

# F1 Analysis - Final Project - Team 174


Sam Wagner, Conner Nations, Braxtyn DeGoyer, Brian Yuhas, Davud Davudov, Edwin Pimentel

## Motivation/Introduction

- 1 | 150-300 sensors with more than 750 million data points during 2-hour race. This corresponds to 300-400 Gb data per car every weekend
- 2 | Gaining insight into the root causes and correlations that lead to success or failure in Formula 1 (F1) racing has long been difficult to achieve.
- 3 | Our analysis utilizes F1 datasets from previous years to provide new insight into F1 racing to highlight key factors that may contribute to success (or failure) in a race
- 4 | End-users can find a satisfying visualization and analysis that can be used to highlight previously unreviewed performance drivers in F1 races.

## F1 Data

 We have built the F1 dashboard by utilizing the FastF1 dataset via API connection. (<https://theoehrly.github.io/Fast-F1/>)

 We have analyzed data collected from all races between 2018 and 2022. Data consists of 141m records in total, with 2.65gb on disk, and around 12gb in memory

## Our approach

### How does it work:

- Collect F1 data through API. Multiple datasets for years between 2018-2022 are cleaned and combined
- Calculate and develop desired metrics needed (overtaking, driver trend).
- Use Streamlit for visualization which allows for customizable and easy web page development.

### How does it solve the problem:

- We analyze and display the overtaking potential of a driver using several methods centering around differences in car speeds and track sector overtaking “friendliness”. We determine how much the introduction of DRS impacted overtakes in F1
- Overtake ability of drivers are determine and clustered.
- The impact of tire selection on race success is studied. Users can analyze lap times by year, race, and driver in order to assess the distributions of lap times across the various tire compositions.
- By combining all these datasets and analysis through F1 Dashboard.

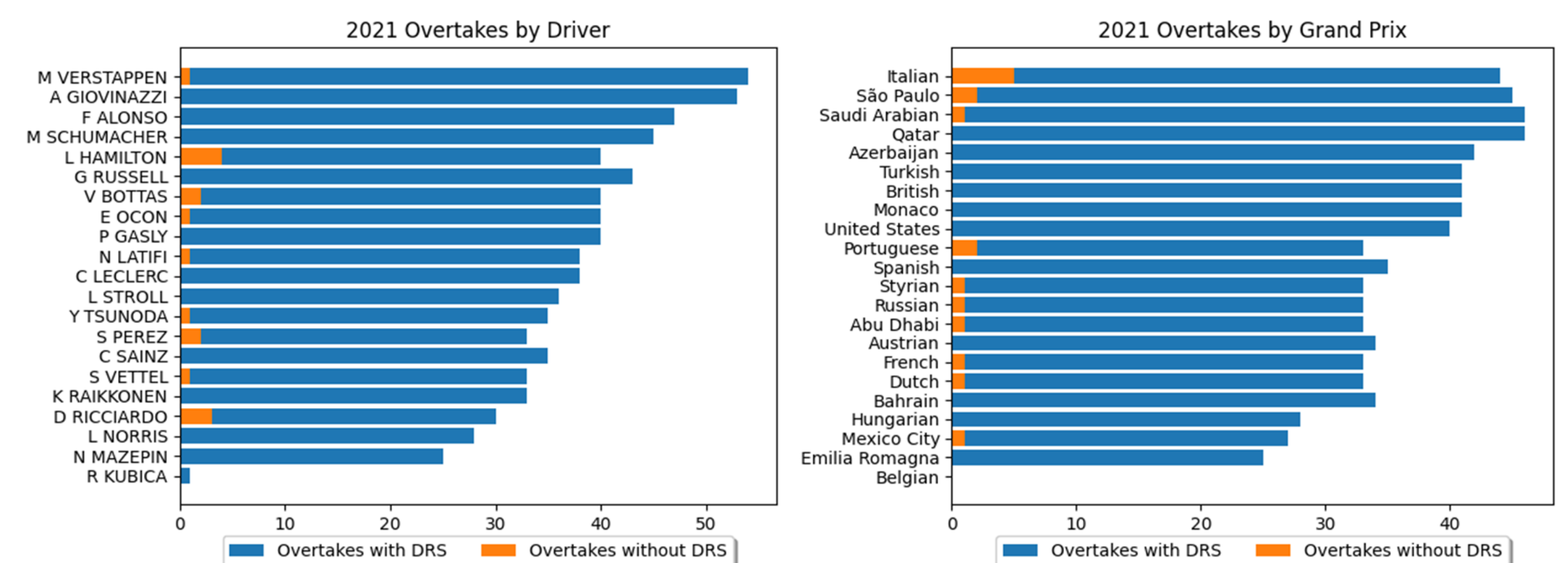
### What is new in our approach:

- Overtaking data is calculated using time gaps between cars and positional coordinate data. Impact of DRS on overtake is analyzed
- Driver performance are studied calculating an “overtake efficiency” metric, and clustering drivers into three tiers.
- A cohesive and interactable visual based webpage which F1 analysis is lacking.

## Experiments & Results

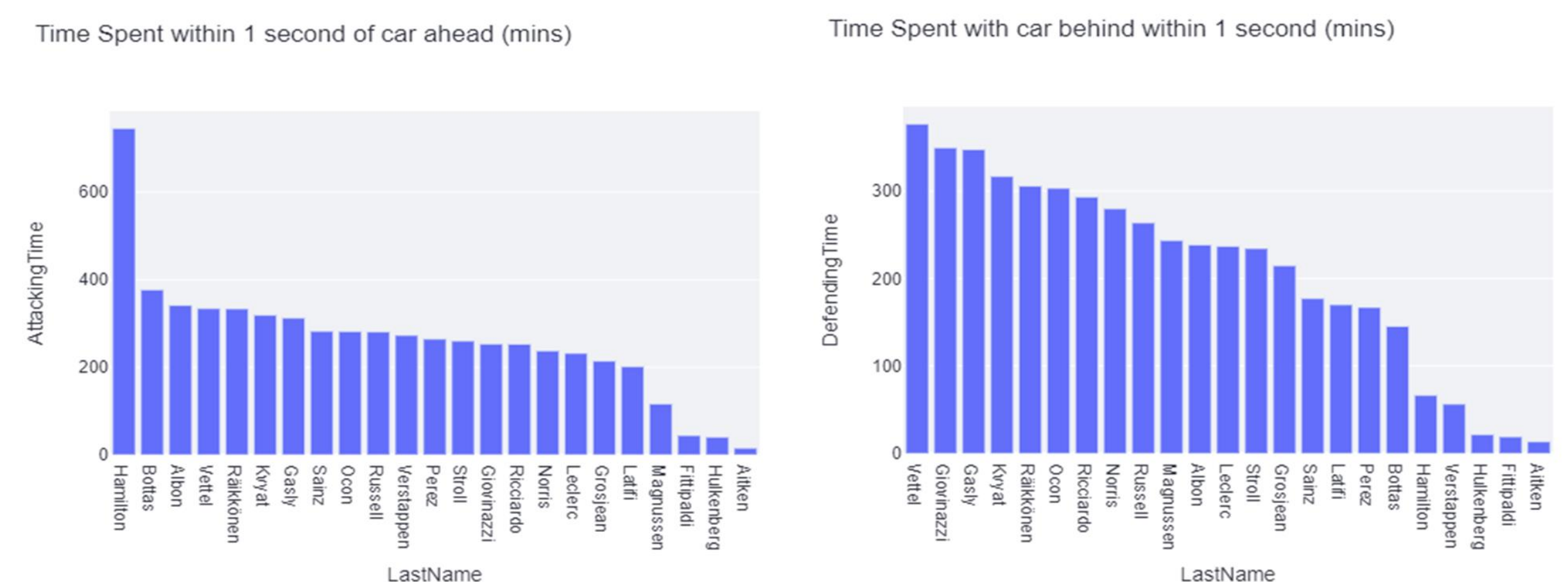
### DRS and Overtakes:

- We observe that most of the overtakes nowadays are performed using DRS, to the point where some drivers perform overtakes exclusively using DRS in most tracks.
- We observed that the track with more overtakes without DRS was the Italian and the hardest to perform overtakes was Emilia Romangna.
- The driver with the most overtakes in the year was Verstappen, and the driver that was able to perform the most overtakes without DRS was Hamilton.



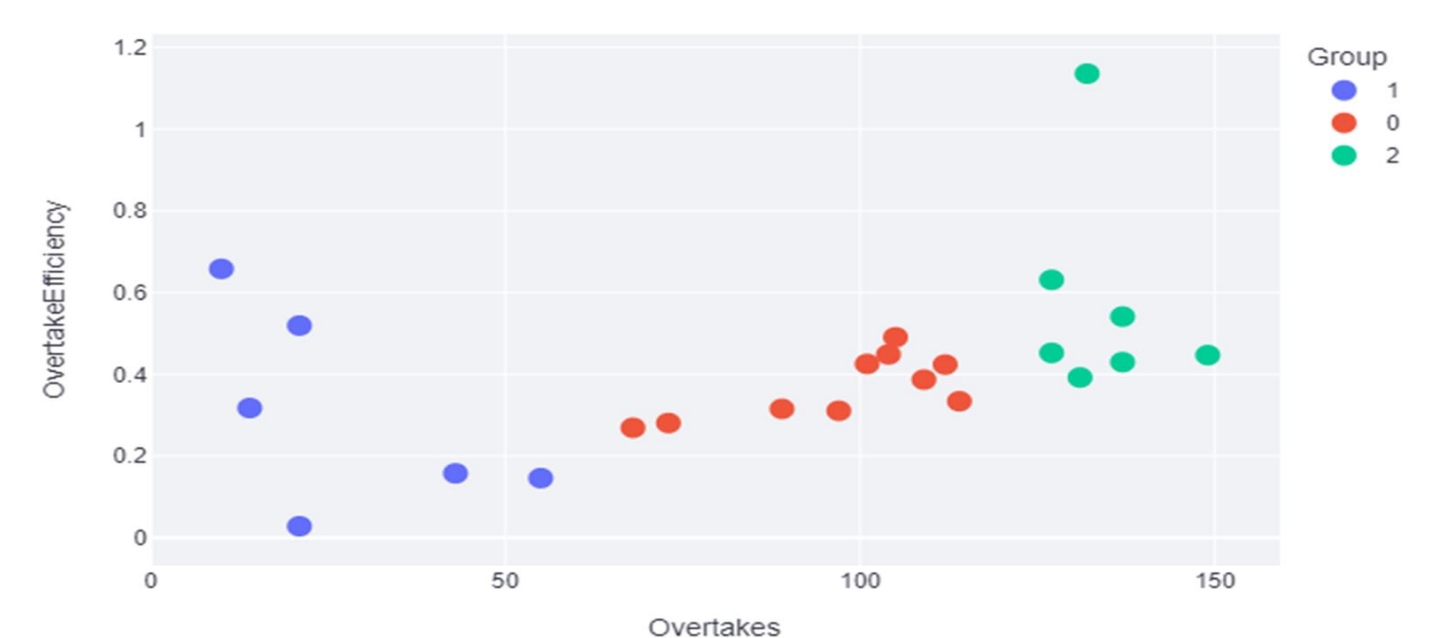
### Driver Trend and Overtakes

- A driver that is able to both keep a car behind for long periods of time while also not wasting time when overtaking will on average be more successful.



- Drivers are grouped based on their overtake ability using K-means algorithm.

Based on several experiments results indicate that most drivers perform at a different level on different tracks, likely a latent variable of how well a particular car is suited to each track layout.



### Tire Analysis

- Experimented tire composition and weather data to analyze the variation in lap times across races, drivers, and weather conditions to determine optimal tire compositions for different situations and different drivers.

The common conception of tire composition heavily affecting lap time might only be applicable on certain tracks. Not all tracks have shown recent evidence of this philosophy holding true.

