

OBJETIVOS Y PLANTEAMENTO INCIAL

"Predecir victoria, empate o derrota de un equipo contra otro con base en antecedentes históricos y estadísticas relevantes".





¿RELEVANTES?









NU.		EUUIPU	IJ	JG	JŁ	JP	6F	GL.	UII	PIS
1	0	TOLUCA	11	7	1	3	16	11	5	22
2	•	GUADALAJARA	11	6	3	2	16	10	6	21
3		MONTERREY	11	5	4	2	17	11	6	19
4	0	TIJUANA	11	5	3	3	24	18	6	18
5	Û	PACHUCA	11	5	3	3	11	10	1	18
6		PUMAS	11	5	2	4	20	17	3	17
7	(3)	AMÉRICA	11	5	2	4	13	12	1	17
8	8	SANTOS	11	3	7	1	15	12	3	16





VARIABLES PROPUESTAS

Resultados recientes Local/Visitante Factores geográficos Factores económicos Posiciones en la tabla

NOCUMINACIONES

- Escándalos
- Directiva
- Medios
- Química
- Rendimiento/situaciones personales
- Estado anímico













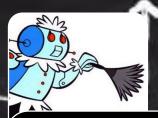
Recopilación de datos

- Búsqueda en Internet
- Determinar forma de descarga
- Acomodo en Google Sheets



Preprocesamiento

- Homogeneización
- Uso de Querys.



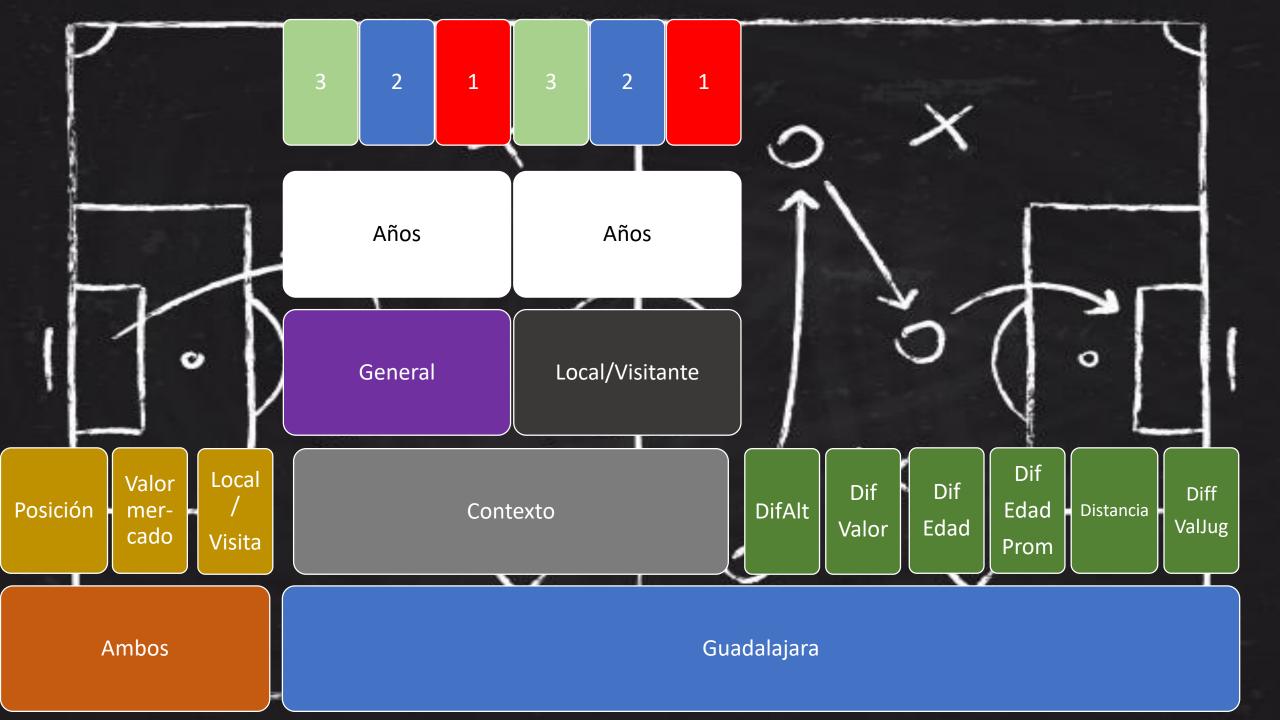
Limpieza de datos

- Manejo de datos faltantes
- Variables categóricas
- Notaciones inadecuadas



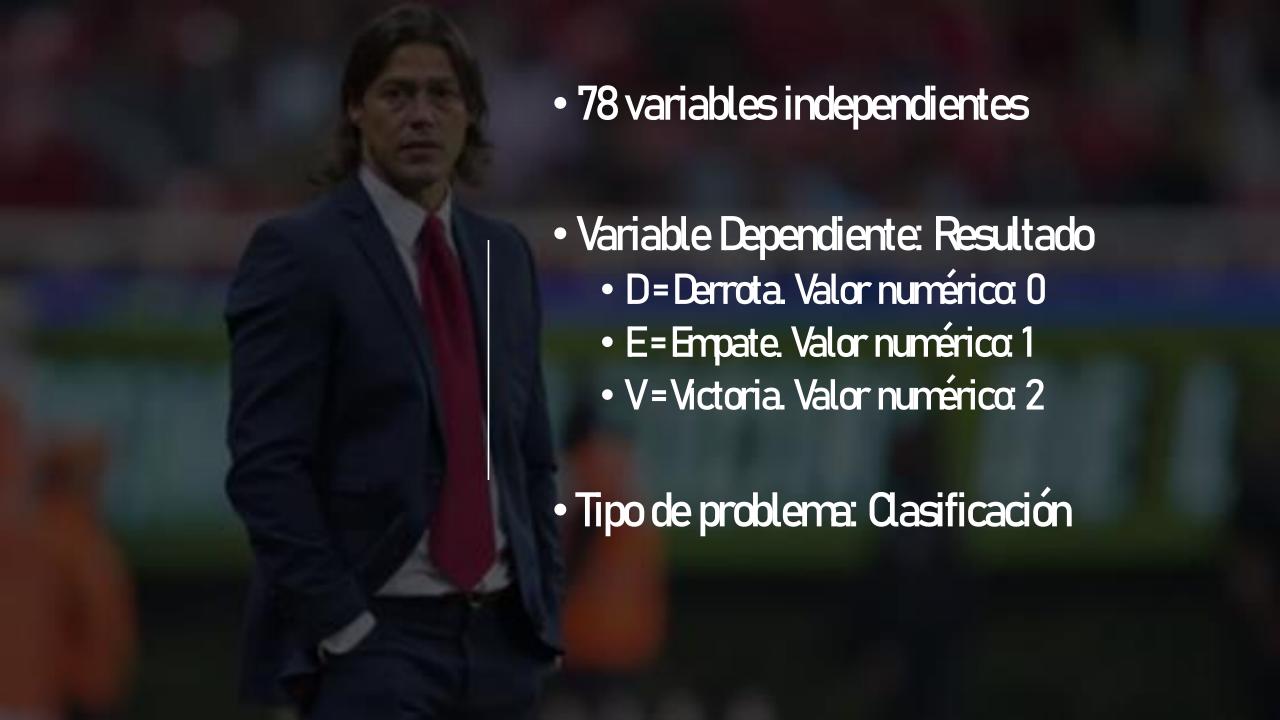
Evaluación

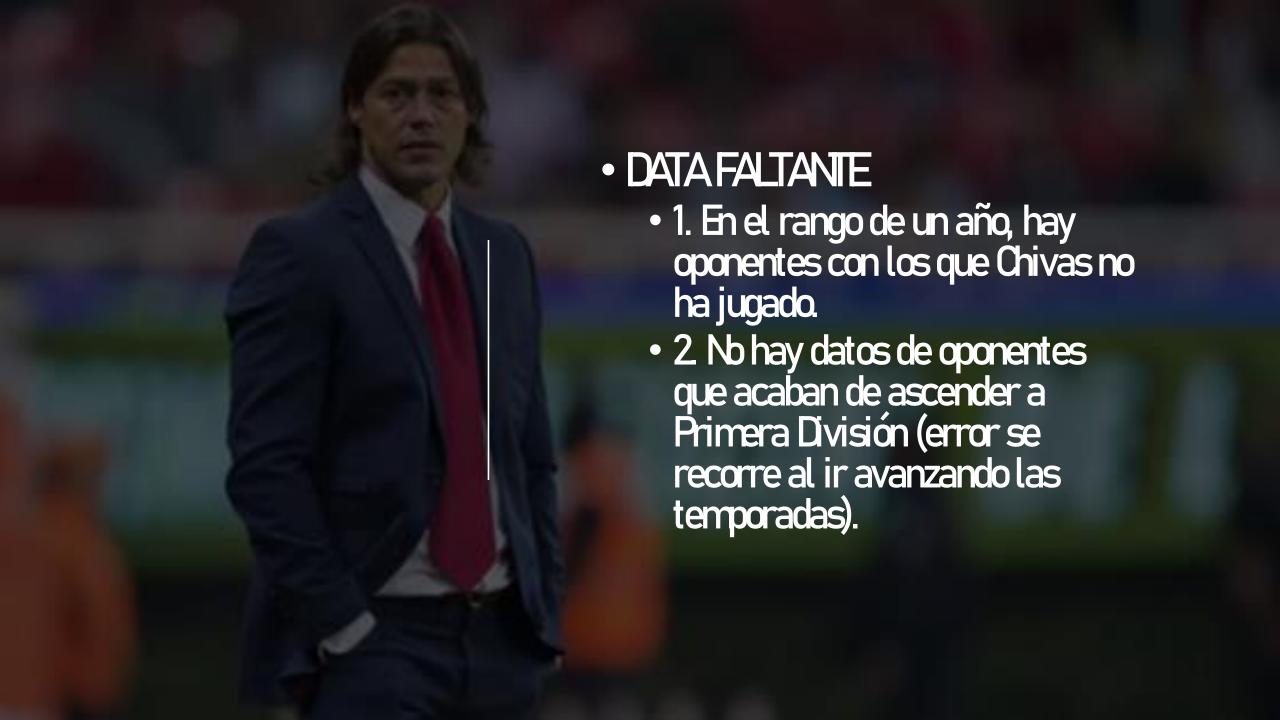
- Importar datos.
- Limpieza de datos en programa
- Clasificadores e histogramas



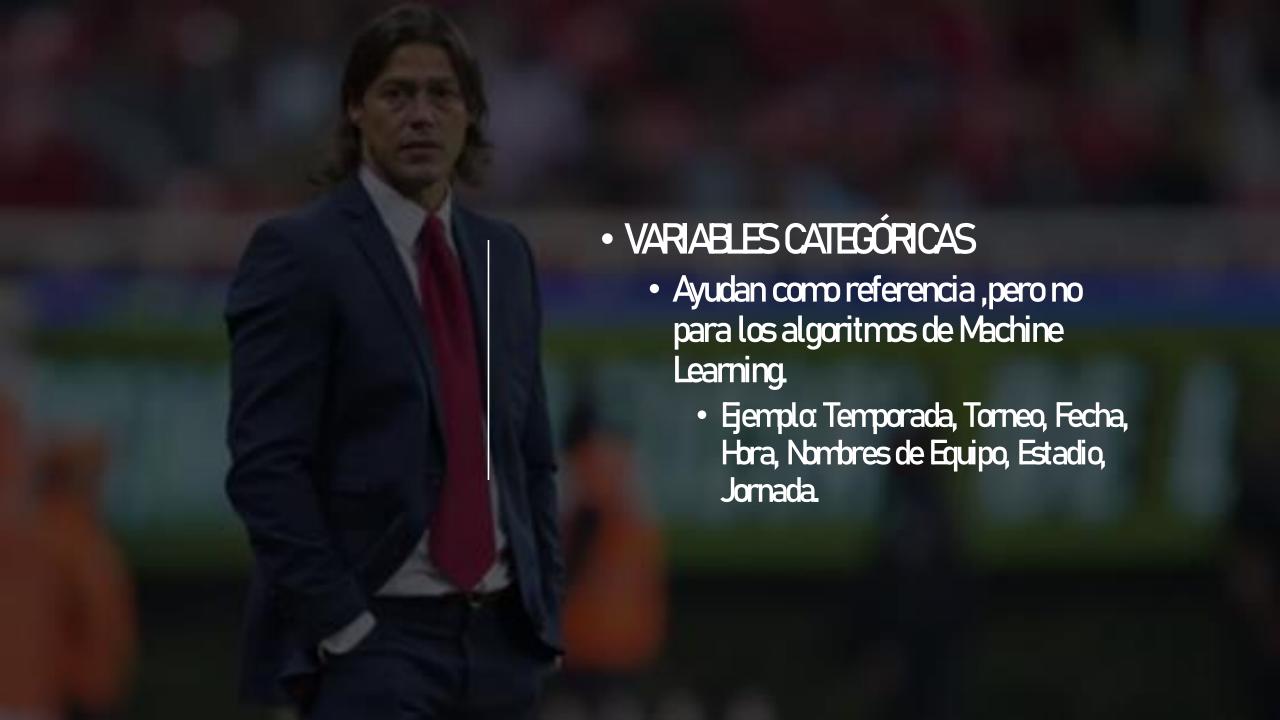
3 años 1 año

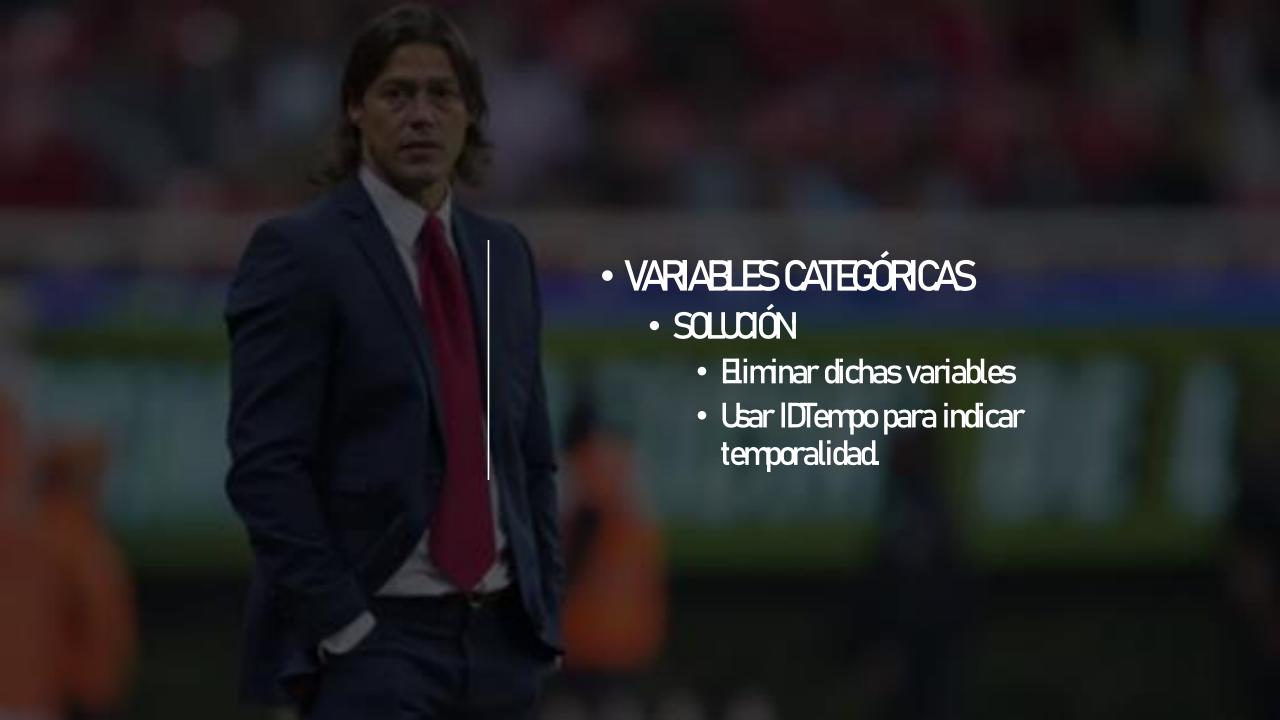
2 años

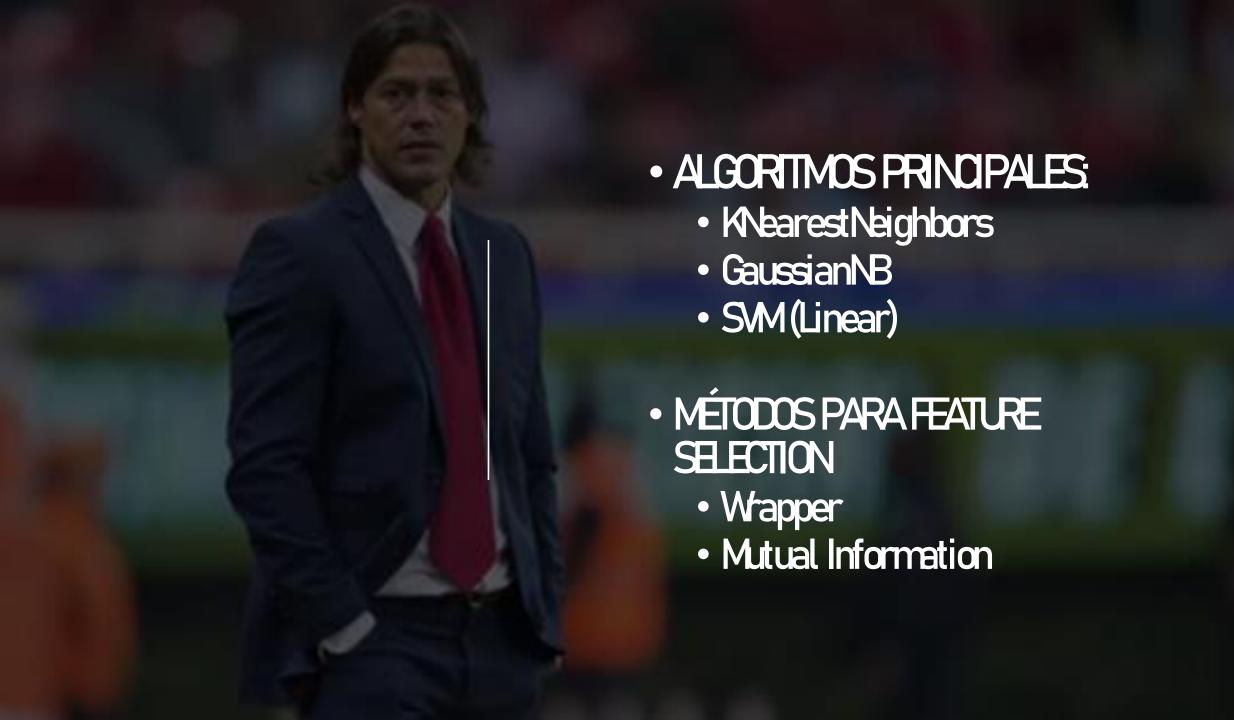


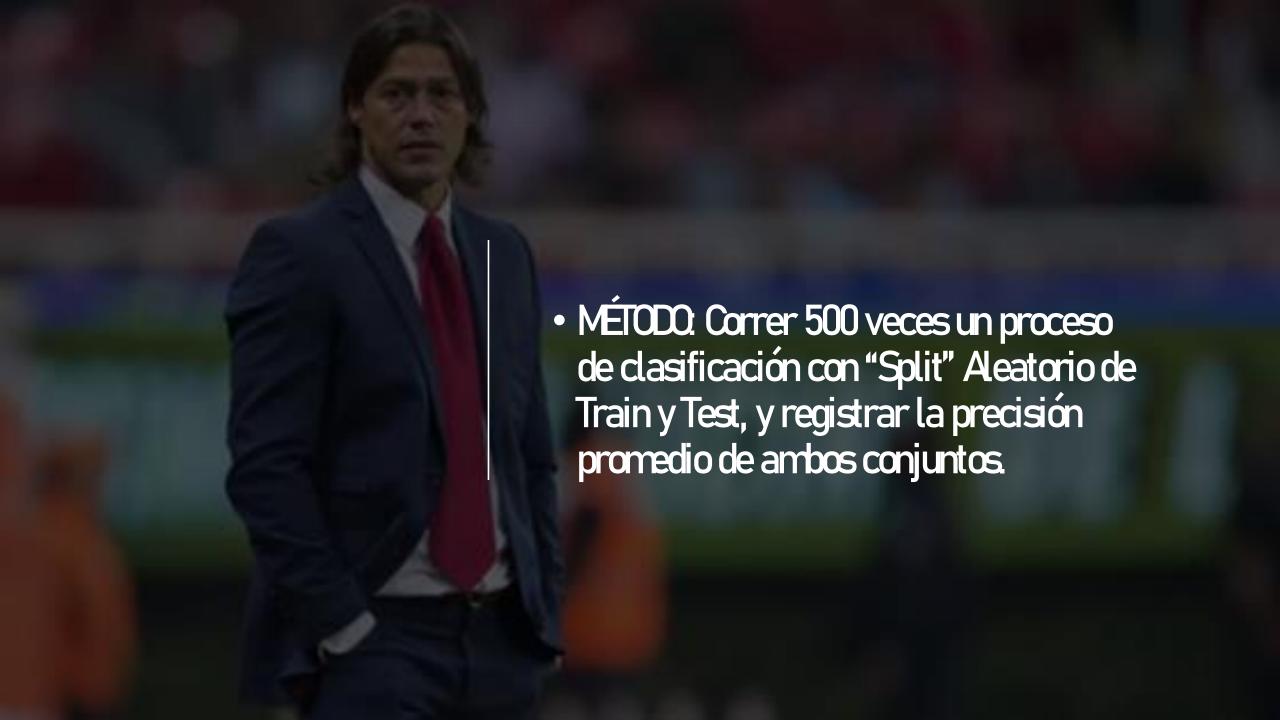


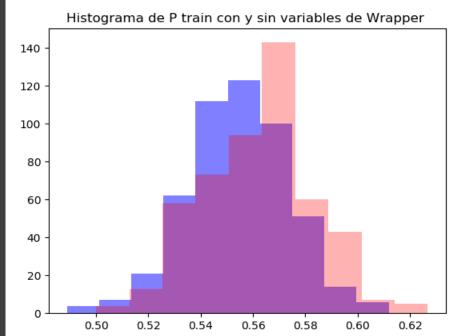


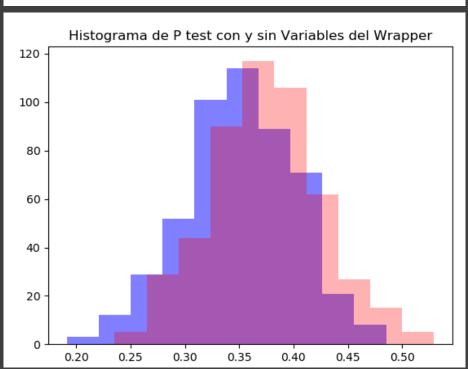




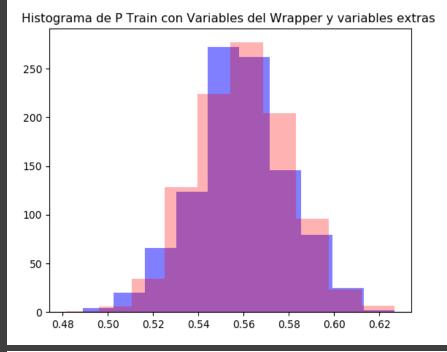


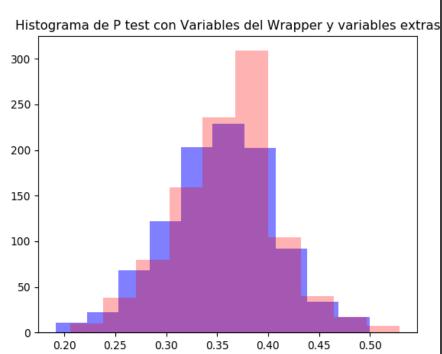




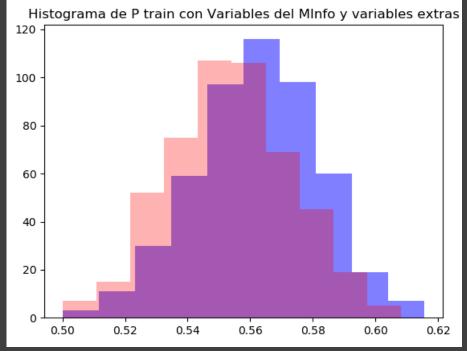


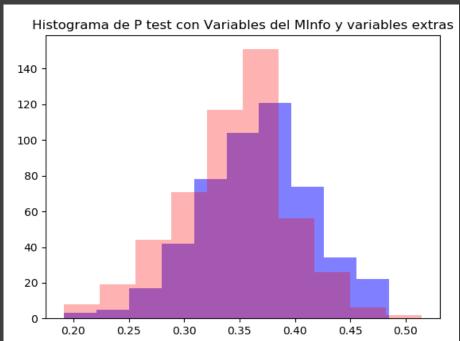








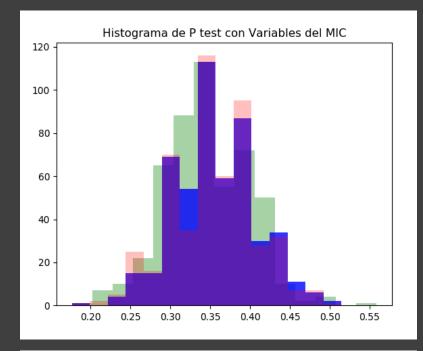


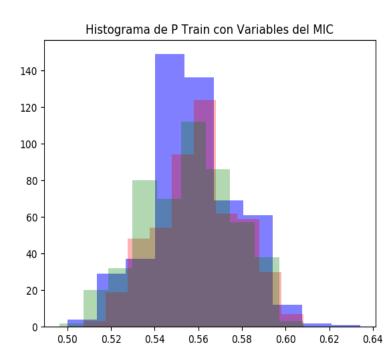












MIC



Todas las variables, las 10 mejores y las 15 mejores (verde, azul, rojo)

VOLVER A LA EXTRACCIÓN DE CARACTERÍSTICAS



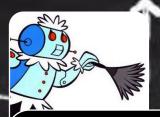
Recopilación de datos

- Búsqueda en Internet
- Determinar forma de descarga
- Acomodo en Google Sheets



Preprocesamiento

- Homogeneización
- Uso de Querys.



Limpieza de datos

- Manejo de datos faltantes
- Variables categóricas
- Notaciones inadecuadas



Evaluación

- Importar datos.
- Limpieza de datos en programa
- Clasificadores e histogramas



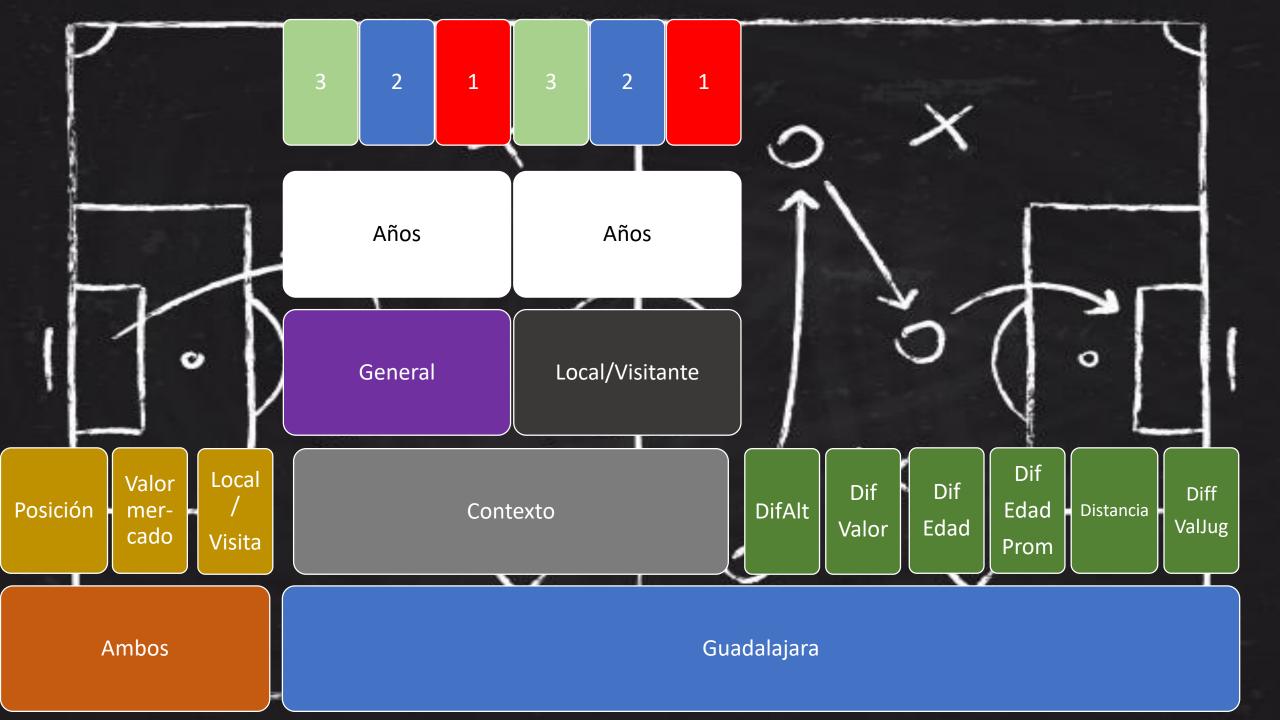






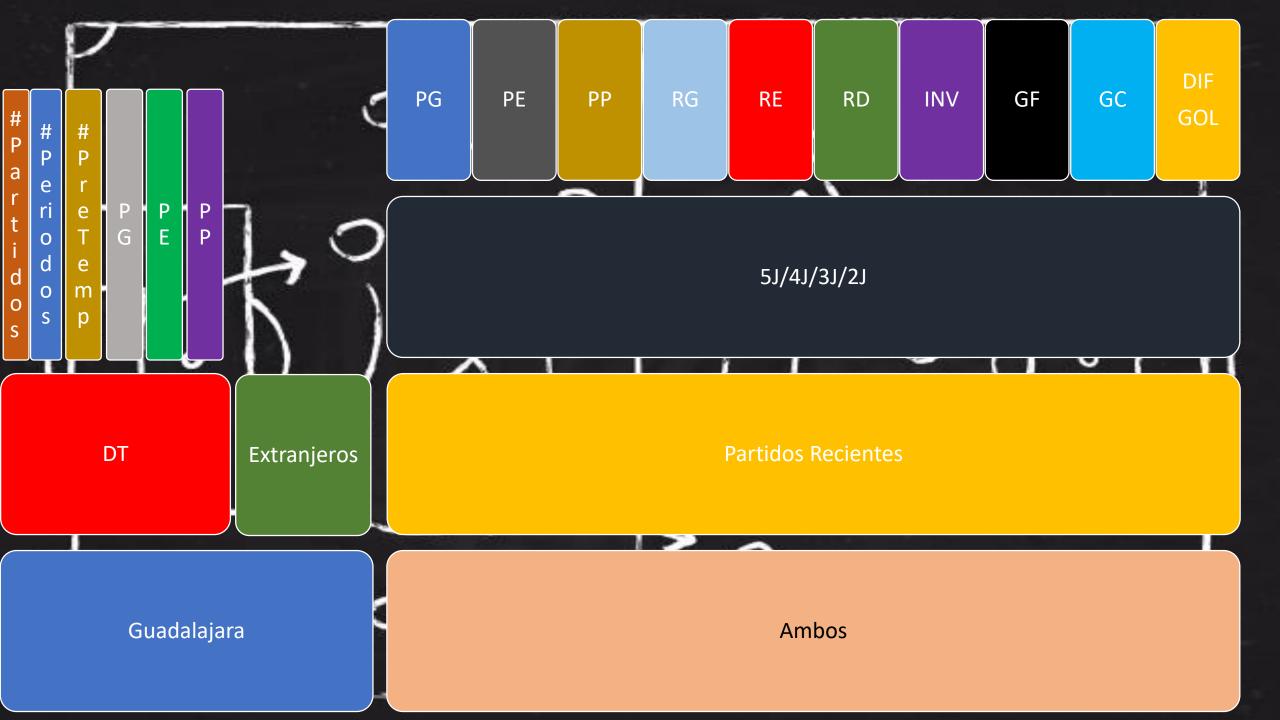


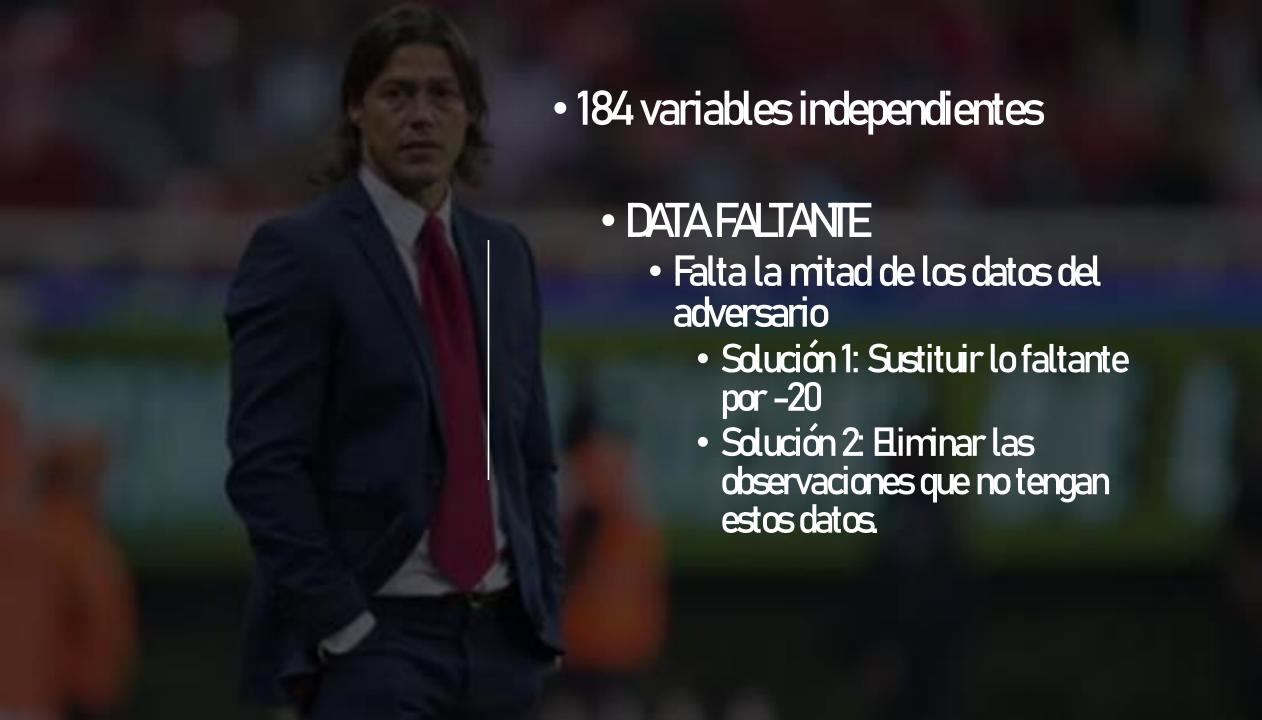
- Director Técnico en turno.
- Número de Partidos dirigidos de forma consecutiva.
- Número de pretemporadas dirigidas.
- Número de extranjeros.
- Información del adversario.
- Rachas e Invictos.



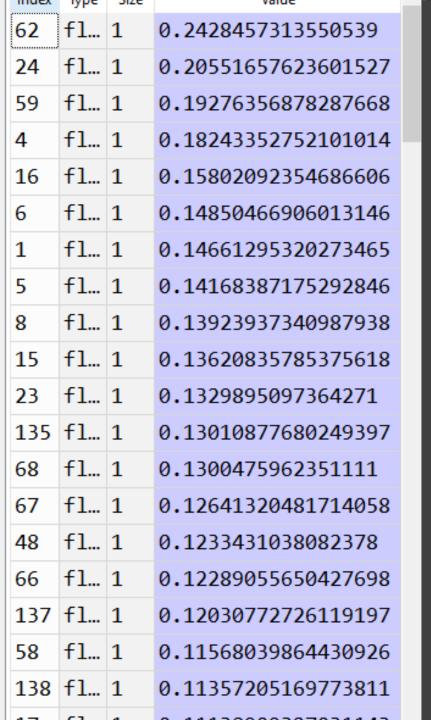
3 años 1 año

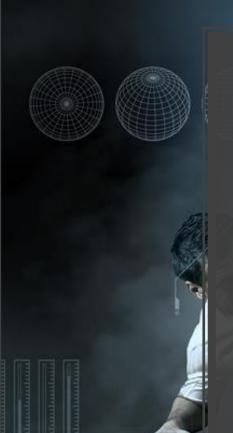
2 años











COEHCIENTES DE MÁXIMA INFORMACIÓN

- Primeras dos o 3 variables con peso considerable
- Pesos importantes en el resto
- 20 variables fueron las que mejor precisión mostraron

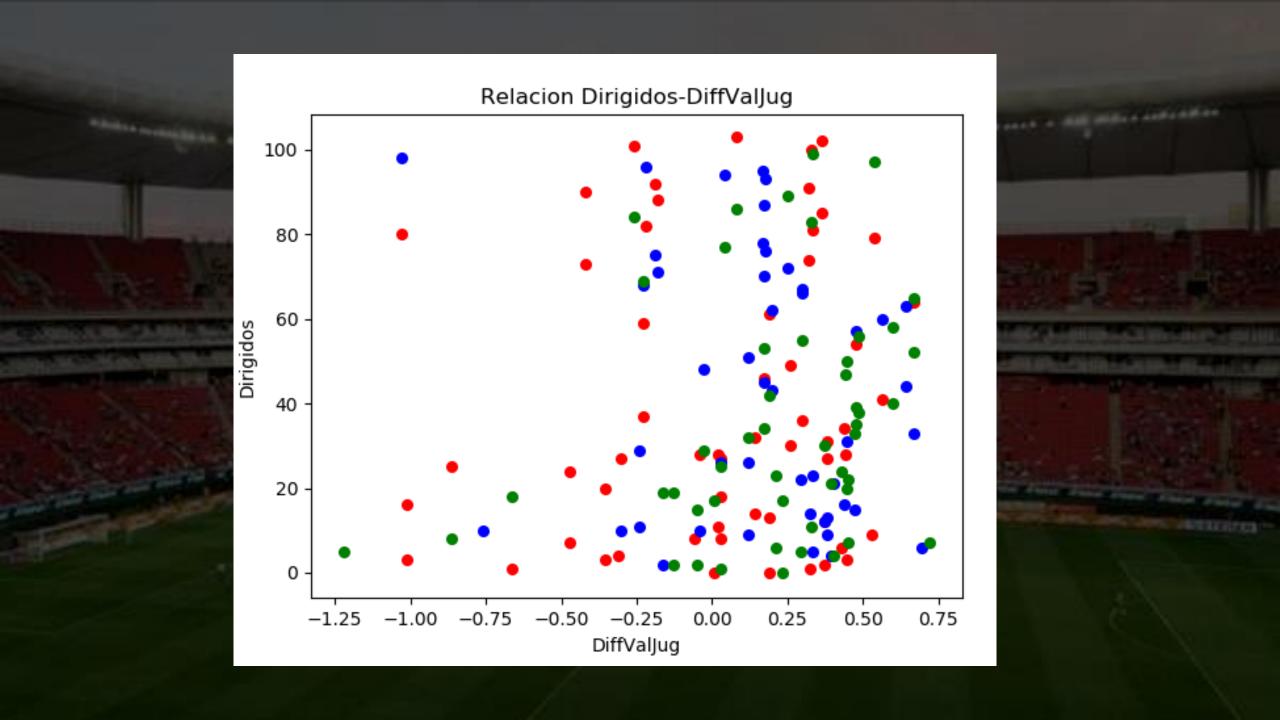
1. #PartidosDir 2GPPORJ2 3.DffValJug 4.ValorLocal 5.G_PPORJ3

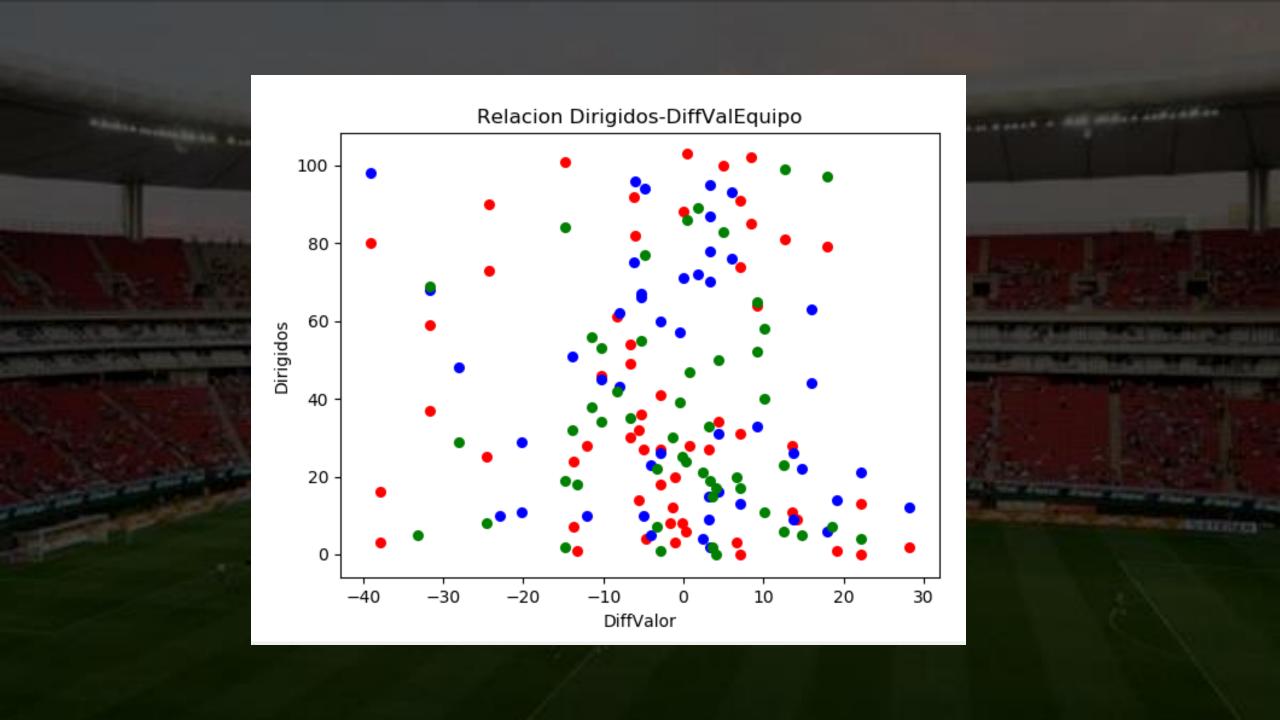
6.DffValor 7.PosLocal 8. Valor Visitante 9.Distancia 10.G PUNIOS3

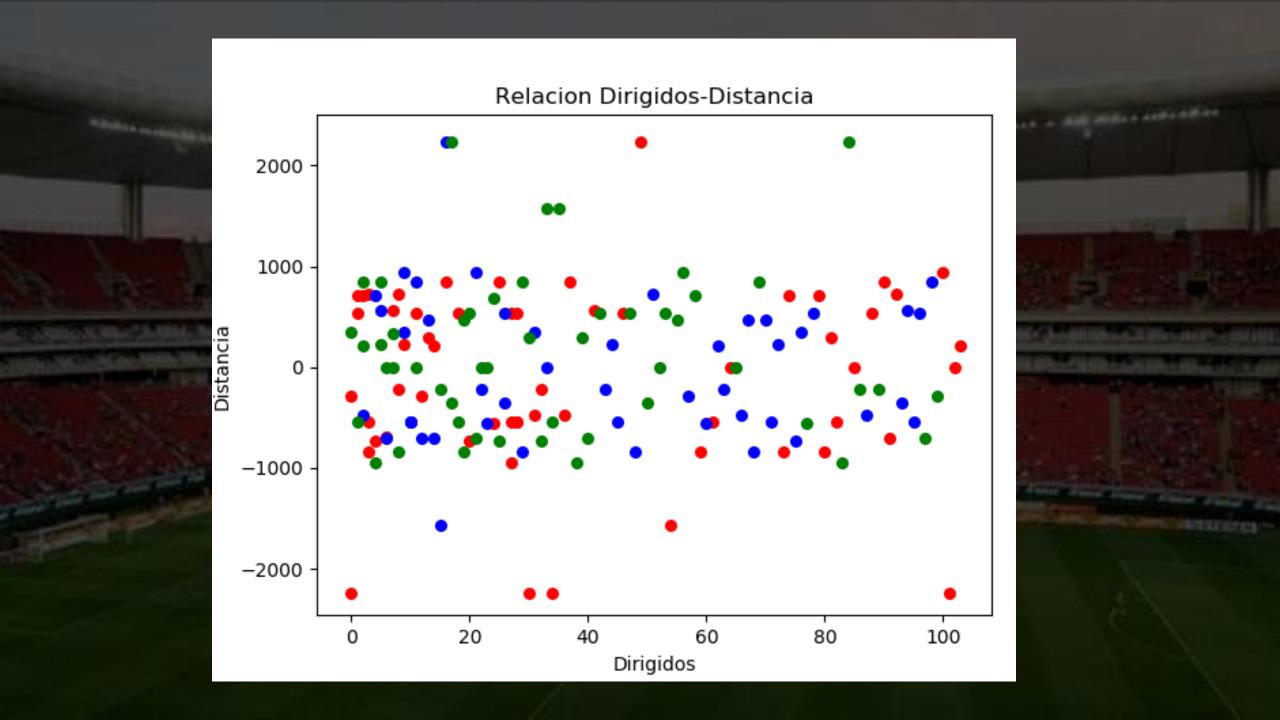
"Los resultados de hace 2 y 3 juegos contra un equipo son importantes."

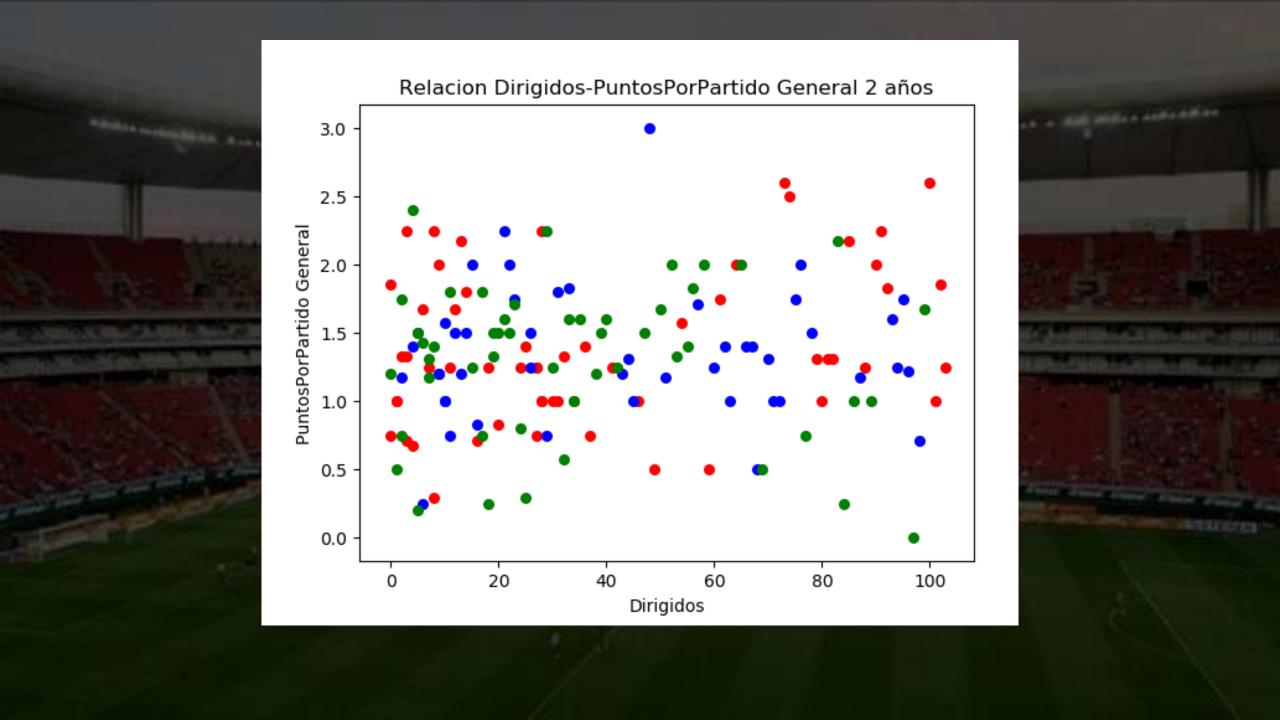
"El valor económico de un club y su posición en la tabla afectan"

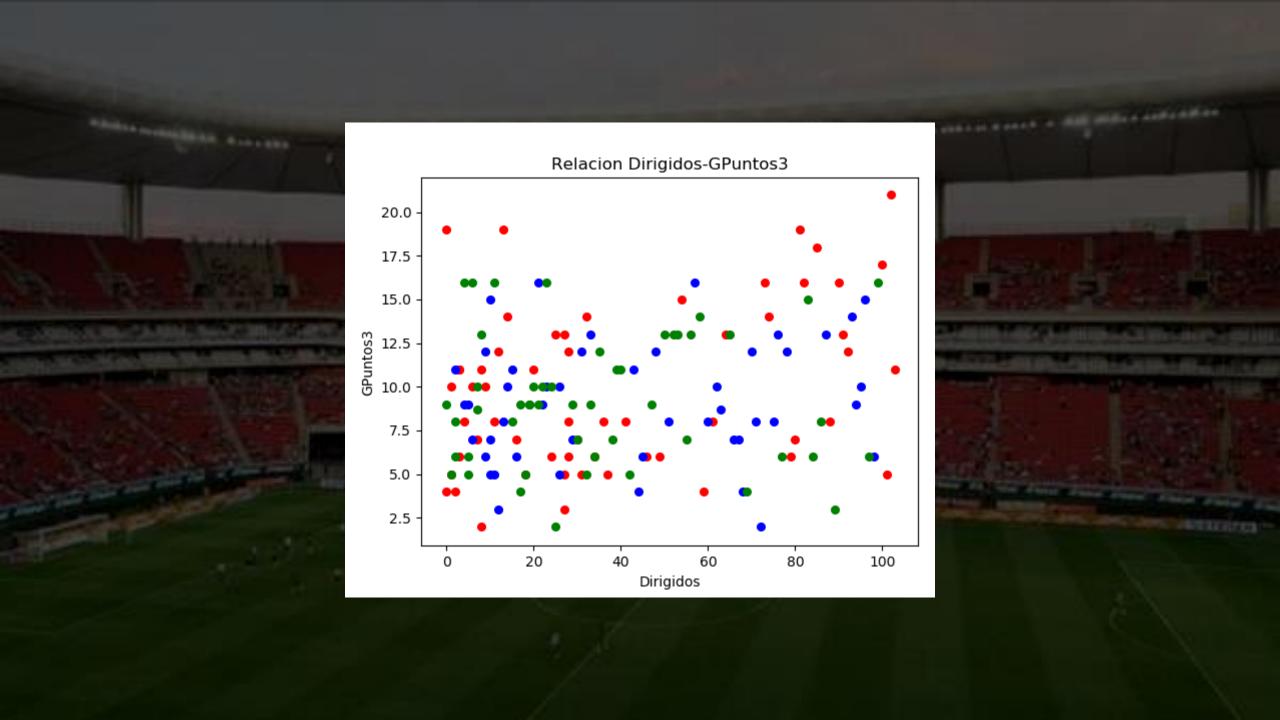
"Los resultados consecutivos de un DT ofrecen mucha información".

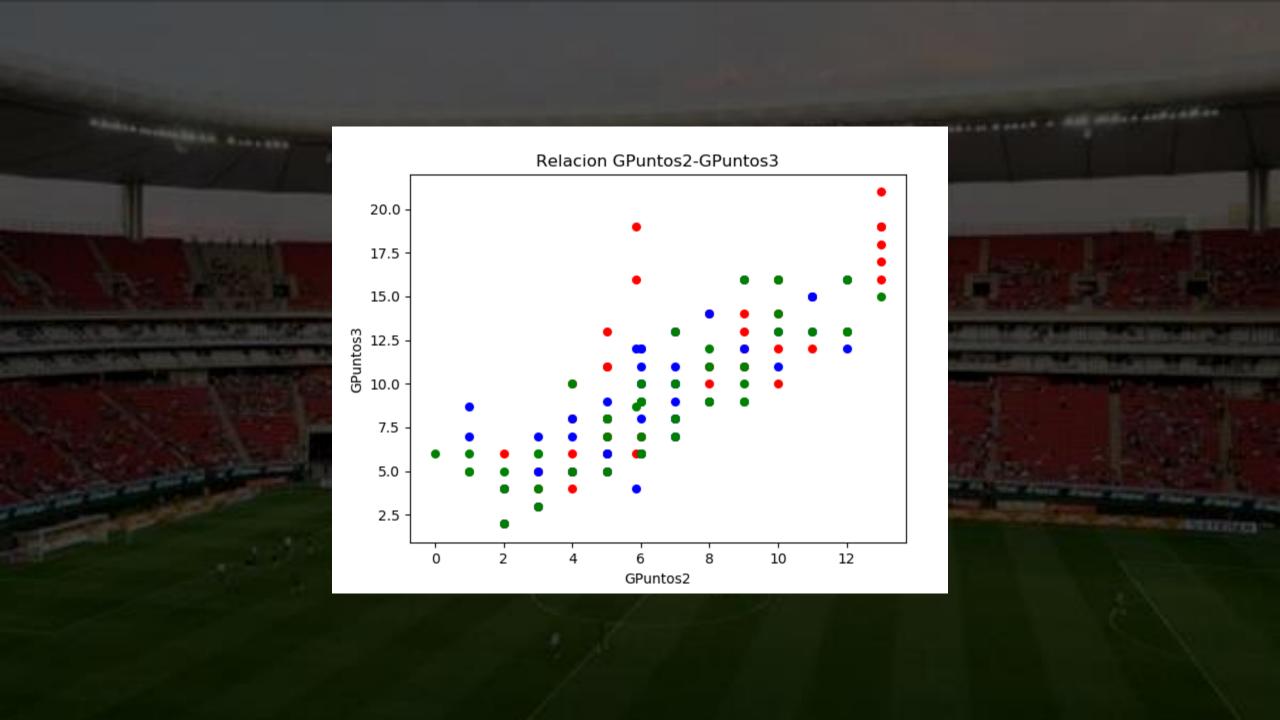












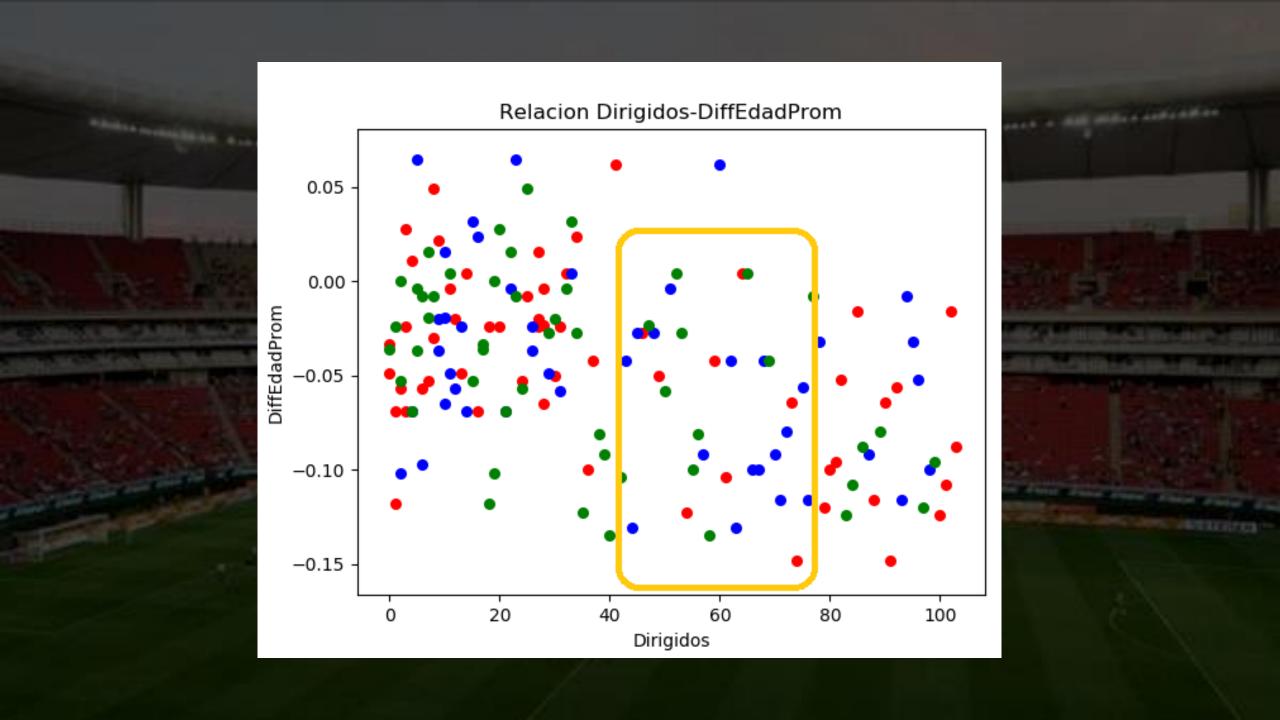
11. G_Puntos2 12 DfG5_Ad 13. DT_PP 14. DT_PE 15. LV PPorJ2

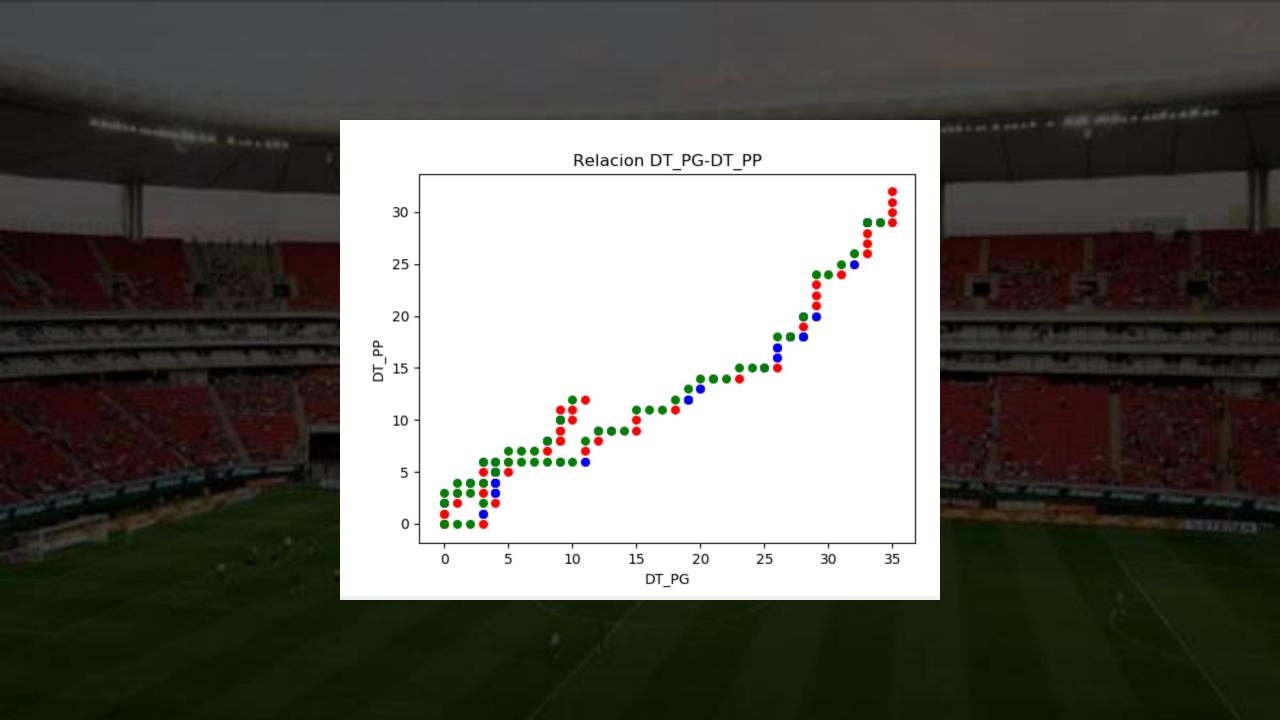
16. DT_PG 17. GC4_Ad 18. DffEdadProm 19. DfG4_Ad 20. G Ffect3

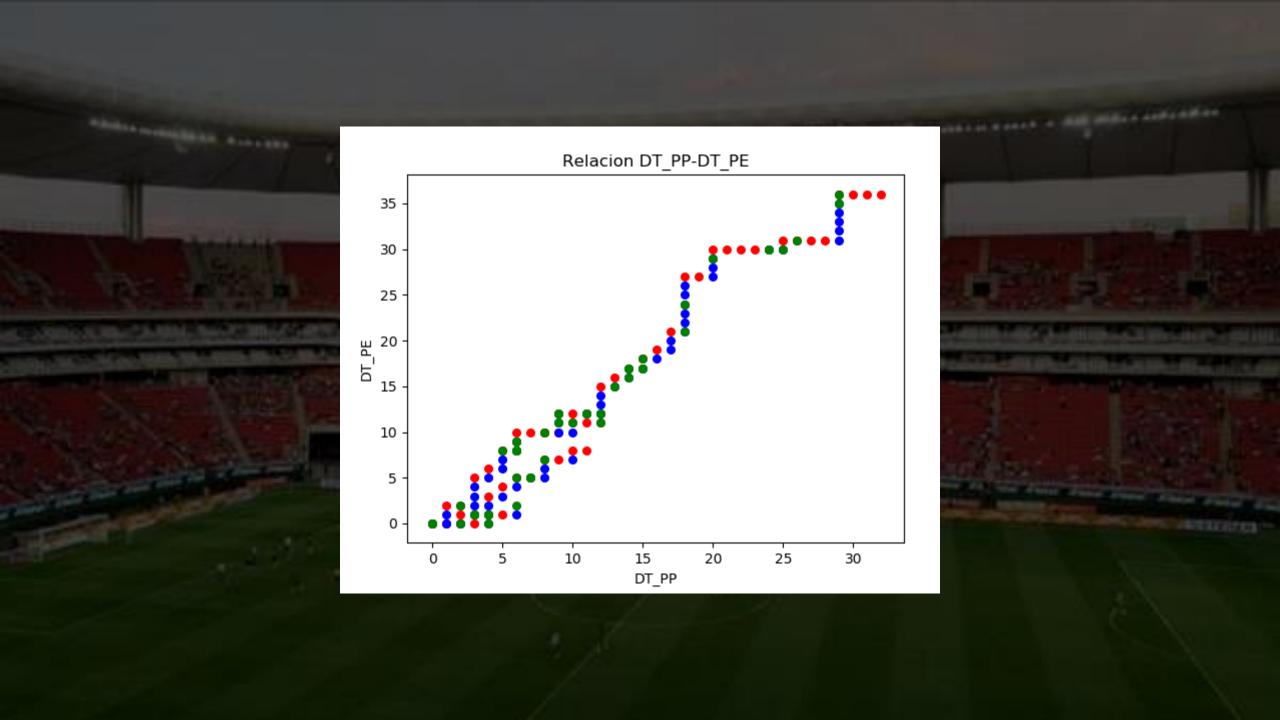
"En el contexto de Local/Msitante, LV_PPorJ2 es el más importante".

"La diferencia de edades entre equipos juega un papel importante"

"La diferencia de goles de los últimos partidos del adversario son relevantes"







21.DfEdad 22 IDTempo 23.G_Fect2 24.LV_DfGd3 25.LV_PPorJ3

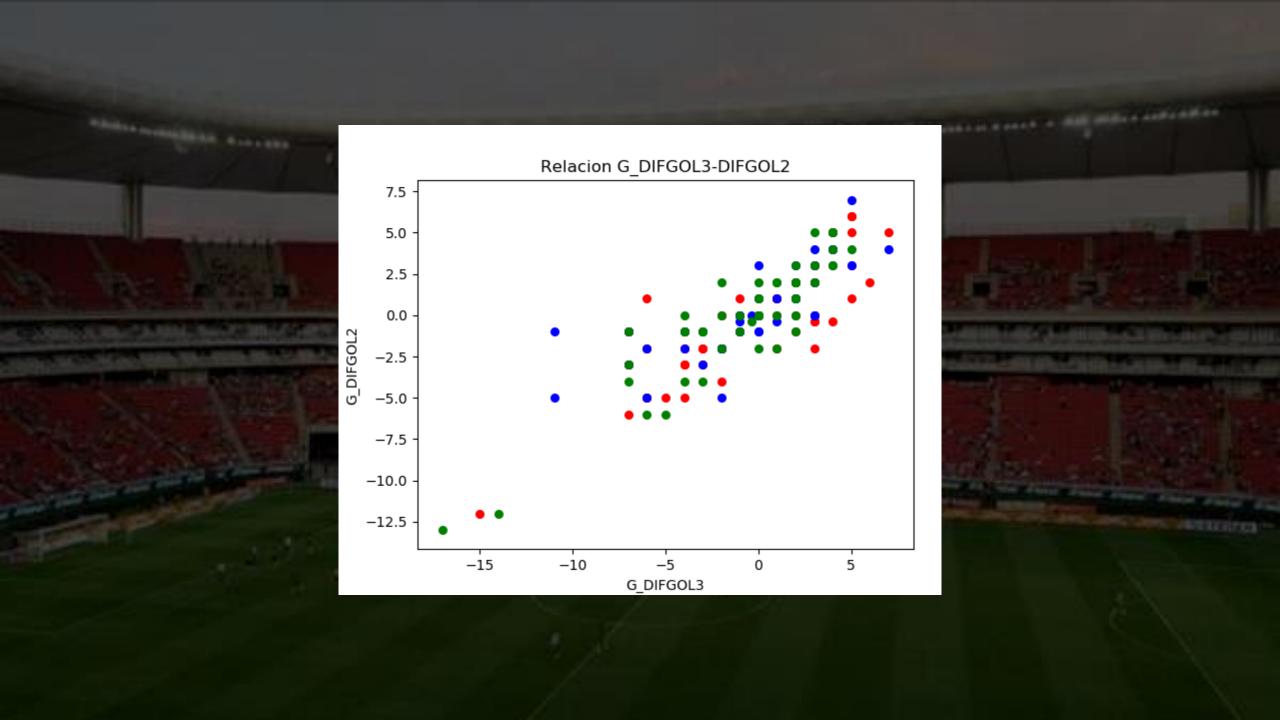
26. DfAt 27.G DfGd3 28.GF5 29.DFG2 30.DFG5

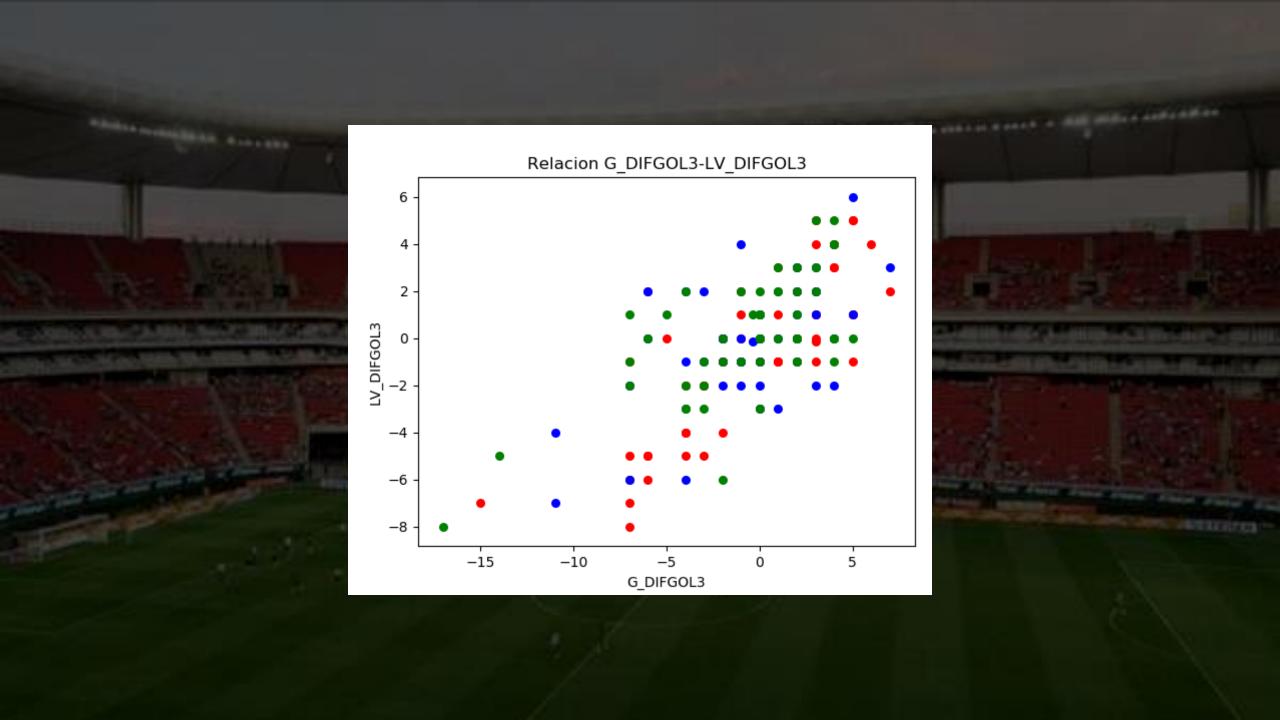
"La diferencia de edades sin promediar aporta menos información que la promediada"

"IDTempo indica qué tan avanzado está el torneo.Tiene relevancia importante"

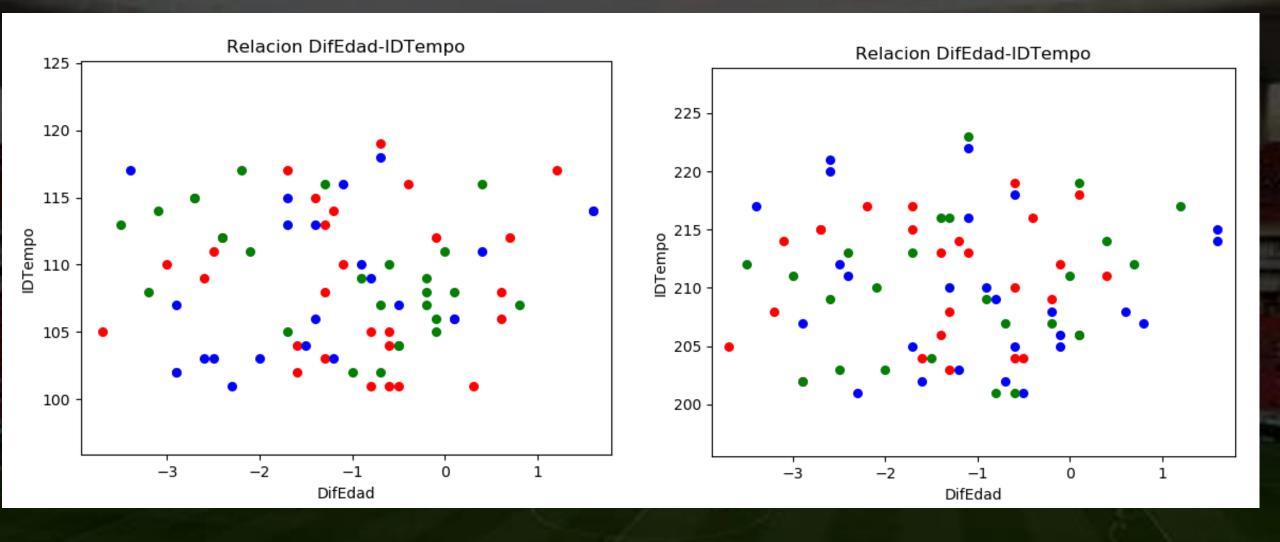
"La diferencia de altura no es insignificante".

"Los resultados de hace 3 años y la diferencia de goles de hace 2 tienen relevancia media"

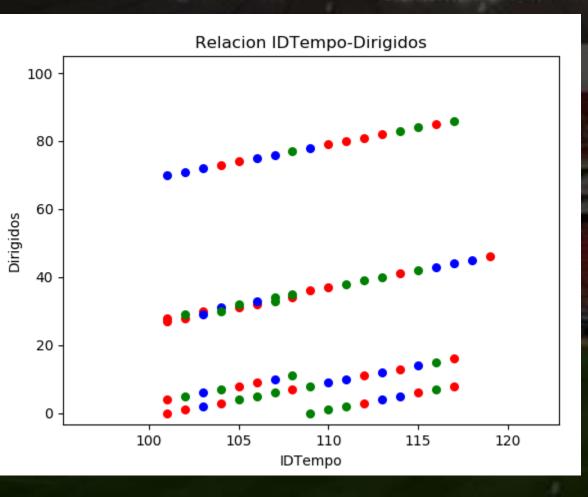


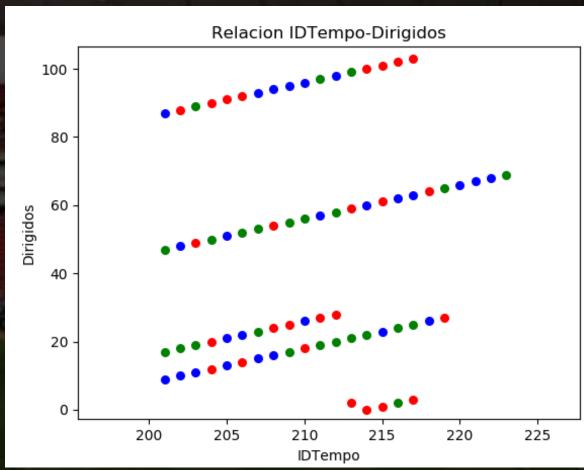


APERTURA / CLAUSURA



IDTEMPO VS.#DIRIGIDOS APERTURA / CLAUSURA



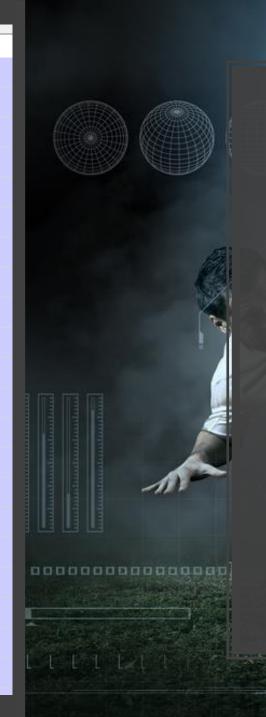


31.GC5 32G DfGd2 33.GF2 34.DfGLAd 35.LV_Puntos3

36. G_Fect1 37.G_DfGd1 38.DfG4 39.LV_Puntos2 40.DfG3



Index	Туре	Size	Value
82	float	1	0.0
86	float	1	0.0
90	float	1	9.360894339446204e-05
95	float	1	0.0004698266404697009
89	float	1	0.0009776085909431187
97	float	1	0.0018060766838135093
3	float	1	0.0024802483990302424
88	float	1	0.0034019510185514974
91	float	1	0.0034019510185514974
94	float	1	0.0036077260891390845
79	float	1	0.0036110087057941026
81	float	1	0.0036170441543933204
78	float	1	0.005938148028337897
85	float	1	0.005938148028338057
93	float	1	0.0062554802142303426
162	float	1	0.007335757745632691
150	float	1	0.007335757745633172
152	float	1	0.007335757745633172
153	float	1	0.007335757745633172
155	float	1	0.007335757745633172
156	float	1	0.007335757745633172



COEMCIENTES DE MÁXIMA INFORMACIÓN

- Primeras dos variables con relevancia nula
- Primeras 5 variables prácticamente despreciables.
- 16-21 muy similares

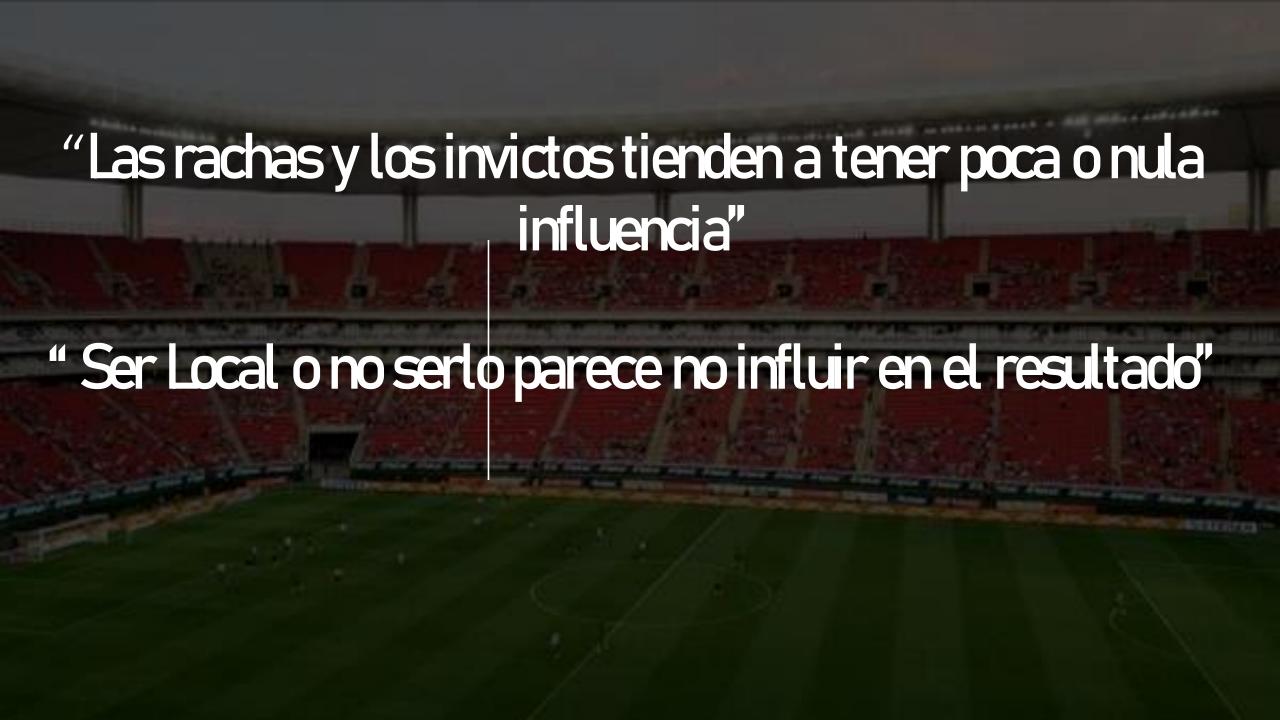
1. RachaG5 2RachaE5 3.RachaP5 4.Invicto4 5.RachaE2

6.Invicto2 7.esLocal 8.RachaE3 9.RachaP4 10.Invicto5

11. PE2 12 PPasado 13. PG2 14. RachaG2 15. RachaP2

16. Invicto3Ad 17. RachaG3Ad 18. RachaE5Ad 19. RachaE4Ad 20. RachaE2Ad

*Msma relevancia



21. RachaP5Ad 22Invicto4Ad 23.Invicto3 24.RachaP3 25.LV_PPI

26.Invicto5Ad 27.RachaG5Ad 28.RachaG4Ad 29.PP2 30.RachaG3

31. RachaE4 32Invicto2Ad 33.RachaP4Ad 34.RachaG4 35.LV_PJ1

36.NumPretemp 37.LV_PE2 38.P5GF 39.PG4 40.PE3

"Las rachas y los invictos siguen teniendo poca o nula influencia"

" El número de pretemporadas dirigidas no afecta tanto en un resultado".

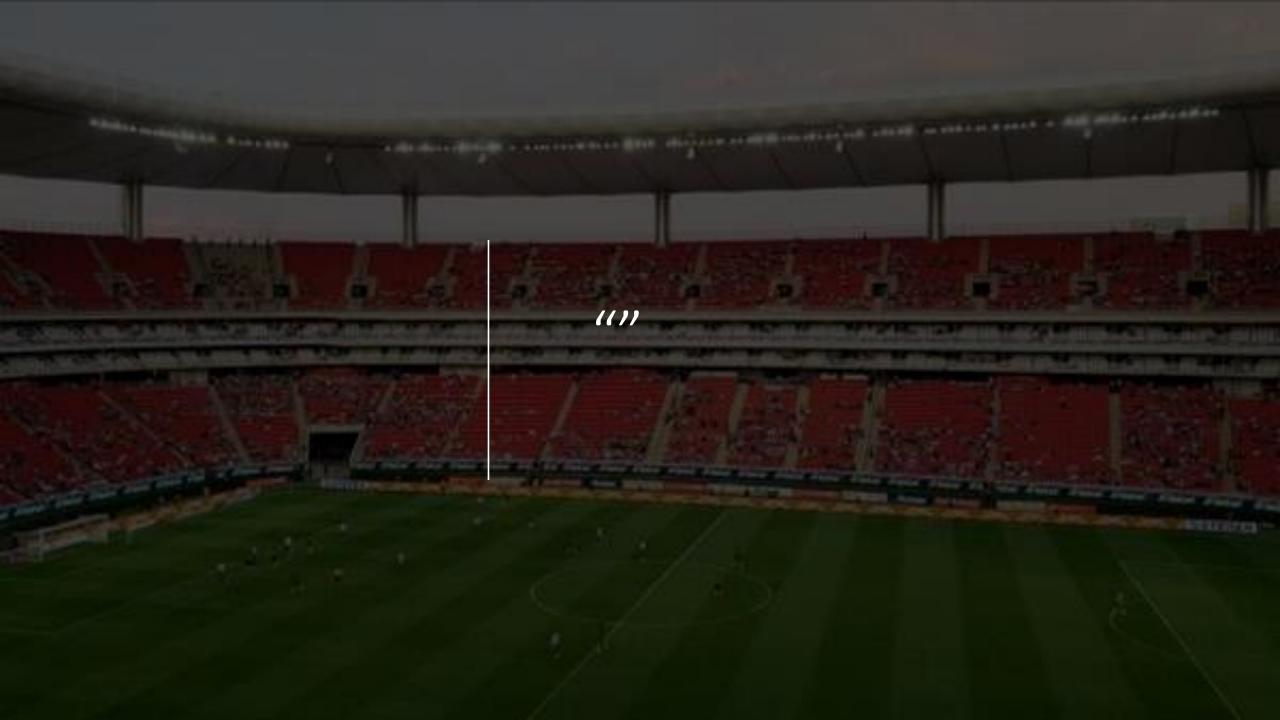
"¿Afectará únicamente el hecho de haber o no haber dirigido al menos una pretemporada?"

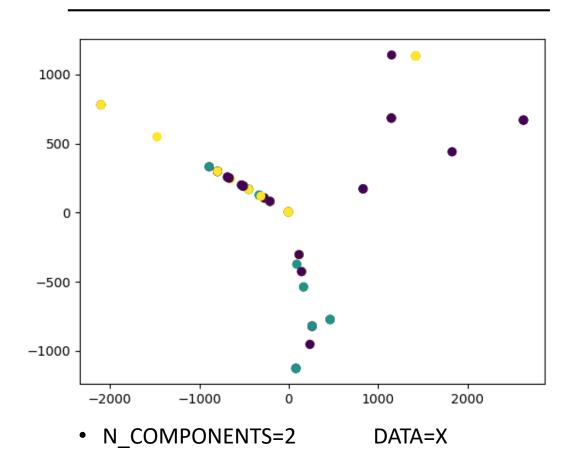
41. AdPPasado 42LV_PEI 43.PE5 44.RachaG2Ad 45.P2GF

46.GCI_Ad 47.AdPP5 48.LV PE3 49.RachaP3Ad 50.AdPE2

51. PP3 52LV_PP3 53.GC2 54.PP4 55.LV_PG3

56.PIGE 57.RachaE3AD 58.AdPG3 59.AdPG2 60.RachaP2Ad

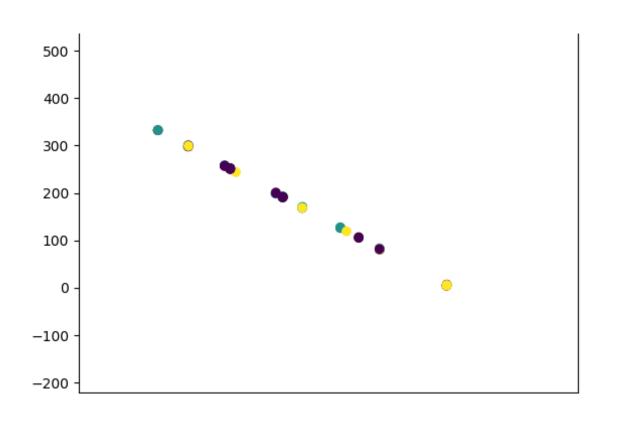


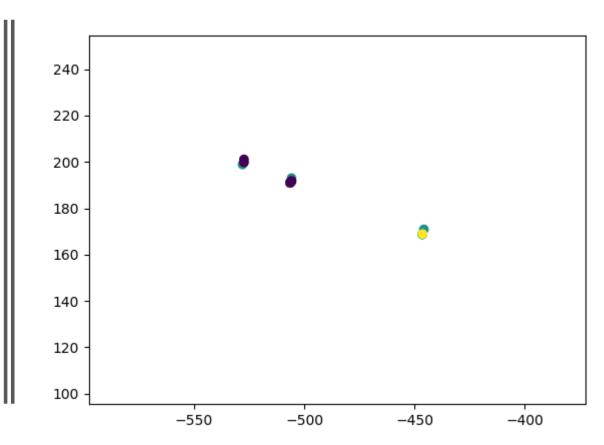




N_COMPONENTS=2

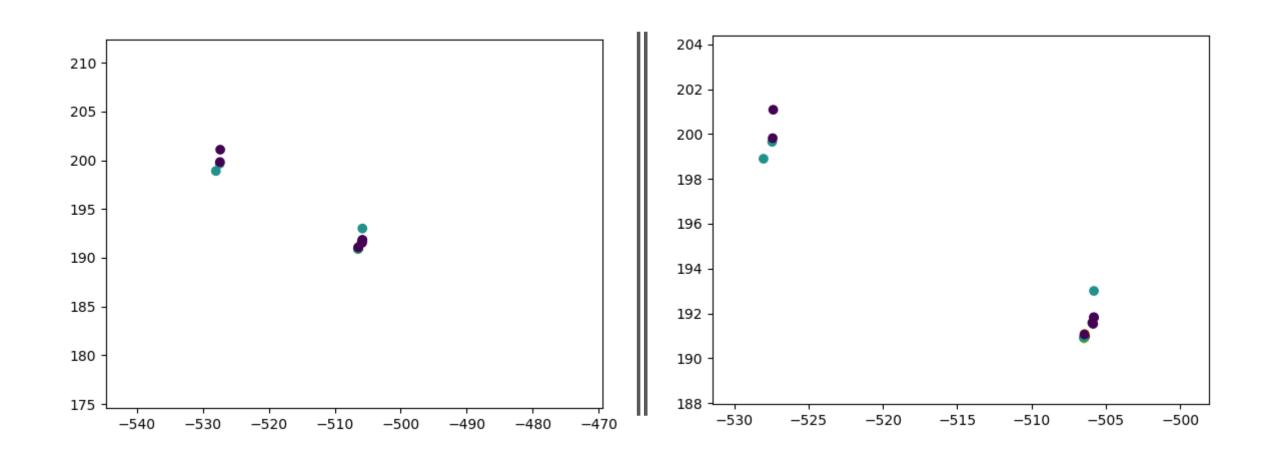
DATA=X

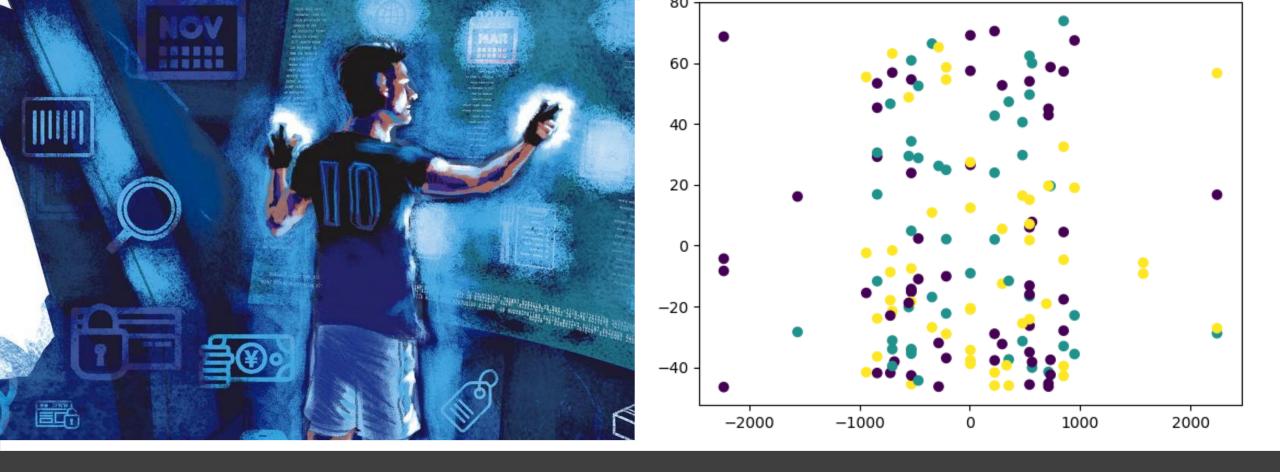




N_COMPONENTS=2

DATA=X

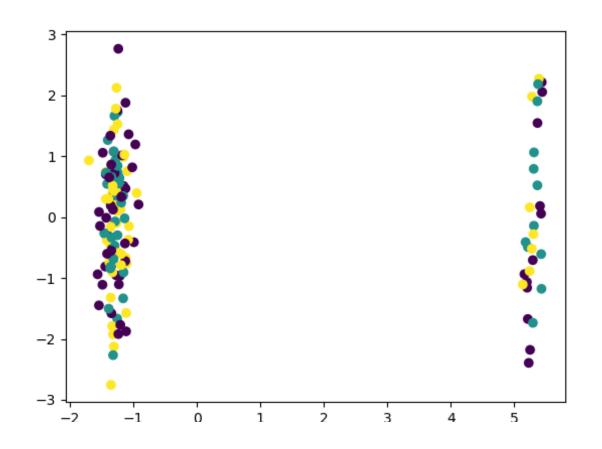




N_COMPONENTS=2

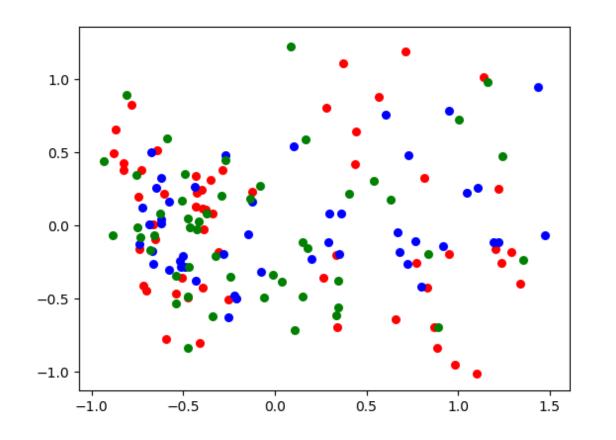
DATA=X_MIC20







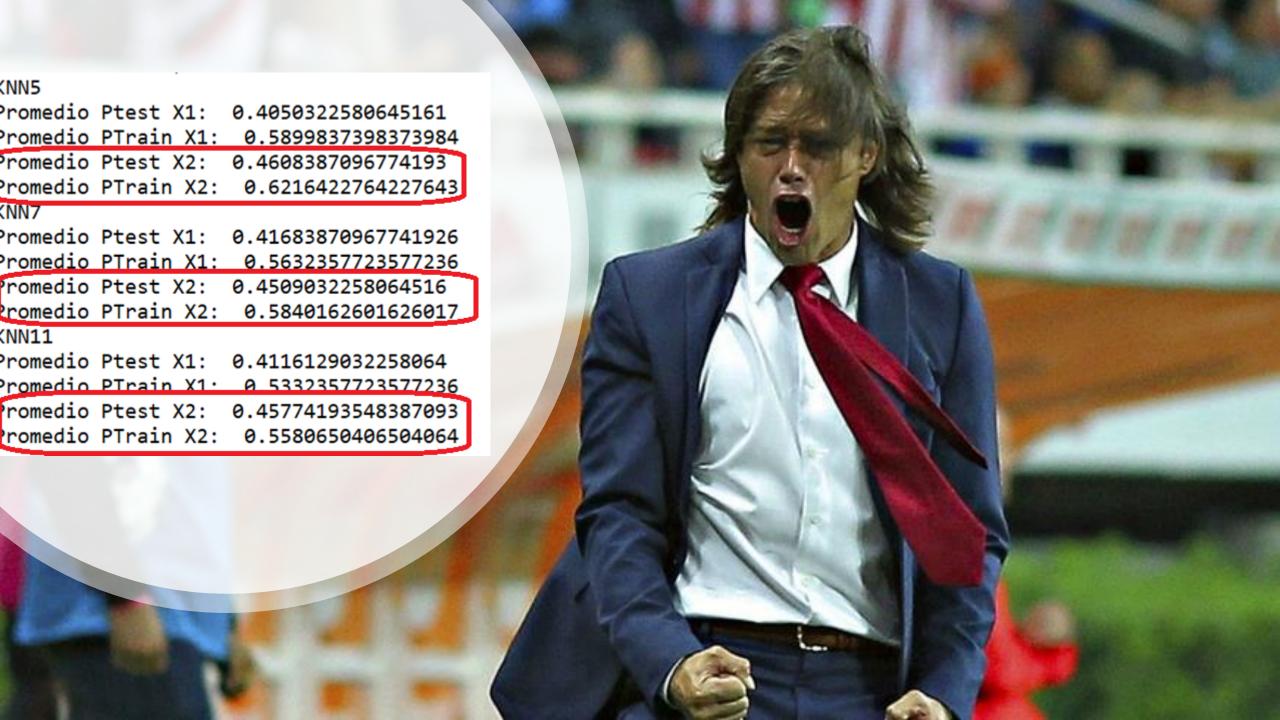
PCA N_COMPONENTS=2 DATA=X SCALER=MinMaxScaler





PCA N_COMPONENTS=2 DATA=X_MIC20 SCALER=MinMaxScaler









- Se probó con PCA de 2 y de 4 componentes.
- Usando KNearestNeighbors con 5 vecinos se consiguieron los mejores resultados.









- Analizar distribuciones de variables relevantes respecto al resultado para obtener conclusiones de interés.
 - Hacer extracción de características del rival en turno.
 - Mejorar la estructura de datos y su extracción.



- Comprobar que las variables relevantes sean las mismas para todos los equipos.
- Obtener datos de casas de apuesta.
 - Validar otras ligas.



- Sistemas de juego (alineaciones) como variable independiente.
- Valor del mercado de ofensiva, defensiva, media y portería.
- Estructura de los clubes



- Clima del estadio en día y hora del partido.
- Balance de los fichajes
- Registro de DT vs DT y
 DT vs club



- Web Scraping y update automático
 - Bases de datos
- Expulsiones y lesiones
 - Ventaneo



