Final Project Presentation

Team: Nexus Nucleus

SCTP - Junior Data Engineer 2024

9 October 2024

ABOUT US

ASMR Research

ASMR Research formed by a group of data enthusiasts specializing in Data Engineering and Analytics.

- **A**sher
- **S**erchen
- **M**in
- **R**ichard

Previous portfolio includes :

Analysis on Singapore Bicycle Parking Locations

Interim Project

Analysis on Singapore's Bicycle Parking Spaces

Objective and Findings

From the 3 estates (Jurong, Woodland, Tampines)…

1 Most HDB bicycle parking lots

Answer : Tampines

2 Most number of sheltered bicycle lots

Answer: Jurong

3 Most bus-stops with bicycle parking facilities

Answer : Tampines

Method/Approach

API & CSV

Extract

Transform

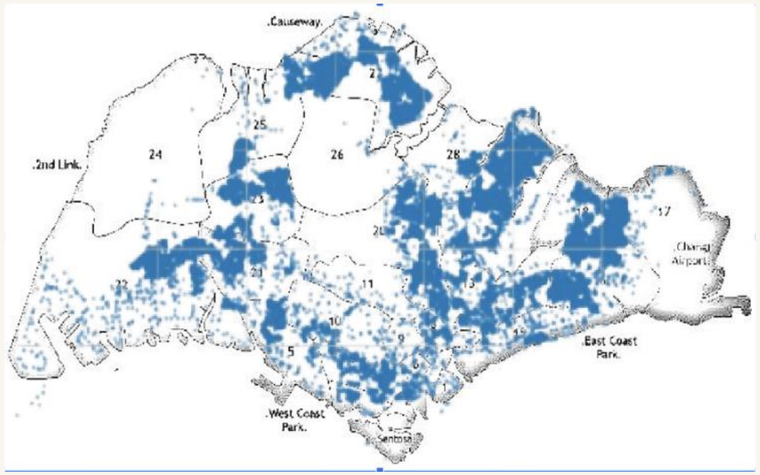
Load

PostgreSQL

Tools & Libraries

Jupyter Notebook
PostgreSQL
Matplotlib.pyplot
SQLalchemy
Requests
Pandas
NumPy

Visualization



SCTP - Junior Data Engineer 2024

NHL Game Data Analysis

TABLE OF CONTENTS

INTRODUCTION

Background

In late 2020, an undisclosed game studio started developing their next NHL computer game.

They had access to detailed NHL data, including play-by-play information with x, y coordinates which they planned to use in their game development

However…..

INTRODUCTION

Problem Statement

The game studio lacked the expertise to analyze the NHL dataset.

And thus, they approached **ASMR Research** to collaborate with them on data analysis and leverage data-driven insights to enhance gameplay mechanics and player dynamics

OBJECTIVES

ETL Pipeline

Effective, scalable ETL pipeline for NHL data

- Ingest
- Transform
- Load

X-Y Coordinates

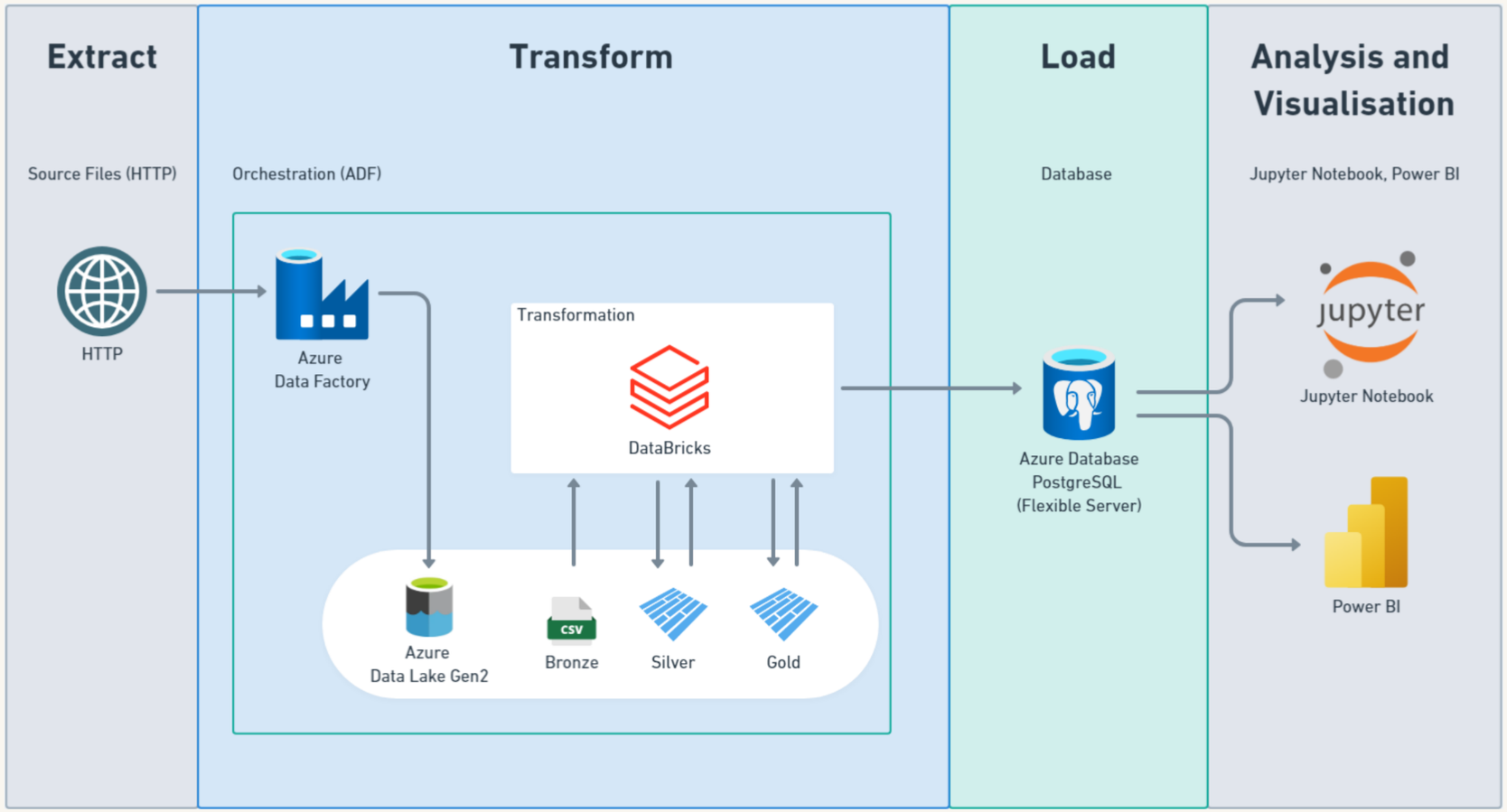
To propose the best attack angles, distances, and shot types with the highest probability of scoring goals in the rink

NPC Dream Team

To identify and propose a dream team of 23 non-playable characters (NPCs), comprising 12 forwards, 8 defensemen, and 3 goalies.

METHODOLOGY

Pipeline Architecture



Services Used

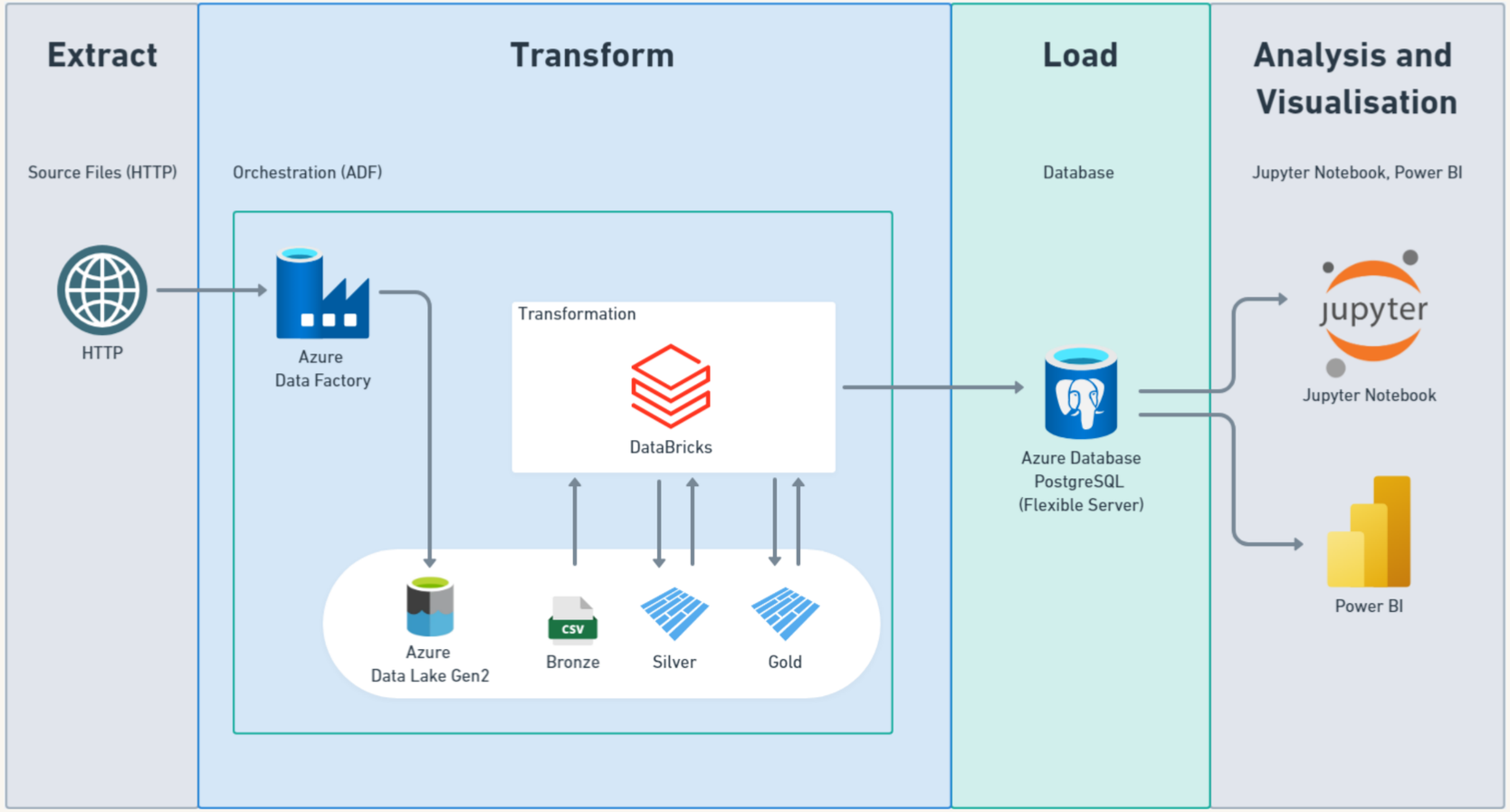
- Azure Data Factory (ADF)
- Azure Data Lake Storage Gen2 (ADLS Gen2)
- Azure Databricks
- Azure Database PostgreSQL

Medallion Architecture

- Bronze: Raw, unprocessed
- Silver: Cleansed, conformed
- Gold: Curated, ready for business

METHODOLOGY

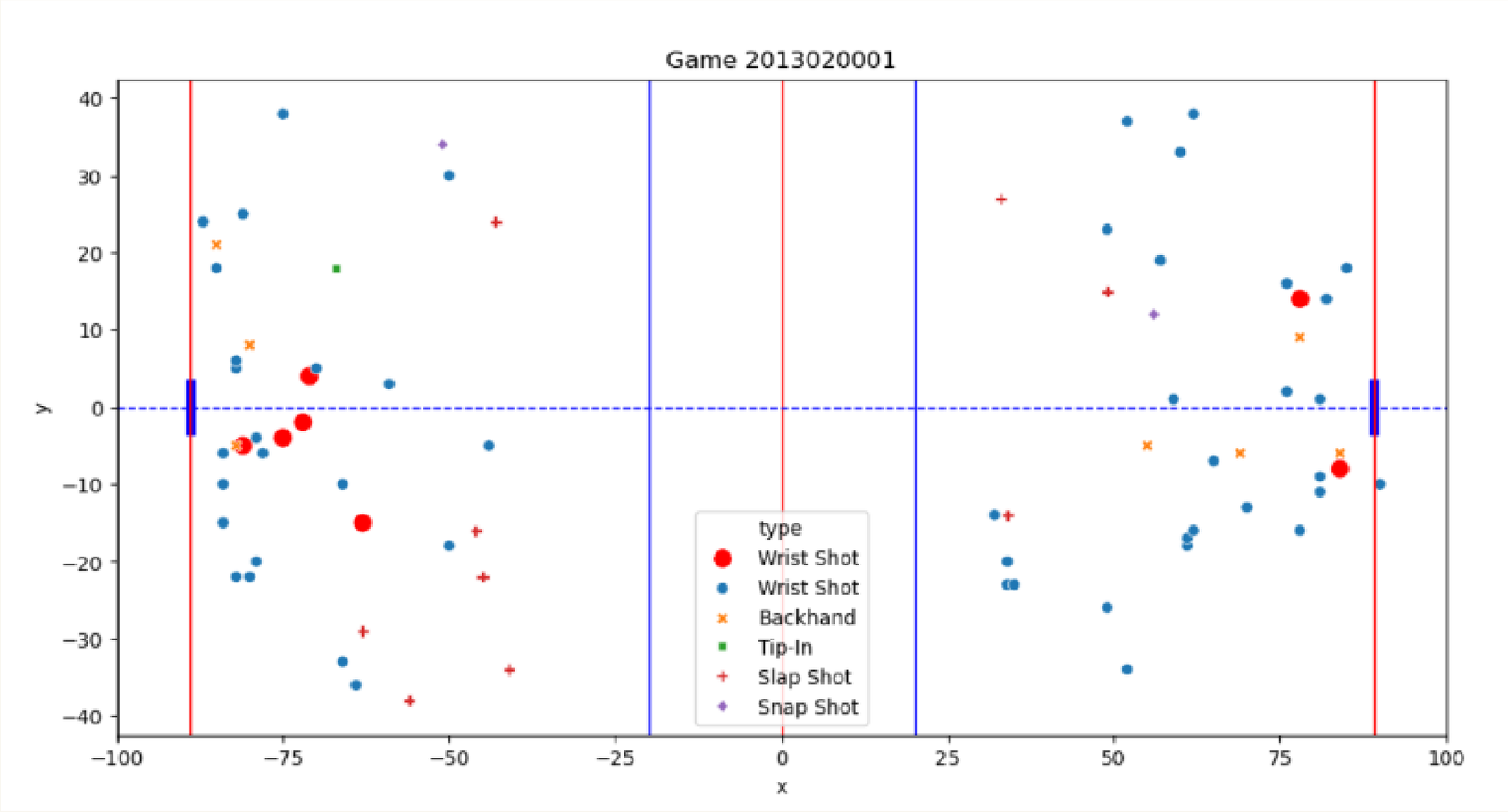
Pipeline Architecture



Approach

1. **Extraction**
   - Azure Data Factory
2. **Transformation**
   - Azure Databricks
3. **Loading**
   - Azure Databricks
   - Azure Database PostgreSQL
4. **Analysis**
   - Jupyter Notebook
   - Power Bi
   - SQL Queries

NHL 101



Game

60 minutes long: 3 periods of 20 minutes each.

Team

23 players in a team
6 players consisting
3 attackers, 2 defenders, 1 goalie

Rink

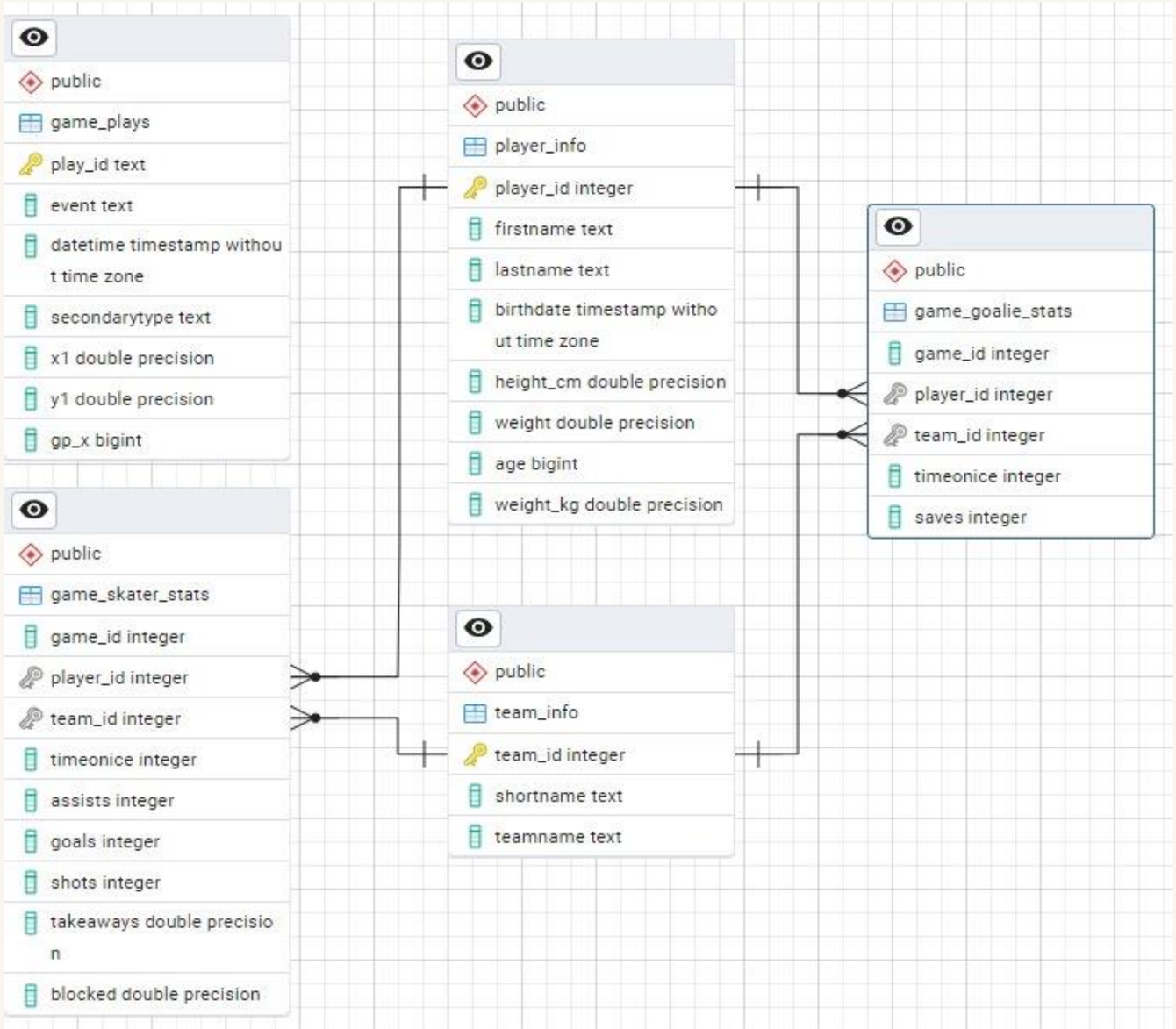
200 ft x 85 ft
Centre: 0,0
Goal -89, 0 and 89 , 0
Left/Right wing

Shot Types

Wrist, Slap, Snap, Tip-In, Backhand, Wrap around,

ANALYSIS

Database ERD



(A) X-Y Analysis

(B) Dream Team Analysis

5 Tables

- game_plays
- game_skater_stats
- game_goalie stats
- player_info
- team_info

Uploaded from Gold layer for analysis in PostgreSQL

ANALYSIS: X-Y

PROBLEM STATEMENT

Which positions on the rink are most likely to result in a goal?

What types of shots are most effective based on distance and angle from the goal post?

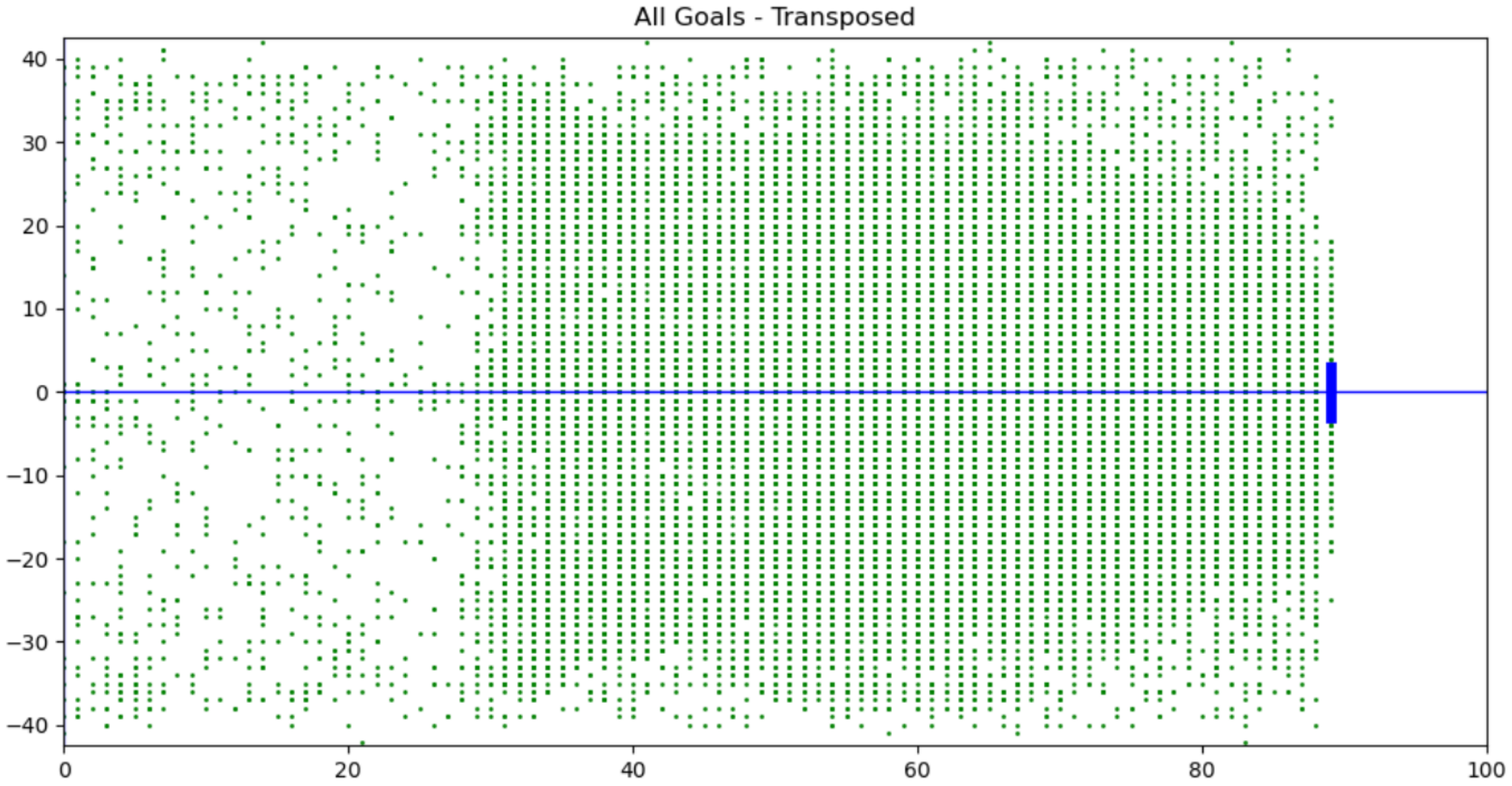


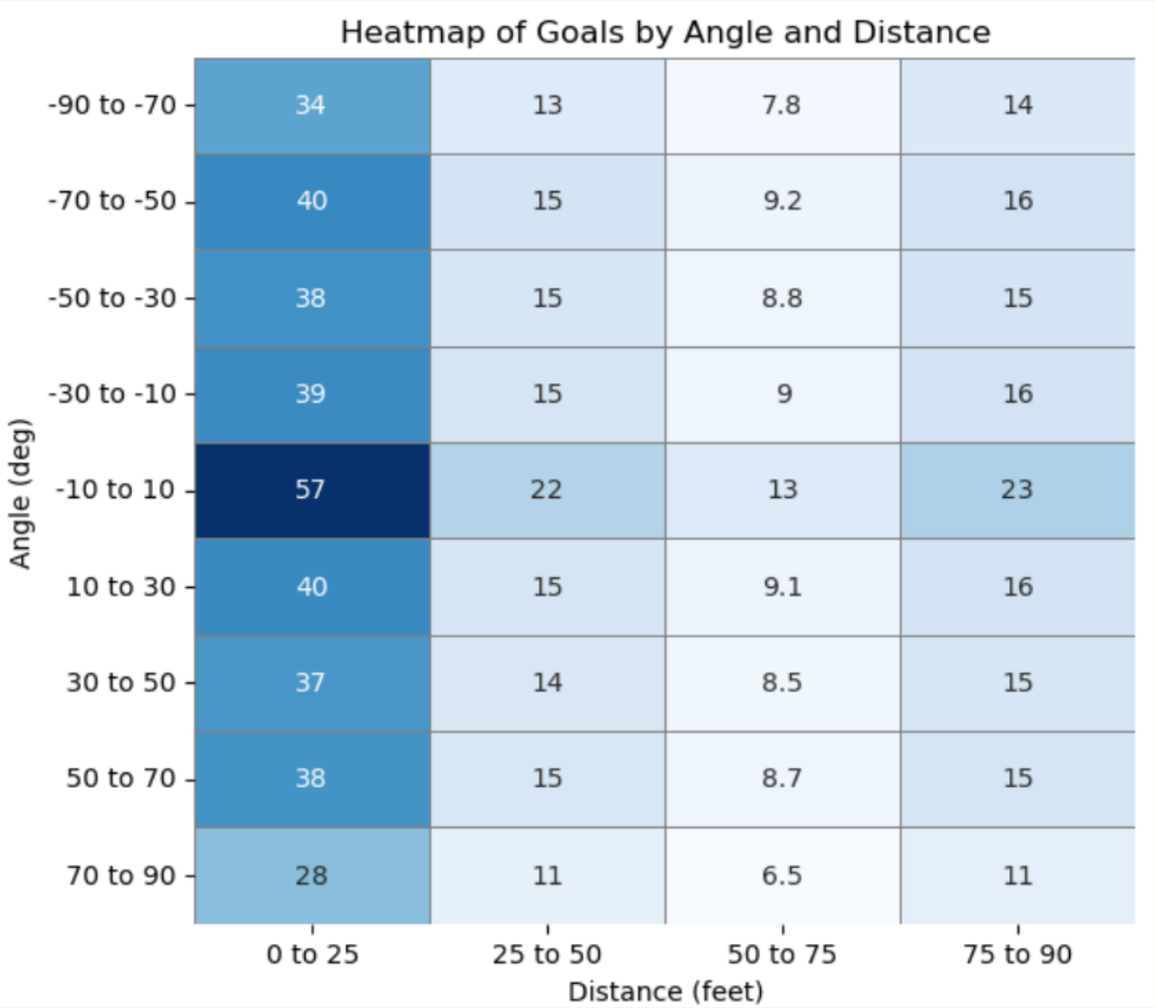
Table Transformation

game_plays Table:

- Transpose x-y to one side of the rink
- Create Goal position at x=89
- Assume all shot at goals are within x=0 to 89
- Create Angle, Distance, Angle-Group/Distance-Group
- Formulate composite index for goal probability

ANALYSIS: X-Y

FINDINGS ON GOAL DISTRIBUTION



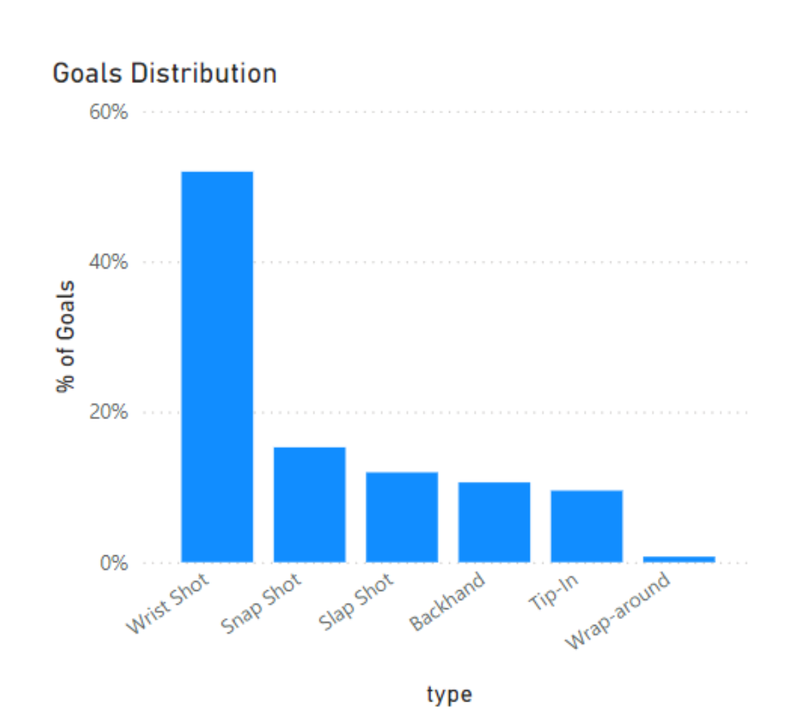
Goal Distance and Angle

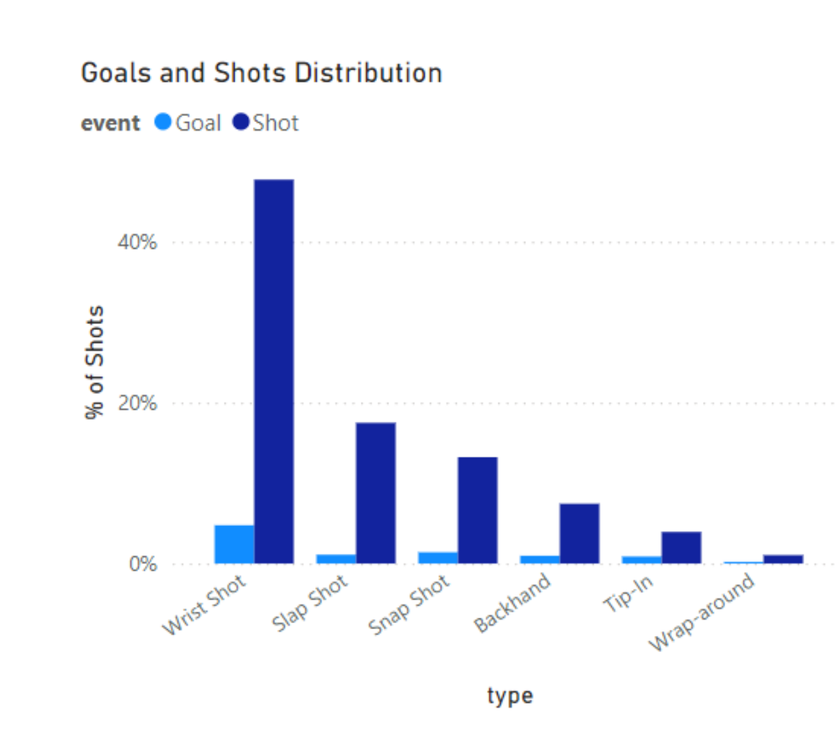
Higher chance of goal at closer range and narrower angles

Generally, distance to goal has more effect than angles to goal

ANALYSIS: X-Y

FINDINGS ON SHOT TYPE DISTRIBUTION





Shot Types

Wrist Shot, Snap/Slap Shot most common

Backhand and Tip-in are less common

Wrap around is the least common

ANALYSIS: X-Y

FINDINGS ON SHOT TYPE OVER DISTANCE AND ANGLE

Tip-In
Shot Type

angle / dist	0 to 25	25 to 50	50 to 75	75 to 100
a -90 to -70	2462	14	8	0
b -70 to -50	18465	105	60	0
c -50 to -30	62781	357	204	0
d -30 to -10	264665	1505	860	0
e -10 to 10	593342	3374	1928	0
f 10 to 30	260972	1484	848	0
g 30 to 50	55395	315	180	0
h 50 to 70	13541	77	44	0
i 70 to 90	1231	7	4	0

angle

- ☐ Backhand
- ☐ Slap Shot
- ☐ Snap Shot
- ■ Tip-In
- ☐ Wrap-around
- ☐ Wrist Shot

distance

- ☐ Backhand
- ☐ Slap Shot
- ☐ Snap Shot
- ■ Tip-In
- ☐ Wrap-around
- ☐ Wrist Shot

Shot Types over Distance and Angle

Heatmap of different shot types over distance and angle

- Tip In
- Slap Shot
- Snap Shot
- Backhand
- Wrap-around
- Wrist Shot

ANALYSIS: X-Y

CONCLUSION

- Dependencies of shot types over both distance and angle from goal post
  - Distance has a greater effect than angle
- Distance from goal post
  - 0 to 25 ft: Tip-In and Backhand
  - 25 to 50 ft: Snap Shot
  - 50 to 75 ft: Slap Shot
- Angles to goal post
  - -30 to 30 deg: Tip-In
  - -50 to –90 or 50 to 90 deg: Wrap Around
  - Other shot types are relatively independent of angle to goal
- Wrist Shot
  - A versatile goal conversion shot, used in wide range of distances and angles

ANALYSIS: NPC DREAM TEAM

APPROACH

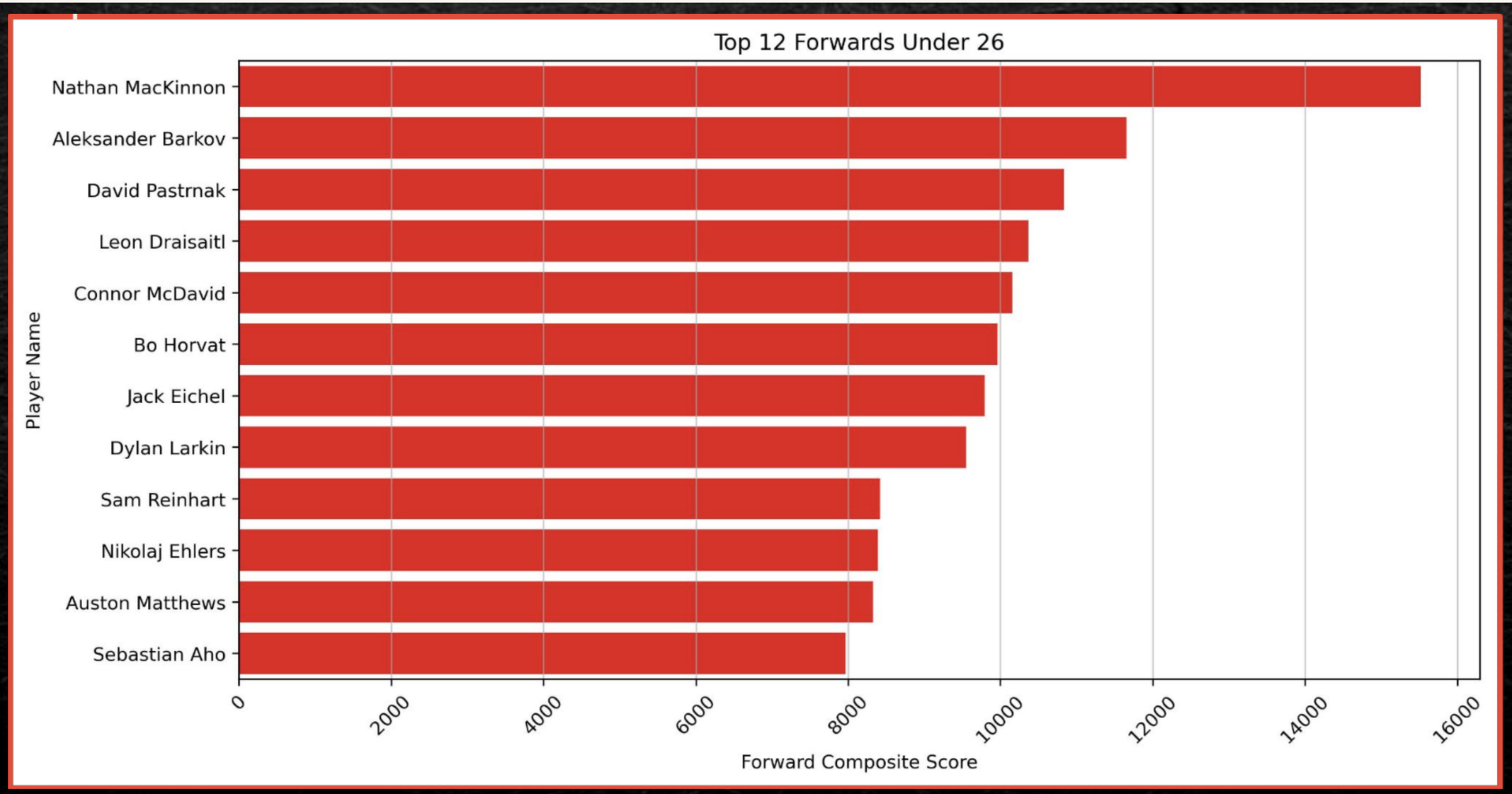
Top 23 Players (under 26)

Categorized into three groups, each group is evaluated based on their recorded performance using a composite score of 10, with the following weighted metrics.

12 Forwards	8 Defensemen	3 Goalies
• Goals: 3 pts • Shots: 3 pts • Assists: 1 pt • Time on Ice: 3 pts	• Takeaways: 3 pts • Blocked: 3 pts • Assists: 1 pt • Time on Ice: 3 pts	• Saves: 7 pts • Time on Ice: 3 pts

ANALYSIS: NPC DREAM TEAM

FINDINGS : TOP 12 FORWARDS



ANALYSIS: NPC DREAM TEAM

1st — 15517.25

Nathan MacKinnon

Age	25
Height	182.88 cm
Weight	92.99 kg
Team	

Colorado Avalanche

2nd — 11655.59

Aleksander Barkov

Age	25
Height	190.50 cm
Weight	96.62 kg
Team	

Florida Panthers

3rd — 10831.29

David Pastrnak

Age	24
Height	182.88 cm
Weight	82.10 kg
Team	

Boston Bruins

4th — 10365.86

Leon Draisaitl

Age	25
Height	187.96 cm
Weight	97.52 kg
Team	

Edmonton Oilers

5th — 10154.56

Connor McDavid

Age	23
Height	185.42 cm
Weight	87.09 kg
Team	

Edmonton Oilers

6th — 9959.67

Bo Horvat

Age	25
Height	182.88 cm
Weight	101.15 kg
Team	

Vancouver Canucks

7th — 9792.83

Jack Eichel

Age	24
Height	187.96 cm
Weight	93.44 kg
Team	

Buffalo Sabres

8th — 9548.94

Dylan Larkin

Age	24
Height	185.42 cm
Weight	89.81 kg
Team	

Detroit Red Wings

9th — 8419.76

Sam Reinhart

Age	25
Height	185.42 cm
Weight	87.09 kg
Team	

Buffalo Sabres

10th — 8392.26

Nikolaj Ehlers

Age	24
Height	182.88 cm
Weight	78.02 kg
Team	

Winnipeg Jets

11th — 8326.55

Auston Matthews

Age	23
Height	190.50 cm
Weight	97.98 kg
Team	

Toronto Maple Leafs

12th — 7966.18

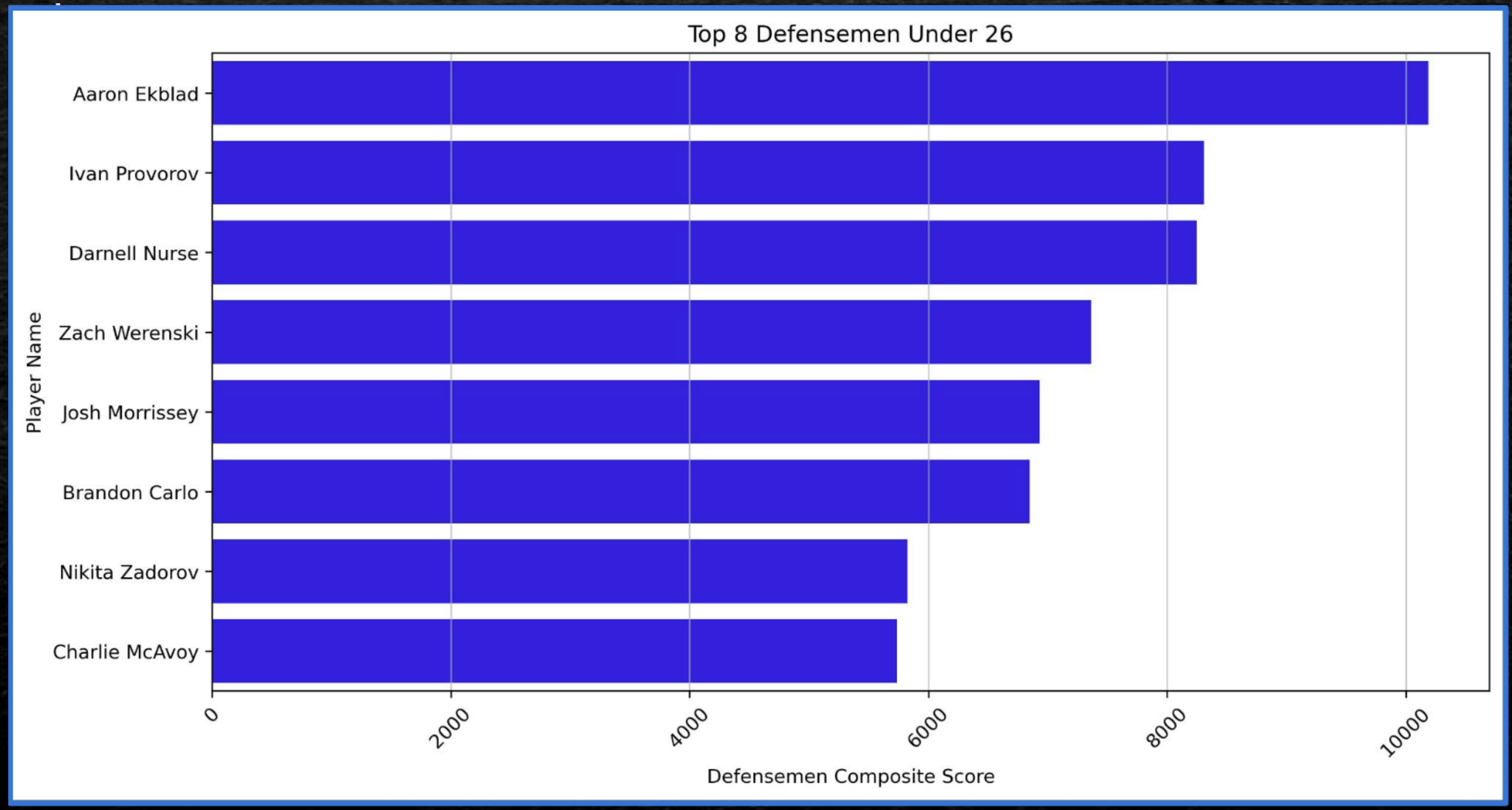
Sebastian Aho

Age	23
Height	180.34 cm
Weight	78.02 kg
Team	

Carolina Hurricanes

ANALYSIS: NPC DREAM TEAM

FINDINGS : TOP 8 DEFENSEMEN



ANALYSIS: NPC DREAM TEAM

1st — 10191.58

Aaron Ekblad

Age	24
Height	193.04 cm
Weight	97.98 kg
Team	

Florida Panthers

2nd — 8312.16

Ivan Provorov

Age	23
Height	185.42 cm
Weight	91.17 kg
Team	

Philadelphia Flyers

3rd — 8250.85

Darnell Nurse

Age	25
Height	193.04 cm
Weight	100.24 kg
Team	

Edmonton Oilers

4th — 7362.31

Zach Werenski

Age	23
Height	187.96 cm
Weight	94.8 kg
Team	

Columbus Blue Jackets

5th — 6931.86

Josh Morrissey

Age	25
Height	182.88 cm
Weight	88.45 kg
Team	

Winnipeg Jets

6th — 6850.29

Brandon Carlo

Age	24
Height	195.58 cm
Weight	94.35 kg
Team	

Boston Bruins

7th — 5823.02

Nikita Zadorov

Age	25
Height	195.58 cm
Weight	104.33 kg
Team	

Colorado Avalanche

8th — 5738.25

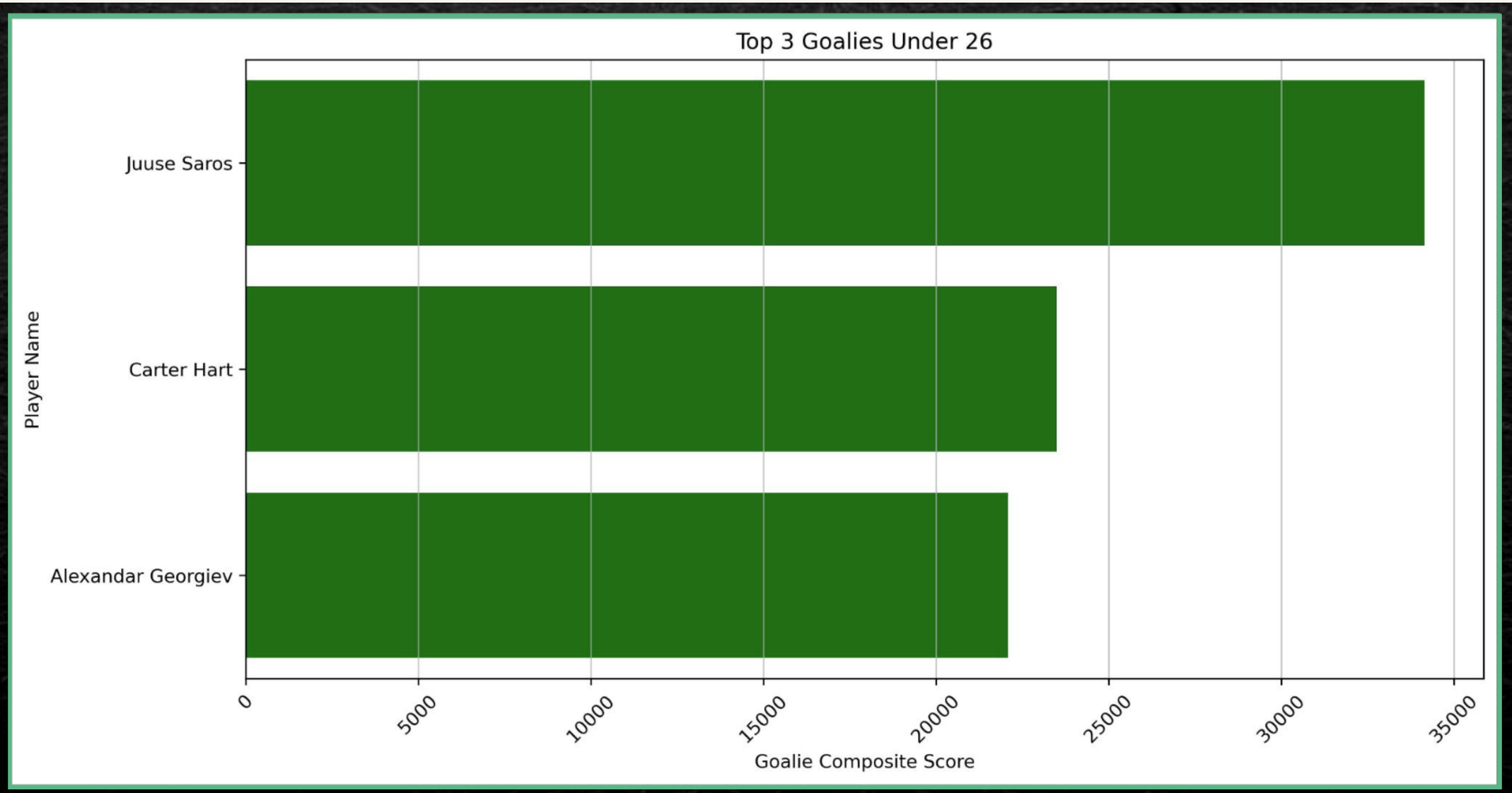
Charlie McAvoy

Age	23
Height	182.88 cm
Weight	94.35 kg
Team	

Boston Bruins

ANALYSIS: NPC DREAM TEAM

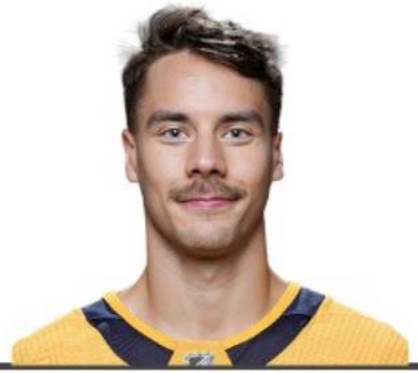
FINDINGS : TOP 3 GOALIES



ANALYSIS: NPC DREAM TEAM

1st — 34155.22

Juuse Saros



Age	25
Height	180.34 cm
Weight	81.65 kg
Team	

Nashville Predators

2nd — 23491.58

Carter Hart

Age	22
Height	187.96 cm
Weight	82.1 kg
Team	

Philadelphia Flyers

3rd — 22080.00

Alexandar Georgiev

Age	24
Height	185.42 cm
Weight	81.65 kg
Team	

NY Rangers

ANALYSIS: NPC DREAM TEAM

CONCLUSION

The analysis identifies young talents in key positions with the highest potential to excel in the rankings and establishes a solid foundation for creating engaging characters in game development.

⏩ “Fast Forward” to today ⏩

Statistics from 2021 to 2023:

- 20 of our proposed 23 players were ranked within the Top 100 in their respective fields.
- Accuracy estimated at **87%**.

CHALLENGES ... SOLUTIONS

Unfamiliar with Azure and NHL

Limited knowledge on ETL pipeline implementation in Azure

Limited knowledge of hockey rules, difficult to contextualize data and select relevant metrics

Collaborative Learning

Studied and selected appropriate services

Spent time learning NHL rules to align data.

Team shared findings to accelerate learning in Azure and NHL data.

Resource Restrictions

Faced budget and resource constraints due to licensing limits

Cost Management

Monitored cloud usage
Deactivated resources when not in use

Tight Timelines

Short delivery timelines

Project Management

Defined roles, parallel task execution, and milestone tracking ensured timely project completion.

CONCLUSION

The project provided the team with important key learnings:

1. Building ETL Pipelines using Microsoft Azure
2. Data-driven Sports Analytics and Importance of domain knowledge
3. Collaborative Learning and Adaptability

Thank You !

Special Thanks to Christine and :

Generation SINGAPORE

tp