Targeting Brand Ambassadors using  
Data from the Olympic Games

Christina Huffaker   
*MS Data Science*  
*Utica College*Utica, USA  
clhuffak@utica.edu

Sujoy Datta Choudhury  
*MS Data Science*  
*Utica College*Utica, USA  
@utica.edu

Zaozian Liu  
*MS Data Science*  
*Utica College*Utica, USA  
@utica.edu

*Abstract*— This project reviews a common marketing strategy of sporting goods companies, focusing on the sponsorship tactics used to engage consumers and increase brand visibility, trust, and positive associations. Reviewing the literature reveals that the current model of sponsorship is overly costly and does not deliver the necessary ROI. Alternative strategies are proposed, based on works which suggest that it is not necessary for consumer emotions associated with a sporting event to be positive in order for purchases to be made.

Keywords—Data Science, Data Visualization, Sport Sponsorships, Marketing

# Introduction

A popular marketing strategy among sporting goods companies aims to increase visibility and positive associations of brands among potential consumers. The primary means of accomplishing this aim is by providing branded team apparel to high-school and college athletic teams, particularly focusing on successful teams with a record of wins [1]. This approach is overly costly based on projected ROI, wasting valuable resources with this model [1], and an alternative model for athlete sponsorships is urgently needed.

Oculytica, LLC will outline which Olympic teams to target for sponsorship, using different criteria to engage high-achieving athletes early in their careers, and expose the brand to a wide potential customer base, outpacing competitors who continue to use the popular, but questionable, model.

# Related Work

## Previous Studies

Jensen et al. [1] writes on the ROI for sporting goods companies on large investments in sponsorships, determining that most companies are unwisely paying more than expected ROI. Additionally, the expense of sponsoring highly-successful teams does not seem to be worthwhile, resulting in unsustainable company spending. Wang & Kaplanidou [2] describe the impact of psychological emotion-lifting on spectators of sporting events. Even negative emotions induced by sports (such as a team loss) can be leveraged to generate purchases, although the impact of negative emotions may result in fewer purchases than positive emotions. Research by Mazodier, Corsi and Quester [3] write about brand association. They show that advertising messages typical of a specific event are most effective in transferring associations of an event to a brand, possibly more so than sponsorships. Unusual marketing campaigns usually do not result in significant instances of brand association [3].

## Synthesis of Previous Studies

Taken together, this research presents an alternative, and not yet popular, tactic for sporting good companies’ marketing strategies. In review of literature, no studies were identified in which a company synthesized the aforementioned works. Oculytica, LLC, in using these works to inform and guide this analysis, hopes to generate a new vein of scholarship on sports sponsorship.

# Data description

## Description of Kaggle Data

## The Kaggle Dataset “120 Years of Olympic History: Athletes and Results” [4] provides 136,000 observations on 15 variables. Originally, the data was scraped from the website Sports Reference by the user rgriffin in May 2018 [4]. Directly downloaded by the Oculytica team from Kaggle, the data required pre-processing to prepare for the analysis.

## Data Pre-Processing

In preparatory data processing, the following variables were selected: Sex, Age, Height, Weight, Team, Year, Season, City, Sport, Medal, Country, and Continent. The data was provided in .csv format, allowing it to be easily imported into R and Tableau without modifying the format. Each row in the data corresponds to an individual athlete competing in an individual event [1]. After data cleaning was performed, a new dataset containing 85,256 rows with 11 features Sex (nominal), Age (ordinal), Height (ordinal), Weight (ordinal), Country, Year, Season, NOC, Sport, and Medal was created. The NOC column was joined with a separate file with Country Code for plotting the world map.

# Methods

### Data preprocessing consisted of extracting the relevant data, then aggregating, sorting, joining, and appending data frames. Following the data cleaning and pre-processing, several preliminary analyses were performed, which were then used to inform the most useful focus for the final analysis. The concept to focus on athletes under age 19 arose from the literature, as it is presumed that younger athletes, who have had fewer opportunities to compete, are naturally less decorated, and therefore less sought after by companies using the typical sponsorship tactics. Following this initial analysis, a subset of data from the year 2000 - 2016 (the most recent year included in the data) was extracted to focus the larger analysis on the most recent, and therefore relevant, trends across players and countries. This subset includes 5 summer Olympic events from 2000 to 2016, and 4 Winter Olympic events from 2002 to 2014. Players under 19 year old, both male and female, from all countries, were included, but the primary focus is on medal winning players who are not from the top 10 medal-winning countries.

### The specific types of visualizations were determined by the data type for each column of interest. Categorical data, which was the main data type in this dataset, is best represented using bar and line charts, so the majority of the graphics are in these formats. Other visualization types, such as pie charts, do not display categorical data as effectively. For the Country column, the data was converted to geospatial types using Tableau, with the purpose of visualizing the data on a map. Within R, the bar and line charts were created, while Tableau was used for the line chart, dot graphs, and map visualization.

### Design considerations also influenced choices in the display of output within both R and Tableau. Color is an important choice in data visualizations. Ward, et al [5] provide examples of common color scales (p. 154), which informed the selection for each of the graphics presented here. Colors need to be representative of the data, but able to be perceived and distinguished easily. The selections were made for each visualization accordingly. According to Tufte [6], there are five principles which guide graphical design, which heavily influenced the design of the graphics in both R and Tableau. With minimal extraneous features, the data is clearly shown, and its depiction is the central focus of each visualization. Features which do not depict the data are minimized for clarity.

# Results

## Results using R

Preliminary line chart results indicate number of medal winners based on gender demographics, shown in Figure 1. Line charts and bar charts were produced for better understanding and easy comparison. Two charts were produced to compare the demographics of male and female players in both the Summer and Winter Olympics for those all athletes who competed and those athletes who were medal-winning.

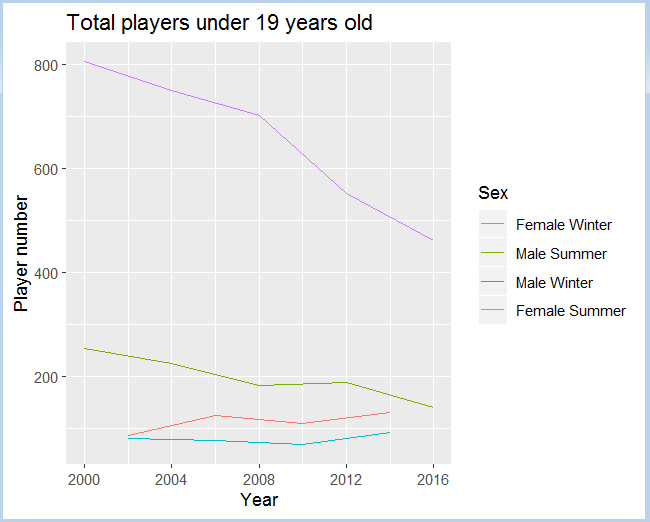


Figure 1. Total numbers of Olympic Athletes under 19 in Summer and Winter Olympics by gender.

##### 

Figure 2. Total numbers of medal-winning Olympic Athletes under 19 in Summer and Winter Olympics by gender.

This figure shows two noteworthy features. First, the gender demographic, dividing the athletes into categories of males and females, reveal that there have been more female participants under age 19 in nearly every year displayed in Figure 1, but during the Winter Olympics, which occurred in 2002, 2006, 2010, and 2014, the disparity is somewhat smaller than during the Summer Games, which occurred in 2000, 2004, 2008, 2012, and 2016. It is also important to note the decline in total number of medal-winning athletes under 19 from 2012 to 2016 during the Summer Games, while the Winter Games show decreases and increases. Though the direct cause of this decline is unclear, at this point, it could indicate a lack of participants in this age category, medal-winning or non-medal-winning alike.

In Figures 3 and 4, bar charts depict the number of all athletes from medal-winning countries in the Summer and Winter Olympic Games under age 19 by country tier. Tiers were defined by grouping the top 10 medal-winning countries as the top tier, and the next 11-20 medal-winning countries into the non-top ten tier.

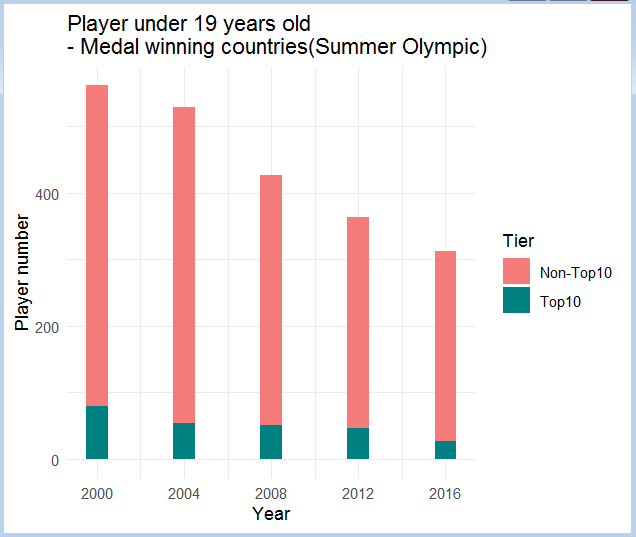


Figure 3. The number of all athletes from medal-winning countries in the Summer Olympic Games under age 19 by country tier.

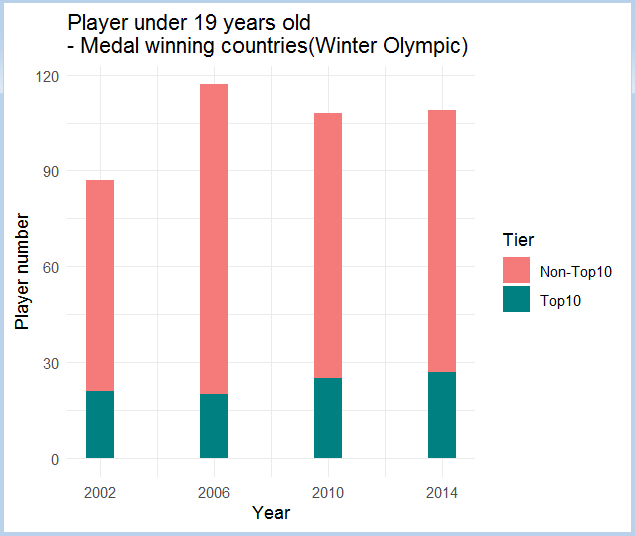


Figure 4. The number of all athletes from medal-winning countries in the Winter Olympic Games under age 19 by country tier.

The bar charts in Figures 5 and 6 show the number of male and female athletes under 19 years old from medal winning countries which are not in the top 10 medal winning countries. They also indicate that there are a greater number of females under 19 participating than males in the same demographic.

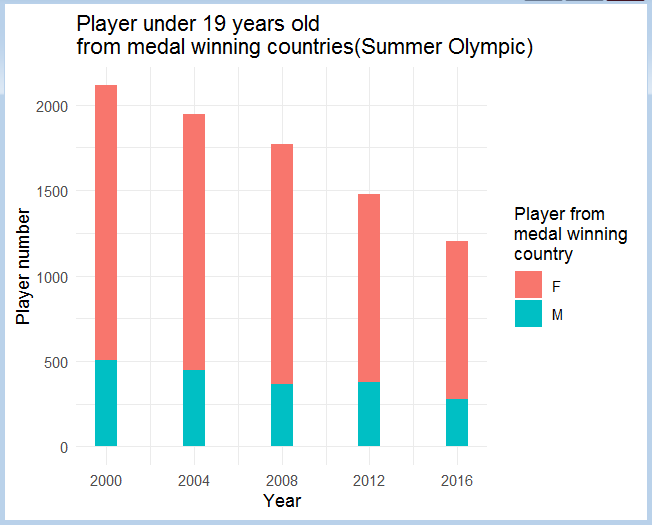


Figure 5. The number of all athletes from medal-winning countries in the Summer Olympic Games under age 19 by gender demographic.

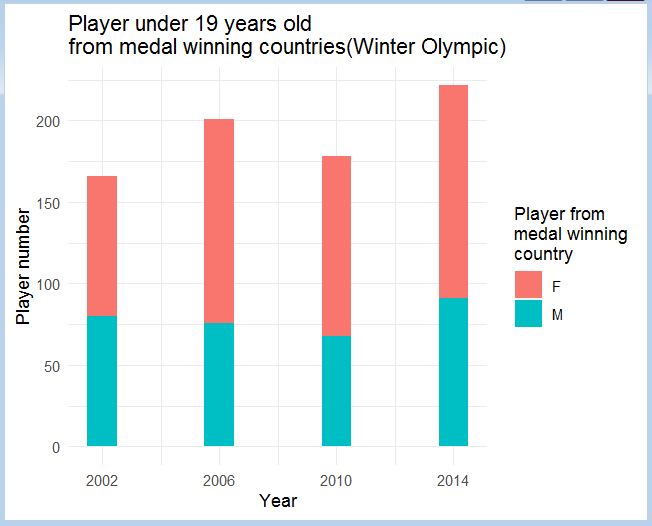


Figure 6. The number of all athletes from medal-winning countries in the Winter Olympic Games under age 19 by gender demographic.

To better understand the gender demographics for athletes in the United States, the bar charts in Figures 7 and 8 were created. Again, the trends represented in the U.S. are similar to those seen in the Figures 1 and 2, which depicts the gender demographics for athletes from all countries, with more female than male athletes. Although the exact reason for this gender disparity is not clear, it may indicate that the male athletes of this age group could be generally under-resourced, and therefore, good candidates for sponsorships.

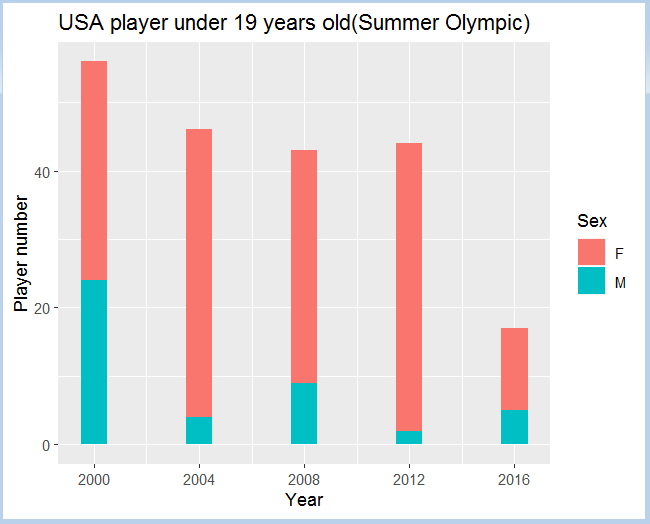


Figure 7. The number of American athletes in the Summer Olympic Games under age 19 by gender demographic.

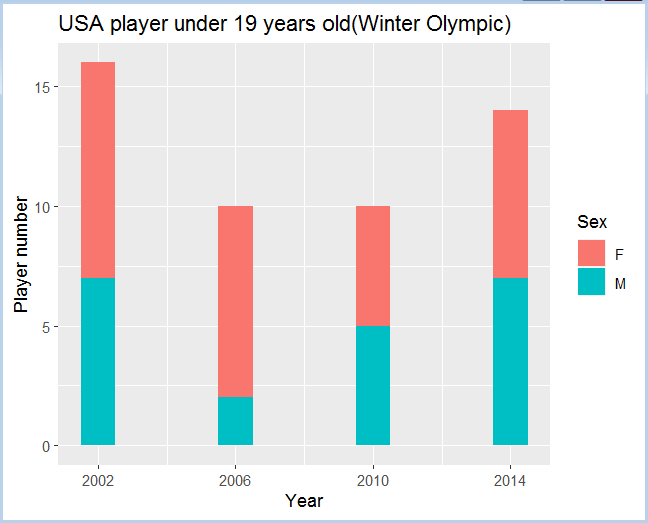


Figure 8. The number of American athletes in the Winter Olympic Games under age 19 by gender demographic.

## Results Using Tableau

Using Tableau, another set of visualizations was produced, and published online. A line chart depicts the count of athletes of all ages who participated in the Olympic Games from 2000-2016, with the gender demographics in each sport shown in corresponding colors for male and female. Animation capabilities highlight that there is a peak in the highly popular sports like athletics (track and field), gymnastics, swimming, and cross-country skiing show high peaks in the graph when filtered to show ages under 19. To depict the medal breakdown of gold, silver, and bronze by individual athlete, the dot graph was created. Each dot is colored to represent the color of the medal won by the individual, while the dot size corresponds to the number of medals won. The final dot graph shows the specific athletes who are recommended to be brand ambassadors. These are the gold-medal winners, with their medal count, country, and sport also visible by hovering.

In a different type of visualization, the map graphics display location data, another type of information vital to building a new marketing strategy. The first chloropleth map shows the average age of athletes in each country, adjustable by using the filter option. In the second map, it shows the total number of athletes under 19 per country, making it clear that Australia, China, Russia, and the United States are primarily locations which should be focused on for sponsorship campaigns, as they have the highest numbers of athletes under age 19.

# Future Work

This analysis has only begun to explore the many possibilities for alternative sponsorship campaigns. Future analyses could be designed around different groups of athletes. For instance, further development of campaigns could center specifically on sports popular in second/third tier countries. Keeping the context & region in mind, this study could be extended further by including data for a targeted age group along with the type of sport and the sex of the athlete.

# Conclusion

Through this analysis, a target demographic has been identified for building a brand ambassador program. Initial sponsorships for this demographic will ensure that athletes are provided with resources, likely increasing numbers of competitors in this age group and enhancing brand visibility. While both male and female groups will benefit from sponsorships, investing more in female athletes will likely benefit the company, because of their dominance in this age group. Focusing on under-represented groups will minimize the cost of sponsorships to the company [1], in turn creating greater ROI, but not negatively impact branding if emotion-lifting tactics are employed [2].In addition to focusing on female athletes, certain countries should be considered strongly when designing marketing campaigns, particularly those countries with a high concentration of athletes under 19: Australia, China, Russia, and the United States. Finally, recommendations for specific brand ambassadors, were created.

Opportunities for future analyses could de designed around different groups of athletes. For instance, further development of campaigns could center specifically on sports popular in second/third tier countries. Keeping the context & region in mind, this study could be extended further by including data for a targeted age group along with the type of sport and the sex of the athlete.

##### Acknowledgment

The authors would like to thank Professor Brandon Bean for his excellent instruction, feedback, and guidance throughout this project.

##### References

1. Jensen, J., Wakefield, L., Cobbs, J., & Turner, B. (2016). Forecasting sponsorship costs: Marketing intelligence in the athletic apparel industry. *Marketing Intelligence & Planning, 34*(2), 281-298.
2. Wang, R., & Kaplanidou, K. (2013). I want to buy more because I feel good: The effect of sport induced emotion on sponsorship. *International Journal of Sports Marketing & Sponsorship, 15*(1).
3. Mazodier, M., Corsi, A., & Quester, P. (2018). Advertisement typicality: A longitudinal experiment - can sponsors transfer the image of a sporting event to their brand? (what we know about sports sponsorships). *Journal of Advertising Research, 58*(3), 268.
4. rgriffin. (2018). 120 years of Olympic history: athletes and results [dataset]. Retrieved from <https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>
5. Ward, M., Grinstein, G., & Keim, D. (2015). *Interactive data visualization : Foundations, techniques, and applications* (Second ed.). Boca Raton: CRC Press.
6. Tufte, E. (2001). *The visual display of quantitative information* (2nd edition, ninth printing ed.). Cheshire, Connecticut: Graphics Press.