

***MAKE-UP OF A
STANLEY CUP
WINNER***



EXECUTIVE SUMMARY

NHL PLAYER & TEAM DATA 2008 / 09 – 2021 / 22

The NHL or National Hockey League is the highest level of professional hockey in North America.

These three data sets represent the player biography data (age, height, weight, birth country, etc) for every NHL player from 2008 until 2022, as well as the player and goalie stats for playoff games.

Money Puck has made available their data for public use so long as they are credited as the source.

These data sources were chosen as they include the biographical and statistical details required to answer the questions I have in mind. The data sources are connected and have common columns in which to merge.

Data Sources

<https://www.money puck.com/data.htm>

<https://www.kaggle.com/datasets/mesadowski/nhl-player-and-team-data-20089202122>

<https://www.hockeyreference.com/>

DATA PROFILE

LIMITATIONS AND ETHICS

- At this point a limitation is that players are given country but not state or province.
- We include player names, nationality, birth place, height and weight; these are essential to the study.
- Data was collected by a number of statisticians, this could affect the quality of data.
- Game Score was calculated for Skaters but not for Goalies.
- Players may any number of positions, the one chosen for this project is their primary position.

DATA PROFILE - ALLPLAYERSLOOKUP (ORIGINAL DATA SET)

■ Original Data Set – 2,914 x 11

Column	Player ID	Name	Position	Team	Birthdate	Weight (lbs)	Height (ft.in)	Nationality	Shoots Catches	Primary Number	Primary Position
Min	8,445,550	N/A	N/A	N/A	N/A	140	N/A	N/A	N/A	1	N/A
Max	8,483,570	N/A	N/A	N/A	N/A	265	N/A	N/A	N/A	98	N/A
Mean	8,474,080	N/A	N/A	N/A	N/A	199.44	N/A	N/A	N/A	36.78	N/A
Count	2,914	2,914	2,914	2,914	2,829	2,829	2,829	2,829	2,829	2,829	2,829
Missing Cells	0	0	0	0	85	85	85	85	85	85	85

DATA PROFILE - ALLPLAYERSLOOKUP (ORIGINAL DATA SET)

Column	Column Description	Data Type	Time Variant
Player ID	Each player is assigned an ID number	Qualitative/Ordinal	No
Name	Player Name	Qualitative/Nominal	No
Position	Their primary position	Qualitative/Nominal	Yes
Team	Team they played for	Qualitative/Nominal	Yes
BirthDate	Year-Month-Day	Qualitative/Ordinal	No
Weight	Weight in Pounds.	Quantitative/Continuous	Yes
Height	Height in Feet & Inches	Quantitative/Continuous	Yes
Nationality	Birth Country (abbreviated)	Qualitative/Nominal	No
Shoots / Catches	Which hand do they shoot (player) or catch (goalie)	Qualitative/Nominal	No
Primary Number	Their primary jersey number	Qualitative/Nominal	Yes
Primary Position	Their primary position	Qualitative/Nominal	Yes

DATA PROFILE

DATA CLEANING & CONSISTENCY CHECKS

Original Data Set: "allplayerslookup"

Prepared Data Set: "playerbio"

- Original Data: 2,914 x 11
- Cleaned Data: 2,914 x 13
- Missing Data found: <https://www.hockeyreference.com/>
- **Cleaning process included:**
 - Finding missing data
- Removed Primary Position (redundant as it's same as Position)
 - Separating "Name" into "First Name" & "Last Name"
 - Changing Height from ft.in to inches
 - Expanding "Birthdate" into "Year", "Month", "Day"
 - Using Excel Filters checked data.

DATA PROFILE - PLAYERBIO (CLEAN DATA SET)

■ Clean Data Set 2,914 x 13

Column	Player ID	First Name	Last Name	Position	Team	Birth Year	Birth Month	Birth Day	Weight (lbs)	Height (inches)	Nationality	Shoots Catches	Primary Number
Min	8,445,550	N/A	N/A	N/A	N/A	1962	1	1	140	64	N/A	N/A	1
Max	8,483,570	N/A	N/A	N/A	N/A	2003	12	31	265	81	N/A	N/A	98
Mean	8,474,080	N/A	N/A	N/A	N/A	1988.88	5.88	15.58	199.36	73.15	N/A	N/A	36.91
Count	2,914	2,914	2,914	2,914	2,914	2,914	2,914	2,914	2,914	2,914	2,914	2,914	2,914
Missing Cells	0	0	0	0	0	0	0	0	0	0	0	0	0

DATA PROFILE - PLAYERBIO (CLEAN DATA SET)

Column	Column Description	Data Type	Time Variant
Player ID	Each player is assigned an ID number	Qualitative/Ordinal	No
First Name	Player First Name	Qualitative/Nominal	No
Last Name	Player Last Name	Qualitative/Nominal	No
Position	Their primary position	Qualitative/Nominal	Yes
Team	Team they played for	Qualitative/Nominal	Yes
Birth Year	Birth Year	Qualitative/Nominal	No
Birth Month	Birth Month	Qualitative/Nominal	No
Birth Day	Birth Day	Qualitative/Nominal	No
Weight	Weight in Pounds.	Quantitative/Continuous	Yes
Height	Height in Inches	Quantitative/Continuous	Yes
Nationality	Birth Country (abbreviated)	Qualitative/Nominal	No
Shoots / Catches	Which hand do they shoot (player) or catch (goalie)	Qualitative/Nominal	No
Primary Number	Their primary jersey number	Qualitative/Nominal	Yes

DATA PROFILE - SKATER PLAYOFFS (CLEAN DATA SET)

■ Clean DataSet—4,932 x 9

Column	Player ID	Season	First Name	Last Name	Team	Position	Games Played	Ice Time	Game Score
Min	8,445,550	2008	N/A	N/A	N/A	N/A	1	49	-4.63
Max	8,482,124	2021	N/A	N/A	N/A	N/A	27	44,853	45.26
Mean	8,472,948.8	2014.64	N/A	N/A	N/A	N/A	9.26	9,480.53	4.15
Count	4,932	4,932	4,932	4,932	4,932	4,932	4,932	4,932	4,932
Missing Cells	0	0	0	0	0	0	0	0	0

DATA PROFILE - SKATER PLAYOFFS

Year	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
Original Data	1,665 x 154	1,710 x 154	1,695 x 154	1,740 x 154	1,745 x 154	1,720 x 154	1,695 x 154
Clean Data	333 x 9	342 x 9	339 x 9	348 x 9	349 x 9	344 x 9	339 x 9

Year	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22
Original Data	1,700 x 154	1,720 x 154	1,665 x 154	1,665 x 154	2,580 x 154	1,685 x 154	1,675 x 154
Clean Data	340 x 9	344 x 9	333 x 9	333 x 9	516 x 9	337 x 9	335 x 9

DATA CLEANING & CONSISTENCY CHECKS

- Cleaning process included:
 - Combined Data Sets from Different Years into one complete Skater Set
 - Separating "Name" into "First Name" & "Last Name"
- Removed unnecessary columns and rows. Each skater had five entries per year, only one was necessary.
 - Through Excel Filters checked data; no missing data.
 - Final Count: skaters clean (4,932 x 9)

DATA PROFILE - SKATER PLAYOFFS (CLEAN DATA SET)

Column	Column Description	Data Type	Time Variant
Player ID	Player ID assigned to each player.	Qualitative/Ordinal	No
Season	Which season are these stats for.	Quantitative/Discrete	Yes
First Name	Player's First Name	Qualitative/Nominal	No
Last Name	Player's Last Name	Qualitative/Nominal	No
Team	Player's Team	Qualitative/Nominal	Yes
Position	Player's Position	Qualitative/Nominal	Yes
Games Played	Amount of games played that season.	Quantitative/Discrete	Yes
Ice Time	Total ice time (in minutes)	Quantitative/Continuous	Yes
Game Score	Score calculated for each player (created by Dom Luszczyzyn)	Quantitative/Continuous	Yes

DATA PROFILE - GOALIE PLAYOFFS (CLEAN DATA SET)

■ Clean Data Set 375 x 8

Column	Player ID	Season	First Name	Last Name	Team	Position	Games	Ice Time
Min	8,445,550	2008	N/A	N/A	N/A	N/A	1	67
Max	8,483,570	2021	N/A	N/A	N/A	N/A	26	103,392
Mean	8,474,080	2014.76	N/A	N/A	N/A	N/A	7.25	25,212.58
Count	375	375	375	375	375	375	375	375
Missing Cells	0	0	0	0	0	0	0	0

DATA PROFILE - GOALIE PLAYOFFS

Year	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
Original Data	125 x 36	135 x 36	125 x 36	115 x 36	115 x 36	140 x 36	130 x 36
Clean Data	25 x 8	27 x 8	25 x 8	23 x 8	23 x 8	28 x 8	26 x 8

Year	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22
Original Data	140 x 36	115 x 36	145 x 36	110 x 36	210 x 36	120 x 36	150 x 36
Clean Data	28 x 8	23 x 8	29 x 8	22 x 8	42 x 8	24 x 8	30 x 8

DATA CLEANING & CONSISTENCY CHECKS

- Cleaning process included:
 - Combined Data Sets from Different Years into one complete Goalie Set
 - Separating "Name" into "First Name" & "Last Name"
- Removed unnecessary columns and rows. Each goalie had five entries per year, only one was necessary.
 - Through Excel Filters checked data; no missing data.
 - Final Count: goalies clean (375 x 8)

DATA PROFILE - GOALIE PLAYOFFS (CLEAN DATA SET)

Column	Column Description	Data Type	Time Variant
Player ID	Player ID assigned to each player.	Qualitative/Ordinal	No
Season	Which season are these stats for.	Quantitative/Discrete	Yes
First Name	Player's First Name	Qualitative/Nominal	No
Last Name	Player's Last Name	Qualitative/Nominal	No
Team	Player's Team	Qualitative/Nominal	Yes
Position	Player's Position	Qualitative/Nominal	Yes
Games Played	Amount of games played that season.	Quantitative/Discrete	Yes
Ice Time	Total ice time (in minutes)	Quantitative/Continuous	Yes

QUESTIONS TO EXPLORE

- What countrymen typically do well in the playoffs?
 - Does it help to have more than one person from the same country?
- Do specific countries produce certain positions?
- Is there a specific age that tend to do well?
- Is it better to have a bigger or smaller team?
- Are there any surprises when it comes to a Stanley Cup winning team?