TDE
DIGITAL CREATIVES IN SPORTS

F1 Data analytics

Improving the Grand Prix experience for F1 viewers at home

S8 Graduation FHICT

4 Sept 2023 - 16 Jan 2024

By Jordi Franssen

20 November 2023

www.tde.nl

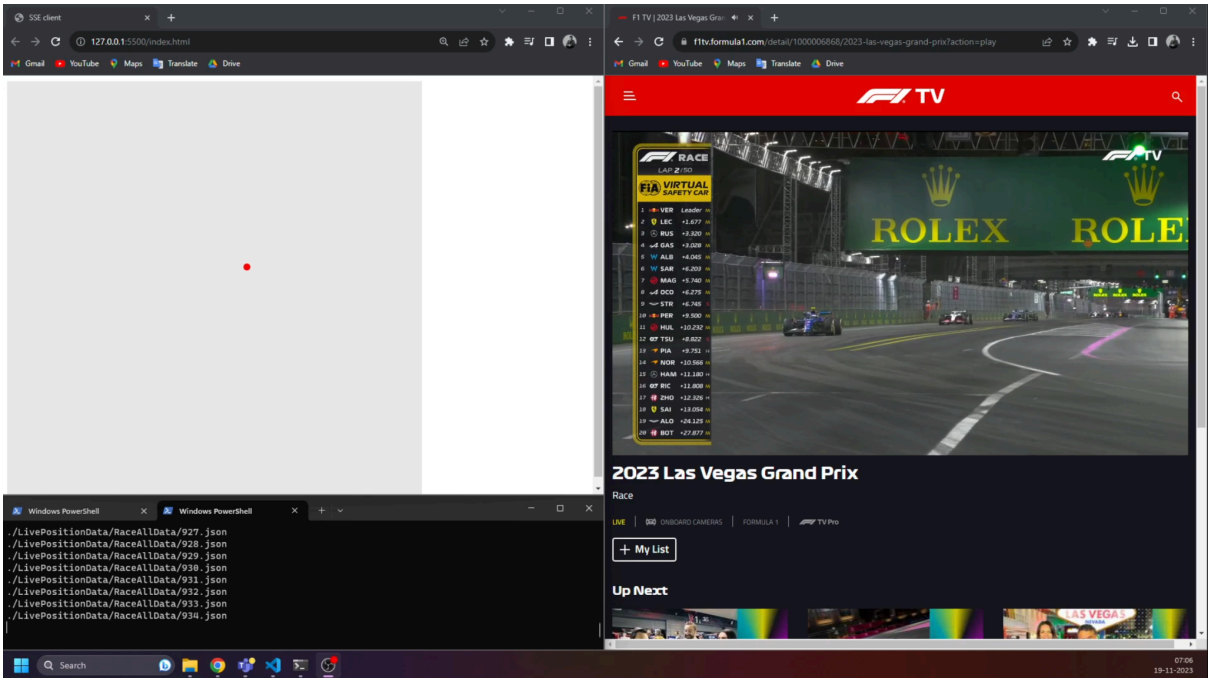
Introduction

For the Vegas GP, I recorded all data from the Formula 1 API and made a screen recording from the live broadcast. This way, I am able to tell which data belongs to which moment in the race and therefore, I can do some data analytics to find out what the variables in the data actually mean. In this document you'll read about the discoveries I made to decide which data to use for detecting events during the race, and what data not to use.

Broadcast delay

From the recording I made, I concluded that there's a 23 second delay on the broadcast compared to the datasource.

I checked this by picking a moment on the broadcast where something happened, like a deployed safety car. Because I logged the filenames of the data that was saved at that moment, I was able to analyze that file and observe that it was indeed visible that the track status was valued as "VSCDeployed". I checked the files that were saved before that file until the track status was valued as "AllClear". This was 23 files, or 23 seconds earlier. Therefore, I've established that there's a 23 second delay. I checked this for the free practice, qualifying and the race. During qualifying, there was a 24 second delay, and during the free practice, there was a 26 second delay. Therefore, I can conclude that this delay is quite consistent, with around 25 seconds.



info@tde.nl
(+31) 040 782 00 01
www.tde.nl

S8 Graduation FHICT
s8-graduation.jordifranssen.com
Jordi Franssen

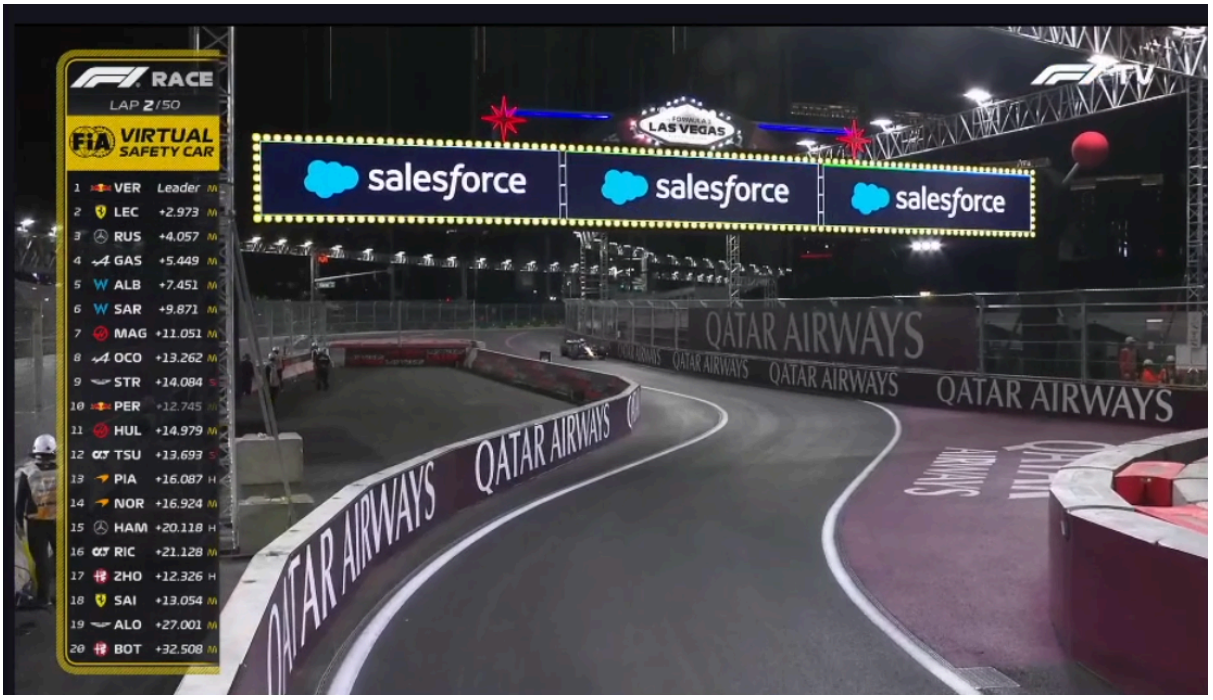
Other discoveries

OnTrack

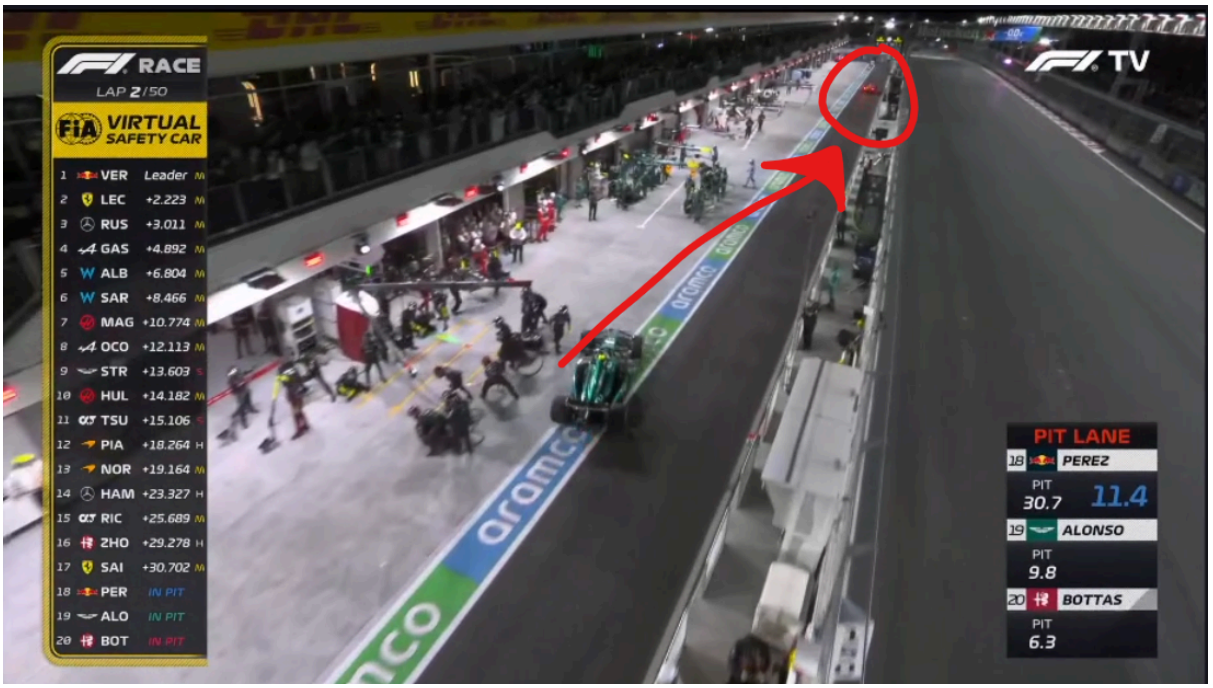
I assumed that the OnTrack value found in the position data indicated whether a car is within the racing lines or not. I checked this with the crash of Norris last weekend. After Norris crashed, he was in the barriers instead of on the track. However, his position status was still OnTrack, instead of OffTrack. This indicates that we cannot use this value to filter out overtakes during a crash for example. Therefore, we need to detect a crash in another way. Norrisses position data didn't change as he crashed of course. Therefore, we can detect a crash based on position data.

InPit and PitOut

We need to detect pit stops, however, I was curious what the InPit and PitOut value exactly mean. I used a pitstop from Perez as reference and picked a moment on the broadcast when Perez was driving to the pitlane. From this moment, InPit is set to true.



InPit stays true until Perez leaves the pitlane. From the moment he accelerates to rejoin the track, InPit is set to False and PitOut is set to true, and at this moment his tire compound updates. PitOut stays true for 30 seconds. This is way longer than Perez took to rejoin the track. Therefore, the time that PitOut is set to true is useless.



Tyre data

The API houses a TyreData variable. This variable updates with the tyre compound the cars are using. This data can be used to detect which tyres are changed during a pit stop.

Overtakes

In the TimingDataF1 variable, the live position of the driver in the race is provided. If this data changes, an overtake must have taken place. Therefore, we can use this data to determine which cars are overtaking each other.

Driverlist

The DriverList variable houses data of the drivers themselves. Everywhere else in the API, drivers are indicated by racing number, not by name. We can use this variable to

match the racing numbers with the full name of the drivers.

IsOut

The TimingDataF1 also houses an IsOut variable. This variable is set to true for a driver if that driver is out of the race. We can use this variable to stop the backend script from logging overtakes for example.

Conclusion

Because of the screen recording I made during the Vegas GP, I was able to analyze the data that came from the Formula 1 API during the race. I can simply check which json file was written when an event happened and check if that event can also be detected in the data itself. Because of this, I was also able to determine a delay between the data source and the live broadcast. This delay turns out to be around 25 seconds. I have also been able to determine how to detect potential crashes, pit stops and tyre changes, overtakes and how to match the racing numbers in the API with the full name of the drivers.

Summary

For the Vegas GP, I recorded all data from the Formula 1 API and made a screen recording from the live broadcast. This way, I am able to tell which data belongs to which moment in the race and therefore, I can do some data analytics to find out what the variables in the data actually mean.

Because of the screen recording I was also able to determine a delay between the data source and the live broadcast. This delay turns out to be around 25 seconds. I have also been able to determine how to detect potential crashes, pit stops and tyre changes, overtakes and how to match the racing numbers in the API with the full names of the drivers.

Learning Outcome Clarification

- Learning Outcome 1: Professional Duties
- Learning Outcome 4: Investigative Problem Solving

This deliverable is a professional duty on a bachelor level in the activity of Analysis as I analyzed the data from the data source and concluded on what data the data actually means and what data can be used to detect events happening during the race. Therefore, Learning Outcome 1: Professional Duties applies.

I used “Data analysis” as a research method from ictresearchmethods.nl for this deliverable. Therefore, Learning Outcome 4: Investigative Problem Solving applies.