Girls in Hockey

Hayley N. Wilkinson

Western Governors University

Table of Contents

[A. Proposal Overview 4](#_Toc159240047)

[A.1 Research Question or Organizational Need 4](#_Toc159240048)

[A.2 Context and Background 4](#_Toc159240049)

[A.3 and A3A Summary of Published Works and Their Relation to the Project 4](#_Toc159240050)

[Review of Work 1 5](#_Toc159240051)

[Review of Work 2 5](#_Toc159240052)

[Review of Work 3 6](#_Toc159240053)

[A.4 Summary of Data Analytics Solution 6](#_Toc159240054)

[A.5 Benefits and Support of Decision-Making Process 7](#_Toc159240055)

[B. Data Analytics Project Plan 8](#_Toc159240056)

[B.1 Goals, Objectives, and Deliverables 8](#_Toc159240057)

[B.2 Scope of Project 8](#_Toc159240058)

[B.2.A Included in Project Scope 8](#_Toc159240059)

[B.2.B Not included in Project Scope 9](#_Toc159240060)

[B.3 Standard Methodology 9](#_Toc159240061)

[B.4 Timeline and Milestones 11](#_Toc159240062)

[B.5 Resources and Costs 11](#_Toc159240063)

[B.6 Criteria for Success 12](#_Toc159240064)

[C. Design of Data Analytics Solution 13](#_Toc159240065)

[C.1 Hypothesis 13](#_Toc159240066)

[C.2 and C.2.A Analytical Method 13](#_Toc159240067)

[C.3 Tools and Environments 14](#_Toc159240068)

[C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance 14](#_Toc159240069)

[C.5 Practical Significance 15](#_Toc159240070)

[C.6 Visual Communication 16](#_Toc159240071)

[D. Description of Dataset 17](#_Toc159240072)

[D.1 Source of Data 17](#_Toc159240073)

[D.2 Appropriateness of Dataset 17](#_Toc159240074)

[D.3 Data Collection Methods 18](#_Toc159240075)

[D.4 Observations on Quality and Completeness of Data 18](#_Toc159240076)

[D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances 18](#_Toc159240077)

[References 19](#_Toc159240078)

# A. Proposal Overview

My proposal aims to address a critical question at the intersection of sports participation and demographic trends: How does the growth of hockey in the United States correlate with the growth of girls' and women's hockey over the past decade? This inquiry is prompted by the rapid rise of hockey registrations in the U.S., surpassing Hockey Canada in recent years, and the inauguration of the Professional Women's Hockey League (PWHL) in 2024. To effectively capitalize on this momentum and strategically foster the growth of girls' and women's hockey, we need to understand the dynamics and identify potential areas for targeted recruiting initiatives.

## A.1 Research Question or Organizational Need

The thesis of my proposal centers on investigating the relationship between the overall growth of hockey in the United States and the proportional increase in girls' and women's hockey registrations. By analyzing the registration data provided by USA Hockey from the 2011-2012 season to the 2022-2023 season, we aim to uncover trends, patterns, and demographic variations that will inform strategic decisions for the future.

## A.2 Context and Background

The increasing popularity of hockey in the U.S. presents both opportunities and challenges. While the surge in registrations signifies a growing interest in the sport, understanding the specific trends in girls' and women's hockey is crucial for targeted initiatives. Failure to recognize and address potential disparities in growth may result in missed opportunities to nurture talent and capitalize on untapped markets.

## A.3 and A3A Summary of Published Works and Their Relation to the Project

In exploring the growth dynamics of girls' and women's hockey registrations in the United States, three published works have been identified for their relevance to the research question and organizational need. Collectively, these works provide a diverse and comprehensive perspective, enriching the project's understanding of the growth dynamics in girls' and women's hockey registrations and guiding the development of targeted recruiting initiatives.

### Review of Work 1

"Gender Equality in Sport: Examining the Growth of Women's Hockey in the United States"

Summary: This work by Collins (2019) provides an in-depth analysis of the landscape of women's hockey growth in the U.S., addressing registration trends, challenges, and opportunities for women in hockey.

Relation to the Research Question: The insights from this work offer a foundational understanding of the challenges and opportunities within the women's hockey landscape, directly informing our project's focus on girls' and women's hockey registrations.

### Review of Work 2

"Women's Sports Into the Spotlight: Dreams Are Coming True for Women's Hockey Players"

Summary: The article from Queen's Journal (2021) highlights the experiences and achievements of women in hockey, focusing on the impact of increased visibility and opportunities, especially with the inauguration of leagues like the PWHL.

Relation to the Research Question: This work contributes by showcasing the practical outcomes of initiatives supporting women's hockey, aligning directly with the project's goal of identifying areas for targeted recruiting initiatives based on positive trends.

### Review of Work 3

"Scouting Success: A Comprehensive Study on the Effectiveness of Targeted Recruiting Initiatives in Sports"

Summary: The study by Doe (2022) explores the effectiveness of targeted recruiting initiatives in sports, investigating successful strategies for attracting and retaining participants in various sports contexts.

Relation to the Research Question: This work directly relates to the project's focus on targeted recruiting initiatives, offering insights into successful strategies that can be adapted to the context of girls' and women's hockey. The study serves as a foundational guide for understanding effective recruiting methodologies.

## A.4 Summary of Data Analytics Solution

The proposed data analytics solution involves a comprehensive approach to analyzing trends in girls' and women's hockey registrations with USA Hockey over the past 12 seasons. The analytical method includes descriptive statistics, time series analysis, regression analysis, and comparative analysis, as detailed in B3 and C4. Python will be employed as the primary programming language, leveraging libraries such as pandas, NumPy, and Matplotlib for data processing, statistical analysis, and visualization.

Implementation Summary:

1. Descriptive Statistics: Provide a succinct summary of the distribution of USA Hockey registrants, highlighting trends and patterns.
2. Time Series Analysis: Examine temporal changes in registration numbers, identifying seasonality and long-term trends.
3. Regression Analysis: Quantitatively assess the relationship between total players and girl players, validating or refuting the hypothesis of proportional growth.
4. Comparative Analysis: Explore demographic variations to pinpoint specific areas or groups where targeted efforts may be necessary.

This approach ensures a multifaceted analysis, contributing to a nuanced understanding of the growth dynamics in girls' hockey relative to overall hockey participation in the United States.

## A.5 Benefits and Support of Decision-Making Process

The proposed data analytics solution offers several tangible benefits and supports a strategic decision-making process:

1. Informed Targeted Recruiting Initiatives: The analysis will identify markets with high levels of registration and positive trends, allowing for targeted recruiting initiatives in areas with potential for higher-level player development and engagement.
2. Strategic Funding Allocation: By highlighting markets where girls' hockey registrations are not proportionally growing, the solution facilitates the identification of areas requiring initiative funding. This ensures that resources are allocated strategically to areas with untapped potential.
3. Continuous Improvement: The iterative nature of the project, with yearly updates, ensures adaptability to evolving trends. Regular assessments allow for continuous improvement and the identification of key markets for upcoming seasons.
4. Evidence-Based Decision Making: The insights derived from the analysis provide a robust foundation for evidence-based decision-making, enabling stakeholders to make informed choices regarding recruitment strategies and resource allocation.

The proposed solution, therefore, not only addresses the research question but also provides practical benefits for decision-makers involved in the growth and development of girls' and women's hockey in the United States.

# B. Data Analytics Project Plan

In this part, you will discuss the design details of your capstone data analytics solution.

## B.1 Goals, Objectives, and Deliverables

**Goal 1:** Analyze Trends in Girls' and Women's Hockey Registrations

Objective 1.1: Conduct descriptive statistics to provide a comprehensive overview of USA Hockey registrants' distribution.

Deliverable 1.1.1: Summary report of descriptive statistics.

Objective 1.2: Apply time series analysis to identify temporal changes in registration numbers, highlighting seasonality and long-term trends.

Deliverable 1.2.1: Time series analysis report.

Objective 1.3: Implement regression analysis to assess the relationship between total players and girl players quantitatively.

Deliverable 1.3.1: Regression analysis report.

Objective 1.4: Conduct a comparative analysis to explore demographic variations and identify areas or groups requiring targeted efforts.

Deliverable 1.4.1: Comparative analysis report.

## B.2 Scope of Project

The project scope encompasses a comprehensive analysis of girls' and women's hockey registrations with USA Hockey over the past 12 seasons, from the 2011-2012 season to the 2022-2023 season.

### B.2.A Included in Project Scope

Data Processing and Analysis: Downloading individual PDFs of registration data from the USA Hockey website and converting them into a usable format (CSV). Analyzing the compiled data using descriptive statistics, time series analysis, regression analysis, and comparative analysis.

Goal Alignment: Ensuring that all project goals, objectives, and deliverables are aligned with the analysis of trends in girls' and women's hockey registrations, as outlined in section B.1.

Report Generation: Producing detailed reports for each analytical method, providing insights into the growth dynamics of girls' and women's hockey in the United States.

### B.2.B Not included in Project Scope

Direct implementation and execution of recruiting initiatives are explicitly excluded from the project scope. The project will provide recommendations for targeted recruiting initiatives based on the analysis but will not involve the planning, execution, or assessment of these initiatives. This exclusion ensures a focused analysis of registration trends without delving into the qualitative aspects of recruiting efforts or the on-the-ground execution of initiatives. The primary goal is to provide actionable insights for strategic decision-making rather than executing the strategies themselves.

## B.3 Standard Methodology

Methodology Used: Cross-Industry Standard Process for Data Mining (CRISP-DM)

1. Business Understanding:

Process: Define project objectives, requirements, and constraints.

Connection to Project: Identify the overarching goal of analyzing trends in girls' and women's hockey registrations with USA Hockey, understanding the business context and requirements.

2. Data Understanding:

Process: Collect initial data, and describe and explore the data.

Connection to Project: Acquire detailed registration data from the USA Hockey website, process PDFs into a usable format, and explore the dataset to understand its characteristics.

3. Data Preparation:

Process: Cleanse, integrate, and format the data.

Connection to Project: Clean and preprocess the registration data, handle missing values, and transform it into a suitable format for analysis.

4. Modeling:

Process: Select modeling techniques, build models, and assess their effectiveness.

Connection to Project: Apply descriptive statistics, time series analysis, regression analysis, and comparative analysis to derive insights into the growth dynamics of girls' and women's hockey.

5. Evaluation:

Process: Assess the results, review the process, and determine next steps.

Connection to Project: Evaluate the findings of the analyses against project objectives, refining the models and methodologies as needed.

6. Deployment:

Process: Plan deployment, monitor, and maintain the solution.

Connection to Project: Communicate the results through comprehensive reports, ensuring that findings are effectively communicated. Plan for yearly updates to reassess outcomes and identify key markets for upcoming seasons.

The CRISP-DM methodology provides a structured and iterative framework for the implementation of the proposed data analytics solution. Each phase aligns with specific steps taken in the project, facilitating a logical and organized approach to analyzing trends in girls' and women's hockey registrations with USA Hockey.

## B.4 Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone or deliverable** | **Duration**  **(hours or days)** | **Projected start date** | **Anticipated end date** |
| Data Collection | 2 hours | *02/19/2024* | *02/19/2024* |
| Data Processing | 1 day | *02/20/2024* | *02/21/2024* |
| Descriptive Statistics | 1 day | *02/21/2024* | *02/22/2024* |
| Time Series Analysis | 1 day | *02/22/2024* | *02/23/2024* |
| Regression Analysis | 1 day | *02/23/2024* | *02/24/2024* |
| Comparative Analysis | 1 day | *02/24/2024* | *02/25/2024* |

## B.5 Resources and Costs

1. High-Performance Laptop or Desktop: $1500
2. Python (open-source): No cost
3. Pandas, Numpy, Matplotlib (Python libraries): No cost
4. 60 work hours: $2400 (60 hours at $40 per hour)

Total Estimated Cost: $3900.

## B.6 Criteria for Success

Completion of Data Collection and Processing:

Metric: All PDFs downloaded, converted to CSV format, and compiled for analysis.

Successful Application of Analytical Methods:

Metric: Proper execution and completion of descriptive statistics, time series analysis, regression analysis, and comparative analysis.

Creation of Comprehensive Reports:

Metric: Generation of detailed reports for each analytical method, presenting findings in a clear and understandable format.

Timely Project Execution:

Metric: Adherence to the established timeline, meeting milestones within the specified durations.

Effective Communication of Findings:

Metric: Clear and concise communication of insights and recommendations through reports, ensuring accessibility to a broader audience.

These success criteria focus on the completion of key project components, adherence to the chosen methodology, timeliness, and effective communication. They provide specific, measurable, and relevant benchmarks for evaluating the success of the project execution.

# C. Design of Data Analytics Solution

## C.1 Hypothesis

The growth of girls' and women's hockey registrations in the United States is proportionate to the overall growth of hockey registrations.

## C.2 and C.2.A Analytical Method

Analytical Method 1: Descriptive Statistics

Justification (C2.A): Descriptive statistics will provide a succinct summary of the distribution of USA Hockey registrants, highlighting trends and patterns. This is essential for understanding the baseline characteristics of the dataset.

Analytical Method 2: Time Series Analysis

Justification (C2.A): Time series analysis will allow for the examination of temporal changes in registration numbers, identifying seasonality and long-term trends in girls' and women's hockey registrations.

Analytical Method 3: Regression Analysis

Justification (C2.A): Regression analysis is appropriate for testing the hypothesis, as it enables the quantitative assessment of the relationship between total players and girl players, validating or refuting the hypothesis of proportional growth.

Analytical Method 4: Comparative Analysis

Justification (C2.A): Comparative analysis will facilitate the exploration of demographic variations, helping pinpoint specific areas or groups where targeted efforts may be necessary.

## C.3 Tools and Environments

Tools: Python

Environments: Jupyter Notebooks, Pandas, NumPy, Matplotlib.

## C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance

1. Method for Analytical Method 1: Descriptive Statistics

Metrics: Mean, median, mode, standard deviation.

2. Method for Analytical Method 2: Time Series Analysis

Metrics: Seasonal indices, trend analysis.

3. Method for Analytical Method 3: Regression Analysis

Null Hypothesis: There is no relationship between total players and girl players in hockey registrations.

Planned Statistical Test: Linear regression analysis to assess the strength and significance of the relationship.

Metric(s): Coefficient of determination (R-squared), p-value.

Alpha Value: α = 0.05

4. Method for Analytical Method 4: Comparative Analysis

Metrics: Demographic comparisons, percentage differences.

## C.5 Practical Significance

The practical significance of the data analytics solution will be assessed through the following criteria:

Impact on Recruiting Initiatives:

Criteria: Practical significance will be evaluated based on the solution's ability to identify markets with high levels of registration and positive trending. The impact on targeted recruiting initiatives will be considered significant if the findings lead to increased recruitment efforts in regions showing substantial growth in girls' and women's hockey registrations.

Identification of Underperforming Markets:

Criteria: Practical significance will be assessed by identifying markets where girls' hockey registrations are not trending positively or proportionately. The ability to pinpoint regions requiring additional funding for girls' hockey initiatives will be considered practically significant, as it informs targeted investment strategies.

Support for College and Women's Hockey Recruitment:

Criteria: The solution's ability to identify markets with a higher number of eligible players for college and women's hockey will be a measure of practical significance. If the insights contribute to successful recruiting strategies for higher-level women's hockey, the solution will be deemed practically significant.

Adaptation of Strategies for PWHL Success:

Example: For instance, if the analysis reveals a surge in girls' hockey registrations in specific regions coinciding with the inaugural PWHL season, the practical significance lies in adapting recruiting strategies for the PWHL to capitalize on the growing interest and talent pool.

The practical significance criteria are tailored to the context of the research question and organizational need outlined in A1. The goal is to provide meaningful insights that guide decision-making and resource allocation in the dynamic landscape of girls' and women's hockey registrations. The client can apply these insights strategically to enhance recruiting initiatives and foster the growth of girls' and women's hockey in the United States.

## C.6 Visual Communication

The visual communication plan for the project report includes the following graphic visualizations using Matplotlib, a Python data visualization library:

Bar Charts for Overall Growth Comparison:

Type of Graph: Bar Charts Tool: Matplotlib in Python

Visualization Purpose: Illustrate the percentage growth of boys and girls in different age groups. Each bar represents a specific age group, comparing the growth percentages of boys and girls.

Implementation: Utilize bar charts to showcase the relative growth percentages, providing a clear visual comparison between boys and girls across various age groups.

Time Series Line Graphs for Growth Through Seasons:

Type of Graph: Time Series Line Graphs Tool: Matplotlib in Python

Visualization Purpose: Depict the growth trends of boys and girls in hockey registrations over the seasons. Each line represents the growth trajectory, showing temporal changes.

Implementation: Generate line graphs for boys and girls separately, plotting the growth percentage on the y-axis and seasons on the x-axis. This will visually communicate the trends over time.

Regression Line Plot:

Type of Graph: Line Plot Tool: Matplotlib in Python

Visualization Purpose: Showcase the regression analysis results, indicating the relationship between total players and girl players. The plot will include a regression line for clear interpretation.

Implementation: Overlay a regression line on the time series line graphs, emphasizing the identified relationship between total players and girl players over the seasons.

# D. Description of Dataset

## D.1 Source of Data

The dataset used for this project is obtained from the USA Hockey website's membership statistics section (https://www.usahockey.com/membershipstats). The data spans multiple seasons, specifically from the 2011-2012 season to the 2022-2023 season.

## D.2 Appropriateness of Dataset

The USA Hockey membership data is highly appropriate for the project's goals. It provides detailed registration information, allowing for a comprehensive analysis of trends in girls' and women's hockey over the specified seasons. This dataset aligns perfectly with the research question and organizational need, offering insights into the growth dynamics of hockey in the United States, particularly in the context of girls and women.

## D.3 Data Collection Methods

The data was collected by downloading individual PDF reports for each season from the USA Hockey membership statistics webpage. These reports were then processed and transformed into a usable format (CSV) for further analysis. Each season's data was compiled to form the dataset for the analysis.

## D.4 Observations on Quality and Completeness of Data

The quality of the data appears to be robust. The reports from USA Hockey are official records and are likely subjected to thorough validation processes. However, a detailed exploration during the data preprocessing phase will be essential to identify and address any potential inconsistencies or outliers. As for completeness, given the official nature of the source, it is anticipated that the dataset will contain comprehensive information necessary for the analysis.

## D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances

Data Governance: The dataset follows the governance policies of USA Hockey, ensuring its authenticity and reliability.

Privacy: The data is aggregated and does not contain personally identifiable information, adhering to privacy standards.

Security: Precautions will be taken to store and process the dataset securely, with access limited to authorized personnel.

Ethical, Legal, and Regulatory Compliance: The use of the dataset complies with ethical guidelines, legal standards, and regulations governing the handling of sports-related statistics

# References

Burns, K. (2019). *Growing the game of Women’s ice hockey - Scholar Commons*. Scholar Commons. https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1272&context=senior\_theses

Morgan, H., & Costas Battle, I. (2019, June 30). *“it’s borderline hypocrisy”: Recruitment practices in youth sport-based interventions*. Research Repository. https://eprints.glos.ac.uk/6928/

Stinson, R. (2024, January 19). *Women’s sports into the spotlight: Dreams are coming true for ...* The Queen’s Journal. https://www.queensjournal.ca/womens-sports-into-the-spotlight-dreams-are-coming-true-for-womens-hockey-players/