# ISyE 6740 - Spring 2021
# Project Proposal

**Team Member Names:** Garrett N. Bushnell (902319940)

**Project Title:** Formula 1 Wreck Prediction

# 1 Problem Statement

Formula 1 is the pinnacle of motor sports. This sport is characterized by its speed, level of danger, and last-but-not-least, the size of the team budgets. The sport is measured in milliseconds - in qualifying, the difference between first and second place is sometimes less than 1/1000 of a second. Since the sports start in the 1950's, 52 drivers have died. Finally, over the course of 11 months, teams will spend upwards of $450 million dollars developing and maintaining their two cars[1]. With such high stakes, it's needless to say that the teams and drivers live on the edge.

Since the start of the 2017 season, there have been 39 wrecks[3]. Given the bespoke nature of each car, wrecks can cause enormous setbacks to teams. Monetarily, car repairs can cost upwards of $1.3 million[2]. Damaged parts may have to be re-manufactured from the factory (often thousands of miles away). The team also loses valuable time gathering data and tuning their car.

The goal of this project is to develop an indicator that can serve as a gauge to how likely a particular driver is to have an accident during the next lap. Although the likelihood ultimately depends on thousands of various factors, I hypothesize that several factors from various groups could provide a reasonable predictor of disaster. Such an indicator would be extraordinarily valuable to ensure driver safety and the limit spending on repairs - as teams could continually monitor the conditions.

# 2 Data Source

A byproduct of the competitive nature is the fact that Formula 1 racing lives and breathes data. During the course of a race, each car generates upwards of 1,500 sensor readings each second, totaling around 3GB of data[4]. Although most of the data is proprietary, several sensor readings and other race-specific data are live streamed via an API. The Fast-F1 project is an open-source project that has captured, cleaned, and archived this data[5].

## 2.1 Environmental Factors

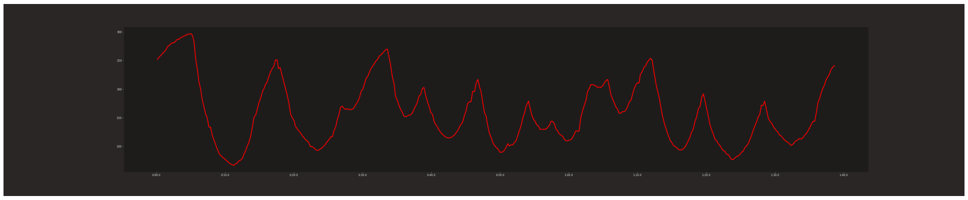
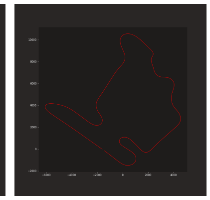
Firstly, environmental factors could influence the likelihood of a wreck. Races routinely experience high winds - which hamper the extremely aerodynamic vehicles. Rain also plays a factor in races, in which teams must select the proper tires for the current and expected track conditions. Having incorrect tire selections has lead to numerous incidents.

## 2.2 Race Factors

There are also race-related situations that could cause incidents to be more prevalent. Following an accident or break-down a safety care is deployed to shepard the drivers around the track until the situation is resolved. As the safety car bunches the cars together, there is an increase of incidents when the cars are released by the safety cars.

## 2.3 Team and Driver Factors

Finally, team and driver decisions will greatly influence the probability of a wreck. For the first few laps following a tire change, the tires are not at optimal temperature and drivers often lose control of their cars during this lull. Drivers may also be more likely to have an accident when they are attempting to overtake an opponent, which can be measured by the delta between their car and the car ahead. Finally, a driver may be at increased risk when they deviate from their normal racing line or become out-of-sync when compared to their previous few laps. Driver data consists of sensor readings denoting metrics such as speed, gear, and RPM over millisecond intervals. Positional data for each driver over each lap is also available. Example plots are included below.

Given that our variables possess different levels of granularity, we will have to have two separate data sources. One, denoted $X_l$ will be variables on a per lap basis. Another matrix of variables $X_d$, will serve to house the sensor data and the car's current position on the track.

$X_l =$

| lap | Rain | High Wind | Safety Car | Tire Age |
|---|---|---|---|---|
| 1 | 0 | 1.3 | 0 | 0 |
| 2 | 1 | 1.2 | 1 | 1 |

$X_d =$

| $Position_X$ | $Position_Y$ | Speed | Gear | RPM |
|---|---|---|---|---|
| 102 | 1256 | 135 | 4 | 1025 |
| 104 | 1324 | 142 | 5 | 998 |

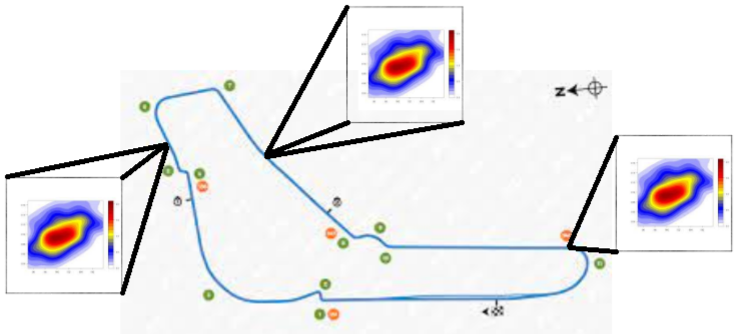
# 3 Methodology

Our problem can be posed as novelty detection. We seek to find the probability of a particular driver having a wreck in the next lap $(t+1)$, given the data in the current and preceding laps ($t$ and $t-1..n$). Our model is meant to be used on a per driver, per race basis. We are only seeking to predict the probability of an accident for driver $d$ in race $r$. By utilizing this type of model, we can ignore differences between drivers and variations between different tracks. We will let $y_{[t+1,d,r]} = 1$ denote the probability that driver $d$ in race $r$ wrecks in the next lap. In short, our model will combine a measure of systematic risk with a measure of driver-specific risk. Our model will take the form:

$$P(y_{[t+1,d,r]} = 1 | X_l, X_d) = \alpha_{[t,r]} * \pi_{[t_{0...-n},d,r]} \tag{1}$$

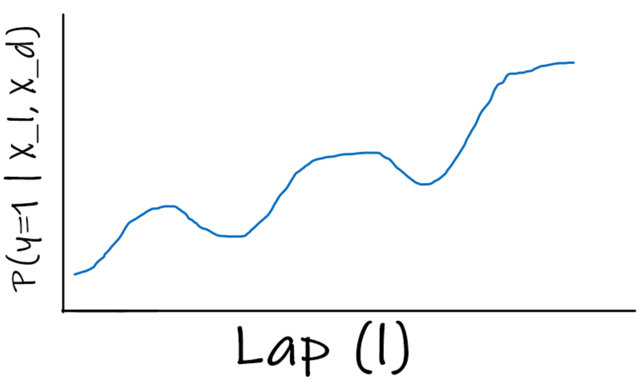
We will let $\alpha_{[t,r]}$ will denote the probability of wrecking given lap-granular data $(X_l)$. This part of the model will capture systematic risk on the track - either from environmental factors (such as rain or wind) or by race factors (such as safety car deployment or age of tires). Except for the tire age, all factors would be universal amongst all drivers for the given track.

Finally, we will let $\pi[t_{0...-n}, d, r]$ denote the likelihood of an accident given the current and past sensor and position readings. This part of the model will work on the $X_d$ data. The most interesting part of this section is that we have no prior data to rely upon. If a driver wrecks, that's the end of their race and we cannot use their "wrecked metrics" to predict another wreck. As we have no prior wrecked-events to help model, we default to using denisty estimation. It's hypothesized that the previous lap's speed, RPM, and gear at a particular point on the track should be relatively stable. If the readings at that particular point on the track during the current lap are outside of a bound or in a low-density area, it's hypothesized that the driver is off their normal line, or out of sync with recent trends - putting them at an increased risk of having an accident. Two modeling approaches were tested. One involved treating all data within a single lap as one observation, and modeling a single multivariate Gaussian Distribution. Another method involved modeling only on specific sections of the track (Ie. subsections of the single lap). For this section, we randomly chose 30 points on the track, and modeled each point as having a specific Gaussian distribution for each metric.



Exasperating the difficulties of our approach, cars continually burn fuel throughout the race and as a result become quicker and complete laps faster. As a result, the speed, RPM, etc., of the car at specific points will change over the course of the race. To account for this, it's hypothesized that a rolling window of the last $n$ laps data should be used. Using data from older periods would not be indicative of the current performance. In our model, an optimal lag of the prior 4 laps was chosen.

Although possible, I will not attempt to normalize our results and return a true probability. Instead, our "probability" will merely be a relative value - lower denoting lower likelihood of wrecking, while increases in the value over time would suggest increasing likelihood of an accident. By plotting our metric continuously throughout the course of the race, we can monitor increases and decreases in risk - similar to a normal control chart.




Our final model becomes:

$$P(y_{[t+1,d,r]} = 1|X_l, X_d) = \alpha_{[t,r]} * \pi_{[t_{0...-n},d,r]} \tag{2}$$

Where

$$\alpha_{[t,r]} = \Pi_{i=1}^{m} P_{d,i} X_{d,i} \tag{3}$$

where $P$ denotes an m-dimensional vector of conditional probabilities [Appendix A].

$$\begin{aligned} \pi_{[t_{0...-n},d,r]} &= \Pi_{j=1}^{k} \pi_j \\ &= \Pi_{j=1}^{k} (\Pi_{i=1}^{30} P(D_i | \mu_{i,t-1..t-4}, \sigma^2_{i,t-1..t-4}) \end{aligned} \tag{4}$$

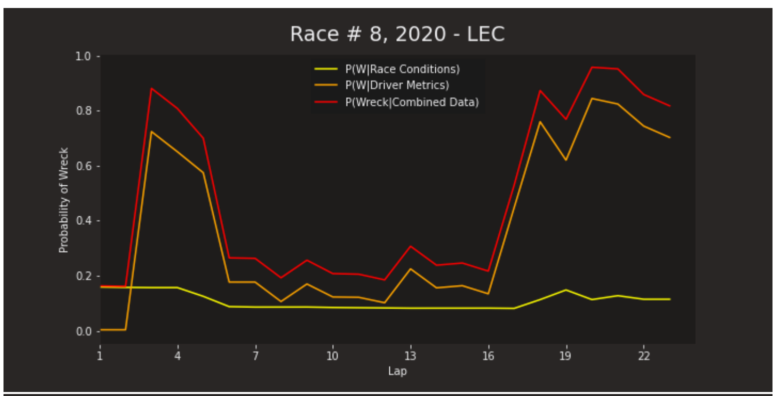
where k denotes the number of metrics (X position, speed, RPM,...)

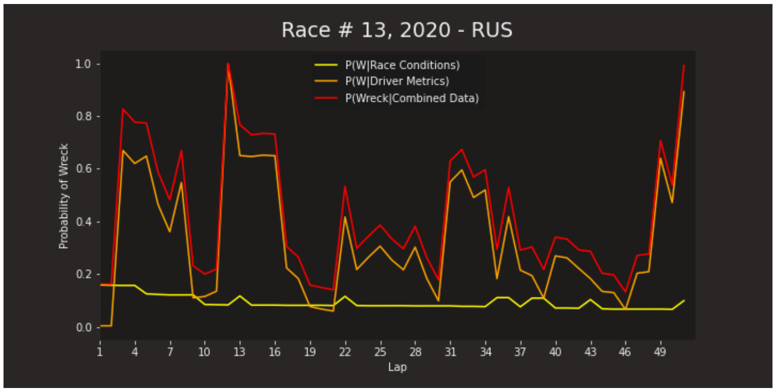
# 4 Evaluation and Final Results

Models will be fit for three random historical wrecks. Over each lap, we will calculate and chart the "probability" of wrecking on the next lap. A successful model would be characterized by a noticeable increase in the probability of an accident, in the laps immediately prior to the actual accident occurring.

## 4.1 Plots

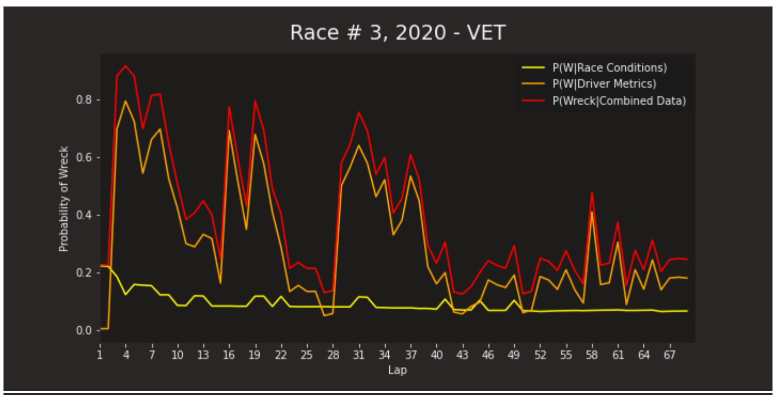
The following figures are results showing the computed probabilities for three wrecks in the 2020 and 2019 seasons.

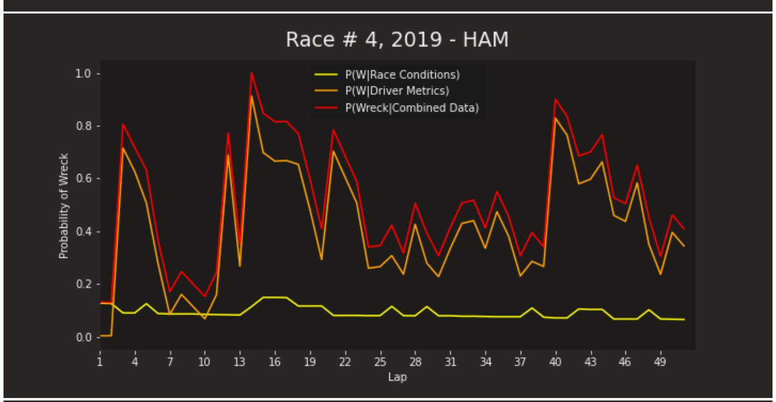


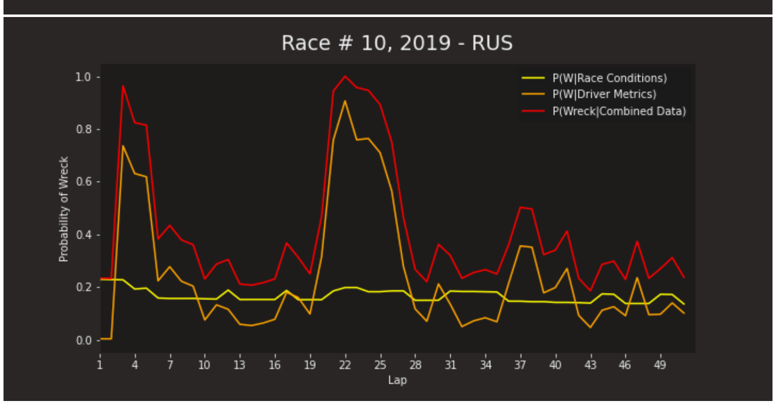




As predicted, the first and third plot show striking evidence that the driver was in danger of having an accident. Although not as striking, the second figure also shows a sharp increase in probability immediately before the driver's wreck.

In contrast, we can show results for three drivers who did not wreck.



### 4.2 Interpretation

It's noted that in all figures, the vast majority of the risk comes from the driver's metrics, with race conditions normally contributing less than 20%. A benefit of using density estimation is our ability to examine exactly what's contributed to our scores with relative ease.

If we dig back into the data, we can see the large spike in our last plot is the result of the driver being in a different gear than he normally had been at specific parts of the track, up until lap 20.

Similarly, the data shows that the increase in LeClerc's probability of wrecking increased after lap 16 due to several factors [Appendix A2]. His normal line was impacted, evidenced by his probability of the Z-coordinate of his car shifting from a 52% to over 80% in the course of 1 lap. On the same lap there was also a 2x increase in his Gear percentage. Combined, it would suggest the driver was severely out of line during this, and the following laps - not taking his normal racing line and not being in the proper gear at various parts of the track.

## 5 Future Expansions and Applications

As noted, the problem of predicting a one-off event is extremely difficult. Our results are promising, but further refining is necessary. For instance, the Probability of wrecking on several occasions in the figures is 100%. It may be preferential to switch from a probability estimate to a score on a bounded range (say from 1 to 5).

In all figures, there always appears to be at least one period of a high-probability of an accident. Further investigation should be done to see exactly what parameters are causing this, and see if their input into the final score should be reduced.

Finally, some factors are much more likely to contribute to wrecks than others. Instead of multiplying all of our probabilities with equal weight, a weighted average may be preferential. Variable selection techniques could be used to pick the most important features, while leaving out less informative ones.

## 6 Appendix

### 6.1 A

| | P(x) |
|---|---|
| P(Wreck) | 0.057018 |
| P(Wreck — High Wind) | 0.070690 |
| P(Wreck — Rain) | 0.062500 |
| P(Wreck — Last Lap was Fastest Lap) | 0.034926 |
| P(Wreck — Fresh Tires) | 0.031854 |
| P(Wreck — Last Lap was Safety Car) | 0.033940 |
| P(W— Just Finished Safety Car & Rain) | 0.108974 |
| P(W— Rain & Fresh Tires) | 0.073698 |
| P(W— Just Finished Safety Car & Fresh Tyres) | 0.046875 |

### 6.2 A1

| | X | Y | Z | RPM | Speed | nGear | Throttle | Brake | total |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 2.328306e-10 |
| 2 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 2.328306e-10 |
| 3 | 0.999696 | 0.999666 | 0.907073 | 0.999724 | 0.996168 | 0.832234 | 0.984439 | 0.995161 | 2.935081e-01 |
| 4 | 0.999645 | 0.999684 | 0.889123 | 0.999540 | 0.993887 | 0.735055 | 0.979030 | 0.993409 | 1.585648e-01 |
| 5 | 0.999748 | 0.999647 | 0.880804 | 0.999411 | 0.994697 | 0.729757 | 0.976225 | 0.992201 | 1.463924e-01 |
| 6 | 0.996306 | 0.998208 | 0.835598 | 0.999210 | 0.971339 | 0.573054 | 0.968665 | 0.500000 | 2.511374e-03 |
| 7 | 0.999841 | 0.999676 | 0.885226 | 0.999292 | 0.988287 | 0.654747 | 0.969732 | 0.500000 | 5.921869e-03 |
| 8 | 0.997391 | 0.999041 | 0.770211 | 0.999231 | 0.962102 | 0.654747 | 0.919859 | 0.500000 | 2.436939e-03 |
| 9 | 0.995313 | 0.998083 | 0.808060 | 0.999011 | 0.947364 | 0.557144 | 0.962759 | 0.500000 | 1.723648e-03 |
| 10 | 0.993752 | 0.996983 | 0.782236 | 0.998311 | 0.941151 | 0.220121 | 0.939924 | 0.500000 | 3.219700e-05 |
| 11 | 0.991285 | 0.991672 | 0.604046 | 0.996801 | 0.936946 | 0.500000 | 0.957001 | 0.500000 | 3.099180e-04 |
| 12 | 0.991229 | 0.992623 | 0.530625 | 0.998449 | 0.947699 | 0.500000 | 0.939116 | 0.500000 | 1.809695e-04 |
| 13 | 0.988352 | 0.991099 | 0.646788 | 0.996830 | 0.898214 | 0.220121 | 0.943065 | 0.500000 | 1.201876e-05 |
| 14 | 0.993279 | 0.994509 | 0.539633 | 0.998035 | 0.940664 | 0.220121 | 0.983291 | 0.500000 | 8.603831e-06 |
| 15 | 0.993147 | 0.995212 | 0.658689 | 0.999268 | 0.930191 | 0.220121 | 0.956247 | 0.500000 | 1.645365e-05 |
| 16 | 0.997838 | 0.998416 | 0.791156 | 0.998923 | 0.912383 | 0.220121 | 0.988595 | 0.500000 | 3.732251e-05 |
| 17 | 0.996181 | 0.996632 | 0.861525 | 0.998544 | 0.868895 | 0.500000 | 0.972096 | 0.500000 | 1.058036e-03 |
| 18 | 0.994758 | 0.995733 | 0.707759 | 0.998188 | 0.930303 | 0.507274 | 0.971806 | 0.500000 | 6.629743e-04 |
| 19 | 0.994854 | 0.996107 | 0.640904 | 0.998862 | 0.911279 | 0.368090 | 0.919500 | 0.500000 | 9.161477e-05 |
| 20 | 0.998556 | 0.998886 | 0.888551 | 0.998637 | 0.953066 | 0.758029 | 0.989249 | 0.500000 | 1.000583e-02 |
| 21 | 0.999776 | 0.999836 | 0.960726 | 0.999702 | 0.997237 | 0.809222 | 0.986560 | 0.992199 | 3.307962e-01 |
| 22 | 0.999966 | 0.999968 | 0.987242 | 0.999801 | 0.998574 | 0.931750 | 0.992562 | 0.994702 | 6.757020e-01 |
| 23 | 0.999897 | 0.999916 | 0.988434 | 0.999685 | 0.994733 | 0.793935 | 0.979124 | 0.993425 | 3.317222e-01 |
| 24 | 0.999944 | 0.999936 | 0.984240 | 0.999830 | 0.994820 | 0.790373 | 0.993606 | 0.993706 | 3.404829e-01 |
| 25 | 0.999926 | 0.999915 | 0.981789 | 0.999684 | 0.993202 | 0.742038 | 0.988625 | 0.993293 | 2.544123e-01 |
| 26 | 0.999903 | 0.999890 | 0.979872 | 0.999356 | 0.992614 | 0.602214 | 0.975211 | 0.989602 | 1.017564e-01 |
| 27 | 0.997043 | 0.996512 | 0.803756 | 0.998954 | 0.980168 | 0.368090 | 0.980999 | 0.984402 | 5.968233e-03 |
| 28 | 0.994918 | 0.995391 | 0.709452 | 0.998118 | 0.933843 | 0.368090 | 0.974954 | 0.500000 | 1.906646e-04 |
| 29 | 0.997037 | 0.996937 | 0.736826 | 0.997305 | 0.925107 | 0.220121 | 0.946792 | 0.500000 | 2.458130e-05 |
| 30 | 0.989391 | 0.991835 | 0.612424 | 0.996541 | 0.944203 | 0.794426 | 0.944613 | 0.500000 | 2.026694e-03 |
| 31 | 0.991659 | 0.995489 | 0.573002 | 0.999009 | 0.943218 | 0.516059 | 0.987293 | 0.500000 | 3.399425e-04 |
| 32 | 0.984132 | 0.991896 | 0.553665 | 0.998050 | 0.934173 | 0.220121 | 0.904901 | 0.500000 | 6.343336e-06 |
| 33 | 0.997152 | 0.998467 | 0.721584 | 0.998922 | 0.909664 | 0.220121 | 1.000000 | 0.500000 | 2.665057e-05 |
| 34 | 0.992306 | 0.995831 | 0.855853 | 0.998365 | 0.942226 | 0.220121 | 0.956342 | 0.500000 | 4.916760e-05 |
| 35 | 0.988349 | 0.995763 | 0.770054 | 0.997788 | 0.859914 | 0.220121 | 0.955877 | 0.500000 | 2.190101e-05 |
| 36 | 0.985849 | 0.992756 | 0.606332 | 0.998470 | 0.949725 | 0.794426 | 0.956313 | 0.500000 | 2.087764e-03 |
| 37 | 0.998337 | 0.998788 | 0.775058 | 0.998267 | 0.935574 | 0.991136 | 0.995830 | 0.500000 | 1.609875e-02 |
| 38 | 0.995856 | 0.996978 | 0.957891 | 0.998270 | 0.932960 | 0.793436 | 0.999894 | 0.500000 | 1.523979e-02 |
| 39 | 0.998516 | 0.999249 | 0.757465 | 0.997954 | 0.923741 | 0.528148 | 0.969439 | 0.500000 | 1.012033e-03 |
| 40 | 0.998461 | 0.999226 | 0.816949 | 0.997771 | 0.954932 | 0.528148 | 0.967405 | 0.500000 | 1.549207e-03 |
| 41 | 0.997978 | 0.998794 | 0.841613 | 0.997605 | 0.866131 | 0.753597 | 0.990550 | 0.500000 | 5.357331e-03 |
| 42 | 0.997471 | 0.998715 | 0.763837 | 0.998179 | 0.924927 | 0.274044 | 0.963042 | 0.500000 | 7.385381e-05 |
| 43 | 0.882924 | 0.991579 | 0.636342 | 0.996526 | 0.868380 | 0.220121 | 0.883688 | 0.500000 | 4.833734e-06 |
| 44 | 0.993466 | 0.995268 | 0.650090 | 0.995805 | 0.805949 | 0.475320 | 0.914332 | 0.500000 | 1.579185e-04 |
| 45 | 0.998533 | 0.999027 | 0.805737 | 0.995649 | 0.835840 | 0.410896 | 0.914332 | 0.500000 | 2.492731e-04 |
| 46 | 0.982239 | 0.994450 | 0.619613 | 0.998698 | 0.839380 | 0.516059 | 0.698529 | 0.500000 | 6.993077e-05 |
| 47 | 0.998042 | 0.998565 | 0.919305 | 0.998716 | 0.998980 | 0.516059 | 1.000000 | 0.500000 | 3.094650e-03 |
| 48 | 0.996591 | 0.997286 | 0.742108 | 0.997889 | 0.894550 | 0.318767 | 0.910115 | 0.500000 | 8.319967e-05 |
| 49 | 0.996080 | 0.997336 | 0.734793 | 0.997443 | 0.922198 | 0.318767 | 0.908875 | 0.500000 | 8.950385e-05 |
| 50 | 0.993831 | 0.995445 | 0.782780 | 0.998594 | 0.966241 | 0.411556 | 0.910938 | 0.500000 | 3.848825e-04 |
| 51 | 0.993895 | 0.995786 | 0.734012 | 0.999117 | 0.960882 | 0.318767 | 0.913050 | 0.500000 | 1.061051e-04 |

### 6.3 A2

| | X | Y | Z | RPM | Speed | nGear | Throttle | Brake | total |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.000015 |
| 2 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.500000 | 0.000015 |
| 3 | 0.999150 | 0.999689 | 0.952203 | 0.999074 | 0.991173 | 0.803590 | 0.963506 | 0.992599 | 0.523926 |
| 4 | 0.998480 | 0.999579 | 0.941897 | 0.999215 | 0.991403 | 0.735541 | 0.957525 | 0.992041 | 0.423362 |
| 5 | 0.997587 | 0.999371 | 0.928165 | 0.999322 | 0.988630 | 0.667118 | 0.951302 | 0.989898 | 0.329844 |
| 6 | 0.987341 | 0.996283 | 0.806133 | 0.998695 | 0.931420 | 0.317648 | 0.762764 | 0.991149 | 0.031378 |
| 7 | 0.973916 | 0.992362 | 0.644386 | 0.997602 | 0.878059 | 0.426485 | 0.765873 | 0.993774 | 0.031357 |
| 8 | 0.971729 | 0.986417 | 0.522251 | 0.998424 | 0.887113 | 0.344442 | 0.703337 | 0.990755 | 0.011325 |
| 9 | 0.968876 | 0.992579 | 0.547433 | 0.998031 | 0.897546 | 0.460879 | 0.790320 | 0.990602 | 0.028954 |
| 10 | 0.981104 | 0.991233 | 0.571444 | 0.997354 | 0.913503 | 0.304675 | 0.800582 | 0.996398 | 0.015143 |
| 11 | 0.967049 | 0.987153 | 0.558849 | 0.997805 | 0.884358 | 0.294106 | 0.886358 | 0.992194 | 0.014826 |
| 12 | 0.987074 | 0.994978 | 0.640368 | 0.998111 | 0.911615 | 0.220121 | 0.816801 | 0.990923 | 0.010395 |
| 13 | 0.957507 | 0.985327 | 0.565572 | 0.996501 | 0.882781 | 0.564704 | 0.857059 | 0.990539 | 0.050640 |
| 14 | 0.975516 | 0.985056 | 0.507379 | 0.997795 | 0.869456 | 0.437682 | 0.850765 | 0.991641 | 0.024394 |
| 15 | 0.972649 | 0.991589 | 0.584171 | 0.997890 | 0.838789 | 0.394170 | 0.888342 | 0.992965 | 0.026886 |
| 16 | 0.970056 | 0.995590 | 0.529608 | 0.997513 | 0.907038 | 0.344442 | 0.849920 | 0.992790 | 0.018091 |
| 17 | 0.964506 | 0.995088 | 0.801952 | 0.998804 | 0.957591 | 0.753580 | 0.814428 | 0.991965 | 0.200867 |
| 18 | 0.999989 | 1.000000 | 0.997370 | 0.999703 | 0.993056 | 0.855432 | 0.902148 | 0.994007 | 0.576896 |
| 19 | 0.997963 | 0.999751 | 0.973347 | 0.999538 | 0.991027 | 0.703939 | 0.921992 | 0.992968 | 0.384337 |
| 20 | 0.998020 | 0.999764 | 0.981566 | 0.999718 | 0.995374 | 0.876623 | 0.995183 | 0.992643 | 0.712291 |
| 21 | 0.999372 | 0.999735 | 0.979876 | 0.999829 | 0.996035 | 0.856739 | 0.992659 | 0.993606 | 0.678723 |
| 22 | 0.999263 | 0.999711 | 0.978704 | 0.999766 | 0.994659 | 0.776233 | 0.992387 | 0.992949 | 0.553040 |
| 23 | 0.993154 | 0.999142 | 0.925222 | 0.999697 | 0.994722 | 0.793918 | 0.975442 | 0.993539 | 0.493451 |

## References


[1] It Cost Mercedes $459M to Win the F1 Championship in 2020
`https://www.thedrive.com/accelerator/40525/it-cost-mercedes-459m-to-win-the-f1-championship-`

[2] How a million pound crash highlighted F1's money squeeze
`https://us.motorsport.com/f1/news/bottas-million-pound-crash-cost-cap-analysis/6433104/`

[3] List of Formula One fatalities
`https://en.wikipedia.org/wiki/List`$_{o}f_{F}ormula_{O}ne_{f}atalities$

[4] Formula 1 Case Study
`https://aws.amazon.com/solutions/case-studies/formula-one/`

[5] The Fast F1 Project
`https://theoehrly.github.io/Fast-F1/fastf1.html`