

QUESTION

**ANY GUESSES WHO WILL
BE POINT FINISHERS IN THE FIRST
RACE OF THE 2024 SEASON?**

SURVEY SAYS:

P1 VER

P2 LEC

P3 SAI

PIA

RIC

NOR

PER

RUS

HAM

ALO



ON THE RIGHT TRACK
WHO WINS?

MSDS 2024
LEARNING TEAM 4

PREDICTING F1 RACE POINT
FINISHERS USING AN ENSEMBLE
TREE MODEL REGRESSOR



FAST FACTS

10
TEAMS ON
THE GRID

2
DRIVERS
PER TEAM

>20
RACES IN A SEASON

3 PRACTICE SESSIONS

1 QUALIFYING SESSION

1 RACE SESSION

TOP 10
WILL SCORE POINTS
AT THE END OF THE RACE

TEAMS AND
DRIVERS

RACES PER
SEASON

RACE
WEEKEND

POINTS
SYSTEM



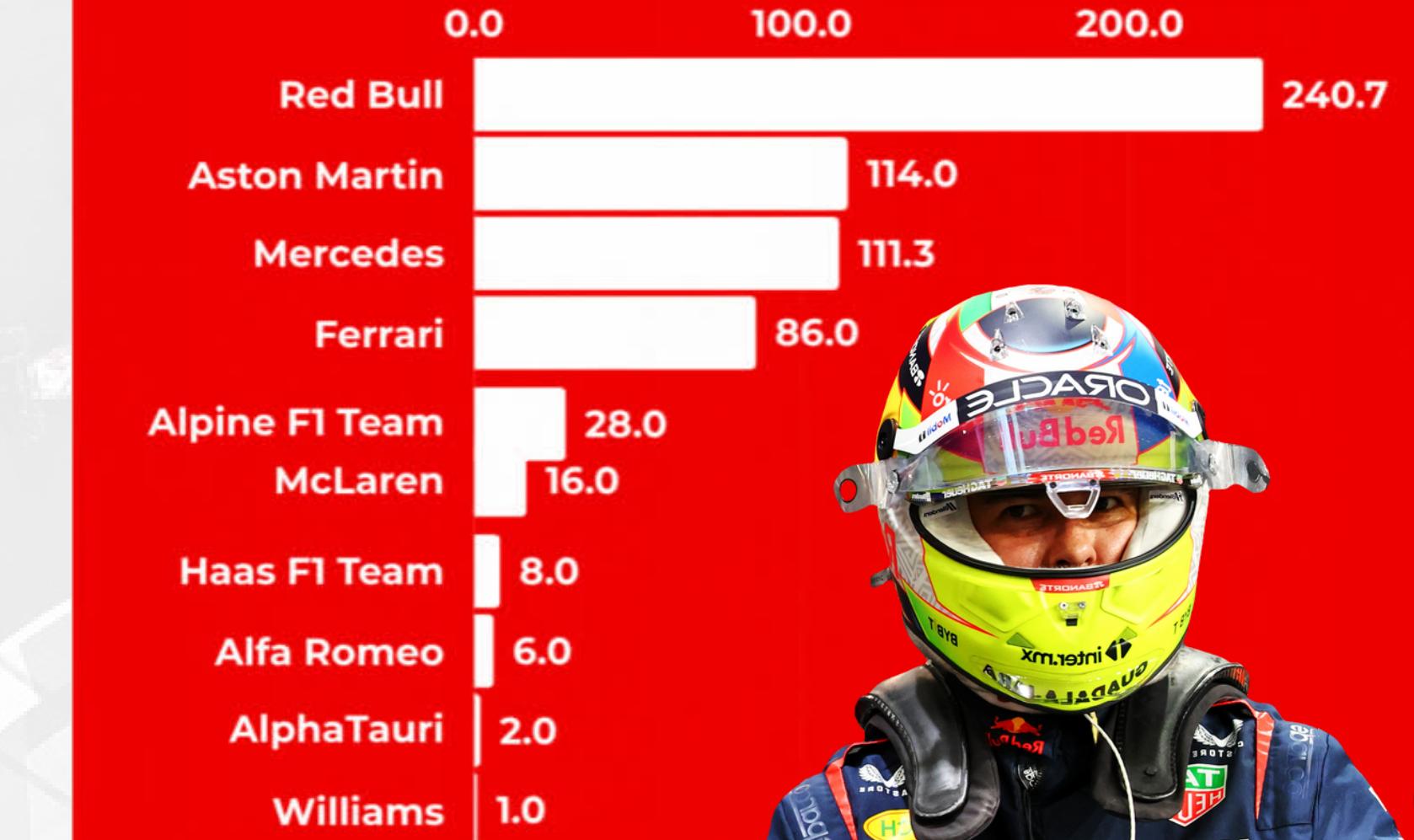
F1 RECAP

EXPLORATORY DATA ANALYSIS

2023 END OF SEASON POINTS



2023 END OF SEASON POINTS



Red bull and Max Verstappen have emerged as the **most dominant** in both the 2022 and 2023 seasons; they are **followed Ferrari drivers, and Mercedes drivers.**



PROJECT

ESSENTIALS & DETAILS

PROJECT OBJECTIVE

PREDICT POINTS FINISHERS

How can we **develop a regression model that predicts Formula 1 drivers likely to finish in the top 10**, considering the criticality of every point towards the championship?



F1 COMMUNITY IN THE PHILIPPINES

0 F1 Club of the Philippines ▾

official-session-chat

GOAL: LVL 2 2/7 Boosts >

@Midnight95 That's it for testing! Apologies for the hiccups earlier. Unreliable yung IPTV ko 😢
zichhh 02/24/2024 6:55 AM
Thank you!

February 25, 2024

moemoe 02/25/2024 9:37 PM
hi mods, can i have an invite link 😊

February 27, 2024

pimpom Yesterday at 7:48 PM
bolaga

@moemoe hi mods, can i have an invite link 😊
Midnight95 Yesterday at 9:26 PM
<https://discord.gg/n8RKqHWM43>

YOU'VE BEEN INVITED TO JOIN A SERVER

F1 Club of the Philippines 0
202 Online • 2,528 Members Joined

GENERAL

ontrack-racers
sim-racers
racemarshals-page
games-chat
content-creators
motorsports-chat

FEEDBACK

discord-suggestions
emoji-suggestions
staff-feedback

21lateboyman Yesterday at 11:04 PM
hello, can i ask what time of Bahrain GP in the Philippines?

February 28, 2024

@21lateboyman hello, can i ask what time of Bahrain GP in the Philippines?
Charles Today at 10:27 PM
11:00 pm po

Message #official-session-chat

F1 Club of the Philippines

Private group · 14.4K members

+ Invite Share

Discussion Look For Players Buy and Sell Featured Members Media

SIGNIFICANT INCREASE IN EXPOSURE AFTER THE RELEASE OF
NETFLIX DRIVE TO SURVIVE

2.5K
Discord
members

14.4K
Facebook
members

B U S I N E S S V A L U E



TEAM STRATEGY

Establish **expectations and strategies** while **identifying areas for improvement**.



SPORTS BETTING

Allow **bettors to make informed decisions on who to bet on** during the race



F1 FANTASY

Aid in **building F1 fantasy teams and planning for transfers** on race weekends

\$75M

Up to 75 Million
Dollars in sponsorship
per year

>\$1B

Estimated market
size of the Formula 1
betting market

\$87B

Fantasy sports
projected to grow
by year 2031

PREVIOUS STUDIES

01

PREDICTING RACE RANKINGS IN 2020

- Only uses historical data specific to the circuit

DIFFERENCE FACTOR

Uses all previous race information to predict values

Landry, M. (2020, July). F1-predictions Version 3. Retrieved February 22, 2024 from <https://www.kaggle.com/code/mlandry/f1-predictions>

02

PREDICTING RACE RANKINGS IN 2019

- Used fixed features without use of any time elements

DIFFERENCE FACTOR

Uses time-series features such as lags

Robert, J., Evrenian, R., & Jiyuan SUN, T. C. (2021, May 26). F1 race predictions. Kaggle. <https://www.kaggle.com/code/julien24/f1-race-predictions>



FEATURES USED FOR PREDICTIONS:

CURRENT RACE

QUALIFYING



PREVIOUS RACE

DRIVER/TEAM PTS

DRIVER/TEAM WINS

DRIVER POSITION

TEAM POSITION

QUALIFYING POS.

FASTEAT LAP RANK

AVE. LAP SPEED

2 - 5 RACES AGO

END RACE POSITION

QUALIFYING POS.



573 ROWS

20 FEATS.

2022 AND 2023 F1 SEASONS DATASET



Data was extracted from the **Ergast Development website**, an **API** that provides **Formula 1 data** since the 1950s World Championship, and **Wikipedia** for additional information.



MODEL DETAILS

LAST FOUR RACES OF THE 2023 SEASON

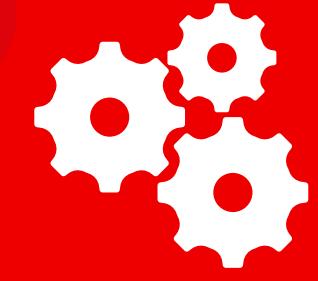
01



SELECT BEST MODEL

Build baseline model: walk-forward
Time series split for assessing robustness

02



IMPLEMENT MODEL

Produce race result predictions
for the final few races of the season

03



EXPLAINABILITY

Use interpretation techniques and
make actionable suggestions



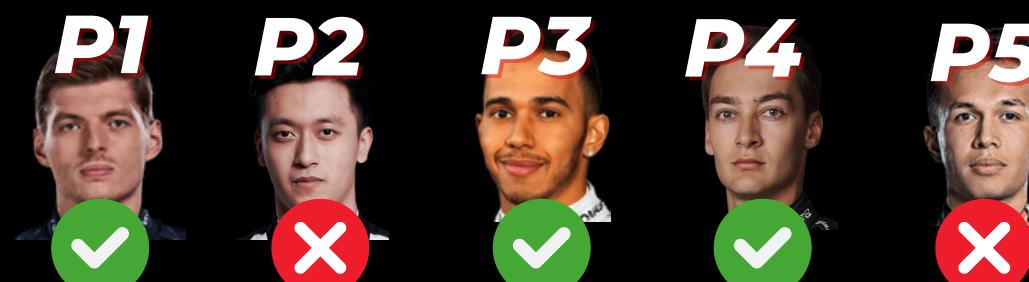
METHODOLOGY

RANKING EVALUATION METRIC

AVERAGE PRECISION @ K

- measures the **percentage of relevant results** among top k results
- Takes into consideration the **order of relevant results**

EXAMPLE



= **0.81**



= **0.53**





BASE-LINE COMPARISON **Walk-forward**



Using **Finishing Position in the Previous Race** as the prediction



Score to beat is **MAP@K score of 80%** which is the average scores of all test splits

TESTING MODEL ROBUSTNESS **Time Series Split**



Maximum test size constrained to the **driver with the lowest number of races**

RANDOM FOREST



n_estimators: 300
max_depth: 1
max_features: 0.3

MAP@K SCORE:
89%

XGBOOST REGRESSOR



n_estimators: 300
max_depth: 8
learning_rate: 0.05

MAP@K SCORE:
83%

GRADIENT BOOSTING



n_estimators: 300
max_depth: 10
learning_rate: 0.1

MAP@K SCORE:
78%

LINEAR REGRESSION



MAP@K SCORE:
64%

COMPARE MODELS

Ensemble Models were mainly used to capture the complexity of the data set

BEST MODEL PREDICTION

Random Forest

Regressor

MAP@K SCORE: **89%**

n_estimators: 300; max_depth: 1; max_features: 0.3

- Average for all splits is **89%**
MAP@K
- The **baseline model was beat by almost 10% MAP@K**



Red Bull



PREDICTIONS

LAST FOUR RACES OF THE 2023 SEASON



PREDICTION STEPS

01

Predict **one race at a time** per driver for all races.

02

Each driver has **different set of historical data**

03

Lags: **most recent completed race** by a driver.

04

Exclude drivers affected by unpredictable race events*

* crashes or engine failures



2023 MEXICO
GRAND PRIX

VER	PIA
LEC	NOR
SAI	OCO
HAM	GAS
RUS	RIC

AP@K: 89%



2023 LAS VEGAS
GRAND PRIX

VER	RUS
LEC	ALO
PER	OCO
HAM	PIA
SAI	STR

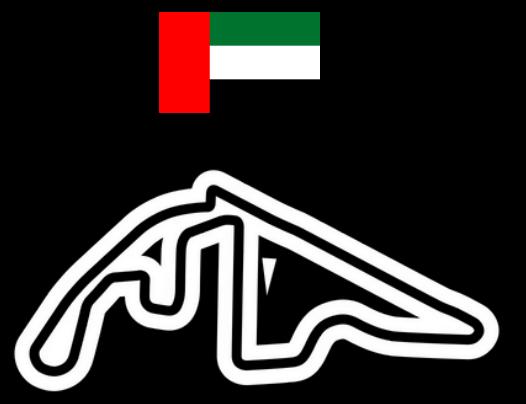
AP@K: 100%



2023 SAO PAULO
GRAND PRIX

VER	NOR
PER	OCO
HAM	PIA
SAI	STR
ALO	GAS

AP@K: 88%



2023 ABU DHABI
GRAND PRIX

VER	RUS
PER	NOR
LEC	ALO
HAM	OCO
PIA	STR

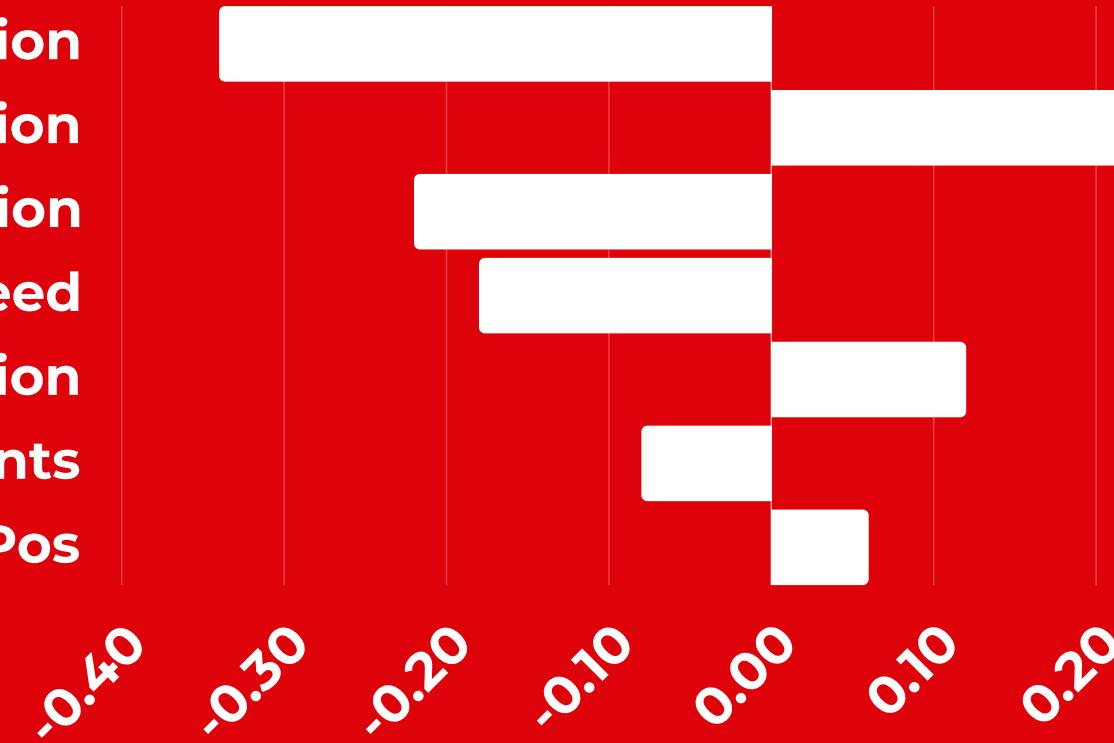
AP@K: 89%

91.5% Predicting the last four
MAP@K races of the 2023 season



GASLY PREDICTED TO BE >10TH PLACE IN THE ABU DHABI GRAND PRIX

qualifyingPosition
5lagPosition
1lag_constructorPosition
1lag_aveLapSpeed
1lag_aveLapPosition
1lag_constructorPoints
1LagQPos



PIERRE GASLY

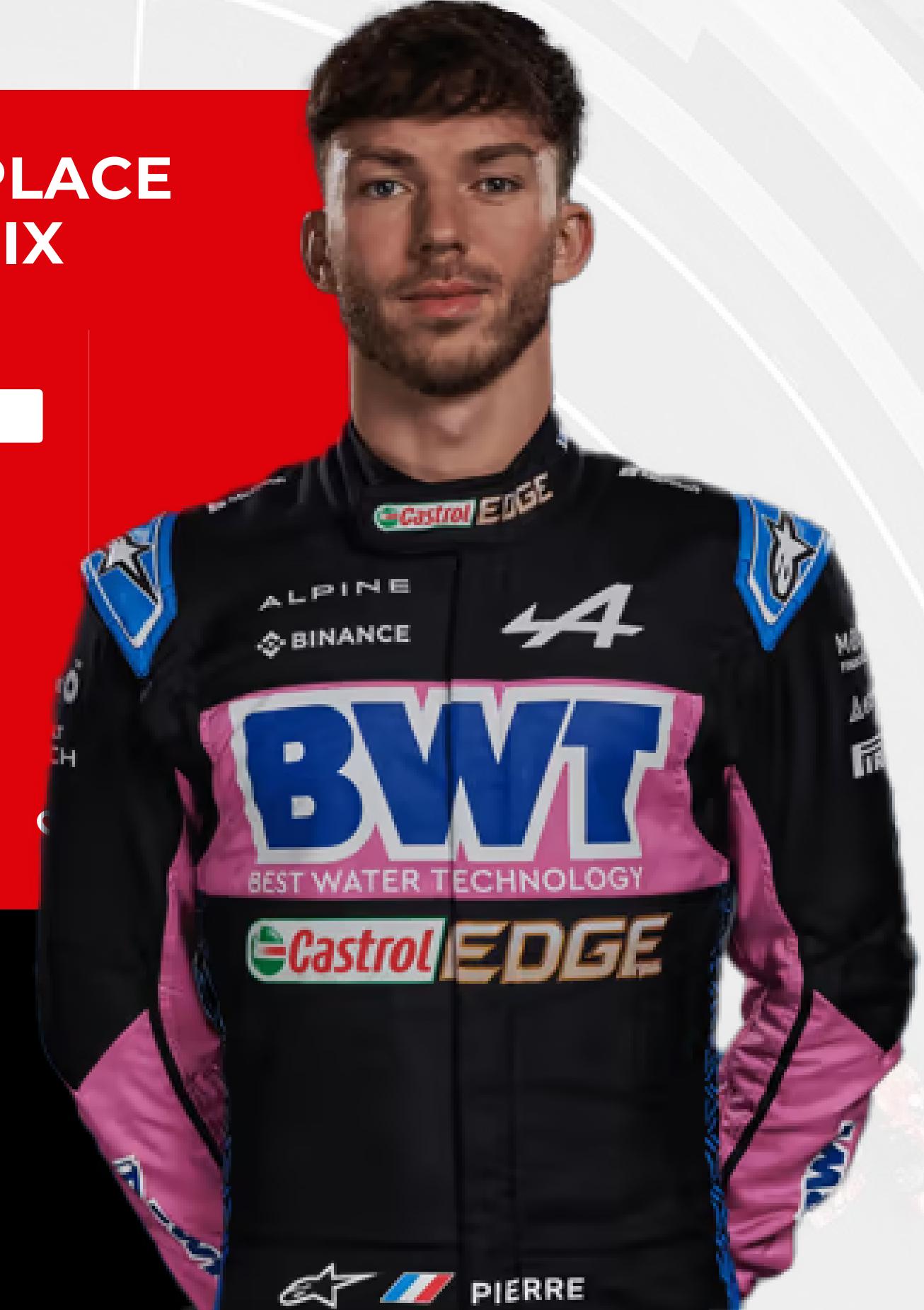
REASONS FOR
PREDICTION

9th - 12th

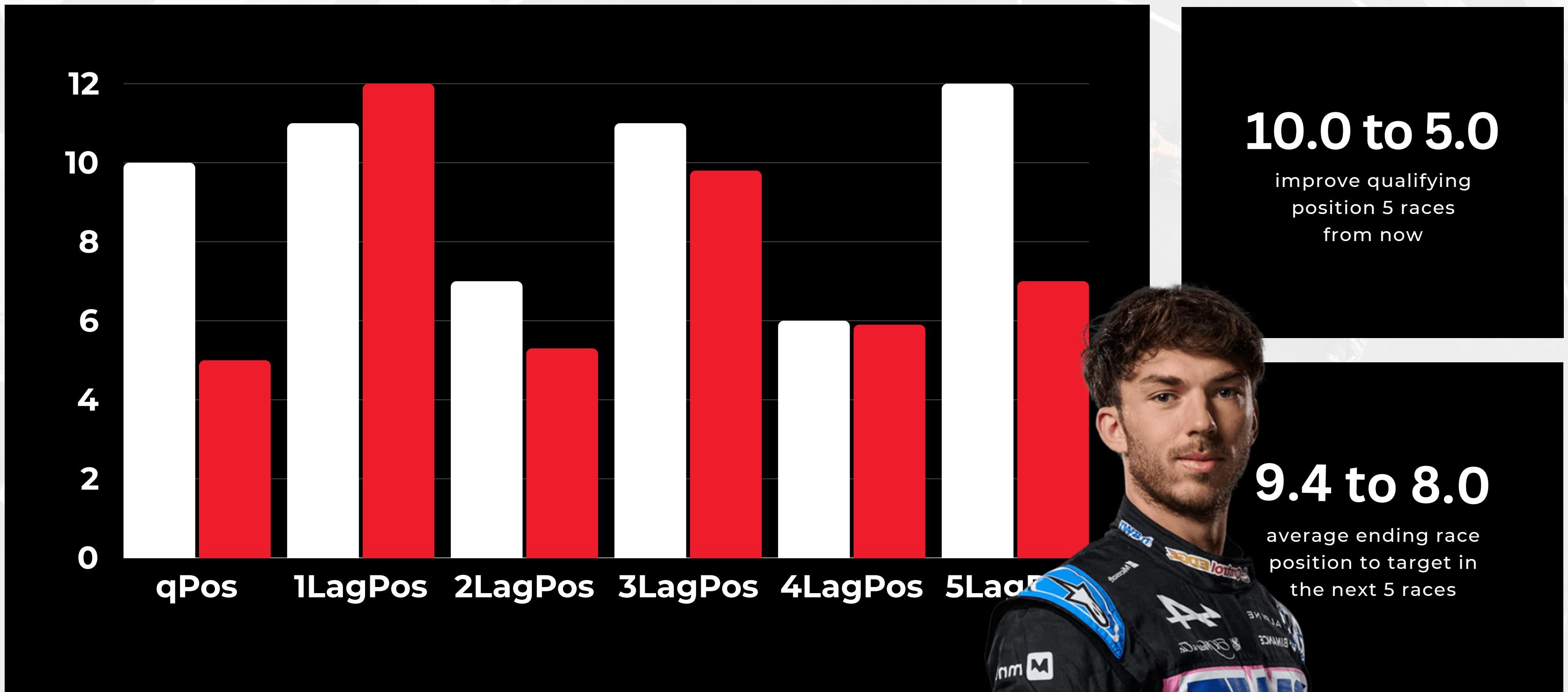
Position in the
current qualifying

11th - 13th

End race position
5 races ago



WHAT CAN GASLY DO TO IMPROVE?





RECOMMENDATIONS

TEAM RADIO

IRIA



“Future researchers can add driver performance metrics as features in predicting values”

Ex. Race craft, pace, etc.

TEAM RADIO

IUY



“Possibility of including car performance metrics such as aerodynamics and engine performance”

TEAM RADIO

IRAM



“Explore models where predictions can be done in batches like multi-output neural networks”

TEAM RADIO

IDEL



“Enrich dataset with live race data (weather, track temp) and qualitative insights (team strategies, interviews) for analysis.”

A photograph of Formula 1 driver Max Verstappen. He is wearing a dark blue Red Bull Racing team jacket over a white shirt. The jacket features the Red Bull logo, TAG Heuer, Mobil 1, Oracle, Honda, and Tezos sponsors. He has his right hand raised to his head, resting his fingers against his temple. He is smiling slightly and looking towards the camera. The background is a blurred indoor setting.

QUESTION

**ANY GUESSES WHO WILL
BE POINT FINISHERS IN THE FIRST
RACE OF THE 2024 SEASON?**

HERE'S WHAT WE THINK...



**2024 BAHRAIN
GRAND PRIX**

VER

PER

LEC

RUS

HAM

VER

1ST PLACE

PER

2ND PLACE

SAI

ALO

PIA

NOR

OCO

LEC

3RD PLACE



THANK YOU!



REFERENCES

- Autocom. (2018). From Racing Legends to Future Stars: Iconic Drivers That Shaped the Formula 1 Betting Market. Retrieved from autocom.co.uk/from-racing-legends-to-future-stars-iconic-drivers/
- Cooper, S. (2023, September 8). 10 Big Sponsorship Deals Generating Huge \$350m Fortune on F1 Grid. Retrieved from www.planetf1.com/features/f1-sponsorship-deals-value-teams
- Efimov, V. (2023, July 29). Comprehensive Guide to Ranking Evaluation Metrics. Retrieved from towardsdatascience.com/comprehensive-guide-to-ranking-evaluation-metrics-7d10382c1025
- Ergast Developer API. (n.d.). The Ergast API Is Now DEPRECATED. Retrieved from ergast.com/mrd/
- Jeddah Corniche Circuit. (2022, March 27). FORMULA 1 STC SAUDI ARABIAN GRAND PRIX 2022 - PRACTICE 1. Retrieved from www.formula1.com/en/results.html/2022/races/1125/saudi-arabia/practice-1.html
- Landry, M. (2020, July). F1-predictions Version 3. Retrieved February 22, 2024 from <https://www.kaggle.com/code/mlandry/f1-predictions/notebook?scriptVersionId=39971201>
- Lutz, Z. (n.d.). Formula 1 World Championship (1950 - 2023). How to Bet on Formula 1 - The Complete Guide for 2024. Retrieved from www.si.com/fannation/betting/how-to-bet/f1-betting-guide
- Wikipedia. (2009). Abu Dhabi Grand Prix. Retrieved from en.wikipedia.org/wiki/Abu_Dhabi_Grand_Prix



GRID SEARCH

```
GBM - (8, 0.01) - 0.75334847981971
GBM - (8, 0.05) - 0.7549921537171757
GBM - (8, 0.1) - 0.7609597794776586
GBM - (10, 0.01) - 0.7517019522731823
GBM - (10, 0.05) - 0.7549921537171757
GBM - (10, 0.1) - 0.7609597794776586
GBM - (12, 0.01) - 0.7517019522731823
GBM - (12, 0.05) - 0.7549921537171757
GBM - (12, 0.1) - 0.7609597794776586
```

```
XGB - (8, 0.01) - 0.8171396570010107
XGB - (8, 0.05) - 0.8316488738276878
XGB - (8, 0.1) - 0.8142960773443358
XGB - (10, 0.01) - 0.8047757690371228
XGB - (10, 0.05) - 0.8316488738276878
XGB - (10, 0.1) - 0.8142960773443358
XGB - (12, 0.01) - 0.8047757690371228
XGB - (12, 0.05) - 0.8316488738276878
XGB - (12, 0.1) - 0.8123642198499782
```

```
RF - (1, 0.2) - 0.8882130540176395
RF - (1, 0.25) - 0.8871839743010598
RF - (1, 0.3) - 0.8901567081515716
RF - (1, 0.35) - 0.8901567081515716
RF - (1, 0.4) - 0.8756917786366422
RF - (2, 0.2) - 0.8861123872232999
RF - (2, 0.25) - 0.8685995798660482
RF - (2, 0.3) - 0.8734351815300448
RF - (2, 0.35) - 0.8720883452387641
RF - (2, 0.4) - 0.8734351815300448
```



SAMPLE DATA

raceld	positionOrder	1Lag_dsPois	1Lag_dsPosition	1Lag_driveWins	1Lag_consTructorPoints	1Lag_consTructorPosition	1Lag_consTructorWinningPosition	1Lag_aveLastlapPosition	1Lag_aveLastlapSpeed	1Lag_fastestLapRank	qualifyingPosition
1079	5	36	6	0	95	3	0	6.40	3.36	7	6
1080	8	46	6	0	120	3	0	11.02	4.66	2	8
1081	4	50	6	0	134	3	0	8.09	3.34	9	7
1082	3	62	6	0	161	3	0	6.33	6.00	3	4
1083	3	77	6	0	188	3	0	3.51	4.36	4	5

raceld	1-5LagPosition						1-5LagQPos					
1079	6	13	4	10	3	6	13	5	16	5		
1080	5	6	13	4	10	6	6	13	5	16		
1081	8	5	6	13	4	8	6	6	13	5		
1082	4	8	5	6	13	7	8	6	6	13		
1083	3	4	8	5	6	4	7	8	6	6		