

Project Title

Community Data Lab Project Report:

Partner Name

Developed by:

Sanam Prince David, Jeethendrinee Ellundhula

*Wichita State university*

In association with:

Yan Fujian, *Wichita State University*

December 7,2023

Contents

Summary 2

Introduction 3

Preparation 5

Analysis 8

Results 13

Recommendations 15

Conclusion 17

References 18

Summary

The study explores the time patterns of adult sports participation in Lawrence, Kansas. By locating demand peaks and analyzing seasonal trends, the study seeks to comprehend the temporal dynamics of adult sports participation.

The findings of the investigation might identify specific times of year when there is a significant demand for adult sports activities; these times could be associated with social, cultural, or seasonal factors. Important insights can also be obtained by determining and assessing the outside variables affecting the acceptance and demand for these activities. Factors such as weather patterns, societal mores, financial pressures, and even the availability of venues and resources can have a big impact on how popular adult sports become.

The results indicate that throughout the previous ten years, there has been a steady increase in the number of activities. There were 120 activities instead of 90 between 2012 and 2015. There was a minor decrease in the quantity of activities in 2016. Sports participation is highest in the summer (29.4%), followed by fall (33.5%), spring (24.3%), and winter (12.7%). Their popularity is explained by the pleasant weather in the summer and spring, while the popularity of indoor and cold-season sports is demonstrated by the significant participation in the fall and winter.

A variety of adult sports stakeholders, such as lawmakers, facility managers, and community organizers, may find informative knowledge of these relationships and factors to be helpful. All year long, adult sports events may become more appealing and have higher participation rates if these findings were applied to marketing, resource allocation, and strategic planning.

Introduction

Background of the Dataset:

Data on adult sports registration is largely derived from the analyzed data. Rather than being merely a set of statistics, this data illustrates the changing interests and commitments of adult consumers to leisure sports. Seeing, comprehending, and forecasting patterns in community sports participation is made possible by this sizable data collection. This data shows how adult consumers' interests and commitments to leisure sports are evolving, making it more than just a collection of numbers. This large data set offers a unique opportunity to see, understand, and predict patterns in community sports participation.

Community partner context:

Our partnership with Park and Rec is based on a shared acknowledgment of the essential impact of sports on enhancing community well-being. The commitment of Park and Rec to promote adult participation in recreational activities is evident through their continuous efforts to offer a diverse array of sports programs. This collaboration signifies a deliberate dedication to utilizing data as a tool to observe and enhance the overall health and vibrancy of community engagement.

Prompt from the partner:

Our community partner's thought-provoking prompts helped to define the parameters of our investigation. These questions functioned as a compass to direct our analysis. Our tasks included analyzing registration data to identify seasonal patterns, charting the enrollment rate's trajectory, and estimating the relative popularity of various sports over time. Our community partner was especially interested in finding out how participation fluctuated over the course of the year and seasons, how popular the different sports were, and whether there were any hidden trends that could guide the creation of new programs.

Data Description:

The dataset comprises 571 rows and 10 columns, with the following headers: Activity, Description, Year, Max Count, Enrolled, Unnamed: 5, Unnamed: 6, Unnamed: 7, Key, Unnamed:9. Notably, the columns Unnamed: 5, Unnamed: 6, and Unnamed: 7 exclusively contain null values. The dataset provides insights over a span of five years, specifically for the years 2012, 2013, 2014, 2015, and 2022.

Preparation

Data Loading

First, we’ve imported the necessary libraries and imported the data into a dataframe using pandas library. The data loading process was critical in ensuring we had enough information for a thorough examination.

Data Exploration

           We’ve then displayed the dataset as below and the output shows the first and last few rows of the dataset. From this we can also get the shape of the dataset which here is 571 rows and 10 columns. We can now understand that there are few columns with NaN values. NaN values in python are referred as empty or Null values. df.info() give us more detailed information about the data, including the data types, if there’re any null values. Looking at the size and info of our dataset, it is evident that there are all null values in 3 columns and there’s no significance with those columns and we’ll drop them.

A screenshot of a table

Description automatically generated

Figure First and last few rows of Dataset

A screenshot of a computer code

Description automatically generated

Figure df.info - Data types of Columns and Non Null Values

Unique values

Knowing the unique values in the columns of the dataset help us to better understand the data and make our analysis easier. This is one of the important steps in the data exploration stage.

Sub setting the Data

As we can see that there’re significant null values in the columns and the dataset appears to be in 2 parts explaining the max count and enrollment of the activities in first part and the sport in the second part respectively. So, we’ve split the dataset into 2 subsets df2 and df3 accordingly.

Data Cleaning

There’re Nan values in the df3 data frame, so we need to clean the data. We’ve understood that the nan values don’t represent any missing values, but the information is just limited to 27 rows.

Data Manipulation

Data manipulation is crucial for cleaning and transforming raw data into a format suitable for analysis, which leads to more accurate and insightful outcomes. We’ve followed some data manipulation steps to correctly map the sports to the activities.

Feature Engineering

Feature engineering enhances the performance of machine learning models by creating more informative features from raw data. We’ve created new columns (Season\_code, Game\_code, Day). Below is the snapshot of the adjusted dataframe till now.

A screenshot of a computer

Description automatically generated

Figure Sample of Data subset after data manipulation

Mapping

Mapping is useful for visualizing geographical data, allowing for the spatial analysis of trends and patterns that would be difficult to discern from raw data alone. We’ve mapped the season after a thorough understanding of the data. We’ve also done the activity and sport mapping so that we can start our analysis and below are first few rows of the final dataframe after all adjustments.

A screenshot of a calendar

Description automatically generated

Figure Data subset after all adjustments

Analysis

Popularity of sports over years

We examined how enrolment percentages varied over time, displaying the changes from 2012 to 2022 using bar charts. The long-term patterns and fluctuations were displayed visually.

A graph with a line and a dotted line

Description automatically generated

Figure Yearly Activity Distribution

The bar chart above reveals that 2014 experienced the highest enrollment in comparison to other years. The missing data between 2016 and 2021 could be attributed to external influences. Subsequent to 2014, there was a sudden decrease in average enrollment, potentially caused by factors such as a pandemic, diminished adult interest, or various other reasons.

Seasonal Enrollment PercentageTop of Form

To determine the seasonal changes in enrollment, we have produced a donut graphic. We may demonstrate the following using the chart below:

A chart with numbers and a number of different colors

Description automatically generated with medium confidence

Figure Seasonal Sport percentage

Of all the seasons, fall has the highest enrollment percentage (33.5%). Summer, on the other hand, comes in second with a 29.4% enrollment rate. With 24.3%, Spring is in third place. Winter comes in last with the lowest percentage of 12.7%. From the foregoing, enrolment in sports is directly impacted by weather conditions.

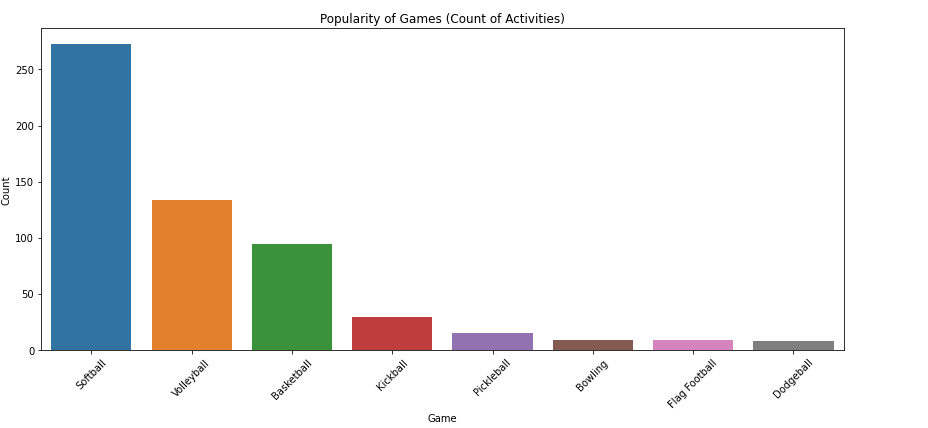
Overall Popularity of games

Figure Popularity of Games

To find out which game is the most popular among adults, we made a bar chart. The findings indicated that, out of all the games, softball is the most popular.

Popularity of sports over seasons

The bar chart shows enrollment numbers in the top 5 sports, with Summer School having the highest enrollment followed by Fall School, Spring School, Fall Weekend, and Winter Weekend.

**A graph of blue rectangular bars with white text

Description automatically generated**

Figure Enrollment of Top 5 Sports

Enrollment Rate by day

The bar chart depicts the average enrollment rate by day of the week, showing Wednesday as the peak day for enrollments, with noticeably lower rates on the weekends.

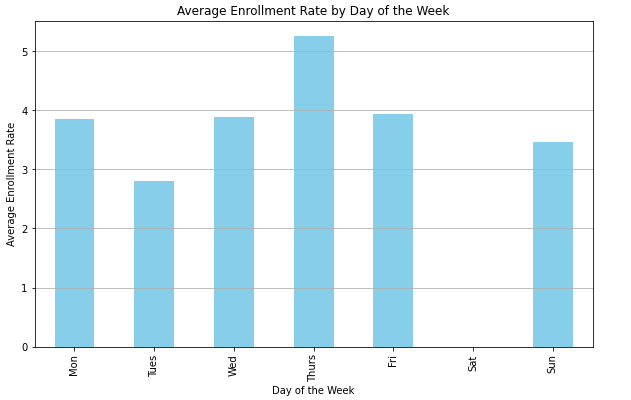


Figure Average Enrollment rate by Day of the week

Results

Popularity of Adult Sports

The bar chart shows that Summer Softball is the most popular sport with the highest count of activities, followed by Spring Softball and Summer Volleyball, indicating a seasonal trend in sports popularity.

A graph of different sports activities

Description automatically generated

Figure Popularity of Sports

The donut chart illustrates that Summer and Fall sports are the most popular, with summer activities making up nearly a third of all sports, indicating a higher engagement in sports activities during warmer seasons.

The bar chart shows a peak in the number of activities in 2014, with a noticeable decline by 2015 and further in 2022, suggesting a possible decrease in program offerings or participation over the years.

A graph of blue bars

Description automatically generated with medium confidence

Figure Yearly Activity Distribution

External Factors Influencing demand of Adult Sports:

Several external elements could impact the demand for adult sports enrollment. Weather patterns, the prevailing economic, climate, and cultural festivities all potentially sway participation rates. Additionally, the availability of alternative leisure pursuits and the success or popularity of nearby sports teams might also affect individuals' interest in signing up for sports activities. Each of these factors contributes to the complex landscape of adult sports engagement, where multiple influences intersect to shape the level of participation.

Optimal Schedule for Adult Sports by Season

The analysis of adult sports by season reveals distinct preferences and participation rates that influence optimal scheduling. During summer, sports like Pickleball and Softball exhibit full enrollments, suggesting they are highly popular and can be scheduled more frequently. In contrast, the winter schedule shows maximum participation in Pickleball, indicating a preference for indoor activities during colder months. Spring activities, such as Pickleball and Softball, also show high enrollment, whereas fall presents a balanced mix, with Kickball and Volleyball being the most enrolled activities.

The bar chart representing average enrollment by sport and season illustrates that certain sports peak in specific seasons, for instance, Basketball in winter and Kickball in fall. This trend suggests that scheduling should capitalize on seasonal popularity by offering more sessions of sports with higher enrollment in their respective peak seasons.

Overall, creating an optimal schedule for adult sports entails not only understanding seasonal preferences but also considering maximum capacity constraints to ensure that the demand for popular sports is met without over-scheduling less popular ones. This data-driven approach can significantly enhance participant satisfaction and optimize resource usage.

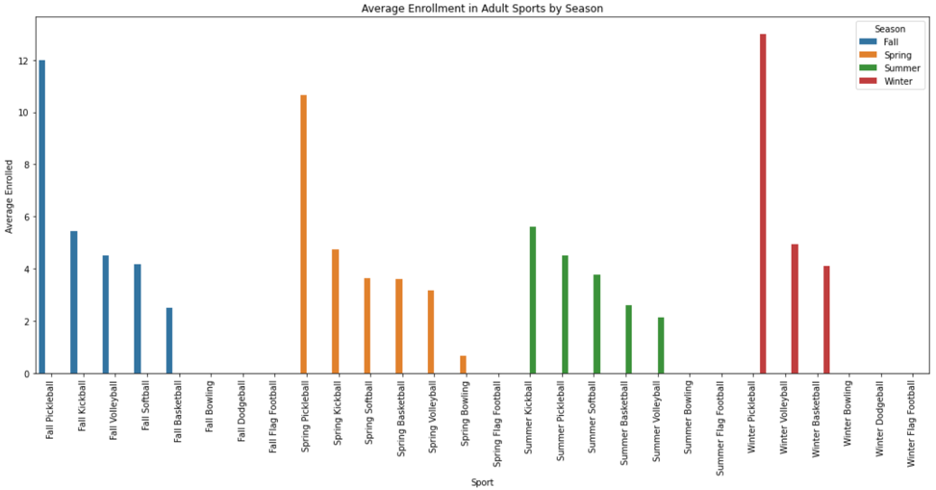


Figure Average enrollment in adult sports by season

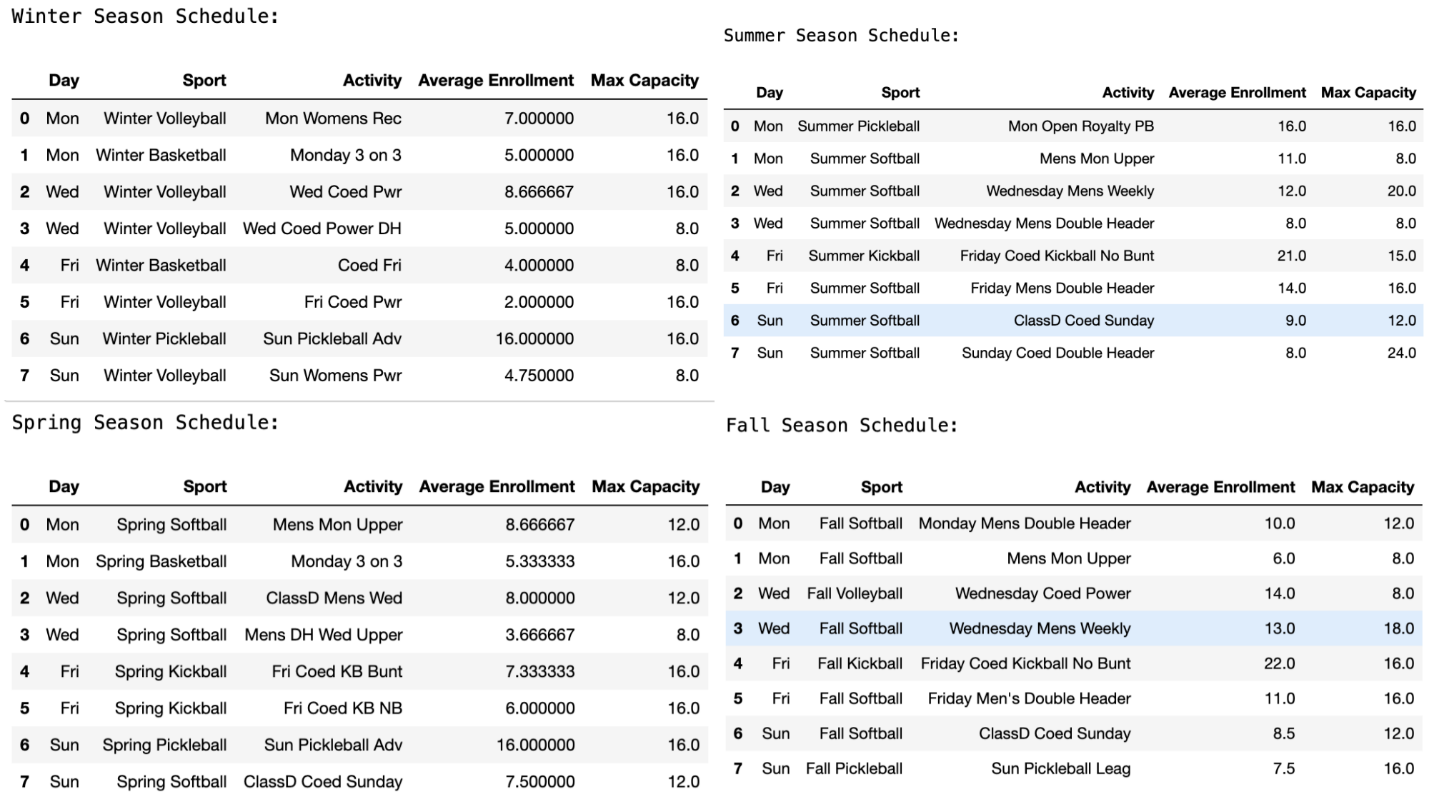


Figure Proposed optimal schedule.

Recommendations

Suggestions for Advancement

Examining sports sign-up data can yield critical insights into patterns of participation throughout the seasons and preferences for various sports. Yet, our project has faced challenges that limit our ability to address all the questions from our associates comprehensively. To navigate past these challenges and reinforce the integrity of our conclusions, we propose the subsequent steps for further investigation:

Collect Further Details

Delving into the personal attributes of registrants, such as age and geographical location, could prove instrumental in understanding participant demographics. Identifying these elements could elucidate preferences for specific sports among different demographic groups.

Assess Additional Influences

Consideration should be given to non-sporting influences such as meteorological conditions, major athletic events, or the economic capacity of individuals to afford sports activities. These influences could sway individuals' decisions to participate in sports.

Maintain Persistent Observation

By diligently tracking sports registration data over a longer timeline, we can better distinguish between sports that consistently engage interest and those that are temporarily popular.

Employ Sophisticated Analytic Techniques

Leveraging cutting-edge predictive analytics might offer foresight into future enrollment trends, which would be invaluable for forward-thinking planning.

Initiate Participant Surveys

Gaining direct insights from participants regarding their reasons for sports involvement or registration can reveal profound motivations.

Enhance Collaborative Efforts

A more thorough comprehension of the sports that capture communal interest can be achieved by building partnerships with local authorities, educational bodies, and sporting organizations.

Conclusion

The developed report can be used to understand and forecast the trends of sports participation. The dashboard can be useful for the stakeholders to quickly access the trends which helps in decision making. If provided, a sports scheduling system that is optimized would represent a data-driven strategy for enhancing adult sports programs.

References

Assumptions for external factors: <https://www.clearinghouseforsport.gov.au/research/ausplay/results>