

PRESENTED BY TEAM 9

FORMULA 1

TENEIKA ASKEW MINI SPRINT VIA CLICKED

BUSINESS PROBLEM

Scenario - F1 and Racing Data

We are data analysts on a team working for the Fédération Internationale de l'Automobile (FIA), owned by the Formula One Group. We are working closely with Mohammed Ben Sulayem, FIA President, to provide data for recruiting agencies to make decisions on recruiting top performers.

Optimizing F1 Driver Recruitment with Historical Data Analysis

Problem Statement: In Formula 1 racing, selecting drivers with winning attributes is critical for success. However, recruitment agencies face challenges in identifying top talent amidst numerous factors influencing race outcomes.

Approach:

1. Data Collection: Gather comprehensive historical F1 race data.
2. Analysis: Explore key performance metrics associated with top-three finishes.
3. Model Development: Build a predictive model for assessing driver potential.
4. Recommendations: Provide actionable insights for recruitment decisions.

Outcomes:

- Analysis of winning attributes and performance metrics.
- Predictive model for driver selection.
- Actionable recommendations for recruitment agencies.

Benefits:

- Improved driver selection accuracy.
- Enhanced team competitiveness.
- Increased race-winning potential.

By leveraging historical F1 data and advanced analytics, we empower recruitment agencies to make informed decisions, ensuring top talent selection.



ABOUT THE TEAM



**Anabelle
Capois Espinal**

- Bachelor, Business and Econ. '14
- MBA, Strategy '22
- Data Science Bootcamp '23
- KaggleX Mentee '23
- Code First Girls '24



**Chloe
Callender**

- BSc. Computer Science
- Post Bacc. Business Analytics
- MSc. Computer Science



**Jalen
Morris**

- B.Sc. Biology
- General Assembly - Data Science Immerserve



**Shana
Prince**

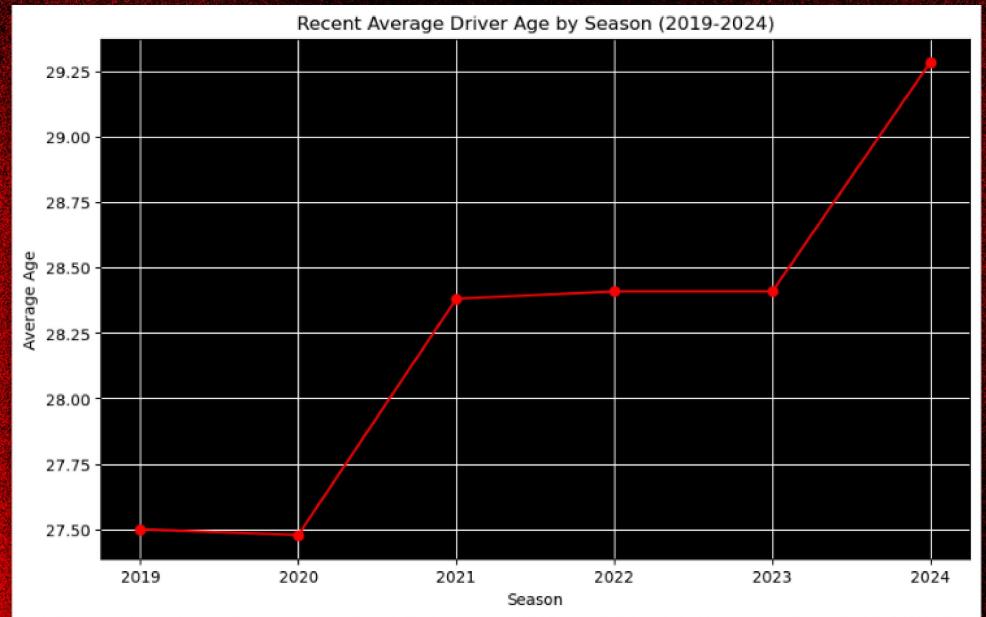
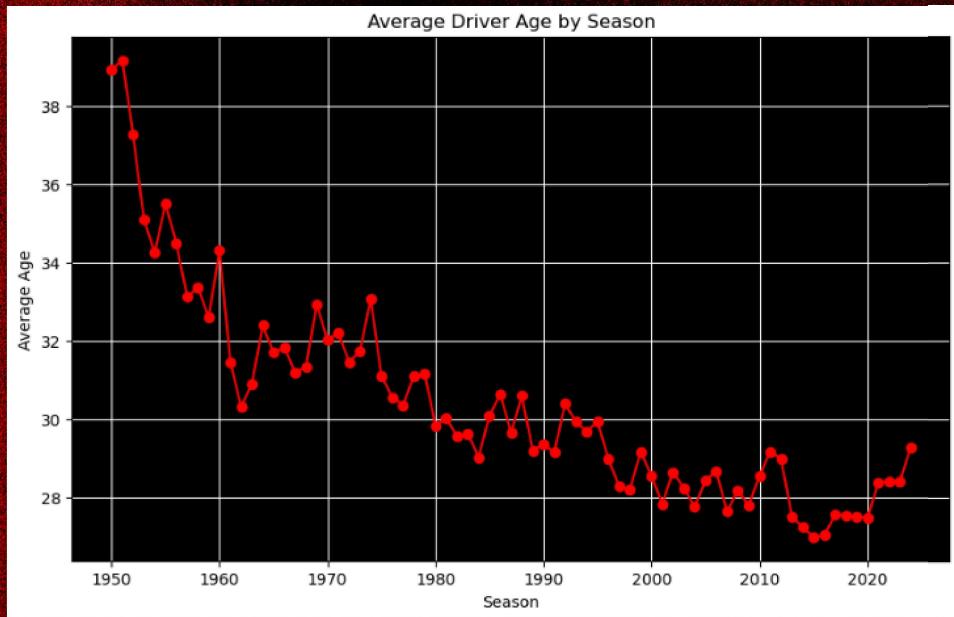
- B.A. in English
- M.S. in Educational & Instructional Technology



EXPLORING THE DATA

AVERAGE DRIVER AGE

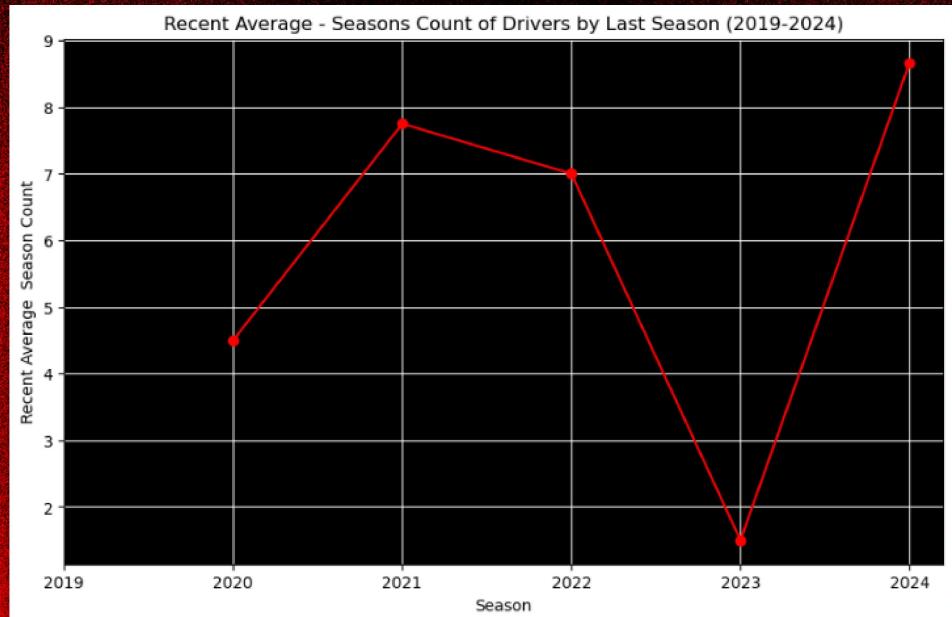
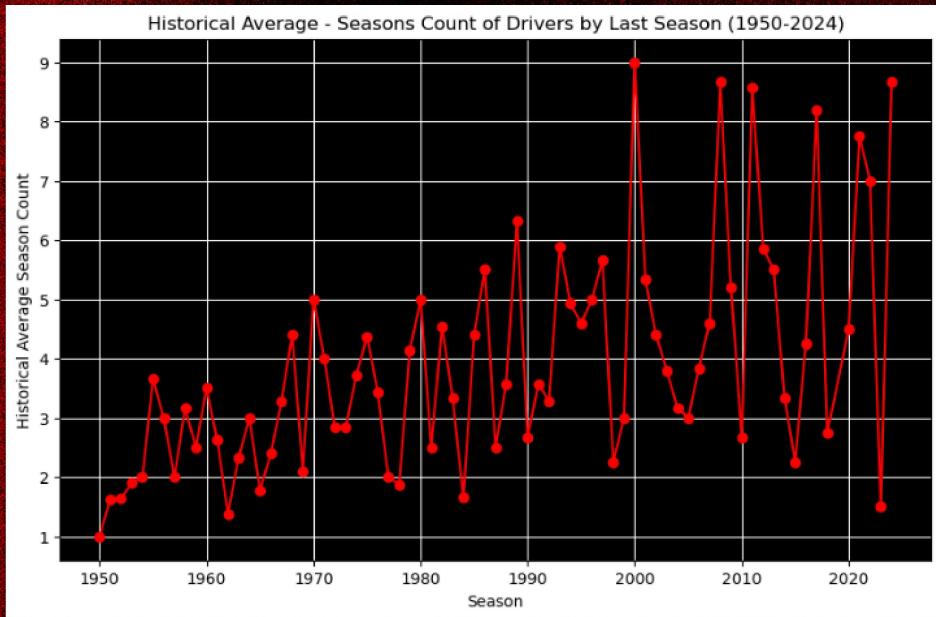
- Since 1950 there has been an overall downward trend in the driver age
- Recent seasons have seen an upward trend
 - A nearly 2 year increase from 2019 to now



*Data considers all status types (Finished, +2Laps, Accident, etc.)

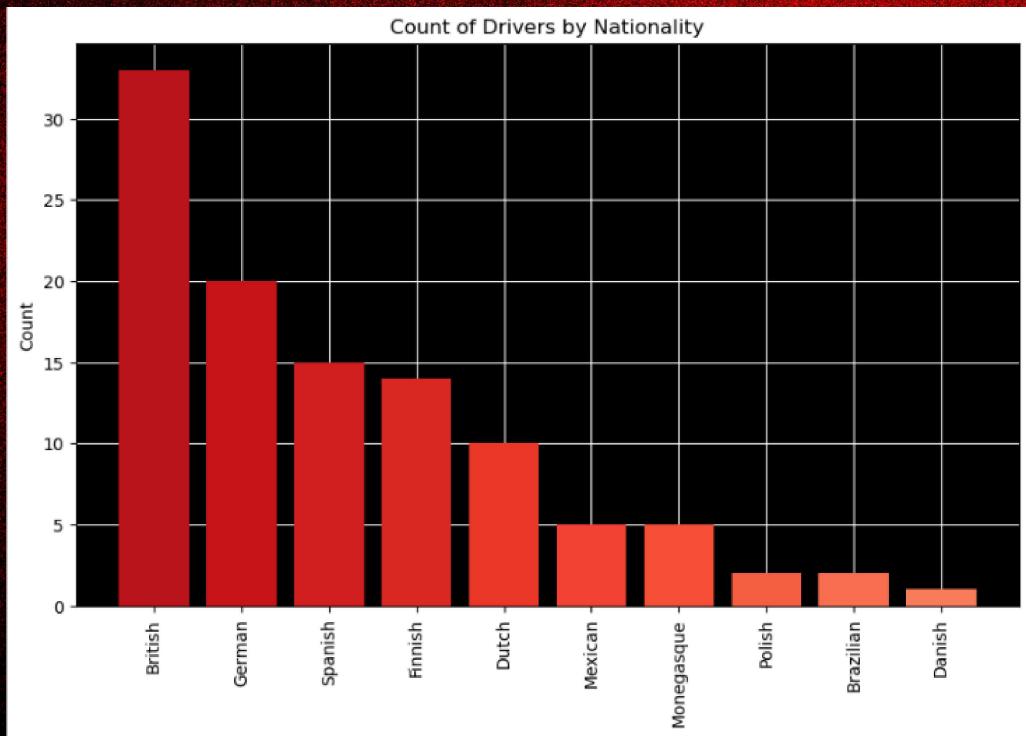
YEARS OF EXPERIENCE

- Average season count by year of latest season participation has seen an upward trend
- Current season drivers have among highest average season participation count
- 2019 was the only year that saw no driver have their last season participation



*Data considers all status types (Finished, +2Laps, Accident, etc.)

DRIVER NATIONALITY



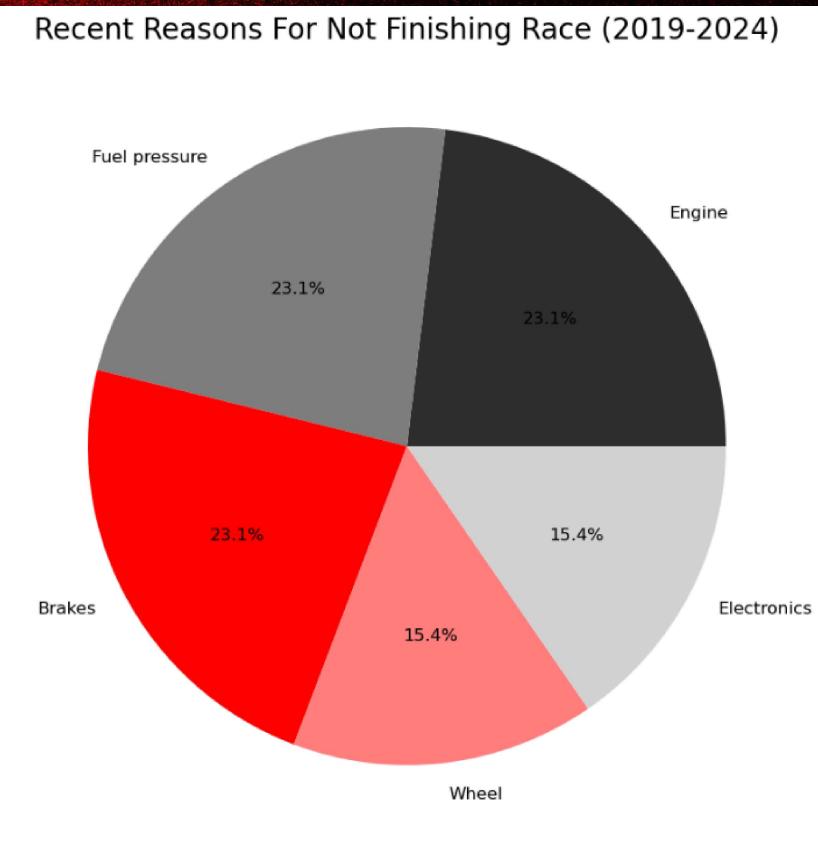
Drivers with these nationalities have historically shown that they can finish the race, placing among the top 3.

00:30:39.958...

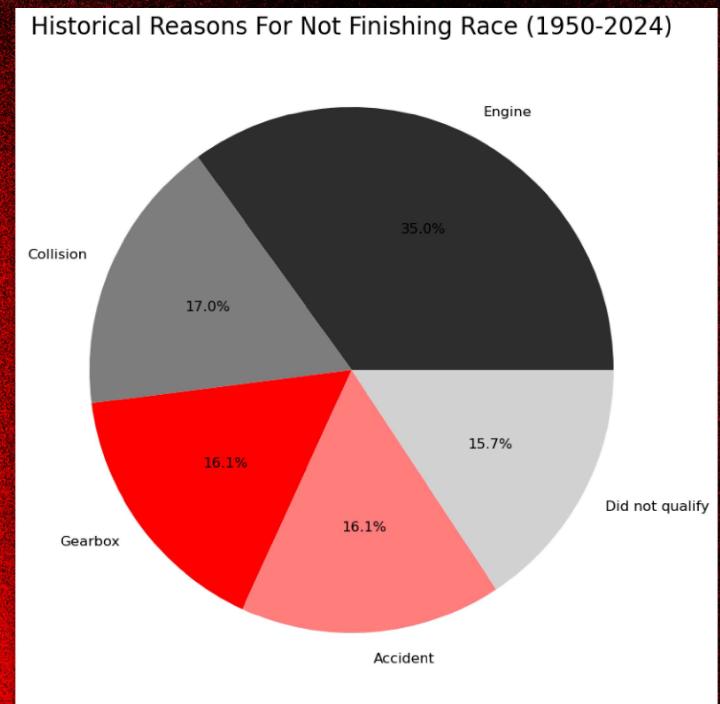
TOP THREE FINISHER STATS

	Historical Mean	Recent Mean
Total RaceTime	00:33:41.378571028	00:30:39.958030555
Fastest Lap Speed	212.891822 milliseconds	222.432028 milliseconds
Laps	56.887850	58.083333
Points	17.355140	19.611111

CAUSE OF NOT FINISHING



- The top reasons for not being able to finish the race has shifted
 - Electronics is present in recent data given car innovation
 - Engine issues remains a major issue
 - Collisions and accidents are not prevalent in recent years

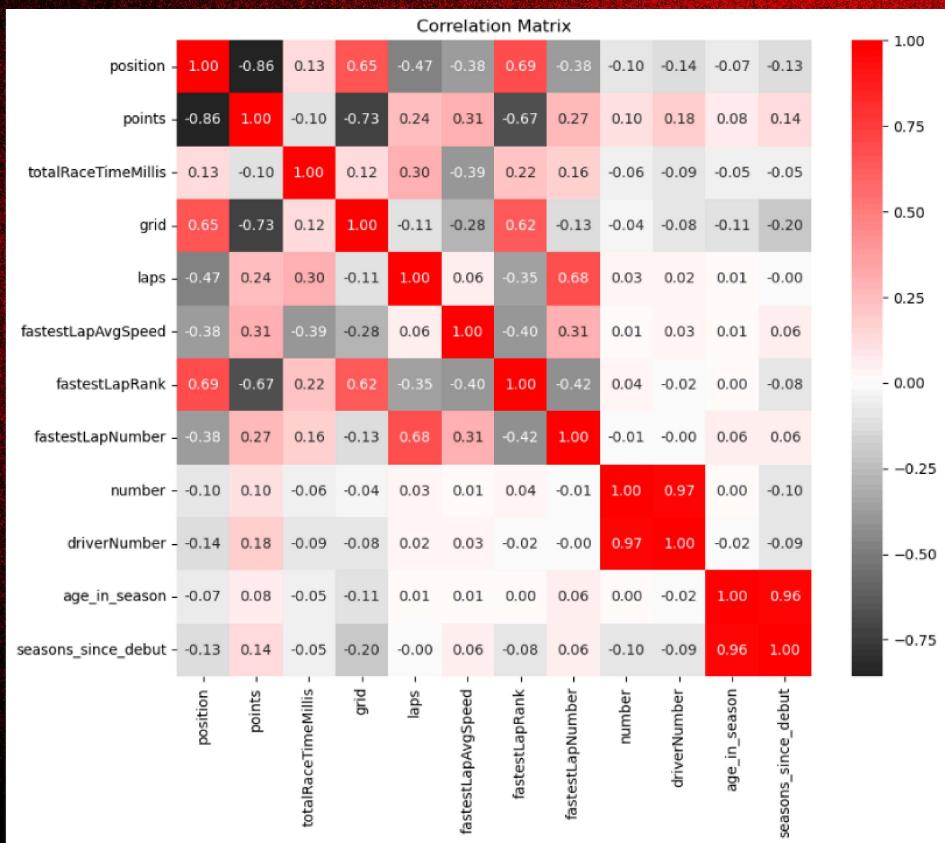




MACHINE LEARNING

PREDICTION ATTRIBUTES

What features had a strong correlation with finishing position for 2019-2025 race finishers?



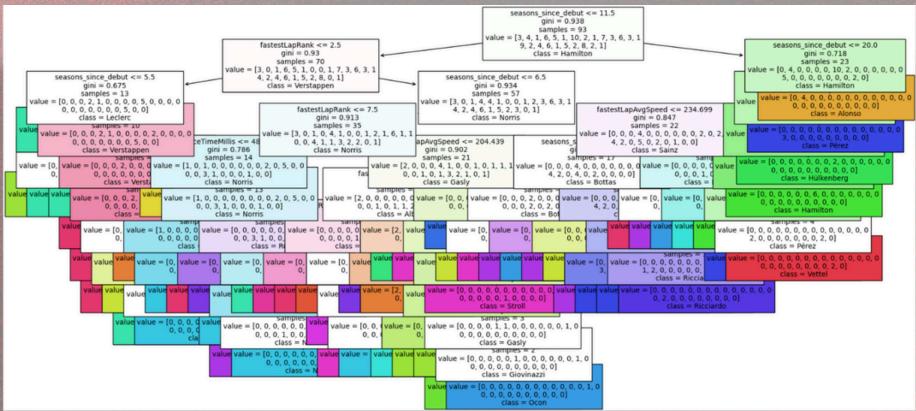
Chosen Attributes for ML Model

01. Winning Time

02. Fastest Lap Average Speed

03. Fastest Lap Rank

04. Driver Season Experience



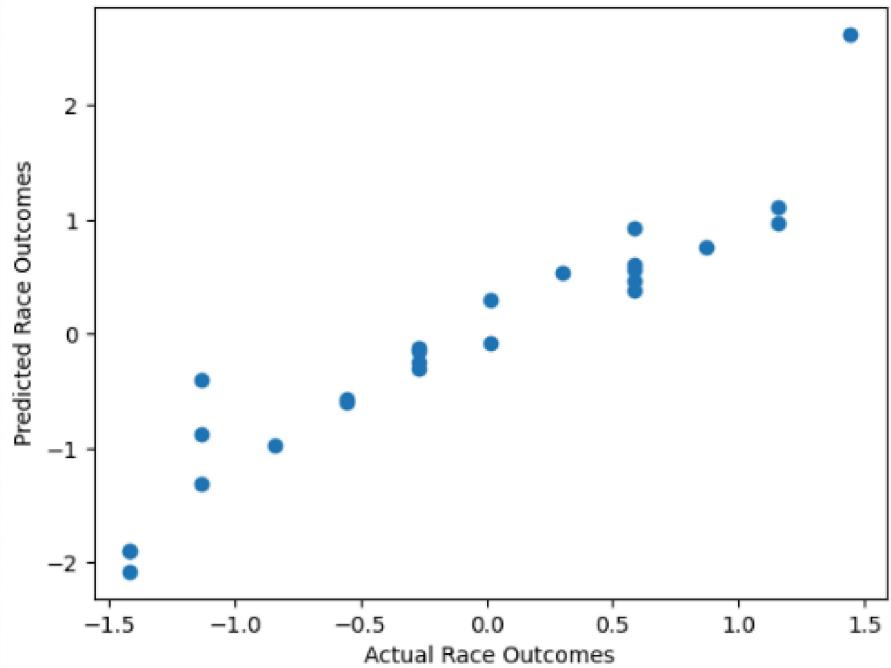
25%

	totalRaceTimeMillis	fastestLapAvgSpeed	fastestLapRank	seasons_since_debut	Actual Winner	Predicted Winner
83	7380177.0	225.473	6.0	4	Lecerc	Russell
4	5185528.0	219.621	4.0	2	Lecerc	Lecerc
94	5868338.0	201.641	8.0	9	Magnussen	Sainz
81	7376598.0	230.403	1.0	15	Hamilton	Hamilton
24	5697363.0	205.313	4.0	13	Vettel	Pérez
129	5726510.0	200.559	15.0	5	Albon	Tsunoda
116	5115417.0	239.629	5.0	4	Norris	Russell
147	4927726.0	238.424	8.0	8	Ocon	Ocon
61	5524642.0	208.984	2.0	7	Verstappen	Verstappen
25	5707049.0	203.422	12.0	1	Norris	Verstappen

DECISION TREE MODEL ACCURACY

- A sport of outliers
- Our model is a work in progress. However, by narrowing the data used to the last 5 years, the accuracy increased by about 14%

Actual vs. Predicted Race Outcomes



88%

LINEAR REGRESSION MODEL ACCURACY

```
Accuracy: 0.88
Confusion Matrix:
[[15  1]
 [ 2  7]]
Classification Report:
precision    recall   f1-score   support
          0       0.88      0.94      0.91      16
          1       0.88      0.78      0.82       9
accuracy                           0.88      25
macro avg       0.88      0.86      0.87      25
weighted avg    0.88      0.88      0.88      25
```



BUSINESS IMPLICATIONS



INFORMED AND STRATEGIC DECISIONS

- 88% ML model accuracy = **strong predictive capabilities**, offering insights into outcomes and driver performances.
- Can leverage this accuracy to **enhance decision-making processes**, such as optimizing race strategies and identifying areas for driver improvement.
- Improved forecasting accuracy can lead to **more engaging experiences for sponsors, broadcasters, and fans**.
- While the 88% accuracy rate is significant, **ongoing refinement and validation of the model** are essential to ensure reliability across diverse conditions and seasons.

PRESENTED BY TEAM 9

THANK YOU.

[CHECK OUT OUR GOOGLE COLAB!](#)