



FORMULA 1

Machine Learning
Project

RAÍSSA ANDRÉ FRANCO

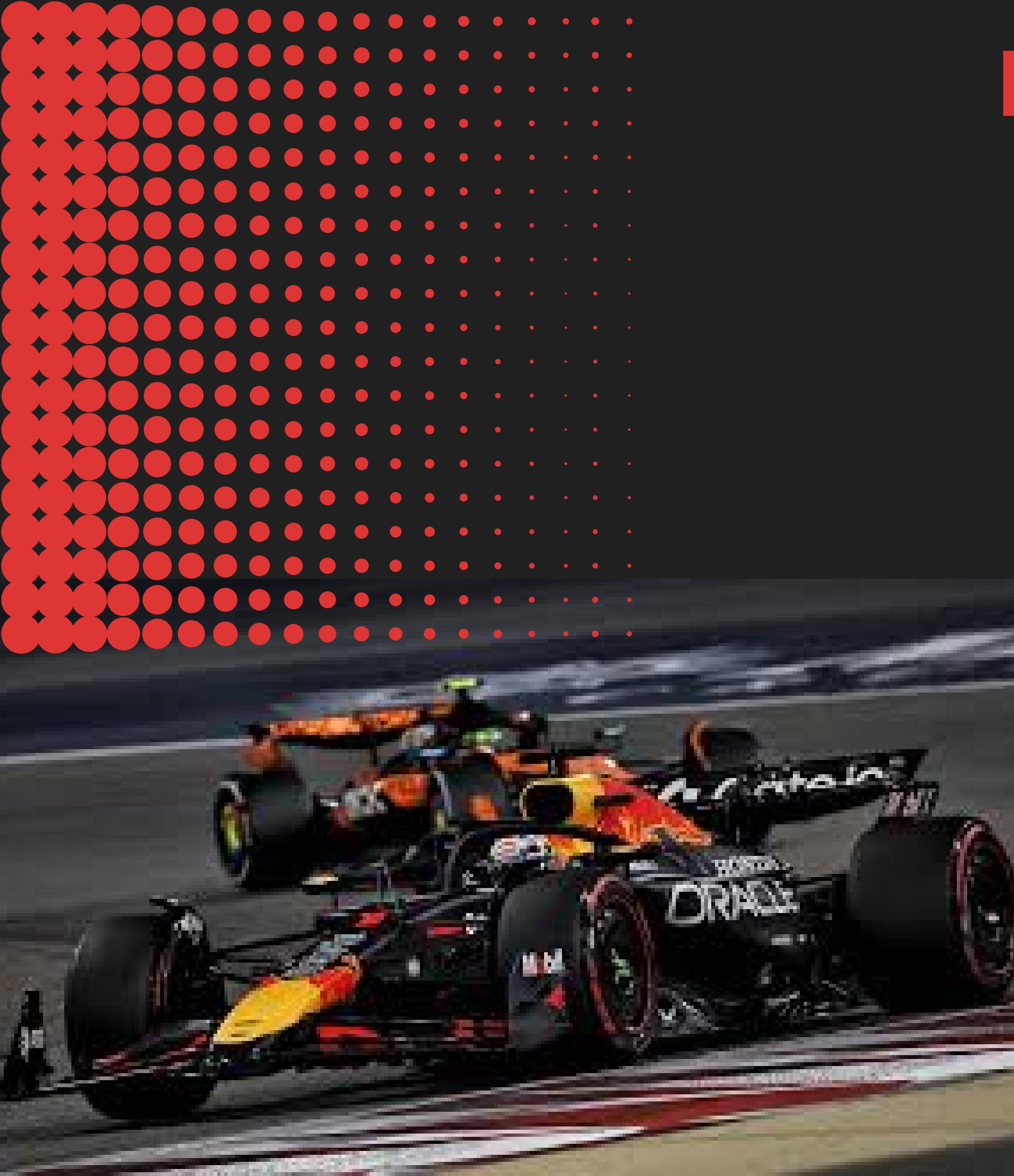
PROJECT OVERVIEW



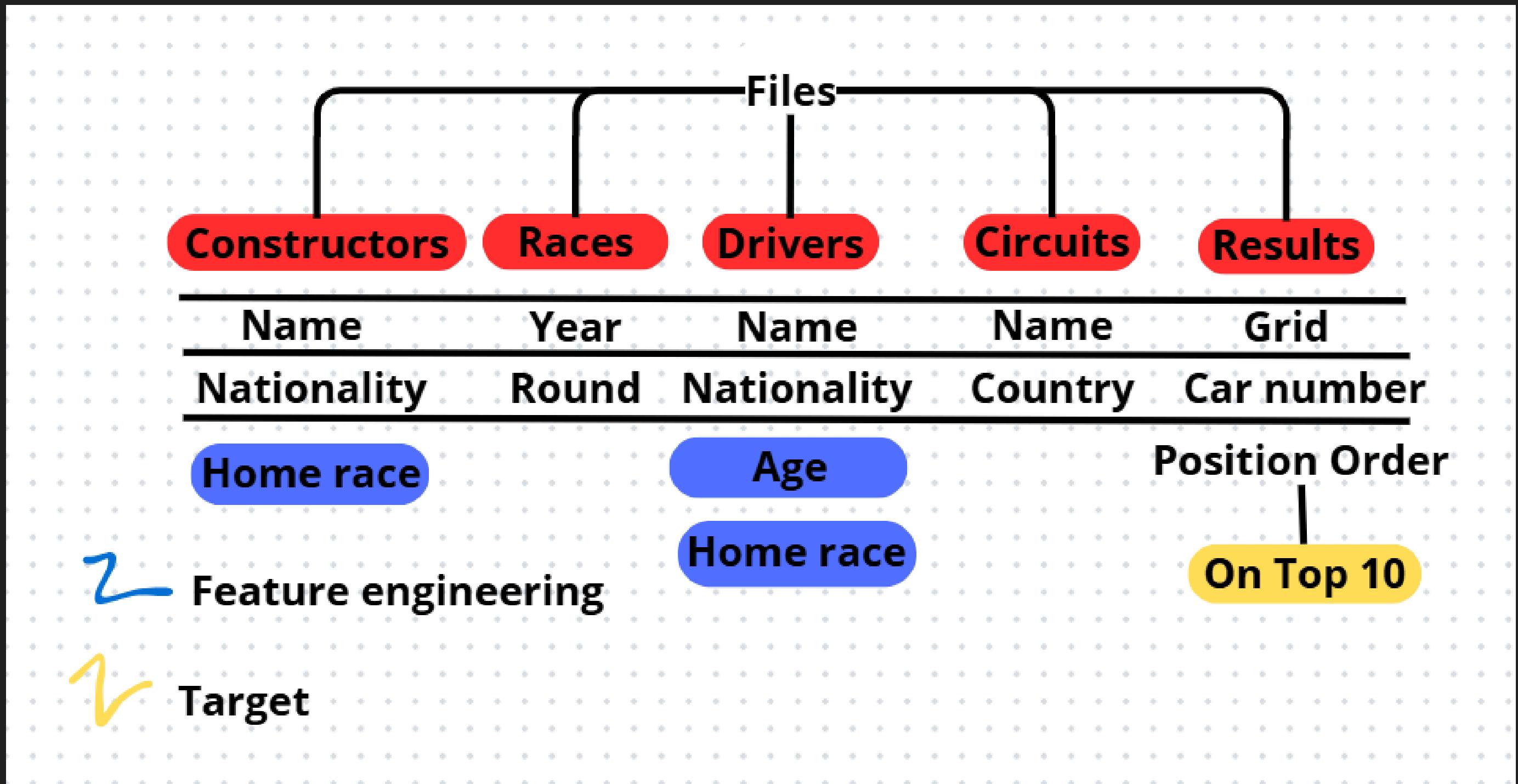
- Predict whether a driver will finish in the top 10 in a Formula 1 race.
- Analyze which features have the most influence on race outcomes.
- Compare the performance of different machine learning models.

DATASET

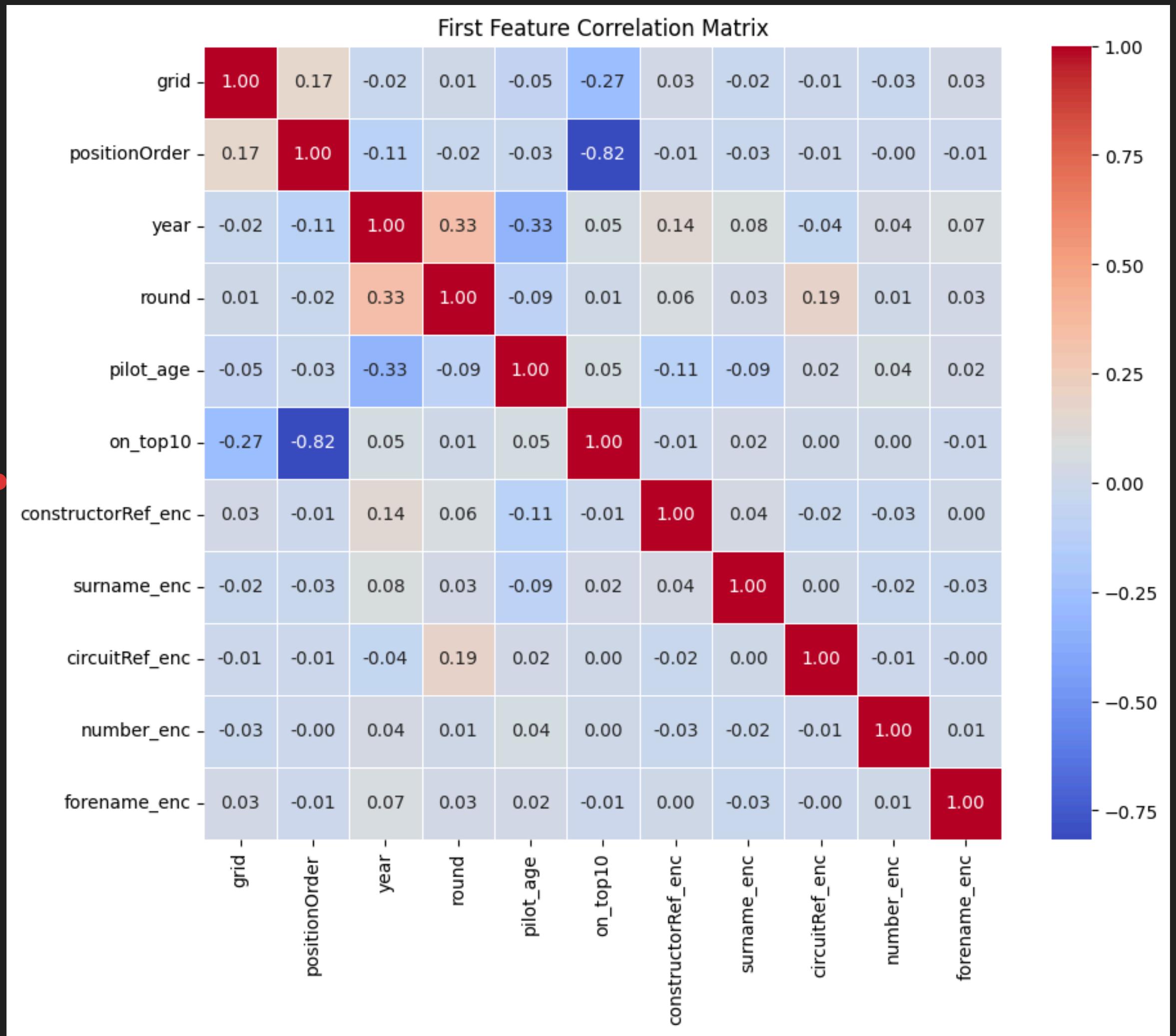
- Formula 1 Race Data
(Kaggle-Ergast)
- Data from 1950 – Present
- 14 files
- + 27.000 races



FEATURES



First Feature Correlation Matrix



EVALUATION

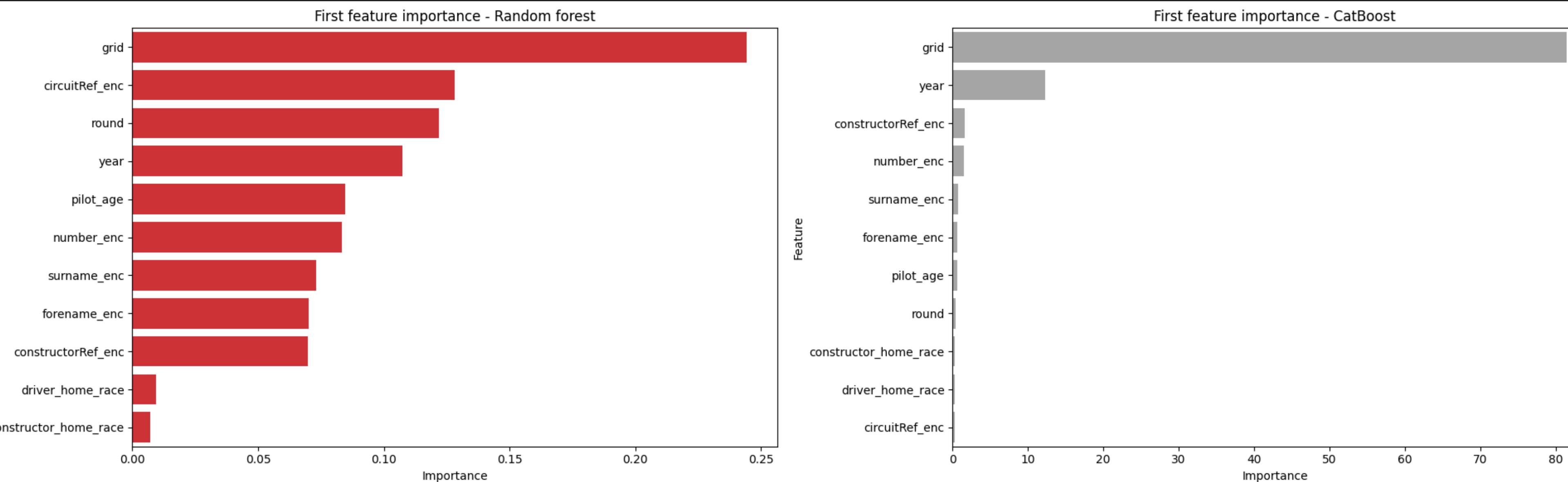
RANDOM FOREST

		precision	recall	f1-score	support
	0	0.72	0.76	0.74	3120
	1	0.64	0.59	0.62	2296
	accuracy			0.69	5416
	macro avg	0.68	0.68	0.68	5416
	weighted avg	0.69	0.69	0.69	5416

CATBOOST

		precision	recall	f1-score	support
	0	0.74	0.75	0.75	3120
	1	0.66	0.65	0.65	2296
	accuracy			0.71	5416
	macro avg	0.70	0.70	0.70	5416
	weighted avg	0.71	0.71	0.71	5416

FEATURE IMPORTANCE





+ FEATURE ENGINEERING

Driver and Constructor



Podium before



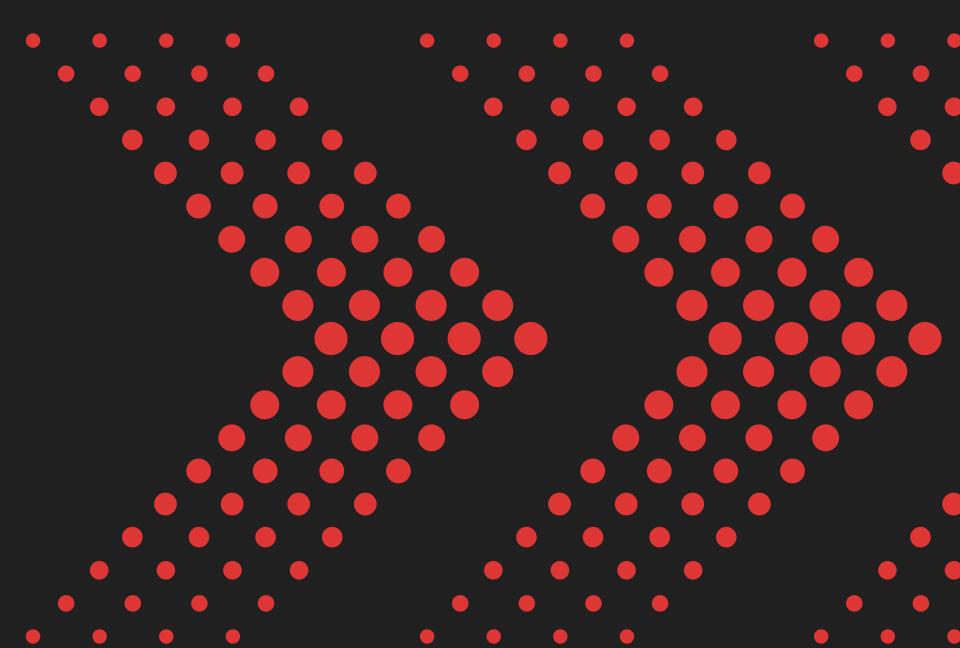
Races before

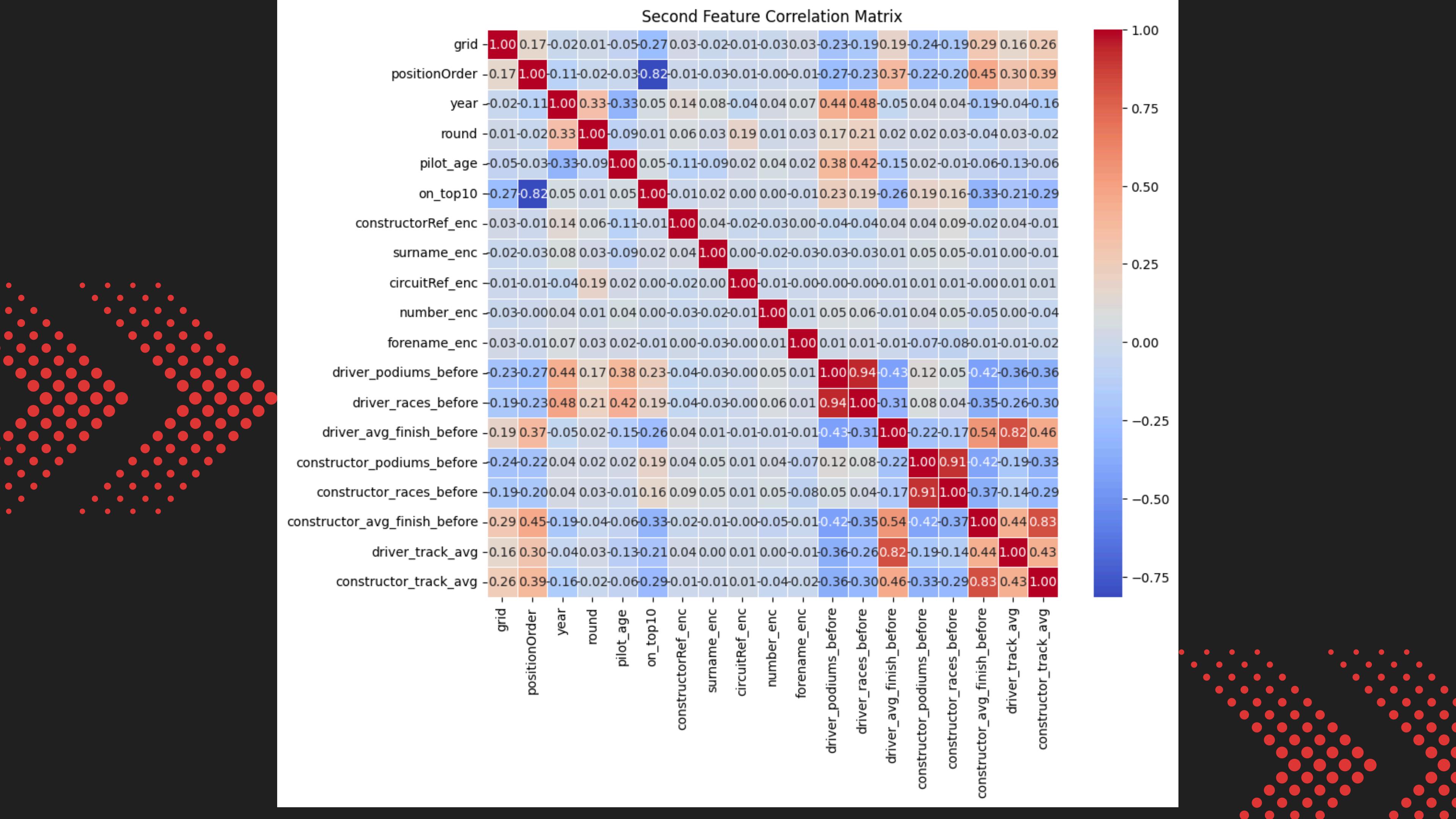


Avg finish before



Track avg before





EVALUATION

RANDOM FOREST

	precision	recall	f1-score	support
0	0.72	0.78	0.75	3120
1	0.67	0.60	0.63	2296
accuracy			0.70	5416
macro avg	0.70	0.69	0.69	5416
weighted avg	0.70	0.70	0.70	5416

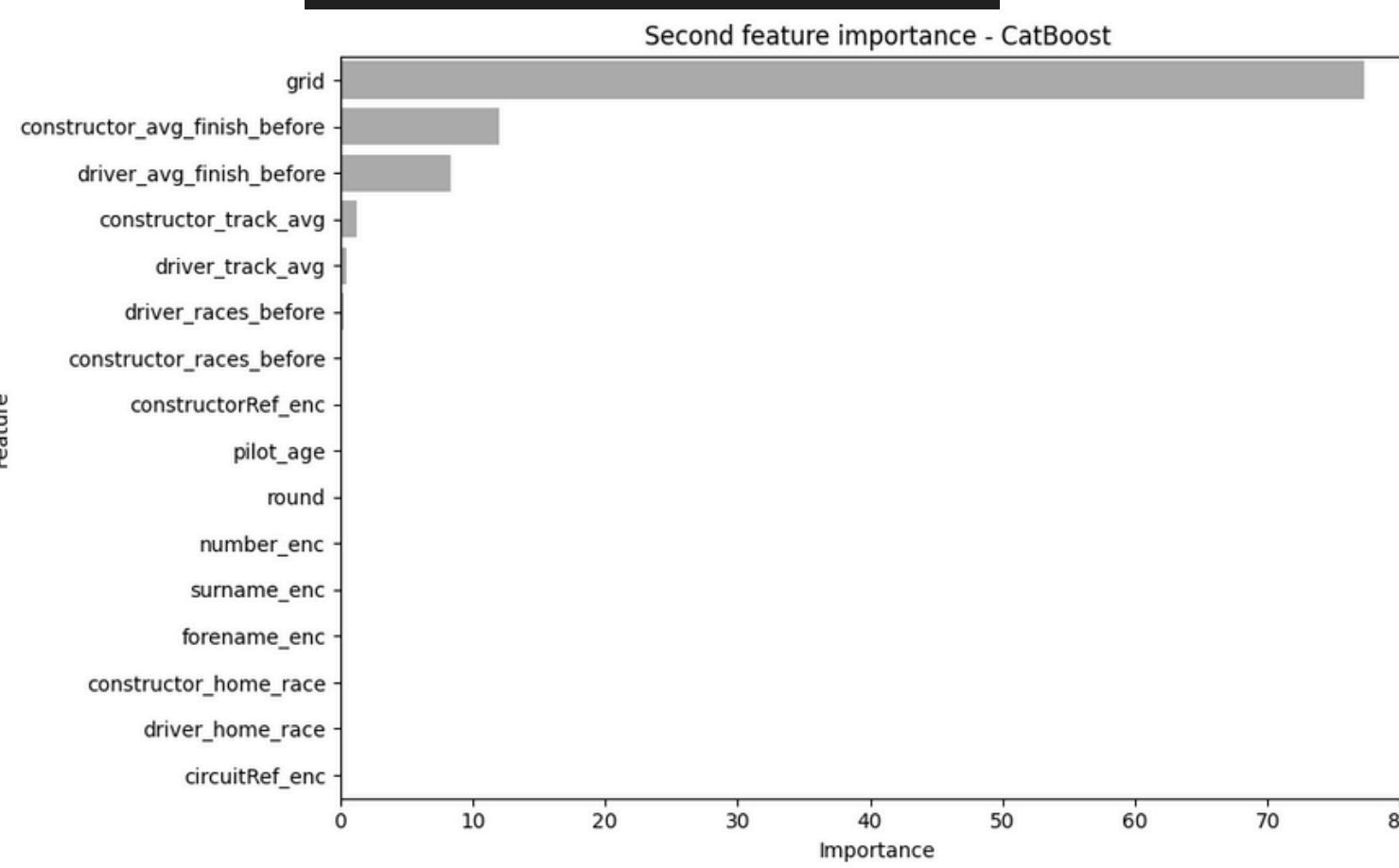
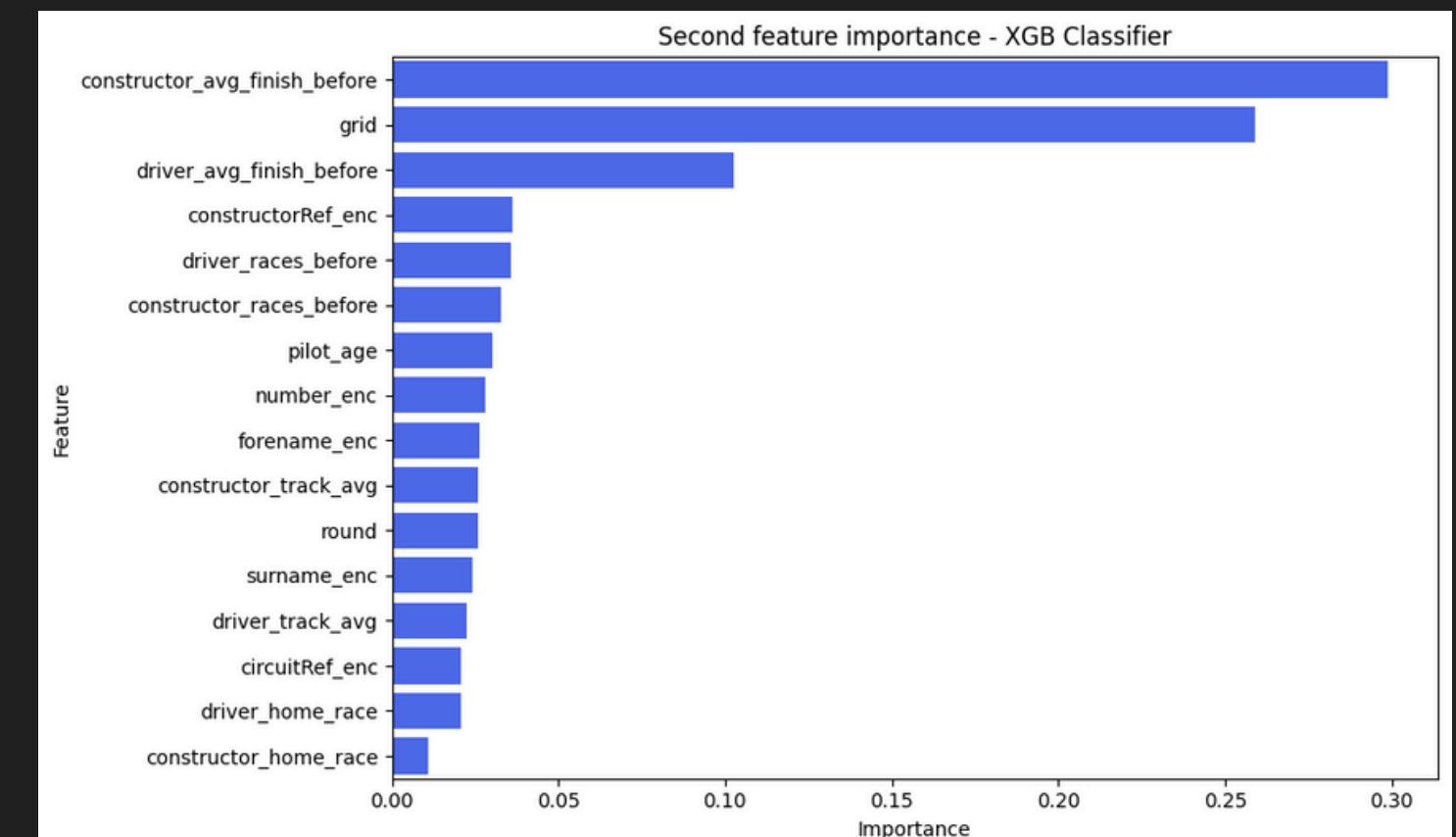
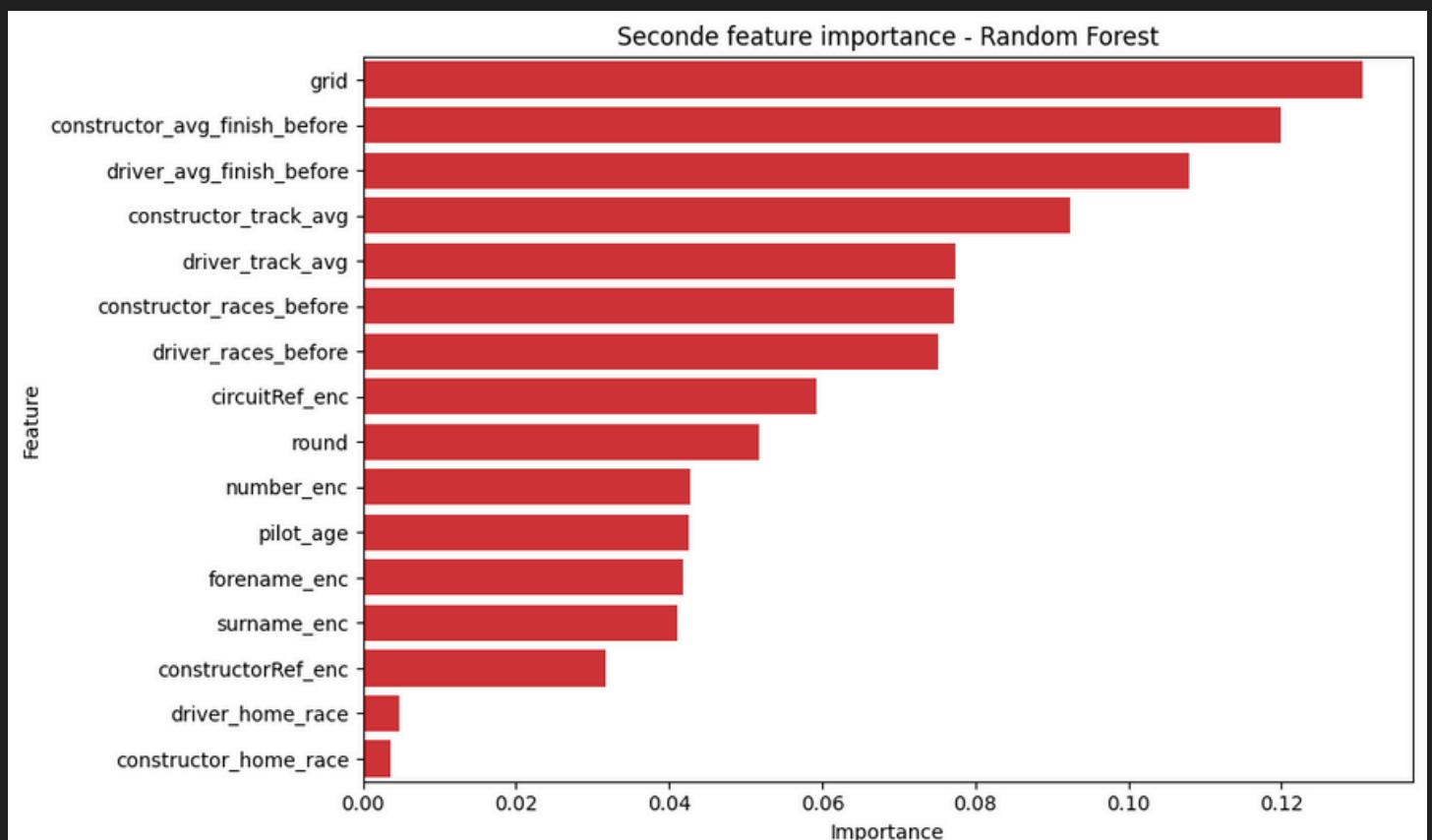
CATBOOST

	precision	recall	f1-score	support
0	0.74	0.77	0.75	3120
1	0.67	0.63	0.65	2296
accuracy			0.71	5416
macro avg	0.70	0.70	0.70	5416
weighted avg	0.71	0.71	0.71	5416

XGB CLASSIFIER

	precision	recall	f1-score	support
0	0.74	0.78	0.76	3120
1	0.67	0.62	0.65	2296
accuracy			0.71	5416
macro avg	0.70	0.70	0.70	5416
weighted avg	0.71	0.71	0.71	5416

FEATURE IMPORTANCE



DATA BALANCING



Techniques that combine oversampling with noise reduction to improve model performance on imbalanced datasets

EVALUATION

SMOTE + ENN

	precision	recall	f1-score	support
0	0.77	0.67	0.72	3120
1	0.62	0.72	0.67	2296
accuracy			0.69	5416
macro avg	0.69	0.70	0.69	5416
weighted avg	0.70	0.69	0.70	5416

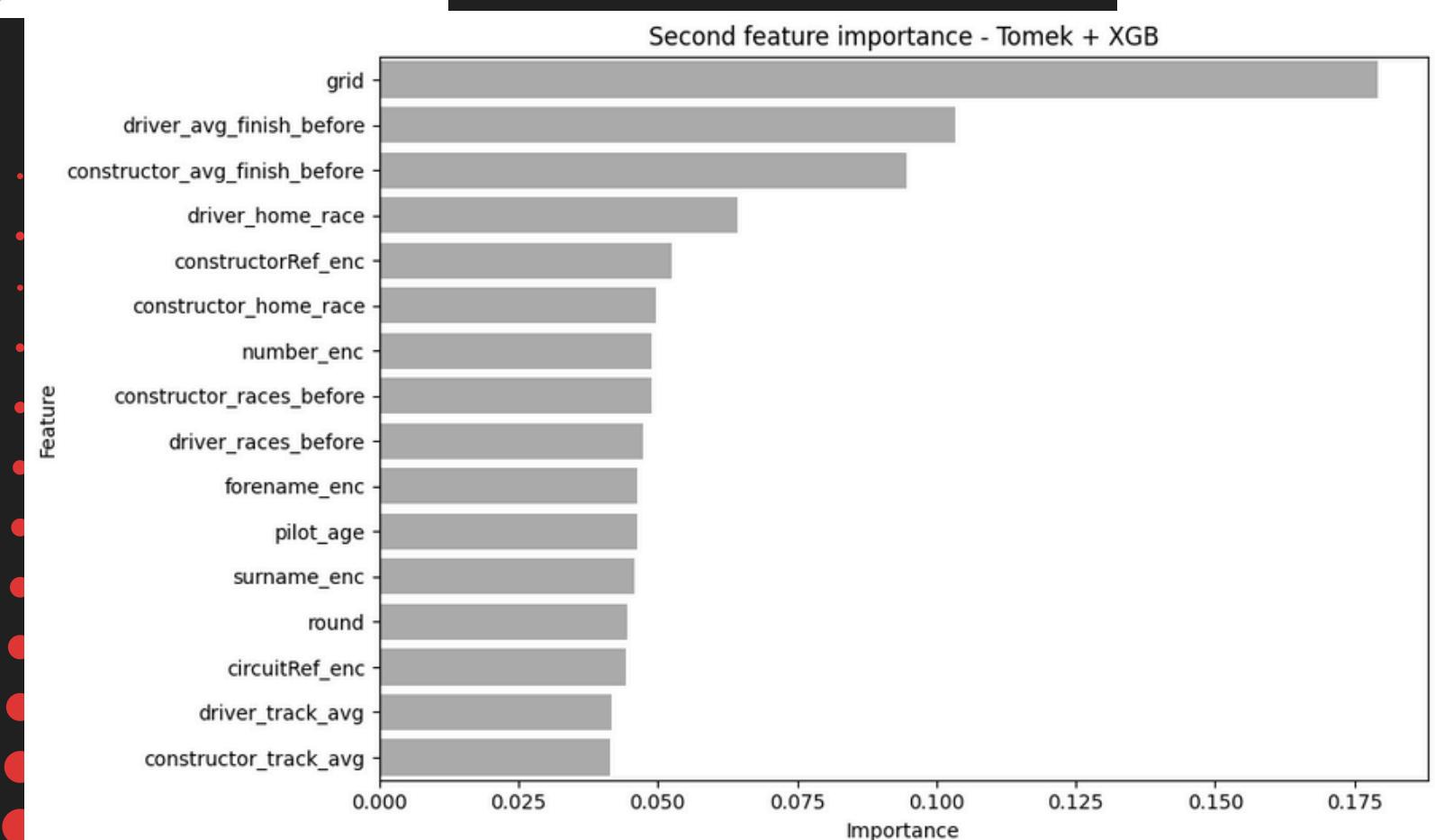
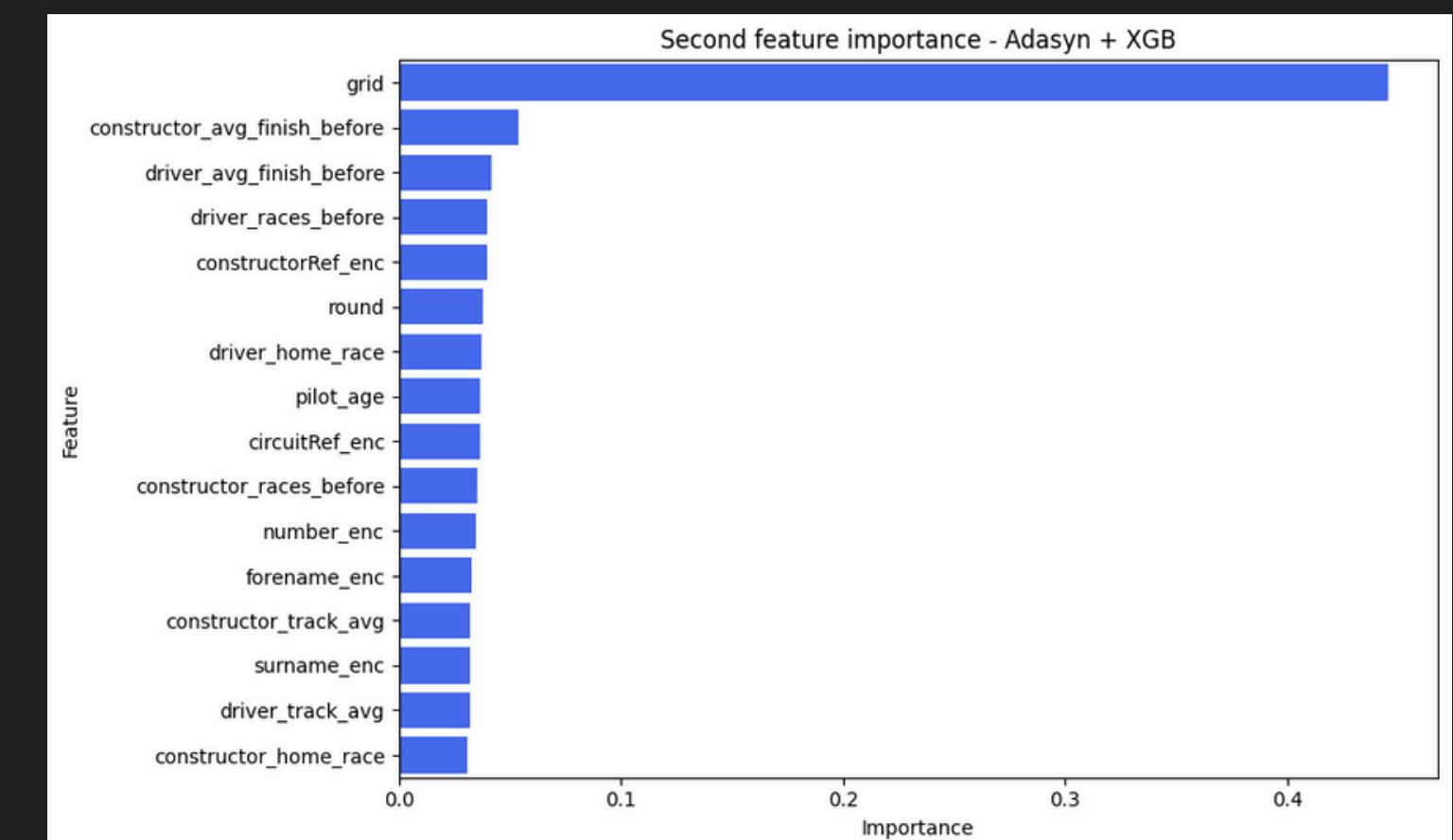
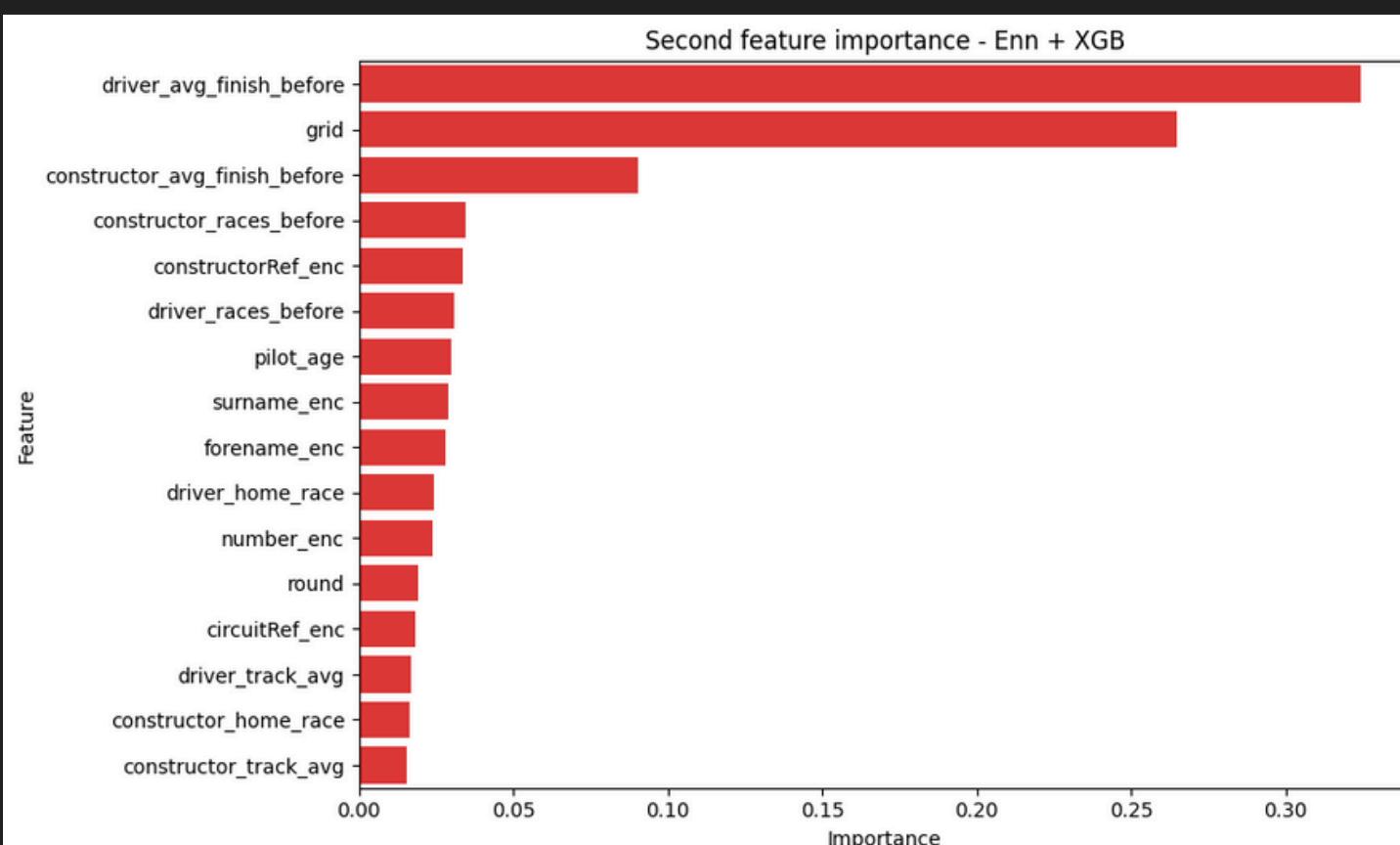
SMOTE + TOMEK

	precision	recall	f1-score	support
0	0.73	0.72	0.73	3120
1	0.63	0.64	0.64	2296
accuracy			0.69	5416
macro avg	0.68	0.68	0.68	5416
weighted avg	0.69	0.69	0.69	5416

ADASYN

	precision	recall	f1-score	support
0	0.76	0.69	0.73	3120
1	0.63	0.71	0.67	2296
accuracy			0.70	5416
macro avg	0.70	0.70	0.70	5416
weighted avg	0.71	0.70	0.70	5416

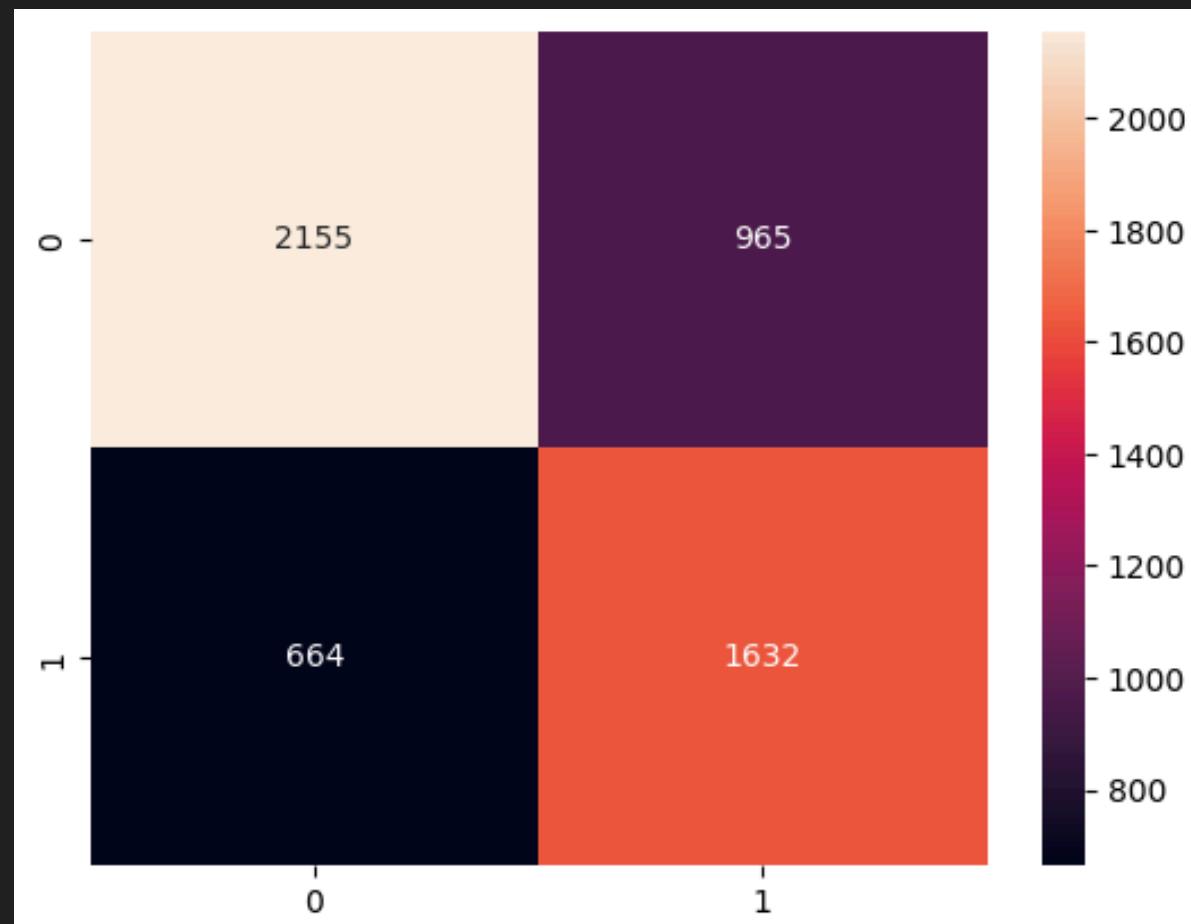
FEATURE IMPORTANCE



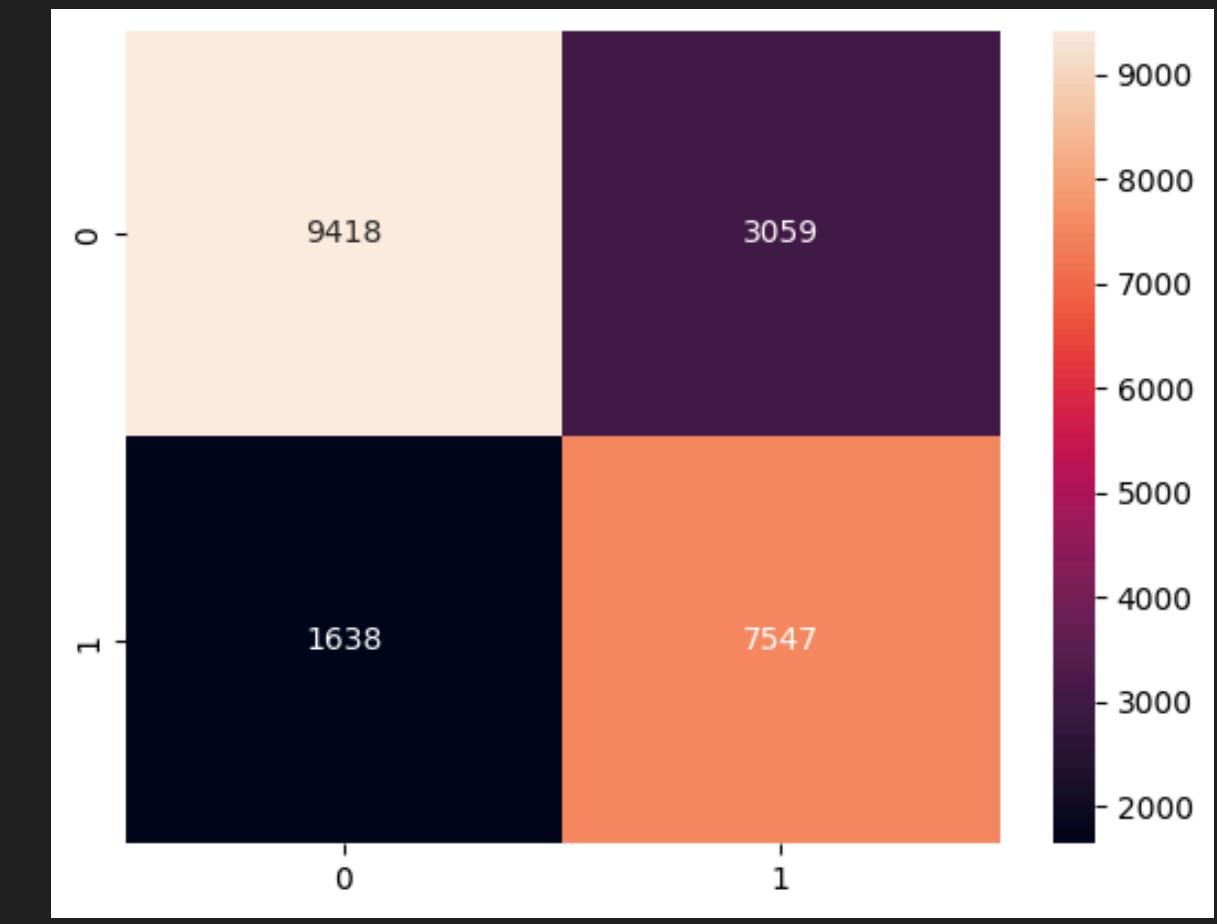
XGB + ADASYN

Test

	precision	recall	f1-score	support
0	0.76	0.69	0.73	3120
1	0.63	0.71	0.67	2296
accuracy			0.70	5416
macro avg	0.70	0.70	0.70	5416
weighted avg	0.71	0.70	0.70	5416



	precision	recall	f1-score	support
0	0.85	0.75	0.80	12477
1	0.71	0.82	0.76	9185
accuracy			0.78	21662
macro avg	0.78	0.79	0.78	21662
weighted avg	0.79	0.78	0.78	21662



Streamlit

Share     

Prediction: Driver in Top 10

Fill in the race information to predict if the driver will finish in the Top 10.

Driver's average previous finishes

- +

Grid position

Press Enter to apply - +

Team's average previous finishes

- +

Number of previous races for the driver

- +

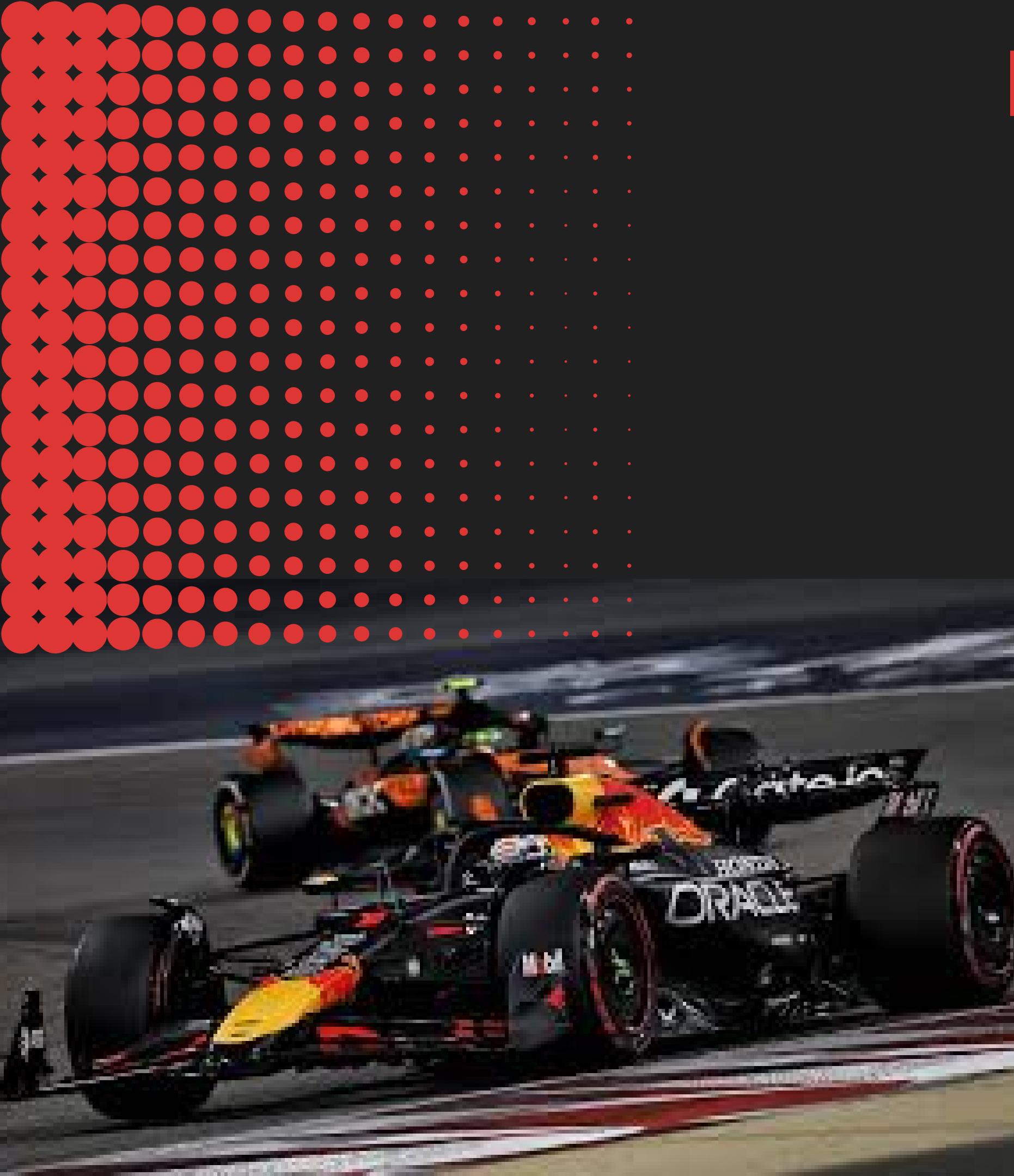
Number of previous races for the team

- +

[Manage app](#)

INSIGHTS

- Grid is the most important feature
- Balancing techniques
- Others features



FORMULA 1

Machine Learning
Project

SEPTEMBER 19, 2025



Thank you