

Minería y Análisis de Datos

Eduardo Gutiérrez Navarrete
Juan Carlos Soto Hernández
Christopher Chávez
Miguel Castañeda Martínez



Contexto de F1



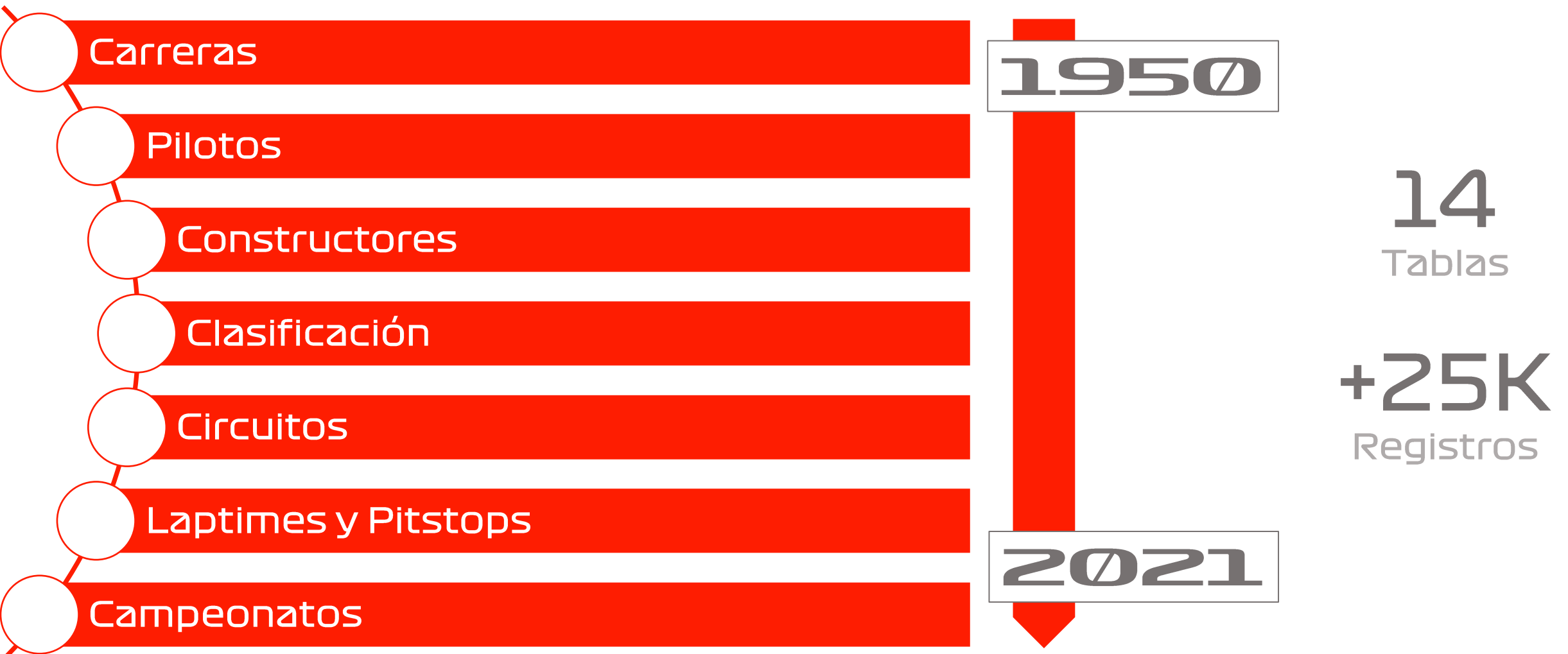
La F1 es la **clase más alta de automovilismo** regulada por la FIA.

El Campeonato Mundial de F1 ha sido uno de los principales en todo el mundo desde su **temporada inaugural en 1950**.

Una **temporada de F1** consiste en una **serie de carreras** (Grandes Premios) que tienen lugar en todo el mundo en circuitos de carreras o urbanos.

Acerca de la base

Se puede encontrar un dataset de **F1 en Kaggle** que contiene información sobre:



Diccionario de variables

Circuits

circuitId	circuitRef	name	location	country	lat	lng	alt	url
Id	Reference name of circuit	Actual name of circuit	City	Country	Latitude	Longitude	Altitude	Wiki URL

Constructor_results

constructorResultsId	raceId	constructorId	points	status
id	Id of Race	Id of Constructor	Points	Status of result

Constructor_standings

constructorStandingsId	raceId	constructorId	points	position	positionText	wins
Id	Id of race	Id of constructor	points	Final position	Final position in text	Number of wins

Lap_times

raceId	driverId	lap	position	time	milliseconds
Id	Id of driver	Lap number	Final position	time	milliseconds

Pit_stops

raceId	driverId	stop	lap	time	duration	milliseconds
Id	Id of driver	stop number	Lap number	Time of pit stop	Duration of pit stop	miliseconds

Qualifying

qualifyId	raceId	driverId	constructorId	number	position	q1	q2	q3
Id	Id of race	Id of driver	Id of constructor	Car number	Final position	Time in Qualifying 1	Time in Qualifying 2	Time in Qualifying 3

Constructors

constructorId	constructorRef	name	nationality	url
Id	Reference name of constructor	Actual name of constructor	Country	Wiki URL

Driver_standings

driverStandingsId	raceId	driverId	points	position	positionText	wins
Id	Id of race	Id of driver	Points	Final Position	Final position in text	Wins

Drivers

driverId	driverRef	number	code	forename	surname	dob	nationality	url
Id	Reference name of driver	Driver number	3-letter alphabet code of driver	First name	Last Name	Date of birth	Country	Wiki URL

Races

raceId	year	round	circuitId	name	date	time	url	fp1_date	fp1_time	fp2_date	fp2_time	fp3_date
Id	Year	Race number	Id of circuit	Name of circuit	Date	Time	Wiki URL					

Results

resultId	raceId	driverId	constructorId	number	grid	position	positionText	positionOrder	points	laps
Id	Id of race	Id of Driver	Id of constructor	Car number	Position in stating grid	Final position	Final position in text	Final rank	Points	Laps

Seasons

year	url
Year	Wiki URL

Diccionario de variables

Sprint_results

resultId	raceId	driverId	constructorId	number	grid	position	positionText	positionOrder	points	laps
Id	Id of race	Id of driver	Id of constructor	Car number	Position in starting grid	Final position	Final position in text	Final rank	Points	Laps

Status

statusId	status
Id	Name

Diagrama entidad relación

Observaciones iniciales.

- 1 Las tablas **sprint_results** y **results** tienen la misma estructura
- 2 Se puede crear una sola tabla principal utilizando **drivers** y **races**
- 3 Una sola tabla principal utilizando **constructors** y **races**
- 4 Una sola tabla principal utilizando **drivers**, **constructors** y **races**



Lo que se busca del EDA

1

Uno de los eventos más significativos para la categoría fue la **muerte de Ayrton Senna** en 1994. ¿Existe alguna **tendencia** del valor **Fatal accident en Status** que se pueda relacionar con esto?

2

¿Existe una **relación** entre la posición de **salida (start_position)** y la **posición de llegada (end_position)** en las carreras?

3

¿Qué **pilotos y/o equipos** lograron el **mejor aprovechamiento** de la **posición de salida (start_position)**?

4

¿Podemos predecir quién va a **ganar independientemente del piloto** con base en las **características del auto** ?

Gracias

