www.associazionenemesis.it

STATISTICA IN POLE

A cura di Bellinaso Marco, Bergamo Tommaso, Bonato Edoardo e Bortoliero Daniele

La Formula Uno è uno sport in cui il talento del pilota non viene risaltato se messo alla guida di una vettura poco performante, così come una vettura veloce ha bisogno di un pilota che riesca a concretizzare le sue potenzialità. Tuttavia, anche se si riesce a trovare il connubio "monoposto-pilota" ideale ci si scontra sempre con la fortuna, capace di portare ad eventi che rendono più imprevedibile e spettacolare questo sport, ma, al contempo, meno meritocratico.
Se si riuscisse a valutare matematicamente solo la pura performance di un uomo e la sua macchina, eliminando tutte le variabili che potrebbero condizionarne la prestazione, chi avrebbe meritato di vincere il campionato mondiale di F1 nel 2021?

Percorsi per le Competenze Trasversali e per l'Orientamento

I Percorsi per le Competenze Trasversali e per l'Orientamento (P.C.T.O.) hanno recentemente preso il posto dell'alternanza scuola-lavoro (A.S.L.), introdotta con la L.107/2015, nota come "Buona Scuola". Le finalità sono però rimaste sostanzialmente inalterate; l'obiettivo dichiarato è indirizzare la didattica verso un metodo di apprendimento in sintonia con le esigenze del mondo esterno. Nel fare questo coinvolge anche gli adulti, siano essi insegnanti (tutor interni) o referenti della realtà ospitante (tutor esterni).

È importante sottolineare che i partner educativi esterni non sono obbligatoriamente imprese ed aziende; possono essere, infatti, anche associazioni sportive, enti culturali e istituzionali e ordini professionali. L'obiettivo è quello di sviluppare in modo condiviso alcune esperienze coerenti alle attitudini e alle passioni degli studenti.

I percorsi realizzati durante il P.C.T.O. sono co-progettati, attuati, verificati ed opportunamente valutati sulla base di opportune convenzioni con imprese, associazioni di rappresentanza, con enti pubblici e privati, inclusi quelli del terzo settore, che siano disposti ad accogliere gli studenti per periodi di apprendimento in situazione lavorativa, che non si configurano, però, in rapporto individuale di lavoro (D.Lgs. 15/4/2005, n. 77).

I licei

Da molti anni ormai, il mondo degli istituti tecnico-professionali si è affacciato all'esterno della scuola, investendo tempo ed energie in percorsi formativi che prevedessero collaborazioni con enti ed aziende del territorio. La novità, già presente con l'A.S.L., è l'estensione di questo tipo di formazione anche ai licei, per i quali però l'obiettivo da perseguire è anche quello di orientare gli studenti verso la prosecuzione dei loro studi. Le attività sono obbligatorie per tutti gli studenti e, nel caso dei licei, prevedono attualmente lo svolgimento di un minimo di 90 ore nell'arco del triennio.

Chi è lo Studente RicercAttore?

È innanzitutto uno *Studente* che si affaccia per la prima volta nel mondo della *Ricerca Scientifica* e vi recita un ruolo da *Attore protagonista*.

Fare scienza non significa chiaramente solo studiare modelli matematici preconfezionati da altri, ma "sporcarsi le mani" raccogliendo dati ed elaborandoli al fine di ottenere delle conclusioni che confermino o smentiscano un'ipotesi preliminare di ricerca.

Come spesso accade per un ricercatore scientifico, può succedere che lo studente durante il P.C.T.O. abbia la sensazione di brancolare nel buio, fatichi a comprendere il significato dei dati in suo possesso; ma non è mai solo, lavora in equipe e collabora con altri, discute, dibatte.

Qui si innesta anche l'azione del tutor che deve riuscire a fornire una stampella in caso di necessità, senza intervenire direttamente nelle dinamiche del gruppo orientandone la ricerca; deve fornire stimoli e strumenti affinché gli studenti sviluppino durante il progetto di P.C.T.O. le necessarie competenze d'indagine, risvegliando in loro la curiosità e sostenendone la fiducia di poter capire.

È necessario risvegliare anche la creatività in campo scientifico, cioè aggredire i problemi utilizzando – se necessario – approcci diversi, attivare connessioni nel pensiero, renderlo flessibile ed in grado di adattarsi alle esigenze del progetto di ricerca, formulare ipotesi e validarle alla luce dei dati raccolti.

È chiaro che, per poter ottenere un simile risultato, non ci si può basare su rigidi schematismi, ma si devono stimolare i ragazzi a formulare congetture; e questo non lo si può certo inventare in pochi minuti: è necessario che la didattica venga rimodellata perché i ragazzi si abituino a proporre idee e nuovi punti di vista.

L'evoluzione scientifica sempre più frenetica che permea di sé la società moderna ci spinge a promuovere una cultura scientifica basata sulla ricerca, l'informazione e la conoscenza.
La scuola deve agire per consolidare quanto più possibile le conoscenze scientifiche degli studenti, per consentire loro di orientarsi al meglio in un mondo che cambia così rapidamente.
Per poterci riuscire non si può rinunciare a ricorrere alla Statistica, una disciplina che acquisisce sempre maggiore rilievo vista la necessità di elaborare grandi moli di dati in vari ambiti della scienza, e di *"farli parlare"* aiutandoci a comprenderne il significato.
Uno degli obiettivi di questo progetto è, pertanto, quello di fornire ai ragazzi quelle basi di statistica necessarie ad orientarsi in ambito scientifico che, però, faticano a trovare spazio nella normale didattica in classe.
Un ricercatore, poi, presenta i risultati della sua ricerca sia scrivendo articoli che rispettino un preciso codice comunicativo sia tenendo conferenze e seminari.
L'attività di ricerca del nostro progetto culmina, quindi, con la stesura di un articolo divulgativo e la presentazione pubblica dei risultati nel corso di alcune conferenze tematiche.
Questo consente ai ragazzi di trovare una modalità comunicativa efficace sia in forma scritta che orale.

In cosa consiste, in sostanza, il nostro progetto di P.C.T.O.?
Come abbiamo visto, gli studenti sono stati impegnati in un *project work* nel quale hanno simulato l'attività di un gruppo di ricerca; hanno analizzato dati provenienti da fonti diverse: sensori presenti in Laboratorio di Fisica, da banche dati open italiane e straniere. Alcuni gruppi hanno potuto beneficiare anche del *know how* e dei materiali resi accessibili dal CERN di Ginevra.
I dati raccolti sono stati elaborati utilizzando gli strumenti della statistica descrittiva ed hanno consentito loro di scrivere delle relazioni come questa, secondo un protocollo di divulgazione scientifica concordato assieme al nostro partner principale (l'Associazione Nemesis). In questa fase gli studenti hanno anche esercitato un "controllo incrociato" sulle relazioni dei vari gruppi, svolgendo attività di *"blind peer reviewer"*, tipica dei comitati di redazione delle riviste scientifiche. L'ultima fase del progetto ha previsto l'esposizione pubblica dei risultati conseguiti, mediante alcune serate di divulgazione aperte alla cittadinanza.

I nostri partner
Ringraziamo innanzitutto l'Associazione Nemesis, che ha tra i suoi obiettivi la diffusione della cultura scientifica nella società civile, con particolare riferimento al mondo della scuola, e che ospita nel suo sito una sezione che raccoglie i lavori preparati dagli studenti.
Un grazie particolare al CERN di Ginevra per aver reso accessibili dati e pubblicazioni, utilizzati dai ragazzi per il percorso di ricerca sulla fisica particellare.
Non meno importante è stato il supporto fornito dal laboratorio Big & Open Data Innovation (http://bodai.unibs.it/) dell'Università di Brescia. Nato grazie al supporto di Fondazione Cariplo con l'obiettivo di creare gruppi di lavoro che studiano e sviluppano – attraverso specifici progetti di ricerca – nuovi metodi, tecniche e strumenti per raccolta, gestione e analisi di dati multidisciplinari, merita la nostra gratitudine per averci fornito, tramite il progetto BDsports, il *know how* necessario per sviluppare il progetto.
Un ringraziamento doveroso va infine al Comune di Thiene, che ci ha messo a disposizione gli spazi idonei a tenere le serate in cui i ragazzi hanno potuto mettersi alla prova come divulgatori.

STATISTICA IN POLE

Indice dei contenuti

1. Introduzione e obiettivi dell'analisi.....5

2. L'analisi.....5

2.1 Scelta dei parametri.....5

2.2 Qualifica.....6

2.3 Gara.....7

3. Risultati e conclusioni.....12

3.1 Qualifica.....12

3.2 Gara.....15

APPENDICE: le formule.....20

A1. Analisi delle qualifiche.....20

A2. Analisi delle gare.....21

A2.1 Preparazione dei giri.....21

A2.2 Selezione dei giri regolari.....22

A2.3 Prodotto tra il numero del giro [clean lap n°] e il tempo sul tiro [clean lap time].....23

A2.4 Pendenza della retta dei minimi quadrati dei giri corrispondenti di ogni stint [a].....23

A2.5 Pendenza delle rette dei minimi quadrati [a_{tot}].....24

A2.6 Baricentro del grafico a dispersione dei giri regolari [M_x, M_y].....24

A2.7 Intercetta della retta dei minimi quadrati dei giri corrispondenti di ogni stint [b_{tot}].....25

A2.8 Tempo percentuale di ogni giro rispetto al tempo sul giro ideale [$\%stint_n$].....25

A2.9 Media dei tempi percentuali [%M].....25

A2.10 Differenza di tempo in percentuale dei giri con pitstop [$\Delta\%_{pitstop}$].....26

A2.11 Media dei tempi percentuali con pitstop [$\%M_{pitstop}$].....26

1. INTRODUZIONE E OBIETTIVI

Il mondiale di Formula 1 del 2021 è stato sicuramente uno dei più movimentati ed emozionanti degli ultimi anni. La lotta per il titolo mondiale tra Lewis Hamilton e Max Verstappen è stata certamente l'evento di maggiore interesse, ma si sono viste altre rivalità minori, come il ritorno dello storico duello tra McLaren e Ferrari.
La Formula Uno è uno sport in cui il talento del pilota non viene risaltato se messo alla guida di una vettura poco performante, così come una vettura veloce ha bisogno di un pilota che riesca a concretizzare le sue potenzialità. Tuttavia, anche se si riesce a trovare il connubio "monoposto-pilota" ideale ci si scontra sempre con la fortuna: l'essere coinvolti in un incidente senza averne colpa, un errore dei meccanici al pitstop (con la conseguente perdita di tempo e posizioni), ma anche un errore giuridico da parte dei commissari di gara. Tutti questi eventi rendono più imprevedibile e spettacolare questo sport, ma, al contempo, meno meritocratico.
Ed è proprio a partire da queste considerazioni che siamo arrivati a formulare l'obiettivo della nostra indagine: eliminando gli eventi eccezionali e ogni tipo di condizionamento esterno, se venisse considerata solamente la costanza del singolo pilota abbinata alla prestazione della sua vettura, chi avrebbe meritato di vincere il campionato mondiale 2021?

2. ANALISI

2.1 SCELTA DEI PARAMETRI

Semplificando la valutazione di un pilota in poche macrocategorie, i parametri fondamentali che indicano le sue capacità sono: velocità sul giro secco, passo gara e racecraft (cioè la capacità di sorpassare, difendersi e approfittare di tutte le occasioni che si presentano). Quest'ultimo indicatore può essere valutato solo attraverso l'analisi di una grande mole di dati, spesso accessibili solo alle scuderie, e in modo almeno in parte soggettivo, quindi è stato escluso da quest'analisi. Per quanto riguarda gli altri due indici, saranno loro i principali oggetti della nostra analisi e verranno ricavati rispettivamente dai dati provenienti dalle qualifiche e dalle gare.
Per effettuare la nostra indagine abbiamo deciso di considerare i Gran Premi dopo la pausa estiva, in quanto le monoposto a questo punto della stagione sono, generalmente, già arrivate a uno stadio di sviluppo avanzato. In questo caso particolare inoltre, possiamo dire che con il 2021 si è concluso un ciclo per passare, nel 2022, a monoposto radicalmente diverse in tutte le aree tecniche. Non potendo essere trasferito molto know-how da una stagione all'altra è ancor

più probabile che i team abbiano fermato il più presto possibile lo sviluppo delle vetture, concentrandosi solo nella loro comprensione e messa a punto.

2.2 Qualifica

La fase di qualificazione si divide in tre momenti: Q1, Q2 e Q3. In ogni sessione, vi è un limite di tempo per tentare di far segnare il giro più veloce, scaduto il quale vengono eliminati i 5 piloti che hanno fatto segnare il tempo più lento. Nelle fasi di Q1 e Q2, ai fini della classifica finale, vengono contati solo i tempi dei piloti eliminati, mentre i tempi dei piloti "salvi" vengono azzerati. Quindi, è solo quando si arriva in Q3 che i piloti hanno la possibilità di lottare per partire in prima posizione.
Da tali considerazioni si può intuire che il tempo segnato in Q3 può non essere il giro più veloce percorso durante le tre sessioni di qualifica. Ciò può dipendere da vari fattori:

1. *mancanza di un set di gomme nuovo;*
2. *errore del pilota;*
3. *errore del team (ad es. uscita troppo ritardata del pilota rispetto al tempo limite);*
4. *problemi tecnici (ad es. problemi nell'accensione della monoposto);*
4. *mutamento delle condizioni della pista.*

Perciò, per valutare la prestazione "pilota-vettura" in qualifica, abbiamo utilizzato il tempo del giro più veloce tra le 3 sessioni di qualifica. Inoltre non abbiamo utilizzato le qualifiche di tutti i 12 Gran Premi che hanno avuto luogo dopo della pausa estiva ma solo delle 10 avvenute completamente in condizioni di pista asciutta.
L'esempio riportato rappresenta le qualifiche del Gran Premio d'Olanda. Si può notare come la classifica ufficiale (Tab.1) differisca da quella prodotta dalla nostra analisi (Tab.2): infatti, certi piloti (evidenziati in colore) non hanno fatto segnare il loro giro più veloce nella fase finale di qualifica e, quindi, sono inevitabilmente scalati in classifica.

Pos	Driver	time Q1	time Q2	time Q3	Q1 (s)	Q2 (s)	Q3 (s)
1	VER	1:10.036	1:09.071	1:08.885	70,036	69,071	68,885
2	HAM	1:10.114	1:09.726	1:08.923	70,114	69,726	68,923
3	BOT	1:10.219	1:09.769	1:09.222	70,219	69,769	69,222
4	GAS	1:10.274	1:09.541	1:09.478	70,274	69,541	69,478
5	LEC	1:09.829	1:09.437	1:09.527	69,829	69,437	69,527
6	SAI	1:10.022	1:09.870	1:09.537	70,022	69,87	69,537
7	GIO	1:10.050	1:10.033	1:09.590	70,05	70,033	69,59
8	OCO	1:10.179	1:09.919	1:09.933	70,179	69,919	69,933
9	ALO	1:10.435	1:10.020	1:09.956	70,435	70,02	69,956
10	RIC	1:10.255	1:09.865	1:10.166	70,255	69,865	70,166
11	RUS	1:10.382	1:10.332		70,382	70,332	
12	STR	1:10.438	1:10.367		70,438	70,367	
13	NOR	1:10.489	1:10.406		70,489	70,406	
14	LAT	1:10.093	1:11.161		70,093	71,161	
15	TSU	1:10.462	1:11.314		70,462	71,314	
16	PER	1:10.530			70,53		
17	VET	1:10.731			70,731		
18	KUB	1:11.301			71,301		
19	SCH	1:11.387			71,387		
20	MAZ	1:11.875			71,875		

Tab. 1 - Risultati qualifiche GP Olanda

Average lap
70,11442

Tab. 1.1 - Tempo medio di qualifica

Pos	Driver	Fastest lap	Fastest lap (%)
1	VER	68,885	98,24655
2	HAM	68,923	98,30075
3	BOT	69,222	98,72719
4	LEC	69,437	99,03383
5	GAS	69,478	99,09231
6	SAI	69,537	99,17646
7	GIO	69,59	99,25205
8	RIC	69,865	99,64426
9	OCO	69,919	99,72128
10	ALO	69,956	99,77405
11	LAT	70,093	99,96945
12	RUS	70,332	100,3103
13	STR	70,367	100,3602
14	NOR	70,406	100,4159
15	TSU	70,462	100,4957
16	PER	70,53	100,5927
17	VET	70,731	100,8794
18	KUB	71,301	101,6923
19	SCH	71,387	101,815
20	MAZ	71,875	102,511

Tab. 2 - Risultati elaborati

Il dato "Fastest lap (%)" (Tab.2) è stato ottenuto dividendo il tempo più veloce per il giro medio (Average lap, Tab.1.1), ovvero per la media di tutti i tempi delle tre sessioni di qualifica.

2.3 Gara

Ancora più importante per un pilota della capacità di centrare un unico giro veloce, è l'abilità di rimanere costantemente veloci per più di 300km. Per valutare le abbinate pilota-vettura sotto questo punto di vista, anche in questo caso abbiamo escluso i Gran Premi corsi anche solo per alcuni giri in condizioni di pista bagnata: in queste situazioni infatti la progressione del tracciato, con l'asciugarsi o il bagnarsi dell'asfalto a seconda del presentarsi o meno di nuove precipitazioni, è l'elemento più influente ma anche imprevedibile, e finisce col determinare un'importanza predominante della strategia. Inoltre non abbiamo utilizzato i dati provenienti dalle Sprint Race in quanto essendo una novità introdotta nella stagione 2021 non sono ancora facilmente accessibili.

L'analisi di una gara è stata più complessa rispetto al lavoro svolto per le qualifiche. Abbiamo dovuto infatti tenere in considerazione un certo numero di variabili: non tutti i giri percorsi durante una gara permettono al pilota di percorrere le traiettorie ideali per poter massimizzare la velocità, quindi abbiamo eliminato i giri percorsi durante queste situazioni:

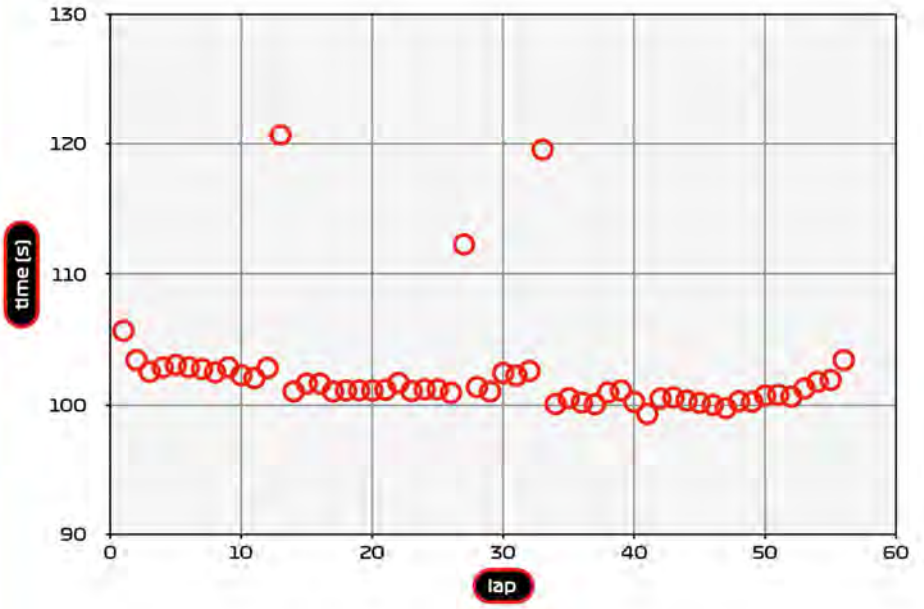
1. *primo giro di gara e giro di partenza con standing start dopo di una bandiera rossa*: in quanto il giro viene iniziato da una posizione statica;
2. *giro della sosta ai box e quello successivo*: in questo caso, il tempo viene condizionato dal tempo della sosta, ma anche dal tempo affinché le gomme entrino in temperatura (in media 1 giro). Si noti che per il Gran Premio degli USA per ogni pitstop abbiamo considerato come non utilizzabile un solo giro, e non due come negli altri casi, in quanto la linea del traguardo si trova molto vicina all'entrata ai box e quindi la tornata di rientro ai box non è significativamente rallentata;
3. *eventuali giri in regime di safety car o virtual safety car*: le vetture sono costrette a rallentare;
4. *giri di rientro da una bandiera rossa:* le vetture girano a velocità ridotta per ispezionare la pista e preparare le gomme;
5. *tutti i giri percorsi al di sotto di 4 secondi dall'auto davanti e 1 secondo dall'auto dietro*: il pilota non è nelle condizioni di poter percorrere le traiettorie ideali a causa delle manovre per attaccare la vettura davanti o difendersi da quella che lo segue e/o è influenzato dalle turbolenze o dalla scia generate della vettura che lo precede. I valori limite sono stati ottenuti da uno studio dell'ingegner Pat Symonds.

Per ognuno degli otto Gran Premi, dopo aver ottenuto i giri di ciascun pilota (Tab. 3 e Graf. 4[1]) ed aver eliminato i piloti ritirati per problemi tecnici, la cui gara risulta compromessa, abbiamo eliminato tutte le tornate non rilevanti (Tab. 5 e Graf. 6), ottenendo così solo i giri pertinenti.

[1] Per gli esempi di questo paragrafo useremo sempre i giri di Leclerc durante il GP degli USA.

Driver: LEC	
lap	time (s)
1	105,695
2	103,439
3	102,544
4	102,876
5	103,07
6	102,894
7	102,758
8	102,504
9	102,881
10	102,274
11	102,056
12	102,831
13	120,732
14	101,037
15	101,634
16	101,629
17	101,035
18	101,107
19	101,12
20	101,107
21	101,164
22	101,691
23	101,065
24	101,198
25	101,194
26	100,933
27	112,284
28	101,371
29	101,033
30	102,445
31	102,246
32	102,603
33	119,567
34	100,065
35	100,519
36	100,184
37	100,061
38	100,957
39	101,103
40	100,229
41	99,303
42	100,484
43	100,573
44	100,335
45	100,162
46	100,01
47	99,753
48	100,281
49	100,245
50	100,712
51	100,778
52	100,632
53	101,265
54	101,776
55	101,912
56	103,442

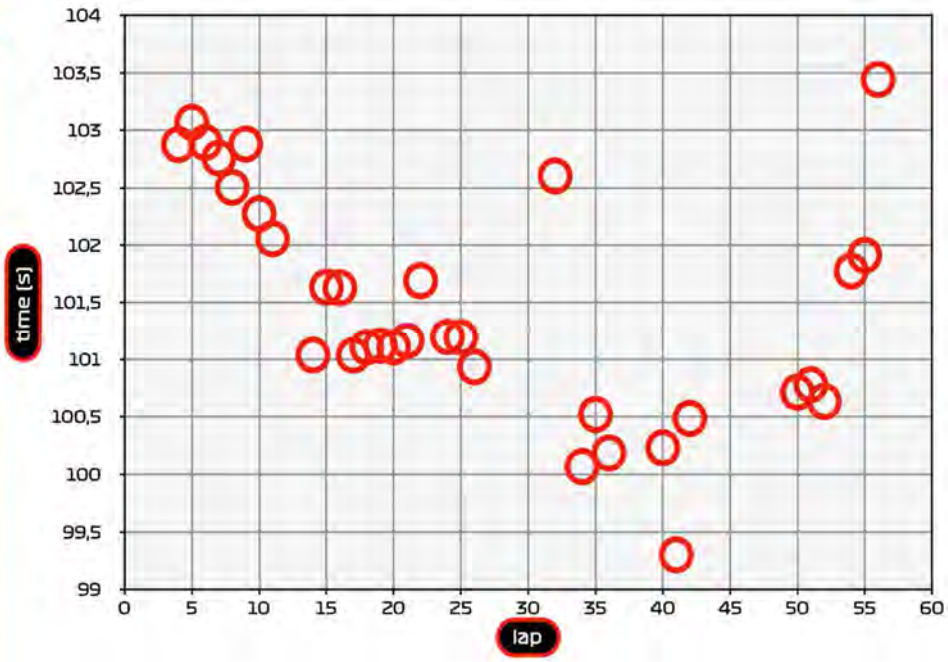
Tab. 3 - Giri Leclerc, GP USA



Graf. 4 - Rappresentazione passo gara complessivo

Driver: LEC	
lap	time (s)
4	102,876
5	103,07
6	102,894
7	102,758
8	102,504
9	102,881
10	102,274
11	102,056
14	101,037
15	101,634
16	101,629
17	101,035
18	101,107
19	101,12
20	101,107
21	101,164
22	101,691
24	101,198
25	101,194
26	100,933
32	102,603
34	100,065
35	100,519
36	100,184
40	100,229
41	99,303
42	100,484
50	100,712
51	100,778
52	100,632
54	101,776
55	101,912
56	103,442

Tab. 5 - Giri puliti Leclerc, GP USA



Graf. 6 - Rappresentazione giri puliti

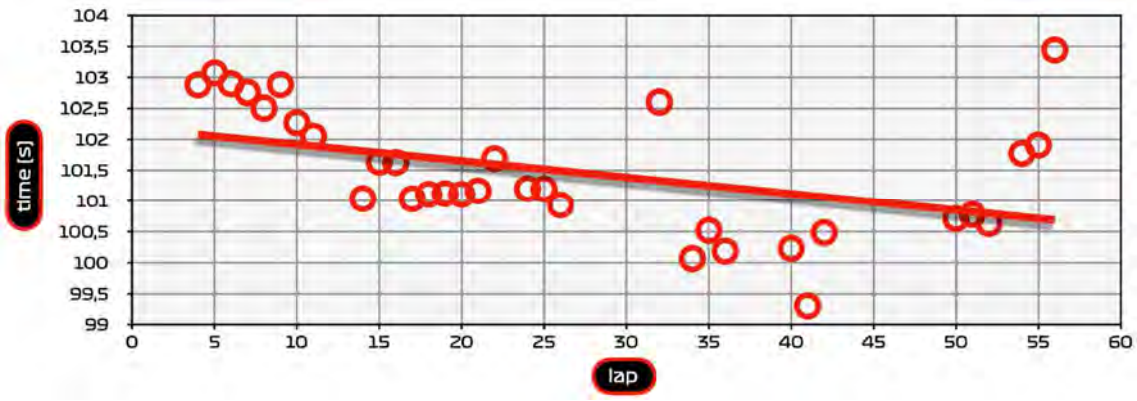
A questo punto però si presenta il problema maggiore dell'analisi del passo gara: non tutti i giri presentano condizioni uguali, quindi i tempi sul giro non hanno tutti lo stesso valore. Due fattori, infatti, permettono di far segnare tempi sempre più brevi a mano a mano che la corsa procede:

1. *le automobili si alleggeriscono,* principalmente a causa del consumo di carburante;
2. *la superficie della pista diventa più aderente,* poiché si deposita su di essa uno strato sempre maggiore di gomma proveniente dagli pneumatici che si consumano (questo è un fattore meno influente del primo, ma comunque da considerare).

Questi due fattori permettono di far segnare tempi sempre più veloci a mano a mano che la corsa procede. Di conseguenza, se l'analisi si fermasse qui, un pilota che avesse fatto segnare molti giri puliti nella prima parte di gara sarebbe molto svantaggiato rispetto ad un altro, magari meno veloce nella realtà, ma che avesse siglato la maggior parte dei suoi giri puliti nel finale di gara. Per risolvere questo problema, abbiamo pensato di usare le linee di tendenza, che abbiamo studiato durante l'anno scolastico.
A colpo d'occhio, potremmo scegliere di applicare il metodo dei minimi quadrati su due diversi set di dati, ovvero:

1. *su tutti i giri del Gran Premio (Graf. 7, in questo caso sono tutti i giri di un pilota ma il discorso è analogo);*
2. *sui giri di ciascuno stint di ogni pilota (Graf. 8[2]).*

In quest'ultimo caso però la linea di tendenza andrebbe a identificare principalmente l'evoluzione dei tempi in base al degrado delle gomme, mentre nel primo caso si andrebbe ad identificare una via di mezzo tra quello che ci interessa e la progressione dovuta alle coperture.

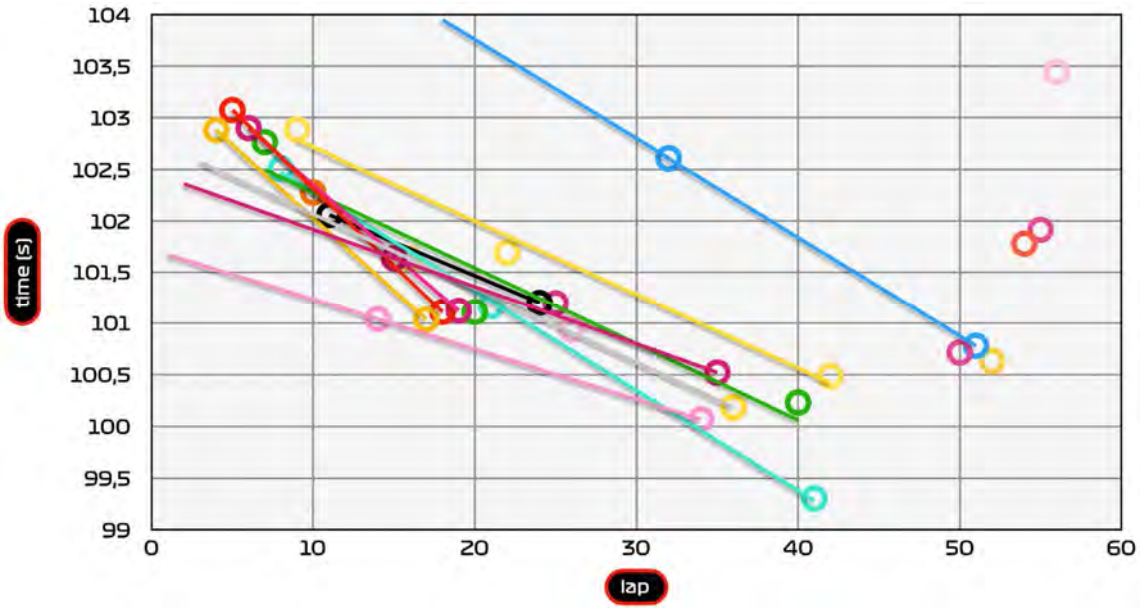


Graf. 7 - Linea di tendenza su tutti i giri puliti

[2] Con stint si intende i giri i percorsi tra due pitstop.

Graf. 8 - Linee di tendenza di ogni stint

Abbiamo quindi scelto di seguire una strada più impervia e calcolare le linee di tendenza tra i giri corrispondenti[3] di ogni stint di ogni pilota (Graf. 9): questo fa ottenere dei segmenti corti e singolarmente poco rilevanti, essendo ottenuti da serie di due o tre dati, ma in un numero molto alto (varie decine per ogni gara).



Graf. 9 - Linee di tendenza calcolate sui giri corrispondenti di ogni stint

Data la pendenza di ciascuna di queste rette, si può determinarne la media. Calcoliamo quindi le coordinate del baricentro del grafico a dispersione dei giri puliti, che è un punto che ha come coordinate la media dei tempi siglati nei giri puliti e la media dei giri nei quali questi tempi sono stati fatti segnare. Ricavati quindi la pendenza della retta e il baricentro dello scattergram, che è un punto attraverso cui la retta deve passare, si può ottenere una linea che indichi l'ideale progressione dei tempi data dalla gommatura della pista e dall'alleggerimento delle monoposto. Una volta ottenuto ciò, si possono esprimere i giri puliti di ciascun pilota come una

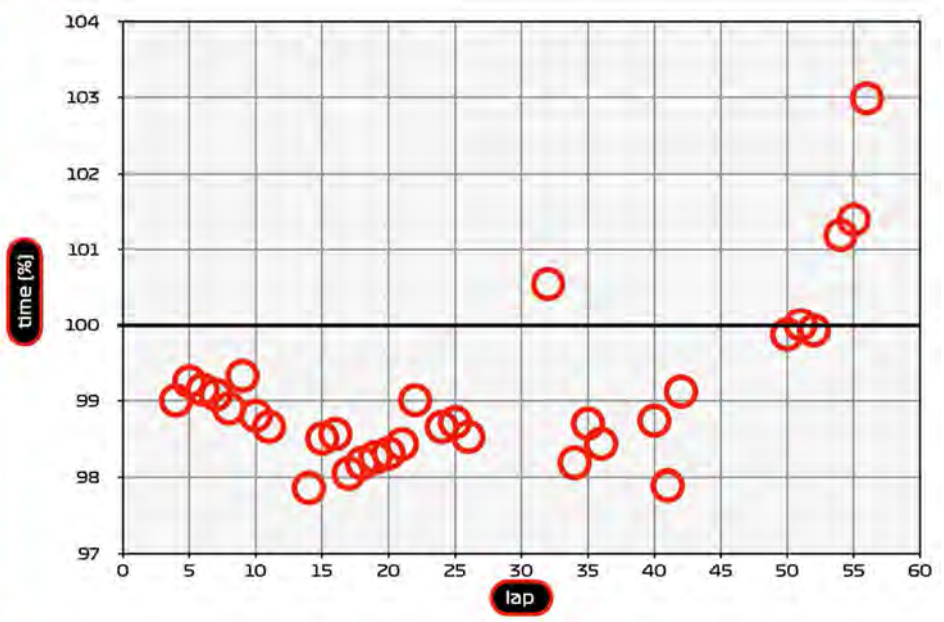
[3] Ad esempio: un pilota ha concluso tre stint. Dovremmo quindi tracciare la linea di tendenza tra il primo giro del primo, secondo e terzo stint. Successivamente la linea di tendenza tra il secondo giro del primo secondo e terzo stint e così via.

percentuale rispetto al tempo sul giro teorico che sta sulla retta al giro corrispondente (Tab. 10 e Graf. 11).

Driver: LEC

lap	time (%)
4	99,00163
5	99,25208
6	99,14632
7	99,07899
8	98,89773
9	99,32543
10	98,80307
11	98,65608
14	97,86045
15	98,50235
16	98,56126
17	98,04866
18	98,18212
19	98,25843
20	98,30956
21	98,42886
22	99,0059
24	98,65413
25	98,71446
26	98,52401
32	100,5472
34	98,18848
35	98,69861
36	98,43418
40	98,7374
41	97,88954
42	99,11894
50	99,86985
51	100,0015
52	99,92279
54	101,1929
55	101,3954
56	102,986

Tab. 10 - Giri puliti espressi come percentuale rispetto al giro di riferimento



Graf. 11 - Rappresentazione dei giri come percentuale

La media di questi giri poi va integrata con l'aggiunta del tempo richiesto dai pitstop[4], ottenendo così il passo gara di ciascuna accoppiata monoposto-pilota.

[4] Per una valutazione equa infatti non si può considerare semplicemente la media dei tempi percentuali, in quanto altrimenti un pilota che farà più pitstop potrà spingere di più nei suoi stint ottenendo statistiche vantaggiose.

3. RISULTATI E CONCLUSIONI

Procediamo ora ad esporre e commentare i risultati delle nostre elaborazioni.

3.1 Qualifica

Cominciamo con l'analisi delle 10 qualifiche (Tab. 12, 13 e 14 e Graf. 15).

	Ungheria		Olanda		Italia		Turchia		USA	
Pos	Driver	avg time [%]	Driver	avg time [%]	Driver	avg time [%]	Driver	avg time [%]	Driver	avg time [%]
1	Ham	98,11602	Ver	98,24655	Bot	98,31026	Ham	97,84692	Ver	97,91239
2	Ver	98,41654	Ham	98,30075	Ham	98,42889	Bot	98,00042	Ham	98,13264
3	Bot	98,52582	Bot	98,72719	Ver	98,81815	Ver	98,23421	Per	98,14845
4	Nor	99,37273	Lec	99,03383	Nor	98,84658	Lec	98,31568	Bot	98,50781
5	Gas	99,38444	Gas	99,09231	Ric	98,85399	Gas	98,3877	Lec	98,64586
6	Per	99,41957	Sai	99,17646	Gaz	99,18147	Alo	98,566	Sai	98,84188
7	Lec	99,51714	Gio	99,25205	Sai	99,43109	Per	98,83639	Ric	98,85874
8	Alo	99,57568	Ric	99,64426	Lec	99,4904	Nor	99,12922	Nor	98,93461
9	Sai	99,71618	Oco	99,72128	Per	99,61521	Tsu	99,24729	Gas	99,18543
10	Oco	99,72139	Alo	99,77405	Gio	99,75733	Str	99,54366	Tsu	100,0285
11	Vet	99,84758	Lat	99,96945	Vet	99,98841	Vet	100,1222	Vet	100,411
12	Ric	100,005	Rus	100,3103	Str	100,1206	Oco	100,1777	Oco	100,5122
13	Str	100,0336	Str	100,3602	Alo	100,1812	Rus	100,3726	Rus	100,9011
14	Rai	100,8922	Nor	100,4159	Oco	100,2232	Sai	100,5733	Alo	100,9116
15	Gio	100,9313	Tsu	100,4957	Rus	100,5803	Sch	100,6004	Gio	100,9517
16	Tsu	101,3684	Per	100,5927	Lati	101,239	Ric	101,4045	Str	101,1508
17	Rus	101,4009	Vet	100,8794	Tsu	101,2983	Lat	101,6466	Lat	101,1635
18	Lat	101,5206	Kub	101,6923	Sch	101,6381	Gio	102,0528	Rai	101,4965
19	Maz	102,6732	Sch	101,815	Kub	101,9866	Rai	103,3457	Sch	101,6946
20			Maz	102,511	Maz	102,2165	Maz	104,4367	Maz	102,0076

Tab. 12 - Risultati elaborazioni qualifiche 1

	Messico		Brasile		Qatar		Ar. Saudi.		Abu Dhabi	
Pos	Driver	avg time [%]	Driver	avg time [%]	Driver	avg time [%]	Driver	avg time [%]	Driver	avg time [%]
1	Bot	97,49608	Ham	98,22329	Ham	98,15826	Ham	98,26869	Ver	98,26459
2	Ham	97,6824	Ver	98,85658	Ver	98,71082	Bot	98,39334	Ham	98,70859
3	Ver	97,94581	Bot	98,93466	Bot	98,94885	Ver	98,42815	Nor	99,24833
4	Per	98,09615	Per	99,01707	Gas	99,14559	Per	98,75717	Per	99,26748
5	Gas	98,24264	Gas	99,44215	Alo	99,18202	Lec	98,87844	Sai	99,32133
6	Tsu	98,55745	Sai	99,513	Nor	99,2561	Gas	98,95817	Bot	99,37399
7	Lec	98,61784	Lec	99,56071	Sai	99,38847	Nor	99,01993	Lec	99,47691
8	Sai	98,63455	Nor	99,73566	Tsu	99,43826	Ric	99,06036	Tsu	99,59419
9	Ric	98,63712	Ric	99,82097	Oco	99,59735	Tsu	99,0671	Oco	99,79644
10	Nor	99,54944	Alo	99,92796	Vet	99,76008	Sai	99,08394	Ric	99,82038
11	Vet	99,5867	Oco	100,0378	Per	100,003	Oco	99,46237	Alo	99,88141
12	Rai	99,72034	Gio	100,2591	Str	100,1414	Gio	99,50953	Gas	99,91612
13	Gio	100,0943	Tsu	100,2706	Lec	100,1451	Rai	99,77903	Str	100,6007
14	Rus	100,1726	Vet	100,3415	Ric	100,3078	Alo	99,8509	Gio	100,6689
15	Oco	100,3885	Rai	100,4919	Rus	100,5009	Rus	99,85764	Vet	100,7969
16	Alo	100,8074	Str	100,7232	Rai	100,9867	Lat	100,1395	Lat	100,9322
17	Lat	101,198	Lat	101,0615	Lat	101,0559	Vet	100,1631	Rus	101,0339
18	Sch	101,3291	Rus	101,1425	Gio	101,1154	Str	100,354	Rai	101,4599
19	Maz	101,9009	Sch	101,6861	Sch	101,2915	Sch	100,4618	Sch	101,6119
20	Str	103,9183	Maz	102,0621	Maz	104,2692	Maz	101,5948	Maz	102,5442

Tab. 13 - Risultati elaborazioni qualifiche 2

In generale si può notare come tutti i piloti siano abbastanza ravvicinati e con risultati discretamente variabili a seconda del circuito; è evidente però il grande distacco del pilota della Haas Nikita Mazepin, avvicinato solo da Raikkonen in Turchia (evidentemente svantaggiato dal mo-

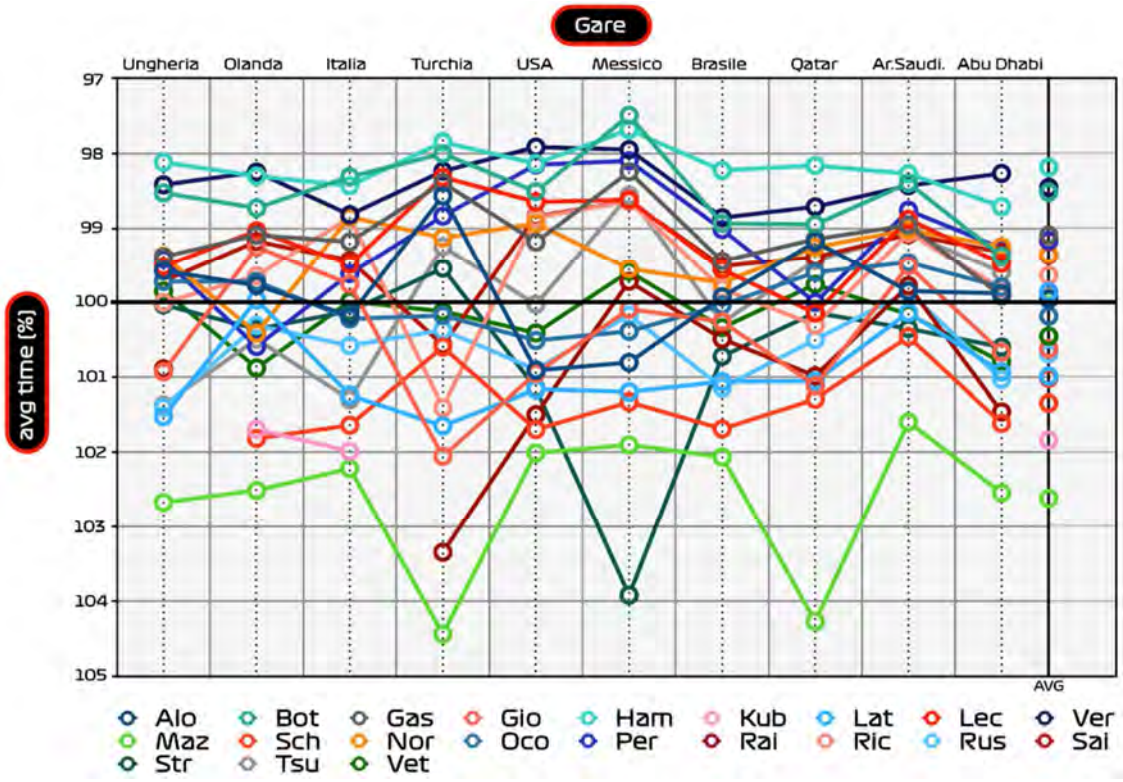
Media tot

Driver	avg time (%)
Ham	98,18664
Ver	98,45513
Bot	98,52184
Gas	99,0936
Lec	99,16819
Per	99,17532
Nor	99,35085
Sai	99,36802
Ric	99,64131
Alo	99,86582
Tsu	99,93659
Oco	99,96383
Vet	100,1897
Gio	100,4592
Rus	100,6273
Str	100,6947
Lat	100,9926
Rai	101,0215
Sch	101,3476
Kub	101,8395
Maz	102,6216

Tab. 14 - Medie risultati elaborazioni qualifiche

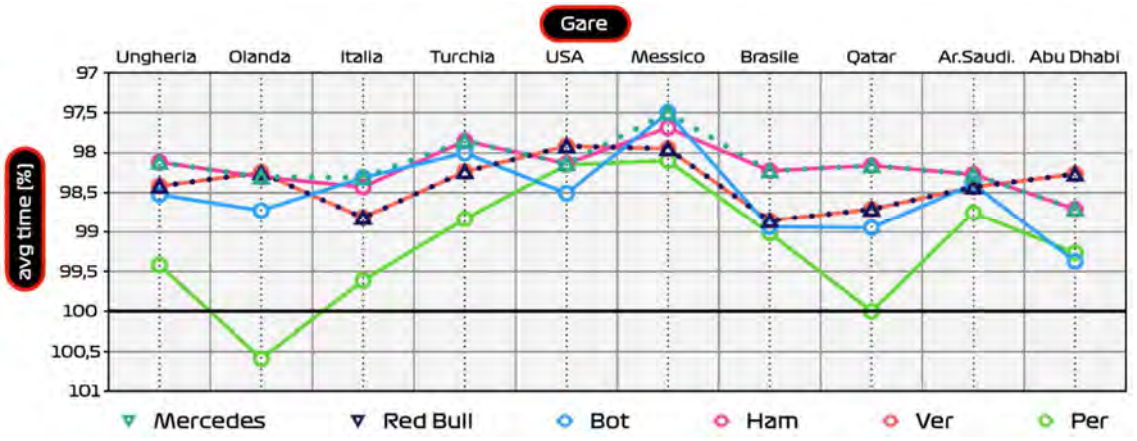
mento che era al suo secondo Gran Premio dopo essere guarito dal covid ed essere stato sostituito da Kubica nelle tappe in Olanda e Italia) e da Stroll in Messico a causa di un incidente del canadese. Ciò non può che alimentare l'ormai comune convinzione che la Haas sia in ritardo rispetto alle altre scuderie e che Mazepin non sia così talentuoso come i suoi avversari. Questo giudizio è avallato anche dalle prestazioni del suo compagno di squadra Mick Schumacher, che riesce tutto sommato a limitare i danni e a rimanere insieme agli altri piloti di coda del gruppo principale. Il driver russo è stato accusato di essere entrato in Formula 1 solo grazie alla sponsorizzazione da parte di suo padre (il miliardario Dmitry Mazepin), e questi risultati non possono che dar credito a questa ipotesi.

Un altro sottogruppo leggermente staccato dal resto, anche se in modo meno evidente, è quello formato dai primi tre piloti, Hamilton, Verstappen e Bottas, che riescono quasi sempre a mantenere un certo gap dal resto delle vetture e ciò si riflette particolarmente nelle medie finali, da cui risultano con un sensibile vantaggio rispetto ai distacchi minimi presenti in uno sport competitivo come la Formula 1.



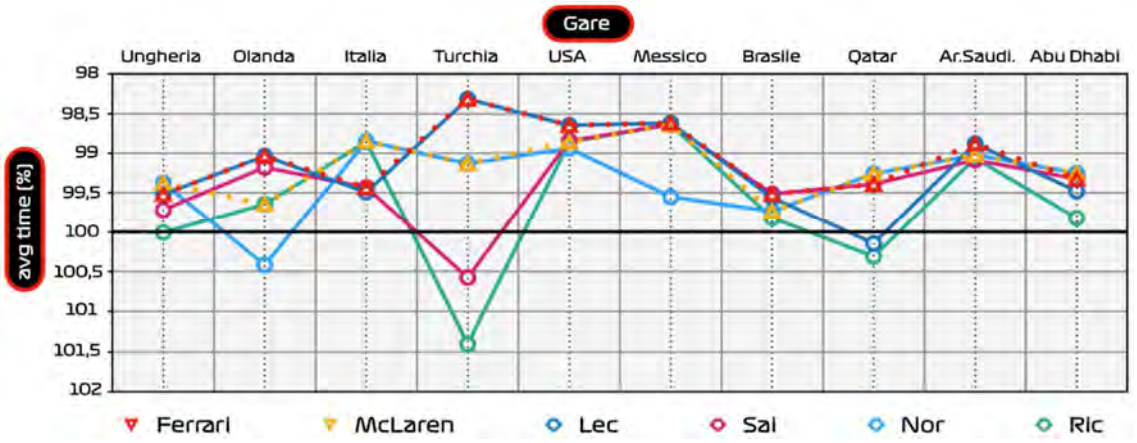
Graf. 15 - Rappresentazione risultati elaborazioni qualifiche

Sempre riguardo alle qualifiche, passiamo ora a concentrarci più specificamente sulle lotte che hanno animato il campionato 2021 e che abbiamo già citato nell'introduzione: quella tra Mercedes e Red Bull per i titoli e quella tra Ferrari e McLaren per il terzo posto.



Graf. 16 - Rappresentazione risultati elaborazioni qualifiche Mercedes e Red Bull

Per quanto riguarda Lo scontro tra Mercedes e Red Bull (Graf. 16), si può notare che tre dei quatto piloti coinvolti, già citati precedentemente, ottengono quasi sempre prestazioni molto ravvicinate, mentre Perez, il secondo pilota Red Bull, fatica molto a mettere insieme un giro ottimale e si avvicina alle performance del compagno di squadra solo nei Gran Premi centrali (USA, Messico, Brasile). È inoltre interessante notare il sensibile distacco che Hamilton infligge a tutti gli altri atleti in occasione delle qualifiche in Brasile e Qatar, forse grazie all'introduzione del "super motore"[5] Mercedes nel primo di questi 2 appuntamenti, ma di come i valori in capo tornino poi simili agli appuntamenti precedenti.



Graf. 17 - Rappresentazione risultati elaborazioni qualifiche Ferrari e McLaren

A proposito alla lotta tra McLaren e Ferrari invece (Graf. 17), si può notare come in quasi tutti i Gran Premi almeno 3 dei 4 piloti abbiano avuto prestazioni abbastanza simili, tranne nel caso della Turchia, in cui Leclerc e Norris hanno sonoramente battuto Sainz e Ricciardo, e del Qatar, in cui la situazione si è invertita. Si può inoltre notare come le prestazioni tra la migliore delle

[5] Durante la stagione 2021 il team Mercedes è stato afflitto da un'anomalia al motore termico delle sue monoposto, non risolvibile a causa del congelamento regolamentare dei propulsori a partire dall'inizio della stagione, che causa una sensibile perdita di potenza nel giro già di poche gare. La scuderia decide quindi di usare più motori del limite regolamentare, preferendo scontare delle penalità ed avere una power unit più potente. L'ultimo dei 5 motori usati da Hamilton (su un limite da regolamento di 3) viene smarcato in Brasile ed è ottimizzato per soli 4 Gran Premi, quindi con una potenza particolarmente elevata.

due Ferrari e la migliore delle due McLaren[6] siano solitamente simili, tranne in parte in Olanda e Italia ma soprattutto in Turchia.

3.2 Gara

Ci concentreremo ora invece sui risultati ottenuti dalle elaborazioni delle gare (Tab. 18, 19 e 20 e Graf. 21). Sfortunatamente il Gran Premio d'Arabia Saudita non ha fornito dati utili[7], quindi non sarà incluso in questo paragrafo.

	Olanda		Italia		USA		Messico		Brasile	
Pos	Driver	avg time (%)	Driver	avg time (%)	Driver	avg time (%)	Driver	avg time (%)	Driver	avg time (%)
1	Ver	98,79157	Ham	99,07828	Ver	98,99956	Ver	98,36704	Bot	99,53996
2	Per	99,52856	Per	99,26367	Bot	99,50258	Per	98,55602	Ver	99,94302
3	Bot	99,77499	Bot	99,51157	Lec	99,82196	Ham	98,88525	Per	100,0753
4	Ham	99,94098	Nor	99,61894	Per	99,91768	Gas	99,82146	Sai	100,1233
5	Gas	100,3003	Ric	99,80254	Nor	100,0863	Sai	99,87816	Ham	100,1439
6	Lec	100,3348	Sai	100,1684	Ham	100,1045	Lec	100,2194	Lec	100,622
7	Vet	100,3867	Lec	100,4694	Ric	100,3755	Nor	100,2875	Alo	100,6781
8	Nor	100,4309	Str	100,4746	Rai	100,5771	Vet	100,3286	Rai	101,0406
9	Sai	100,5771	Gio	100,5479	Vet	101,1073	Alo	100,3465	Oco	101,0538
10	Oco	100,7133	Lat	100,5973	Str	101,2302	Rai	100,4589	Gas	101,6791
11	Alo	100,8383	Kub	100,8082	Tsu	101,3067	Ric	100,9377	Vet	101,8235
12	Ric	101,1527	Vet	100,9443	Sai	101,411	Str	101,0316	Rus	101,9756
13	Gio	101,401	Oco	101,1395	Rus	101,7892	Oco	101,2408	Gio	102,1613
14	Kub	101,5739	Ver	101,3024	Lat	101,926	Bot	101,2411	Lat	102,2054
15	Sch	102,621	Sch	102,8164	Gio	102,014	Gio	101,2532	Maz	102,5634
16	Lat	103,3973	Maz	103,7517	Sch	102,2402	Rus	101,5107	Nor	102,7546
17					Maz	103,2509	Lat	102,0069	Sch	102,8583
18							Maz	103,4337	Tsu	103,3458

Tab. 18 - Risultati elaborazioni gare 1

	Qatar		Abu Dhabi	
Pos	Driver	avg time (%)	Driver	avg time
1	Ham	99,60069	Ham	99,07189
2	Ver	100,1227	Ver	100,1204
3	Sai	100,6346	Bot	100,3494
4	Lec	100,6758	Sai	100,5002
5	Vet	100,7242	Nor	100,6695
6	Nor	100,8166	Vet	100,8938
7	Oco	100,8799	Oco	100,9357
8	Per	100,9349	Lec	100,9382
9	Alo	100,9386	Tsu	100,9819
10	Ric	101,2957	Alo	101,1584
11	Gas	101,468	Str	101,8693
12	Tsu	101,6356	Lat	102,0748
13	Rai	101,7278	Ric	102,5463
14	Sch	101,9866	Sch	102,7031
15	Rus	102,0089		
16	Lat	102,5655		
17	Gio	102,5727		
18	Maz	103,5136		

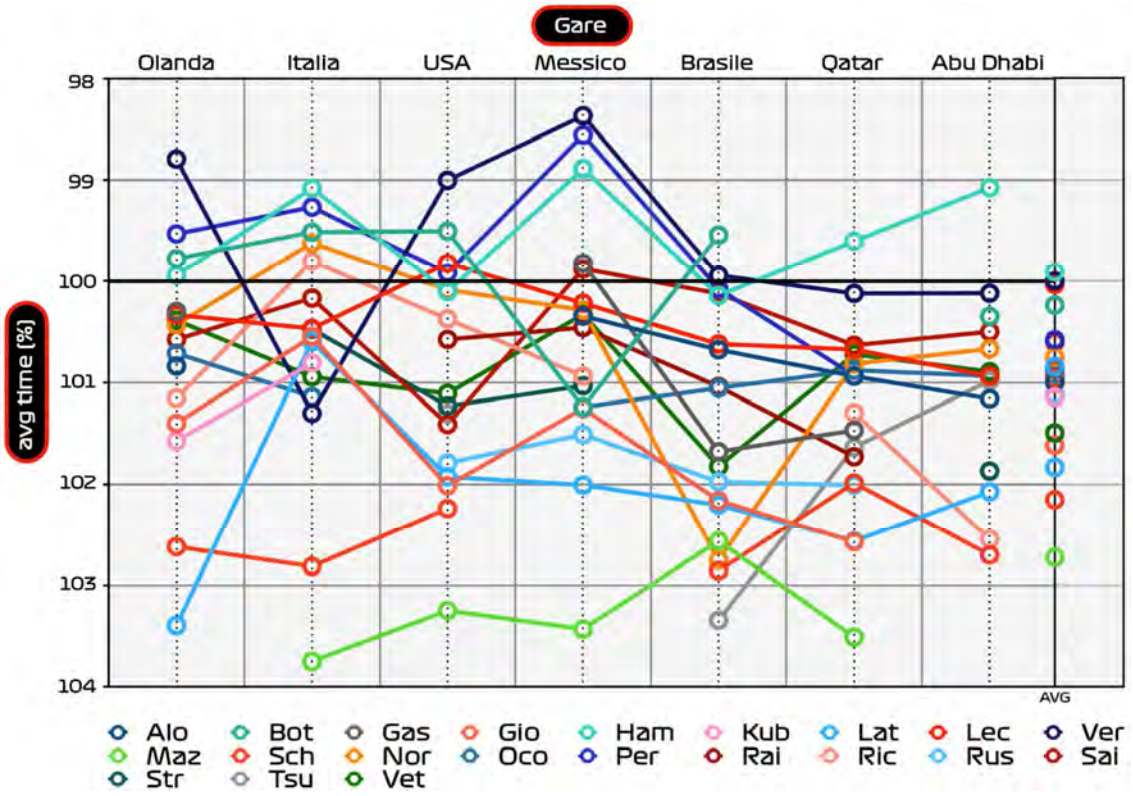
Tab. 19 - Risultati elaborazioni gare 2

Media tot	
Driver	avg time (%)
Ham	99,5465
Ver	99,66383
Per	99,71269
Bot	99,9866
Lec	100,4402
Sai	100,4704
Nor	100,6663
Alo	100,792
Gas	100,8172
Vet	100,8869
Rai	100,9511
Oco	100,9938
Ric	101,0184
Str	101,1514
Kub	101,191
Gio	101,6583
Tsu	101,8175
Rus	101,8211
Lat	102,1105
Sch	102,5376
Maz	103,3027

Tab. 20 - Medie risultati elaborazioni gare

[6] Di fatto, a parità di macchina, più un pilota va forte e più sta guidando al limite della monoposto, quindi sta sfruttando meglio il potenziale della vettura. Per questo, se si vuole valutare il veicolo, è più rilevante prendere la migliore prestazione tra quelle dei piloti che lo guidano piuttosto che la loro media.

[7] In gara si sono verificate due bandiere rosse ravvicinate, ai giri 13 e 15, che hanno fatto sì che quasi tutti i piloti effettuassero il loro primo e unico pitstop durante queste due sospensioni forzate. Si è creata così per quasi tutti una gara a due stint con il primo dei quali non sufficientemente lungo da ottenere un distanziamento delle monoposto tale da far segnare dei giri puliti. Come risultato, si veniva a creare un'unica linea di tendenza, peraltro anomala, quindi non rilevante.

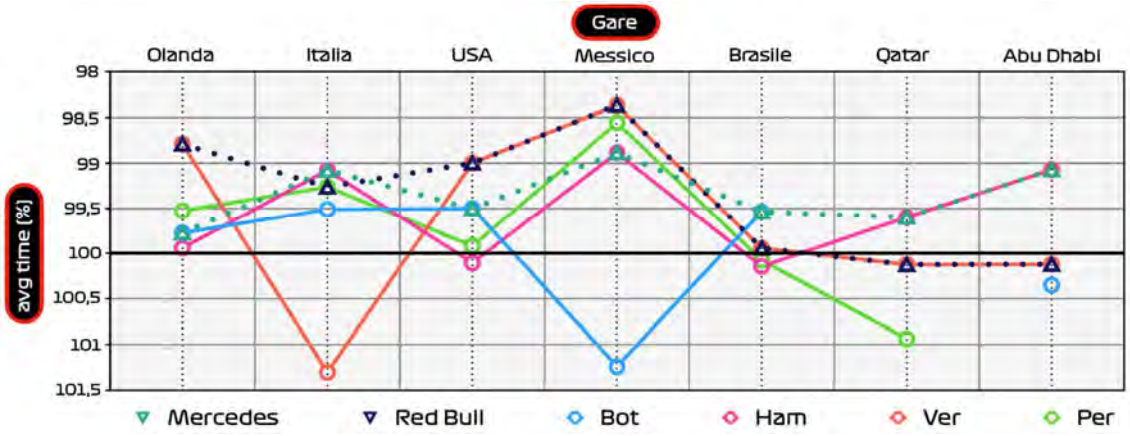


Graf. 21 - Rappresentazione risultati elaborazioni gare

Ad un'analisi generale dei dati ottenuti balzano subito all'occhio alcune peculiarità, in parte comuni con i risultati delle qualifiche e in parte no. Per prima cosa il range dei dati è più ristretto rispetto alle qualifiche, si va infatti da 98,37% a 103,8%, nelle gare, mentre nelle qualifiche si va da 97,50% a 104,4%. Ciò indica che in gara le prestazioni dei piloti sono ancora più vicine e la competitività è ancora più alta. Si può inoltre notare che in generale i piloti siano inizialmente riuniti in un unico gruppo compatto (a meno di qualche eccezione che sottolineeremo dopo) ma che via via che la stagione procede questi si distanzino sempre di più, tanto che al Gran Premio del Messico si può individuare un gruppo di testa (Verstappen, Perez, Hamilton) e a quello di Abu Dhabi un gruppo di coda (Schumacher, Ricciardo, Latifi, Stroll). Peraltro col procedere della stagione si rileva una sempre maggiore tendenza dei tempi di andare oltre il 100%. Inoltre è evidente anche qui la lentezza di Mazepin, a conferma di quanto già scritto precedentemente, che riesce a stare insieme al resto dei piloti più attardati solo in Brasile. Presentano lacune molto evidenti, anche se non ai livelli del pilota russo, anche Latifi nel GP d'Olanda e Schumacher in Olanda, Italia e USA. Dall'altra parte del grafico invece oltre al già citato terzetto che domina la gara in Messico, si possono notare i 2 acuti di Verstappen in Olanda e Stati Uniti, il discreto distacco che Hamilton infligge a Verstappen in Qatar con quest'ultimo che a sua volta stacca tutto il resto dei piloti, ma soprattutto lo stradominio di Hamilton nel finale di stagione, dove supera di più dell'1% il suo concorrente più prossimo. Per quanto riguarda infine le medie generali, mentre nelle qualifiche oltre all'evidente distacco dei primi 3 piloti e di Mazepin dietro la gran parte dei concorrenti presenta distacchi più o meno omogenei, si delineano delle suddivisioni abbastanza precise:

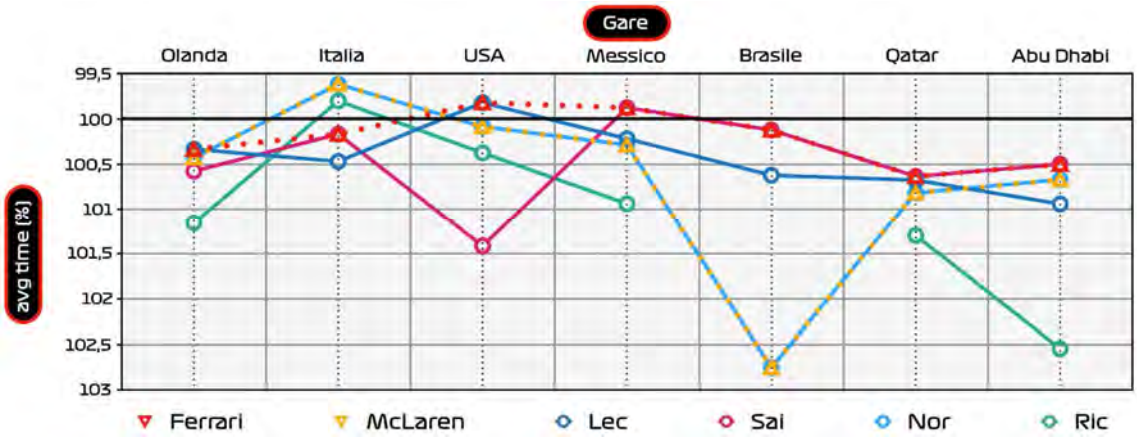
1. *il gruppo dei "top team"* formato dai quattro piloti di Mercedes e Red Bull;

2. *il raggruppamento dei "midfield team"* formato dagli alfieri Ferrari, McLaren, Alpine, Aston Martin, Alfa Romeo (tre in questo caso) e da Gasly (Alpha Tauri);
3. *l'insieme dei "backmarkers"* costituito da Tsunoda (Alpha Tauri), dalle due Williams e dalle due Haas, con Mazepin sensibilmente staccato.



Graf. 22 - Rappresentazione risultati elaborazioni gare Mercedes e Red Bull

Passando ora all'osservazione dei dati riguardanti la rivalità tra Mercedes e Red Bull (Graf. 22), si può notare come Perez non sia più l'anello debole del quartetto ma anzi dei quattro sia Bottas ora il più arretrato, come sottolineato dalle medie generali. Oltre a quanto già detto precedentemente sugli spunti in positivo osservabili, si possono notare le prestazioni negative di Verstappen al Gran Premio d'Italia, di Bottas in Messico e in parte di Perez in Qatar.



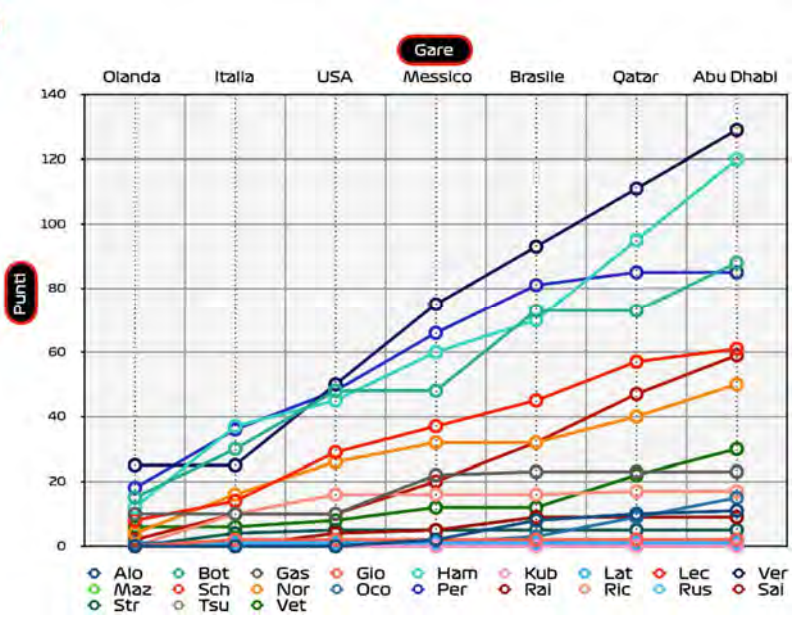
Graf. 23 - Rappresentazione risultati elaborazioni gare Ferrari e Mclaren

Procedendo poi con la lotta tra Ferrari e McLaren (Graf. 23), si vede che anche qui non ci sono piloti che mostrano grandissime lacune e il gruppo è compatto. Gli unici punti particolari sono le prestazioni particolarmente negative di Sainz in USA, Norris in Brasile e Ricciardo ad Abu Dhabi.

Punteggio gare teorico

Pos	Driver	Punti tot
1	Ver	129
2	Ham	120
3	Bot	88
4	Per	85
5	Lec	61
6	Sai	59
7	Nor	50
8	Vet	30
9	Gas	23
10	Ric	17
11	Oco	15
12	Alo	11
13	Rai	9
14	Str	5
15	Gio	2
16	Tsu	2
17	Lat	1
18	Kub	0
19	Rus	0
20	Sch	0
21	Maz	0

Tab. 24 - Punteggi totali in base al passo gara

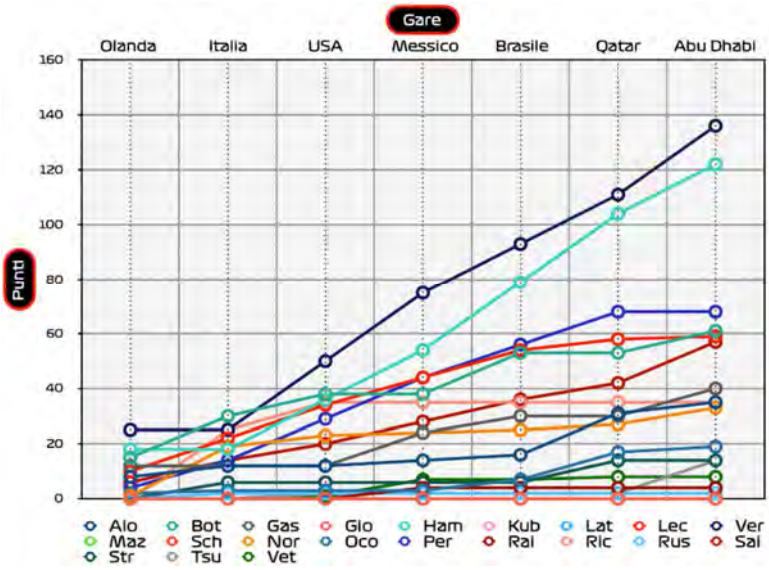


Graf. 25 - Progressione punteggi in base al passo gara

Punteggio gare reale

Pos	Driver	Punti tot
1	Ver	136
2	Ham	122
3	Per	68
4	Bot	61
5	Lec	59
6	Sai	57
7	Gas	40
8	Alo	35
9	Ric	35
10	Nor	33
11	Oco	19
12	Str	14
13	Tsu	14
14	Vet	8
15	Rai	4
16	Rus	2
17	Gio	0
18	Kub	0
19	Lat	0
20	Sch	0
21	Maz	0

Tab. 26 - Punteggi totali in base all'ordine di arrivo reale



Graf. 27 - Progressione punteggi in base all'ordine di arrivo reale

Infine, per curiosità, pur non contribuendo questo a valutare l'effettiva velocità delle accoppiate monoposto-pilota, abbiamo calcolato i punteggi ottenuti in queste gare se l'ordine di arrivo finale fosse determinato solo dal passo gara (Tab. 24 e Graf. 25) e l'abbiamo confrontato con i risultati delle elaborazioni delle gare (Tab. 20) e con i punteggi ottenuti con gli ordini d'arrivo reali (Tab. 26 e Graf. 27). Sia per i punteggi basati sull'ordine di arrivo teorico che per i punteggi calcolati sull'ordine di arrivo reale, abbiamo usato il sistema di punteggio 2010-2018[8], ovve-

[8] Punti per i primi 10, in ordine 25 per il 1° classificato, 18 per il 2°, 15 per il 3°, 12 per il 4°, 10 per il 5°, 8 per il 6°, 6 per il 7°, 4 per l'8°, 2 per il 9° e 1 per il 10°.

ro quello dell'anno scorso ma senza il punto aggiuntivo per il giro veloce e senza tenere conto delle gare sprint. Concentrandoci sulle prime quattro posizioni (quelle successive non presentano per la maggior parte grandi differenze di punteggi e inoltre più ci si avvicina al fondo della classifica e più i punti si ottengono per casi fortuiti e singoli eventi), un primo elementi di rilievo è il maggiore distacco dei primi due piloti da tutti gli altri e la maggiore vicinanza dei loro scudieri al resto dei piloti nei punteggi "reali" rispetto a quelli teorici. È interessante inoltre notare come sia la classifica "reale" che quella teorica dia in vantaggio Verstappen su Hamilton nonostante quest'ultimo risulti più veloce sia sul passo gara che sul giro secco secondo le nostre statistiche. Ciò dimostra che non basta avere una media finale delle prestazioni migliore, ma anche una distribuzione adeguata delle performance; a causa della grande differenza di punti assegnati tra le posizioni di testa (ad esempio 25 per il primo posto e 18 per il secondo, ben 7 punti di differenza), fare molti secondo posti e pochi primi è meno conveniente che fare molti primi posti e qualche risultato molto negativo (o persino qualche ritiro). Inoltre vincere con un vantaggio molto grande sul secondo (ad esempio l'1% di Hamilton su Verstappen ad Abu Dhabi) dà tanti punti quanti una vittoria di misura.

APPENDICE : le formule

L'analisi dei dati relativi alle gare di formula 1 nella sua complessità ha richiesto l'utilizzo di alcuni strumenti, tra cui Microsoft Excel e lo sviluppo di alcune formule, dalle più elementari alle più complesse.

A1. Analisi delle qualifiche

La fase di qualificazione è suddivisa in tre sessioni denominate Q_1, Q_2, e Q_3, durante ognuna delle quali i piloti si battono per tentare di far segnare il giro più veloce entro un limite di tempo ben definito; scaduto il termine, i cinque piloti più lenti vengono eliminati. Nelle fasi di Q_1 e Q_2 vengono presi in considerazione solo i tempi dei piloti eliminati, mentre i tempi dei piloti salvi vengono azzerati ai fini della classifica finale. Si può dedurre allora come i piloti si diano testa solo nell'ultima fase per partire in prima posizione. Da tali considerazioni, il tempo segnato in Q_3 può non essere il giro più veloce percorso durante le tre sessioni di qualifica.
Per poter analizzare i tempi segnati dai piloti, però, è necessario che essi siano disponibili in formato decimale usando come unità di misura i secondi [s] diversamente da come forniti dalla nostra fonte, l'*API Ergast*, in formato m:ss.sss ovvero minuti, secondi e millesimi di secondo. Per effettuare le conversioni abbiamo usato la formula (1) qui di seguito:

```
=60*STRINGA.ESTRAI(REF1;1;1)+SOSTITUISCI(STRINGA.ESTRAI(REF1;
                        3; 6); "."; ",")
```
(1)

nella quale abbiamo inserito in sostituzione di `REF1`, le coordinate della cella contenente il valore da convertire. In questo modo si moltiplica per 60 il numero di minuti nel tempo sul giro, ovvero la prima cifra del valore estrapolata attraverso il comando `STRINGA.ESTRAI`; si somma poi a quanto ottenuto la parte già in formato decimale, ovvero le 6 cifre a partire dal terzo grafema, ottenuta analogamente utilizzando il comando `STRINGA.ESTRAI`, riscritto però sostituendo al punto [.] una virgola [,][9] attraverso il comando `SOSTITUISCI`.
Per un'analisi delle prestazioni pilota-vettura in qualifica si è utilizzato il tempo del giro più veloce tra le tre sessioni estrapolato attraverso la formula (2):

```
=MIN(REF1)
```
(2)

[9] Nel formato decimale statunitense il separatore decimale è il punto [.], mentre in Italia e in gran parte d'Europa è la virgola [,]

in cui abbiamo in sostituzione di REF_1 le coordinate delle celle contenenti i tempi di un singolo pilota. Abbiamo poi rapportato il dato ottenuto come percentuale all'*average lap*, la media di tutti i tempi delle tre sessioni di qualifica. A seguire la formula (3) utilizzata:

$$=REF_1/MEDIA(REF_2)*100 \quad (3)$$

con le coordinate delle celle contenenti il tempo più veloce di ciascun pilota in sostituzione a REF_1 e quelle delle celle con tutti i tempi fatti segnare in fase di qualifica al posto di REF_2.
In sostanza la (3) è la "traduzione" della formula analitica:

$$Fastest\ lap\ [\%] = \frac{Fastest\ lap\ [s]}{Average\ lap\ [s]} \cdot 100 = \frac{Fastest\ lap\ [s]}{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i} \cdot 100$$

A1. Analisi delle gare

A2.1 Preparazione dei giri

La nostra fonte dei dati fornisce per ogni tornata della gara il numero del giro e per ciascun pilota la sua posizione e il suo tempo sul giro. Manca quindi un dato molto importante per selezionare i giri puliti, e cioè il distacco dalle monoposto circostanti, valore che dobbiamo ricavare noi.
Per prima cosa vanno convertiti i valori dei tempi sul giro dal formato in cui vengono forniti (m:ss:sss) nel formato decimale attraverso la formula (1) già spiegata precedentemente.
Successivamente, per definire il tempo totale abbiamo utilizzato la funzione Excel SOMMA.SE, che somma il contenuto delle celle specificate secondo una condizione o criterio assegnato. Nel nostro caso, essendo

```
=SOMMA.SE(intervallo;criterio;[int_somma])
```

avremo:

$$=SOMMA.SE(REF_1:REF_2;REF_2;REF_3) \quad (4)$$

nella quale abbiamo inserito, in sostituzione di REF_1, la cella da cui inizia la colonna dei nomi di piloti, di REF_2 la cella in cui è riportato il nome del pilota al giro desiderato e di REF_3 l'intervallo in cui sono riportati i tempi sul giro in formato decimale. In questo modo la formula ispeziona tutta la colonna dei nomi dei piloti dall'inizio alla riga desiderata, trova tutte le volte che il nome del pilota desiderato si ripete e somma tutti i tempi sul giro contenuti nelle celle corrispondenti.
Di seguito un esempio di applicazione della formula per il pilota Max Verstappen:

```
=SOMMA.SE($B$2:B2;B2;$E$2:E2)
```

Ora dunque possiamo calcolare la distanza della monoposto con la vettura che precede (5) e quella che segue (6) rispettivamente con le formule:

$$=\text{MIN}(\text{SE}(REF_1-REF_2>0;REF_1-REF_2)) \quad (5)$$

$$=-\text{MAX}(\text{SE}(REF_1-REF_2<0;REF_1-REF_2)) \quad (6)$$

dove con REF_1 indichiamo le coordinate della cella relativa al tempo totale del pilota allo specifico giro, mentre con REF_2 l'intervallo contenente tutti i valori dei tempi totali.
La formula (5), quindi, seleziona attraverso il comando SE tutte le differenze di tempo positive rispetto al riferimento e poi da queste sceglie la minore, come si nota dall'utilizzo del comando MIN, così da ottenere il distacco dalla vettura immediatamente precedente. Allo stesso modo, la (6) individua, sempre con il comando SE, tutte le differenze di tempo negative rispetto al riferimento e con il comando MAX viene estratto il valore massimo (ovvero quello minimo in valore assoluto essendo tutti numeri minori di zero) ricavandone il modulo cambiando di segno grazie al − posto all'inizio della formula. Pertanto non vengono considerati solo i distacchi dalle vetture appena davanti e dietro in classifica, ma anche i doppiati e quindi la distanza dalle monoposto fisicamente più vicine, ovvero quelle che influiscono realmente sul tempo.
Di seguito due esempi, rispettivamente del gap_{front} e gap_{back}:

```
=MIN(SE(F5-$F$2:$F$1361>0;F5-$F$2:$F$1361))
```

```
=-MAX(SE(F5-$F$2:$F$1361<0;F5-$F$2:$F$1361))
```

A2.2 Selezione dei giri regolari

Come già spiegato nel paragrafo 2.3, una volta eliminati manualmente le tornate affette da *safety car*, *VSC*, partenze, ripartenze e pitstop, per ottenere i *giri regolari* è necessario selezionare solo quei giri *condotti ad una distanza superiore a 4s dalla vettura che segue e a 1s da quella che precede*. A tale scopo abbiamo sviluppato e impiegato le formule (7) e (8) che riportiamo qui di seguito:

$$=\text{SE}(gap_{front}<\text{gap min}_{front};"";\text{SE}(gap_{back}<\text{gap min}_{back};"";\text{lap})) \quad (7)$$

$$=\text{SE}(gap_{front}<\text{gap min}_{front};"";\text{SE}(gap_{back}<\text{gap min}_{back};"";\text{time})) \quad (8)$$

La funzione SE consente di eseguire confronti logici tra un valore e un risultato previsto, dando in output uno di due possibili risultati: il primo risultato si ottiene se il confronto è *vero*, il secondo se è *falso*. In tal caso, le due formule prodotte consentono di verificare se in un determinato giro le condizioni scelte affinché si possa determinare un giro regolare si realizzano o meno. Nel caso in cui queste si verifichino nelle celle corrispondenti verranno inseriti rispettivamente: nella (7) il numero del giro pulito (*lap*) e nella (8) il tempo segnato su tale giro (*time*). In caso contrario entrambe le caselle rimarranno vuote, come si può notare a colpo d'occhio dalla presenza di "". Riportiamo alcuni esempi di applicazione:

```
=SE(D3<$AN$1;"";SE(E3<$AN$2;"";B3))
```

```
=SE(D3<$AN$1;"";  SE(E3<$AN$2;"";C3))
```

A2.3 Prodotto tra il numero del giro [clean lap n°] e il tempo sul tiro [clean lap time]

Questa operazione non ha un'utilità in sé, ma servirà successivamente per calcolare il valore a, ovvero la pendenza della retta dei minimi quadrati.A tal fine abbiamo utilizzato la formula (9):

$$=\text{SE.ERRORE}(REF_1*REF_2;"") \tag{9}$$

sostituendo a REF_1 le coordinate della cella contenente il numero del giro di riferimento e a REF_2 quelle del tempo sul giro, come nell'esempio riportato qui sotto:

```
=SE.ERRORE(F3*G3;"")
```

La funzione `SE.ERRORE` restituisce un valore specificato dall'utente se la formula restituisce un errore. In caso contrario, restituisce il risultato della formula. Dunque, se i due fattori della moltiplicazione rispettano le condizioni imposte, ovvero se il giro di riferimento è un giro regolare, all'interno della cella si otterrà il risultato della moltiplicazione, altrimenti questa rimarrà vuota.

A2.4 Pendenza della retta dei minimi quadrati dei giri corrispondenti di ogni stint [a]

Una retta dei minimi quadrati ha solitamente una forma del tipo$y = ax + b$, dove$a$sta ad indicareil coefficiente angolare della retta, mentre b la sua intercetta con l'asse y. Note le coordinate dei punti dello *scattergram*, si può calcolare la pendenza della sua linea di tendenza con la formula (10):

$$a = \frac{cov(X,Y)}{var(X)} \tag{10}$$

dove$cov(X,Y)$ èla *covarianza* delle serie statistiche X e Y e$var(X)$ è la *varianza* della serie X. La covarianza, a sua volta,la si può esprimere come:

$$cov(X,Y) = M(XY) - M(X)M(Y)$$

ove M è la media. Dato che la serie Xriguarda l'insieme dei numeri dei giri regolari(*clean lap n°*) e la serie Y l'insieme dei tempi dei giri regolari (*clean lap time*), per il calcolo di a abbiamo definito la seguente formula (11):

```
=SE.ERRORE((MEDIA(REF1)-
(MEDIA(REF2)*MEDIA(REF3)))/VAR.P(REF2);"")
```
(11)

inserendo in sostituzione di REF_1 le coordinate delle celle contenenti i prodotti tra i numeri dei giri regolari e i tempi sul giro, di REF_2 le celle dei numeri dei giri e di REF_3 le celle dei tempi sul giro. Per il calcolo della varianza abbiamo usato la funzione VAR.P la quale svolge il calcolo basandosi sull'intera popolazione e non su stime campionarie. Si può notare come ancora una volta si è utilizzata la funzione SE.ERRORE , che restituisce un valore specificato dall'utente (in questo caso non avremo alcun valore di output "") se la formula restituisce un errore; in caso contrario, restituisce il risultato della formula. A seguire un esempio:

```
=SE.ERRORE((MEDIA(H2;O2;V2;AC2;AJ2)-(MEDIA(F2;M2;T2;AA2;AH2)
    *MEDIA(G2;N2;U2;AB2;AI2)))/VAR.P(F2;M2;T2;AA2;AH2);"")
```

A2.5 Pendenza delle rette dei minimi quadrati [a_{tot}]

$$=\text{MEDIA}(a_1:a_n) \tag{12}$$

La formula (12) è stata utilizzata con l'obiettivo di calcolare la media *di tutte le pendenze delle rette dei minimi quadrati dei giri corrispondenti di ogni stint*, ovvero la media di tutti i valori di a calcolati precedentemente, che ci fornisce una panoramica sulla progressione dei tempi in funzione del consumo di carburante e della gommatura della pista.
Di seguito riportiamo la formula in azione:

```
=MEDIA($AK$2:$AK$544)
```

A2.6 Baricentro del grafico a dispersione dei giri regolari [Mx, My]

Il baricentro dello scattergram è un punto attraverso il quale deve passare la *linea di tendenza* di un grafico a dispersione. Il baricentro del grafico a dispersione dei giri regolari, quindi, ci è utile per definire la *retta generale dei minimi quadrati*, e avrà come coordinate X e Y rispettivamente la media delle ascisse (13) e delle ordinate (14) dei punti che compongono il grafico, ossia la media dei valori dei numeri dei giri regolari (*clean lap n°*) e dei tempi sul giro (*clean lap time*):

```
=MEDIA(clean lap n°)
```
(13)

```
=MEDIA(clean lap time)
```
(14)

Ancora una volta riportiamo le formule utilizzate:

```
=MEDIA($F$2:$F$544;$M$2:$M$544;$T$2:$T$544;$AA$2:$AA$544;$AH$2:$AH$
544)
```

```
=MEDIA($G$2:$G$577;$N$2:$N$577;$U$2:$U$577;$AB$2:$AB$577;$AI$2:$AI$
577)
```

A2.7 Intercetta della retta dei minimi quadrati dei giri corrispondenti di ogni stint [b_{tot}]

Sapendo che$y = ax + b$, la formula che useremo per ottenere il valore del termine noto b è la seguente:

$$b = y - ax$$

Conoscendo a_{tot} e sapendo che il baricentro dello scattergram sta sulla linea di tendenza, possiamo calcolare b_{tot} con la formula:

$$=M_y-(a_{tot}*M_x) \tag{15}$$

Sostituendo alla formula analitica le coordinate delle celle si ottiene:

```
=$AN$5-($AN$3*$AN$4)
```

A2.8 Tempo percentuale di ogni giro rispetto al tempo sul giro ideale [%stint$_n$]

Dopo aver definito la *retta generale dei minimi quadrati* possiamo esprimere tutti i tempi sui giri reali R come un rapporto percentuale $R\%$ rispetto al tempo T che sta al giro corrispondente λ sulla retta di pendenza a e intercetta b attraverso la relazione:

$$R\% = \frac{100R}{T} = \frac{100R}{a\lambda + b}$$

Per calcolare, quindi, quello che vogliamo ottenere, abbiamo sviluppato la formula (16):

```
=SE(SE.ERRORE(REF1/(REF2*REF3+REF4)*100;"")=0;"";SE.ERRORE(REF1/(REF2*REF3+REF4)*100;""))
```
(16)

in cui abbiamo inserito, in sostituzione di `REF1`, il tempo sul giro di riferimento, di `REF2` il numero del giro di riferimento, di `REF3` il valore di a e di `REF4` quello di b. Anche questa volta i comandi `SE` e `SE.ERRORE` sono stati usati per lasciare la cella vuota in caso di errori o mancanza di dati.

A2.9 Media dei tempi percentuali [%M]

La formula (17) che segue ci permette di calcolare la media di tutti i tempi percentuali dei giri rispetto al tempo sul giro ideale ottenendo, così, il *passo gara di ciascun pilota senza prendere in considerazione i pitstop*:

$$=MEDIA(\%stint_i:\%stint_f) \tag{17}$$

Si riporta un esempio della formula elaborata in azione:

=MEDIA(AP2:AT36)

A2.10 Differenza di tempo in percentuale dei giri con pitstop [$\Delta\%_{pitstop}$]

Arrivati a questo punto ci siamo concentrati sul calcolo della differenza di tempo in percentuale dei giri con pitstop, funzione della media tra tutte le coppie dei giri entrata/uscita dai box in regime regolare. Tale valore si riferisce al tempo medio impiegato per svolgere un pitstop espresso come percentuale rispetto alla media dei tempi dei giri regolari. Esso è ricavabile attraverso l'impiego della formula:

$$=MEDIA(REF_1)/REF_2*200-200 \quad (18)$$

dove con REF_1 abbiamo indicato, per l'appunto, le coordinate di tutte quelle celle contenenti le coppie formate da giro d'entrata e di uscita dai box svolti in regime di gara normale, mentre con REF_2 la media dei tempi fatti segnare nei giri regolari. Da notare come, nel caso dei Gran Premi come quello degli Stati Uniti in cui il traguardo è posizionato affinché solo il tempo di un giro sia influenzato dal pitstop, la formula (18) andrebbe sostituita dalla seguente:

$$=MEDIA(REF_1)/REF_2*100-100 \quad (19)$$

A2.11 Media dei tempi percentuali con pitstop [$\%M_{pitstop}$]

Con la formula (20) siamo stati in grado di definire il passo gara *reale* del pilota integrando i pitstop nella media dei tempi percentuali:

$$=(\%M*n+\Delta\%_{pitstop}*m)/n \quad (20)$$

Come si evince, la formula calcola il prodotto tra la media dei tempi percentuali del pilota $\%M$ e il numero di giri regolari portati a termine dal pilota in gara n, ovvero quei giri privi di *safety car* o *VSC*, sommando poi la differenza di tempo in percentuale dei giri con pitstop $\Delta\%_{pitstop}$ moltiplicati perla loro frequenza m, ovvero il numero di *stint* (ossia il lasso di tempo, o di giri, che intercorre tra due pitstop) meno uno, dividendo il tutto per il numero di giri regolari.
A seguire un esempio della formula applicata al pilota di F1 Lewis Hamilton:

$$=(AU122*64+\$AX\$2*3)/64$$

Acknowledgements

Research carried out in collaboration with the Big&Open Data Innovation Laboratory (BODaI-Lab), University of Brescia (project nr. 03-2016, title “Big Data analytics in sports”, www.bodai.unibs.it/bdsports/), granted by Fondazione Cariplo and Regione Lombardia.

Riferimenti bibliografici e sitografia

[1] S. Borra, A. Di Ciaccio, Statistica. Metodologie per le scienze economiche e sociali, Mc Graw-Hill, Milano (2008)

[2] P. Newbold, W.L. Carlson, B. Thorne, Statistica, 2° ed., Pearson Italia, Milano (2010)

[3] Manisera M., Zuccolotto P. - “Sport analytics : la statistica divertente”, in Induzioni: demografia, probabilità, statistica a scuola : 53, 2, 2016 - Fabrizio Serra Ed.

[4] Ergast Developer API
(http://ergast.com/mrd/)

[5] Wikipedia, *Gran Premio d’Olanda 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_d%27Olanda_2021)

[6] Formulapassion.com, *GP Olanda 2021, Gara: Verstappen manda in delirio Zandvoort*
(https://www.formulapassion.it/motorsport/live/gp-olanda-2021-gara-diretta-zandvoort-verstappen-hamilton-leclerc-sainz-giovinazzi-alonzo-vettel-kubica-norris-ricciardo-redbull-mercedes-ferrari-alfa-romeo-583116.html)

[7] Wikipedia, *Gran Premio d’Italia 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_d%27Italia_2021)

[8] Formulapassion.com, *F1 / GP Italia 2021, Gara: Ricciardo-Norris, 1-2 McLaren*
(https://www.formulapassion.it/motorsport/live/gp-italia-2021-gara-diretta-monza-verstappen-hamilton-ricciardo-norris-leclerc-vettel-sainz-ferrari-mclaren-mercedes-red-bull-giovinazzi-alonso-alfa-romeo-aston-martin-584451.html)

[9] Wikipedia, *Gran Premio degli Stati Uniti d’America 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_degli_Stati_Uniti_d%27America_2021)

[10] Formulapassion.com, *F1/ GP USA 2021, Gara: Verstappen piega Hamilton, 3° Perez*
(https://www.formulapassion.it/motorsport/live/f1-gp-stati-uniti-2021-gara-diretta-austin-hamilton-verstappen-leclerc-sainz-alonso-vettel-raikkonen-giovinazzi-bottas-perez-mercedes-red-bull-ferrari-alfa-romeo-aston-martin-590212.html)

[11] Wikipedia, *Gran Premio di Città del Messico 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_di_Città_del_Messico_2021)

[12] Formulapassion.com, *F1/ GP Messico 2021, Gara: Verstappen suona la nona in Messico*
(https://www.formulapassion.it/motorsport/live/f1-gp-messico-2021-gara-diretta-mexico-city-ferrari-leclerc-sainz-redbull-verstappen-perez-mercedes-hamilton-raikkonen-vettel-giovinazzi-norris-alonso-mclaren-alfa-romeo-591468.html)

[13] Wikipedia, *Gran Premio di San Paolo 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_di_San_Paolo_2021)

[14] Formulapassion.com, *F1/ GP Brasile 2021, Gara: Hamilton firma un capolavoro*
(https://www.formulapassion.it/motorsport/live/f1-gp-brasile-2021-gara-diretta-san-paolo-vettel-leclerc-sainz-alonso-raikkonen-verstappen-hamilton-ferrari-mercedes-redbull-alfa-romeo-giovinazzi-norris-ricciardo-592584.html)

[15] Wikipedia, *Gran Premio del Qatar 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_del_Qatar_2021)

[16] Formulapassion.com, *F1/ GP Qatar 2021, Gara: Lewis davanti a Max, 3° un eroico Alonso*
(https://www.formulapassion.it/motorsport/live/f1-gp-qatar-2021-gara-diretta-losail-ferrari-mercedes-redbull-verstappen-hamilton-leclerc-sainz-perez-bottas-raikkonen-alonso-vettel-alfa-romeo-giovinazzi-norris-ricciardo-mclaren-593335.html)

[17] Wikipedia, *Gran Premio d'Arabia Saudita 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_d%27Arabia_Saudita_2021)

[18] Formulapassion.com, *F1/ GP Arabia Saudita 2021, Gara: Lewis batte Max, pareggio in classifica*
(https://www.formulapassion.it/motorsport/live/f1-gp-arabia-saudita-2021-gara-diretta-gedda-verstappen-hamilton-mercedes-redbull-ferrari-sainz-leclerc-alonso-vettel-raikkonen-giovinazzi-norris-ricciardo-perez-bottas-595090.html)

[19] Wikipedia, *Gran Premio di Abu Dhabi 2021*
(https://it.wikipedia.org/wiki/Gran_Premio_di_Abu_Dhabi_2021)

[20] Formulapassion.com, *F1/ GP Abu Dhabi 2021, Gara: Verstappen 'dieci' e titolo*
(https://www.formulapassion.it/motorsport/live/f1-gp-abu-dhabi-2021-gara-hamilton-vs-verstappen-per-il-titolo-diretta-ultima-raikkonen-giovinazzi-yas-marina-mercedes-vs-redbull-ferrari-leclerc-sainz-alonso-vettel-596529.html)

STATISTICA IN POLE

Un progetto di:
Bellinaso Marco, Bergamo Tommaso, Bonato Edoardo e Bortoliero Daniele

Tutor scolastico:
Meneghini Lorenzo

ASSOCIAZIONE NEMESIS

BODAI - LAB
BIG&OPEN DATA INNOVATION LABORATORY

Math is everywhere
Corradini in research

Corradini