F1WP – Formula 1 Winner Prediction

Progetto per il corso di Fondamenti di Intelligenza Artificiale A.A. 2021/2022 di Giorgio Angelo Esposito, matricola 0512107389

https://github.com/giorgio-angelo-esposito/F1WP.git

1. INTRODUZIONE

La Formula Uno è uno sport automobilistico nato ufficialmente nel 1950 ed è attualmente lo sport automobilistico di più alta categoria per quanto riguarda le vetture monoposto a ruota scoperta da corsa su circuito.

Il termine "Formula" si riferisce all'insieme di regole che i partecipanti (team e piloti) devono rispettare.

Nel tempo le vetture si sono evolute molto diventando, nell'ultimo decennio, l'apice della tecnologia nelle corse automobilistiche.

Una gara di Formula Uno non comprende però, solo la gara stessa: un *Gran Premio* occupa un intero *weekend*: si inizia il giovedì con le interviste ai piloti, prove libere il venerdì e il sabato mattina, qualifiche il sabato pomeriggio e gara la domenica.

Data la natura estremamente dinamica dello sport, fatta di sorpassi, incidenti e guasti, viene naturale provare a prevedere chi tra i partecipanti alla gara sarà il vincitore, cosa non sempre scontata.

2. DESCRIZIONE DELL'AGENTE

L'obiettivo del progetto è quindi quello di realizzare un agente capace di determinare quale pilota vincerà un Gran Premio.

Andiamo ora a definire la misura PEAS (**P**erformance, **E**nvironment, **A**ctuators, **S**ensors) dell'agente:

TABELLA DELLA MISURA PEAS	
PERFORMACE	La misura di performance dell'agente è la sua capacità di predire correttamente il vincitore di un Gran Premio di Formula Uno
ENVIRONMENT	 L'ambiente in cui opera l'agente è: OSSERVABILE: l'agente ha sempre accesso a tutti i dati che ha a disposizione DISCRETO: l'agente ha un numero limitato di dati da cui apprendere AGENTE SINGOLO STATICO: l'ambiente non cambia mentre l'agente sta apprendendo
ACTUATORS	L'attuatore dell'agente corrisponde alla predizione effettuata
SENSORS	I sensori dell'agente sono l'insieme di dati passati in input al machine learner per apprendere

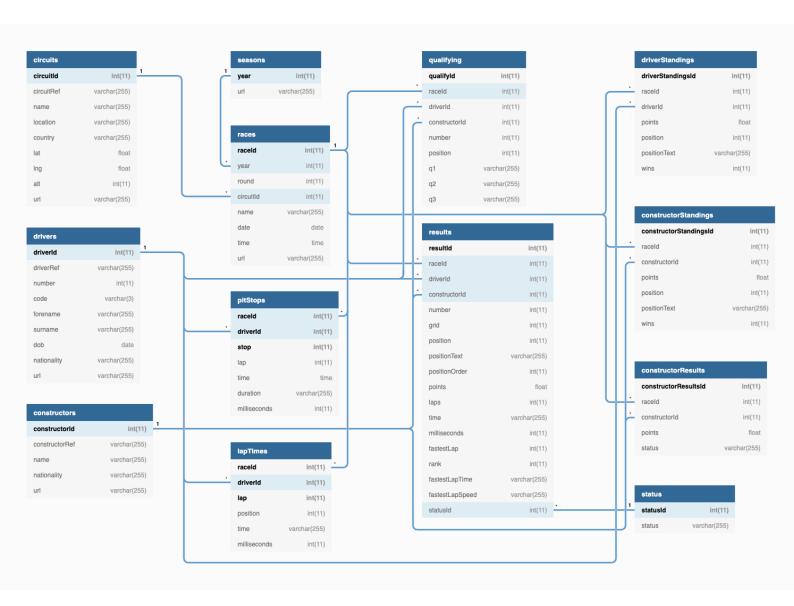
3. RACCOLTA E ANALISI DEI DATI

3.1 SCELTA DEL DATASET

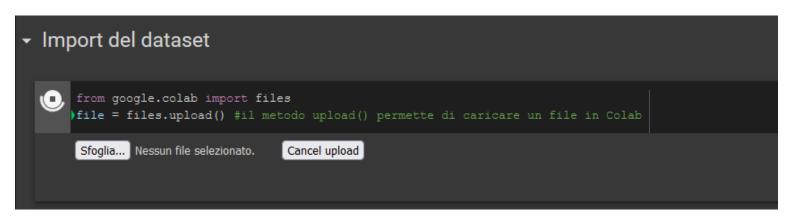
Passiamo ora al dataset che andremo a utilizzare per i nostri scopi. Dopo aver cercato sui vari siti dedicati (Kaggle, Google Dataset Search, ecc...), la scelta è ricaduta su un web service chiamato Ergast Developer API (http://ergast.com/mrd/). Ergast fornisce dati storici relativi alle corse automobilistiche e, oltre a poter interrogare il database mediante query parametriche, permette di scaricare i dati in formato .csv (comma separeted value).

Trattandosi di un database relazionale, avremo diverse tabelle su cui eseguire le tecniche di Data Analysis.

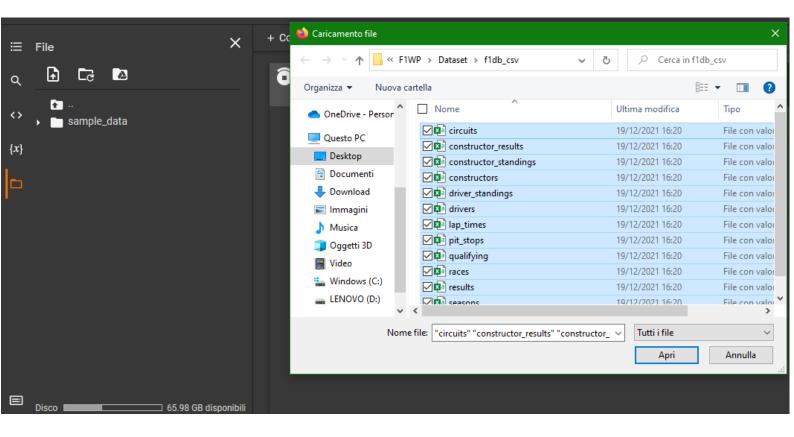
Di seguito, la struttura del database.



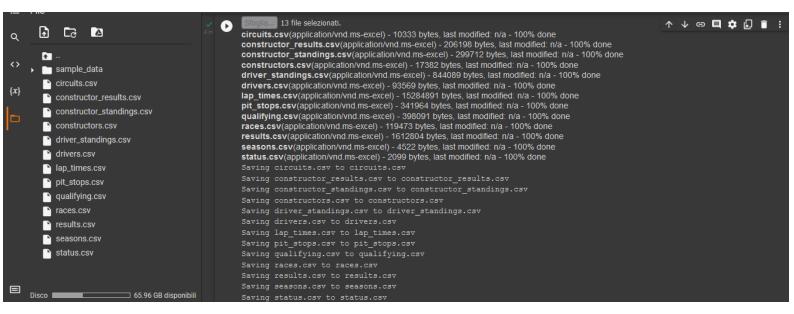
Procediamo quindi ad importare i vari file .csv e a descrivere i dati che contengono: per importare i file in Colab utilizziamo il metodo upload() di files, nativo di Colab.



Selezioniamo quindi i file.



Infine, i file sono importati correttamente.



Procediamo ora ad analizzare tabella per tabella, esaminando i campi in esse contenuti, in maniera tale da scegliere quali sono utili ai nostri scopi.

3.2 DESCRIZIONE E ANALISI TABELLE

Andiamo adesso a descrivere e esaminare le tabelle che compongono il database. Ergast mette a disposizione una descrizione delle sue tabelle, e per ognuna di queste andremo a esaminare il contenuto, visualizzando i dati, cercando relazioni tra di essi ed eventuali dati mancanti.

Come strumenti useremo il linguaggio Python e le sue librerie pandas e matplotlib.

TABELLA CIRCUITS

circuitRef var	pe Nu		+		
circuitRef var	·i	11 Key	Default	Extra	Description
location var country var lat flo lng flo alt int			NULL NULL NULL NULL NULL	auto_increment 	Primary key Unique circuit identifier Circuit name Location name Country name Latitude Longitude Altitude (metres) Circuit Wikipedia page

La tabella circuits riporta le informazioni sui circuiti.

La tabella circuits contiene 79 record, nessuno dei quali contiene valori nulli.

TABELLA CONSTRUCTOR

Field	Type	Null	Key	Default	Extra	Description
constructorId constructorRef name nationality url		NO YES	PRI UNI 	NULL NULL	auto_increment 	

La tabella constructor riporta le informazioni sui costruttori, ovvero i team.

TABELLA CONSTRUCTOR RESULTS

Field	Type	Null	Key	Default	Extra	Description
constructorResultsId	+ int(11)	+	+	+	+ auto increment	+ Primarv kev
raceId	int(11)	NO		1 0	, <u>-</u>	Foreign key link to races table
constructorId	int(11)	NO		0	i	Foreign key link to constructors tab
points	float	YES		NULL	i	Constructor points for race
status	varchar(255)	YES		NULL	i	"D" for disqualified (or null)

La tabella constructor_results riporta le informazioni sui risultati ottenuti dai costruttori nelle gare.

	constructorRes	ultsId	raceId	constructorId	points	status
0		1	18	1	14.0	\N
1		2	18	2	8.0	\N
2		3	18	3	9.0	\N
3		4	18	4	5.0	\N
4		5	18	5	2.0	\N
11945		16445	1073	214	6.0	/N
11946		16446	1073	117	0.0	\N
11947		16447	1073	210	0.0	\N
11948		16448	1073	3	0.0	\N
11949		16449	1073	51	0.0	/N
[11950	rows x 5 colum	ns]				
constr	uctorResultsId	1195	0			
raceId		1195	0			
constr	uctorId	1195	0			
points		1195	0			
status		1195	0			
dtype:	int64					

All'interno della tabella sono presenti 11950 record, nessuno dei quali contiene valori nulli. Ma si può subito osservare che la colonna status sembra contenere tutti valori pari a \N, ma così non è infatti da un analisi più approfondita si può notare che alcuni record contengono anche valori diversi:

	constructorResultsId	raceId	constructorId	points	status
185	186	36	1	14.0	D
195	196	37	1	18.0	D
207	208	38	1	12.0	D
218	219	39	1	14.0	D
228	229	40	1	18.0	D
239	240	41	1	12.0	D
250	251	42	1	18.0	D
262	263	43	1	8.0	D
273	274	44	1	14.0	D
283	284	45	1	10.0	D
294	295	46	1	15.0	D
306	307	47	1	10.0	D
316	317	48	1	18.0	D
328	329	49	1	11.0	D
338	339	50	1	10.0	D
350	351	51	1	8.0	D
361	362	52	1	8.0	D

TABELLA CONSTRUCTOR STANDINGS

Field	Type	Nul	l Key	Ţ	Default	Extra	Description
constructorStandingsId	int(11)	NO	PRI	i	NULL	auto increment	Primary key
raceId	int(11)	NO				_	Foreign key link to races table
constructorId	int(11)	NO					Foreign key link to constructors table
points	float	NO					Constructor points for season
position	int(11)	YES			NULL		Constructor standings position (integer)
positionText	varchar(255)	YES			NULL		Constructor standings position (string)
wins	int(11)	NO			0		Season win count

La tabella constructor_standings riporta le informazioni sulla classifica costruttori.

	constructorStand	ingsId	raceId	positionText	wins
0		1	18	1	1
1		2	18	3	0
2		3	18	2	0
3		4	18	4	0
4		5	18	5	0
12711		27938	1074		0
12712		27939	1074		0
12713		27940	1074		0
12714		27941	1074		0
12715		27942	1074		0
[12716 :	rows x 7 columns	1			
	ctorStandingsId	12716	5		
raceId		12716			
constru	ctorId	12716	5		
points		12716	5		
position	n	12716			
position		12716			
wins		12716			
dtype:	int64				

La tabella contiene 12716 record e nessuno di essi contiene valori nulli.

Bisogna però fare una precisazione: nell'immagine, gli ultimi cinque valori di positionText sono contrassegnati come "-": questo perché i record relativi al raceId 1074 sono inerenti alla prima gara del Campionato del 2022, che comincerà a Marzo.

TABELLA DRIVERS

drivers table						
Field	Type	Null	Кеу	Default	Extra	Description
driverId driverRef number code forename surname dob nationality url	int(11) varchar(255) int(11) varchar(3) varchar(255) varchar(255) date varchar(255) varchar(255)	NO NO YES YES NO NO YES YES	PRI I I I I I I I I I I I I	NULL NULL NULL NULL NULL	auto_increment	Primary key Unique driver identifier Permanent driver number Driver code e.g. "ALO" Driver forename Driver surname Driver date of birth Driver nationality Driver Wikipedia page

La tabella drivers riporta le informazioni sui piloti.

TABELLA DRIVER STANDINGS

river_standings tab	le					
Field	Type	Null	Key	Default	Extra	Description
driverStandingsId raceId driverId points position positionText wins	int(11) int(11) int(11) int(11) float int(11) varchar(255) int(11)	NO NO NO NO YES YES	 PRI 	NULL 0 0 0 NULL NULL	auto_increment - - - -	Primary key Foreign key link to races table Foreign key link to drivers table Driver points for season Driver standings position (integer) Driver standings position (string) Season win count

La tabella driver_standings riporta le informazioni sulla classifica ottenuta dai piloti in una gara.

	driverStandingsId	raceId	driverId	points	position	positionText	wins
0	1	18	1	10.0	1	1	1
1	2	18	2	8.0	2	2	0
2	3	18	3	6.0	3	3	0
3	4	18	4	5.0	4	4	0
4	5	18	5	4.0	5	5	0
33389	70776	1074	840	0.0	16		0
33390	70777	1074	852	0.0	17		0
33391	70778	1074	830	0.0	18		0
33392	70779	1074	20	0.0	19		0
33393	70780	1074	855	0.0	20		0
[33394	rows x 7 columns]	<u> </u>					

La tabella contiene 33394 record, nessuno contenente valori nulli.

Allo stesso modo di constructor_standings i valori relativi a raceId 1074 sono relativi alla prima gara del Campionato del 2022 che comincerà a Marzo.

TABELLA LAP TIMES

lap_times table						
Field	Type	Null	Key	Default	Extra	Description
raceId driverId lap position time milliseconds	int(11) int(11) int(11) int(11) int(11) varchar(255) int(11)	NO NO NO YES YES YES	PRI PRI PRI PRI	NULL NULL NULL NULL NULL	 	Foreign key link to races table Foreign key link to drivers table Lap number Driver race position Lap time e.g. "1:43.762" Lap time in milliseconds

La tabella lap_times riporta le informazioni sui tempi ottenuti dai piloti durante le gare.

TABELLA PIT STOPS

pit_stops table							
Field	Type	Nu	111	Key	Default	Extra	Description
raceId driverId stop lap time duration milliseconds	int(11) int(11) int(11) int(11) int(11) time varchar(255) int(11)	NC		PRI PRI PRI	NULL NULL NULL NULL NULL NULL		Foreign key link to races table Foreign key link to drivers table Stop number Lap number Time of stop e.g. "13:52:25" Duration of stop e.g. "21.783" Duration of stop in milliseconds

La tabella pit_stops riporta le informazioni sui pit stops effetuati dai piloti durante le gare.

TABELLA QUALIFYING

Field	Type		Null	Key	Default	Extra	Description
qualifyId	int(11)		NO	+ PRI	NULL	auto increment	Primary key
raceId	int(11)	ı	NO			ı -	Foreign key link to races table
driverId	int(11)	1	NO				Foreign key link to drivers table
constructorId	int(11)	1	NO				Foreign key link to constructors tabl
number	int(11)	1	NO				Driver number
position	int(11)	1	YES		NULL		Qualifying position
q1	varchar(255)	YES		NULL		Q1 lap time e.g. "1:21.374"
q2	varchar(255)	YES		NULL		Q2 lap time
q3	varchar(255)	YES		NULL		Q3 lap time

La tabella qualifying riporta le informazioni sulle qualifiche dei Gran Premi.

qual	ifyId	raceId	driverId	q1	q2	q3
0	1	18	1	1:26.572	1:25.187	1:26.714
1	2	18	9	1:26.103	1:25.315	1:26.869
2	3	18	5	1:25.664	1:25.452	1:27.079
3	4	18	13	1:25.994	1:25.691	1:27.178
4	5	18	2	1:25.960	1:25.518	1:27.236
9130	9171	1073	849	1:24.338	\N	\N
9131	9172	1073	847	1:24.423	\N	\N
9132	9173	1073	8	1:24.779	/N	\N
9133	9174	1073	854	1:24.906	/N	\N
9134	9175	1073	853	1:25.685	\N	\N
[9135 rows	х 9 с	olumns]				
qualifyId		9135				
raceId		9135				
driverId		9135				
constructo	rId	9135				
number		9135				
position		9135				
q1		9127				
q2		9001				
q3		8880				
dtype: int	64					

Nella tabella qualifying sono presenti 9135 record, ma risulta subito evidente che nella tabella mancano dei dati, maggiormente nelle colonne q1 e q2: questo è motivato dal fatto che le metodologie di qualifiche negli anni sono cambiate e di conseguenza i tempi non coincidono con la struttura data alla tabella. Inoltre è presente il valore speciale "\N".

	qualifyId	raceId	driverId	constructorId	number	position	q1	q2	q 3
3880	3882	114	23	3	4	1	1:15.259	NaN	NaN
3881	3883	114	8	1	6	2	1:15.295	NaN	NaN
3882	3884	114	31	3	3	3	1:15.415	NaN	NaN
3883	3885	114	15	4	7	4	1:15.500	NaN	NaN
3884	3886	114	30	6	1	5	1:15.644	NaN	NaN
8669	8710	1046	825	210	20	16	0:54.705	NaN	NaN
8670	8711	1046	849	3	6	17	0:54.796	NaN	NaN
8671	8712	1046	851	3	89	18	0:54.892	NaN	NaN
8672	8713	1046	8	51	7	19	0:54.963	NaN	NaN
8673	8714	1046	850	210	51	20	0:55.426	NaN	NaN
255 rov	vs × 9 column:	s							

Sono presenti 255 record che contengono valori NaN

TABELLA RACES

Field	Type		Null	Ke	7	Default	Extra	Description
raceId	+ int(11)	-+·	NO	+ PRI	·-+ []	NULL	+ auto increment	Primary key
year	int(11)		NO	i				Foreign key link to seasons table
round	int(11)		NO	i				Round number
circuitId	int(11)		NO	i				Foreign key link to circuits tabl
name	varchar(255)		NO	i				Race name
date	date		NO	İ		0000-00-00		Race date e.g. "1950-05-13"
time	time		YES	ı		NULL		Race start time e.g."13:00:00"
url	varchar(255)		YES	UN1	1	NULL		Race Wikipedia page

La tabella races riporta le informazioni sui Gran Premi.

Sono presenti 1080 record e nessuno di questi contiene valori nulli. Da un'analisi più approfondita non risultano esserci dati con valori speciali.

TABELLA RESULTS

results table	+	+	+		+	
Field	Type	Null	Key	Default	Extra	Description
resultId raceId	int(11) int(11)	NO NO	PRI	0	auto_increment 	Primary key Foreign key link to races table
	int(11) int(11) int(11)	NO NO YES		0 0 NULL	 	Foreign key link to drivers table Foreign key link to constructors table Driver number
• •	int(11) int(11)	NO YES		0 NULL	i	Starring grid position Official classification, if applicable
positionOrder	varchar(255) int(11) float	NO NO NO		 0 0	 	Driver position string e.g. "1" or "R" Driver position for ordering purposes Driver points for race
	int(11) varchar(255)	NO YES		0 NULL		Number of completed laps Finishing time or gap
fastestLap	int(11) int(11)	YES YES YES		NULL NULL	 	Finishing time in milliseconds Lap number of fastest lap
rank fastestLapTime fastestLapSpeed		YES YES		O NULL NULL	 	Fastest lap rank, compared to other drivers Fastest lap time e.g. "1:27.453" Fastest lap speed (km/h) e.g. "213.874"
statusId 	int(11)	ио 	i +	0	; ;	Foreign key link to status table

La tabella results riporta le informazioni sui risultati dei Gran Premi.

	resi	ultId		driverId		fastestLapSpeed	
0		1	18	1	1:27.452	218.300	1
1		2	18	2	1:27.739	217.586	1
2		3	18	3	1:28.090	216.719	1
3		4	18	4	1:28.603	215.464	1
4		5	18	5	1:27.418	218.385	1
		25400	1073	815	1:26.419	219.993	5
		25401	1073	849	1:29.293	212.912	3
		25402	1073	841	1:29.442	212.557	6
		25403	1073	847	1:30.647	209.732	6
25	398 2	25404	1073		1:29.698	211.951	23
_	5399 rows	в ж 18]			
	sultId		25399				
	ceId		25399				
dr	iverId		25399				
co	nstructor	rId	25399				
nu	mber		25399				
gr	id		25399				
po	sition		25399				
_	sitionTe		25399				
_	sitionOrd	der	25399				
po	ints		25399				
la	ps		25399				
	me		25399				
mi	llisecond	ds	25399				
fa	stestLap		25399				
ra	nk		25399				
fa	stestLapl	Time	25399				
fa	stestLap	Speed	25399				
st	atusId		25399				
dt	ype: int@	64					
						-	

Come possiamo vedere sono presenti 25399 record nella tabella e nessuno di questi contiene dati mancanti.

Da un'analisi più approfondita notiamo, però, che all'interno della tabella è presente all'interno di più colonne il valore speciale "\N".

	resultId	raceId	driverId	constructorId	number	grid	position	positionText	positionOrder	points	laps	time	milliseconds
0		18			22					10.0	58	1:34:50.616	5690616
1	2	18	2	2	3	5	2	2	2	8.0	58	+5.478	5696094
2		18			7	7				6.0	58	+8.163	5698779
3	4	18	4	4	5	11	4	4	4	5.0	58	+17.181	5707797
4	5	18	5		23	3	5	5	5	4.0	58	+18.014	5708630
5	6	18	6	3	8	13		6	6	3.0	57	\N	W
6	7	18	7	5	14	17	7	7	7	2.0	55	\N	W
7	8	18	8	6	1	15	8	8	8	1.0	53	\N	W
8	9	18	9	2	4	2	\N	R	9	0.0	47	\N	W
9	10	18	10	7	12	18	\N	R	10	0.0	43	\N	۱N
10	11	18	11	8	18	19	\N	R	11	0.0	32	\N	W
11	12	18	12	4		20	\N	R	12	0.0	30	\N	\N
12	13	18	13		2	4	\N	R	13	0.0	29	\N	W
13	14	18	14	9	9	8	\N	R	14	0.0	25	\N	\N

TABELLA SEASON

Field Typ						
IICIG IVE	oe I	Null	Кеу	Default	Extra	Description
year int	(11) char(255)	NO	PRI UNI	0		Primary key e.g. 1950 Season Wikipedia page

La tabella season riporta le informazioni sulla singola stagione.

TABELLA STATUS

status table	=					
Field	Type	Null	Key	Default	Extra	Description
statusId status status	int(11) varchar(255)		PRI 	NULL	auto_increment	Primary key Finishing status e.g. "Retired"

La tabella status riporta le informazioni riguardanti lo stato dei piloti a fine gara.

- 4. ALGORITMI UTILIZZATI
- 5. VALUTAZIONI
- 6. SVILUPPI FUTURI
- 7. GLOSSARIO
- **Formula**: si riferisce all'insieme di regole che i partecipanti (team e piloti) devono rispettare.
- Prove Libere: le prove libere sono tre sessioni durante le quali i piloti possono prendere confidenza con la pista e gli ingegneri aggiustare e adattare meglio l'assetto della vettura. Attualmente ogni sessione dura 60 minuti, contro i 90 e i 45 adottati nelle precedenti stagioni.
- Qualifiche: le qualifiche servono a stabilire l'ordine di partenza della gara della domenica, la griglia di partenza. Attualmente la qualifica si compone di tre sessioni, denominate Q1, Q2 e Q3 della durata di 18, 15 e 12 minuti rispettivamente, con un sistema knock-out: le 5 vetture più lente al termine della Q1 sono eliminate, allo stesso modo le ultime 5, tra le vetture rimaste, al termine delle Q2 sono eliminate e le restanti combattono per la Pole Position nella Q3.
 - Il formato delle qualifiche è quello che ha subito più variazioni nel corso della storia della Formula Uno.
- Pole Position: è il termine che indica la prima posizione nella griglia di partenza.
- **Gran Premio**: erroneamente confuso con la gara della domenica, il Gran Premio indica tutti gli eventi che si svolgono dal giovedì alla domenica.
- **Costruttore**: un Costruttore può essere visto come il team stesso. Il termine deriva dal fatto che, inizialmente, un team costruiva sia telaio che motore per le monoposto di Formula Uno. Col tempo però, a causa delle grandi quantità di denaro da dover spendere per poter sviluppare una vettura, molti team hanno deciso di utilizzare un motore costruito da altri team, costruendo però da se il telaio, così da poter prendere parte al Mondiale Costruttori.
- Mondiale Piloti: il Mondiale Piloti è il titolo che si contendono ogni anno i
 piloti di Formula Uno. Alla fine di ogni gara, i primi dieci classificati ottengono
 un punteggio proporzionato alla posizione (1° = 25 punti, 2° = 18 punti, ..., 10°

- = 1 punto, secondo l' assegnazione attuale). Il pilota che a fine Campionato avrà totalizzato più punti si aggiudicherà il titolo.
- Mondiale Costruttori: il Mondiale Costruttori è l'analogo del Mondiale Piloti per i team di Formula Uno. I punteggi sono calcolati come la somma dei punti ottenuti dai piloti di un team al termine di una gara. Il team che avrà totalizzato più punti a fine Campionato avrà totalizzato più punti si aggiudicherà il titolo.