practica2

December 31, 2018

```
In [1]: import decimal
        import math
        import matplotlib.pyplot as plt
        import numpy as np
        import pandas as pd
        from scipy.stats import linregress
        from scipy.optimize import curve_fit
```

En primer lugar, procederemos a leer los datos del fichero[1]. En este paso convertiremos los datos en un data frame, para que sea más cómodo trabajar con ellos. Para ello, tenemos que solucionar el problema de que hay filas con más campos, debido a que el campo `Finish'' puede tener el caracter`,", y al leer el fichero se interpreta como dos campos independientes.

```
In [2]: # Abrimos el fichero con los datos
        file = open('datos_F1_1950_2018_prac2.csv', 'r', encoding="utf-16")

        # Leemos el fichero
        data = file.readlines()

        #>Cerramos el fichero
        file.close()

        # Eliminamos los caracteres de saltos de línea y dividimos por ","
        data = [(d.strip('\n')).split(',') for d in data]

        # Corregimos aquellas filas en las que hay una "," en el campo Finish
        for d in data:
            if len(d) > 10:
                d[8] = d[8] + ':' + d[9]
                d.pop(9)

        # Convertimos los datos en Data Frame
        df_data = pd.DataFrame.from_records(data[1:], columns=data[0])

        # Lo mostramos en pantalla
```

[1]Esta práctica se realizará usando Python 3.7. En el directorio de github se acompaña el código en python así como un fichero de Jupyter notebook

```
df_data
#set(list(df_data['GP']))
```

Out[2]:

	#	Year	GP	Driver	Number	\
0		1950	great britain	Nino Farina	2	
1		1950	great britain	Luigi Fagioli	3	
2		1950	great britain	Reg Parnell	4	
3		1950	great britain	Yves Giraud Cabantous	14	
4		1950	great britain	Louis Rosier	15	
5		1950	great britain	Bob Gerard	12	
6		1950	great britain	Cuth Harrison	11	
7		1950	great britain	Philippe Ãtancelin	16	
8		1950	great britain	David Hampshire	6	
9		1950	great britain	Joe Fry Brian Shawe Taylor	10	
10		1950	great britain	Johnny Claes	18	
11		1950	great britain	Juan Manuel Fangio	1	
12		1950	great britain	Joe Kelly	23	
13		1950	great britain	Prince Bira	21	
14		1950	great britain	David Murray	5	
15		1950	great britain	Geoff Crossley	24	
16		1950	great britain	Toulo de Graffenried	20	
17		1950	great britain	Louis Chiron	19	
18		1950	great britain	EugÃÍne Martin	17	
19		1950	great britain	Peter Walker Tony Rolt	9	
20		1950	great britain	Leslie Johnson	8	
21		1950	monaco	Juan Manuel Fangio	34	
22		1950	monaco	Alberto Ascari	40	
23		1950	monaco	Louis Chiron	48	
24		1950	monaco	Raymond Sommer	42	
25		1950	monaco	Prince Bira	50	
26		1950	monaco	Bob Gerard	26	
27		1950	monaco	Johnny Claes	6	
28		1950	monaco	Luigi Villoresi	38	
29		1950	monaco	Philippe Ãtancelin	14	
...		...	...	...	...	
23077		2018	brazilian f1	Brendon Hartley	28	
23078		2018	brazilian f1	Carlos Sainz Jr.	55	
23079		2018	brazilian f1	Pierre Gasly	10	
23080		2018	brazilian f1	Stoffel Vandoorne	2	
23081		2018	brazilian f1	Esteban Ocon	31	
23082		2018	brazilian f1	Sergey Sirotkin	35	
23083		2018	brazilian f1	Fernando Alonso	14	
23084		2018	brazilian f1	Lance Stroll	18	
23085		2018	brazilian f1	Nico HÃlkenberg	27	
23086		2018	brazilian f1	Marcus Ericsson	9	
23087		2018	abu dhabi f1	Lewis Hamilton	44	
23088		2018	abu dhabi f1	Sebastian Vettel	5	
23089		2018	abu dhabi f1	Max Verstappen	33	

23090	2018	abu dhabi f1	Daniel Ricciardo	3
23091	2018	abu dhabi f1	Valtteri Bottas	77
23092	2018	abu dhabi f1	Carlos Sainz Jr.	55
23093	2018	abu dhabi f1	Charles Leclerc	16
23094	2018	abu dhabi f1	Sergio PÃľrez	11
23095	2018	abu dhabi f1	Romain Grosjean	8
23096	2018	abu dhabi f1	Kevin Magnussen	20
23097	2018	abu dhabi f1	Fernando Alonso	14
23098	2018	abu dhabi f1	Brendon Hartley	28
23099	2018	abu dhabi f1	Lance Stroll	18
23100	2018	abu dhabi f1	Stoffel Vandoorne	2
23101	2018	abu dhabi f1	Sergey Sirotkin	35
23102	2018	abu dhabi f1	Pierre Gasly	10
23103	2018	abu dhabi f1	Esteban Ocon	31
23104	2018	abu dhabi f1	Marcus Ericsson	9
23105	2018	abu dhabi f1	Kimi RÃďikkÃűnen	7
23106	2018	abu dhabi f1	Nico HÃlkenberg	27

	Team	Grid position	Final position	Points	Finish	\
0	Alfa Romeo	1	1	9	02:13:23.600	
1	Alfa Romeo	2	2	6	02:13:26.200	
2	Alfa Romeo	4	3	4	02:14:15.600	
3	Talbot-Lago	6	4	3	+2 laps	
4	Talbot-Lago	9	5	2	+2 laps	
5	ERA	13	6	0	+3 laps	
6	ERA	15	7	0	+3 laps	
7	Talbot-Lago	14	8	0	+5 laps	
8	Maserati	16	9	0	+6 laps	
9	Maserati	2020	10	00	+6 laps	
10	Talbot-Lago	21	11	0	+6 laps	
11	Alfa Romeo	3	12	0	Oil Leak	
12	Alta	19	13	0	Not classified	
13	Maserati	5	14	0	Out of fuel	
14	Maserati	18	15	0	Engine	
15	Alta	17	16	0	Transition	
16	Maserati	8	17	0	Engine	
17	Maserati	11	18	0	Clutch	
18	Talbot-Lago	7	19	0	Oil pressure	
19	ERA	1010	20	00	Gearbox	
20	ERA	12	21	0	Compressor	
21	Alfa Romeo	1	1	9	03:13:18.700	
22	Ferrari	7	2	6	+1 lap	
23	Maserati	8	3	4	+2 laps	
24	Ferrari	9	4	3	+3 laps	
25	Maserati	15	5	2	+5 laps	
26	ERA	16	6	0	+6 laps	
27	Talbot-Lago	19	7	0	+6 laps	
28	Ferrari	6	8	0	Rear Axle	

29	Talbot-Lago	4	9	0	Oil leak
...	...	...	...	...	...
23077	Toro Rosso	16	11	0	01:28:08.494
23078	Renault	15	12	0	01:28:09.355
23079	Toro Rosso	9	13	0	01:28:12.798
23080	McLaren	20	14	0	01:28:14.332
23081	RP Force India	18	15	0	01:28:15.651
23082	Williams	14	16	0	01:27:25.980
23083	McLaren	17	17	0	01:27:27.461
23084	Williams	19	18	0	01:27:44.176
23085	Renault	13	19	0	Power unit
23086	Sauber	6	20	0	Collisiion damage
23087	Mercedes	1	1	25	01:39:40.382
23088	Ferrari	3	2	18	01:39:42.963
23089	Red Bull	6	3	15	01:39:53.088
23090	Red Bull	5	4	12	01:39:55.761
23091	Mercedes	2	5	10	01:40:28.339
23092	Renault	11	6	8	01:40:52.930
23093	Sauber	8	7	6	01:41:11.171
23094	RP Force India	14	8	4	01:41:11.657
23095	Haas	7	9	2	01:40:04.140
23096	Haas	13	10	1	01:40:05.675
23097	McLaren	15	11	0	01:40:23.266
23098	Toro Rosso	16	12	0	01:40:30.045
23099	Williams	20	13	0	01:40:31.261
23100	McLaren	18	14	0	01:40:31.945
23101	Williams	19	15	0	01:40:47.948
23102	Toro Rosso	17	16	0	Power unit
23103	RP Force India	9	17	0	Power unit
23104	Sauber	12	18	0	Power unit
23105	Ferrari	4	19	0	Power unit
23106	Renault	10	20	0	Collision

	Laps
0	70
1	70
2	70
3	68
4	68
5	67
6	67
7	65
8	64
9	3232
10	64
11	62
12	57
13	49

14	44
15	43
16	36
17	26
18	8
19	55
20	2
21	100
22	99
23	98
24	97
25	95
26	94
27	94
28	63
29	38
...	...
23077	70
23078	70
23079	70
23080	70
23081	70
23082	69
23083	69
23084	69
23085	43
23086	22
23087	55
23088	55
23089	55
23090	55
23091	55
23092	55
23093	55
23094	55
23095	54
23096	54
23097	54
23098	54
23099	54
23100	54
23101	54
23102	47
23103	46
23104	24
23105	7
23106	0

```
[23107 rows x 10 columns]
```

```
In [3]: LW = [d[0] == 'Lewis Hamilton' for d in df_data['Driver']]
        #print(df_data['Driver'][LW])
        puntos = [int(p[0]) for p in df_data['Points'][LW]]

        for y in range(2000, 2019):
            dummy = df_data['# Year'] == str(y)
            print(y, (set(list(df_data['GP'][dummy]))))
```

```
2000 {'san marino', 'france', 'malaysia', 'hungary', 'italy', 'spain', 'australia', 'great brit
2001 {'the united states', 'san marino', 'malaysia', 'france', 'hungary', 'italy', 'spain', 'au
2002 {'the united states', 'san marino', 'malaysia', 'france', 'hungary', 'italy', 'spain', 'au
2003 {'the united states', 'san marino', 'malaysia', 'france', 'hungary', 'italy', 'spain', 'au
2004 {'china', 'the united states', 'san marino', 'malaysia', 'france', 'hungary', 'italy', 'sp
2005 {'china', 'the united states', 'san marino', 'malaysia', 'france', 'hungary', 'italy', 'sp
2006 {'china', 'the united states', 'san marino', 'malaysia', 'france', 'hungary', 'italy', 'sp
2007 {'china', 'the united states', 'malaysia', 'france', 'hungary', 'italy', 'spain', 'brazil
2008 {'china', 'malaysia', 'hungary', 'italy', 'spain', 'australia', 'great britain', 'germany
2009 {'china', 'malaysia', 'hungary', 'italy', 'abu dhabi', 'spain', 'australia', 'great brita
2010 {'china', 'malaysia', 'hungary', 'italy', 'abu dhabi', 'spain', 'australia', 'great brita
2011 {'china', 'malaysia', 'korea', 'hungary', 'italy', 'abu dhabi', 'india', 'spain', 'austra
2012 {'china', 'brazilian', 'malaysia', 'korea', 'hungary', 'italy', 'abu dhabi', 'india', 'spa
2013 {'brazilian f1 gp', 'bahrain f1', 'chinese', 'spanish f1', 'indian f1 gp', 'abu dhabi f1 g
2014 {'australian f1 gp', 'brazilian f1 gp', 'austrian f1 gp', 'bahrain f1 gp', 'italian f1 gp
2015 {'australian f1 gp', 'austrian f1 gp', 'bahrain f1 gp', 'italian f1 gp', 'hungarian f1 gp
2016 {'japanese f1', 'malaysian f1', 'usa f1', 'british f1', 'hungarian f1', 'austrian f1', 'ca
2017 {'russian f1', 'monaco f1', 'chinese f1', 'bahrain f1', 'belgian f1', 'japanese f1', 'spa
2018 {'japanese f1', 'british f1', 'hungarian f1', 'austrian f1', 'canadian f1', 'russian f1',
```

```
In [4]: # Eliminamos los caracteres erróneos
        for j in range(len(df_data)):
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ã', 'É')
            df_data['Driver'][j] = df_data['Driver'][j].replace('ÃÍ', 'è')
            df_data['Driver'][j] = df_data['Driver'][j].replace('ÃĽ', 'é')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãą', 'á')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãă', 'á')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ã', 'Ó')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãş', 'ó')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãď', 'ä')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãű', 'ö')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ã', 'ü')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãŋ', 'í')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãğ', 'ç')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãť', 'ô')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãÿ', 'ø')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãž', 'ú')
```

```
            df_data['Driver'][j] = df_data['Driver'][j].replace('Ãś', 'ñ')
            df_data['Driver'][j] = df_data['Driver'][j].replace('Åą', '')
            df_data['Driver'][j] = df_data['Driver'][j].replace('van der', 'vander')

        list(set(df_data['Driver']))
```

```
Out[4]: ['Martin Brundle',
         'Dick Gibson',
         'Franco Forini',
         'Alberto Uria Oscar González',
         'Leo Kinnunen',
         'Bruno Senna',
         'Yves Giraud Cabantous',
         'Tom Belsø',
         'Sébastien Bourdais',
         'Bill Moss',
         'Daniil Kvyat',
         'René Arnoux',
         'Mark Blundell',
         'George Fonder',
         'Josef Peters',
         'Helmut Niedermayr',
         'Peter Westbury',
         'Erwin Bauer',
         'Rob Schroeder',
         'Hideki Noda',
         'Johnny Dumfries',
         'Lloyd Ruby',
         'Pat Flaherty Jim Rathmann',
         'Niki Lauda',
         'Bob Anderson',
         'Tony Settember',
         'Mauro Baldi',
         'Mike Hawthorn José Froilán González',
         'Yuji Ide',
         'Oswald Karch',
         'Johnny Thomson',
         'Elmer George',
         'Andy Linden Chuck Stevenson Jerry Hoyt',
         'Jean-Pierre Beltoise',
         'Troy Ruttman',
         'Bill Holland',
         'Massimiliano Papis',
         'Carlo Abate',
         'Juan Jover',
         'Emilio de Villota',
         'Tim Parnell',
         'Tony Maggs',
```

```
'Nino Vaccarella',
'Alan Rees',
'Vittorio Brambilla',
'Vic Elford',
'Jean-Louis Schlesser',
'Gianmaria Bruni',
'Timmy Mayer',
'Juan Manuel Bordeu',
'Alfonso de Portago Peter Collins',
'A.J. Foyt',
'Alessandro Pesenti-Rossi',
'John Fitch',
'Domenico Schiattarella',
'Jim Rathmann Eddie Johnson',
'Graham McRae',
'Masten Gregory',
'Derek Bell',
'Ignazio Giunti',
'Don Branson',
'Johnny Mantz Walt Faulkner',
'Trevor Blokdyk',
'Kurt Kuhnke',
'Jimmy Reece',
'André Milhoux',
'Ernesto Brambilla',
'Nanni Galli',
'Bobby Unser',
'Gerry Ashmore',
'Gabriele Tarquini',
'Nino Farina',
'Jan Magnussen',
'Pablo Birger',
'Helmuth Koinigg',
'Jean-Marc Gounon',
'Denny Hulme',
'Tony Brise',
'Don Edmunds',
'Ken Kavanagh',
'John James',
'Masahiro Hasemi',
'Tony Brooks Stuart Lewis-Evans',
'Sergey Sirotkin',
'Al Keller',
'Sergio Pérez',
'Fred Agabashian',
'George Follmer',
'Ingo Hoffmann',
'Peter Arundell',
```

```
'Renato Pirocchi',
'Alex Soler-Roig',
'Jan Flinterman',
'Bob Veith',
'Louis Chiron',
'Rubens Barrichello',
'Adrian Sutil',
'Piero Taruffi Juan Manuel Fangio',
'Maurício Gugelmin',
'Antonio Creus',
'Len Duncan George Fonder',
'Bruce Halford',
'Clemar Bucci Carlos Menditeguy Harry Schell',
'Gary Hocking',
'Louis Rosier',
'Divina Galica',
'Pastor Maldonado',
'José Dolhem',
'Alan Brown',
'Raul Boesel',
'Tommy Byrne',
'Luigi Fagioli',
'Jackie Holmes',
'Peter Walker Tony Rolt',
'Beppe Gabbiani',
'Jerry Unser',
'Jacques Pollet',
'Len Sutton',
'Stirling Moss',
'Gerino Gerini Chico Landi',
'Robin Widdows',
'Wilson Fittipaldi',
'Umberto Maglioli José Froilán González',
'Patrick Friesacher',
'Nasif Estéfano',
'John Watson',
'Herbert MacKay-Fraser',
'Alfonso de Portago',
'Eugène Chaboud Philippe Étancelin',
'Karl Kling',
'Bruce Kessler',
'Claudio Langes',
'Kevin Cogan',
'Peter Revson',
'Jackie Oliver',
'Chuck Arnold',
'Ed Elisian Eddie Russo',
'Jim Clark',
```

```
'Perry McCarthy',
'Tony Crook',
'Alessandro Zanardi',
'Jimmy Davies Jim Rathmann Sam Hanks',
'Bud Tingelstad',
'Ken Wharton',
'Brausch Niemann',
'Ernesto Prinoth',
'Ronnie Bucknum',
'Hans Herrmann',
'Richard Robarts',
'Lucas di Grassi',
'Andrea de Adamich',
'Jim Hurtubise',
'Emanuele Pirro',
'Duncan Hamilton',
'François Hesnault',
'Larry Perkins',
'Guy Edwards',
'Tiff Needell',
'Pedro Diniz',
'Ray Reed',
'Chico Landi',
'Nino Farina André Simon',
'Karl Wendlinger',
'Justin Wilson',
'Consalvo Sanesi',
'Stephen South',
'David Hobbs',
'Mike Hailwood',
'Rodney Nuckey',
'Alan Jones',
'Jack McGrath',
"Fritz d'Orey",
'Narain Karthikeyan',
'Franco Rol',
'José Froilán González Felice Bonetto',
'Sam Posey',
'Vitantonio Liuzzi',
'Rudi Fischer Peter Hirt',
'Giancarlo Baghetti',
'Gino Munaron',
'Jaime Alguersuari',
'Mike Magill',
'Chris Amon',
'Cliff Griffith',
'Duane Carter Marshall Teague Tony Bettenhausen Jimmy Jackson',
'Mike Sparken',
```

```
'Skip Barber',
'Toshio Suzuki',
'Mike Nazaruk',
'Luigi Fagioli Juan Manuel Fangio',
'George Abecassis',
'Felice Bonetto Juan Manuel Fangio',
'Pedro de la Rosa',
'Jackie Stewart',
'Joe Kelly',
'Alain de Changy',
'Michael May',
'Jochen Mass',
'Ricardo Zunino',
'Fernando Alonso',
'Oscar Larrauri',
'Rudolf Schoeller',
'Luigi Musso Juan Manuel Fangio',
'Ludovico Scarfiotti',
'Pierre-Henri Raphanel',
'Conny Andersson',
'Jack Fairman',
'Vincenzo Sospiri',
'Lance Stroll',
'Satoru Nakajima',
'Alain Prost',
'Ian Burgess',
'Tom Bridger',
'Johnnie Parsons',
'François Picard',
'Ed Elisian',
'Brian Naylor',
'Doug Serrurier',
'Ernie McCoy',
'Alex Caffi',
'Toranosuke Takagi',
'Brian Redman',
'Lamberto Leoni',
'Felice Bonetto',
'Guy Tunmer',
'Mike Beuttler',
'Gilles Villeneuve',
'Giedo vander Garde',
'Chris Lawrence',
'Luciano Burti',
'Sebastian Vettel',
'Ian Stewart',
'Dennis Taylor',
'Harry Blanchard',
```

```
'Mike Taylor',
'Andrea de Cesaris',
'Siegfried Stohr',
'Jean-Pierre Jabouille',
'Gregor Foitek',
'Adolf Brudes',
'Jean Behra Cesare Perdisa',
'Carel Godin de Beaufort',
'Philippe Adams',
'Stuart Lewis-Evans',
'Nico Hülkenberg',
'Dennis Poore',
'Henry Taylor',
'Felipe Massa',
'Syd vander Vyver',
'Julian Bailey',
'Luigi Piotti',
'Ricardo Londoño',
'Lee Wallard',
'Roger Penske',
'Jack Brabham',
'Tony Bettenhausen Joie Chitwood',
'Spider Webb',
'Luigi Musso Jean Behra Sergio Mantovani',
'Ralph Firman',
'Walt Brown',
'Mario Andretti',
'Sam Hanks',
'Henry Banks',
'Olivier Grouillard',
'Jean-Denis Délétraz',
'Harry Schell Ken Wharton',
'Daniel Ricciardo Valtteri Bottas',
'Shorty Templeman',
'Gerino Gerini',
'Romain Grosjean',
'Bob Bondurant',
'Patrick Nève',
'Nicolas Kiesa',
'Luigi Villoresi',
'Christian Klien',
'José Froilán González',
'Henri Louveau',
'Nelson Piquet Jr.',
'Jean Behra Roberto Mieres',
'Bob Scott',
'Don Beauman',
'Enrique Bernoldi',
```

```
'George Amick',
'Patrick Gaillard',
'Michel Leclère',
'Bruce McLaren',
'Alex Ribeiro',
'Dieter Quester',
'Graham Whitehead',
'Ernst Loof',
'Teo Fabi',
'Naoki Hattori',
'Ian Raby',
'Alan Stacey',
'Jim Rathmann',
'Hiroshi Fushida',
'Ray Crawford',
'Rudi Fischer',
'Pierre Levegh',
'Arturo Merzario',
'Helmut Marko',
'Bill Cheesbourg',
'Guy Ligier',
'Stéphane Sarrazin',
'Myron Fohr',
'Carroll Shelby Masten Gregory',
'Massimo Natili',
'Bruce Johnstone',
'Torsten Palm',
'George Constantine',
'Duke Dinsmore',
'Wayne Weiler',
'Michele Alboreto',
'Élie Bayol André Pilette',
'Peter Ashdown',
'Giorgio Francia',
'Lance Reventlow',
'Mike Parkes',
'Ronnie Peterson',
'Shinji Nakano',
'Elio de Angelis',
'Mark Donohue',
'Fritz Riess',
'Jody Scheckter',
'Geoff Duke',
'Roelof Wunderink',
'Mike Hawthorn Eugenio Castellotti',
'Giorgio Scarlatti Harry Schell',
'Paul Pietsch',
'Andre Lotterer',
```

```
'André Simon',
'Paul Belmondo',
'Phil Cade',
'Johnny Cecotto',
'Cesare Perdisa Peter Collins Wolfgang von Trips',
'Slim Borgudd',
'Franck Lagorce',
'Pedro Chaves',
'Gus Hutchison',
'Robert Doornbos',
'Jesús Iglesias',
'Robert Drake',
'Stirling Moss Jack Fairman',
'Ernie de Vos',
'Geoff Lees',
'Duane Carter Troy Ruttman',
'Georges Grignard',
'Wolfgang von Trips',
'Paddy Driver',
'Maria Teresa de Filippis',
'Aldo Gordini',
'Jim Hall',
'John Nicholson',
'Vern Schuppan',
'Nicola Larini',
'Roberto Guerrero',
'Jean Alesi',
'Mike Hawthorn',
'Ukyo Katayama',
'Philip Fotheringham-Parker',
'Jud Larson',
'Jean Lucas',
'Richie Ginther',
'Lewis Hamilton',
'Michael Schumacher',
'Carlos Pace',
'David Purley',
'Piero Monteverdi',
'Loris Kessel',
'André Testut',
'Anthony Davidson',
'Ernst Klodwig',
'Nino Farina Umberto Maglioli Maurice Trintignant',
'Ernie Pieterse',
'Juan Manuel Fangio Johnny Claes',
'Chuck Weyant',
'Piercarlo Ghinzani',
'Patrick Depailler',
```

```
'Allan McNish',
'Eddie Sachs',
'Damien Magee',
'Pascal Wehrlein',
'Raymond Sommer',
'André Pilette',
'Mark Webber',
'Rikky von Opel',
'Carlos Menditeguy',
'Luki Botha',
'Roland Ratzenberger',
'Stirling Moss Maurice Trintignant',
'Guy Mairesse',
'Mike Wilds',
'Jo Gartner',
'Éric Bernard',
'Jan Flinterman Chico Landi',
'Paul Russo Tony Bettenhausen',
'Kunimitsu Takahashi',
'Esteban Gutierrez',
'Jochen Rindt',
'Mike Fisher',
'Gianfranco Brancatelli',
'Bobby Rahal',
'Pierluigi Martini',
'Peter de Klerk',
'Brett Lunger',
'Lucien Bianchi',
'John Riseley-Prichard',
'Peter Gethin',
'Luigi Taramazzo',
'Brian Gubby',
'Manny Ayulo Jack McGrath',
'Clive Puzey',
'Olivier Gendebien',
'Desiré Wilson',
'Carlo Gimax Franchi',
'Albert Scherrer',
'Thomas Monarch',
'Gino Bianco',
'Desmond Titterington',
'Leslie Thorne',
'Nino Farina Felice Bonetto',
'Corrado Fabi',
'Esteban Tuero',
'Tony Brooks Stirling Moss',
'Tony Bettenhausen',
'Eddie Irvine',
```

```
'Sakon Yamamoto',
'Georges Berger',
'Tiago Monteiro',
'Peter Whitehead',
'Jimmy Stewart',
'Esteban Ocon',
'Hans Von Stuck',
'Sam Hanks Duane Carter',
'Alex Blignaut',
'Johnny Herbert',
'Chuck Daigh',
'Franck Montagny',
'Luigi Villoresi Eugenio Castellotti',
'Günther Seiffert',
'Emilio Zapico',
'Tony Rolt',
'Paolo Barilla',
'Tomá Enge',
'Jackie Lewis',
'Silvio Moser',
'Luigi Musso',
'Ottorino Volonterio André Simon',
'Robert Manzon',
'Eitel Cantoni',
'Roberto Mieres',
'Fred Agabashian Paul Russo',
'Rodger Ward Eddie Johnson',
'Harry Schell Jean Behra',
'Neville Lederle',
'Ayrton Senna',
'Walt Hansgen',
'Bob Said',
'Jacky Ickx',
'Eric Brandon',
'Paul Russo',
'Peter Hirt',
'Hans Heyer',
'Jenson Button',
'Luigi Villoresi Alberto Ascari',
'Umberto Maglioli Jean Behra',
'Keke Rosberg',
'Bob Christie',
'Eric van de Poele',
'Mack Hellings',
'Nico Rosberg',
'Walt Ader',
'Jimmy Jackson',
'Robin Montgomerie-Charrington',
```

```
'Jyrki Järvilehto',
'Alessandro de Tomaso',
'Maurice Trintignant',
'Max Verstappen',
'Chris Bristow',
'Michael Andretti',
'Maurice Trintignant Harry Schell',
'Nello Pagani',
'Ian Scheckter',
'Jack Brabham Mike MacDowel',
'Mika Salo',
'Marcel Balsa',
'Graham Hill',
'Kevin Magnussen',
'Art Cross',
'Charles Pic',
'Brian Henton',
'Kenny Acheson',
'Ivor Bueb',
'Eugène Martin',
'John Taylor',
'Horace Gould',
'Allen Berg',
'Johnny Servoz-Gavin',
'Gianni Morbidelli',
'Stefan Johansson',
'Mika Häkkinen',
'Ricardo Rosset',
'Kamui Kobayashi',
'Reg Parnell',
'Jim McWithey',
'Ted Whiteaway',
'Frank Dochnal',
'Renzo Zorzi',
'Dave Kennedy',
'Reine Wisell',
'Paul di Resta',
'Ralf Schumacher',
'Bill Schindler',
'Stefan Bellof',
'Érik Comas',
'Roger Laurent',
'Jimmy Daywalt',
'Miguel Angel Guerra',
'Pete Lovely',
'Andy Linden',
'Eugenio Castellotti Alfonso de Portago',
'Christian Fittipaldi',
```

```
'Marcus Ericsson',
'Larry Crockett',
'Antônio Pizzonia',
'Sébastien Buemi',
'Eddie Keizan',
'Jimmy Davies',
'Hans Klenk',
'Kurt Ahrens Jr.',
'Willi Heeks',
'Aguri Suzuki',
'John Campbell-Jones',
'Giovanna Amati',
'Ken Downing',
'Mike Hawthorn Wolfgang von Trips',
'David Piper',
'Toulo de Graffenried Ottorino Volonterio',
'Jo Siffert',
'Emanuele Naspetti',
'Jarno Trulli',
'Rio Haryanto',
'Juan Manuel Fangio Eugenio Castellotti',
'Dempsey Wilson',
'Dave Morgan',
'Duane Carter',
'Giovanni Lavaggi',
'Brendon Hartley',
'Paul England',
'David Murray',
'Peter Collins Juan Manuel Fangio',
'François Mazet',
'Ron Flockhart',
'Cal Niday',
'Heinz-Harald Frentzen',
'Felipe Nasr',
'Roberto Merhi',
'Damon Hill',
'Red Amick',
'John Barber',
'Rodger Ward Andy Linden Duke Dinsmore',
'Wolfgang Seidel',
'Jerry Hoyt',
'Gene Hartley',
'Jean-Pierre Jarier',
'Bertil Roos',
'Óscar Alfredo Gálvez',
'Alessandro Nannini',
'Johnny Thomson Jackie Holmes Spider Webb',
'Heinz Schiller',
```

```
'Phil Hill',
'Giorgio Pantano',
'Rudolf Krause',
'Daniel Ricciardo',
'David Clapham',
'Roberto Bonomi',
'Roberto Moreno',
'Bruno Giacomelli',
'Maurice Trintignant Peter Collins',
'Bill Brack',
'Art Bisch',
'Franco Comotti',
'Jay Chamberlain',
'Onofre Marimón',
'Cecil Green',
'Theo Fitzau',
'John Love',
'Mike Hawthorn Harry Schell',
'Frank Armi George Fonder',
'Bill Homeier',
'Luigi Musso Sergio Mantovani',
'Alberto Uria',
'Adolfo Schewelm Cruz',
'Carlos Reutemann',
'Dries vander Lof',
'Walt Faulkner Chuck Stevenson',
'Rodger Ward',
'Keith Greene',
'Piero Taruffi Paul Frère',
'Michael Bartels',
'Cliff Allison',
'Huub Rothengatter',
'Thierry Boutsen',
'Giorgio Scarlatti',
'Cesare Perdisa',
'Luigi Villoresi Jo Bonnier',
'Hans Joachim Stuck',
'Jorge Daponte',
'Otto Stuppacher',
'Kazuki Nakajima',
'Dave Charlton',
'Stirling Moss Cesare Perdisa',
'Jacques Villeneuve',
'Charles de Tornaco',
'Ken Richardson',
'Cesare Perdisa Jean Behra',
'Walt Faulkner Bill Homeier',
'Henri Pescarolo',
```

```
'Riccardo Paletti',
'Antonio Giovinazzi',
'Chico Serra',
'Marc Gené',
'Eddie Cheever',
'Juan Pablo Montoya',
'Ian Ashley',
'Adrián Campos',
'Ottorino Volonterio',
'Eugène Chaboud',
'Innes Ireland',
'Bill Vukovich',
'Paul Russo Jerry Hoyt',
'Piero Scotti',
'Edgar Barth',
'Stoffel Vandoorne',
'Yannick Dalmas',
'Gastón Mazzacane',
'Paul Hawkins',
'Jules Bianchi',
'David Brabham',
'Roberto Bussinello',
'Peter Collins',
'Brian Shawe Taylor',
'Piero Carini',
'Keith Andrews',
'Nick Heidfeld',
'Giancarlo Fisichella',
'Chet Miller',
'John Miles',
'Bill Holland Jim Rathmann',
'Max Chilton',
'Bill Aston',
'Archie Scott Brown',
'Mike Thackwell',
'Jean-Eric Vergne',
'Piero Taruffi',
'Paco Godia',
'Christian Goethals',
'Leslie Johnson',
'Christijan Albers',
'Scott Speed',
'Harry Schell',
'Bobby Grim',
'Kazuyoshi Hoshino',
'Bill Cantrell Bayliss Levrett',
'Carl Forberg',
'Mauri Rose',
```

```
'Geoff Crossley',
'Gene Hartley Tony Bettenhausen Chuck Stevenson',
'Boy Hayje',
'Les Leston',
'John Rhodes',
'Alexander Wurz',
'Giovanni de Riu',
'Clemente Biondetti',
'Louis Rosier Charles Pozzi',
'Sergio Mantovani Luigi Musso Harry Schell',
'Kenneth McAlpine',
'Eugenio Castellotti',
'Ivan Capelli',
'Al Pease',
'Andy Linden Johnny Thomson Jimmy Daywalt',
'James Hunt',
'Clemar Bucci',
'Carl Scarborough',
'Howden Ganley',
'Alberto Colombo',
'David Walker',
'Hans Binder',
'George Eaton',
'Jean Lucienbonnet',
'Jonathan Williams',
'Bill Mackey',
'Mike Harris',
'Robert Kubica',
'Alexander Rossi',
'Roberto Lippi',
'Brian Hart',
'François Migault',
'Emerson Fittipaldi',
'Riccardo Patrese',
'Cuth Harrison',
'Toulo de Graffenried',
'Azdrubal Fontes',
'Peter Walker',
'Richard Attwood',
'Taki Inoue',
'Jack Turner',
'Alberto Rodriguez Larreta',
'Tony Gaze',
'Jan Lammers',
'Jimmy Bryan',
'Tom Pryce',
'Roger Loyer',
'Danny Sullivan',
```

```
'Bob Gerard',
'Vic Wilson',
'David Hampshire',
'Luiz Bueno',
'Jacques Laffite',
'Arthur Legat',
'Jonathan Palmer',
'Ricardo Rodríguez',
'Fabrizio Barbazza',
'Lorenzo Bandini',
'Bernie Ecclestone',
"Jérôme d'Ambrosio",
'Mikko Kozarowitzky',
'Toni Ulmen',
'Eliseo Salazar',
'Manny Ayulo',
'Gijs Van Lennep',
'David Coulthard',
'Michael Bleekemolen',
'Cristiano da Matta',
'Andrea Chiesa',
'Noritake Takahara',
'Lance Macklin',
'Duke Nalon',
'Philippe Alliot',
'André Guelfi',
'Willy Mairesse',
'Kurt Ahrens',
'Robert La Caze',
'Olivier Panis',
'Jo Bonnier',
'Stefano Modena',
'Bernd Schneider',
'Alex Yoong',
'Johnnie Tolan',
'Paul Emery',
'Jac Nelleman',
'Danny Ongais',
'Prince Bira',
'Umberto Maglioli',
'Bill Whitehouse',
'Carlos Sainz Jr.',
'Joe James',
'Johnny Claes',
'Colin Davis',
'Alan Rollinson',
'Kimi Räikkönen',
'Joe Fry Brian Shawe Taylor',
```

```
'Andrea Montermini',
'Eddie Johnson',
'Heini Walter',
'Danny Kladis Spider Webb',
'Johnny Boyd',
'Bob Sweikert',
'Alberto Ascari',
'Theo Helfrich',
'Eugenio Castellotti Luigi Musso',
'Tony Trimmer',
'Leslie Marr',
'Hap Sharp',
'Jim Rigsby',
'Olivier Beretta',
"Pat O'Connor",
'Gerhard Berger',
'Giorgio Bassi',
'Ricardo Zonta',
'Marco Apicella',
'Bob Evans',
'Carroll Shelby',
'Eddie Russo',
'Ed Elisian Bob Scott',
'Élie Bayol',
'Tony Brooks',
'Hubert Hahne',
'Christian Danner',
'Colin Chapman',
'Karl Kling Stirling Moss Hans Herrmann',
'Pierre Gasly',
'Billy Garrett',
'Jim Crawford',
'Tarso Marques',
'Al Herman',
'Max Jean',
'Ettore Chimeri',
'John Cordts',
'Gérard Larrousse',
'Bertrand Gachot',
'Jos Verstappen',
'Karun Chandhok',
'Gene Hartley Marshall Teague',
'Heikki Kovalainen',
'Tony Shelly',
'Carl Scarborough Bob Scott',
'Harald Ertl',
'Will Stevens',
'Marc Surer',
```

```
'Patrick Tambay',
'George Connor',
'Gunnar Nilsson',
'Valtteri Bottas',
'Pascal Fabre',
'Tony Marsh',
'Enrico Bertaggia',
'Giulio Cabianca',
"Robert O'Brien",
'Roger Williamson',
'Jean-Eric Vergne Marcus Ericsson',
'Günther Bechem',
'Piers Courage',
'Eric Thompson',
'Gaetano Starrabba',
'Philippe Streiff',
'Andy Linden Bob Scott',
'Paul Frère',
'Alfonso de Portago José Froilán González',
'Dick Rathmann',
'Max de Terra',
'Xavier Perrot',
'Pat Flaherty',
'Manfred Winkelhock',
'Martin Donnelly',
'Harry Schell Giorgio Scarlatti',
'Hernando da Silva Ramos',
'Jacques Villeneuve Sr.',
'Sergio Mantovani',
'Chris Irwin',
'Gene Force',
'Kurt Adolff',
'Nelson Piquet',
'Timo Glock',
'Moisés Solana',
'Luca Badoer',
'Derek Daly',
'Rob Slotemaker',
'Fred Wacker',
'José Froilán González Alberto Ascari',
'Johnnie Parsons Andy Linden Art Cross Jimmy Davies Sam Hanks',
'Walt Faulkner',
'Philippe Étancelin',
'Jo Schlesser',
'Teddy Pilette',
'Jean-Christophe Boullion',
'Ben Pon',
'David Prophet',
```

```
'Eppie Wietzes',
'Didier Pironi',
'Chris Craft',
'Jean Behra',
'Mario de Araujo Cabral',
'Jacques Swaters',
'Paul Goldsmith',
'Toulo de Graffenried Harry Schell',
'Roy Salvadori',
'Juan Manuel Fangio',
'Luis Perez-Sala',
'Dorino Serafini Alberto Ascari',
'Jolyon Palmer',
'Takuma Sato',
'Charles Leclerc',
'Chuck Stevenson',
'Jackie Pretorius',
'Frank Gardner',
'Bobby Ball',
'Jo Vonlanthen',
'Warwick Brown',
'Vitaly Petrov',
'Dan Gurney',
'Nigel Mansell',
'Norberto Fontana',
'Gerhard Mitter',
'Hector Rebaque',
'Markus Winkelhock',
'Mike Spence',
'Giacomo "Geki" Russo',
'François Cevert',
'Tom Jones',
'Clay Regazzoni',
'Stirling Moss Tony Brooks',
'Rolf Stommelen',
'Zsolt Baumgartner',
'Trevor Taylor',
'Toni Branca',
'Rupert Keegan',
'Derek Warwick',
'Tim Schenken',
'Sam Tingle',
'Pedro Lamy',
'Marshall Teague',
'Nino Farina José Froilán González Maurice Trintignant',
'John Surtees',
'Lella Lombardi',
'Peter Broeker',
```

```
         'Johnny McDowell',
         'Hermann Lang',
         'Ron Flockhart Prince Bira',
         'Bernard Collomb',
         'Don Freeland',
         'Pedro Rodríguez',
         'Alfredo Pián']
```

In [5]:
```
# Estimamos el número de pilotos en cada campo
N_drivers = [math.floor(len(d.split())/2) for d in df_data['Driver']]

# Buscamos aquellos campos con un solo nombre
indice_1  = np.array(N_drivers)==1
drivers_list = list(set(df_data['Driver'][indice_1]))
```

In [6]:
```
drivers_list.append('Joie Chitwood')
drivers_list.append('Dries van der Lof')
drivers_list.append('Bill Cantrell')
drivers_list.append('Bayliss Levrett')
drivers_list.append('Hernando da Silva Ramos')
drivers_list.append('Carel Godin de Beaufort')
drivers_list.append('Maria Teresa de Filippis')
drivers_list.append('Mario de Araujo Cabral')
drivers_list.append('Syd van der Vyver')
drivers_list.append('Pedro de la Rosa')
drivers_list.append('Giedo van der Garde')
drivers_list.append('Eric van de Poele')
drivers_list.append('Joe Fry')
drivers_list.append('Brian Shawe Taylor')
drivers_list.append('Dorino Serafini')
drivers_list.append('Alberto Ascari')
drivers_list.append('Johnny Mantz')
drivers_list.append('Walt Faulkner')
drivers_list.append('Frank Armi')
drivers_list.append('George Fonder')
drivers_list.append('Danny Kladis')
drivers_list.append('Spider Webb')
drivers_list.append('Len Duncan')
drivers_list.append('George Fonder')
drivers_list.append('Charles Pozzi')
drivers_list.append('Alberto Uria')
drivers_list.append('Oscar González')
drivers_list.append('Mike MacDowel')
```

In [7]:
```
print(sorted(drivers_list))
```

```
['A.J. Foyt', 'Adolf Brudes', 'Adolfo Schewelm Cruz', 'Adrian Sutil', 'Adrián Campos', 'Aguri S
```

In [8]:
```
# Creamos una lista donde almacenar el resultado
sep_drivers = []

# Para cada registro...
for i in range(len(df_data['Driver'])):
    # Definimos una variable para simplificar el código
    d = df_data['Driver'][i]

    # Variable auxiliar donde almacenamos la posición
    # inicial del nombre de los pilotos
    limits = []

    # Para cada nombre en la lista...
    for dr in drivers_list:
        # Comprobamos si el nombre está en el registro
        if dr in d:
            # En caso afirmativo, lo almacenamos
            limits.append(d.find(dr))

    # Ordenamos las posiciones de forma ascendente
    limits = sorted(limits)
    # Y añadimos la longitud de la cadena
    limits.append(len(d))

    # Creamos una variable para almacenar los resultados parciales
    dummy = []
    # Extraemos los nombres del campo del registro
    for j in range(len(limits)-1):
        dummy.append(d[limits[j]:limits[j+1]].strip(' '))

    # Limpiamos los resultados
    dummy = [st for st in dummy if st != '']

    # Y los añadimos a la variable creada para ello
    sep_drivers.append(dummy)
```

In [9]:
```
# Creamos una variable para almacenar los pilotos que faltan en la lista
missing = []

# Para cada registro...
for i in range(len(df_data['Driver'])):
    # Comparamos la longitud del campo original con la
    # suma de las longitudes de los nombres separados
    if len(' '.join(sep_drivers[i])) != len(df_data['Driver'][i]):
        # Si son diferentes, añadimos el registro a la variable
        missing.append(df_data['Driver'][i])

# Mostramos en pantalla los pilotos que faltan
```

```
        print(list(set(missing)))

        #[s_d == [] for s_d in sep_drivers]
        #print(df_data['Driver'][indices])
```

```
[]
```

```
In [10]: df_data['Driver'] = sep_drivers
```

```
In [11]: # Identificamos el número de pilotos en cada registro
         n_items = [len(d) for d in df_data['Driver']]

         # Creamos unas variables que almacenan los datos que nos interesan
         Laps = []
         Grid = []
         Points = []

         # Para cada registro
         for i in range(len(n_items)):
             # Atributo "Laps"
             # Determinamos la longitud del atributo para cada piloto
             n=int(len(df_data['Laps'][i])/n_items[i])
             # Dividimos la cadena de texto en subcadenas de la longitud calculada
             dummy_1 = [df_data['Laps'][i][j * n:(j + 1) * n] for j in range((len(df_data['Laps
             # Almacenamos el resultado
             Laps.append(dummy_1)

             # Atributo "Grid Position"
             # Determinamos la longitud del atributo para cada piloto
             n=int(len(df_data['Grid position'][i])/n_items[i])
             # Si el atributo está vacío
             if (len(df_data['Grid position'][i]) == 0):
                 # Lo cambiamos por la palabra "Pitlane"
                 dummy_2 = [np.nan] * n_items[i]
             # Si el atributo tiene una longitud menor que el número de pilotos
             elif (len(df_data['Grid position'][i]) < n_items[i]):
                 # Suponemos que el atributo aplica a todos ellos
                 dummy_2 = [df_data['Grid position'][i]] * n_items[i]
             # En otro caso
             else:
                 # Dividimos la cadena de texto en subcadenas de la longitud calculada
                 dummy_2 = [df_data['Grid position'][i][j * n:(j + 1) * n] for j in range((len
             # Almacenamos el resultado
             Grid.append(dummy_2)

             # Atributo "Points"
             # Determinamos la longitud del atributo para cada piloto
```

```
            n=int(len(df_data['Points'][i])/n_items[i])
            # Si el atributo está vacío
            if (df_data['Points'][i] == ''):
                # Lo cambiamos por el valor "0"
                dummy_3 = [0] * n_items[i]
            # Si el atributo tiene una longitud menor que el número de pilotos
            elif (len(df_data['Points'][i]) < n_items[i]):
                # Suponemos que el atributo aplica a todos ellos
                dummy_3 = df_data['Points'][i] * n_items[i]
            # En otro caso
            else:
                # Dividimos la cadena de texto en subcadenas de la longitud calculada
                dummy_3 = [df_data['Points'][i][j * n:(j + 1) * n] for j in range((len(df_data
            # Almacenamos el resultado
            Points.append(dummy_3)

        # Sustituimos las columnas en el data frame por las nuevas columnas
        df_data['Laps'] = Laps
        df_data['Grid position'] = Grid
        df_data['Points'] = Points
```

In [12]:
```
df_data['GP'] = [d.replace('indianapolis 500', 'Indianapolis 500') for d in df_data['
df_data['GP'] = [d.replace('australian', 'australia') for d in df_data['GP']]
df_data['GP'] = [d.replace('austrian', 'austria') for d in df_data['GP']]
df_data['GP'] = [d.replace('belgian', 'belgium') for d in df_data['GP']]
df_data['GP'] = [d.replace('brazilian', 'brazil') for d in df_data['GP']]
df_data['GP'] = [d.replace('british', 'great britain') for d in df_data['GP']]
df_data['GP'] = [d.replace('canadian', 'canada') for d in df_data['GP']]
df_data['GP'] = [d.replace('caesars palace', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace('chinese', 'china') for d in df_data['GP']]
df_data['GP'] = [d.replace('dallas', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace('detroit', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace('european', 'europe') for d in df_data['GP']]
df_data['GP'] = [d.replace('french', 'france') for d in df_data['GP']]
df_data['GP'] = [d.replace('german f1', 'germany') for d in df_data['GP']]
df_data['GP'] = [d.replace('germanyy', 'german') for d in df_data['GP']]
df_data['GP'] = [d.replace('hungarian', 'hungary') for d in df_data['GP']]
df_data['GP'] = [d.replace('indian', 'india') for d in df_data['GP']]
df_data['GP'] = [d.replace('italian', 'italy') for d in df_data['GP']]
df_data['GP'] = [d.replace('japanese', 'japan') for d in df_data['GP']]
df_data['GP'] = [d.replace('the pacific', 'japan') for d in df_data['GP']]
df_data['GP'] = [d.replace('korean', 'korea') for d in df_data['GP']]
df_data['GP'] = [d.replace('malaysian', 'malaysia') for d in df_data['GP']]
df_data['GP'] = [d.replace('mexican', 'mexico') for d in df_data['GP']]
df_data['GP'] = [d.replace('pescara', 'italy') for d in df_data['GP']]
df_data['GP'] = [d.replace('russian', 'russia') for d in df_data['GP']]
df_data['GP'] = [d.replace('spanish', 'spain') for d in df_data['GP']]
df_data['GP'] = [d.replace('the netherland', 'the netherlands') for d in df_data['GP'
```

```
df_data['GP'] = [d.replace('the netherlandss', 'the netherlands') for d in df_data['GP']]
df_data['GP'] = [d.replace('the united states west', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace('the united states', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace('united states', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace('us f1', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace('usa f1', 'u.s.a.') for d in df_data['GP']]
df_data['GP'] = [d.replace(' gp', '') for d in df_data['GP']]
df_data['GP'] = [d.replace(' f1', '') for d in df_data['GP']]

for i in range(len(df_data)):
    df_data['GP'][i] = df_data['GP'][i].title()

set(list(df_data['GP']))
```

Out[12]:
```
{'Abu Dhabi',
 'Argentina',
 'Australia',
 'Austria',
 'Azerbaijan',
 'Bahrain',
 'Belgium',
 'Brazil',
 'Britain',
 'Canada',
 'China',
 'Europe',
 'France',
 'Germany',
 'Great Britain',
 'Hungary',
 'India',
 'Indianapolis 500',
 'Italy',
 'Japan',
 'Japanse',
 'Korea',
 'Luxembourg',
 'Malaysia',
 'Mexico',
 'Monaco',
 'Morocco',
 'Portugal',
 'Russia',
 'San Marino',
 'Singapore',
 'South Africa',
 'Spain',
 'Sweden',
```

```
 'Switzerland',
 'The Netherlands',
 'Turkey',
 'U.S.A.'}
```

In [13]:
```
# Creamos variables para almacenar los atributos de cada registro
Year   = []
GP     = []
Driver = []
Number = []
Team   = []
Grid   = []
Final  = []
Points = []
Finish = []
Laps   = []

# Para cada registro...
for i in range(len(df_data)):
    # Y por cada nombre en el campo "Driver"...
    for j in range(len(df_data['Driver'][i])):
        # Añadimos los atributos de cada registro
        # a la variable corresondiente
        Year.append(int(df_data['# Year'][i]))
        GP.append(df_data['GP'][i])
        Driver.append(str(np.array(df_data['Driver'][i][j]).flatten()[0]))
        Number.append(int(df_data['Number'][i]))
        Team.append(df_data['Team'][i])
        Grid.append(float(np.array(df_data['Grid position'][i][j]).flatten()[0]))
        Final.append(int(np.array(df_data['Final position'][i]).flatten()[0]))
        Points.append(float(np.array(df_data['Points'][i][j]).flatten()[0]))
        Finish.append(df_data['Finish'][i])
        Laps.append(int(np.array(df_data['Laps'][i][j]).flatten()[0]))

# Creamos un data frame con estas variables
df = pd.DataFrame.from_items([('Year', Year),
        ('GP', GP),
        ('Driver', Driver),
        ('Number', Number),
        ('Team', Team),
        ('Grid', Grid),
        ('Final', Final),
        ('Points', Points),
        ('Finish', Finish),
        ('Laps', Laps),
         ])
```

```
/opt/local/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/ipyker
```

```
In [14]: df
```

Out[14]:

	Year	GP	Driver	Number	Team
0	1950	Great Britain	Nino Farina	2	Alfa Romeo
1	1950	Great Britain	Luigi Fagioli	3	Alfa Romeo
2	1950	Great Britain	Reg Parnell	4	Alfa Romeo
3	1950	Great Britain	Yves Giraud Cabantous	14	Talbot-Lago
4	1950	Great Britain	Louis Rosier	15	Talbot-Lago
5	1950	Great Britain	Bob Gerard	12	ERA
6	1950	Great Britain	Cuth Harrison	11	ERA
7	1950	Great Britain	Philippe Étancelin	16	Talbot-Lago
8	1950	Great Britain	David Hampshire	6	Maserati
9	1950	Great Britain	Joe Fry	10	Maserati
10	1950	Great Britain	Brian Shawe Taylor	10	Maserati
11	1950	Great Britain	Johnny Claes	18	Talbot-Lago
12	1950	Great Britain	Juan Manuel Fangio	1	Alfa Romeo
13	1950	Great Britain	Joe Kelly	23	Alta
14	1950	Great Britain	Prince Bira	21	Maserati
15	1950	Great Britain	David Murray	5	Maserati
16	1950	Great Britain	Geoff Crossley	24	Alta
17	1950	Great Britain	Toulo de Graffenried	20	Maserati
18	1950	Great Britain	Louis Chiron	19	Maserati
19	1950	Great Britain	Eugène Martin	17	Talbot-Lago
20	1950	Great Britain	Peter Walker	9	ERA
21	1950	Great Britain	Tony Rolt	9	ERA
22	1950	Great Britain	Leslie Johnson	8	ERA
23	1950	Monaco	Juan Manuel Fangio	34	Alfa Romeo
24	1950	Monaco	Alberto Ascari	40	Ferrari
25	1950	Monaco	Louis Chiron	48	Maserati
26	1950	Monaco	Raymond Sommer	42	Ferrari
27	1950	Monaco	Prince Bira	50	Maserati
28	1950	Monaco	Bob Gerard	26	ERA
29	1950	Monaco	Johnny Claes	6	Talbot-Lago
...	...	...	...	...	...
23202	2018	Brazil	Brendon Hartley	28	Toro Rosso
23203	2018	Brazil	Carlos Sainz Jr.	55	Renault
23204	2018	Brazil	Pierre Gasly	10	Toro Rosso
23205	2018	Brazil	Stoffel Vandoorne	2	McLaren
23206	2018	Brazil	Esteban Ocon	31	RP Force India
23207	2018	Brazil	Sergey Sirotkin	35	Williams
23208	2018	Brazil	Fernando Alonso	14	McLaren
23209	2018	Brazil	Lance Stroll	18	Williams
23210	2018	Brazil	Nico Hülkenberg	27	Renault
23211	2018	Brazil	Marcus Ericsson	9	Sauber
23212	2018	Abu Dhabi	Lewis Hamilton	44	Mercedes
23213	2018	Abu Dhabi	Sebastian Vettel	5	Ferrari
23214	2018	Abu Dhabi	Max Verstappen	33	Red Bull
23215	2018	Abu Dhabi	Daniel Ricciardo	3	Red Bull

23216	2018	Abu Dhabi	Valtteri Bottas	77	Mercedes
23217	2018	Abu Dhabi	Carlos Sainz Jr.	55	Renault
23218	2018	Abu Dhabi	Charles Leclerc	16	Sauber
23219	2018	Abu Dhabi	Sergio Pérez	11	RP Force India
23220	2018	Abu Dhabi	Romain Grosjean	8	Haas
23221	2018	Abu Dhabi	Kevin Magnussen	20	Haas
23222	2018	Abu Dhabi	Fernando Alonso	14	McLaren
23223	2018	Abu Dhabi	Brendon Hartley	28	Toro Rosso
23224	2018	Abu Dhabi	Lance Stroll	18	Williams
23225	2018	Abu Dhabi	Stoffel Vandoorne	2	McLaren
23226	2018	Abu Dhabi	Sergey Sirotkin	35	Williams
23227	2018	Abu Dhabi	Pierre Gasly	10	Toro Rosso
23228	2018	Abu Dhabi	Esteban Ocon	31	RP Force India
23229	2018	Abu Dhabi	Marcus Ericsson	9	Sauber
23230	2018	Abu Dhabi	Kimi Räikkönen	7	Ferrari
23231	2018	Abu Dhabi	Nico Hülkenberg	27	Renault

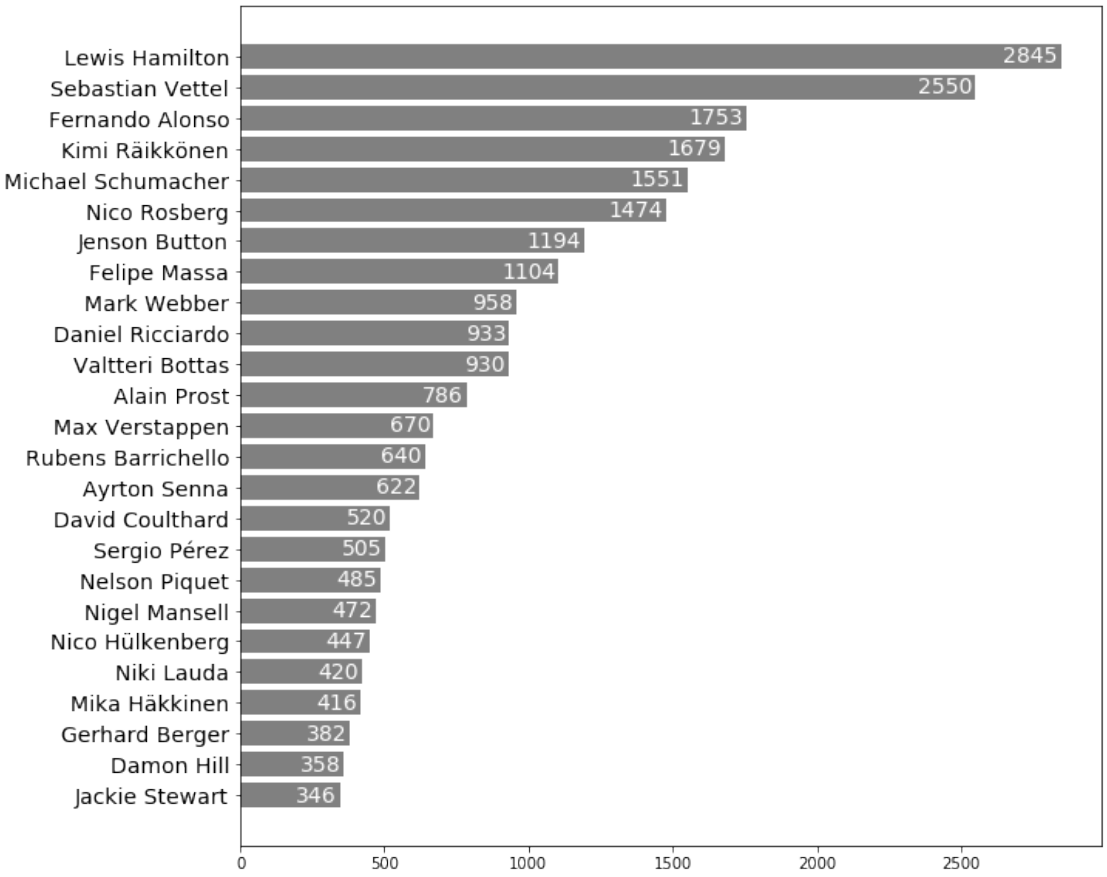
	Grid	Final	Points	Finish	Laps
0	1.0	1	9.0	02:13:23.600	70
1	2.0	2	6.0	02:13:26.200	70
2	4.0	3	4.0	02:14:15.600	70
3	6.0	4	3.0	+2 laps	68
4	9.0	5	2.0	+2 laps	68
5	13.0	6	0.0	+3 laps	67
6	15.0	7	0.0	+3 laps	67
7	14.0	8	0.0	+5 laps	65
8	16.0	9	0.0	+6 laps	64
9	20.0	10	0.0	+6 laps	32
10	20.0	10	0.0	+6 laps	32
11	21.0	11	0.0	+6 laps	64
12	3.0	12	0.0	Oil Leak	62
13	19.0	13	0.0	Not classified	57
14	5.0	14	0.0	Out of fuel	49
15	18.0	15	0.0	Engine	44
16	17.0	16	0.0	Transition	43
17	8.0	17	0.0	Engine	36
18	11.0	18	0.0	Clutch	26
19	7.0	19	0.0	Oil pressure	8
20	10.0	20	0.0	Gearbox	5
21	10.0	20	0.0	Gearbox	5
22	12.0	21	0.0	Compressor	2
23	1.0	1	9.0	03:13:18.700	100
24	7.0	2	6.0	+1 lap	99
25	8.0	3	4.0	+2 laps	98
26	9.0	4	3.0	+3 laps	97
27	15.0	5	2.0	+5 laps	95
28	16.0	6	0.0	+6 laps	94
29	19.0	7	0.0	+6 laps	94

```
...     ...    ...     ...                ...   ...
23202  16.0     11     0.0       01:28:08.494    70
23203  15.0     12     0.0       01:28:09.355    70
23204   9.0     13     0.0       01:28:12.798    70
23205  20.0     14     0.0       01:28:14.332    70
23206  18.0     15     0.0       01:28:15.651    70
23207  14.0     16     0.0       01:27:25.980    69
23208  17.0     17     0.0       01:27:27.461    69
23209  19.0     18     0.0       01:27:44.176    69
23210  13.0     19     0.0         Power unit    43
23211   6.0     20     0.0  Collisiion damage    22
23212   1.0      1    25.0       01:39:40.382    55
23213   3.0      2    18.0       01:39:42.963    55
23214   6.0      3    15.0       01:39:53.088    55
23215   5.0      4    12.0       01:39:55.761    55
23216   2.0      5    10.0       01:40:28.339    55
23217  11.0      6     8.0       01:40:52.930    55
23218   8.0      7     6.0       01:41:11.171    55
23219  14.0      8     4.0       01:41:11.657    55
23220   7.0      9     2.0       01:40:04.140    54
23221  13.0     10     1.0       01:40:05.675    54
23222  15.0     11     0.0       01:40:23.266    54
23223  16.0     12     0.0       01:40:30.045    54
23224  20.0     13     0.0       01:40:31.261    54
23225  18.0     14     0.0       01:40:31.945    54
23226  19.0     15     0.0       01:40:47.948    54
23227  17.0     16     0.0         Power unit    47
23228   9.0     17     0.0         Power unit    46
23229  12.0     18     0.0         Power unit    24
23230   4.0     19     0.0         Power unit     7
23231  10.0     20     0.0          Collision     0

[23232 rows x 10 columns]
```

In [15]:
```
selection_1 = (df['Points'] >0) & (np.isfinite(df['Points']))
```

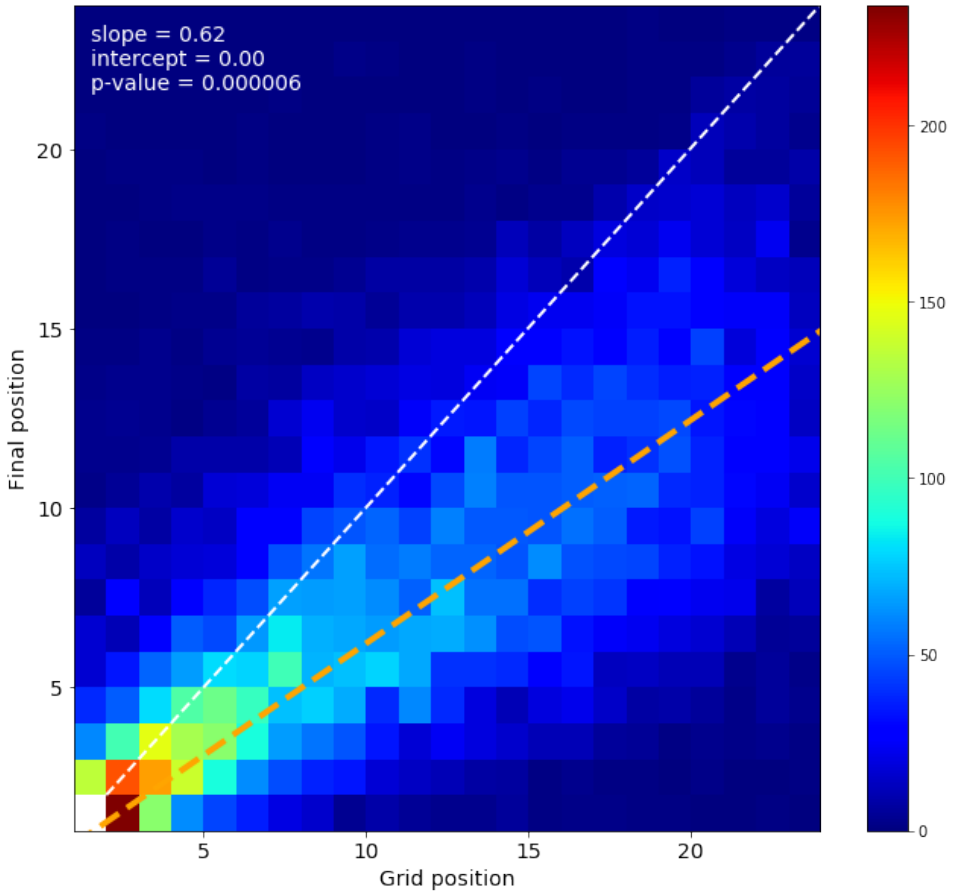
```
# creamos una lista en la que cada piloto aparezca una vez
scoring_drivers = np.array(list(set(df['Driver'][selection_1])))

# creamos una variable para almacenar los datos
drivers_score = []
# Para cada piloto...
for driver in scoring_drivers:
    # sumamos sus puntos
    dummy = np.sum(df['Points'][df['Driver']==driver])
    # y los almacenamos
    drivers_score.append(dummy)
```

```
# array con índices de ordenación
sort_ind = np.argsort(drivers_score)

# variables para el gráfico con los 25 valores de puntuación más altos
plot_score = np.array(drivers_score)[sort_ind][-25:]
plot_drivers = scoring_drivers[sort_ind][-25:]
y_pos = np.arange(1, len(plot_score)+1, 1)

# Creamos el gráfico
fig, ax1 = plt.subplots(figsize=(10,10))
ax1.barh(y_pos, plot_score, align='center',
         color='gray', ecolor='black')
ax1.set_yticks(np.arange(1, len(plot_score)+1, 1))
ax1.set_yticklabels(plot_drivers, fontsize=14)

for i in range(len(plot_score)):
    plt.text(plot_score[i]-10, y_pos[i]-0.2, str(int(plot_score[i])),
             horizontalalignment='right', color='w', fontsize=14)
plt.show()
```



In [16]:
```
# Suprimimos los signos '+' del format ''+x laps'
selection_2 = [d.replace('+', '') for d in df['Finish']]
# En algunos campos aparece un signo '-', también lo quitamos
selection_2 = [d.replace('-', '') for d in selection_2]
# Miramos si el primer carácter es un dígito
selection_2 = [d[0].isdigit() for d in selection_2]
# Seleccionamos también aquellos datos no nulos en la posición de salida
selection_2 = selection_2 & (np.isfinite(df['Grid']))

# Variables para la representación gráfica con la selección de los datos
x = df['Grid'][selection_2]
y = df['Final'][selection_2]
# Calculamos un ajuste lineal
def func(x, a):
    return x*a

dummy = np.histogram2d(x,y, bins=[32, 23], normed=False)
reg = [d>10 for d in dummy[0]]

x_fit = []
y_fit = []

for i in range(len(dummy[0])):
    for j in range(len(dummy[0][i])):
        for k in range(int(dummy[0][i][j])):
            x_fit.append(dummy[1][i])
            y_fit.append(dummy[2][j])

p0 = curve_fit(func, x_fit, y_fit)

# Representamos los puntos en escala de grises y el ajuste
fig, ax2 = plt.subplots(figsize=(11,10))
#plt.plot(x, y, 'ok', alpha=0.025)
plt.hist2d(x, y, bins=[32, 23], normed=False, cmax=400, cmap='jet')
plt.colorbar()
plt.axis([1, 24, 1, 24])
plt.plot([1, 25], [1,25], '--w', linewidth=2)
plt.plot([1, 25], [1*p0[0], 25*p0[0]], '--', color='orange', linewidth=4)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.xlabel('Grid position', fontsize=14)
plt.ylabel('Final position', fontsize=14)
plt.text(1.5, 23.5, 'slope = ' + str("%.2f" % p0[0]) + '\n' +
         'intercept = ' + str("%.2f" % 0.00) + '\n' +
         'p-value = ' + str("%.6f" % p0[1]),
         horizontalalignment='left', verticalalignment='top',
         color='w', fontsize=14)
```

```
plt.show()
```


slope = 0.62
intercept = 0.00
p-value = 0.000006
Final position
Grid position
5
10
15
20
0
50
100
150
200

In [ ]: