

Formula 1 & Data Science with ORACLE

ORACLE: Reto Enseña 2.0

López Romero Diana
García Andrade Juan Carlos

Junio-2024

Análisis de Datos de los Puntajes de la F1



Introducción

Como primer punto, resolveremos las pregunta que se nos hicieron previamente:

¿Cuáles son las variables que el aficionado percibe que tienen un impacto positivo en el interés de una carrera de Fórmula 1?

Consideramos que la gran mayoría de los aficionados a este deporte se enfocan más en aquellos elementos que llegan a ser visuales; ya que son pocos los que llegan a tener un gran interés en partes más complejas que influyen en la carrera. Es por ello, que nosotros, basándonos en los datos que nos han proporcionado hemos decidido que aquellas variables que se pueden tomar en cuenta son:

- Los lugares del Gran Premio: El lugar donde se lleva a cabo la carrera tiene muchas implicaciones que pueden depender del gusto del fanático, desde el país que puede ser turísticamente atractivo, por la forma de la pista, o inclusive por el tipo de clima que le guste al fanático.
- El clima: La estructura de la parte de los carros para la F1 están hechos con la finalidad de que pueda sobrellevar los percances que pueda haber a lo largo de una carrera, sin embargo, hay cosas como la lluvia o fuertes vientos que pueden llegar a ser imprevistos que consigan alterar el funcionamiento de las llantas, la pista o ambas que no permitan una carrera óptima.
- Forma del circuito: Al haber distinción en los tipos de circuitos, hay algunas que por su estructura llegan a tener más curvas que otras, lo que aumenta las probabilidades de choques u otros tipos de accidentes lo cual propiciaría un mayor interés en la carrera.

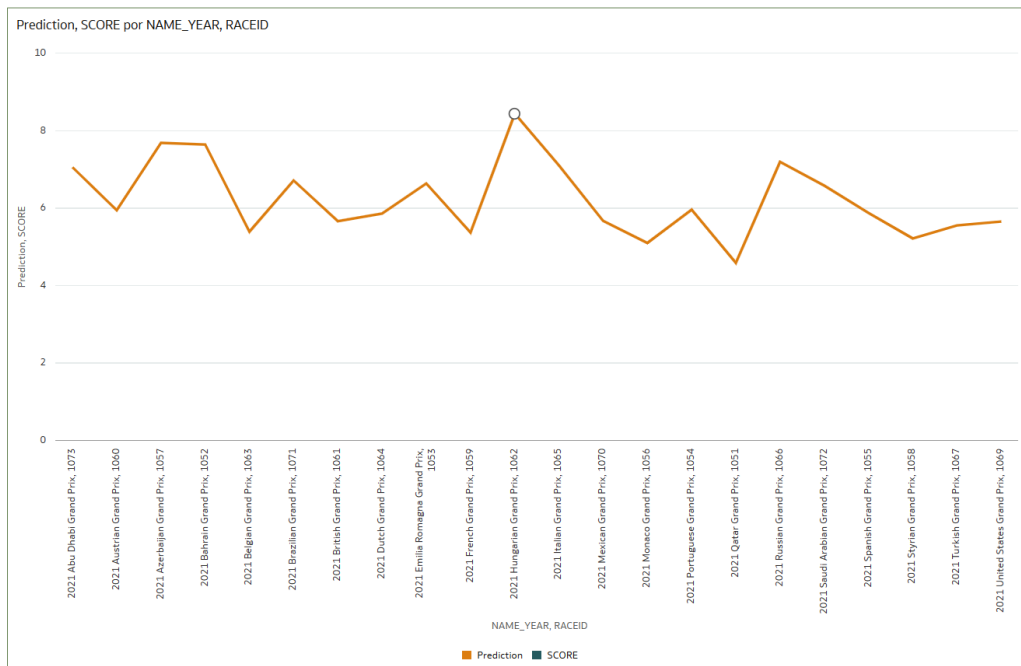
Sin embargo, a través de este análisis, descubriremos si es que este tipo de factores son en realidad motivos para que ciertas carreras en la F1 sean de mayor relevancia que otras.

¿Qué carrera de F1 puede ser la más interesante para el usuario el siguiente año?

Para 2021 : Gran Premio de Hungría

NAME_YEAR	Prediction ▼
2021 Hungarian Grand Prix	8,44
2021 Azerbaijan Grand Prix	7,68
2021 Bahrain Grand Prix	7,65
2021 Russian Grand Prix	7,20
2021 Italian Grand Prix	7,11
2021 Abu Dhabi Grand Prix	7,06

La justificación a la respuesta anterior se expresa en la gráfica, donde podemos ver la predicción que se realizó para el año 2021, en donde se puede ver que el pico de score más alto, señalado con un pequeño punto blanco:



Análisis Exploratorio

De primera instancia, si basamos nuestro análisis en el score(puntaje) de cada carrera, primero hemos de establecer de qué año son la mejor y la peor carrera. En las siguientes tablas podemos observar tanto los 10 mejores como los peores puntajes tanto como los 10 mejores.

NAME_YEAR, SCORE

NAME_YEAR	SCORE ▲
2020 Abu Dhabi Grand Prix	3,20
2010 German Grand Prix	3,74
2017 Abu Dhabi Grand Prix	3,74
2008 European Grand Prix	3,98
2018 Canadian Grand Prix	4,24
2019 Canadian Grand Prix	4,40
2020 Spanish Grand Prix	4,40
2008 Chinese Grand Prix	4,45
2020 Belgian Grand Prix	4,50
2015 Brazilian Grand Prix	4,57

NAME_YEAR, SCORE

NAME_YEAR	SCORE ▼
2012 Brazilian Grand Prix	9,45
2019 German Grand Prix	9,40
2011 Chinese Grand Prix	9,24
2014 Canadian Grand Prix	9,19
2014 Hungarian Grand Prix	9,14
2014 Bahrain Grand Prix	9,10
2015 United States Grand Prix	9,10
2011 Canadian Grand Prix	9,10
2020 Turkish Grand Prix	8,90
2012 Abu Dhabi Grand Prix	8,85

A continuación presentamos una tabla con la información de tales carreras. Y encontramos que como mejor tenemos a la carrera del año 2012 con específico Brazilian Grand Prix, y como peor tenemos la del 2020 Abu Dhabi Grand Prix.

Si vemos en la segunda columna, se muestra el score de cada una de estas, y las columnas siguientes presentan otra información característica de las mismas, tales como:

- El clima, tanto una descripción corta así como una variable binaria que representa si llovió o no.
- DNFs
- Adelantamientos de esa carrera
- La fecha de tal carrera.
- Un ID que realiza la relación con las otras tablas.

Si nos damos cuenta podemos ver que a pesar del puntaje hay otros puntajes en relevancia que podemos ver que es el:

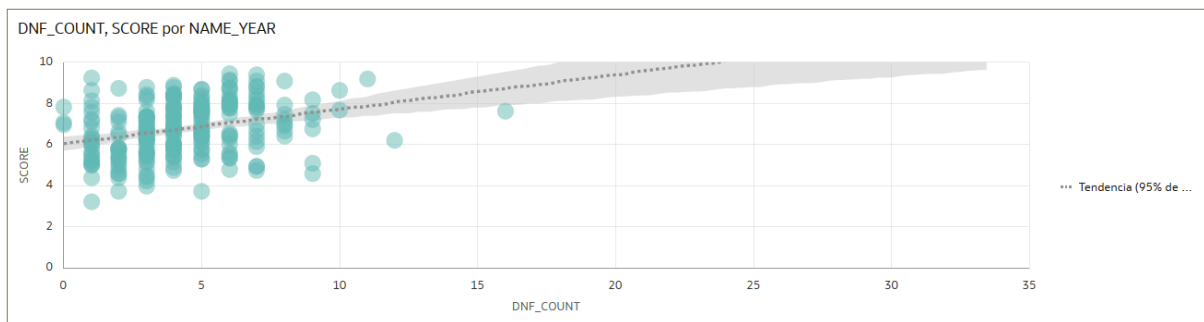
- DNF_COUNT(DC)
- DNF_DUE_TO_ACCIDENT_COUNT(DDTAC)
- OVERTAKEN_POSITIONS_TOTAL(OPT).

NAME_YEAR	SCORE	OVERTAKEN_POSITIONS_TOTAL	DNF_COUNT	DNF_DUE_TO_ACCIDENT_COUNT	WEATHER	WEATHER_WET	TIME	NAME	F1DATE	CIRCUITREF	RACEID
2012 Brazilian Grand Prix	9,45	367	6	6	drops of rain, light rain shower getting heavier at the end. [2] Air Temp 19 °C (66 °F)[2] Track Temp 21 °C (70 °F)[2]	Y		Brazilian Grand Prix	25/11/2012	interlagos	879
2020 Abu Dhabi Grand Prix	3,20	79	1	0	Clear	N		Abu Dhabi Grand Prix	13/12/2020	yas_marina	1.047

Primero haremos enfoque en los DNFs; como sabemos estos pueden implicar que hayan ocurrido accidentes por diferentes causas desde fallas mecánicas hasta choques, y como consecuencia hacen que el piloto ya no sea capaz de competir en la carrera. Sin embargo, hay una clasificación que nos hacen en esta tabla, donde los DDTAC son para aquellos accidentes que ocurren cuando no son causados por fallas mecánicas.

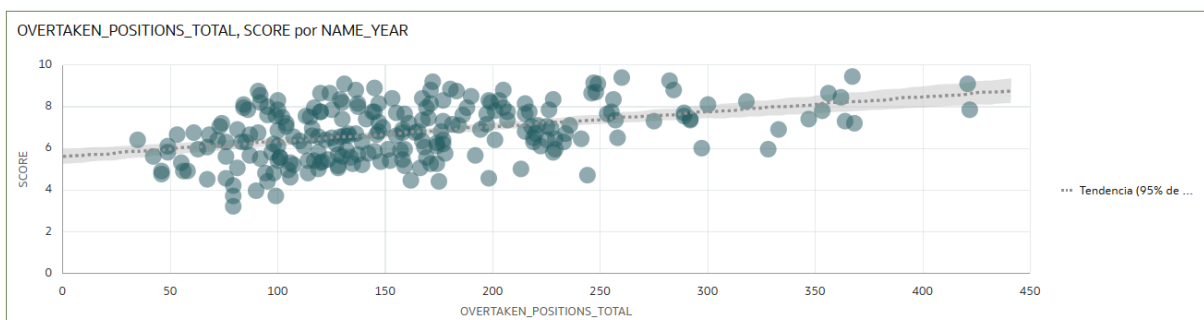
Como vemos en el año 2012, todos los DNFs registrados fueron accidentes que no tuvieron causa mecánica o de error humano, todos fueron accidentes.

Con referente a lo anterior, tenemos en nuestros registros que las carreras que se tienen abarcan desde el año de 1950 hasta 2021, y siguiendo con los DNF que señalan los accidentes; podemos percatarnos en la siguiente gráfica que en los primeros años, los accidentes eran muy frecuentes independientemente de la causa de estos, es por eso que deducimos que la F1 a lo largo de todo este tiempo ha desarrollado y evolucionado sus normas y reglas para que los accidentes sean menores, y por otro lado, consideramos que los equipos han evolucionado y mejorado sus carros con la finalidad de tener mejores desempeños con una alta seguridad.



Realizando un poco de investigación, en la carrera del Gran Premio de España 2012 en efecto hubo un gran accidente que marcó toda la carrera, debido a eso se consideraba que había pilotos que hacían que las carreras fueran más emocionantes, debido a sus formas de conducción. Y posteriormente, eso hizo que hubiera una gran disputa entre si el accidente fue realizado con intención o no, ya que esto provocó que los pilotos favoritos de ese momento se vieron afectados. Lo que agudizó las rivalidades de los para las posteriores carreras, y por ende la de mayor relevancia fue la final de ese año.

De igual forma, tenemos otra variable de relevancia OPT. Los overtaken, que conocemos como aquellos momentos en la carrera donde hay una disputa entre dos o más pilotos por obtener un mejor puesto. Como veremos a continuación hemos hecho una relación entre estos overtaken versus el score por cada carrera. Como podemos ver, aunque han ocurrido muchos sucesos de éste tipo a lo largo de los años, no obstante, en la gráfica se proyecta que en los últimos años, éste tipo de acontecimiento ha disminuido de una manera notable. Probablemente igual se deba al constante cambio de las normativas que se aplican para la carrera, ya que como sabemos están incluso cambian cada año, por lo que podríamos considerarlo como una hipótesis.



A continuación tenemos la siguiente gráfica donde se muestran las posiciones de cada piloto a lo largo de las carreras mejor(superior) y peor(inferior) valorada por los usuarios, de acuerdo a cada piloto(distinguido por colores). Por consiguiente, hemos generado la hipótesis de que la posición de inicio si afecta el desempeño del piloto durante todas las carreras de ese año; si observamos con detención, en ambos años los pilotos que han ocupado los primeros puestos; conservan estos, bajan de puesto(no mayor a 5 lugares de su posición inicial) o vuelven a sus posiciones iniciales. Por otro lado, si contemplamos debajo

de nuestra gráfica veremos los colores que representan a cada piloto, y se notan pocos colores dentro de ambas gráficas, ya que solo damos enfoque a los primeros 5 lugares (ya que como sabemos, solo hay tres lugares importantes al final de todo), por lo que vemos que pilotos con posiciones más bajas no pueden llegar a alcanzar los primeros lugares.

Por otra parte, pensamos que este hecho solo podemos dejarlo en hipótesis, de igual manera, ya que se necesitan otras variables y enfoques en los datos para poder analizar este hecho; tales como, algunas habilidades del piloto, su forma de conducción, el carro, las llantas del mismo, etc.

Por consiguiente al tener una mayor cantidad de cambio de posiciones en la carrera mayor será la aceptación que al mantener la mayoría de la posición durante toda la carrera.

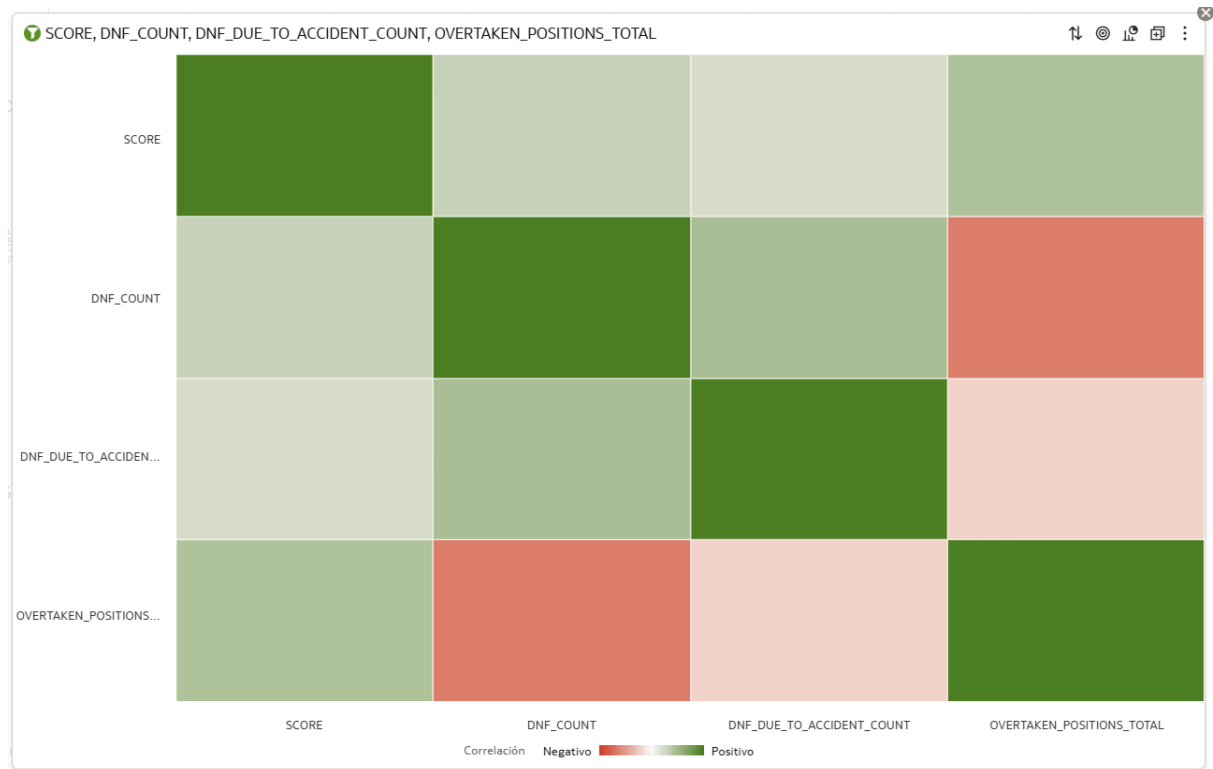


Finalmente se ha realizado una matriz de correlación, esto con el objetivo de poder conocer alguna posible relación entre algunas de las variables, para este caso hemos elegido las variables con las que hemos trabajado con anterioridad, para saber si estas variables (DNFs y OPT) se relacionan con el puntaje, es decir, si este tipo de eventos son los que más entretienen a los fanáticos, y de ser así pueden contemplarse normas, que puedan llegar a hacer las carreras seguras, sin que generen pérdida de emoción entre los fanáticos.

Reflexionando con respecto a la gráfica, podemos observar que el puntaje no tiene una relación dependiente lo suficientemente fuerte con las variables que hemos estado analizando. Tal podría ser el caso de que, OPT influyan sobre puntaje de manera un tanto relevante al igual que los DNF, independientemente del tipo de causa de este.

Podríamos generar la hipótesis de que, para algunos fanáticos si es emocionante el hecho de los overtake o inclusive los accidentes que puedan ocurrir durante una carrera, pero eso no significa que no haya otras variables que influyen con mayor intensidad.

Por otro lado, tenemos algunas relaciones negativas: OPT contra ambos DNFs, podríamos ver que los accidentes que ocurren en la pista no afecta de ninguna manera a los overtake, al contrario, podría ser el caso de que entre menos percances en la pista, genera un mejor desempeño en los pilotos, creando así momentos de tensión entre los mismos para obtener un mejor puesto.



Modelo de ML

Con las variables que tenemos en la tabla del archivo RACES, hemos decidido crear nuestro modelo. Por lo tanto, utilizando la plataforma de Oracle Cloud y Oracle Analytics, hemos optado por hacerlo de manera automática.

“Un modelo predictivo de Oracle Analytics aplica un algoritmo específico a un juego de datos para predecir valores o clases, o para identificar grupos en los datos.

También puede usar los modelos de aprendizaje automático de Oracle para realizar predicciones de los datos.

Oracle Analytics incluye algoritmos para ayudarle a entrenar modelos predictivos para diferentes fines. Algunos ejemplos de algoritmos son los árboles de clasificación y regresión (CART), la regresión logística y k-means.” (Fry et al., pág. 437)

Primero hemos entrenado nuestro modelo, para ello, la plataforma de Oracle nos ofrece una manera muy fácil de hacerlo. Como veremos a continuación, únicamente hemos de hacer selección de las variables que queremos predecir(lado derecho), qué es score; el tipo de modelo que queremos aplicar, en nuestro caso es regresión, y otras variables que como podemos ver depende de nuestro criterio.

The screenshot shows the Oracle Analytics model configuration interface. It includes fields for 'Origen de datos' (F1.RACES_TRAIN), 'Predecir' (SCORE), 'Tipo de predicción' (Regresión), and 'ID de caso' (RACEID). There is a 'Configuración adicional' section with a 'Restablecer' button and settings for 'Número máximo de modelos principales' (5), 'Duración máxima de ejecución (horas)' (8), 'Nivel de servicio de base de datos' (Bajas), and 'Métrica de modelo' (Error absoluto medio negativo).

Y debido a que hemos seleccionado modelos de regresión, podemos elegir cuales de los siguientes algoritmos queremos que se creen, nosotros hemos elegido todos, pues queremos el mejor de todas las opciones posibles:

The screenshot shows the 'Algoritmos' section of the Oracle Analytics interface. It lists five algorithms, all of which are selected with checkboxes: 'Modelo Lineal Generalizado', 'Modelo lineal generalizado (regresión Ridge)', 'Red neuronal', 'Máquina de vector de soporte (Gauss)', and 'Máquina de vector de soporte (lineal)'.

Como podemos ver en la siguiente imagen, se tomó al “Error Absoluto Medio Negativo” como métrica para poder evaluar la eficacia de cada modelos que se generaría de manera automática.

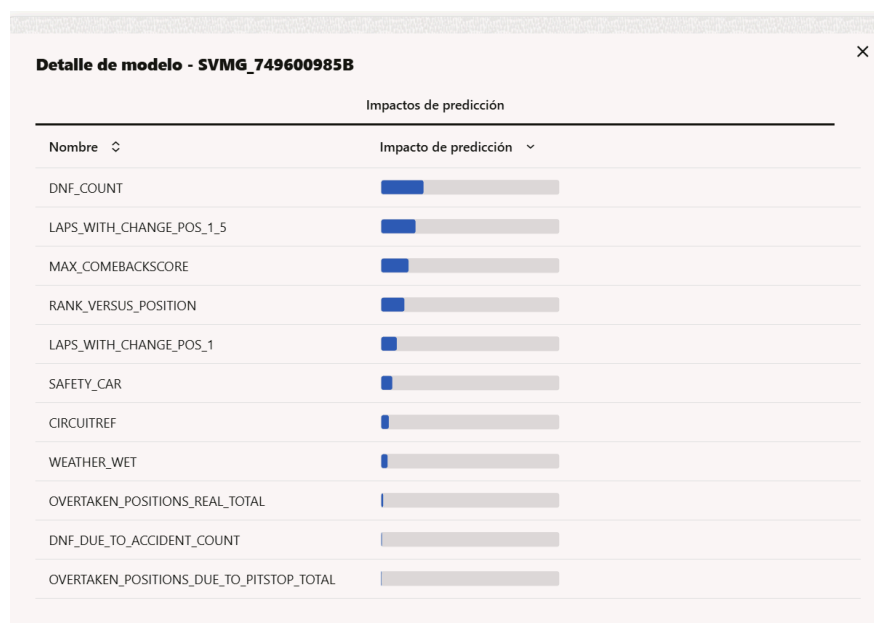
Al utilizar la herramienta de AutoML, se han generado los cinco modelos, y estos se han evaluado por igual con la métrica que escogimos con anterioridad . Esto quiere decir, que el modelo ya ha sido entrenado y que podemos aplicarlo a los futuros datos para poder hacer nuestras predicciones.

Hemos hecho la elección del modelo: SVMG_74600985B; ya que su valor en la métrica tiene el valor más bajo.

Clasificación		
Desplegar	Rename	Crear bloc de notas
Métricas		
Algoritmo	Nombre de modelo	Error absoluto medio negativo
Máquina de vector de soporte (Gauss)	SVMG_749600985B	0.6686
Modelo lineal generalizado (regresión Ridge)	GLMR_7AA2C863D6	0.6717
Modelo Lineal Generalizado	GLM_2130C3E73D	0.6719
Red neuronal	NN_BCBA6B6C50	0.6736
Máquina de vector de soporte (lineal)	SVML_76EBFCDE65	0.7737

Ya elegido, ahora solo debemos guardarlo para poder utilizarlo posteriormente.

Aquí podemos distinguir cuales son las variables que tienen un mayor impacto en nuestro modelo, y que por ende, son las que afectan de una manera significativa a nuestra variable score.



Debido a que decidimos calcular la predicción de los puntajes para el año 2022, creímos que bastarían los datos que teníamos para predecir ambos años. Sin embargo, después de establecer el filtro de la predicción de 2021 a 2022, al momento de graficar no nos aparecía ninguna predicción para 2022.

Así que nos dimos a la tarea de hacer los registros para el año 2022:

	A	B	C	D	E	F	G	H	I	J	
1	RACEID	YEAR	ROUND	NAME	F1DATE	TIME	URL	SCORE	DNF_COUNT	DNF_DUE_TO_	WEATHER
2	1095	2022	22	Abu Dhabi Grand Prix	20/11/2022		https://en.wikipedia.org/wiki/2022_Abu_Dhabi_Grand_Prix		3	0	Clear
3	1094	2022	21	Brazilian Grand Prix	13/11/2022		https://en.wikipedia.org/wiki/2022_S%C3%A3o_Paulo_Grand_Prix		1	0	Cloudy
4	1093	2022	20	Mexican Grand Prix	30/10/2022		https://en.wikipedia.org/wiki/2022_Mexico_City_Grand_Prix		2	1	Partly cloudy
5	1092	2022	19	United States Grand Prix	23/10/2022		https://en.wikipedia.org/wiki/2022_United_States_Grand_Prix		3	1	Partly cloudy
6	1091	2022	18	Japanese Grand Prix	9/10/2022		https://en.wikipedia.org/wiki/2022_Japanese_Grand_Prix		2	2	Rain
7	1090	2022	17	Singapore Grand Prix	2/10/2022		https://en.wikipedia.org/wiki/2022_Singapore_Grand_Prix		6	3	Partly cloudy in a wet
8	1089	2022	16	Italian Grand Prix	11/9/2022		https://en.wikipedia.org/wiki/2022_Italian_Grand_Prix		4	0	Sunny
9	1088	2022	15	Dutch Grand Prix	4/9/2022		https://en.wikipedia.org/wiki/2022_Dutch_Grand_Prix		2	0	Partly cloudy
10	1087	2022	14	Belgian Grand Prix	28/8/2022		https://en.wikipedia.org/wiki/2022_Belgian_Grand_Prix		2	2	Sunny
11	1086	2022	13	Hungarian Grand Prix	31/7/2022		https://en.wikipedia.org/wiki/2022_Hungarian_Grand_Prix		1	0	Over cast with drizzle
12	1085	2022	12	French Grand Prix	24/7/2022		https://en.wikipedia.org/wiki/2022_French_Grand_Prix		5	3	Sunny
13	1084	2022	11	Austrian Grand Prix	10/7/2022		https://en.wikipedia.org/wiki/2022_Austrian_Grand_Prix		3	1	Partly cloudy

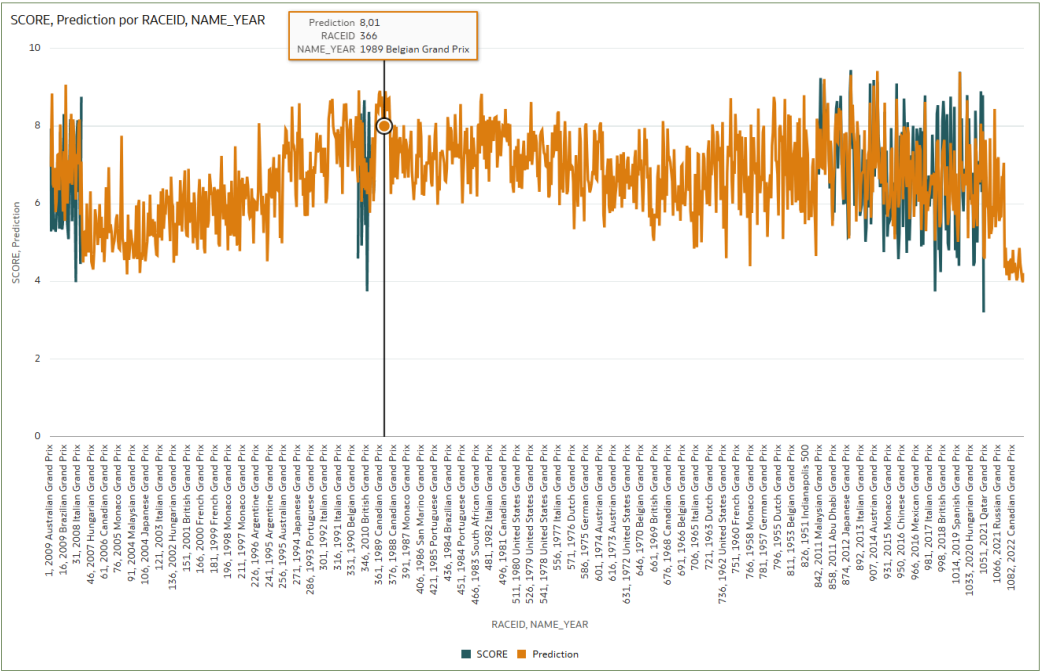
Desgraciadamente no pudimos obtener toda la información de los registros, nos faltaron variables como TIME y OVERTAKE_POSITIONS_TOTAL. Y esto debido a que no encontramos el significado de la variable TIME, ni encontramos la fuente de esta información. A pesar de ello, fuimos capaces de obtener una predicción estimada con los datos que obtuvimos, ya que como vimos en la imagen anterior, estas variables no tienen gran relevancia para la estimación del puntaje, lo cual se verá en la siguiente sección.

RESULTADOS

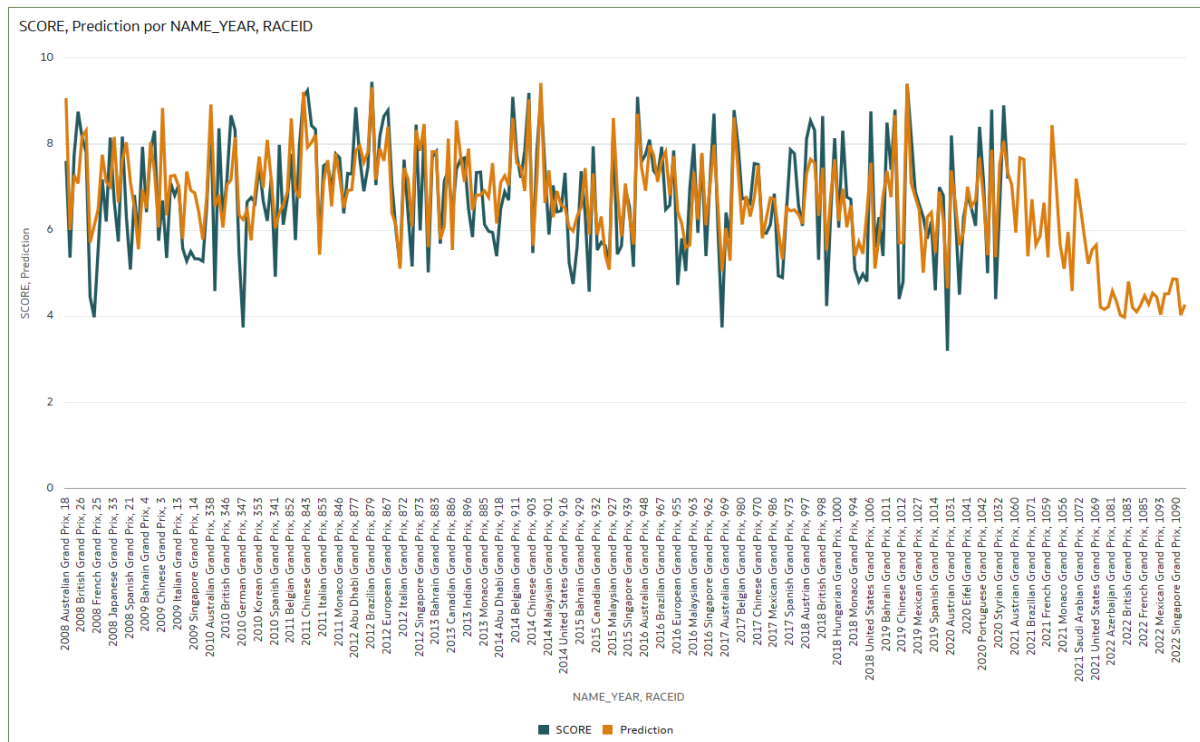
Para poder saber si los puntajes predictivos son similares a los que ya se tenían registrados, creamos las siguientes tablas en donde comparamos estos últimos, de diferentes años:

NAME_YEAR	SCORE	Prediction	NAME_YEAR	SCORE	Prediction	NAME_YEAR	SCORE	Prediction
2008 Australian Grand Prix	7,61	9,07	2009 European Grand Prix	5,36	6,35	2013 Italian Grand Prix	5,84	6,47
2008 Bahrain Grand Prix	5,36	6,00	2009 German Grand Prix	7,10	7,25	2013 Japanese Grand Prix	7,34	6,82
2008 Belgian Grand Prix	7,74	7,28	2009 Hungarian Grand Prix	6,81	7,27	2013 Korean Grand Prix	7,35	6,82
2008 Brazilian Grand Prix	8,76	7,09	2009 Italian Grand Prix	7,05	7,02	2013 Monaco Grand Prix	6,13	6,92
2008 British Grand Prix	8,16	8,17	2009 Japanese Grand Prix	5,58	5,80	2013 Singapore Grand Prix	5,97	6,77
2008 Canadian Grand Prix	7,81	8,32	2009 Malaysian Grand Prix	5,28	7,37	2013 Spanish Grand Prix	5,95	7,56
2008 Chinese Grand Prix	4,45	5,71	2009 Monaco Grand Prix	5,50	6,93	2013 United States Grand Prix	5,40	6,16
2008 European Grand Prix	3,98	6,10	2009 Singapore Grand Prix	5,34	6,86	2014 Abu Dhabi Grand Prix	6,49	7,10
2008 French Grand Prix	5,55	6,49	2009 Spanish Grand Prix	5,33	6,42	2014 Australian Grand Prix	6,89	7,26
2008 German Grand Prix	7,18	7,75	2009 Turkish Grand Prix	5,28	5,78	2014 Austrian Grand Prix	6,70	7,04
2008 Hungarian Grand Prix	6,20	7,12	2010 Abu Dhabi Grand Prix	6,60	6,79	2014 Bahrain Grand Prix	9,10	8,61
2008 Italian Grand Prix	8,15	7,01	2010 Australian Grand Prix	8,64	8,92	2014 Belgian Grand Prix	7,76	7,56
2008 Japanese Grand Prix	6,66	8,17	2010 Bahrain Grand Prix	4,59	6,55	2014 Brazilian Grand Prix	7,22	7,47
2008 Malaysian Grand Prix	5,74	6,65	2010 Belgian Grand Prix	8,37	6,82	2014 British Grand Prix	7,85	6,91
2008 Monaco Grand Prix	8,18	7,63	2010 Brazilian Grand Prix	6,25	6,06	2014 Canadian Grand Prix	9,19	9,04
2008 Singapore Grand Prix	6,30	8,04	2010 British Grand Prix	7,20	7,06	2014 Chinese Grand Prix	5,47	5,64
2008 Spanish Grand Prix	5,09	7,13	2010 Canadian Grand Prix	8,67	7,17	2014 German Grand Prix	7,86	7,20
2008 Turkish Grand Prix	6,82	6,52	2010 Chinese Grand Prix	8,33	8,16	2014 Hungarian Grand Prix	9,14	9,42
2009 Abu Dhabi Grand Prix	5,79	5,56	2010 European Grand Prix	5,45	6,35	2014 Italian Grand Prix	7,42	6,63
2009 Australian Grand Prix	7,94	6,95	2010 German Grand Prix	3,74	6,25	2014 Malaysian Grand Prix	5,90	7,39
2009 Bahrain Grand Prix	6,42	6,50	2010 Hungarian Grand Prix	6,65	6,47	2014 Monaco Grand Prix	7,04	6,30
2009 Belgian Grand Prix	7,85	8,05	2010 Italian Grand Prix	6,76	5,76	2014 Singapore Grand Prix	6,43	6,91
2009 Brazilian Grand Prix	8,31	7,35	2010 Japanese Grand Prix	6,64	6,89	2014 Spanish Grand Prix	6,45	6,58
2009 British Grand Prix	5,76	6,07	2010 Korean Grand Prix	7,54	7,71	2014 United States Grand Prix	7,33	6,52
2009 Chinese Grand Prix	6,69	8,84	2010 Malaysian Grand Prix	6,68	6,98	2015 Abu Dhabi Grand Prix	5,25	6,07
2009 European Grand Prix	5,36	6,35	2010 Monaco Grand Prix	6,22	8,10	2015 Australian Grand Prix	4,75	5,98

A pesar de tener todos estos registros, no podemos checar uno por uno, si tenemos en cuenta que van desde 1950; si no que pensamos que es mejor verlo a través de una gráfica:



Debido a que tenemos muchos registros sin score registrado no es muy posible notar la relación entre estos, es por ello que en la siguiente gráfica decidimos hacer un filtro, donde solo visualizamos desde el año 2008 en donde ya hay registros del score de la carrera para ver si podemos tener un mejor enfoque:

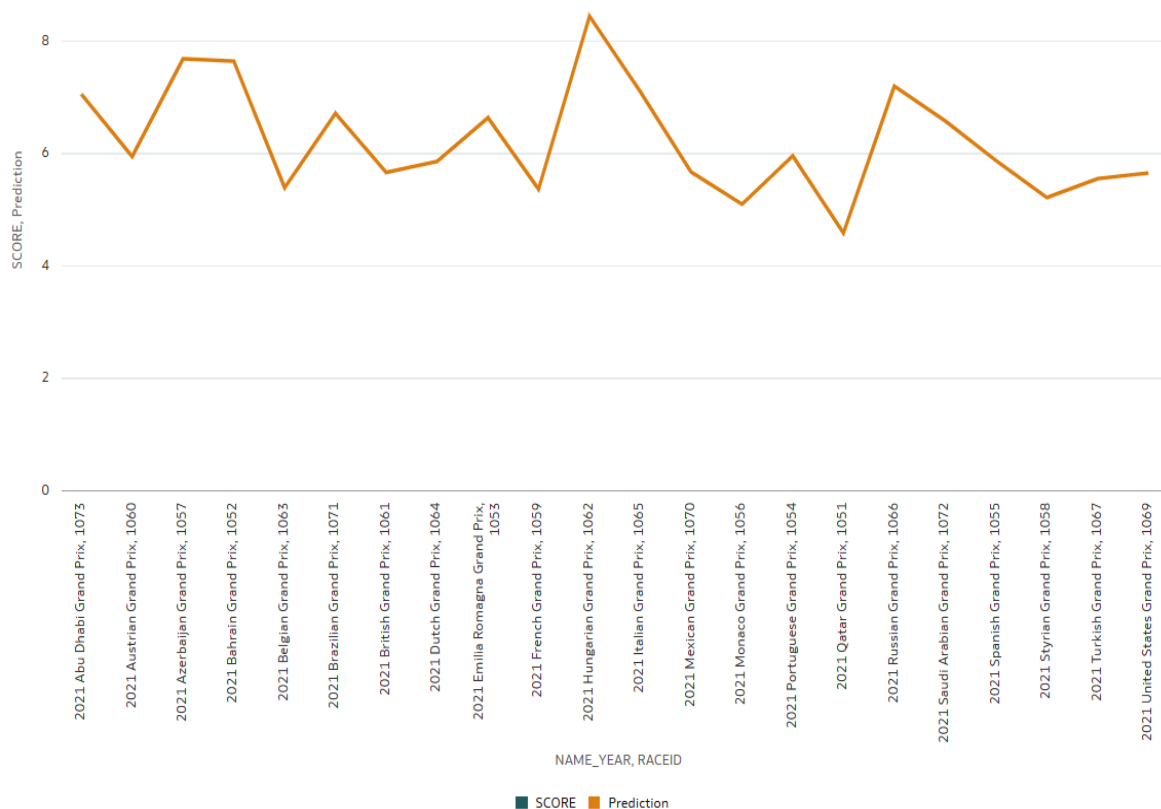


De esta forma podremos ver de manera más detallada que el modelo elegido fue una buena opción o todo lo contrario. Ya aquí, somos capaces de observar una precisión bastante buena comparada con la real, sabemos que no es posible obtener un modelo que genere de manera idéntica todos los valores, pero consideramos que el modelo escogido tiene una buena proximidad a estos, lo que nos permite confiar en las predicciones para los años sugeridos. Observando detenidamente, nos daremos cuenta que al final de la gráfica la línea naranja que representa las predicciones, es más larga que la azul que muestra los valores que ya tenemos. Y esto es porque aquí ya estamos contemplando las predicciones pertenecientes a los años 2021 y 2022.

Ahora bien, podemos analizar detenidamente la predicción para el año 2021. De primera instancia veremos todas las predicciones para ese año:

NAME_YEAR	Prediction ▼
2021 Hungarian Grand Prix	8,44
2021 Azerbaijan Grand Prix	7,68
2021 Bahrain Grand Prix	7,65
2021 Russian Grand Prix	7,20
2021 Italian Grand Prix	7,11
2021 Abu Dhabi Grand Prix	7,06
2021 Brazilian Grand Prix	6,72
2021 Emilia Romagna Grand Prix	6,64
2021 Saudi Arabian Grand Prix	6,58
2021 Portuguese Grand Prix	5,96
2021 Austrian Grand Prix	5,94
2021 Spanish Grand Prix	5,88
2021 Dutch Grand Prix	5,86
2021 Mexican Grand Prix	5,67
2021 British Grand Prix	5,66
2021 United States Grand Prix	5,65
2021 Turkish Grand Prix	5,55
2021 Belgian Grand Prix	5,39
2021 French Grand Prix	5,37
2021 Styrian Grand Prix	5,22
2021 Monaco Grand Prix	5,10
2021 Qatar Grand Prix	4,59

De manera gráfica lo tenemos de la siguiente manera:

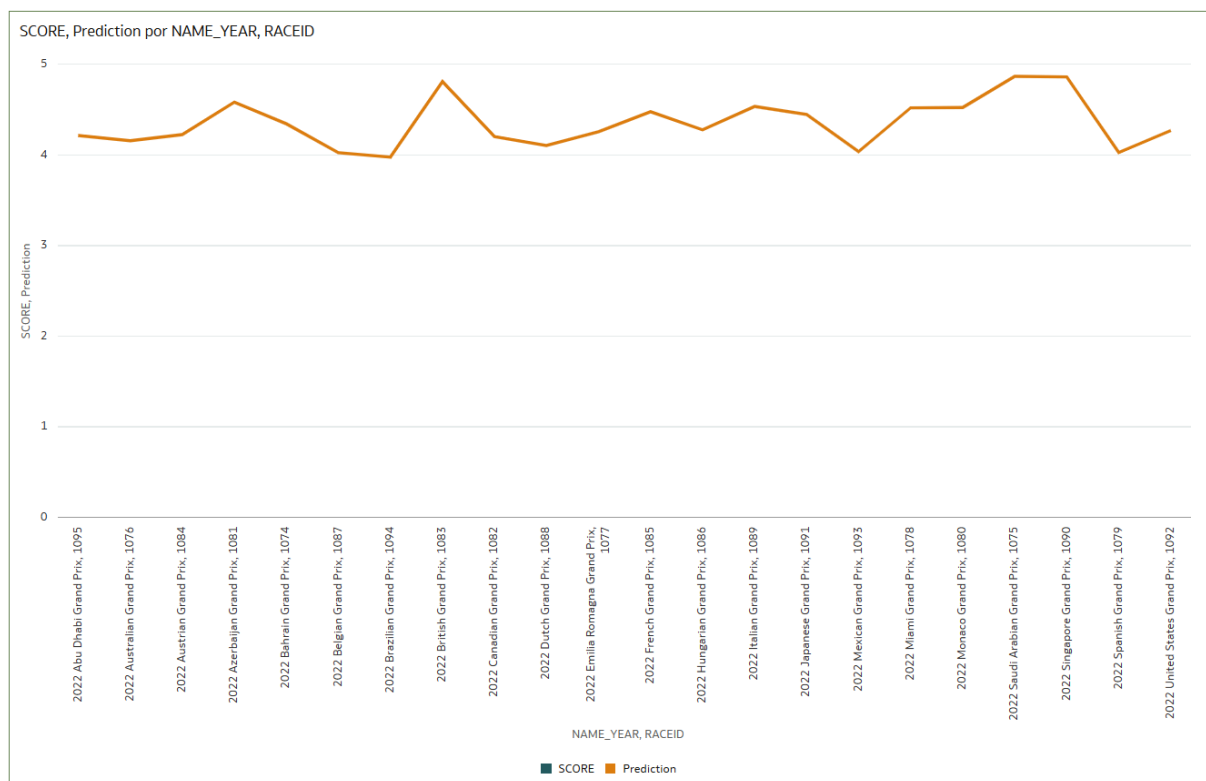


Donde observamos que el punto más alto se encuentra en el Gran Premio de Hungría, y por lo tanto concluimos que para ese año, esa fue la carrera más relevante.

Para el siguiente año(2022), igualmente tenemos una tabla con todos los puntajes de las carreras para ese año:

NAME_YEAR	Prediction ▼
2022 Saudi Arabian Grand Prix	4,87
2022 Singapore Grand Prix	4,86
2022 British Grand Prix	4,81
2022 Azerbaijan Grand Prix	4,58
2022 Italian Grand Prix	4,54
2022 Monaco Grand Prix	4,52
2022 Miami Grand Prix	4,52
2022 French Grand Prix	4,48
2022 Japanese Grand Prix	4,45
2022 Bahrain Grand Prix	4,34
2022 Hungarian Grand Prix	4,28
2022 United States Grand Prix	4,27
2022 Emilia Romagna Grand Prix	4,26
2022 Austrian Grand Prix	4,22
2022 Abu Dhabi Grand Prix	4,22
2022 Canadian Grand Prix	4,20
2022 Australian Grand Prix	4,16
2022 Dutch Grand Prix	4,11
2022 Mexican Grand Prix	4,04
2022 Spanish Grand Prix	4,03
2022 Belgian Grand Prix	4,02
2022 Brazilian Grand Prix	3,98

Puede que de esta manera nos tardemos un poco más en localizar dónde se encuentra el puntaje más alto es por ello que hemos graficado la predicción de la siguiente manera:



Como se muestra, el pico más alto que tenemos se encuentra en el Gran Premio de Arabia Saudita. Aunque los puntajes son muy similares, los decimales nos demuestran que aún así, existe una gran diferencia entre ellos.

Conclusiones

Análisis F1

A lo largo de este análisis pudimos trabajar con diversos factores que se involucran en la carrera de la F1. Si bien, nuestro enfoque fue dirigido hacia el score, pudimos lograr un modelo que generaba valores lo suficientemente parecidos a los que se nos otorgaron inicialmente, y aunque sí había cierta diferencia entre estos, no era demasiado grande para decir que el modelo no funciona o no sea el más óptimo en nuestro caso; aunque tampoco descartamos la idea de que pueda haber un modelo todavía más cercano.

Así mismo, descartamos la idea de que a los fanáticos les gusta ver momentos trágicos durante la carrera, no decimos que no sean un factor que genere un mayor notoriedad; porque probablemente lo que sea relevante de la carrera sea solo ese momento, pero no es suficiente para atraer a los verdaderos fanáticos. Inicialmente creímos que sería todo lo contrario, pero lo descartamos en cuanto vimos la relación que tenían los DNFs con la variable score.

Adicional

Con este trabajo se nos ocurrió la idea de que podríamos llegar a crear un modelo que nos ayude a elegir, dependiendo de diferentes factores, un piloto que sea probable sea el primer lugar en la carrera, sin embargo, necesitaríamos variables diferentes y algunas de las que ya tenemos aquí. Algunas de las variables que consideramos podrían afectar son :

- La posición de salida.
- El tiempo tomado por parada en pits.
- El lugar y la pista de la carrera.
- El clima de ese día.
- Los puntos acumulados.

Estas variables las encontramos en algunas tablas de los archivos que nos fueron entregados , pero no son los únicos que creemos que afectarían:

- El tipo de pista: Algunos pilotos podrían ser mejores en pistas con más curvas u otros en pistas más rectas.
- Habilidades/Características del piloto: Es cierto que hay muchos factores que dependen para la carrera pero si no hay un buen piloto no hay éxito asegurado. Además al haber corredores de diferentes países sabemos que hay cosas como la altura y el peso que son variables muy diferentes unos de otros.

- Características de los carros: Señalamos el diseño de estos, el tipo de llantas, tipo de motor; e inclusive hasta el equipo de los mismos.

Consideramos que sería un gran análisis, pero por el momento solo se quedará como una idea para el futuro.

Trabajo/Experiencia

A través de este análisis nos hemos encontrado con diversos obstáculos que nos han dificultado la exploración de los datos, pero por otro lado hemos conocido una herramienta de gran utilidad y poder, que consideramos tiene potencial para el futuro, si no es que ya es necesario para nuestro presente.

En primer lugar, al recibir los datos nos encontramos con el percance de que no contábamos con la descripción de los mismos. En otras palabras, no contamos con un informe que nos describiera los mismos datos. Y aunque la mayoría de los datos no eran difíciles de descifrar, hubo otros que no supimos que señalaban con exactitud.

Otra parte fundamental de los datos es el origen de los mismos. Para nuestra variable de importancia score, no concuerda con los puntajes que se encuentran publicados en páginas deportivas o incluso de fanáticos de este deporte. Es por ello, que no logramos encontrar el lugar de donde se obtuvieron otros datos y de todos los relacionados a ellos. Incluso tenemos registros de tiempo, nosotros supusimos que se trataba de la hora del inicio de la carrera, pero ese no fue el caso, y su nombre registrado tampoco nos fue de utilidad, por lo que tampoco pudimos saber si tal dato era relevante o no.

Por otro lado, tuvimos el placer de conocer y usar una nueva plataforma. La parte de la visualización es muy similar a su forma de uso como es el caso de Power BI, sin embargo, a diferencia de ese recurso, nosotros tenemos conectados la parte de otro recurso poderoso como es la implementación del aprendizaje automático. Y una gran ventaja de ello es que puede hacerse de manera automática o no, tanto la división de los datos para entrenamiento y prueba, así como la aplicación del mismo para predicciones futuras. De igual manera la mejor parte de esto es que la misma plataforma te ofrece la mejores opciones de modelos que se pueden realizar en base a una métrica que puede elegir el usuario. Y si por el contrario se desea hacer el modelo de manera manual, la plataforma igual ofrece un espacio para ello en donde puedes hacer uso del lenguaje de tu preferencia.

Además en internet puede encontrarse materiales tanto en inglés como en español para hacer uso de su plataforma, que publicados oficialmente por Oracle; sin embargo, considero que podría ayudar material audiovisual, pero eso podemos considerarlo como un gusto personal.

Bibliografías

https://en.wikipedia.org/wiki/2022_Abu_Dhabi_Grand_Prix
https://en.wikipedia.org/wiki/2022_S%C3%A3o_Paulo_Grand_Prix
https://en.wikipedia.org/wiki/2022_Mexico_City_Grand_Prix
https://en.wikipedia.org/wiki/2022_United_States_Grand_Prix
https://en.wikipedia.org/wiki/2022_Japanese_Grand_Prix
https://en.wikipedia.org/wiki/2022_Singapore_Grand_Prix
https://en.wikipedia.org/wiki/2022_Italian_Grand_Prix
https://en.wikipedia.org/wiki/2022_Dutch_Grand_Prix
https://en.wikipedia.org/wiki/2022_Belgian_Grand_Prix
https://en.wikipedia.org/wiki/2022_Hungarian_Grand_Prix
https://en.wikipedia.org/wiki/2022_French_Grand_Prix
https://en.wikipedia.org/wiki/2022_Austrian_Grand_Prix
https://en.wikipedia.org/wiki/2022_British_Grand_Prix
https://en.wikipedia.org/wiki/2022_Canadian_Grand_Prix
https://en.wikipedia.org/wiki/2022_Azerbaijan_Grand_Prix
https://en.wikipedia.org/wiki/2022_Monaco_Grand_Prix
https://en.wikipedia.org/wiki/2022_Spanish_Grand_Prix
https://en.wikipedia.org/wiki/2022_Miami_Grand_Prix
https://en.wikipedia.org/wiki/2022_Emilia_Romagna_Grand_Prix
https://en.wikipedia.org/wiki/2022_Australian_Grand_Prix
https://en.wikipedia.org/wiki/2022_Saudi_Arabian_Grand_Prix
https://en.wikipedia.org/wiki/2022_Bahrain_Grand_Prix

Fry N., Brownbridge P., Harvey R., Gill S., Rhone S. (2017) *Oracle Cloud Visualización de datos y generación de informes en Oracle Analytics Cloud* . (2024) Oracle.