriers for them to follow the MLS.

Cluster 3, also known as ' Neutral Sports Fans,' consists of individuals who are crazy about the sports, rating each aspect of the sport a grade of 4. Also, they’re not ruling out watching MLS matches due to the unfamiliarity of rules or due to being loyal to other sports. Similar to the previous cluster, they might have a higher possibility of following the MLS if there is an opportunity for them to watch the MLS matches. If the MLS can improve the game's contents, there would be a higher incentive for individuals in cluster 3 to follow the MLS.

Remaining Variables and Our Clusters

When we conducted the two-way chi-squared on personal information and both MLS fans’ clusters and other sports fans’ clusters, our clusters had no significant characteristics. Thus, it’d be better for the MLS league to clarify the actual proportion of those clusters to construct the promoting strategies instead of focusing on other external factors. Once we know the actual proportion of the segments, we will then be able to know how to allocate the marketing cost for each aspect of the game.

Appendix

Survey Questions

1. Please Rate the Excitement Level of These Aspects of Sports Games and Events.

2. Do you follow any sports in daily life?

3. How likely would you watch a sports game under the following circumstances?

4. How challenging are these barriers preventing you from being a sports fan?

5. Are you now following MLS?

6. When did you first follow MLS, and why did you follow them?

7. How long have you been an Major League Soccer (MLS) fan?

8. Do you have a favorite MLS team? If so, which team is it, and why do you support them?

9. Are there specific MLS players you admire or consider your favorites? If yes, please share their names and what you appreciate about them.

10. How do you watch MLS games?

11. What's your ticket plan, and about how much do you spend on going to MLS games?

12. What's your MLS online livestream plan, and about how much is it?

13. In your opinion, what improvements or changes could MLS make to enhance the fan experience and attract more supporters?

14. How Significantly Do These Factors Prevent You From Following Major League Soccer (MLS)?

15. What factors may attract your attention in following MLS in the future?

16. Do you watch any professional sports games other than MLS?

17. How do you watch those games?

18. What's your ticket plan, and about how much do you spend on going to those games?

19. What's your online live stream plan, and about how much is it?

20. What is your age?

21. What is your Gender?

22. What's your ethnicity?

23. What's your employment status?

24. How much is your annual household income?

25. If you have any additional comments, thoughts, or suggestions, please share them in the space provided below. (Not required)

Note: Question 3-4 are for non-sports fans only, and question 5 is for MLS fans and other sports fans, question 6-13 are for MLS fans only, question 14-16 are for non-sports fans only and other sports fans, question 17-19 are for other sports fans only.

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Age * Follow_MLS_or_not	132	100.0%	0	0.0%	132	100.0%
Gender * Follow_MLS_or_not	131	99.2%	1	0.8%	132	100.0%
Annual_Household_Income * Follow_MLS_or_not	132	100.0%	0	0.0%	132	100.0%

1. Age x Following MLS

Note: 1: Under24, 2: 25-34, 3: 35-44, 4: 45-54, 5: 55-64, 6: Over 65

Crosstab

Count

		Follow_MLS_or_not		
		0	1	Total
Age	1	1	4	5
	2	19	28	47
	3	24	18	42
	4	14	14	28
	5	4	2	6
	6	4	0	4
Total		66	66	132

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9.047[a]	5	.107
Likelihood Ratio	10.746	5	.057
Linear-by-Linear Association	6.161	1	.013
N of Valid Cases	132		

a. 6 cells (50.0%) have expected count less than 5. The minimum expected count is 2.00.

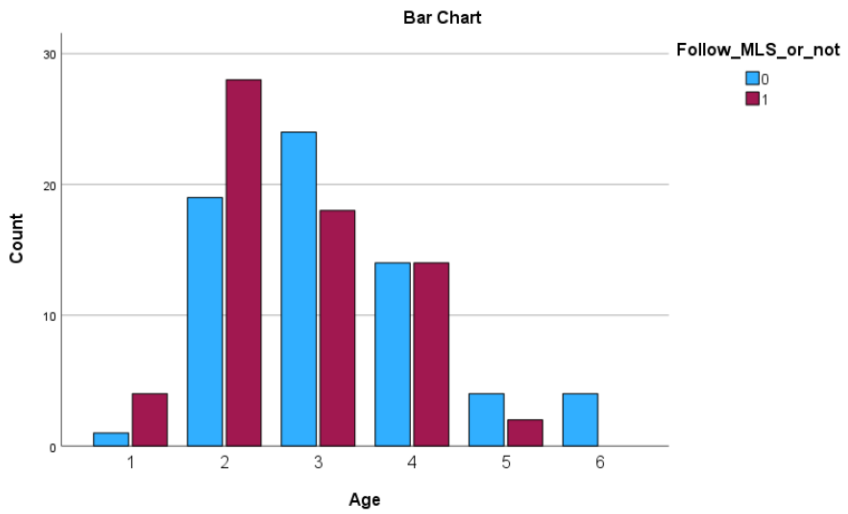
Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Interval by Interval	Pearson's R	-.217	.079	-2.533	.013[c]
Ordinal by Ordinal	Spearman Correlation	-.196	.084	-2.284	.024[c]
N of Valid Cases		132			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.



2. Gender x Following MLS

Note: 0:Female, 1: Male

Crosstab

Count

		Follow_MLS_or_not		
		0	1	Total
Gender	0	26	28	54
	1	39	38	77
Total		65	66	131

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.079[a]	1	.778		
Continuity Correction[b]	.011	1	.917		
Likelihood Ratio	.079	1	.778		
Fisher's Exact Test				.860	.459
Linear-by-Linear Association	.079	1	.779		
N of Valid Cases	131				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 26.79.

b. Computed only for a 2x2 table

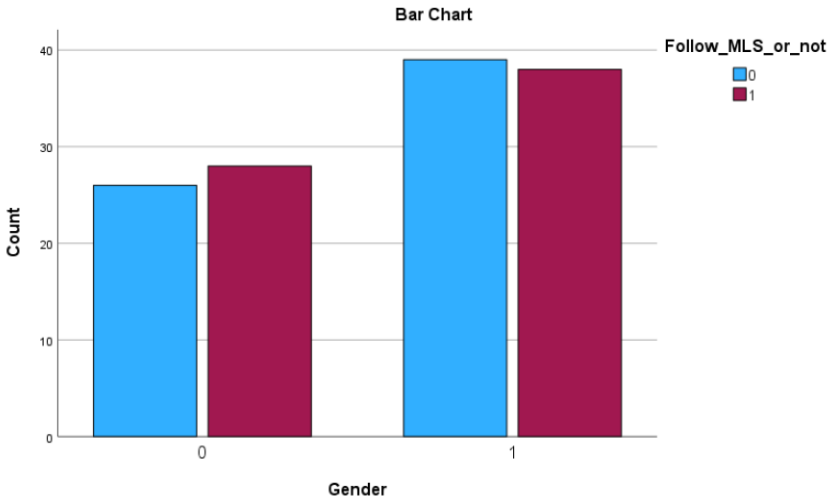
Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Interval by Interval	Pearson's R	-.025	.087	-.280	.780[c]
Ordinal by Ordinal	Spearman Correlation	-.025	.087	-.280	.780[c]
N of Valid Cases		131			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.



3. Annual Household Income x Following MLS

Note: 'Under $25,000': 1, '$25,000 - $49,999': 2, '$50,000 - $74,999': 3, '$75,000 - $99,999': 4, '$100,000 - $149,999': 5, '$150,000 or more': 6

Crosstab

Count

		Follow_MLS_or_not		
		0	1	Total
Annual_Household_Income	1	10	5	15
	2	12	7	19
	3	10	17	27
	4	9	10	19
	5	15	16	31
	6	10	11	21
Total		66	66	132

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	4.930[a]	5	.425
Likelihood Ratio	4.999	5	.416
Linear-by-Linear Association	1.277	1	.258
N of Valid Cases	132		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.50.

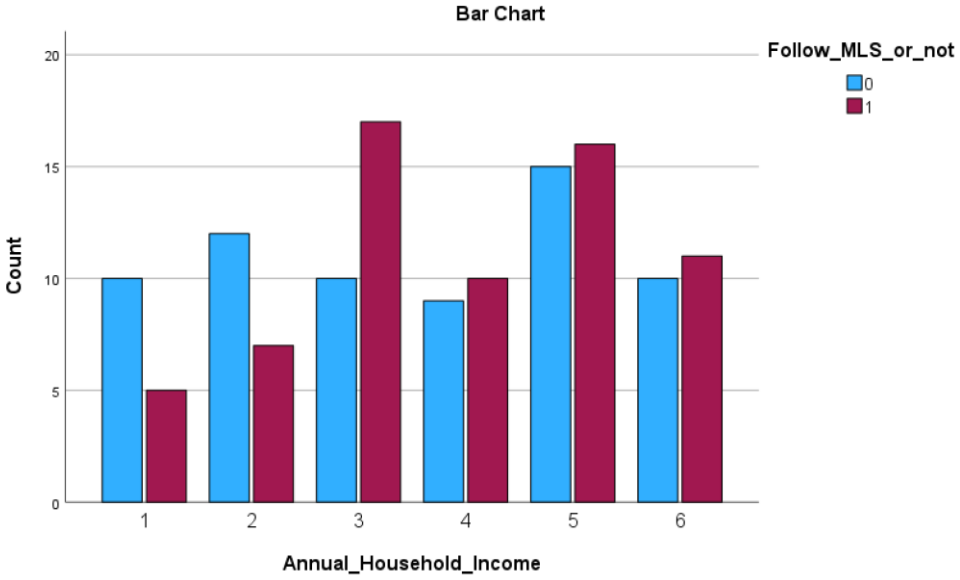
Symmetric Measures

		Value	Asymptotic Standard Error [a]	Approximate T [b]	Approximate Significance
Interval by Interval	Pearson's R	.099	.086	1.131	.260[c]
Ordinal by Ordinal	Spearman Correlation	.093	.087	1.065	.289[c]
N of Valid Cases		132			

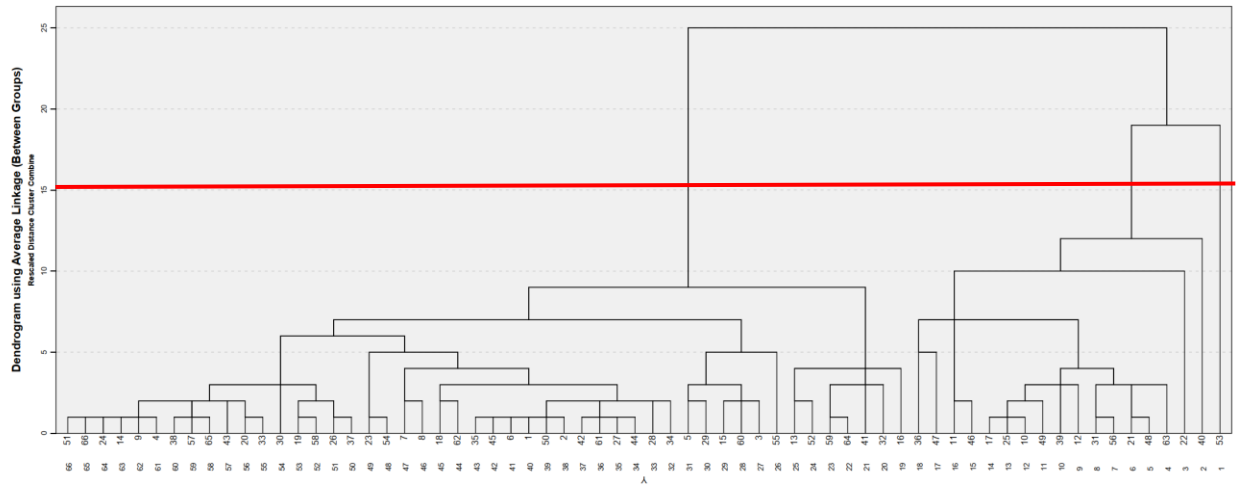
a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.



MLS Fan Dendrogram



MLS Fans K-Mean Cluster, K = 3

Final Cluster Centers

	Cluster 1	Cluster 2	Cluster 3
Matches_Excitement	5	4	2
Supporting_Favorite_Team	5	4	2
Specific_Players	5	4	2
Community_Camaraderie	5	3	2
Sports_Related_Content_n_News_Engagement	5	3	3

Number of Cases in each Cluster

Cluster	1	38.000
	2	13.000
	3	15.000
Valid		66.000
Missing		.000

ANOVA

	Cluster Mean Square	Cluster df	Error Mean Square	Error df	F	Sig.
Matches_Excitement	46.129	2	.345	63	133.669	<.001
Supporting_Favorite_Team	31.778	2	.321	63	99.101	<.001
Specific_Players	31.319	2	.448	63	69.903	<.001
Community_Camaraderie	33.405	2	.353	63	94.709	<.001
Sports_Related_Content_n_News_Engagement	26.239	2	.582	63	45.056	<.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

MLS Fans' Age, Gender, Income vs Clusters

Age * Cluster Number of Case

Crosstab

Count

		Cluster Number of Case			
		1	2	3	Total
Age	1	3	0	1	4
	2	17	5	6	28
	3	10	4	4	18
	4	8	3	3	14
	5	0	1	1	2
Total		38	13	15	66

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	3.968[a]	8	.860
Likelihood Ratio	5.438	8	.710
Linear-by-Linear Association	.776	1	.378
N of Valid Cases	66		

a. 10 cells (66.7%) have expected count less than 5. The minimum expected count is .39.

Gender * Cluster Number of Case

Crosstab

Count

		Cluster Number of Case			
		1	2	3	Total
Gender	0	14	5	9	28
	1	24	8	6	38
Total		38	13	15	66

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2.465[a]	2	.292
Likelihood Ratio	2.444	2	.295
Linear-by-Linear Association	2.029	1	.154
N of Valid Cases	66		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.52.

Annual_Household_Income * Cluster Number of Case

Crosstab

Count

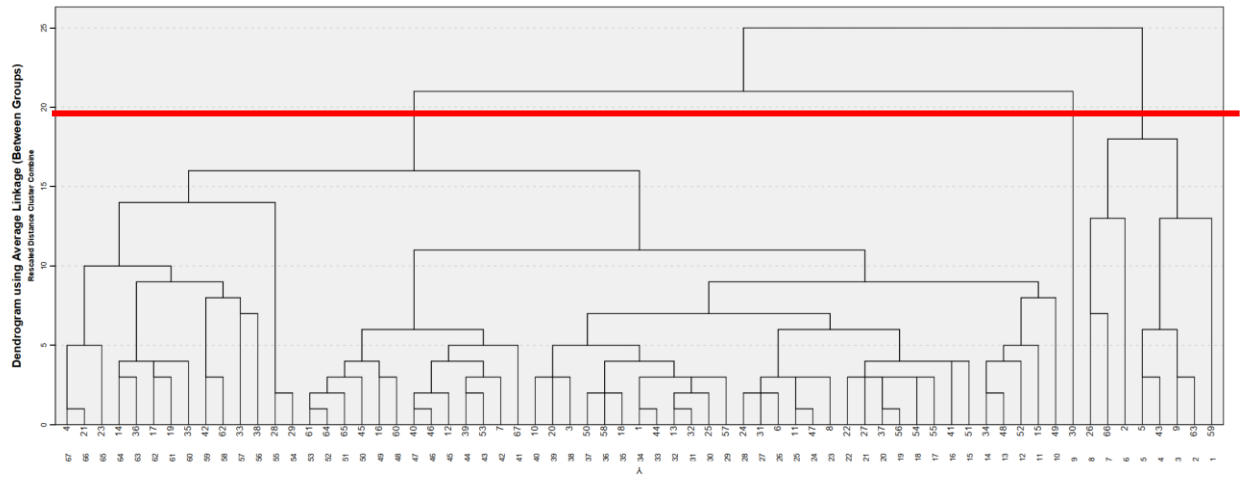
		Cluster Number of Case			
		1	2	3	Total
Annual_Household_Income	1	3	2	0	5
	2	4	1	2	7
	3	10	4	3	17
	4	6	3	1	10
	5	10	1	5	16
	6	5	2	4	11
Total		38	13	15	66

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	7.225[a]	10	.704
Likelihood Ratio	8.610	10	.569
Linear-by-Linear Association	1.011	1	.315
N of Valid Cases	66		

a. 14 cells (77.8%) have expected count less than 5. The minimum expected count is .98.

Other Sports Fans Dendrogram



Other Sports Fans K-Mean Cluster, K = 3

Final Cluster Centers

	Cluster		
	1	2	3
Matches_Excitement	4	2	4
Supporting_Favorite_Team	5	2	4
Specific_Players	5	2	4
Community_Camaraderie	5	2	4
Sports_Related_Content_n_News_Engagement	4	2	3
No_MLS_Unfamiliar_with_rules_of_soccer	4	2	2
No_MLS_have_other_preferred_sports	4	2	3

Number of Cases in each Cluster

Cluster	1	23.000
	2	9.000
	3	35.000
Valid		67.000
Missing		.000

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Matches_Excitement	17.975	2	.617	64	29.115	<.001
Supporting_Favorite_Team	22.031	2	.718	64	30.693	<.001
Specific_Players	18.025	2	.530	64	33.978	<.001
Community_Camaraderie	16.427	2	.817	64	20.098	<.001
Sports_Related_Content_n_News_Engagement	13.445	2	.996	64	13.501	<.001
No_MLS_Unfamiliar_with_rules_of_soccer	31.457	2	.922	64	34.125	<.001
No_MLS_have_other_preferred_sports	18.169	2	1.749	64	10.386	<.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Other Sports Fans' Age, Gender, Income vs Clusters

Age * Cluster Number of Case

Crosstab

Count

		Cluster Number of Case			
		1	2	3	Total
Age	1	0	0	1	1
	2	7	4	9	20
	3	7	3	14	24
	4	5	2	7	14
	5	2	0	2	4
	6	2	0	2	4
Total		23	9	35	67

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	3.816[a]	10	.955
Likelihood Ratio	5.133	10	.882
Linear-by-Linear Association	.232	1	.630
N of Valid Cases	67		

a. 13 cells (72.2%) have expected count less than 5. The minimum expected count is .13.

Gender * Cluster Number of Case

Crosstab

Count

		Cluster Number of Case			
		1	2	3	Total
Gender	0	10	3	13	26
	1	13	6	21	40
Total		23	9	34	66

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.318[a]	2	.853
Likelihood Ratio	.320	2	.852
Linear-by-Linear Association	.133	1	.715
N of Valid Cases	66		

a. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 3.55.

Annual_Household_Income * Cluster Number of Case

Crosstab

Count

		Cluster Number of Case			
		1	2	3	Total
Annual_Household_Income	1	6	1	3	10
	2	5	1	6	12
	3	3	1	7	11
	4	5	1	3	9
	5	3	4	8	15
	6	1	1	8	10
Total		23	9	35	67

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	11.950[a]	10	.288
Likelihood Ratio	11.918	10	.291
Linear-by-Linear Association	4.646	1	.031
N of Valid Cases	67		

a. 12 cells (66.7%) have expected count less than 5. The minimum expected count is 1.21.