Attrition Rate in Formula 1

<u>Overview</u>

Formula 1 recently introduced a short-format Sprint race, which is held alongside the main Grand Prix on the same weekend at selected tracks. The 2023 Qatar Sprint was very entertaining and saw five cars spin out, three safety car periods, and the crowning of Max Verstappen as 2023 Driver's Champion.

While watching, I wondered if it seemed more exciting because the short number of laps (19, compared to 57 for the main race) concentrated the incidents and led to many retirements per lap - a high attrition rate.

Sprint races are also shorter in duration than regular Grands Prix. Therefore it would also be possible to define attrition rate as retirements per minute.

Questions

- Which races in history had the highest attrition rate?
- How did the 2023 Qatar Sprint compare to other races?
- Do Sprints generally have higher attrition rates than Grands Prix?

<u>Data and Method</u>

Data source
http://ergast.com/mrd/

Method

Working in Microsoft SQL Server, I wrote a SQL query to identify retirements, total laps and total duration for each Grand Prix and Sprint, and then join the two sets of results into one table. I used PowerBI and PowerQuery to transform and explore the data further, and produced visualisations. I also used Excel to produce some charts, and PowerPoint to produce summary infographics of my results.

Calculating retirements

A retirement during a race is essentially whenever a car fails to finish. This could mean the car crashed, broke down on the circuit, or simply pulled into the pits and stopped. For the purposes of this analysis, I was interested in cars that started the race but didn't finish it. The database gives some information about the retirements.

"The value of the position attribute in the Result element is always an integer, giving the finishing order of all drivers. The value of the positionText attribute is either an integer (finishing position), "R" (retired), "D" (disqualified), "E" (excluded), "W" (withdrawn), "F" (failed to qualify) or "N" (not classified)."

Unfortunately in Formula 1, a car is counted as a finisher if it completes 90% of the race winner's laps. This wasn't helpful for this analysis - I wanted to count all cars that didn't complete 100% of the laps as retirements. Luckily, the database also lists a status for each result - for example 'Finished' or 'Disqualified', as well as the car's position on the grid. I wrote a CASE WHEN statement to count the number of cars that had a non-zero grid position (that is, they started the race) and a status that wasn't 'Finished', '+n Laps' or 'Disqualified'.

Calculating total laps and race duration

By definition, the winning car will always complete the total number of laps in the race. Likewise, the winner's race time is equal to the race duration, as the race ends when the winning car completes the final lap. I was able to calculate total laps and race duration by writing a SQL subquery to select the time and laps fields for records where the finishing position was 1.

Results

Retirements Per Lap (RPL) reached a peak in the late 1980s

Figure 1 shows that the attrition rate - measured by RPL - increased to a peak in the late 1980s, before gradually decreasing again. The most recent F1 seasons have the lowest RPL figures in history.

The other main trend shown by these data is that some races in the early decades of F1 had astonishingly high attrition rates. Some have an RPL of over 1, meaning that - on average - more than one car retired every lap.
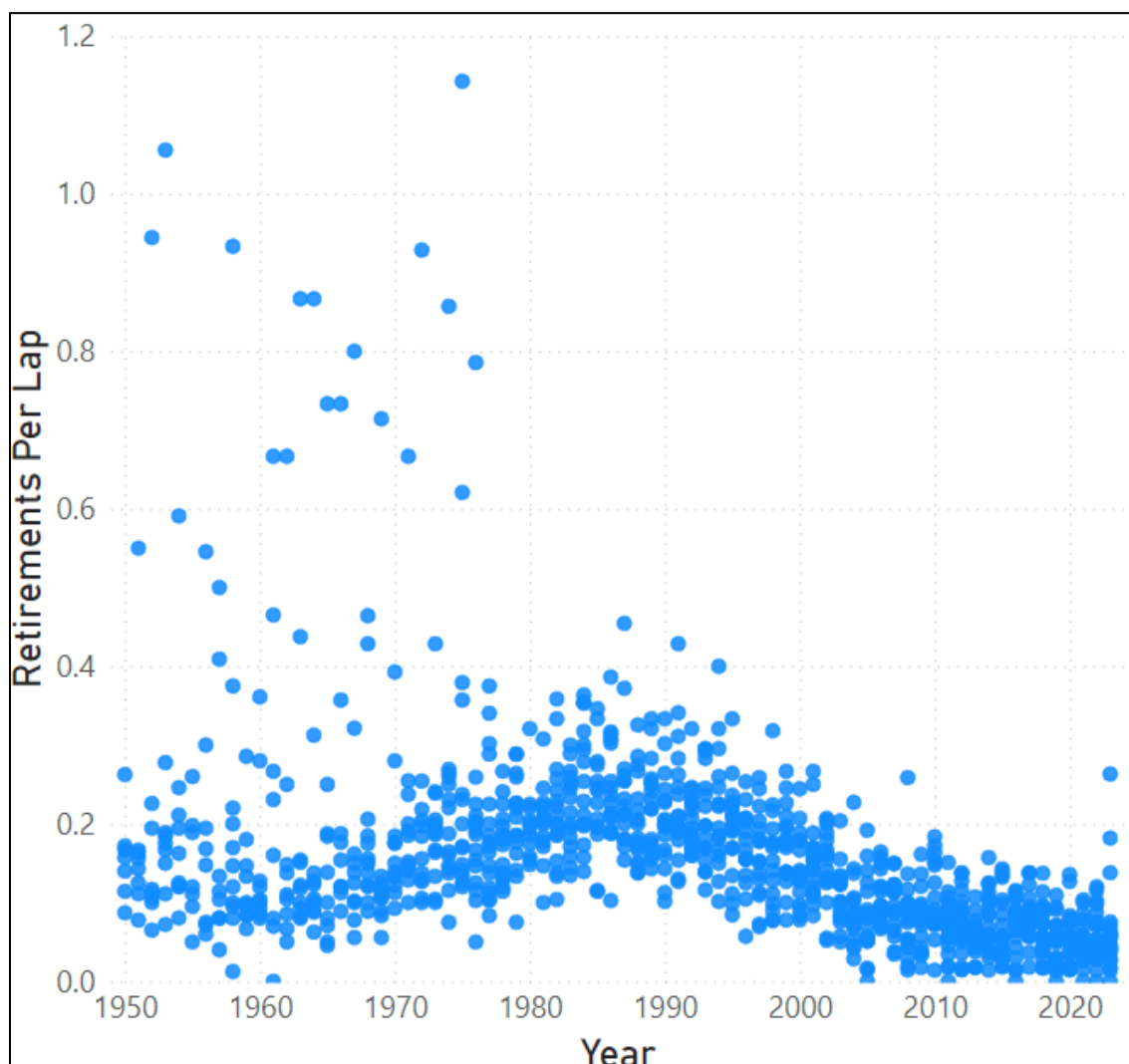

Figure 1: *Retirements Per Lap (RPL) for every F1 Grand Prix and Sprint*

The fewer laps in a race, the higher the RPL

A closer look at those early races with a high RPL showed that they were all held at the Nürburgring (Figure 2). From 1951 until 1976, almost all German Grands Prix used the Nürburgring's Nordschleife circuit - a narrow strip of road that twisted for 23 km through the forest. As Formula 1 races a run over a specified total distance, a longer lap means that the race will have fewer laps in total. The extraordinary length of the Nordschleife meant that races there averaged just 16 laps. In terms of retirements per lap, the races with the highest attrition rates of all time were all German Grands Prix on the Nürburgring Nordschleife.
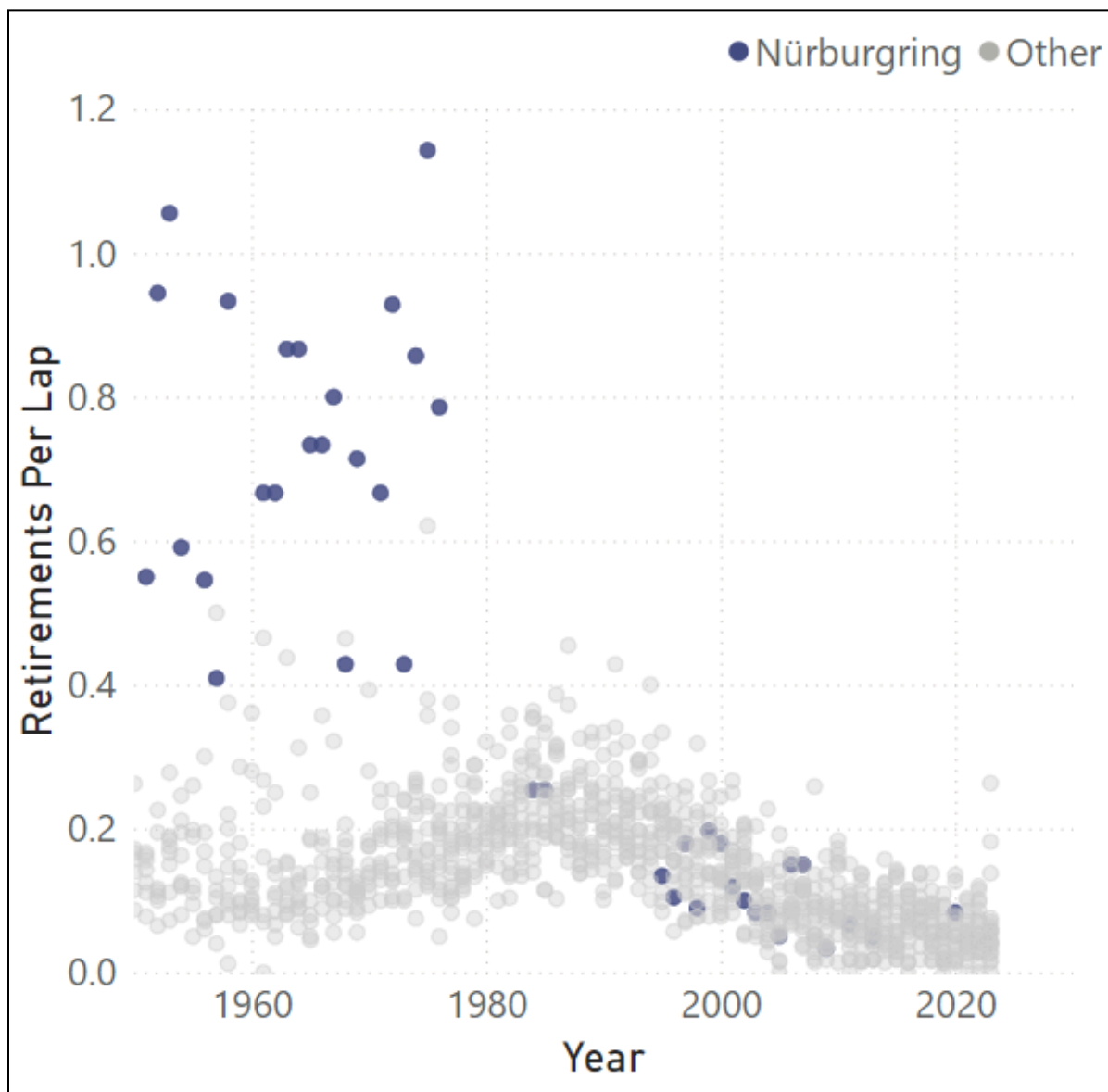

Figure 2: RPL for races at the Nürburgring compared to all other races

Although the Nürburgring Nordschleife races are an extreme example, they do represent one end of a spectrum. Figure 3 shows that race length has a weak negative correlation with RPL. The shorter the race, the higher the attrition rate. Figure 3 also shows that Sprints have lower attrition rates than expected given their small number of laps.
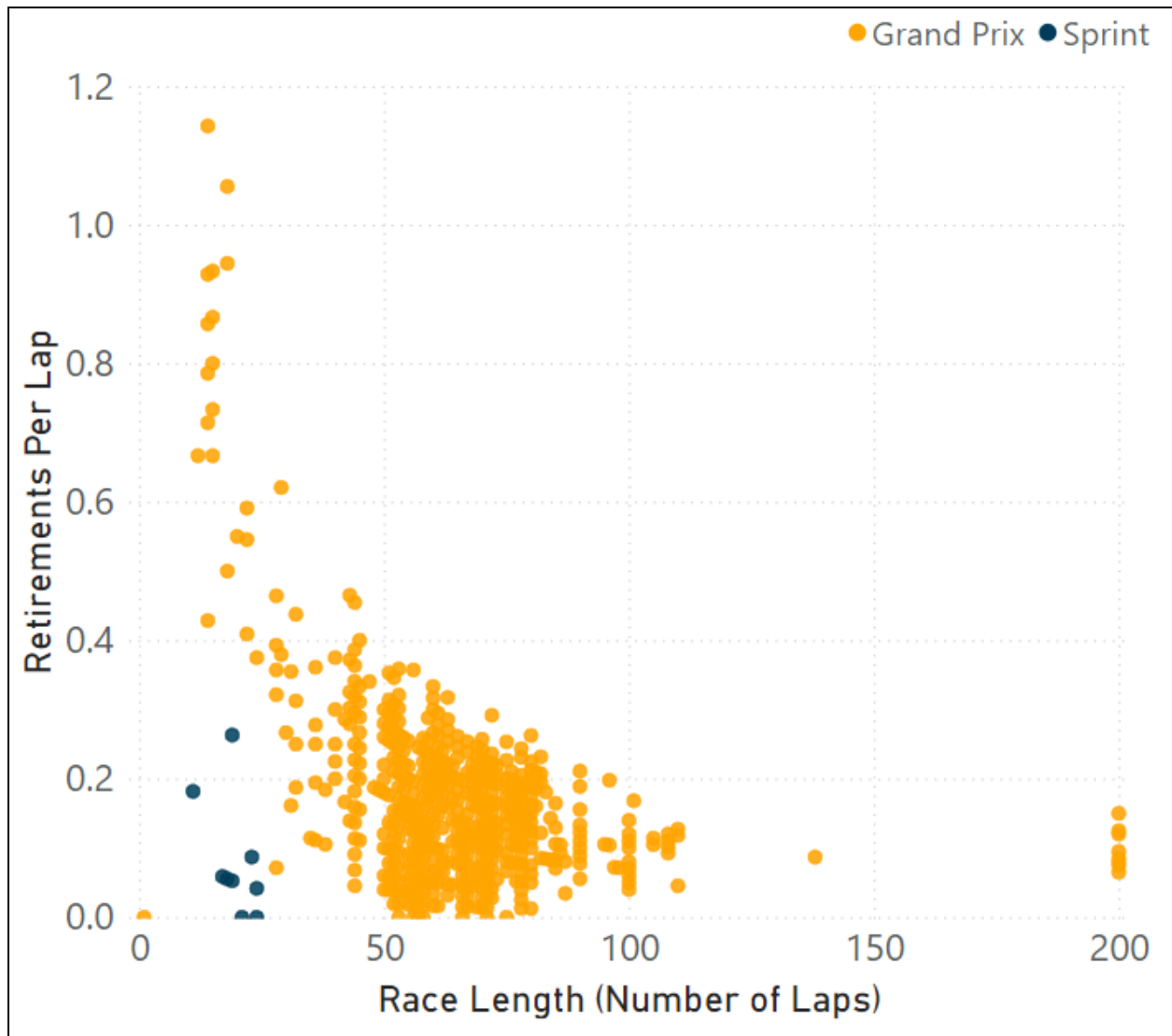

Figure 3: The relationship between laps in a race, and the retirements per lap

Retirements Per Minute (RPM) shows the same trend as RPL

Measuring attrition rate by retirements per minute (RPM) shows the same overall pattern as RPL. Figure 4 shows that RPM increases up to a peak in the late 1980s, before decreasing to record low levels in the most recent seasons. The range of attrition rates for the early seasons is much lower for RPM compared to RPL. This is because races with a small number of very long laps - such as those at the Nürburgring Nordschleife - took roughly the same amount of time to complete as the other races which had a larger number of shorter laps.

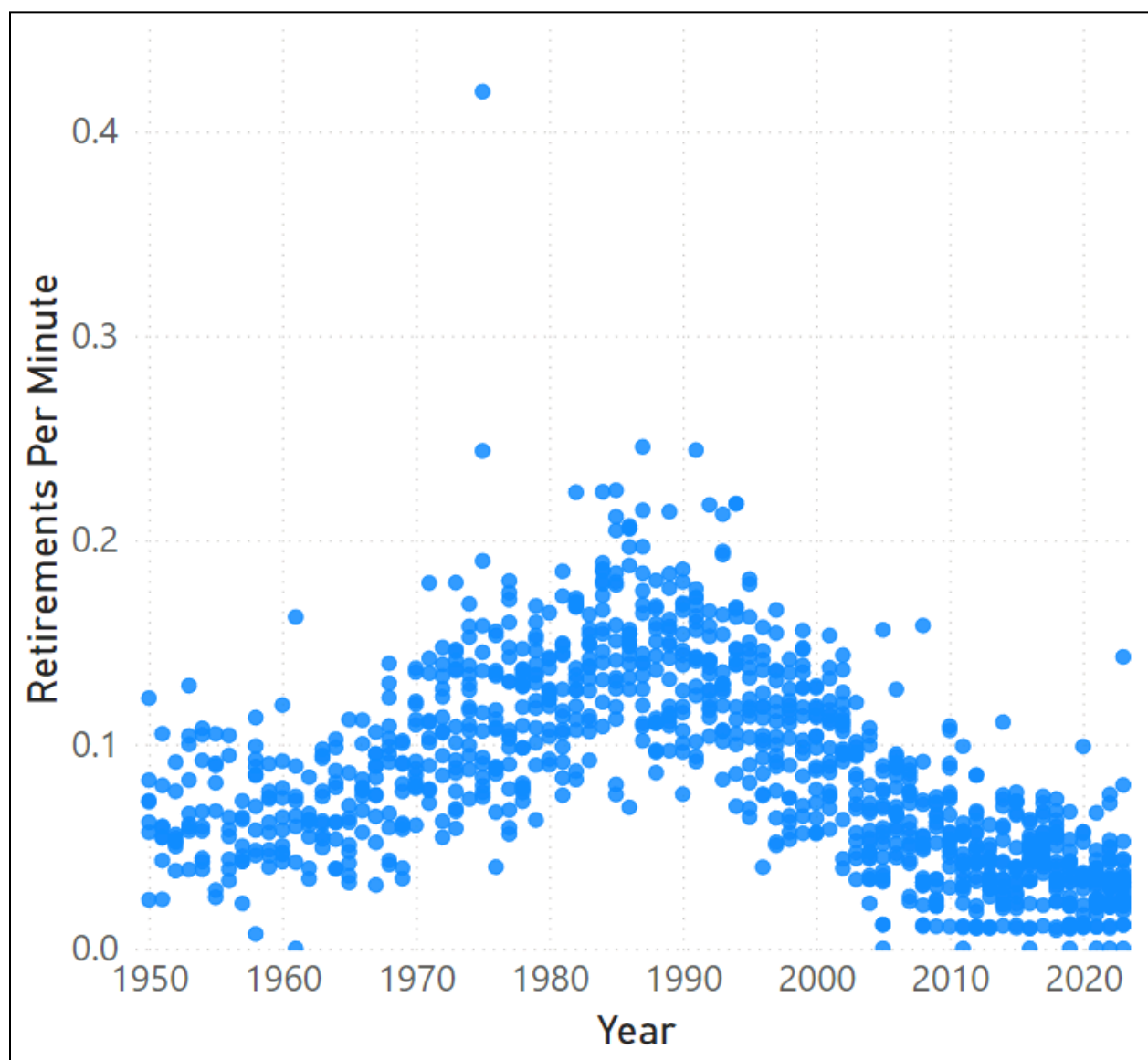The 1975 Spanish Grand Prix is an obvious outlier with 0.42 retirements per lap

Figure 4: *Retirements Per Minute (RPM) for every F1 Grand Prix and Sprint*

The relationship between race duration and RPM isn't clear

Figure 5 shows a very weak negative correlation between race duration and RPM. However, the graph also shows that the majority of races lasted between 75 and 125 minutes, and that those races had a wide range of attrition rates.

The graph appears to show that Sprints have a lower than expected RPM, however there are so few data points for Grands Prix with durations of an hour or less that it is very difficult to draw conclusions.
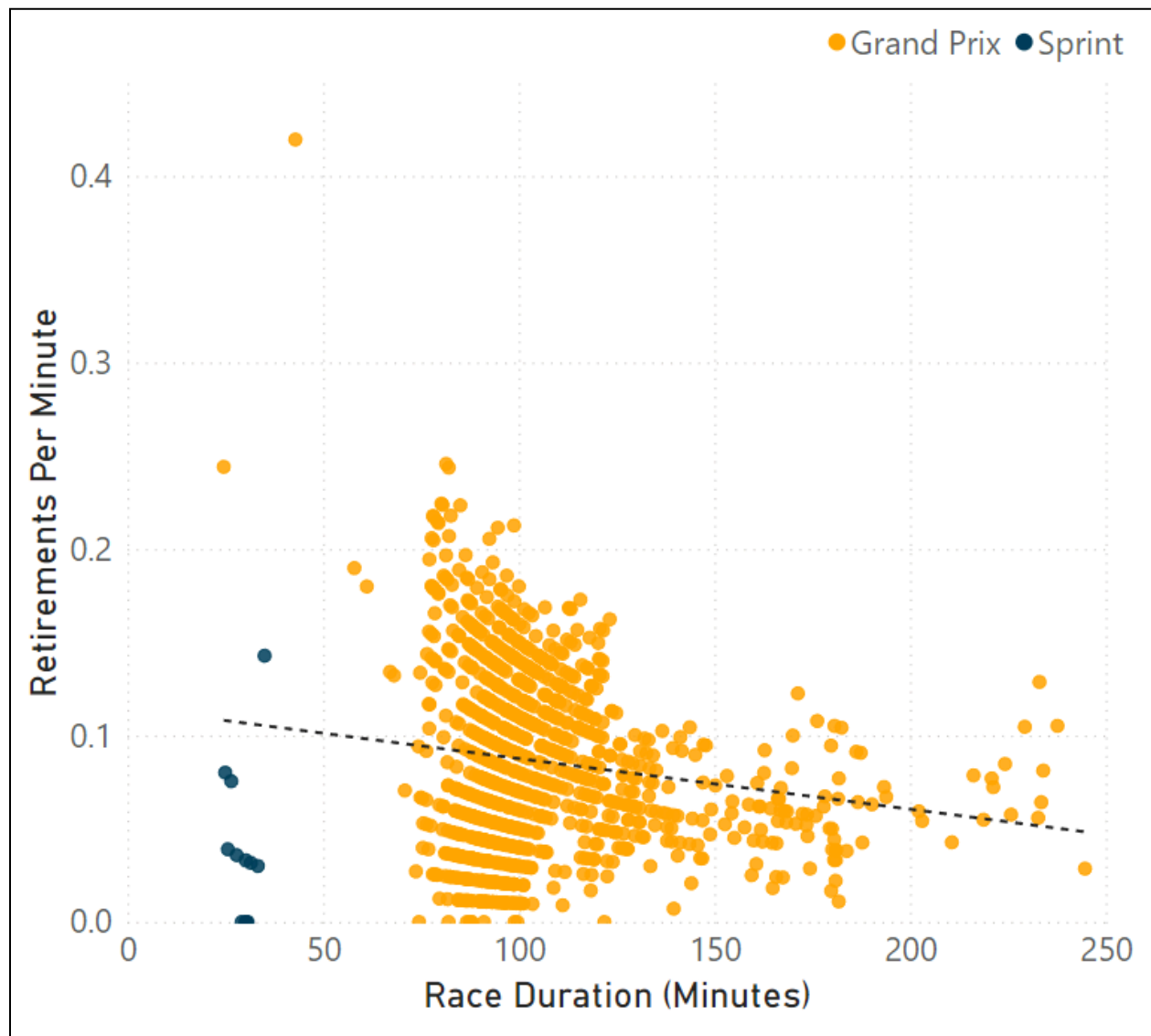

Figure 5: The relationship between race duration and RPM. This graph excludes the aborted 2021 Belgian GP.

The 1975 Spanish Grand Prix had the highest RPM of all time

Year	Name	RPM
1975	Spanish Grand Prix	0.42
1987	German Grand Prix	0.25
1991	Australian Grand Prix	0.24
1975	British Grand Prix	0.24
1985	Austrian Grand Prix	0.22
1984	Italian Grand Prix	0.22
1982	Austrian Grand Prix	0.22
1994	German Grand Prix	0.22
1994	Italian Grand Prix	0.22
1992	Italian Grand Prix	0.22

Figure 6: *Top 10 highest attrition rates (measured by retirements per minute) in Formula 1*

Conclusion

Which races had the highest attrition rates?

In terms of retirements per lap (RPL), the races with the highest attrition rates in F1 history are all German Grands Prix held at the Nürburgring Nordschleife circuit before 1977. The extreme length of a lap at this circuit means that races there had an average of just 16 laps, significantly fewer than any other race. Hence even races with a minimal number of retirements would have a relatively high RPL.

The 1975 German Grand Prix had the highest RPL of all time (1.14), which corresponds to more than one retirement every lap. Excluding the Nürburgring Nordschleife races, the race with the highest attrition rate by retirements per lap is the 1975 Spanish Grand Prix (RPL = 0.62).

Measuring attrition rate as retirements per minute (RPM) is a better way to look at the data, as there is less variation in race duration compared to the number of laps in a race. The 1975 Spanish Grand Prix also has the highest attrition rate by this measure (RPM = 0.42)

How did the 2023 Qatar Sprint compare to other races?

Compared to races throughout F1 history, the attrition rate during the 2023 Qatar Sprint was unremarkable. So far Sprints have only been held in the 2021, 2022 and 2023 seasons. In common with Grand Prix in those seasons, their attrition rates are low in comparison with races from earlier decades.

Do Sprints generally have higher attrition rates than Grands Prix?

By both measures considered in this analysis, sprints have a lower-than-expected attrition rate. Sprints have a below average RPL given their number of laps, and a below average RPM given their duration.

If we consider only the hybrid era of F1, Sprints have a similar average attrition rate to Grands Prix, but a wider range. The 2023 Qatar Sprint had the highest attrition rate for 25 years, but a quarter of all Sprints have had zero retirements - a far greater proportion than Grands Prix in the current era.

Evaluation

Issues with the data

The main issue with the date is that there have only been 12 Sprints so far, compared to 1,101 Grands Prix. It's difficult to draw any meaningful conclusions from such a small sample size.

Furthermore, Sprints have only featured in the three most recent F1 seasons. As they are being trialled, the format of these races has changed significantly in those years. For example, in 2021 only the top three drivers in the Sprint were awarded points, compared to the top eight in 2022 and 2023. In the first two seasons, the finishing order of the Sprint set the grid for the Grand Prix the following day, whereas in 2023 there was a standalone qualifying session for both Sprint and Grand Prix.

These factors could have led drivers or teams to take a different approach to Sprints - for example if a driver was running in 9th place in a 2021 Sprint, there would be no incentive to take a risk and attack the driver in 8th (potentially leading to collision and retirement), as doing so wouldn't gain any points.

Another issue with the data is that Formula 1 has gone through a series of different eras, each of which is significantly different to the others. The cars, circuits, drivers and attitudes in 1950 were all very different to 1980, and to 2023. This makes comparisons between Formula 1 eras notoriously difficult.

Further work

After measuring attrition rate by retirements per lap and per minute, retirements per kilometre would be the next logical step. The length of a race in kilometres could be found by multiplying the circuit length by the number of laps. Unfortunately the database does not contain distance data, and the issue is further complicated as the same circuits have changed configuration - and therefore length - between seasons. For example, every race in Singapore has used the Marina Bay circuit, but some years the circuit has been updated to remove corners and shorten the lap.

I would also like to add analysis on reasons behind trends highlighted in this report. For example, why did the attrition rate peak in the 1980s? What factors caused the 2023 Qatar Sprint to have such a high attrition rate, whereas other Sprints had zero retirements?

On reflection, retirements per lap and retirements per minute might be better communicated as "time between retirements" or "laps between retirements". It would be easier for people to understand that there was a crash every 2 minutes, as opposed to the idea of a fractional number of crashes per minute.

For this project I taught myself how to use PowerBI to transform the data and produce the visualisations used in this report. Going forward, I would like to use PowerBI's features more fully to produce an interactive, information-rich dashboard.