

Max Verstappen was the best driver of the 2021 Formula One season.*

Athaliah Biju

30 April 2022

Abstract

This paper will be an analysis on the 2021 Formula One season. 2021 champion, Max Verstappen was crowned under controversial circumstances. Considering other variables independent of total point count such as race wins, podiums and pole positions, Verstappen was the best driver of the 2021 season.

1 Introduction

Formula One (F1) is the world's most prestigious motorsport competition. The first ever Formula One Driver's Championship was held in 1950 (Judde, Booth, and Brooks 2013). The sport has seen an explosion in popularity in recent years, with a cumulative viewership of 1.5 billion during the 2021 season (2022a). The final race, held in Abu Dhabi, drew 108.7 million viewers (2022a). The surge in interest in the sport can be ascribed to the hit Netflix documentary series, *Formula 1: Drive to Survive*. F1 consists of 10 teams: McLaren, Red Bull, Mercedes, Ferrari, Aston Martin, Alpine, Alpha Tauri, Williams, Alpha Romeo and Haas. Each team has two drivers each, with a total of 20 drivers competing for the World Drivers Championship (WDC) as seen in Table 1. The teams, on the other hand, compete for the World Constructor's Championship (WCC). The driver and/or constructor with the most points at the end of the season are crowned WDC and WCC respectively.

A typical race weekend sees three days of action on track. Friday consists of two Free Practice sessions (FP1 and FP2). These critical 60 minute sessions give teams the chance to test different setups and collect data to determine the optimal tyre and race strategy (2022b). Saturday sees an additional FP session (FP3) as well as Qualifying (2022b). Qualifying is a 60 minute session that determines the grid position for the main Grand Prix on Saturday. This is determined through race pace (which is how fast the car can go) alone (2022b). Qualifying is further divided into three sessions: Q1, Q2 and Q3. The slowest five cars get eliminated in Q1 and determine positions 16-20 (2022b). The subsequent slowest five cars determine positions 11-15 (2022b). Therefore, the goal is to be the fastest car in Q3 to secure pole position (first place on the grid) (2022b). The main event - the Grand Prix - is held on Sunday. The exact number of laps depends on the circuit but must have a minimum race distance of 305 km and must not exceed two hours (Kallor 2022). The first to cross the finish line when the chequered flag is waved is deemed the winner. The top 10 race finishers receive points, as seen in Table 2 (2010). The driver with the fastest lap of the race is awarded an extra point, contingent on a top 10 finish.

The 2021 season saw the introduction of F1 Sprint Qualifying, which sees a modified schedule. The number of FP sessions is reduced to two. On Fridays, FP1 and Qualifying is held. The Qualifying session sets the grid for the Sprint. FP2 is held on Saturday before the Sprint. The Sprint is a 100 km race that determines the race grid for the main Grand Prix (2021a). This was done to spice up a familiar format. The top three drivers of the Sprint qualifying will earn three, two and one point respectively (2021a). The Sprint format was seen at the British GP, Italian GP and Brazilian GP.

Lewis Hamilton, who drives for Mercedes, has won the WDC every single year since 2014, except for 2016 - which was won by teammate Nico Rosberg. Since the genesis of the turbo-hybrid era, Mercedes has won the WCC. Few were able to prove a worthy competitor to Hamilton and Mercedes, that is, until 2021. Red

*Code and data are available at: [LINK](#).

Bull and driver Max Verstappen were able to match the Mercedes' pace. The 2021 season was one of the most dramatic and unpredictable seasons in recent years, with the championship lead changing hands often. Verstappen and Hamilton were in a league of their own. Verstappen and Hamilton had equal points going into the last race. As a result, the season finale race served as the championship decider. Verstappen ended up winning the race, albeit under controversial circumstances. As a result, the Max Verstappen was crowned the 2021 Formula One champion. Mercedes won the WCC for the eighth year in a row. There has been much debate surrounding the final race and whether Verstappen was a worthy champion. This paper seeks to determine who the best driver in the 2021 season was when considering other variables such as race wins, number of podiums and qualifying results in addition to the total accumulated WDC points.

2 Data

The analysis was conducted using R (R Core Team 2020). The dataset was obtained from an R package called `formula1data` made by 'arkraieski' on GitHub (Kraieski 2022). This package was imported using `devtools` (Wickham et al. 2021). The `formula1data` package imports Formula 1 data from the 'Ergast' API (n.d.). I chose to use this package as opposed to the API directly as the package would allow for cleaner analysis. Further cleaning was done using `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2021). The graphs were created using the `ggplot2` package (Wickham 2016). Tables were constructed using `knitr` (Xie 2022).

The data frames I worked with the most were the Race Results, Qualifying Results and Final F1 Standings. I chose these three data frames as I felt it would allow me to tell the story of the 2021 season better. I wanted to focus on the driver performance, I excluded data about the constructors. While much of the data was clean, a lot of manipulation had to be done to extract the desired variables out of a massive dataset. For example, to construct the graph showing the points progression, I had to extract race data individually and then recombine them in a manner that would allow me to show the cumulative points the drivers earned over the course of the season.

3 Results

The 2021 Formula One season consisted of 22 races. The championship standings were decided at the end of the Abu Dhabi Grand Prix, as seen in Table 3.

Table 1: Drivers competing in the Formula 1 World Driver’s Championship during the 2021 season

Name	Constructor
Max Verstappen	Red Bull Racing
Sergio Perez	Red Bull Racing
Lando Norris	McLaren
Daniel Ricciardo	McLaren
Lewis Hamilton	Mercedes
Valtteri Bottas	Mercedes
Charles Leclerc	Ferrari
Carlos Sainz Jr.	Ferrari
Fernando Alonso	Alpine Renault
Esteban Ocon	Alpine Renault
Pierre Gasly	Alpha Tauri
Yuki Tsunoda	Alpha Tauri
Sebastian Vettel	Aston Martin
Lance Stroll	Aston Martin
Kimi Räikkönen	Alfa Romeo
Antonio Giovinazzi	Alfa Romeo
Mick Schumacher	Haas
Nikita Mazepin	Haas
George Russell	Williams
Nicholas Latifi	Williams

Table 2: Formula 1 Points System

Position	Points
1st	25
2nd	18
3rd	15
4th	12
5th	10
6th	8
7th	6
8th	4
9th	2
10th	1

3.1 Race Winners

The 2021 season saw six different race winners, as seen in Figure 1. Max Verstappen won the most races, winning 10 of the 22. Lewis Hamilton won 8. Valtteri Bottas, Sergio Perez, Esteban Ocon and Daniel Ricciardo each won one race.

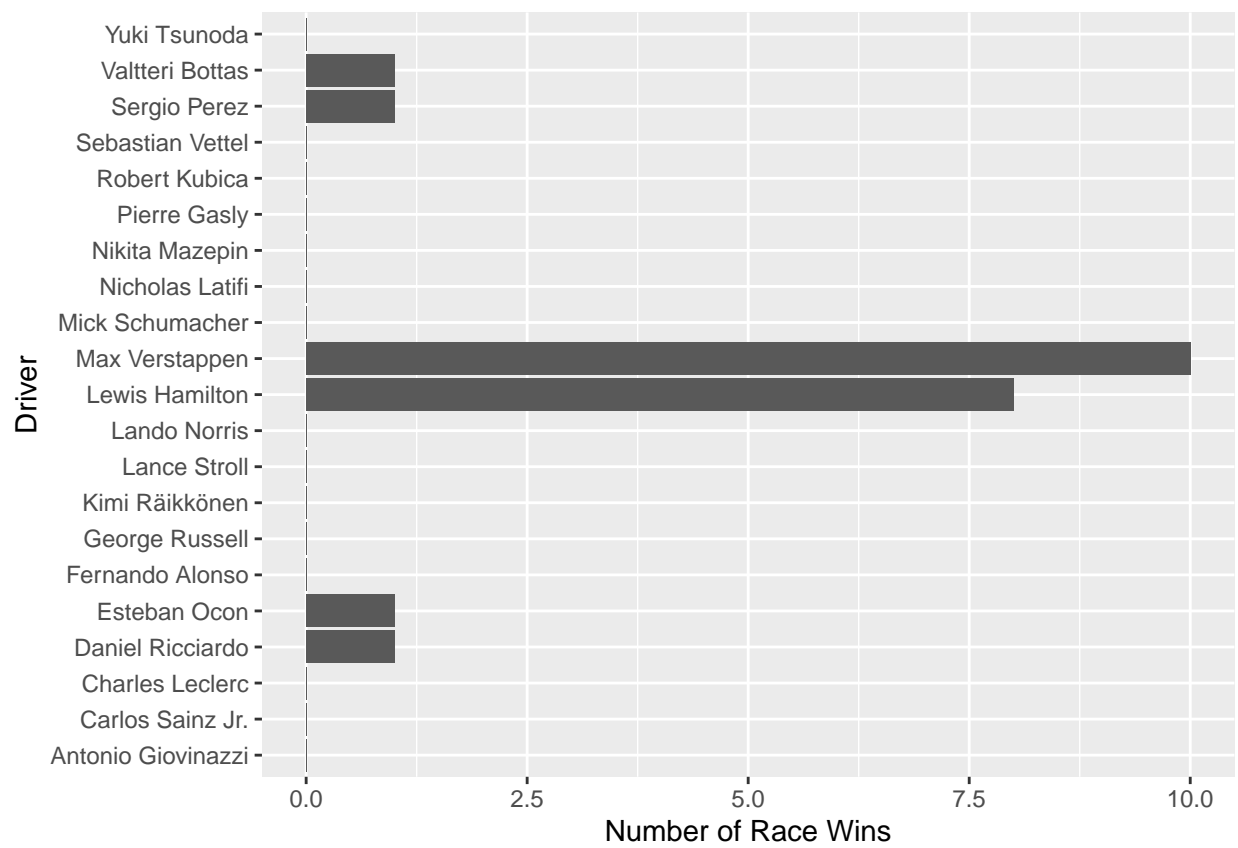


Figure 1: Number of race wins for each driver in the 2021 Formula 1 season

Table 3: Driver Standings at the end of the 2021 Formula 1 World Drivers Championship

Position	Driver	Points
1	Max Verstappen	395
2	Lewis Hamilton	387
3	Valtteri Bottas	226
4	Sergio Perez	190
5	Carlos Sainz Jr.	164
6	Lando Norris	160
7	Charles Leclerc	159
8	Daniel Ricciardo	115
9	Pierre Gasly	110
10	Fernando Alonso	81
11	Esteban Ocon	74
12	Sebastian Vettel	43
13	Lance Stroll	34
14	Yuki Tsunoda	32
15	George Russell	16
16	Kimi Räikkönen	10
17	Nicholas Latifi	7
18	Antonio Giovinazzi	3
19	Mick Schumacher	0
20	Robert Kubica	0
21	Nikita Mazepin	0

3.2 Podiums

As seen in Figure 2, the 2021 season saw 12 different podium sitters. Max Verstappen had the most, with 18. Coming in at second, Lewis Hamilton finished on the podium at 17 races. Valtteri Bottas had 11. Sergio Perez finished the season with 5 podiums. Lando Norris and Carlos Sainz Jr. each stepped on the podium 4 times. Daniel Ricciardo, Sebastian Vettel, Charles Leclerc, Fernando Alonso, Pierre Gasly, Esteban Ocon and George Russell each stepped on the podium one time.

3.3 Pole Positions

As seen in Figure 3, the 2021 season saw five different pole sitters. Max Verstappen and Lewis Hamilton both qualified P1 at eight different occasions. Valtteri Bottas had three pole positions. Charles Leclerc qualified P1 twice. Lando Norris achieved pole position once.

3.4 Points Progression

Figure 4 shows the points progression of the 2021 Formula One Championship. Max Verstappen had the most number of points at the end of the season. A quick note about Kubica: Alfa Romeo reserve driver, Robert Kubica was brought in to race at two Grand Prix: the Dutch GP and the Italian GP. This was because Kimi Räikkönen tested positive for COVID-19 and was therefore unable to compete (2021b).

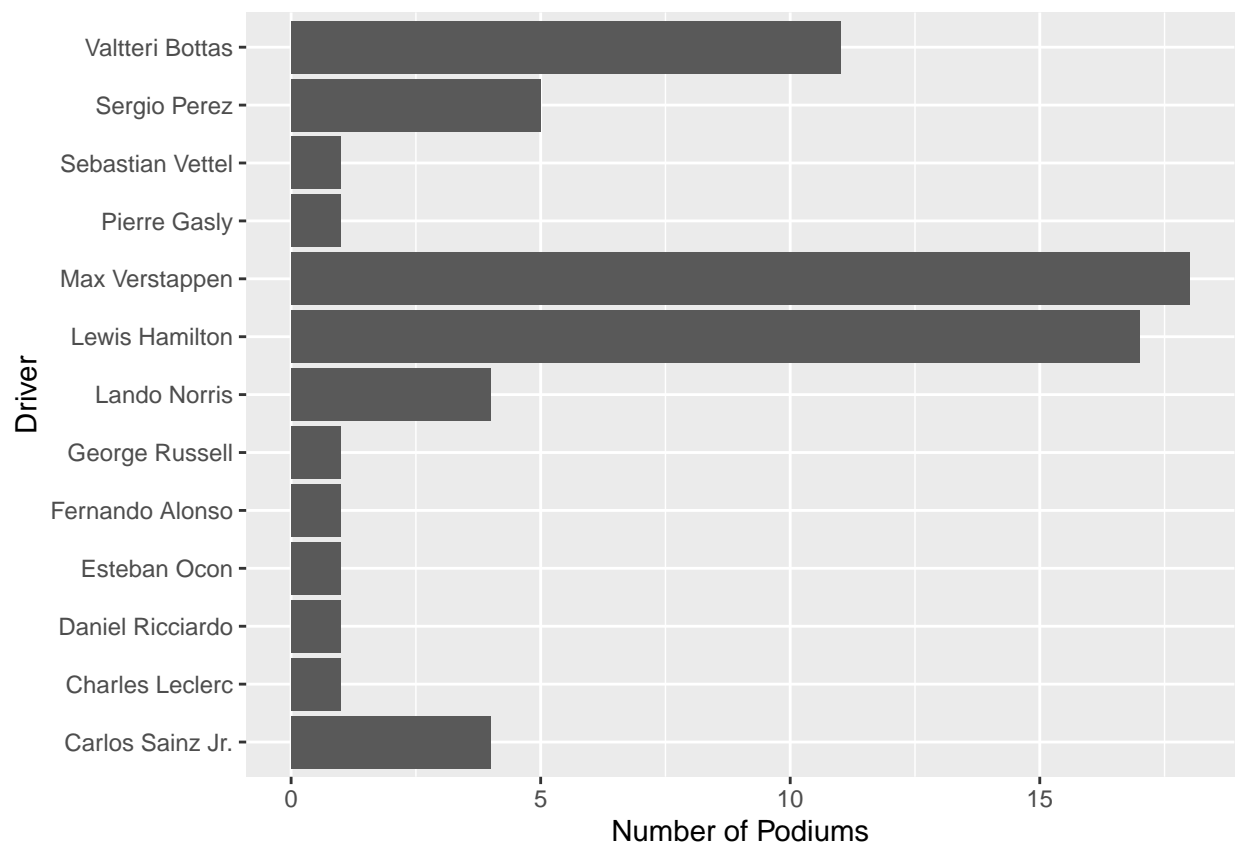


Figure 2: Total number of podiums amongst podium finishers in the 2021 Formula 1 season

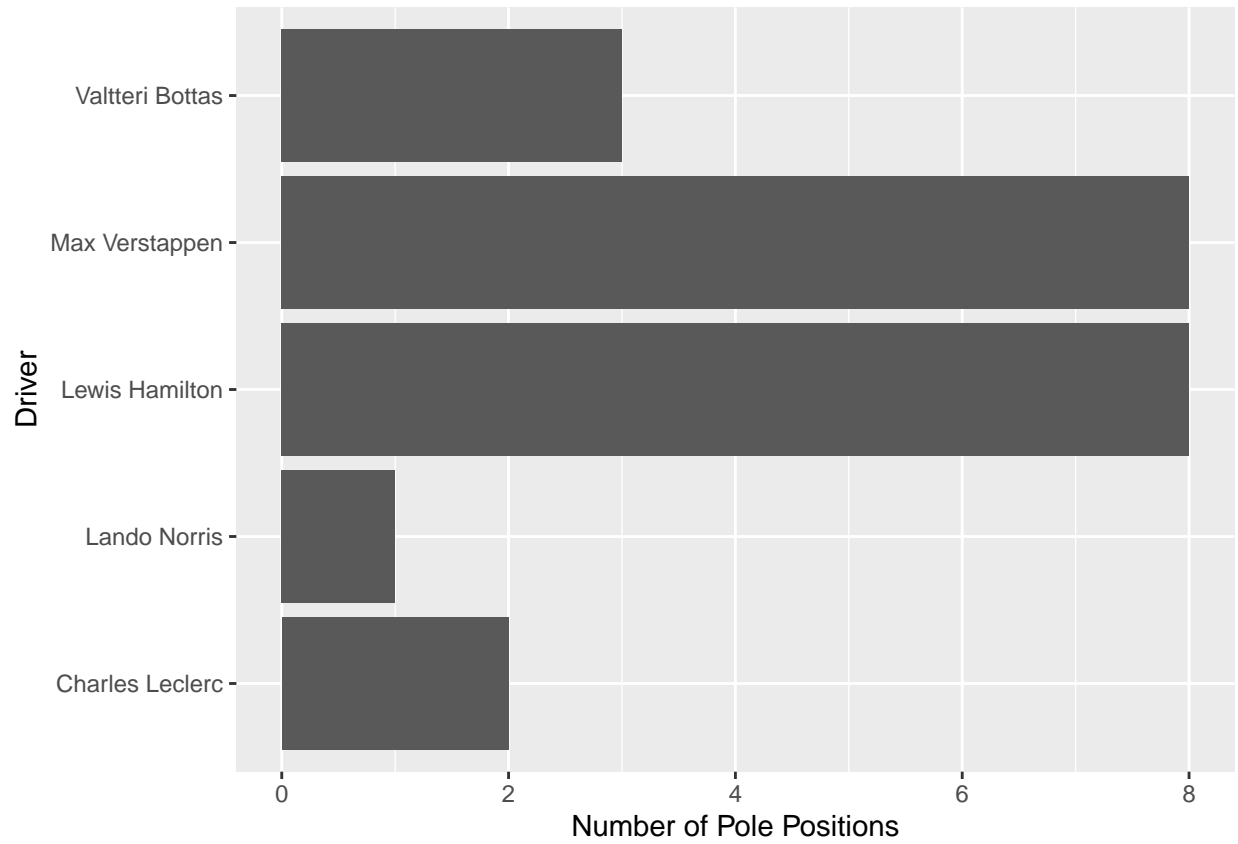


Figure 3: Total number of pole positions amongst the drivers that have qualified P1

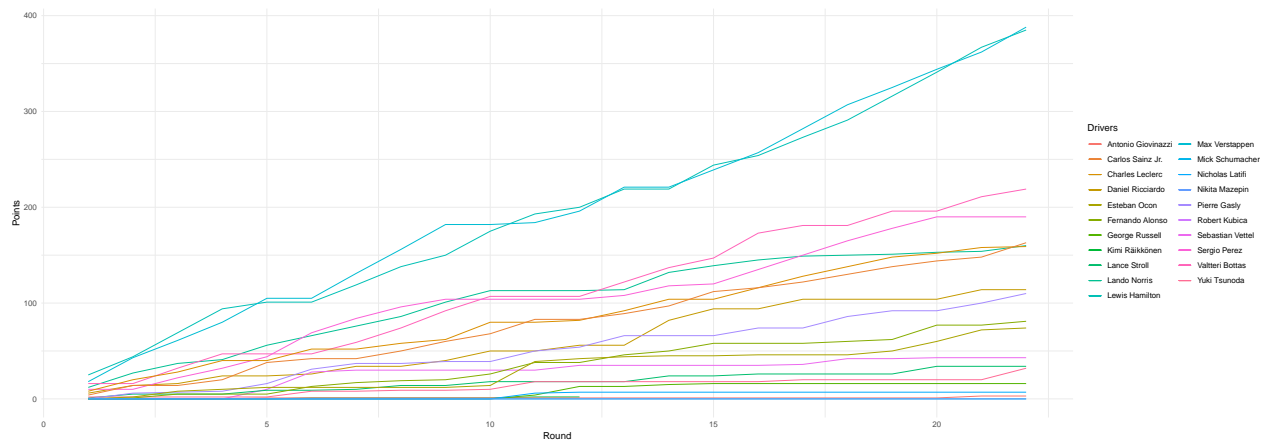


Figure 4: World Drivers Championship points progression of the 2021 Formula 1 season

4 Discussion

4.1 Verstappen and Hamilton were in their own league

The Formula 1 grid is comprised of 20 of the fastest drivers in the world. However, the 2021 season showed that the championship battle involved nobody but Max Verstappen and Lewis Hamilton. As Figure 4 shows, their points totals were far beyond the rest of the grid. Both Verstappen and Hamilton finished the season with 395 and 387 points respectively (Table 3). Their closest competitor was Valtteri Bottas, who finished third with 226 points, as seen in Table 3. Verstappen won the most races in the 2021 season, with 10 wins. Hamilton had 8 race wins. They were also the only drivers to win multiple races, as shown in Figure 1. Verstappen also made the most appearance on the podium, with 18. Hamilton was a close second, with 17 appearances [Figure 2]. When comparing qualifying, both Verstappen and Hamilton had similar results. They were both on pole position 8 times [Figure 3]. However, this does not take into consideration the Sprint qualifying format, which is discussed further below.

4.2 Context about Sprint Races and Qualifying Results

When looking at Figure 3, it must be noted that for data collection purposes, the results of the Sprint are not filed under the Qualifying category. However, the driver who was on pole during a Saturday Qualifying was on pole for the Sprint race and not the Grand Prix. The Sprint format was held at the British GP, Italian GP and Brazilian GP. While Lewis Hamilton was on pole for the inaugural Sprint race at the British GP, Max Verstappen won the Sprint race (2021c). Therefore, he was on pole for the Sunday GP. At the Italian GP, Valtteri Bottas was on pole both at the end of Qualifying as well as the Sprint (2021d). At the final Sprint event of the 2021 season in Brazil, Hamilton was once again on pole for the Sprint, but his teammate Bottas won his second consecutive Sprint in a row (2021e). In terms of success at the Sprint format, Hamilton is the clear winner.

5 Conclusion

Max Verstappen was crowned the 2021 F1 champion, and deservedly so. With the most race wins and podium appearances, his numbers back up his performance. Additionally, he led the most laps out of anyone on the grid (Matthew 2021). 652 laps to be exact, which is double the number of laps Hamilton led (297) (Matthew 2021). He also had drove 6 of the fastest laps of the season (Matthew 2021).

Appendix

.1 Datasheet for the Dataset

.1.1 Motivation

1. For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description. The dataset was created to enable analysis of Formula One data.
2. Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)? Alex Kraieski created the package but the Ergast API is run by a user that goes by Chris.
3. Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number. It is unknown if the creation was funded.

.1.2 Composition

1. What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description. This depends on the dataset you use. For example, the Final F1 Standings dataset for the 2021 season consists of the final standing, driver name, constructor name, points total and number of wins.
2. How many instances are there in total (of each type, if appropriate)? Queries can be made in 12 different categories.
3. Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable). Race data is available for the very first year of the F1 world championship in 1950.
4. What data does each instance consist of? “Raw” data (for example, unprocessed text or images) or features? In either case, please provide a description. Some datasets consist of numbers (race results, qualifying results, etc). Some are links to Wikipedia pages of individuals and circuits. It depends.
5. Is there a label or target associated with each instance? If so, please provide a description. Each Grand Prix has its own ID. The season corresponds to the year it is held in.
6. Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text. Some race results are not available for a certain driver if they did not take part in the race.
7. Are relationships between individual instances made explicit (for example, users’ movie ratings, social network links)? If so, please describe how these relationships are made explicit. No.
8. Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them. No.
9. Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description. No.
10. Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources

that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate. Self-contained.

11. Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description. No, all data were gathered from public sources.
12. Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why. No.
13. Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset. Yes, name, age, date of birth, nationality and car number.
14. Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how. Yes, individuals are identified by name.
15. Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description. No.

.1.3 Collection process

1. How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how. The data were gathered from the Australian Parliamentary Handbook in the first instance, and this was augmented with information from other parliaments, especially Victoria and New South Wales, and Wikipedia.
2. What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated? Scraping and parsing using R.
3. If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)? The dataset is not a sample.
4. Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)? Rohan Alexander. Paid as a post-doc and an assistant professor, although this was not tied to this specific project.
5. Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created. Three years, and then updated from time to time.
6. Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation. No.
7. Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)? Third parties in almost all cases.
8. Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself. No.

9. Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented. No.
10. If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate). Consent was not obtained.
11. Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation. No.
12. Any other comments? No.

.1.4 Preprocessing/cleaning/labeling

1. Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section. Yes cleaning of the data was done.
2. Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data. No. But the package can be accessed at <https://github.com/arkraieski/formula1data>. The API can be accessed at <http://ergast.com/mrd/>.
3. Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point. R was used.
4. Any other comments? No

###Uses

1. Has the dataset been used for any tasks already? If so, please provide a description. It has been used for various applications. They can be seen here: <http://ergast.com/mrd/gallery/>
2. Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point. No
3. What (other) tasks could the dataset be used for? iPhone widget or Smart speaker app
4. Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms? No.
5. Are there tasks for which the dataset should not be used? If so, please provide a description. No.
6. Any other comments? No.

.1.5 Distribution

1. Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description. The dataset is available through GitHub.
2. How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)? API

3. When will the dataset be distributed? The dataset is available now.
4. Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions. No copyright.
5. Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions. None that are known.
6. Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation. None that are known.
7. Any other comments? No.

.1.6 Maintenance

1. Who will be supporting/hosting/maintaining the dataset? Ergast API
2. How can the owner/curator/manager of the dataset be contacted (for example, email address)?
mailto://admin@ergast.com
3. Is there an erratum? If so, please provide a link or other access point. No, the dataset is just updated.
4. Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)? Yes, 2 hours after every track session.
5. If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced. No.
6. Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers. No the dataset is just updated. Although a history is available through GitHub.
7. If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description. Forum on Ergast website
8. Any other comments? No

References

- 2010, February. http://news.bbc.co.uk/sport2/hi/motorsport/formula_one/8493082.stm.
- 2021a. <https://www.formula1.com/en/latest/article.everything-you-need-to-know-about-f1s-new-sprint-qualifying-format-including.1Lawf6r6Ab0y8ha0satSjx.html>.
- 2021e. <https://www.motorsport.com/f1/news/f1-grand-prix-sprint-results-wins-for-brazilian-gp-pole/6778402/>.
- 2021c. <https://www.motorsport.com/f1/news/grand-prix-sprint-results-f1/6632310/>.
- 2021d. *RacingNews365*. <https://racingnews365.com/2021-f1-italian-grand-prix-sprint-qualifying-results>.
- 2021b. *Sauber Group*. <https://www.sauber-group.com/motorsport/f1-news/robert-kubica-to-compete-in-the-2021-dutch-grand-prix/>.
- 2022a. <https://www.formula1.com/en/latest/article.formula-1-announces-tv-race-attendance-and-digital-audience-figures-for-2021.1YDpVJIOHGnuok907sWcKW.html>.
- 2022b. <https://www.motorsport.com/f1/news/insiders-guide-how-an-f1-race-weekend-works/6744616/>.
- n.d. <http://ergast.com/mrd/>.
- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Judde, Chris, Ross Booth, and Robert Brooks. 2013. “Second Place Is First of the Losers: An Analysis of Competitive Balance in Formula One.” *Journal of Sports Economics* 14 (4): 411–39. <https://doi.org/10.1177/1527002513496009>.
- Kallor, Amber. 2022. “The Beginner’s Guide to F1.” *The Ritz-Carlton*. <https://www.journey.ritzcarlton.com/travel-memories/mercedes/F1-beginners-guide>.
- Kraieski, Alex. 2022. *Formula1data: Imports Formula 1 Data Using the 'Ergast' API*. <https://github.com/arkraieski/formula1data>.
- Matthew, Clayton. 2021. “The Numbers Behind Max Verstappen’s F1 Title Win.” *Red Bull*. <https://www.redbull.com/ke-en/max-verstappen-f1-world-championship-statistics>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester, Winston Chang, and Jennifer Bryan. 2021. *Devtools: Tools to Make Developing r Packages Easier*.
- Xie, Yihui. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.