University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

Electrical & Computer Engineering and Computer Science

Electrical & Computer Engineering & Computer Science (ECECS)

TECHNICAL REPORT

SPRING 22

CONTENTS

Technical Report

Executive Summary

The goal of this project is to develop an optimal strategy for allocating sponsorship funds to Olympic teams, considering a holistic approach that incorporates historical medal performance, current prospects, and potential for future success. By leveraging data analytics and predictive modeling, the project aims to provide a data-driven framework for sponsors to allocate their funds strategically, maximizing the return on investment and aligning sponsorship efforts with teams that demonstrate a high likelihood of success.

Team Members:

Venkata Sai Mohan
Jitendra Prasad
Bikram Chand
Jefrina Jahangir

Olympics Data Analysis

Highlights of Project

Comprehensive Strategy: Integrating historical performance, current prospects, and future predictions for optimal sponsorship allocation.

Data-Driven Decision Making: Utilizing advanced analytics to inform objective sponsorship decisions. Empowering sponsors with actionable insights for strategic decision-making

Positive Impact: Aiming for mutually beneficial relationships and elevating the quality of Olympic sports sponsorships.

Submitted on:

12/6/2023

Abstract

This project introduces a pioneering methodology for optimizing Olympic sponsorship allocations, utilizing K-Means clustering and Google Data Studio. Beginning with a meticulous collection of historical performance data, K-Means clustering intelligently categorizes teams based on shared characteristics, providing insights into historical success, current capabilities, and future potential. Machine learning models, seamlessly integrated with K-Means insights, predict future team performance and conduct a scenario analysis on funding impacts. The optimization phase, driven by K-Means clustering, ensures effective sponsorship fund allocation considering budget constraints and agreements, maximizing medal count or achieving specified performance targets. Google Data Studio transforms complex insights into captivating visual representations, enhancing stakeholder understanding and creating a visually compelling pitch. Validation using out-of-sample data ensures model robustness, concluding with a comprehensive, data-driven strategy offering actionable recommendations for sponsors.

The unique synergy of K-Means clustering, and Google Data Studio visualization provides a holistic and visually intuitive approach to enhance the strategic sponsorship allocation process. The refined pitch perception is tailored for stakeholder engagement and sponsor enticement, ensuring resonance, and maximizing the return on investment in Olympic team sponsorships.

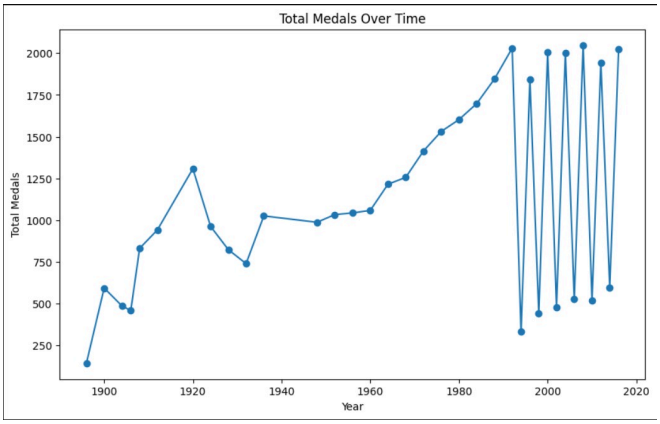
Methodology

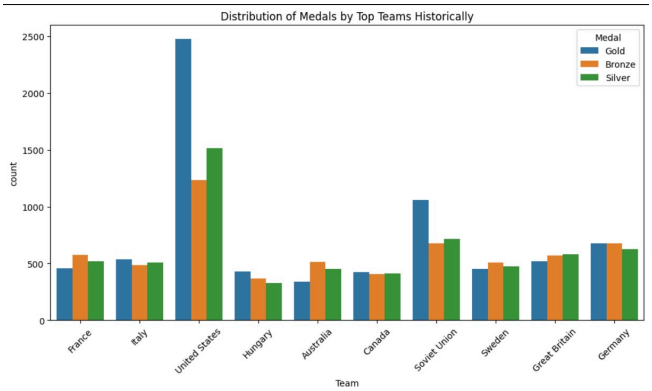
Data collection

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. The dataset contains 271116 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event. The columns are ID, Name, Sex, Age, Height, Weight, Team, NOC, Games, Year, Season, City, Sport, Event and Medal.

Data Cleaning and Preprocessing Steps:

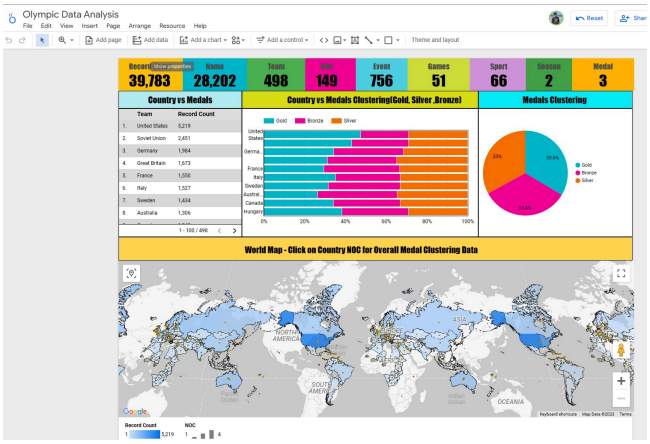
Identifying missing values in each column. Filled null values with mean (Age, Height, Weight) remove duplicate rows to ensure data integrity. Clustering of medals (Gold, Silver, Bronze) using one hot encoder.

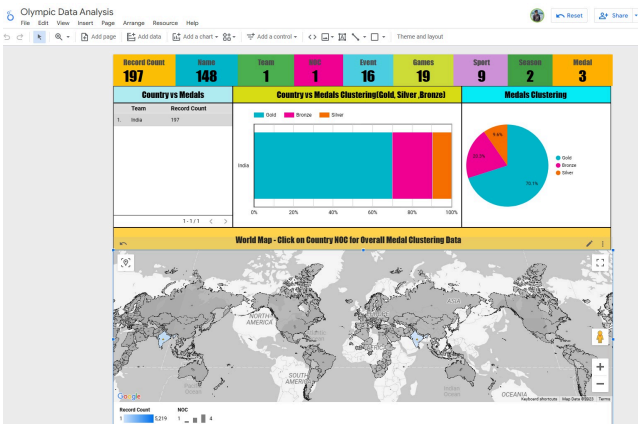




Data Visualization:

We used google data studio to visualize the data.





Modeling Process:

We utilized K-Means clustering as an unsupervised machine learning algorithm. We focused on the Total Number of Medals as the primary clustering feature. We grouped teams into clusters based on their Total Number of Medals and then associated each team with its respective cluster. We utilized historical cluster characteristics to predict sponsorship probability for new data points. The metrics we used are:

Low Sponsorship Probability: Teams in clusters with historically lower medal counts

Medium Sponsorship Probability: Teams in moderately successful clusters.

High Sponsorship Probability: Teams in clusters with a strong historical record of medal achievements.

The K-Means clustering model, driven by historical medal performances, provides a robust basis for predicting the sponsorship probability for each country. By categorizing teams into clusters and assigning sponsorship probabilities, stakeholders can make informed decisions on resource allocation, maximizing the impact of sponsorships based on historical trends. The iterative nature of the process ensures adaptability to evolving dynamics in Olympic sports.

Results Section

This approach provides a systematic way to leverage machine learning for making sponsorship decisions in the context of historical Olympic medal performance. We analyzed all the historical medal data and using k-means clustering our team finds that the “United States” are performing good in all recent years and they have “High Probability” to continue the same performance in future over all countries.

Conclusion

the Olympic data analysis project has unveiled intriguing insights into the historical trends and performances across various dimensions. The implications of this study extend to future Olympic Games, suggesting opportunities for enhancing athlete training, addressing gender disparities, and fostering innovation in sports. Overall, this analysis contributes to a deeper understanding of the dynamic landscape of the Olympics and sets the stage for further research and exploration in the realm of sports analytics.