



EDA - Sports

DATA SCIENCE LAB

WHO IS THE BEST F1 DRIVER?

14기 신태희 구기현 신동준 이건일
15기 김시원 박현진





CONTENTS

1. TOPIC SELECTION

2. BACKGROUND

3. DATASET USED

4. EVALUATION METRIC

5. DATA ANALYSIS

6. VALUATION

7. INSIGHTS

8. LIMITATIONS



WHY THIS TOPIC?

일반적인 평가는 주로

- 시즌 포인트
- 우승 횟수 / 포디움
- 챔피언십 순위

하지만 F1은 차의 성능 차이가 매우 커 같은 실력이라도 팀에 따라, 상황에 따라 결과가 왜곡됨.

→ 순수 드라이버 실력 평가 필요





EDA - Sports

WHAT'S DIFFERENT?

기준 평가	본 분석
결과 중심 (points)	과정 중심 (lap-level data)
차 성능 미통제	팀/차 효과 통제
정성적 인상 평가	정량적 성과 지표
단일 지표	다차원 성과 분해

단순히 “누가 잘한다”가 아니라 “누가, 왜 잘하는가”를 분석

드라이버의 순수 실력 순위는?





EDA - Sports

VOTE FOR THE WORLD CHAMPION



투표!!





F1 OVERVIEW

01 Qualifying-Race

- Qualifying (예선): Q1 → Q2 → Q3
각 라운드 당 가장 빠른 1랩 기준으로 레이스 출발 그리드 결정
→ 추월이 어려운 트랙일수록 예선이 매우 중요
- Race:
본선, 실제 점수 획득
연료 무거움, 타이어 마모, 추월, 사고 등 변수 다수





F1 OVERVIEW

02 Tyre Types

- Soft: 빠르지만 빨리 닳음
- Medium: 균형
- Hard: 느리지만 오래감
→ 레이스 중 최소 2종 사용 의무 (비 오는 경우 제외)

Stint:

한 타이어로 달리는 랩의 수





F1 OVERVIEW

03 DNF, Lap Deletion

- DNF (Did Not Finish, Retire):
사고, 기계 결함, 페널티 등으로 완주 실패
- Lap Deletion (랩 삭제):
4 타이어 모두 트랙 선 밖으로 나가면 해당 랩 무효



04 Car Performance

같은 팀 드라이버는 거의 동일한 차량, 하지만 팀 간 성능 차이 큼



DATASET USED

Fast F1

Formula 1 공식 API를 기반으로 한 Python 라이브러리
FIA / F1이 제공하는 공식 세션 데이터를 구조화하여 제공
팬 사이트 요약 데이터가 아니라 랩, 섹터, 타이어, 포지션 단위의 raw data 접근 가능

사용한 데이터 범위:

- 시즌: 2018-2024
기술 규정 안정화 구간, 하이브리드 시대 중 후반부 → 성능 비교 용이
- 세션: Qualifying, Race

Driver	DriverNumber	Team	LapNumber	LapTime	Position	Stint	Compound	TyreLife	FreshTyre	Sector1Time	Sector2Time	Sector3Time
ALO	14	McLaren	1.0	0 days 00:01:41.528000	10.0			TRUE			0 days 00:00:24.692000	0 days 00:00:38.129000
ALO	14	McLaren	2.0	0 days 00:01:31.565000	10.0	1.0	ULTRASOFT	1.0	TRUE	0 days 00:00:30.482000	0 days 00:00:24.593000	0 days 00:00:36.490000
ALO	14	McLaren	3.0	0 days 00:01:31.304000	10.0	1.0	ULTRASOFT	2.0	TRUE	0 days 00:00:30.619000	0 days 00:00:24.376000	0 days 00:00:36.309000
ALO	14	McLaren	4.0	0 days 00:01:30.551000	10.0	1.0	ULTRASOFT	3.0	TRUE	0 days 00:00:30.031000	0 days 00:00:24.386000	0 days 00:00:36.134000
ALO	14	McLaren	5.0	0 days 00:01:31.910000	10.0	1.0	ULTRASOFT	4.0	TRUE	0 days 00:00:30.524000	0 days 00:00:24.550000	0 days 00:00:36.836000
ALO	14	McLaren	6.0	0 days 00:01:31.731000	10.0	1.0	ULTRASOFT	5.0	TRUE	0 days 00:00:30.531000	0 days 00:00:24.504000	0 days 00:00:36.696000
ALO	14	McLaren	7.0	0 days 00:01:31.167000	10.0	1.0	ULTRASOFT	6.0	TRUE	0 days 00:00:30.449000	0 days 00:00:24.307000	0 days 00:00:36.411000
ALO	14	McLaren	8.0	0 days 00:01:30.926000	10.0	1.0	ULTRASOFT	7.0	TRUE	0 days 00:00:30.468000	0 days 00:00:24.244000	0 days 00:00:36.214000
ALO	14	McLaren	9.0	0 days 00:01:30.407000	10.0	1.0	ULTRASOFT	8.0	TRUE	0 days 00:00:30.157000	0 days 00:00:24.264000	0 days 00:00:35.986000

DATASET USED

2026-02-03 접속.

Penalty



<https://www.fia.com/documents/championships>

https://en.wikipedia.org/wiki/FIA_Super_Licence

Overtaking



<https://openf1.org/#overtakes>

<https://docs.fastf1.dev/>

Salary

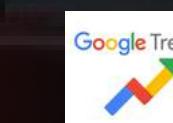


<https://racingnews365.com>

<https://www.spotrac.com/formula1/rankings/player>

<https://www.forbes.com/sites/brettknight/2025/12/09/formula-1s-highest-paid-drivers-2025/>

Popularity



<https://trends.google.co.kr/trends/>

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



EDA - Sports

EVALUATION METRIC

2

Pace



Tyre
Management



Stability



Overtaking



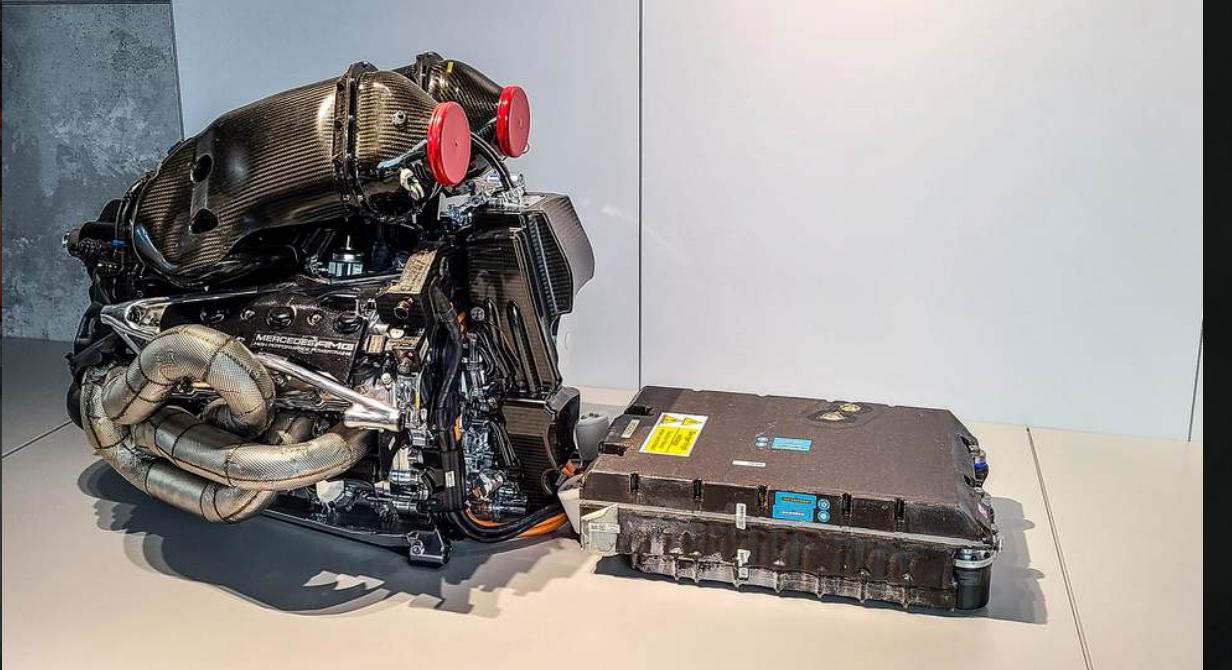


EDA - Sports

EVALUATION METRIC



Car Performance Control





EDA - Sports

EVALUATION METRIC



Car Performance Control

직선 최고 속도 -- 엔진의 성능



RACEFANS.NET

© XPB Images





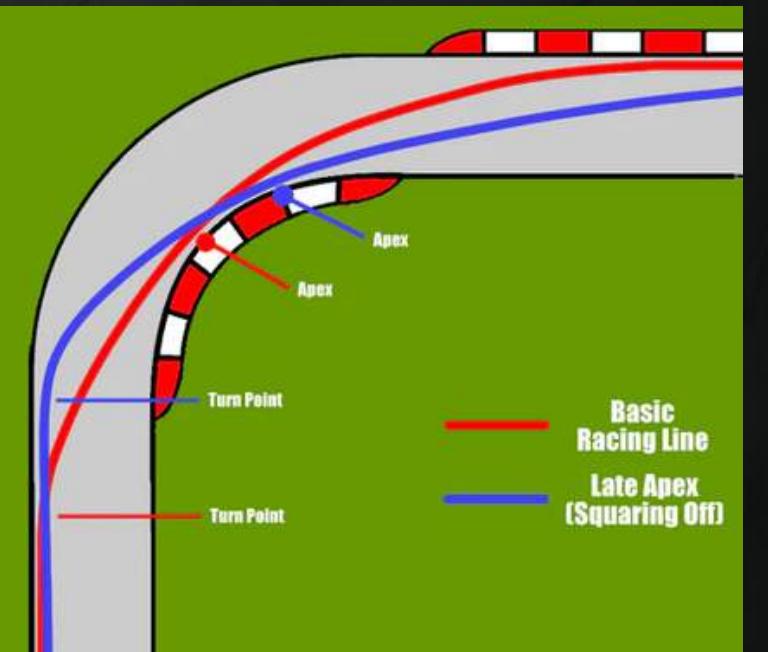
EDA - Sports

EVALUATION METRIC



Car Performance Control

코너 탈출 시 속도 == Aero Dynamic 성능





EDA - Sports

PACE

Qualifying Pace(QP)

$$QP_{i,r,t} = Q_{i,r,t}^{best} - median(Q_{r,t})$$

Race Pace(RP)

$$RP_{i,r,t} = \sum_{s=1}^{S_{i,r,t}} w_{i,r,t,s} \times mean(L_{i,r,t,s}^{corrected})$$

트랙별 비교를 위해 1-10점으로 Min-Max 정규화

→ Overall Pace Score = (QP Score + RP Score) / 2





TYRE MANAGEMENT

Standard Pirelli F1 Tire Compounds

Hard



C1

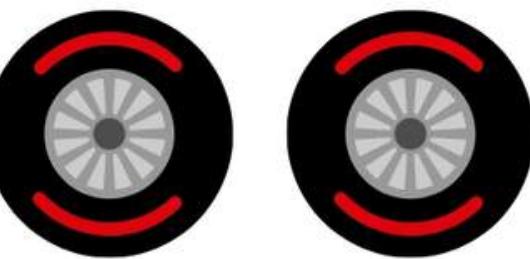
Durable, lower grip,
best for extreme
conditions and
longer races

Medium



C2

Soft



C3

Soft



C4

Less durable,
less grip for
increased speed
in shorter stints



C5



랩타임 변화에서 타이어 마모가 미치는 영향은 매우 큼. (Perantoni et al. , 2014)

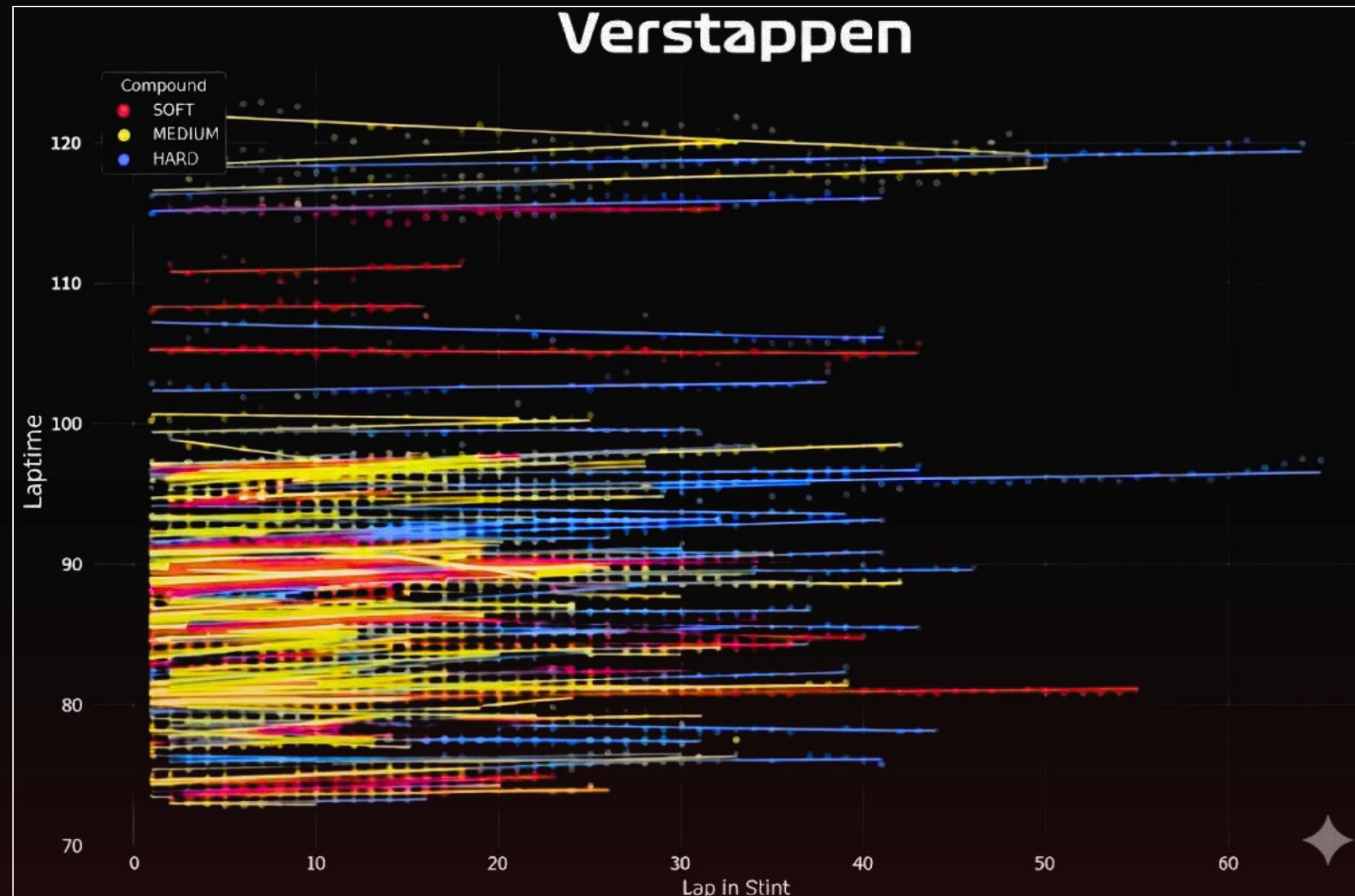
선수가 한 세트의 타이어로 주행한 랩타임 변화양상을 ‘타이어 관리 능력’으로 봄

만약 랩타임이 급격하게 올라간다면?
→ 타이어 관리를 못한 것

랩타임이 급격하게 올라가지 않는다면?
→ 타이어 관리를 잘함



TYRE MANAGEMENT



X: 스티트 내 주행 랩 수

Y: 랩타임 (트랙 길이에 따른 보정을 거침)

기울기: 타이어 노후에 따른 랩타임 변화율

예) 한 경기에 타이어를 3세트 사용했다면,
총 3개의 기울기 데이터 확보.



TYRE MANAGEMENT

타이어 종류	Mean	Std.Dev	Median	Min	Max
HARD	0.37	4.28	0.03	-14.24	77.10
HYPERSOFT	1.72	9.41	0.03	-2.13	58.58
INTERMEDIATE	4.92	15.89	-0.10	-3.65	73.47
MEDIUM	0.46	4.93	0.02	-45.67	67.16
SOFT	1.33	8.25	0.02	-35.66	63.34
SUPERSOFT	0.20	2.98	0.00	-1.31	43.59
ULTRASOFT	0.49	4.82	-0.03	-2.71	45.15
WET	12.75	24.37	-0.96	-4.27	82.04

STABILITY

사고·리타이어·운영상 문제 없이 얼마나 안정적으로 완주하는지를 측정

$$\text{Instability} = 0.6 \cdot \text{DNF Rate} + 0.4 \cdot \text{Deleted Lap Rate}$$

1. DNF Rate

- 시즌 중 출전 레이스 대비 완주 실패 비율
- 선수 실책으로 인한 DNF만 포함

2. Deleted Lap Rate

- 폐널티, 트랙 리미트 위반 등으로 무효 처리된 랩 비중
- 레이스 운영 규율 측면의 불안정성 반영



OVERTAKING

Overtaking 능력 = 추월 성공 + 추월 방어

- 얼마나 추월 기회가 있을 때 이를 잘 캐치했는지?
- 얼마나 상대의 추월 시도를 잘 방어해냈는지?

Track Difficulty Weight : Track Weight_{race} = $\frac{\text{Average Overtakes for the Year}}{\text{Overtakes in this Specific Race}}$





OVERTAKING

Total Score : $\sum(21 - \text{Position}) \times \text{Track Weight}_{\text{race}}$



가중치 →
평균 대비 얼마나 추월이 적게/많이 일어났는가?
를 이용하여 추월과 가중치를 반비례 관계로 설정.

상위 포지션일수록 추월이 어렵기 때문에,
추월 후 position 값에 따라 점수를 선형적으로 설정



DATA PREPROCESSING

각각의 실력 지표 계산식으로 계산한 지표를 min-max scaling 후 0~10으로 표준화



STABILITY

베이스 테이블에 포함된 Instability를 min-max(0~10) 후 반전($10-x$)



OVERTAKING

추월 데이터가 없는 행은 0으로 채워서 진행



DRIVER&EVENT

드라이버는 전체 이름 대신 중계용 식별 코드, 그랑프리는 연도별 매핑



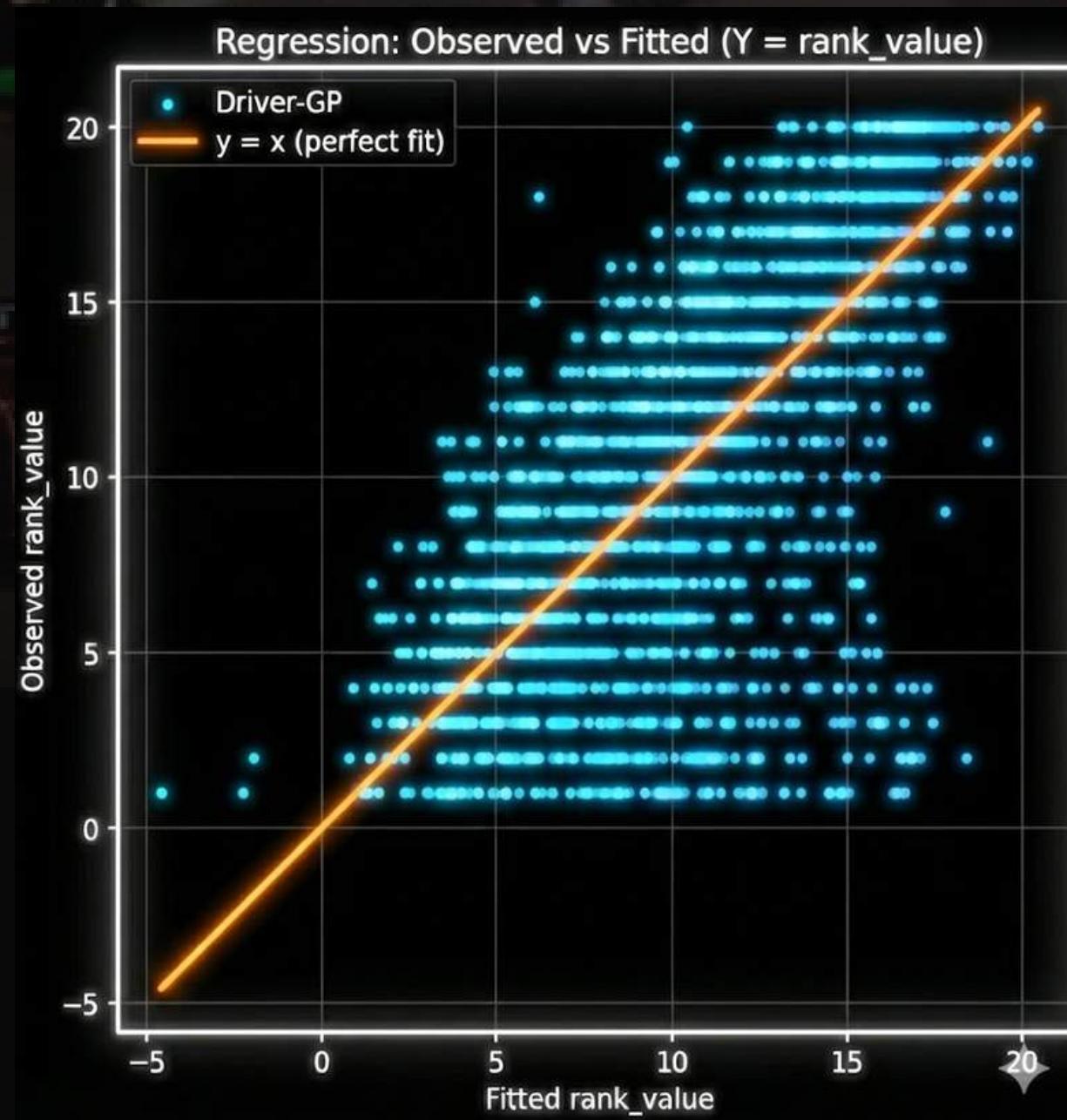
RANKING

작을수록 큰 결과치가 도출되는 RANK VALUE 정의





MODEL FITTING



- Y(종속변수): rank_value
- X(설명변수): Tyre_Mgmt, overall_pace_score, overtake_total_score, Stability (모두 0~10 min-max 스케일)
- 컨디셔닝/교란변수: Car_Score_Scaled_GP – 차량 성능

항목	값
관측 수	2,758 (Driver-GP 행, 결측 제외)
R^2	0.538
Adj. R^2	0.537
F-statistic	640.6



SLOPE INTERPRETATION

변수	coef	P> t	VIF
Tyre_Management_min max_0_10	0.8354	0.000	1.0
overall_pace_score_min max_0_10	1.3186	0.000	1.44
overtake_total_score_mi nmax_0_10	0.4592	0.000	1.1
Stability_minmax_0_10	0.5262	0.000	1.31
Car_Score_Scaled_GP (교 란)	0.0126	0.674	1.16

- 차량 성능 통제 후에도 타이어·페이스·추월·안정성은 모두 순위(rank_value) 개선에 유의한 영향 (다중공선성 문제 없음)
- 효과 크기:
페이스(1.32) > 타이어(0.84) > 안정성(0.53) > 추월(0.46)
- 차량 변수: Car_Score_Scaled_GP는 비유의 ($p=0.67$)
→ 실력지표 효과 해석 타당



EDA - Sports

DRIVER STYLE CLASSIFICATION

전체 표본에서 추정한 기준 회귀기울기와
드라이버별 표본에서 추정한 회귀기울기를 비교해
평균 대비 영향력이 큰 지표를 가진 드라이버를 식별 분류



Driver	Strength	Driver	Strength
ALB	Tyre	SAI	Tyre
ALO	Tyre	SAR	Overtake
BOT	Stability	STR	Tyre
GAS	Tyre	TSU	Overtake
GIO	Overtake	VER	Pace
GRO	Overtake	VET	Stability
HAM	Stability	ZHO	Overtake



DRIVER HIDDEN ABILITY

- 모델의 지표로 설명되지 않는 선수 고유의 잠재력 (Hidden Ability) 을 추정
- 드라이버별 더미변수를 포함한 회귀모형을 적합한 후 드라이버 더미 계수 크기 비교
→ 지표에 잡히지 않는 드라이버 고유의 잠재 역량 평가

$$Y = \beta_1 \cdot pace + \beta_2 \cdot tire + \beta_3 \cdot stability + \\ \beta_4 \cdot overtake + \text{Dummies}_{driver}$$

Driver	Hidden	Driver	Hidden
VER	4.599	NOR	3.031
HAM	4.367	ZHO	2.875
OCO	3.650	ALO	2.820
LEC	3.505	GAS	2.739
SAI	3.408	RIC	2.615
PER	3.379	VET	2.604
BOT	3.157	STR	2.483



DRIVER PERFORMANCE SCORE

전체 표본 회귀모형에서 추정한
기울기(계수) 비율을 각 실력지표의 가중치로
설정한 후 드라이버별 정규화된
실력지표(차량 성능 제외)에 가중치를 곱해
합산한 뒤, 총점이 100이 되도록 재정규화

→ 차량 성능을 통제한 드라이버별 종합적인 실력
평가

Driver	n_races	mean_Tyre_0_100	mean_Pace_0_100	mean_Overtake_0_100	mean_Stability_0_100
VER	142	87.9683	89.7474647544	18.9324807543	88.9350920106
HAM	144	87.8548571279	84.7954751516	21.7020110792	94.3076389472
PIA	45	88.1584871030	68.29581	42.3510282692	82.6767207997
LEC	140	87.42663	73.1976443178	21.3218648924	83.4385865843
SAI	138	88.10309	68.1339348445	21.9482717658	86.8494077309
NOR	125	87.9945237305	67.0450494393	22.2654248256	87.7849622025
PER	141	88.0295989836	64.4722563280	23.6805098153	84.6413931347
VET	99	88.0106367232	64.1952773514	11.9488973145	86.1540060446
BOT	146	87.40168	63.2936828395	16.9108876791	70.4884164964
RUS	126	87.95782	53.1389619085	21.8028208256	84.4341457138
ALO	108	87.4449440405	54.8899203332	25.2189114639	76.9518112609



EDA - Sports



CHAMPION RESULT





EDA - Sports



SALARY VS. SKILL



$$\text{Salary}_i = \alpha + \beta \cdot \text{Skill}_i + e_i$$

- 회귀선 위: overpaid
- 회귀선 아래: underrated

Spearman ρ (연봉순위 vs 실력순위) = 0.754

(Salary vs Hidden: $\rho=0.593$)



SALARY VS. SKILL



Overpaid / Premium

- Hamilton, Verstappen
- 실력 최상위 + 연봉 최상위
→ 가성비는 낮지만 전략 자산



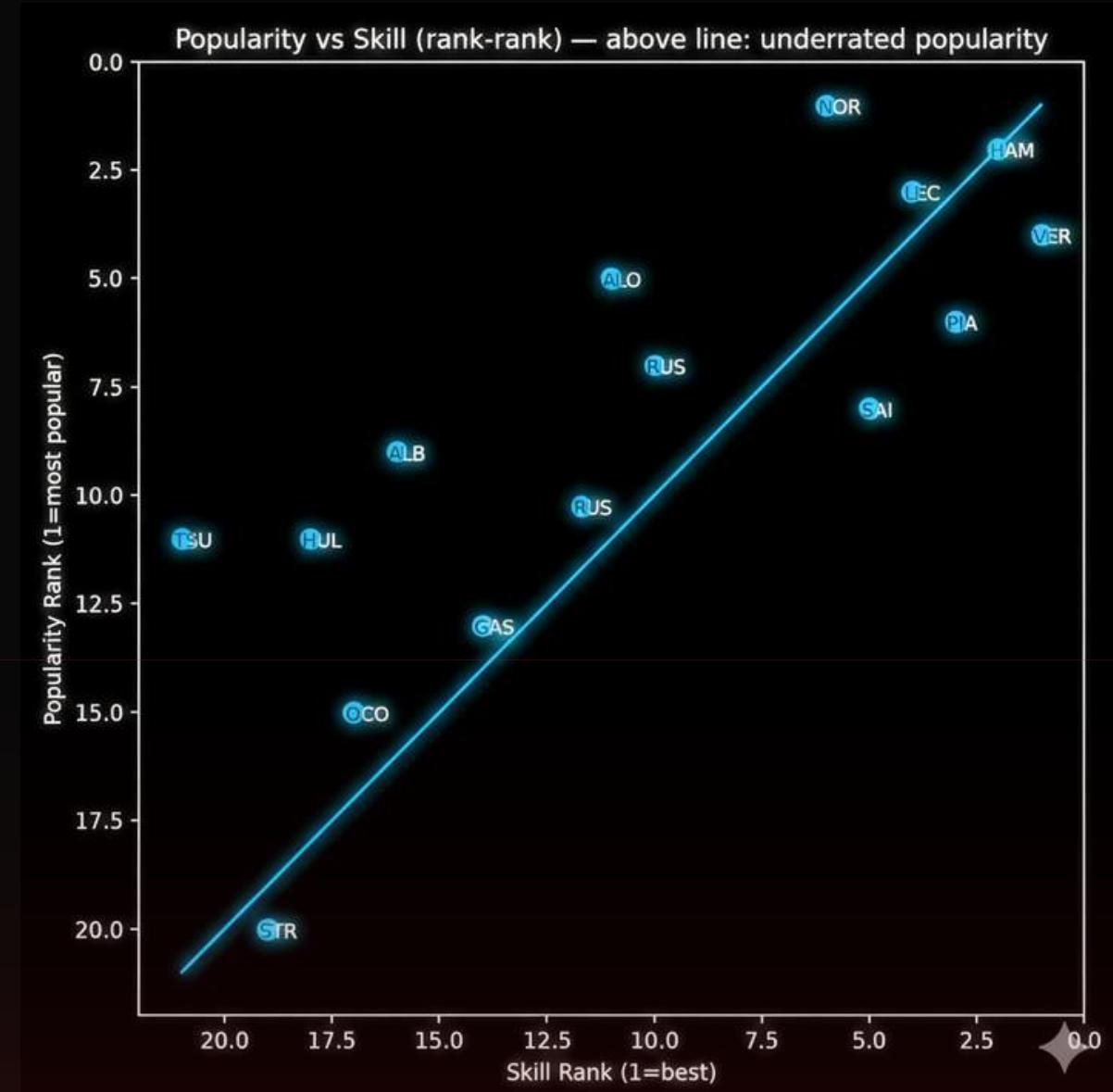
Underrated

- Piastri, Sainz
- 실력은 상위인데 연봉이 아직 낮음
→ 팀 입장에서 가장 매력적인 계약





POPULARITY VS. SKILL



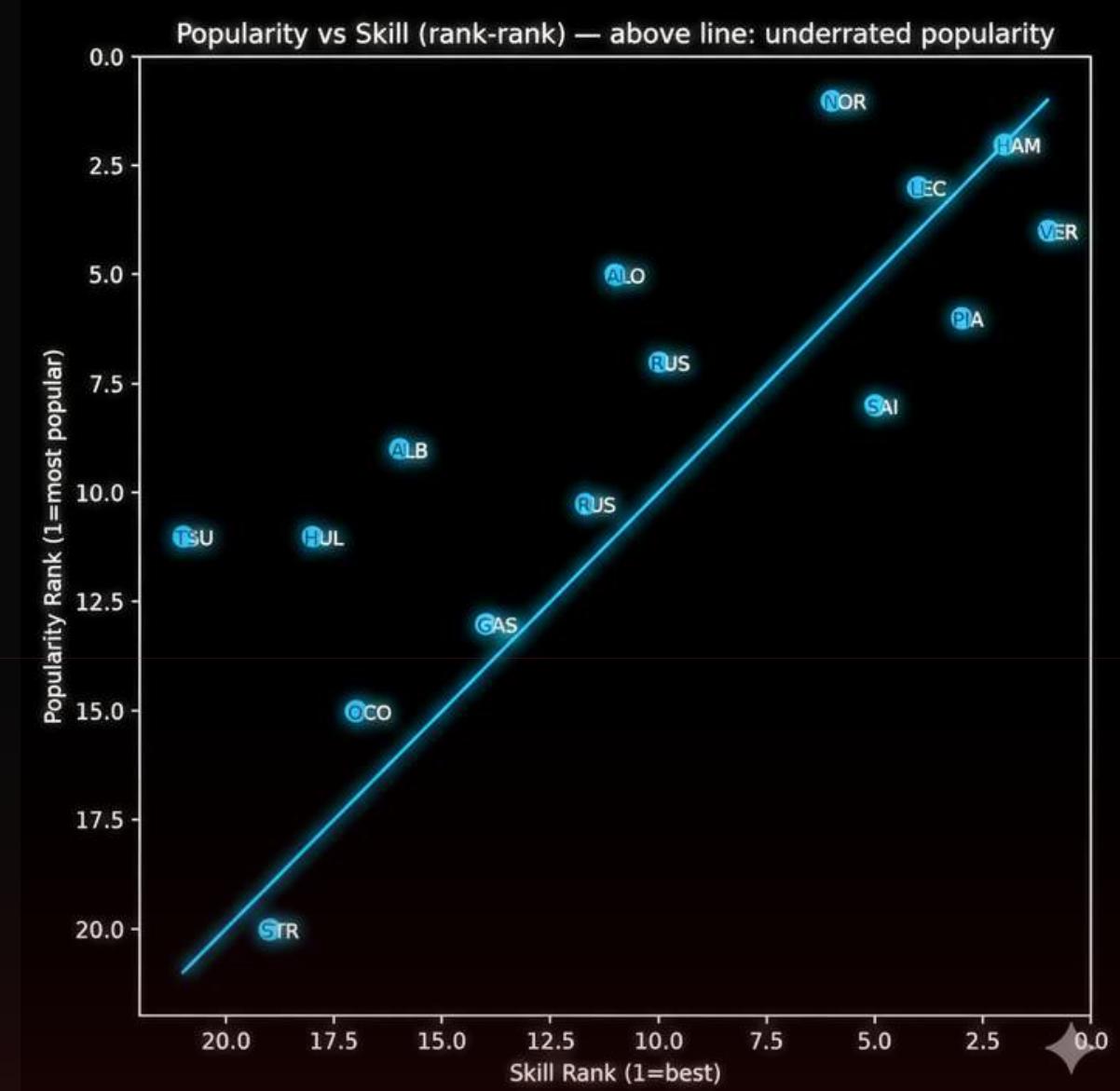
Rank-Rank (둘 다 1등이 best) + $y=x$
• 선 위/아래로 “실력 대비 인기가 과대/과소”를 바로 보여
줌

Spearman ρ (인기순위 vs 실력순위) = 0.799
→ 인기 역시 실력과 꽤 동행
(Popularity vs Hidden: $\rho=0.466$)

라인에서 벗어난 거리 = 인기 프리미엄 / 디스카운트



POPULARITY VS. SKILL



Overhyped (인기 > 실력)

- 특정 스타 드라이버 Alonso, Tsunoda
→ 마케팅 가치는 크지만 성과 리스크



Under the radar (실력 > 인기)

- Sainz, Piastri 등
→ 실력 대비 미디어 노출 부족
→ 저평가된 브랜드 자산





INSIGHTS

누가 제일 좋은 드라이버인가?



COMPLETE:
H & S 모두 높음
→ VER, HAM 등
팀의 핵심 자산



RAW TALENT:
H 높지만, S 낮음
→ OCO: 환경이 바뀌면 실력
급상승 가능성, 투자 기회





INSIGHTS

누가 제일 좋은 드라이버인가?



EXECUTOR:

S 높지만, H 낮음

→ RUS, PIA 등

현재 환경에서는 실력발휘 잘 함,
전략적으로 활용가능한 세컨 드라이버



DEVELOPMENT:

H & S 모두 낮음

→ SAR, MAG, TSU 등 최저 실력,

F1레벨을 따라오지 못함,

새로운 드라이버에게 투자하는 것이 나을 수 있음

OR 저비용 옵션으로 보유





INSIGHTS

Hidden Ability & Skill Score 활용

영입 의사결정 시 실력적으로 고려할 사항

Hidden >> Skill

- 해석: 환경에 의해 실력이 가려진 "원석"
- 예시: OCO (Hidden 3위, Skill 17위), ZHO (Hidden 9위, Skill 25위)
- 전략: 좋은 차량/환경 제공 시 폭발적 성과 기대
- 리스크: 낮음 (잠재력 검증됨)

Skill >> Hidden

- 해석: 측정 가능 기술은 있으나 레이스 지능 부족 "실행형"
- 예시: RUS (Hidden 20위, Skill 10위)
- 전략: 세컨드 드라이버 또는 팀 플레이어 역할
- 리스크: 에이스로 기용 시 기대 미달

Hidden ≈ Skill (둘 다 높음)

- 해석: "완성형 드라이버"
- 예시: VER, HAM, LEC, SAI
- 전략: 핵심 자산으로 장기 계약

Hidden ≈ Skill (둘 다 낮음)

- 해석: 전반적 실력 향상 필요,
- 예시: MSC, LAT, SAR
- 전략: 신인 육성 또는 드라이버 교체





DEEP INSIGHTS

Skill Score + Hidden 데이터 어떻게 활용 가능할까?

공식

드라이버 가치 = $\alpha \times \text{Hidden} + \beta \times \text{Skill Score} + \text{Brand Premium}$

영입 체크리스트

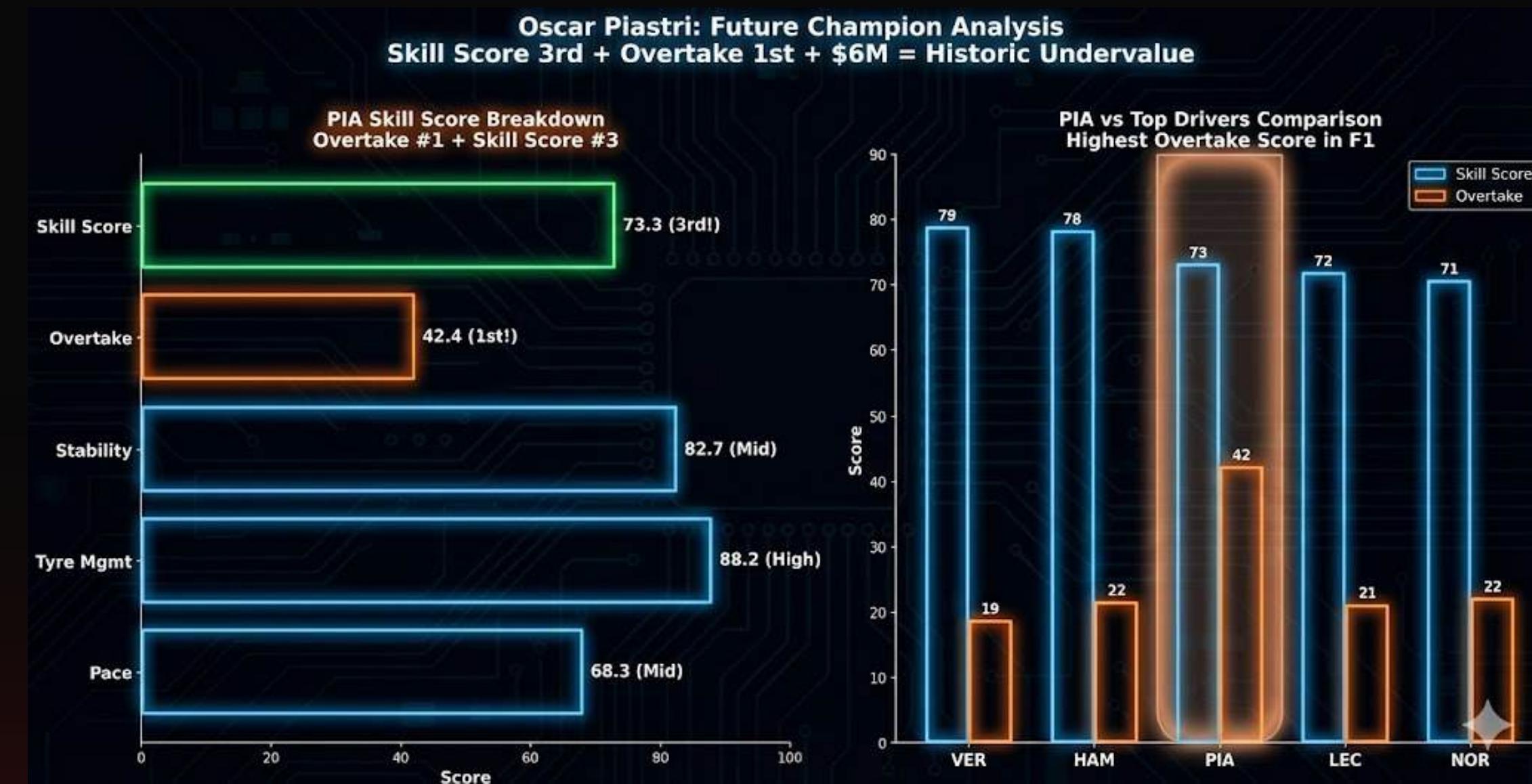
- Hidden Ability 순위는? (숨겨진 실력, 레이스 지능)
- Skill Score 순위는? (측정 가능 종합 실력)
- Hidden vs Skill Score 비교 (원석인가? 실행형인가?)
- Skill score의 4가지 관측 지표 중 우리 팀에 필요한 강점이 있는가?
- 연봉 요구 수준 vs Hidden + Skill 순위 비교





DEEP INSIGHTS

다음 세대 스타 챔피언은 누구인가?





DEEP INSIGHTS

다음 세대 스타 챔피언은 누구인가?



- skill score **전체 3위**
- 24시즌 종료 시점 기준 f1 데뷔 2년만에 전체 관측 가능한 실력 지표 3위에 등극
- 주목할 점은 **overtake 부문 전체 1위**
→ 추월을 가장 잘하는 선수로 추월이 주는 카타르시스가 큰 f1에서 매우 큰 팬덤 형성 예측 가능
- 현재 6M의 salary로 매우 좋은 가격으로 McLaren이 보유 중이며 성장폭과 실력을 미루어보았을 때 미래의 챔피언 가능성 높후
- 따라서 주목해야 할 선수!



EDA - Sports

DEEP INSIGHTS

Red Bull Racing

- VER: \$55M(모델 예측\$41.3M + 챔피언 프리미엄 \$13.7M)
→ 프리미엄은 브랜드 가치 + 마케팅 효과
- PER \$14M(모델 예측 \$21.8M)
→ \$7.8M 저평가
→ PER은 실력 대비 효율적 계약



전략 1: PER 유지 + 역할 명확화

- PER은 \$14M으로 hidden 6.94 확보 → 우수한 가성비
- PER을 확실한 "팀 플레이어"로 재정의
- 컨스트럭터 포인트 안정화 우선

전략 2: 세컨드 드라이버 업그레이드

- 영입 타겟: SAI (Hidden 7.01) 또는 OCO (Hidden 7.62)
- OCO 영입 시: \$6M으로 Hidden 7.62 확보 = PER 대비 +0.68점 상승
- 목표: VER 지원 + 독자 포인트 획득



DEEP INSIGHTS

McLaren

- PIA: \$9.1M 저평가 = F1 역사상 최고 가성비 드라이버
- PIA Skill 73.27 (3위) + Overtake 42.35 (1위) = 공격형 스타
- NOR은 모델 예측과 거의 일치 = 적정 연봉
- 팀 합산 \$26M으로 Skill 144.03 확보 = 최고 효율



전략: 현 체제 유지 + 장기 Lock-in

- PIA 장기 계약 필수 (2027년 이후까지)
- PIA 연봉 인상 불가피: \$6M → \$12-15M 예상
→ 모델 예측 \$15.1M 기준 협상
- NOR와의 건강한 경쟁 환경 유지

투자 우선순위

1. PIA 계약 연장 (가장 중요)
2. 차량 개발 (경쟁력 유지)
3. 엔지니어링 역량 강화

비용 구조 예측

- NOR: \$20M 유지 (적정)
- PIA: \$6M → \$12-15M (인상 불가피)
- 목표 총액: \$32-35M
- ROI: Skill 144.03 / \$35M = 4.1점/\$M (최고 효율)



LIMITATIONS

추월 데이터 한계: 추월 횟수는 Grid 순위와 강한 상관관계 보임

Grid 순위가 낮으면 추월의 기회가 많고, Grid 순위가 높으면 추월의 기회가 적음.

이에 따라, 퀄리파잉 결과가 좋지 못한 선수의 경우 추월의 기회가 더 많을 가능성.

순위에 따른 점수로 보정하였어도, 상위권 선수의 경우 추월에서 높은 점수를 얻지 못할 가능성을 시사함.

예시로, Grid 1등, 추월을 허용하지 않고 1등으로 완주한 선수가 있다 가정하면, 이 선수는 추월에서는 0점을 획득하게 된다는 한계가 존재.

이를 보완하기 위해 일관성 지표와 함께 다루어야 한다는 시사점을 제시.

Grid 순위	평균 추월
1-5위	4.7회
6-10위	4.7회
11-15위	8.0회
16-20위	8.2회

LIMITATIONS

Dirty Air : 하위권에게 주어지는 패널티가 실존하는가?

1. 인과관계의 딜레마 (ENDOGENEITY)

- 페이스가 느려서 순위가 낮은가? (PACE → POSITION)
- 순위가 낮아서 페이스가 느려지는가? (POSITION → PACE)
- 앞차의 난류가 다운포스를 감소시켜 코너링 기록(S2) 악화

2. 통계적 유의성과 낮은 설명력의 괴리

- 회귀 결과: P-VALUE < 0.001 / F-통계량 311.83
- 설명력: 에타 제곱(ETA²) 1.98%
- $S2 = 0.551 + (0.054 * POSITION) - (0.083 * CAR SCORE)$

순위	Δ Sector 1	Δ Sector 2	Δ Sector 3
P2	+0.08s	+0.08s	+0.06s
P5	+0.35s	+0.30s	+0.23s
P10	+0.51s	+0.67s	+0.52s

3. 데이터 수집의 기술적 제약

- 기류 및 하중 센서 등 세부 센서 데이터 부재
- 계산에 사용한 순위는 대리 변수로서 희석 발생





EDA - Sports

DATA SCIENCE LAB

THANK YOU FOR
LISTENING

A dark, blurred background image showing two individuals wearing headsets and working at a computer, suggesting a data science or sports analysis environment.