

# USING LAP TIMES TO EVALUATE FORMULA 1 DRIVERS

## **Introduction:**

Auto racing is an exciting sport that entertains millions of fans around the globe, with Formula 1 being the highest level of competition. Both fans, pundits, and team personnel use various metrics to evaluate drivers for purposes like debates and filling vacancies. These metrics include points scored, average finish, and the frequency of a driver finishing in a certain range (wins, podiums, points finishes, etc). This paper will observe a metric that uses a driver's lap times in a race and assesses how a driver stacks up against the competition.

## **Obtaining Data**

The lap time data I used for this project is from <https://motorsportstats.com>. To get the information, I went to each race page, went to the statistics tab, then the lap time tab, copied the table element with Javascript, put it in a local HTML file, and scraped the race data into a CSV with the "scrape.py" file.

After all the races have their CSV file, they are imported into an SQLite database table (times table). I added tables that held information about each driver's team (roster table) and each race's location (races table). The following query created the result that was put in a comprehensive CSV file (called fulldata):

```
SELECT x.race,r.name,x.driver,t.team,lap,time
FROM times x, roster t, race r
WHERE race=r.raceID AND x.driver=t.driver AND race >=
t.start AND race <= t.end
```

## **Cleaning Data**

The goal for this phase is to clean the dataset of data that would lead to an incorrect evaluation. These include the following laps:

- Laps where a driver has retired and/or crashed
- Laps where a driver has entered and/or exited the pits
- Laps where the race is slowed due to a safety car or a virtual safety car

The cleaning solution used is to remove any lap times that are over 20% slower than the fastest lap time of the respective race and completely void any driver's race if they have

completed less than 20 laps. The solution is represented in “clean.py”:

1. Create two dataframes with Python’s Pandas module: one that contains all the data from “fulldata”, and an empty one that will be used to store “clean” data.
2. Find the fastest lap for each race.
3. For each driver in a given race:
  - a. Ensure they have completed at least 20 laps.
  - b. Filter the fulldataframes to get a dataframe with all lap times by a driver that meets the aforementioned criteria.
  - c. Add all the records from the new dataframes to the clean dataframes. Repeat for all driver-race combinations.
  - d. Once the clean dataframes have been completed, write it to a CSV file that will be used for analysis.

## Analysis

To analyze drivers, I compared their lap times with an “optimal pace”. The optimal pace is a hypothetical race where each lap has the fastest time that was recorded (fastest lap in lap 1, fastest lap in lap 2, ..., fastest lap in lap n = lap 1 of optimal pace, lap 2 of optimal pace, ..., lap n of optimal pace).

I created a metric to represent the difference between the optimal pace and actual pace of a driver: Lap Time Differential (LTD). A driver’s LTD for a given race can be represented with the following equation:  $LTD = \frac{Driver\ Pace}{Adjusted\ Optimal\ Pace} \star 100\%$ .

The following algorithm is used to calculate Driver Pace and Adjusted Optimal Pace:

1. Load the data from the CSV file containing cleaned data into a dataframe, and create another dataframe with the full data.
2. Create a dataframe with the fastest time for each lap for each race.
3. For each driver:
  - a. Find all the driver’s clean laps in the clean dataframes
  - b. Sum the times of all clean laps. This is the Driver Pace.
  - c. Sum the times of the optimal laps for each lap completed cleanly by the driver. This is the Adjusted Optimal Pace.
  - d. Calculate LTD based on the above formula.
4. Store all LTDs in a single dataframe and export them to a CSV.

From here, I will use Google Sheets to visualize the data. Note that the lowest possible LTD is 100%, meaning that the driver’s pace is the optimal pace.

## Overview

Driver	Average LTD	Best Race	LTD	Worst Race	LTD
Max Verstappen	100.76%	AUT	100.20%	HUN	102.22%
Lewis Hamilton	100.77%	UK	100.29%	MON	101.44%
Sergio Perez	101.32%	ITL	100.53%	NET	102.42%
Valtteri Bottas	101.49%	MON	100.47%	IML	103.59%
Charles Leclerc	101.75%	ITL	100.60%	QAT	102.72%
Lando Norris	101.78%	ITL	100.68%	BRA	102.99%
Carlos Sainz	101.86%	MON	100.78%	NET	103.02%
Pierre Gasly	102.09%	TUR	101.22%	ESP	103.14%
Daniel Ricciardo	102.31%	ITL	100.48%	NET	103.25%
Sebastian Vettel	102.48%	HUN	101.28%	SA	103.85%
Fernando Alonso	102.48%	ITL	101.34%	ESP	103.80%
Lance Stroll	102.60%	ITL	101.28%	BRA	104.09%
Esteban Ocon	102.61%	HUN	101.23%	ESP	103.68%
Yuki Tsunoda	102.64%	AZR	101.68%	NET	103.79%
Antonio Giovinazzi	102.83%	ITL	101.43%	ESP	103.76%
Raikkonen-Kubica	102.84%	ITL	101.72%	NET	104.57%
George Russell	103.10%	ITL	101.38%	ESP	104.16%
Nicholas Latifi	103.27%	ITL	101.51%	NET	104.48%
Mick Schumacher	103.82%	AZR	102.82%	NET	105.40%
Nikita Mazepin	104.59%	ITL	103.37%	USA	105.39%

\*Kimi Raikkonen and Robert Kubica's LTDs were combined

\*\*There were cases where a driver did not have an LTD for a race. To combat this, I substituted missing values with their overall average, as to not affect their already-present LTD.

\*\*\*The Belgian GP was not included, as only one lap was completed.

**See Appendix for race-by-race LTD (chart + table) and 5-race average (chart + table)**

## Interpretation

From my observations, I grouped all drivers into three categories:

### *The Championship Battle: Max Verstappen vs Lewis Hamilton:*

The championship battle between Mercedes' Lewis Hamilton and Red Bull's Max Verstappen this year was one of the tightest battles in recent history, so it's no surprise the two are extremely close in average LTD. In the early stages, Verstappen had a lower five-race average LTD than Hamilton, but this changed after his accidents in the United Kingdom and Hungary. Verstappen's excellent performances in Turkey, America, and Mexico brought the two very close, and his strong effort in Qatar gave him a five-race average LTD lower than Hamilton's and maintained that until the end of the Abu Dhabi Grand Prix (see Appendix B).

### *The Fight for Scraps*

Until the last third of the season, the battle for third place behind Verstappen and Hamilton was heated: Red Bull's Sergio Perez and the departing Mercedes driver Valtteri Bottas repeatedly swapped the third and fourth lowest five-race average LTD, and McLaren's Lando Norris, Ferrari's Charles Leclerc and Carlos Sainz, and Alpha Tauri's Pierre Gasly were close behind (see Appendix B). Bottas had the fastest LTD in Monaco and Turkey, while Norris had it in Russia (see Appendix A). By the end of the season, however, Perez drove away from everyone else to comfortably be the third-fastest driver on the grid thanks to having at least the third-lowest LTD for each of the last six races (see Appendix B).

### *Everyone Else*

As for the remainder of the field, many observations can be made. Some findings include:

- McLaren's Daniel Ricciardo, Aston Martin's Sebastian Vettel, and Alpine's Fernando Alonso resided in the midfield. Alonso had a poor start to the season but was able to recover to be in the top 10 of five-race averages for nine of the remaining 11 seasons.
- Aston Martin's Lance Stroll, Alpha Tauri's Yuki Tsunoda, and Alpine's Esteban Ocon formed their cluster further down. Tsunoda, the Japanese rookie, had a solid start to the season but crumbled to a lowly 18th in five-race averages at Russia and Turkey. Tsunoda was able to somewhat recover to as high as 13th.
- Despite contracting Covid-19 and missing races, Alfa Romeo's Kimi Raikkonen had solid performances. His five-race average was as high as 12th, putting him ahead of drivers with better equipment.
- Williams' George Russell had a bright summer stretch where he overachieved with his equipment to get as high as 14th in five-race average, but struggled late in the season, remaining close to his teammate Nicolas Latifi.
- Aside from costing Gene Haas millions of dollars in damages, Mick Schumacher and Nikita Mazepin were considerably slower than the rest of the grid for the entire season. The duo always had the two highest five-race average LTDs while Mazepin was the

slowest driver in all races except at the Dutch Grand Prix, only beating Schumacher (see Appendix B).

## **Conclusion**

In this paper, we observed a metric that takes a driver's lap times, compares it to an optimal pace, and evaluates the driver based on how much slower they were from that. From the analysis, drivers can be ranked to see how they stacked up.

This project was also a learning experience for me. I learned how to use SQL to organize data, Python to filter out dirty data for a better analysis, and spreadsheet software to analyze and visualize data. Also, I learned how to observe, interpret, and present data.

Despite the success of the project, I believe improvements could have been made: I could have used Python or even R to analyze data better, and try out a visualizing library or other software instead of relying on spreadsheets.

## **Appendix**

- A. LTD (by race) table
- B. Five-race average LTD rank table
- C. LTD (by race) chart
- D. Five-race average LTD chart

## A: LTD by Race Table

Driver	Average LTD	BAH	IML	POR	ESP	MON	AZR	FRA	STR	AUT	UK	HUN
Max Verstappen	100.76%	<b>100.68%</b>	<b>100.64%</b>	<b>100.46%</b>	<b>100.91%</b>	<b>100.61%</b>	<b>100.58%</b>	<b>100.27%</b>	<b>100.35%</b>	<b>100.20%</b>	<b>100.76%</b>	102.22%
Lewis Hamilton	100.77%	<b>100.65%</b>	<b>100.98%</b>	<b>100.31%</b>	<b>100.75%</b>	101.44%	<b>100.58%</b>	<b>100.65%</b>	<b>100.71%</b>	101.19%	<b>100.29%</b>	<b>100.72%</b>
Sergio Perez	101.32%	101.27%	101.94%	101.14%	102.21%	100.84%	<b>100.79%</b>	<b>100.78%</b>	<b>100.98%</b>	101.59%	101.70%	101.32%
Valtteri Bottas	101.49%	<b>100.80%</b>	103.59%	<b>100.51%</b>	<b>101.51%</b>	<b>100.47%</b>	102.12%	100.91%	101.31%	<b>100.92%</b>	100.77%	101.49%
Charles Leclerc	101.75%	101.84%	<b>101.33%</b>	101.52%	101.99%	101.75%	101.56%	102.51%	101.76%	101.83%	<b>100.67%</b>	101.75%
Lando Norris	101.78%	101.52%	101.52%	101.42%	102.99%	101.01%	101.84%	101.83%	101.81%	<b>100.87%</b>	101.12%	101.78%
Carlos Sainz	101.86%	101.78%	101.59%	102.00%	102.44%	<b>100.78%</b>	101.67%	102.53%	101.99%	101.71%	101.24%	101.40%
Pierre Gasly	102.09%	101.96%	102.53%	101.96%	103.14%	101.61%	101.46%	102.02%	102.09%	101.49%	101.91%	101.83%
Daniel Ricciardo	102.31%	101.80%	102.59%	101.91%	102.39%	102.53%	102.11%	102.09%	102.76%	101.83%	101.44%	103.10%
Sebastian Vettel	102.48%	103.23%	103.24%	102.24%	103.48%	101.42%	101.43%	102.12%	102.79%	102.05%	102.42%	<b>101.28%</b>
Fernando Alonso	102.48%	102.97%	103.32%	101.74%	103.80%	102.54%	102.29%	102.10%	102.61%	102.16%	102.10%	101.41%
Lance Stroll	102.60%	102.21%	102.82%	102.29%	103.44%	101.89%	101.44%	102.35%	102.59%	102.25%	102.16%	102.60%
Esteban Ocon	102.61%	102.57%	103.10%	101.65%	103.68%	102.56%	102.61%	102.77%	102.70%	102.61%	102.16%	<b>101.23%</b>
Yuki Tsunoda	102.64%	101.97%	102.23%	102.60%	102.64%	102.98%	101.68%	102.62%	102.63%	101.99%	102.23%	102.58%
Antonio Giazzi	102.83%	102.27%	103.14%	102.12%	103.76%	102.52%	102.52%	102.85%	102.80%	102.60%	102.65%	102.75%
Raikkonen-Kubica	102.84%	102.13%	103.07%	102.84%	103.73%	102.53%	102.94%	102.91%	102.65%	102.28%	102.74%	102.27%
George Russell	103.10%	102.61%	103.63%	103.45%	104.16%	102.97%	102.62%	102.71%	103.34%	102.35%	102.56%	102.53%
Nicholas Latifi	103.27%	104.15%	103.27%	104.01%	103.98%	102.97%	102.54%	103.00%	103.48%	102.87%	102.63%	102.55%
Mick Schumacher	103.82%	103.38%	103.39%	103.71%	104.28%	104.57%	102.82%	103.42%	103.95%	103.53%	103.99%	103.26%
Nikita Mazepin	104.59%	<b>104.59%</b>	104.97%	104.45%	105.29%	105.05%	103.67%	103.69%	104.81%	104.24%	103.92%	<b>104.59%</b>

(continued next page)

Driver	Average LTD	NET	ITL	RUS	TUR	USA	MEX	BRA	QAT	SA	UAE
Max Verstappen	100.76%	100.44%	100.72%	101.74%	100.81%	100.48%	100.39%	101.06%	100.50%	101.65%	100.41%
Lewis Hamilton	100.77%	100.54%	100.66%	100.93%	101.31%	100.43%	100.65%	100.92%	100.91%	101.23%	100.42%
Sergio Perez	101.32%	102.42%	100.53%	101.56%	101.14%	101.18%	100.67%	101.14%	102.06%	101.32%	101.07%
Valtteri Bottas	101.49%	101.47%	100.55%	102.26%	100.48%	101.85%	102.76%	101.15%	102.22%	101.91%	102.33%
Charles Leclerc	101.75%	102.44%	100.60%	101.68%	101.16%	101.38%	101.82%	101.87%	102.72%	102.10%	102.54%
Lando Norris	101.78%	102.96%	100.68%	100.73%	101.40%	101.94%	102.21%	102.99%	102.49%	101.85%	102.35%
Carlos Sainz	101.86%	103.02%	100.95%	101.66%	101.13%	101.89%	101.88%	101.88%	102.56%	102.70%	102.28%
Pierre Gasly	102.09%	102.38%	102.09%	102.26%	101.22%	102.09%	101.50%	102.48%	102.42%	102.96%	102.48%
Daniel Ricciardo	102.31%	103.25%	100.48%	101.66%	102.38%	101.81%	102.86%	102.63%	103.21%	102.82%	102.88%
Sebastian Vettel	102.48%	103.52%	101.54%	102.06%	102.06%	102.87%	102.02%	102.82%	102.58%	103.85%	103.11%
Fernando Alonso	102.48%	103.00%	101.34%	101.64%	102.46%	102.88%	102.19%	103.17%	102.46%	103.38%	102.59%
Lance Stroll	102.60%	103.81%	101.28%	102.12%	101.77%	103.39%	102.99%	104.09%	102.75%	103.14%	103.22%
Esteban Ocon	102.61%	103.12%	101.39%	102.23%	102.76%	103.31%	103.14%	103.14%	102.80%	102.54%	102.77%
Yuki Tsunoda	102.64%	103.79%	102.64%	102.69%	102.65%	102.84%	102.64%	103.67%	102.78%	103.26%	102.28%
Antonio Giovinazzi	102.83%	103.68%	101.43%	102.69%	102.44%	102.95%	102.97%	103.63%	103.61%	103.24%	102.76%
Raikkonen-Kubica	102.84%	104.57%	101.72%	102.30%	102.54%	103.38%	102.08%	102.85%	102.85%	103.89%	103.30%
George Russell	103.10%	103.81%	101.38%	102.09%	102.75%	103.50%	104.14%	103.98%	103.49%	103.10%	103.87%
Nicholas Latifi	103.27%	104.48%	101.51%	102.52%	102.50%	103.80%	103.98%	103.66%	103.77%	103.42%	103.58%
Mick Schumacher	103.82%	105.40%	103.11%	103.26%	104.38%	104.26%	103.82%	104.23%	104.10%	103.82%	103.52%
Nikita Mazepin	104.59%	104.99%	103.37%	104.16%	104.66%	105.39%	105.36%	105.06%	105.02%	104.59%	104.59%

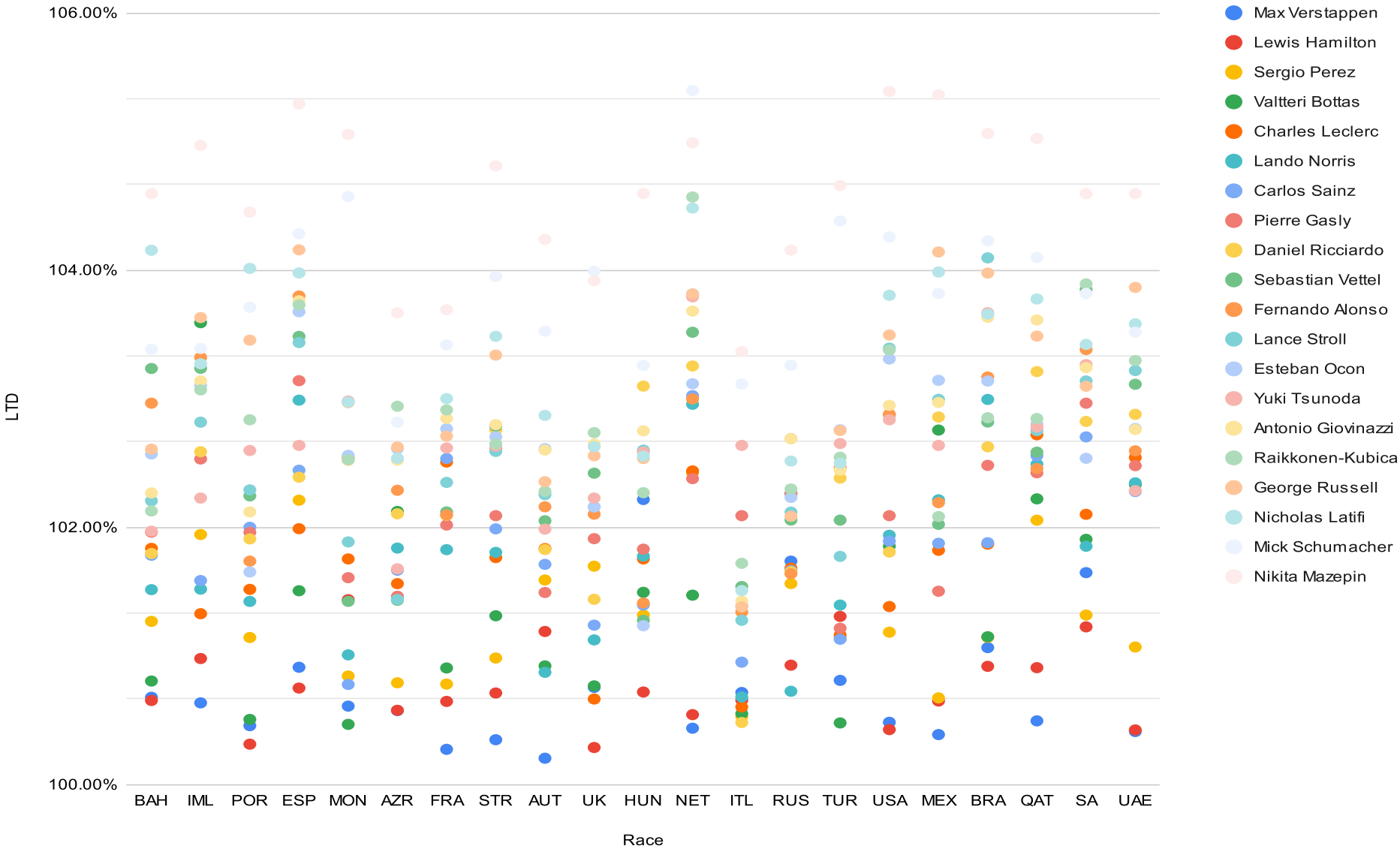
Red = Missing data; substituted with overall average LTD

### B: Five-race average LTD rank table

[illegible]



C: LTD by Race Chart



D: 5-Race Average LTD by race

