



VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

FACVLTADE CIENCIAS

GRADº DE ESTADÍSTICA

TRABAJO DE FIN DE GRADO: ANEXO

**ANÁLISIS ESPACIAL DEL TIRO EN BALONCESTO PROFESIONAL
Y SU TRANSFORMACIÓN A LO LARGO DEL TIEMPO**

Autor: Rodrigo Valladares Alonso

Tutor: José Luis Vicente Villardón

Salamanca, 2020



**VNiVERSiDAD
D SALAMANCA**
CAMPUS DE EXCELENCIA INTERNACIONAL



FACULTAD DE CIENCIAS

GRADO DE ESTADÍSTICA

TRABAJO DE FIN DE GRADO: ANEXO

**ANÁLISIS ESPACIAL DEL TIRO EN BALONCESTO PROFESIONAL
Y SU TRANSFORMACIÓN A LO LARGO DEL TIEMPO**

Autor: Rodrigo Valladares Alonso

Tutor: José Luis Vicente Villardón

Salamanca, 2020

ÍNDICE DE CONTENIDOS

1. Clustering:	1
1.1. 1980	1
1.1.1. Triples	1
1.1.2. Todos	2
1.1.3. Por minuto	4
1.1.4. Intentados	6
1.2. 1995	9
1.2.1. Triples	9
1.2.2. Todos	10
1.2.3. Por minuto	12
1.2.4. Intentados	13
1.3. 2010	16
1.3.1. Triples	16
1.3.2. Todos	17
1.3.3. Por minuto	19
1.3.4. Intentados	21
1.4. 2019	23
1.4.1. Triples	23
1.4.2. Todos	25
1.4.3. Por minuto	27
1.4.4. Intentados	29
2. Clasificación	30
2.1. Variables de tiro	30
2.1.1. 1995	30
2.1.1.1 Pruebas con misma temporada	30
2.1.1.2 Pruebas con distintas temporadas	31
2.1.2. 2010	33
2.1.2.1 Pruebas con misma temporada	33
2.1.2.2 Pruebas con distintas temporadas	35
2.1.3. 2019	37
2.1.3.1 Pruebas con misma temporada	37
2.1.3.2 Pruebas con distintas temporadas	39
2.2. Variables de estadística avanzada	40
2.2.1. 1995	40
2.2.1.1 Pruebas con misma temporada	40

2.2.1.2 Pruebas con distintas temporadas.....	43
2.2.2. 2010.....	45
2.2.2.1 Pruebas con misma temporada.....	45
2.2.2.2 Pruebas con distintas temporadas.....	48
2.2.3. 2019.....	50
2.2.3.1 Pruebas con misma temporada.....	50
2.2.3.2 Pruebas con distintas temporadas.....	53
3. Regresión.....	56
3.1. Todos los tiros.....	56
3.1.1. Variables iniciales.....	56
3.1.2. No lineales añadidas.....	58
3.1.3. Alturas y pesos añadidas.....	60
3.1.4. Tiros añadidos/Todas las variables.....	63
3.1.5. Mínimo de variables importantes.....	65
3.2. Tiros a más de 5 pies.....	68
3.2.1. Variables iniciales.....	68
3.2.2. No lineales añadidas.....	70
3.2.3. Alturas y pesos añadidas.....	72
3.2.4. Tiros añadidos/Todas las variables.....	75
3.2.5. Mínimo de variables importantes.....	77
4. Código en R.....	82
4.1. Clustering.....	82
4.2. Clasificación.....	85
4.3. Regresión.....	87

En este Anexo están incluidos todos los gráficos y tablas empleados en los análisis llevados a cabo en el trabajo, en el mismo orden en el que han sido realizados en éste, así como el código usado en *R Studio* para generarlos. Todas las figuras se describen en su encabezado.

1. CLUSTERING

1.1. 1980

1.1.1. Triples

Tabla 1: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de triples para la temporada 1979/1980. Generado con *R Studio*.

cluster	grupo real				
	C	PF	PG	SF	SG
1	0	4	11	9	12
2	0	0	7	2	5
3	0	2	3	2	8
4	36	33	17	31	10
5	0	0	1	1	3

Gráfico 1: Gráfico del método *PAM* para el *subset* de triples para la temporada 1979/1980. Generado con *R Studio*.

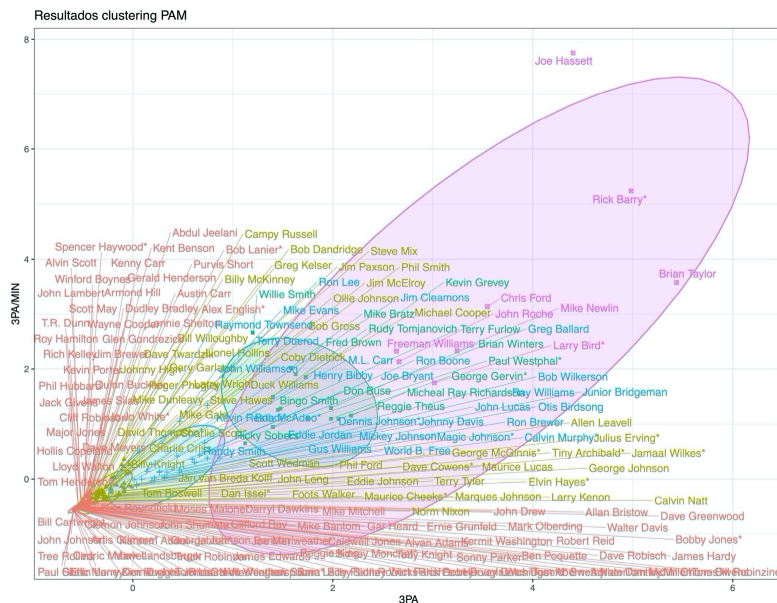
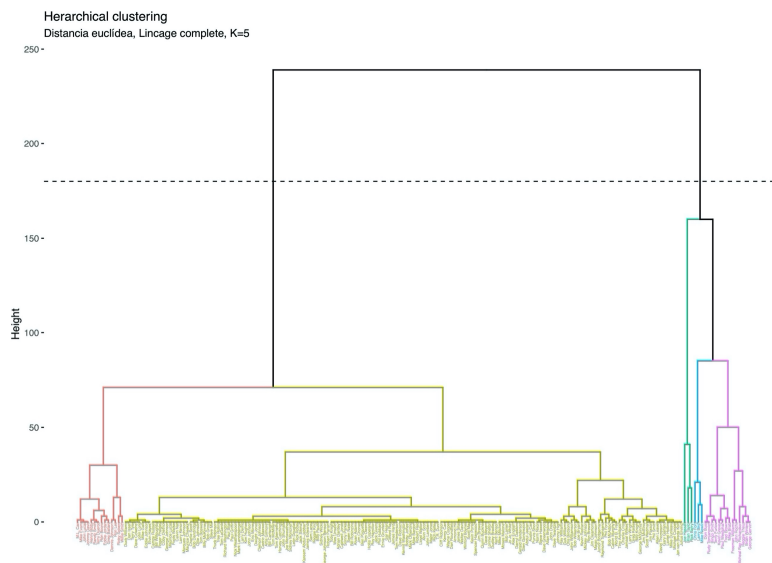


Gráfico 2: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de triples para la temporada 1979/1980. Generado con *R Studio*.



1.1.2. Todos

Gráfico 3: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de todos los tiros para la temporada 1979/1980. Generado con *R Studio*.

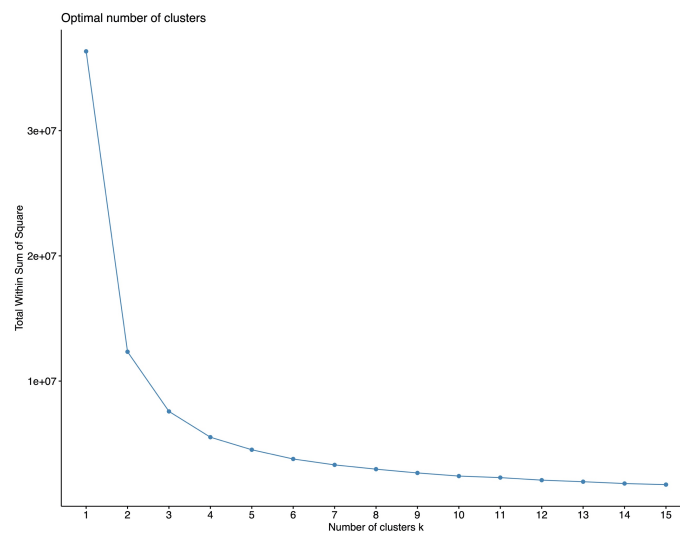


Gráfico 4: Representación del método *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 1979/1980. Generado con *R Studio*.

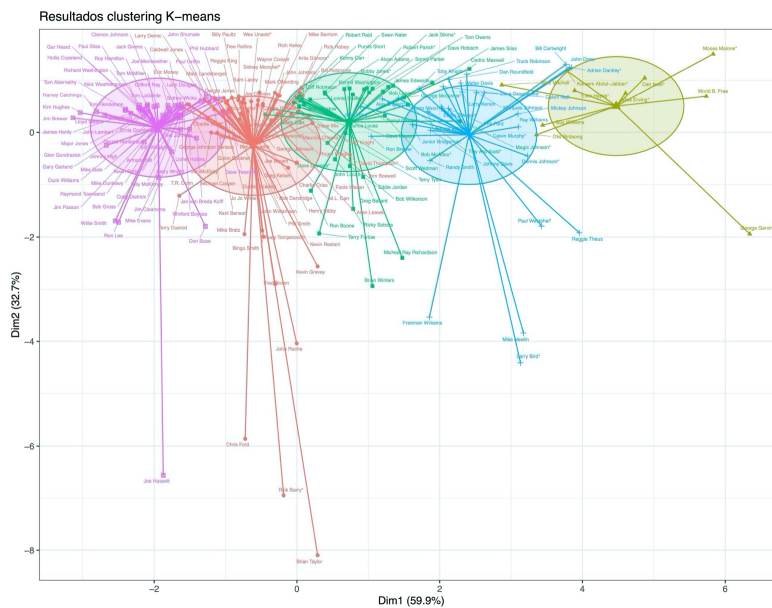


Tabla 2: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 1979/1980. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	13	11	11	15	13
2	3	1	1	2	3
3	10	9	5	9	6
4	1	6	4	9	10
5	9	12	18	10	6

Gráfico 5: Gráfico del método *PAM* para el *subset* de todos los tiros para la temporada 1979/1980. Generado con *R Studio*.

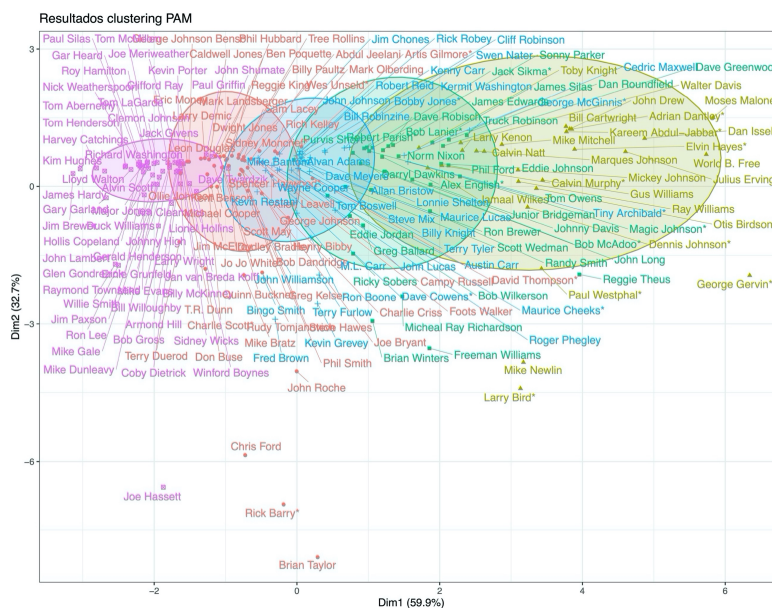
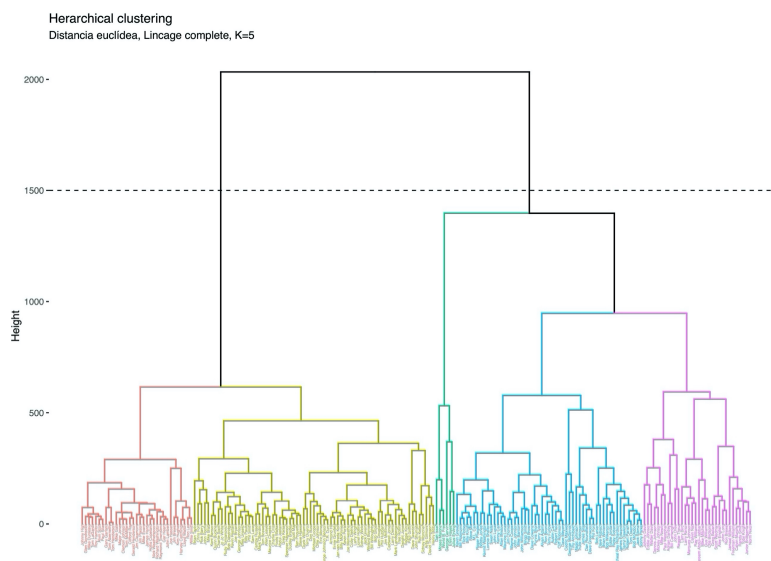


Gráfico 6: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de todos los tiros para la temporada 1979/1980. Generado con *R Studio*.



1.1.3. Por minuto

Gráfico 7: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de los tiros por minuto para la temporada 1979/1980. Generado con *R Studio*.

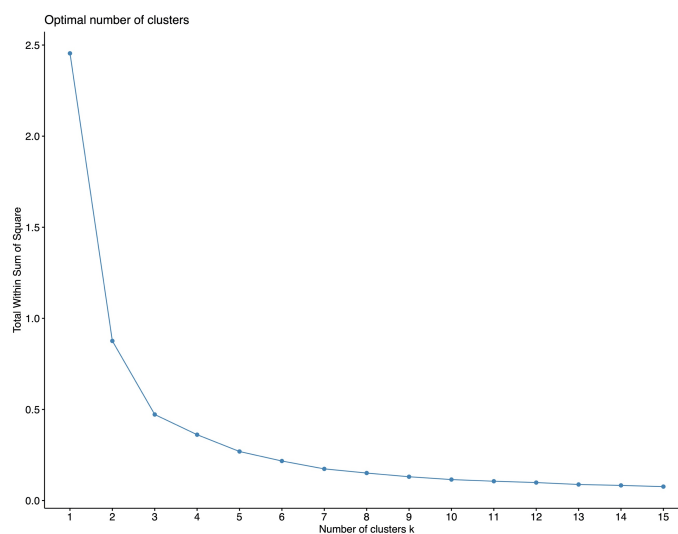


Gráfico 8: Representación del método *K-Means* con $k = 5$ para el *subset* de los tiros por minuto para la temporada 1979/1980. Generado con *R Studio*.

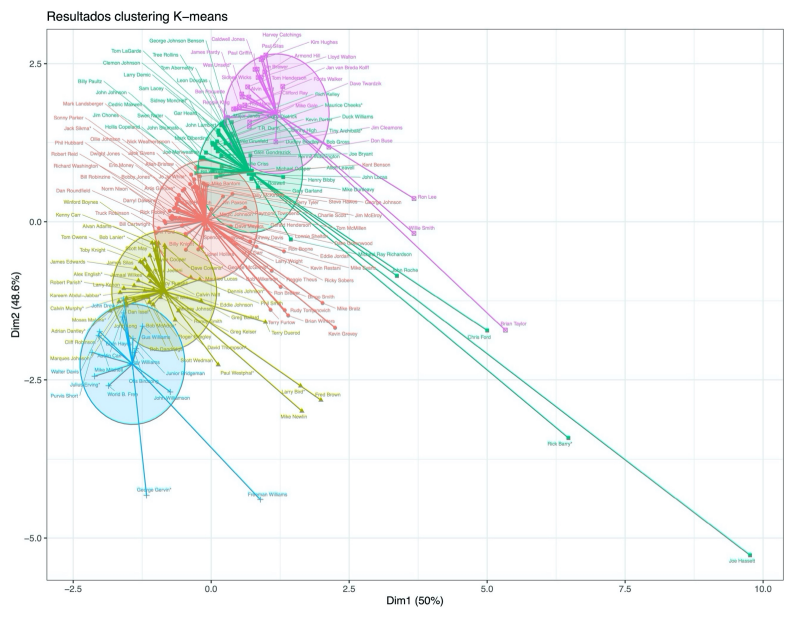


Tabla 3: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de los tiros por minuto para la temporada 1979/1980. Generado con *R Studio*.

cluster	C	PF	PG	SF	SG
1	11	10	12	9	7
2	9	15	12	14	12
3	0	1	2	6	6
4	9	8	1	14	13
5	7	5	12	2	0

Gráfico 9: Gráfico del método *PAM* para el *subset* de los tiros por minuto para la temporada 1979/1980. Generado con *R Studio*.

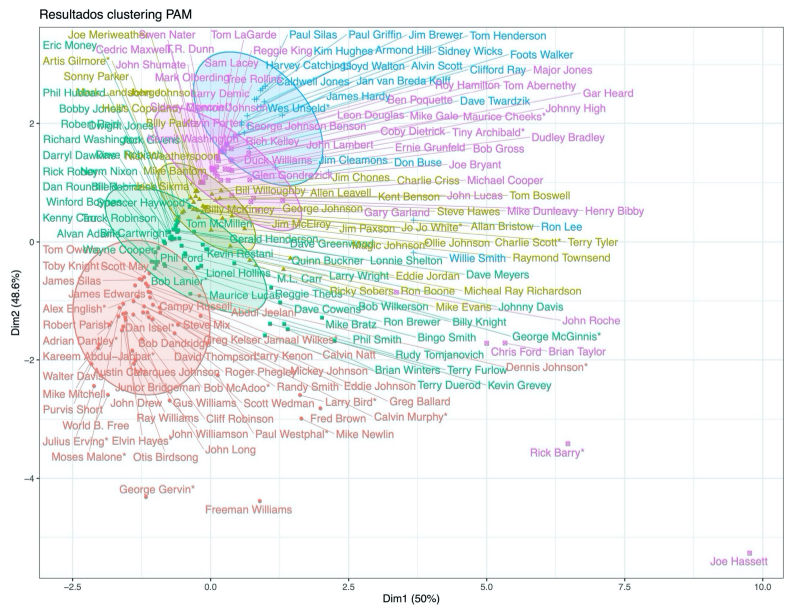
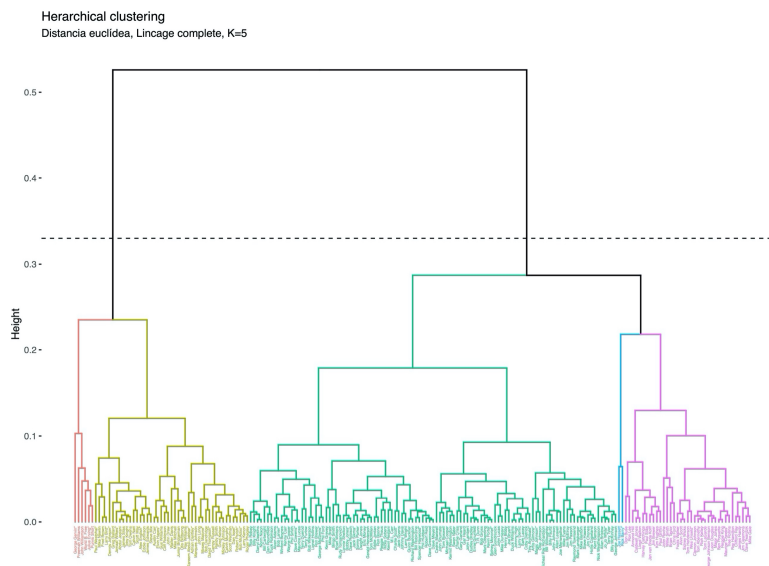


Gráfico 10: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de los tiros por minuto para la temporada 1979/1980. Generado con *R Studio*.



1.1.4. *Intentados*

Gráfico 11: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de los tiros intentados para la temporada 1979/1980. Generado con *R Studio*.

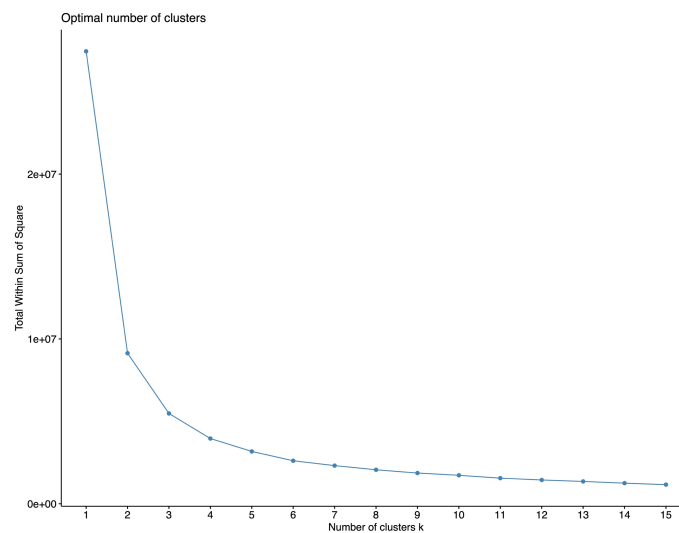


Gráfico 12: Representación del método *K-Means* con $k = 5$ para el *subset* de los tiros intentados para la temporada 1979/1980 Generado con *R Studio*.

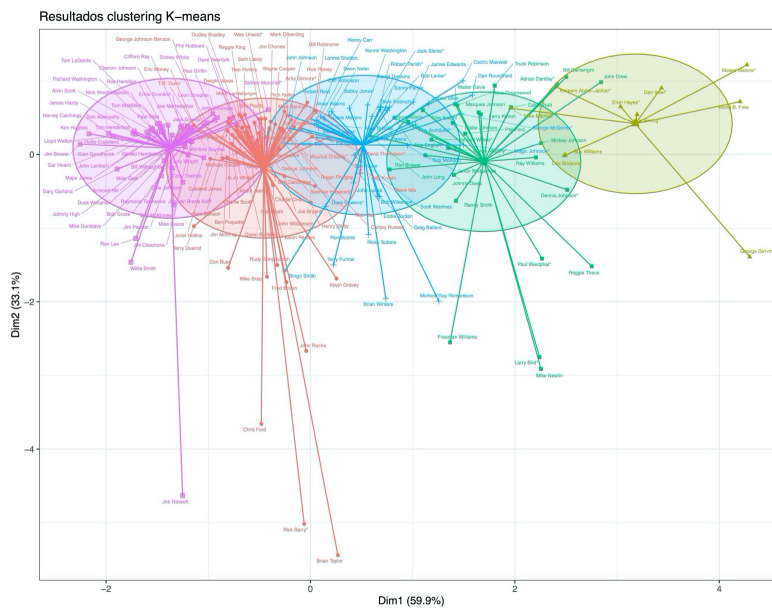


Tabla 4: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de los tiros intentados para la temporada 1979/1980. Generado con *R Studio*.

grupo real	cluster	C	PF	PG	SF	SG
1	13	11	12	13	13	
2	3	1	1	2	3	
3	1	5	4	9	10	
4	10	10	5	11	6	
5	9	12	17	10	6	

Gráfico 13: Gráfico del método *PAM* para el *subset* de los tiros intentados para la temporada 1979/1980. Generado con *R Studio*.

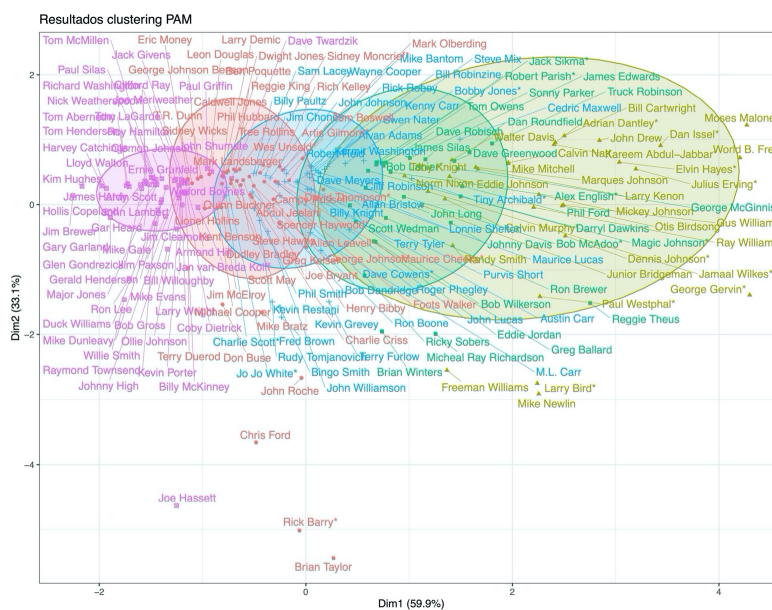


Gráfico 14: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de los tiros intentados para la temporada 1979/1980. Generado con *R Studio*.

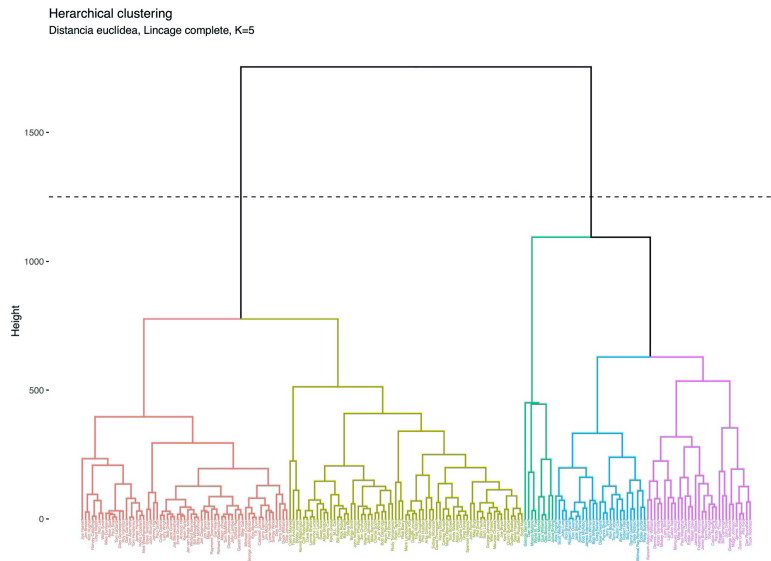


Tabla 5: Tabla de asignación de clusters a los jugadores por posición con la técnica *Hierarchical Clustering* con $k = 5$ para el *subset* de los tiros intentados para la temporada 1979/1980. Generado con *R Studio*.

cluster	grupo real				
	C	PF	PG	SF	SG
1	13	18	12	17	9
2	2	7	5	8	9
3	9	1	2	8	7
4	10	12	19	10	10
5	2	1	1	2	3

Tabla 6: Tabla de asignación de clusters a los jugadores por categoría exterior/interior con la técnica *Hierarchical Clustering* con $k = 2$ para el *subset* de los tiros intentados para la temporada 1979/1980. Generado con *R Studio*.

cluster	grupo real	
	EXT	INT
1	77	53
2	45	22

1.2. 1995

1.2.1. Triples

Gráfico 15: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de triples para la temporada 1994/1995. Generado con *R Studio*.

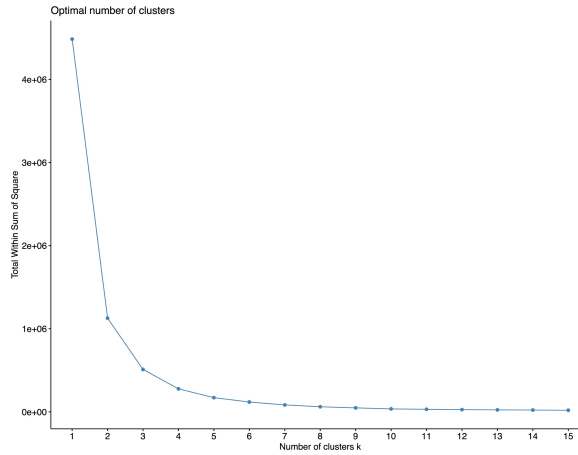


Tabla 7: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de triples para la temporada 1994/1995. Generado con *R Studio*.

cluster	C	PF	PG	SF	SG
1	0	4	19	8	8
2	0	0	4	3	9
3	1	1	3	13	10
4	40	38	5	11	6
5	0	5	17	10	12

Gráfico 16: Gráfico del método *PAM* para el *subset* de triples para la temporada 1994/1995. Generado con *R Studio*.

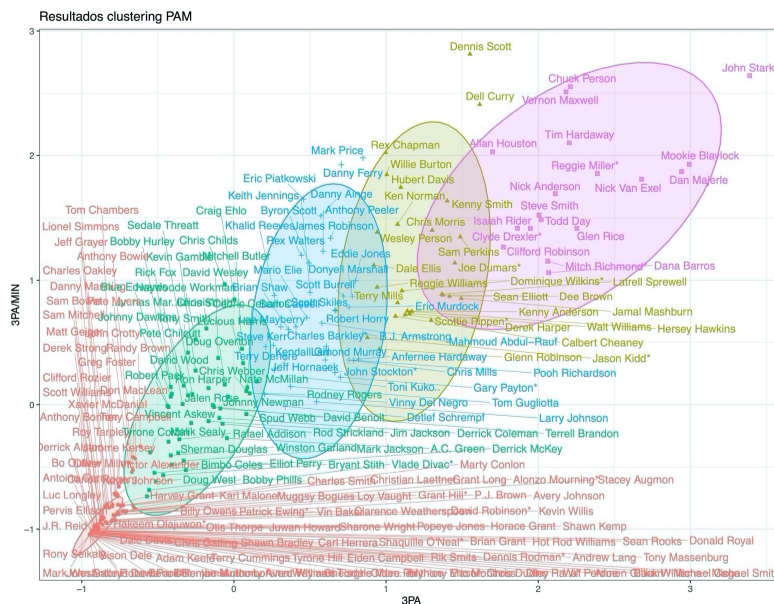
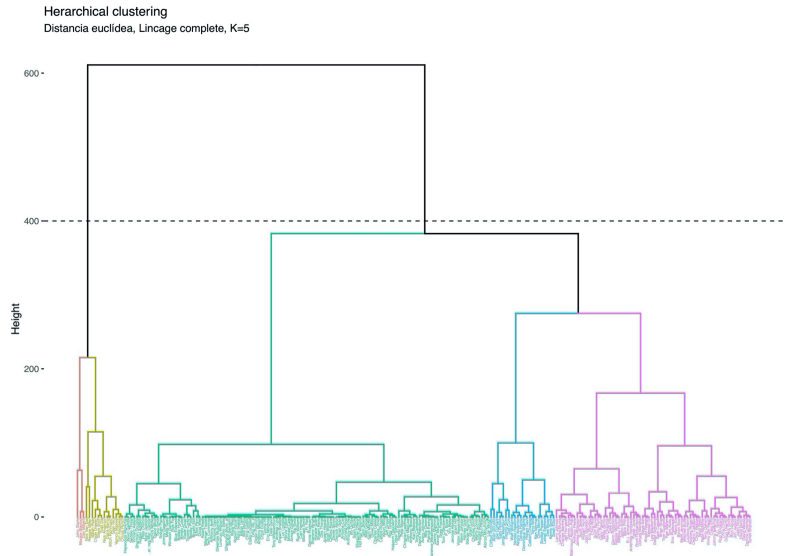


Gráfico 17: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de triples para la temporada 1994/1995. Generado con *R Studio*.



1.2.2. Todos

Gráfico 18: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de todos los tiros para la temporada 1994/1995. Generado con *R Studio*.

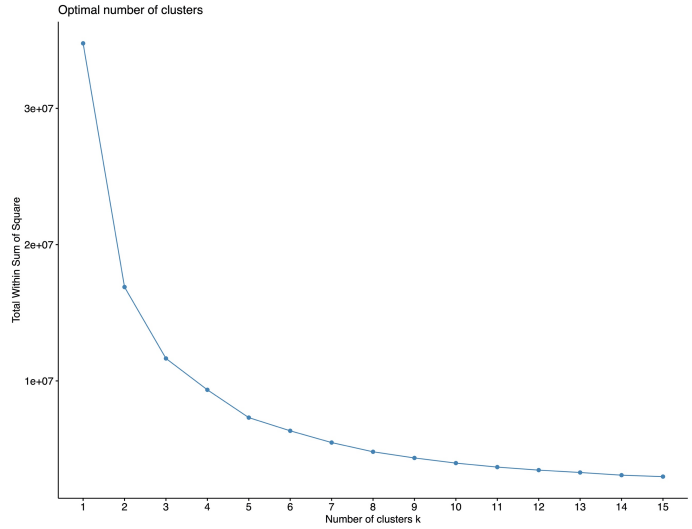


Gráfico 19: Representación del método *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 1994/1995. Generado con *R Studio*.

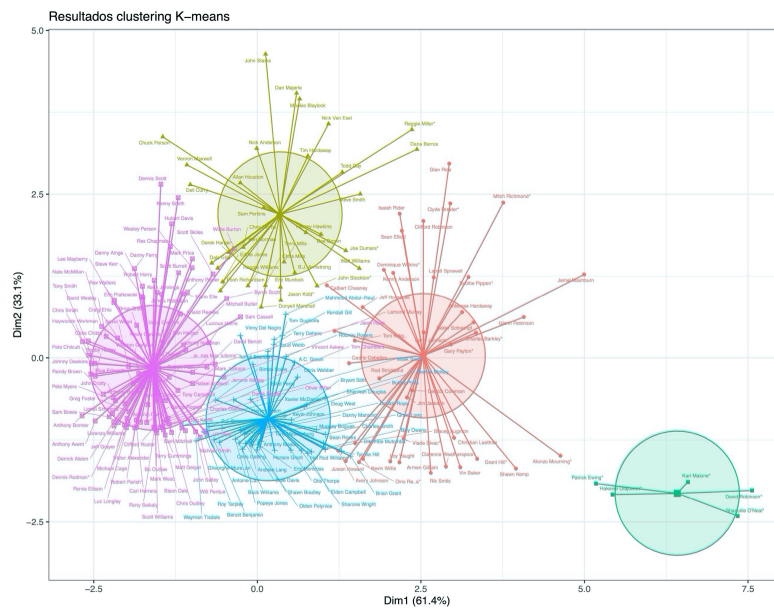
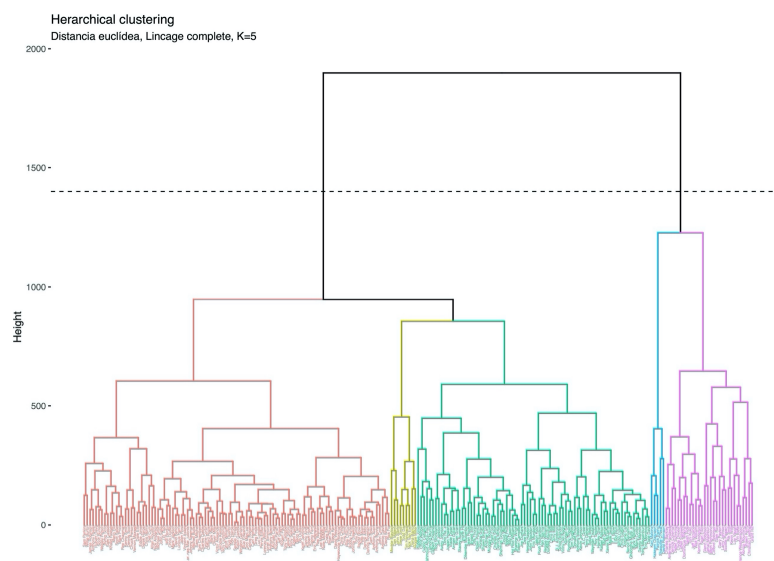


Tabla 8: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 1994/1995. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	4	10	5	13	8
2	1	1	10	9	12
3	4	1	0	0	0
4	15	16	9	6	7
5	17	20	24	17	18

Gráfico 20: Dendrograma generado a partir de un *Hierarchical Clustering* para el *subset* de todos los tiros para la temporada 1994/1995. Generado con *R Studio*.



1.2.3. Por minuto

Gráfico 21: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de los tiros por minuto para la temporada 1994/1995. Generado con *R Studio*.

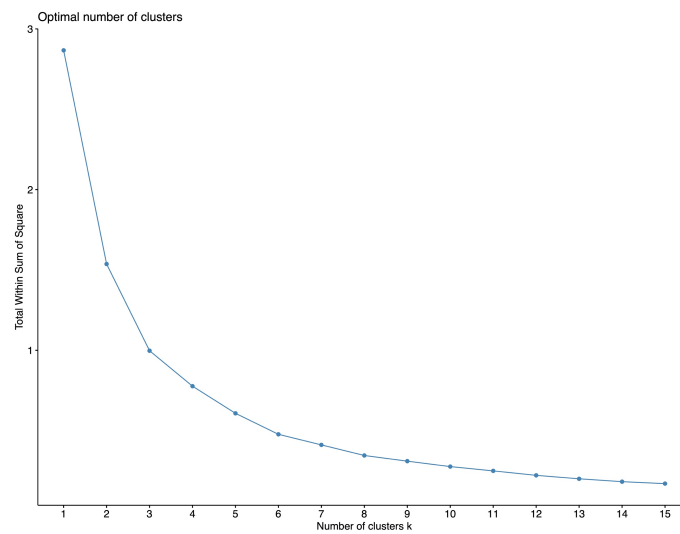


Gráfico 22: Representación del método *K-Means* con $k = 5$ para el *subset* de los tiros por minuto para la temporada 1994/1995. Generado con *R Studio*.

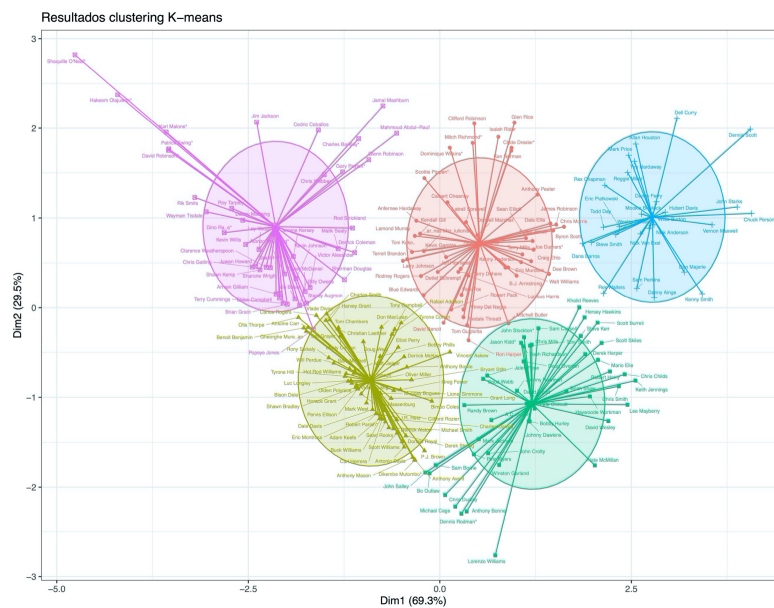
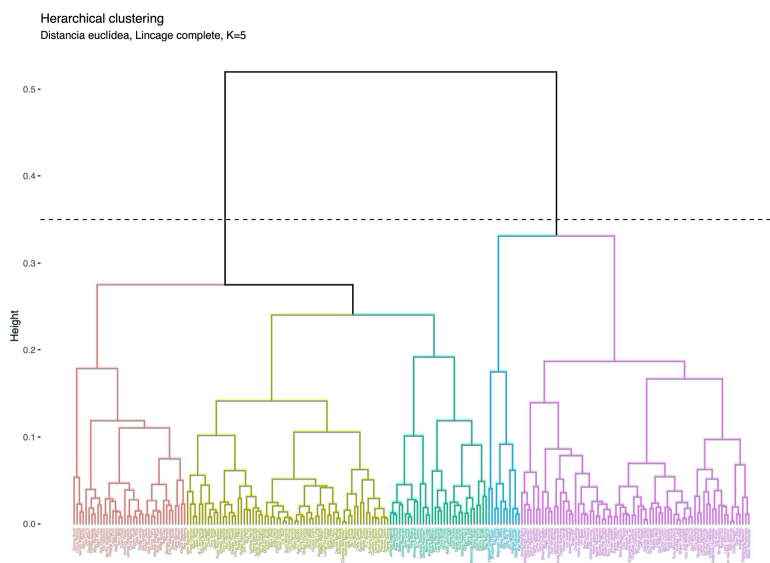


Tabla 9: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de los tiros por minuto para la temporada 1994/1995. Generado con *R Studio*.

cluster	C	PF	PG	SF	SG
1	0	3	10	19	17
2	26	20	4	9	6
3	5	6	23	6	4
4	1	1	6	4	14
5	9	18	5	7	4

Gráfico 23: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de datos de la temporada 1994/1995 y las variables de tiros por minuto. Generado con *R Studio*.



1.2.4. Intentados

Gráfico 24: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de los tiros intentados para la temporada 1994/1995. Generado con *R Studio*.

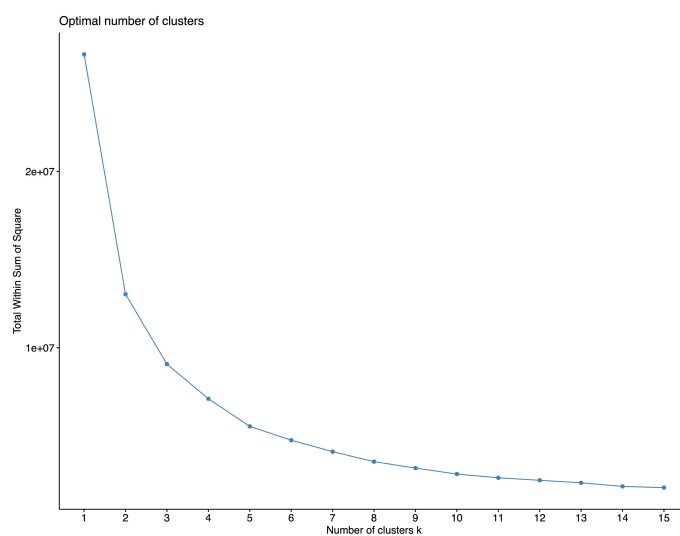


Gráfico 25: Representación del método *K-Means* con $k = 5$ para el *subset* de los tiros intentados para la temporada 1994/1995. Generado con *R Studio*.

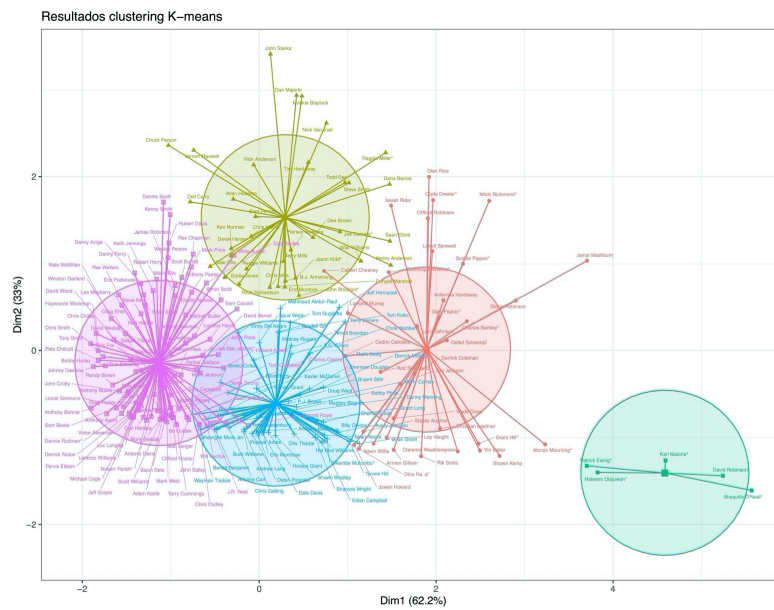


Tabla 10: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de los tiros intentados para la temporada 1994/1995. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	4	10	3	11	7
2	1	1	11	10	12
3	4	1	0	0	0
4	15	16	10	6	8
5	17	20	24	18	18

Gráfico 26: Gráfico del método *PAM* para el *subset* de los tiros intentados para la temporada 1994/1995. Generado con *R Studio*.

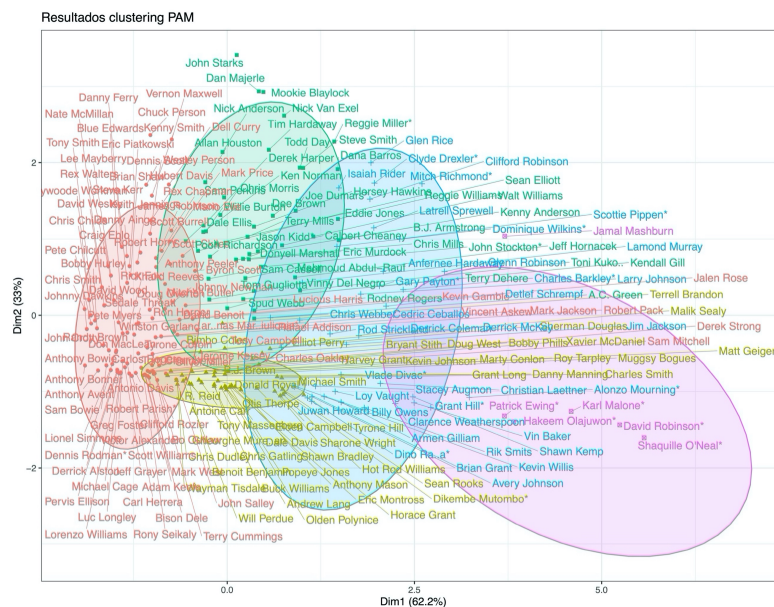


Gráfico 27: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de los tiros intentados para la temporada 1994/1995. Generado con *R Studio*.

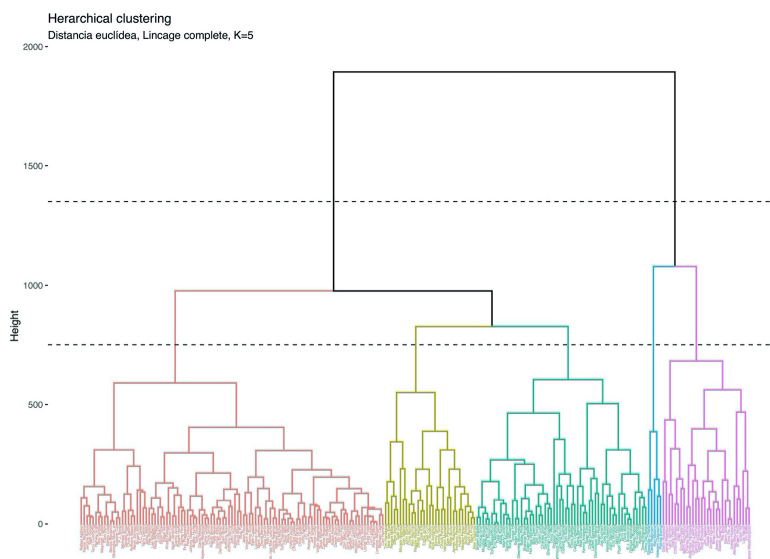


Tabla 11: Tabla de asignación de clusters a los jugadores por posición con la técnica *Hierarchical Clustering* con $k = 5$ para el *subset* de los tiros intentados para la temporada 1994/1995. Generado con *R Studio*.

cluster	grupo real				
	C	PF	PG	SF	SG
1	22	24	25	17	15
2	10	12	11	11	14
3	1	0	8	9	13
4	4	11	4	8	3
5	4	1	0	0	0

Tabla 12: Tabla de asignación de clusters a los jugadores por posición con la técnica *Hierarchical Clustering* con $k = 2$ para el *subset* de los tiros intentados para la temporada 1994/1995. Generado con *R Studio*.

cluster	grupo real	
	EXT	INT
1	123	69
2	15	20

1.3. 2010

1.3.1. Triples

Gráfico 28: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de triples para la temporada 2009/2010. Generado con *R Studio*.

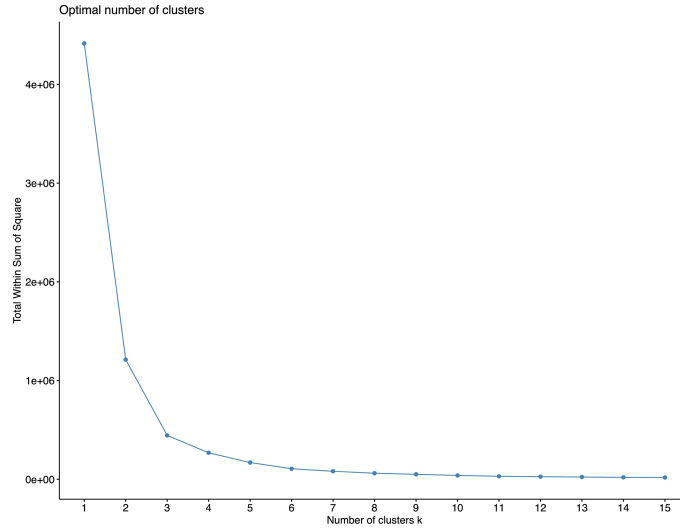


Tabla 13: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de triples para la temporada 2009/2010. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	1	3	26	14	15
2	2	7	12	9	8
3	1	6	6	10	16
4	1	2	6	9	9
5	40	32	7	7	3

Gráfico 29: Gráfico del método *PAM* para el *subset* de triples para la temporada 2009/2010. Generado con *R Studio*.

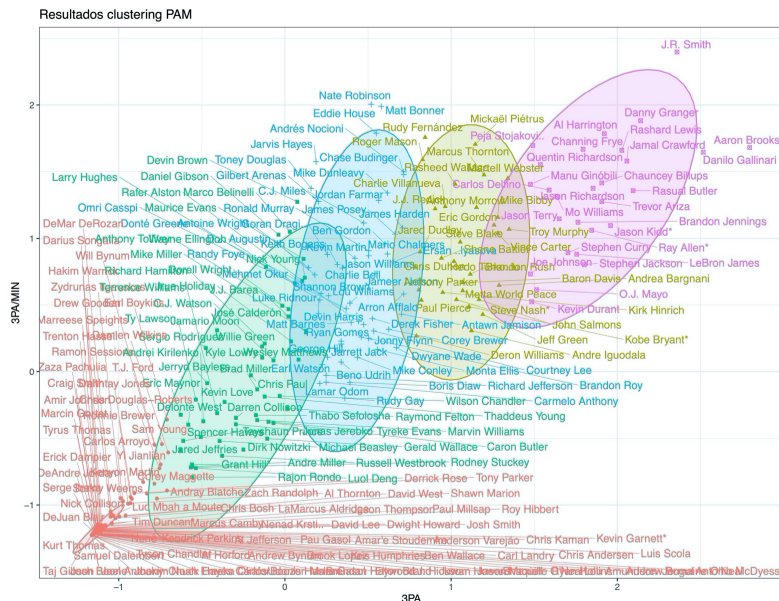
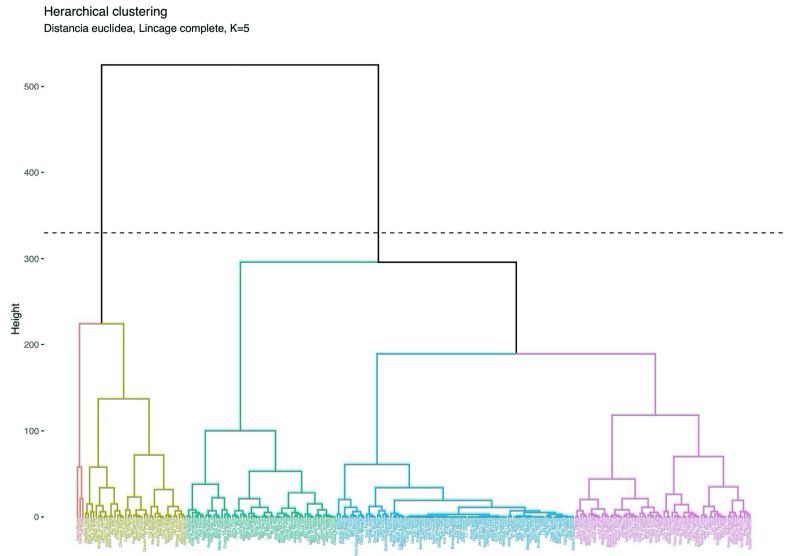


Gráfico 30: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de triples para la temporada 2009/2010. Generado con *R Studio*.



1.3.2. Todos

Gráfico 31: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de todos los tiros para la temporada 2009/2010. Generado con *R Studio*.

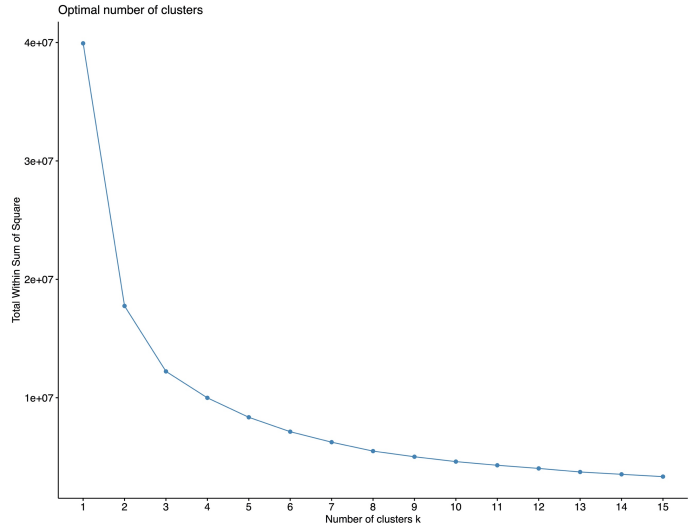


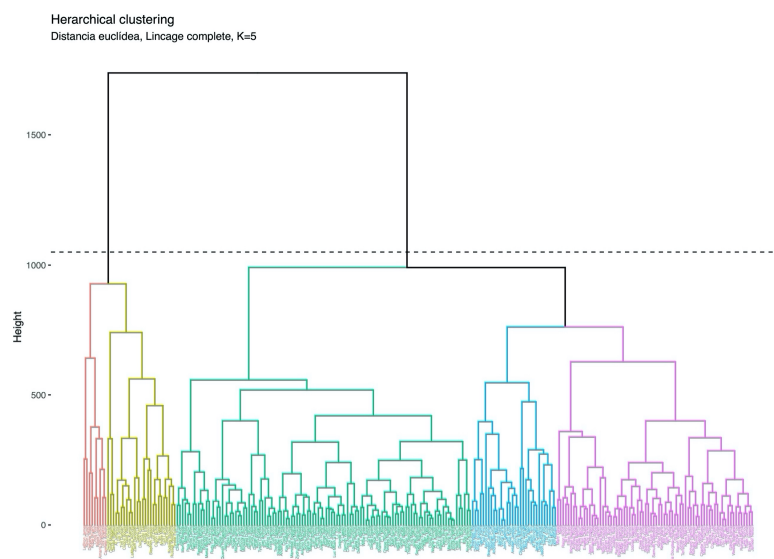
Gráfico 32: Representación del método *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 2009/2010. Generado con *R Studio*.



Tabla 14: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 2009/2010. Generado con *R Studio*.

cluster	C	PF	PG	SF	SG
1	3	3	0	3	2
2	0	5	10	6	13
3	14	10	28	23	22
4	8	10	5	4	5
5	20	22	14	13	9

Gráfico 33: Dendrograma generado a partir de un *Hierarchical Clustering* para el *subset* de todos los tiros para la temporada 2009/2010. Generado con *R Studio*.



1.3.3. Por minuto

Gráfico 34: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de tiros por minuto para la temporada 2009/2010. Generado con *R Studio*.

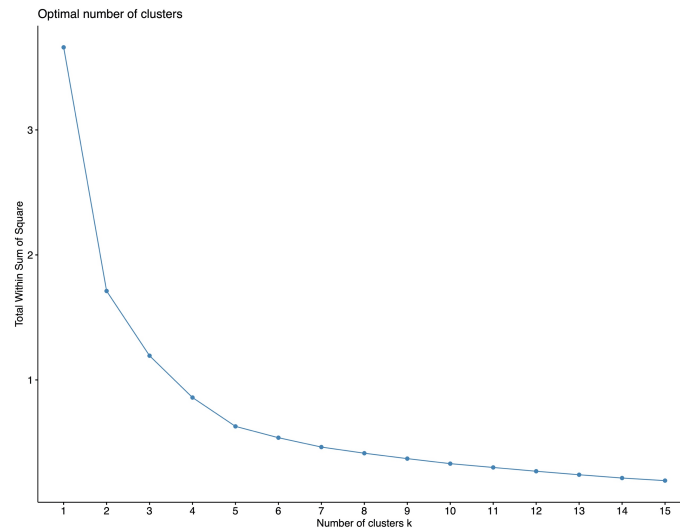


Gráfico 35: Representación del método *K-Means* con $k = 5$ para el *subset* de tiros por minuto para la temporada 2009/2010. Generado con *R Studio*.

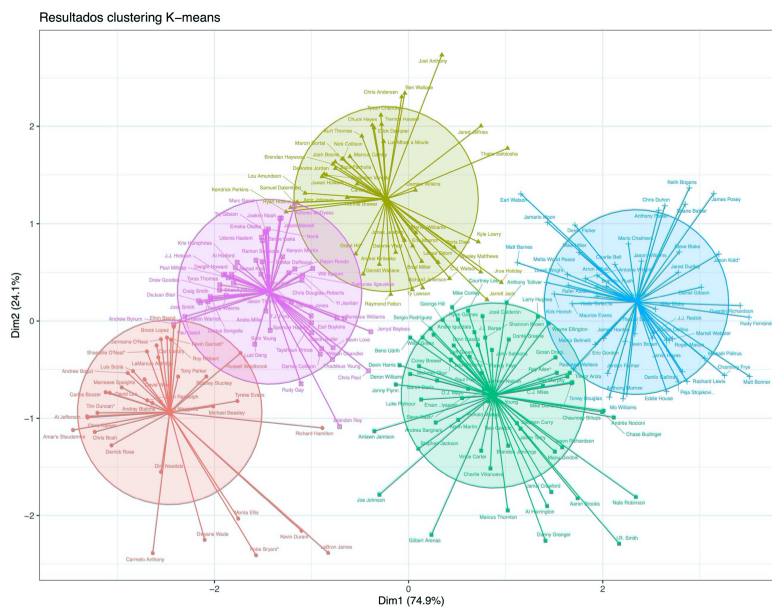


Tabla 15: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de tiros por minuto para la temporada 2009/2010. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	11	13	4	3	5
2	16	11	8	7	4
3	1	8	22	11	23
4	2	2	15	18	14
5	15	16	8	10	5

Gráfico 36: Gráfico del método *PAM* para el *subset* de tiros por minuto para la temporada 2009/2010. Generado con *R Studio*.

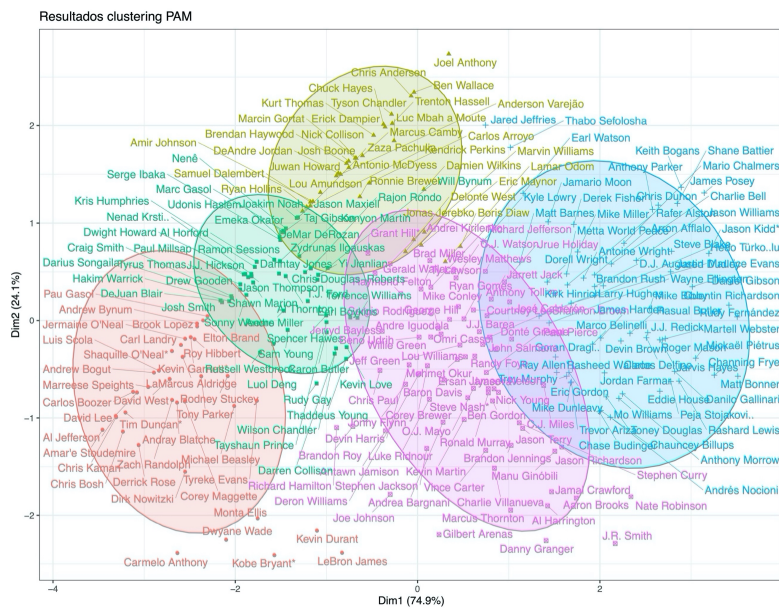
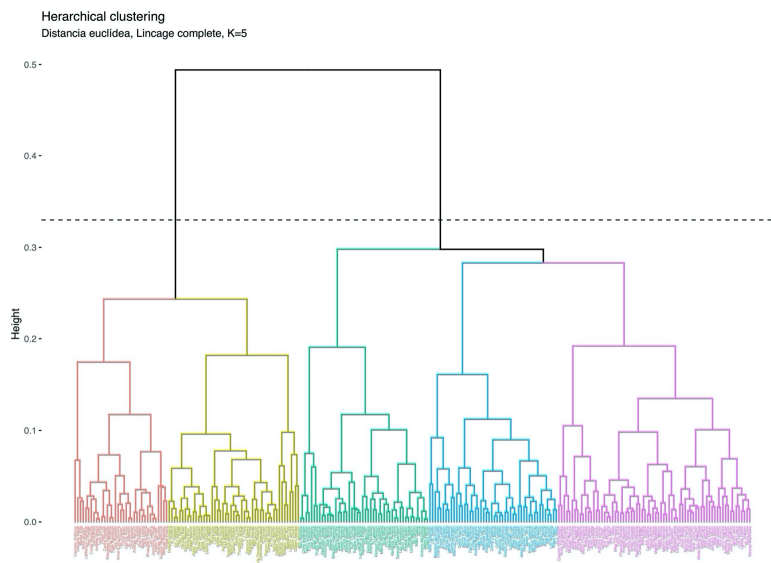


Gráfico 37: Dendrograma generado a partir de un *Hierarchical Clustering* para el *subset* de tiros por minuto para la temporada 2009/2010. Generado con *R Studio*.



1.3.4. Intentados

Gráfico 38: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de tiros intentados para la temporada 2009/2010. Generado con *R Studio*.

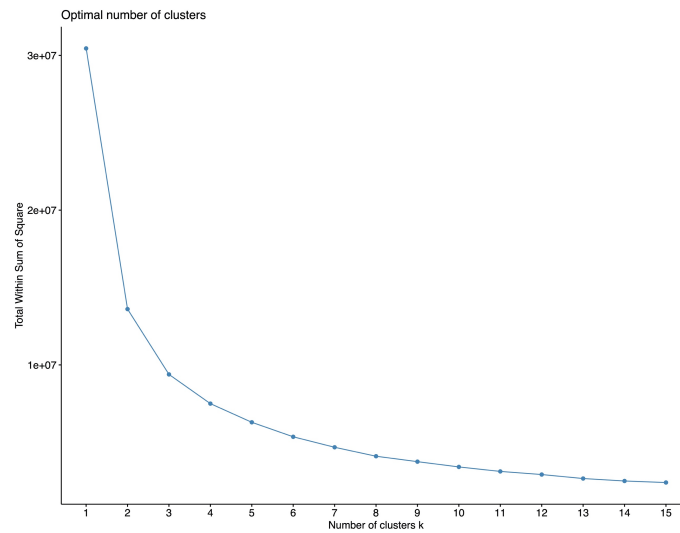


Gráfico 39: Representación del método *K-Means* con $k = 5$ para el *subset* de tiros intentados para la temporada 2009/2010. Generado con *R Studio*.

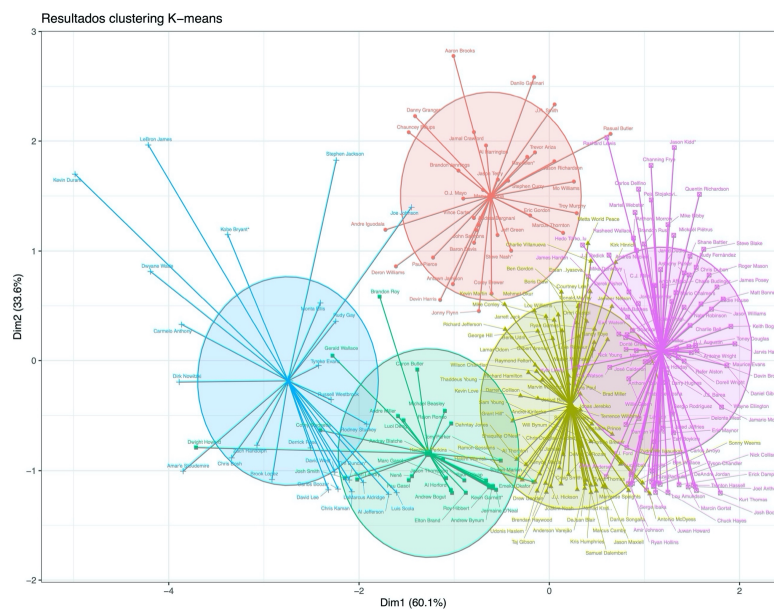


Tabla 16: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de tiros intentados para la temporada 2009/2010. Generado con *R Studio*.

cluster	C	PF	PG	SF	SG
1	0	5	10	6	12
2	14	20	14	13	12
3	11	8	3	4	1
4	6	7	3	4	6
5	14	10	27	22	20

Gráfico 40: Gráfico del método *PAM* para el *subset* de tiros intentados para la temporada 2009/2010. Generado con *R Studio*.

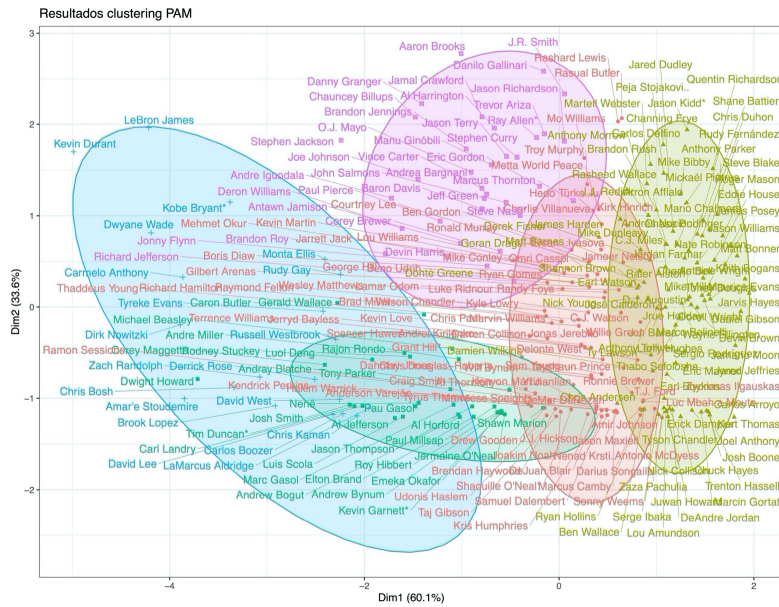


Gráfico 41: Dendograma generado a partir de un *Hierarchical Clustering* para el *subset* de tiros intentados para la temporada 2009/2010. Generado con *R Studio*.

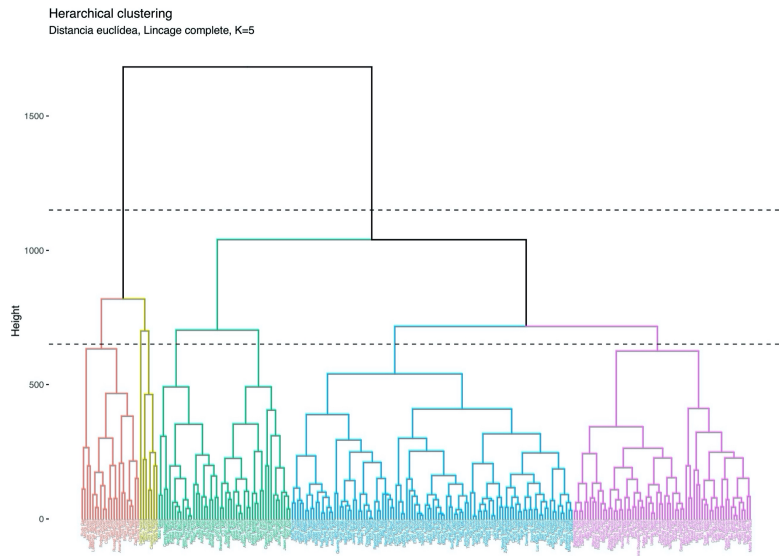


Tabla 17: Tabla de asignación de clusters a los jugadores por posición con la técnica *Hierarchical Clustering* con $k = 5$ para el *subset* de tiros intentados para la temporada 2009/2010. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	18	14	28	22	24
2	10	17	14	15	11
3	10	10	12	8	10
4	6	8	3	1	4
5	1	1	0	3	2

Tabla 18: Tabla de asignación de clusters a los jugadores por posición con la técnica *Hierarchical Clustering* con $k = 2$ para el *subset* de tiros intentados para la temporada 2009/2010. Generado con *R Studio*.

	grupo real	
cluster	EXT	INT
1	144	79
2	13	16

1.4. 2019

1.4.1. Triples

Gráfico 42: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de triples para la temporada 2018/2019. Generado con *R Studio*.

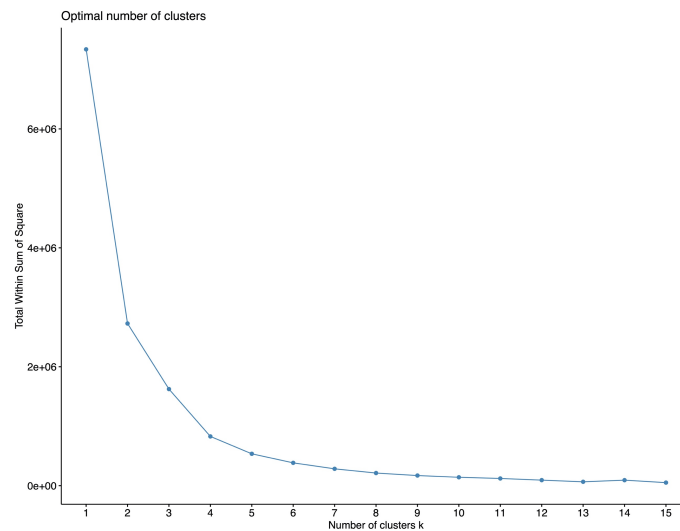


Tabla 19: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de triples para la temporada 2018/2019. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	4	15	8	12	23
2	9	24	26	20	17
3	34	3	8	5	4
4	1	9	11	9	22
5	0	0	5	1	6

1.4.2. Todos

Gráfico 45: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de todos los tiros para la temporada 2018/2019. Generado con *R Studio*.

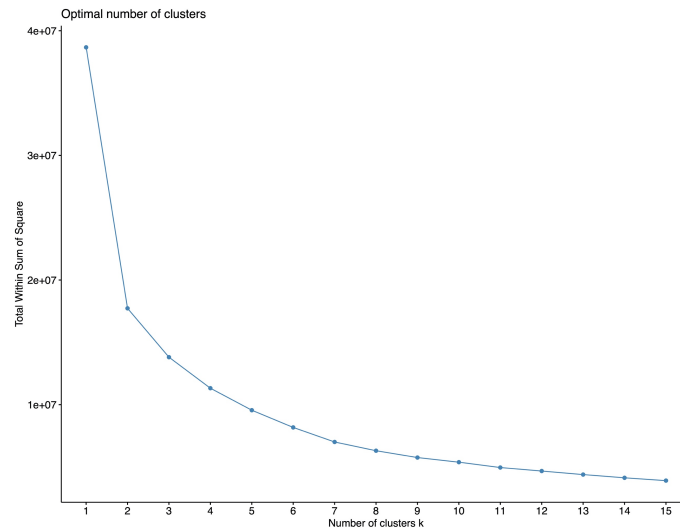


Gráfico 46: Representación del método *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 2018/2019. Generado con *R Studio*.

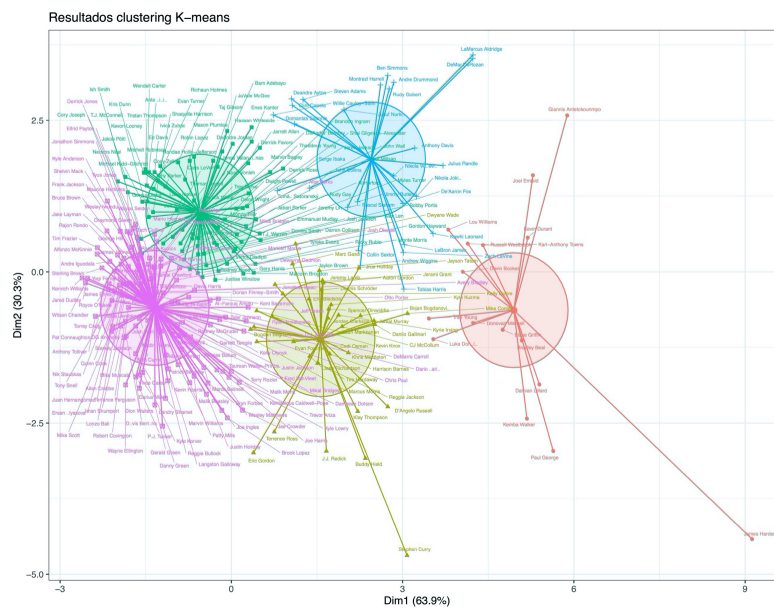


Tabla 20: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de todos los tiros para la temporada 2018/2019. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	2	2	5	2	5
2	1	7	9	6	14
3	28	11	21	5	10
4	14	4	3	4	2
5	3	27	20	30	41

Gráfico 47: Gráfico del método *PAM* para el *subset* de todos los tiros para la temporada 2018/2019. Generado con *R Studio*.

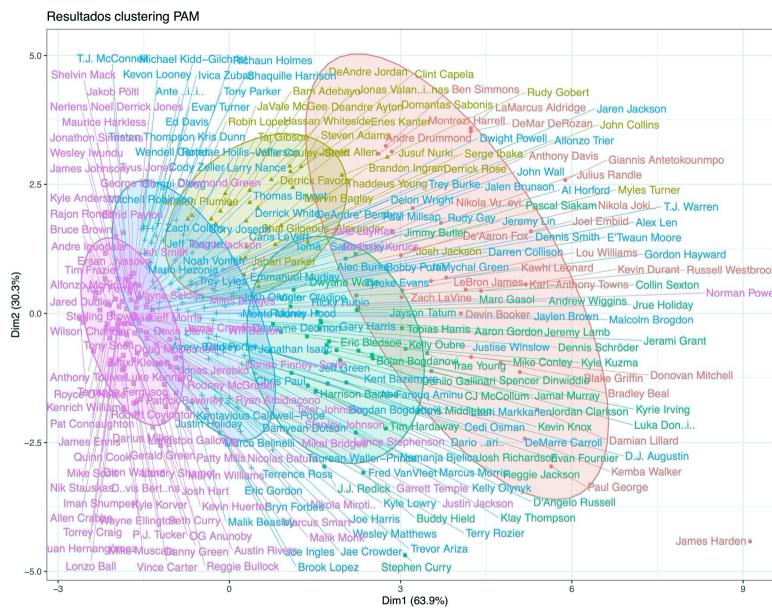
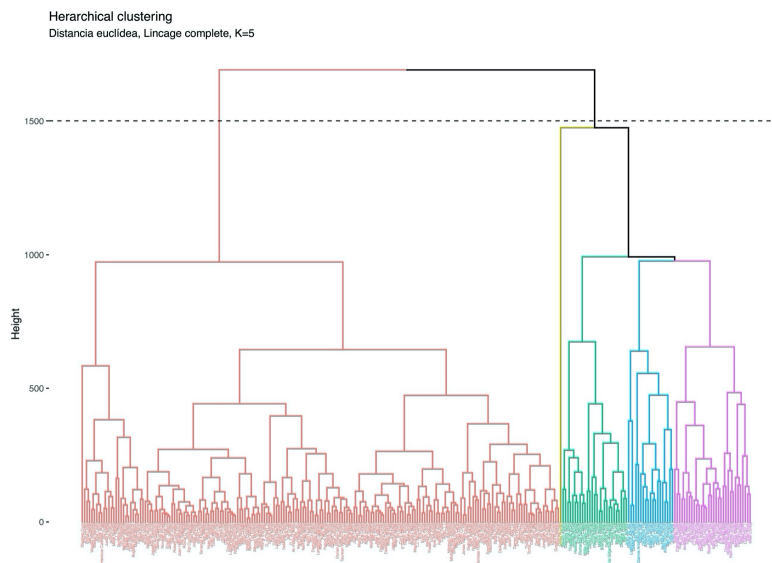


Gráfico 48: Dendrograma generado a partir de un *Hierarchical Clustering* para el *subset* de todos los tiros para la temporada 2018/2019. Generado con *R Studio*.



1.4.3. Por minuto

Gráfico 49: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de tiros por minuto para la temporada 2018/2019. Generado con *R Studio*.

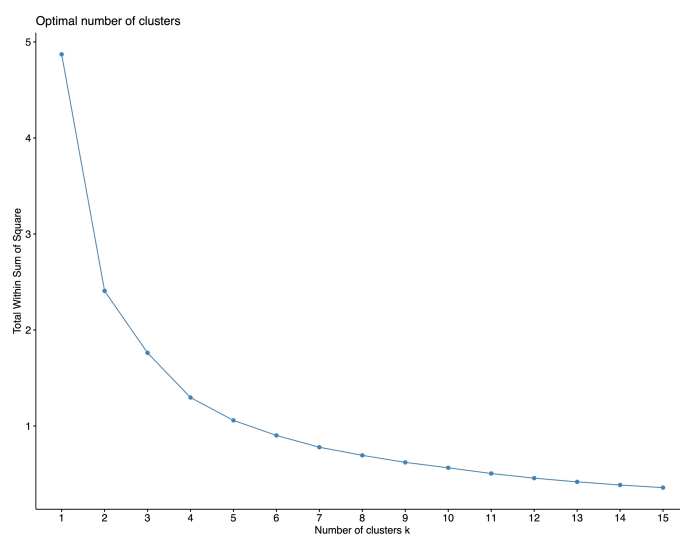


Gráfico 50: Representación del método *K-Means* con $k = 5$ para el *subset* de tiros por minuto para la temporada 2018/2019. Generado con *R Studio*.

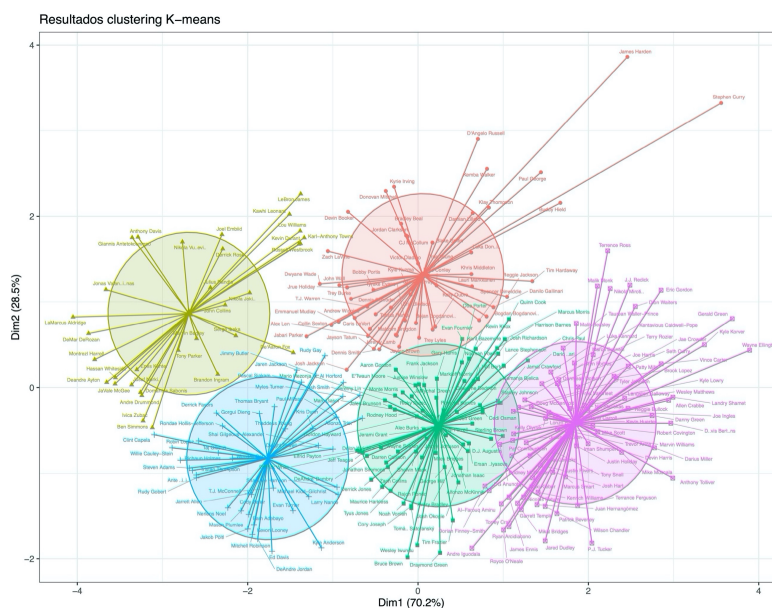


Tabla 21: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de tiros por minuto para la temporada 2018/2019. Generado con *R Studio*.

	grupo real				
cluster	C	PF	PG	SF	SG
1	1	7	18	9	19
2	17	4	5	4	2
3	3	16	21	13	17
4	26	7	5	3	2
5	1	17	9	18	32

Gráfico 51: Gráfico del método *PAM* para el *subset* de tiros por minuto para la temporada 2018/2019. Generado con *R Studio*.

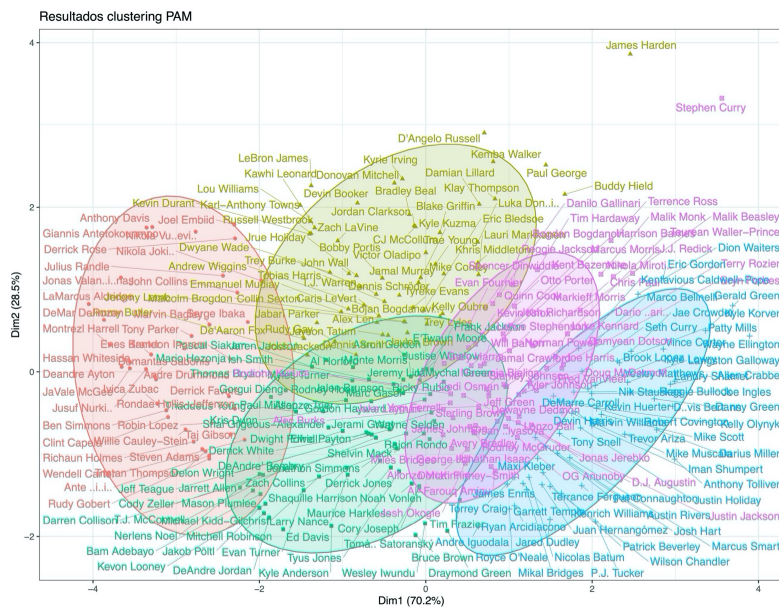
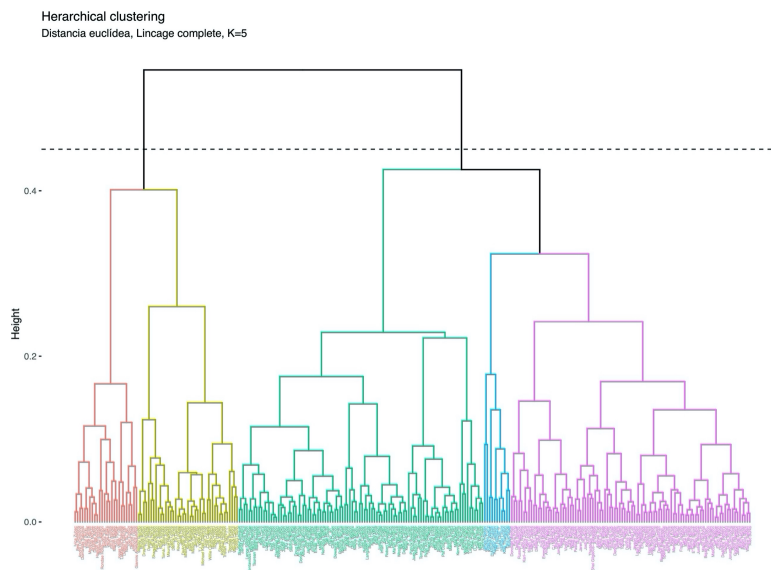


Gráfico 52: Dendrograma generado a partir de un *Hierarchical Clustering* para el *subset* de tiros por minuto para la temporada 2018/2019. Generado con *R Studio*.



1.4.4. Intentados

Gráfico 53: Gráfico de la suma de cuadrados de los errores y los posibles clusters a ser formados para el *subset* de tiros intentados para la temporada 2018/2019. Generado con *R Studio*.

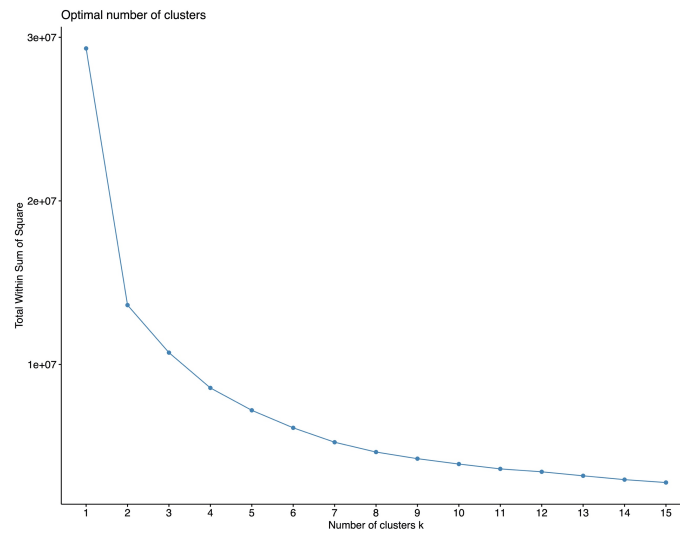


Gráfico 54: Representación del método *K-Means* con $k = 5$ para el *subset* de tiros intentados para la temporada 2018/2019. Generado con *R Studio*.

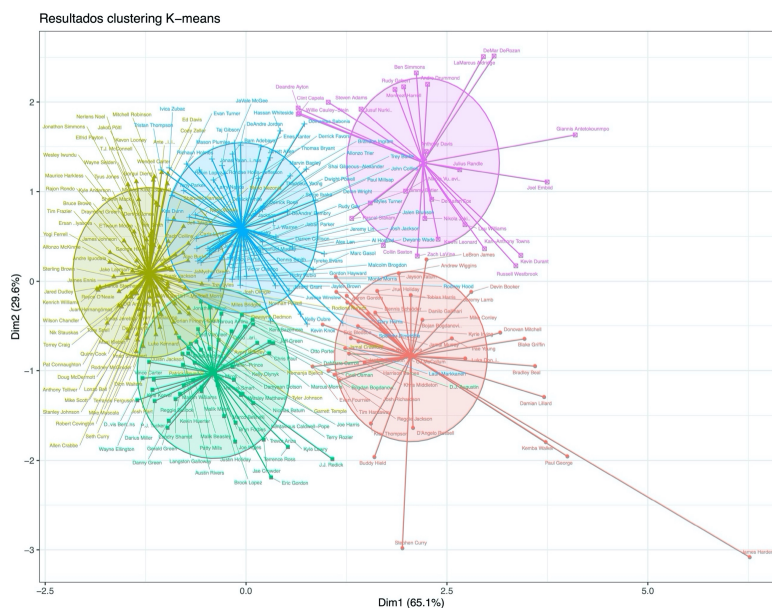
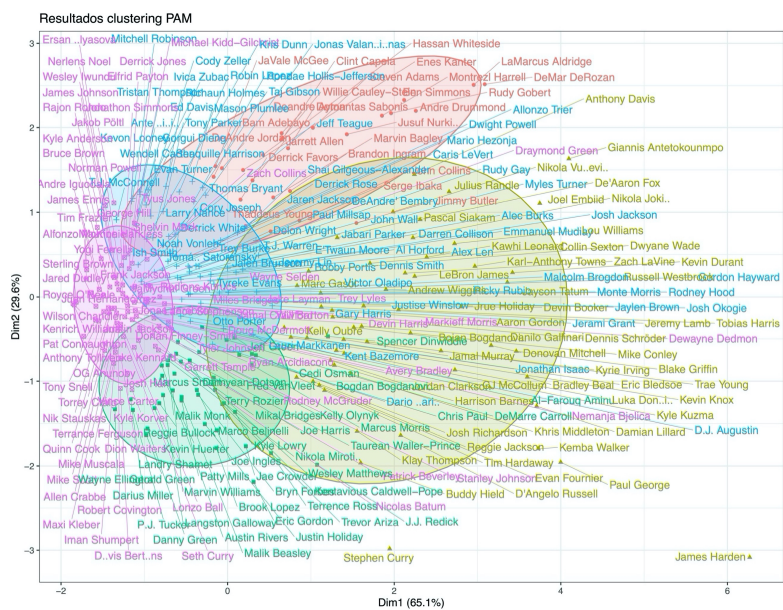


Tabla 22: Tabla de asignación de clusters a los jugadores por posición con la técnica *K-Means* con $k = 5$ para el *subset* de tiros intentados para la temporada 2018/2019. Generado con *R Studio*.

cluster	C	PF	PG	SF	SG
1	0	5	12	7	13
2	11	17	18	23	23
3	1	13	6	9	23
4	22	13	18	5	10
5	14	3	4	3	3

Gráfico 55: Gráfico del método PAM para el subset de tiros intentados para la temporada 2018/2019. Generado con R Studio.



2. CLASIFICACIÓN

2.1. Variables de tiro

2.1.1. 1995

2.1.1.1 Pruebas con misma temporada

Gráfico 56: Gráfico del árbol de decisión generado a partir del modelo para la temporada 1994/1995 para las variables de tiro. Generado con R Studio.

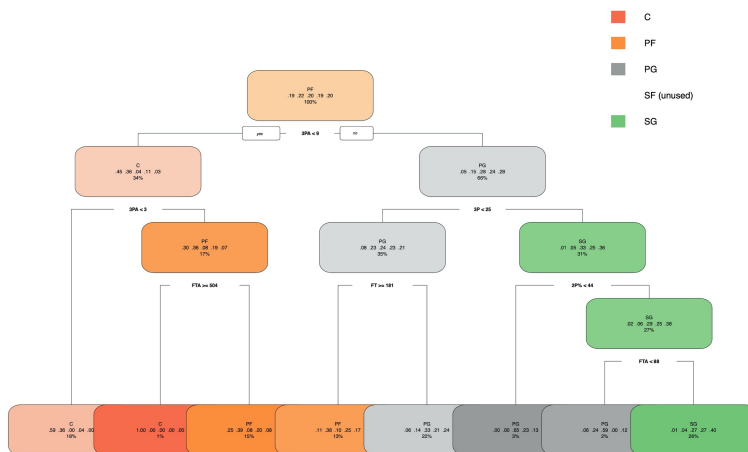


Gráfico 57: Gráfico del *árbol de decisión C5.0* creado a partir del modelo generado para una muestra del 2% de jugadores de la temporada 1994/1995 para las variables de tiro. Generado con *R Studio*.

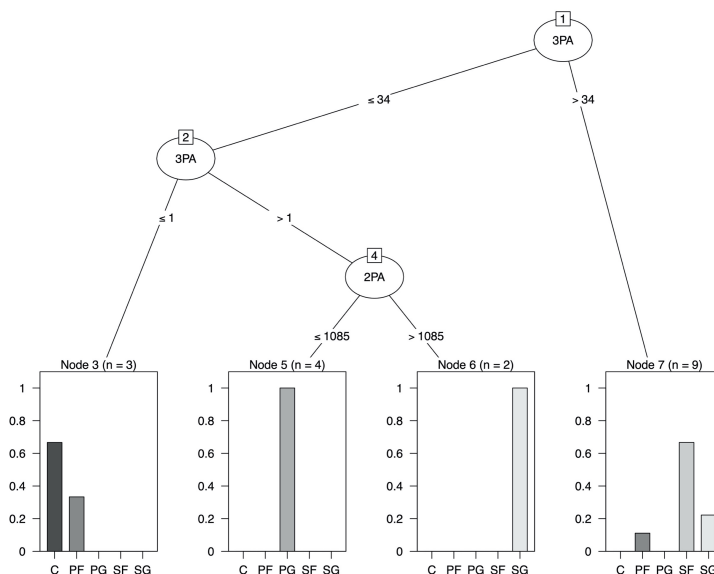


Tabla 23: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 1994/1995. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		19	7	1	1	0
PF		12	25	1	4	5
PG		4	4	25	14	12
SF		3	8	10	9	5
SG		3	4	11	17	23

Overall Statistics

Accuracy : 0.4449
 95% CI : (0.3792, 0.5121)
 No Information Rate : 0.2115
 P-Value [Acc > NIR] : 3.025e-15

2.1.1.2 Pruebas con distintas temporadas

Tabla 24: Matriz de confusión entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 1994/1995. Generada con *Excel*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		22	12	0	0	0
PF		17	16	5	2	0
PG		1	6	26	20	18
SF		0	0	0	0	0
SG		5	16	26	27	33

Overall Statistics

Accuracy : 0.3849
 95% CI : (0.3245, 0.448)
 No Information Rate : 0.2262
 P-Value [Acc > NIR] : 1.082e-08

Tabla 25: Matriz de confusión entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 1994/1995. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 18 12 0 0 0
PF 14 14 5 3 0
PG 4 9 23 21 17
SF 6 2 12 8 12
SG 3 13 17 17 22

Overall Statistics
Accuracy : 0.3373
95% CI : (0.2792, 0.3993)
No Information Rate : 0.2262
P-Value [Acc > NIR] : 3.703e-05
    
```

Tabla 26: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 1994/1995. Generada con *Excel*.

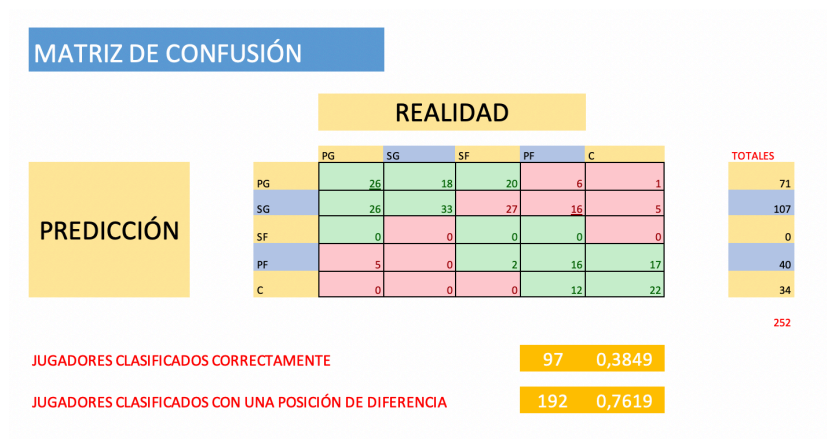


Tabla 27: Matriz de confusión entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 1994/1995. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 13 0 0 0 0
PF 7 0 1 0 1
PG 12 18 21 21 30
SF 0 0 0 0 0
SG 16 33 36 26 41

Overall Statistics
Accuracy : 0.2717
95% CI : (0.2201, 0.3283)
No Information Rate : 0.2609
P-Value [Acc > NIR] : 0.3623
    
```

Tabla 28: Matriz de confusión entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 1994/1995. Generada con *R Studio*.

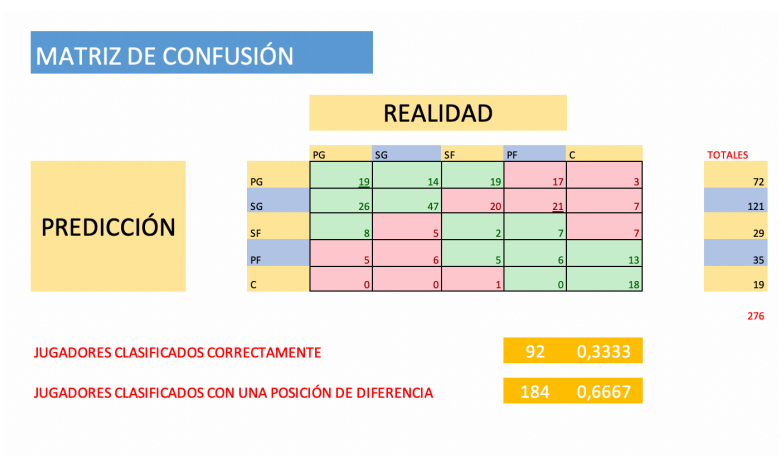
```

Reference
Prediction C PF PG SF SG
C 18 0 0 1 0
PF 13 6 5 5 6
PG 3 17 19 19 14
SF 7 7 8 2 5
SG 7 21 26 20 47

Overall Statistics

Accuracy : 0.3333
95% CI : (0.278, 0.3923)
No Information Rate : 0.2609
P-Value [Acc > NIR] : 0.004461
    
```

Tabla 29: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 1994/1995. Generada con *Excel*.



2.1.2. 2010

2.1.2.1 Pruebas con misma temporada

Tabla 30: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2009/2010. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 24 14 0 1 0
PF 15 13 3 2 1
PG 0 2 26 13 16
SF 3 15 15 19 13
SG 3 6 13 14 21

Overall Statistics

Accuracy : 0.4087
95% CI : (0.3474, 0.4722)
No Information Rate : 0.2262
P-Value [Acc > NIR] : 7.65e-11

Kappa : 0.2595
    
```

Gráfico 58: Gráfico del *árbol de decisión* C5.0 creado a partir del modelo generado para una muestra del 2% de jugadores de la temporada 2009/2010 para las variables de tiro. Generado con *R Studio*.

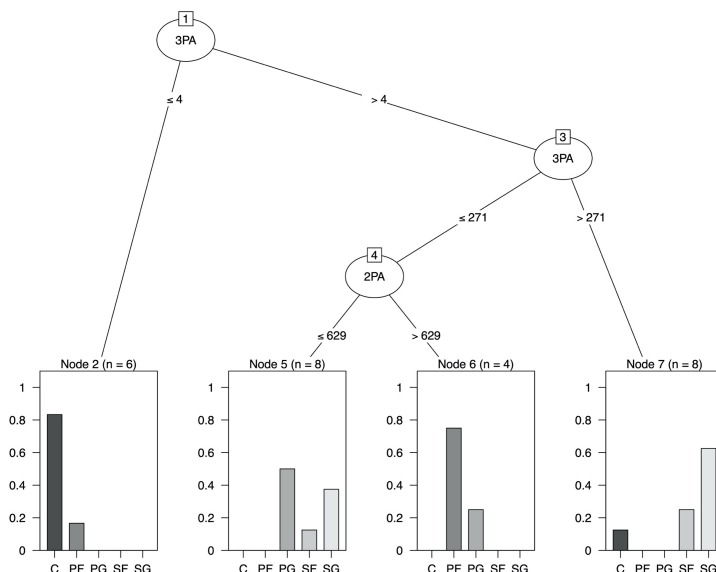


Tabla 31: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 2009/2010. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		27	19	3	3	3
PF		12	16	4	6	2
PG		2	3	15	9	12
SF		2	5	16	17	18
SG		2	7	19	14	16

Overall Statistics

Accuracy : 0.3611
 95% CI : (0.3018, 0.4238)
 No Information Rate : 0.2262
 P-Value [Acc > NIR] : 8.521e-07

Kappa : 0.2028

Tabla 32: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2009/2010. Generada con *Excel*.



2.1.2.1 Pruebas con distintas temporadas

Tabla 33: Matriz de confusión entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2009/2010. Generada con *R Studio*.

		Reference				
		C	PF	PG	SF	SG
Prediction	C	27	25	0	2	0
	PF	10	11	3	3	2
	PG	2	3	13	7	9
	SF	1	6	15	19	11
	SG	1	3	17	14	23

Overall Statistics

Accuracy : 0.4097
 95% CI : (0.3451, 0.4767)
 No Information Rate : 0.2115
 P-Value [Acc > NIR] : 1.185e-11

Tabla 34: Matriz de confusión entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 2009/2010. Generada con *R Studio*.

		Reference				
		C	PF	PG	SF	SG
Prediction	C	23	16	1	8	2
	PF	13	16	7	3	8
	PG	2	4	12	9	10
	SF	1	4	11	13	9
	SG	2	8	17	12	16

Overall Statistics

Accuracy : 0.3524
 95% CI : (0.2904, 0.4184)
 No Information Rate : 0.2115
 P-Value [Acc > NIR] : 7.11e-07

Tabla 35: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2009/2010. Generada con *Excel*.

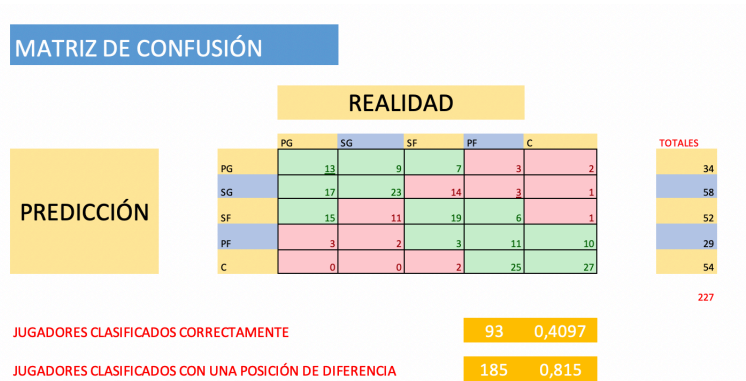


Tabla 36: Matriz de confusión entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2009/2010. Generada con *R Studio*.

```

Prediction C PF PG SF SG
C 14 0 1 0 0
PF 16 2 1 1 1
PG 0 2 10 4 14
SF 10 17 7 9 11
SG 8 30 39 33 46

Overall Statistics

Accuracy : 0.2935
95% CI : (0.2404, 0.351)
No Information Rate : 0.2609
P-Value [Acc > NIR] : 0.1227

Kappa : 0.0859
    
```

Tabla 37: Matriz de confusión entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 2009/2010. Generada con *R Studio*.

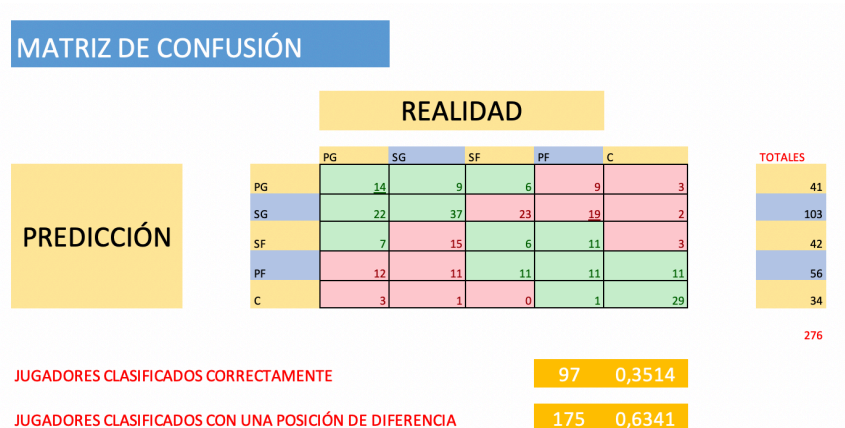
```

Reference
Prediction C PF PG SF SG
C 29 1 3 1 0
PF 11 11 12 11 11
PG 3 9 14 6 9
SF 3 11 7 6 15
SG 2 19 22 23 37

Overall Statistics

Accuracy : 0.3514
95% CI : (0.2952, 0.4109)
No Information Rate : 0.2609
P-Value [Acc > NIR] : 0.0005486
    
```

Tabla 38: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 2009/2010. Generada con *Excel*.



2.1.3. 2019

2.1.3.1 Pruebas con misma temporada

Tabla 39: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2018/2019. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 24 0 1 0 0
PF 21 18 11 8 8
PG 1 1 17 3 12
SF 2 14 5 18 12
SG 0 18 24 18 40

Overall Statistics

Accuracy : 0.4239
95% CI : (0.3649, 0.4846)
No Information Rate : 0.2609
P-Value [Acc > NIR] : 3.168e-09

Kappa : 0.2691
    
```

Tabla 40: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2018/2019. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	17	12	3	1	1	34
	SG	24	40	18	14	2	98
	SF	5	12	18	18	0	53
	PF	11	8	8	18	21	66
	C	1	0	0	0	24	25
							276

JUGADORES CLASIFICADOS CORRECTAMENTE	117	0,42391
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA	208	0,75362

Gráfico 59: Gráfico del *árbol de decisión* creado a partir del modelo generado para la temporada 2018/2019 para las variables de tiro. Generado con *R Studio*.

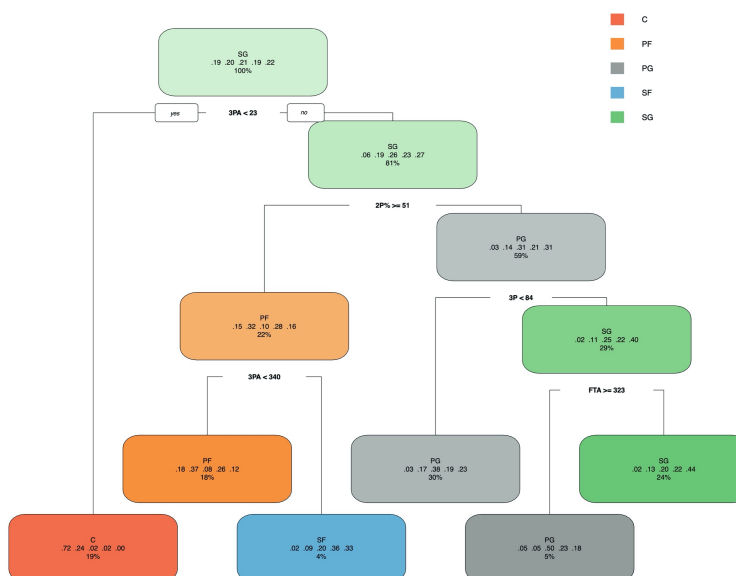


Tabla 41: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 2018/2019. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		32	7	3	5	2
PF		9	16	15	13	12
PG		0	6	23	7	13
SF		6	14	7	11	9
SG		1	8	10	11	36

Overall Statistics

Accuracy : 0.4275
 95% CI : (0.3684, 0.4882)
 No Information Rate : 0.2609
 P-Value [Acc > NIR] : 1.488e-09
 Kappa : 0.2817

Tabla 42: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 2018/2019. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	23	13	7	6	0	49
	SG	10	36	11	8	1	66
	SF	7	9	11	14	6	47
	PF	15	12	13	16	9	65
	C	3	2	5	7	32	49

276

JUGADORES CLASIFICADOS CORRECTAMENTE	118	0,42754
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA	198	0,71739

2.1.3.2 Pruebas con distintas temporadas

Tabla 43: Matriz de confusión entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2018/2019. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		34	30	1	6	2
PF		4	6	8	9	4
PG		2	9	30	21	26
SF		1	0	0	0	2
SG		0	3	9	9	11

Overall Statistics

Accuracy : 0.3568
 95% CI : (0.2945, 0.4229)
 No Information Rate : 0.2115
 P-Value [Acc > NIR] : 3.403e-07

Tabla 44: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2018/2019. Generada con *Excel*.

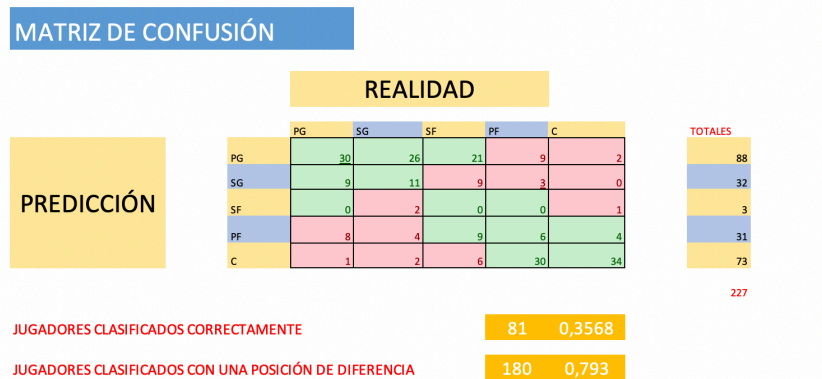


Tabla 45: Matriz de confusión entre la posición real de los datos de la temporada 2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2018/2019. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		39	26	1	2	2
PF		1	5	13	6	4
PG		4	13	32	27	24
SF		0	0	0	1	2
SG		1	6	11	13	19

Overall Statistics

Accuracy : 0.381
 95% CI : (0.3207, 0.444)
 No Information Rate : 0.2262
 P-Value [Acc > NIR] : 2.333e-08

Tabla 46: Matriz de confusión entre la posición real de los datos de la temporada 2010 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de tiro en la temporada 2018/2019. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 29 18 3 5 2
PF 11 15 11 5 6
PG 2 4 21 8 19
SF 2 6 7 12 15
SG 1 7 15 19 9

Overall Statistics

Accuracy : 0.3413
95% CI : (0.2829, 0.4034)
No Information Rate : 0.2262
P-Value [Acc > NIR] : 2.059e-05
    
```

Tabla 47: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de tiro en la temporada 2018/2019. Generada con *Excel*.



2.2. Variables de estadística avanzada

2.2.1. 1995

2.2.1.1 pruebas con misma temporada

Tabla 48: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 18 11 0 1 0
PF 22 31 0 10 0
PG 0 0 41 1 3
SF 1 5 2 20 14
SG 0 2 5 12 28

Overall Statistics

Accuracy : 0.6079
95% CI : (0.5411, 0.6719)
No Information Rate : 0.2159
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 49: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *Excel*.

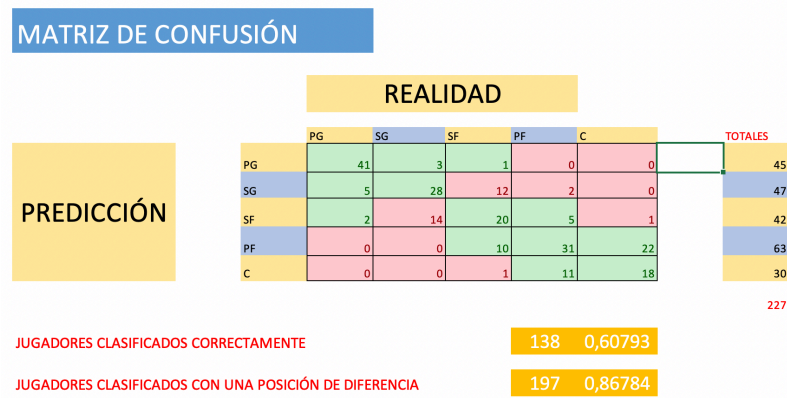


Gráfico 60: Gráfico del *árbol de decisión* creado a partir del modelo generado para la temporada 1994/1995 para las variables de estadística avanzada. Generado con *R Studio*.

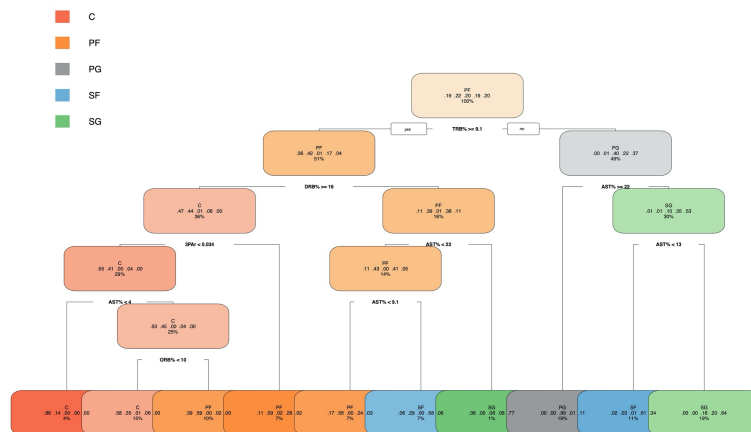


Gráfico 61: Gráfico del *árbol de decisión* C5.0 creado a partir del modelo generado para una muestra del 2% de jugadores de la temporada 1994/1995 para las variables de estadística avanzada. Generado con *R Studio*.

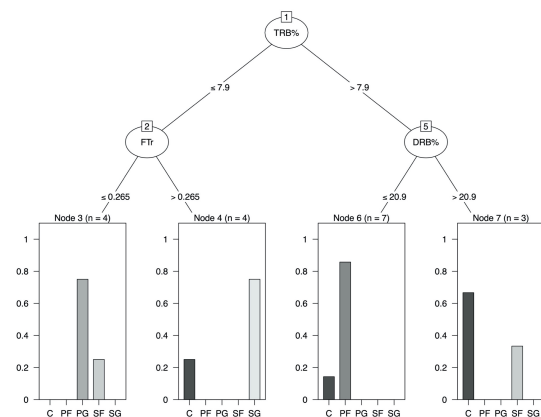


Tabla 50: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 24 15 0 1 0
PF 16 24 0 9 1
PG 0 0 41 1 5
SF 1 9 2 20 10
SG 0 1 5 13 29

Overall Statistics

Accuracy : 0.6079
95% CI : (0.5411, 0.6719)
No Information Rate : 0.2159
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 51: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *Excel*.

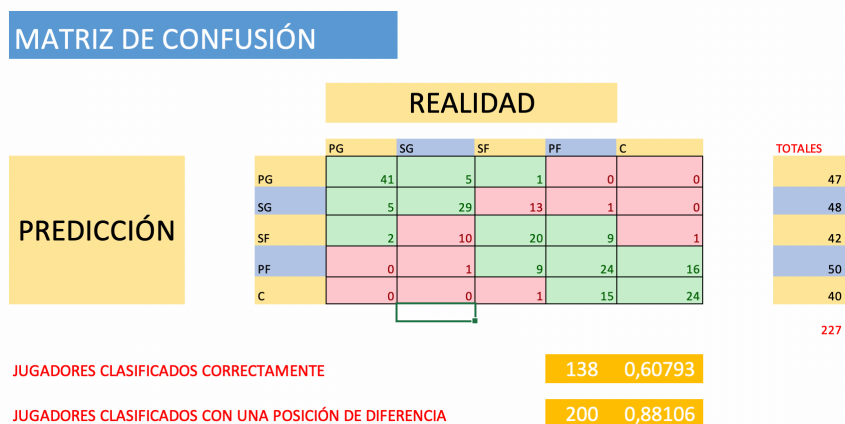


Tabla 52: Matriz de confusión ajustada 2 entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *Excel*.



2.2.1.2 pruebas con distintas temporadas

Tabla 53: Matriz de confusión entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		25	15	0	0	0
PF		19	32	0	14	3
PG		0	0	46	0	8
SF		1	2	2	28	22
SG		0	1	9	7	18

Overall Statistics

Accuracy : 0.5913
 95% CI : (0.5278, 0.6526)
 No Information Rate : 0.2262
 P-Value [Acc > NIR] : < 2.2e-16

Tabla 54: Matriz de confusión entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		20	18	0	0	0
PF		22	21	0	15	0
PG		0	0	41	1	5
SF		3	10	2	16	17
SG		0	1	14	17	29

Overall Statistics

Accuracy : 0.504
 95% CI : (0.4405, 0.5673)
 No Information Rate : 0.2262
 P-Value [Acc > NIR] : < 2.2e-16

Tabla 55: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *Excel*.



Tabla 56: Matriz de confusión ajustada 2 entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *Excel*.

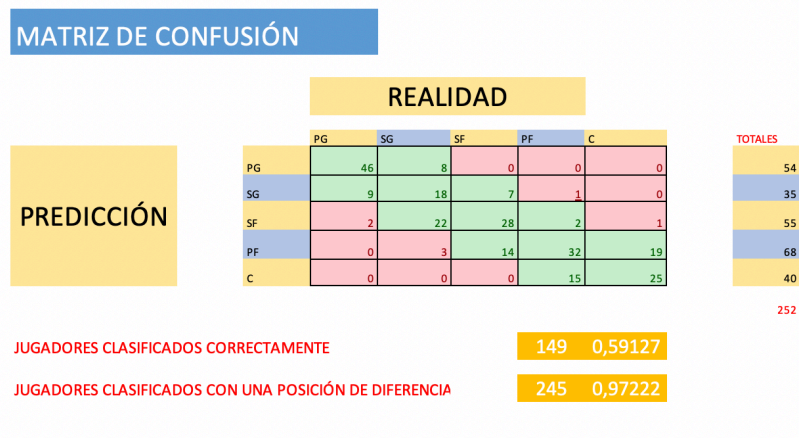


Tabla 57: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 5 0 1 0 0
PF 41 37 4 22 7
PG 0 1 34 1 7
SF 2 12 2 18 40
SG 0 1 17 6 18

Overall Statistics

Accuracy : 0.4058
95% CI : (0.3473, 0.4663)
No Information Rate : 0.2609
P-Value [Acc > NIR] : 1.117e-07
    
```

Tabla 58: Matriz de confusión entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 26 2 0 1 0
PF 18 32 3 24 8
PG 0 0 34 3 9
SF 3 7 3 5 15
SG 1 10 18 14 40

Overall Statistics

Accuracy : 0.4964
95% CI : (0.4359, 0.5569)
No Information Rate : 0.2609
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 59: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	34	9	3	0	0	46
	SG	18	40	14	10	1	83
	SF	3	15	5	7	3	33
	PF	3	8	24	32	18	85
	C	0	0	1	2	26	29
							276

JUGADORES CLASIFICADOS CORRECTAMENTE	137	0,4964
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA	221	0,8007

Tabla 60: Matriz de confusión ajustada 2 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 1994/1995. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	34	9	3	0	0	46
	SG	18	40	14	10	1	83
	SF	3	15	5	7	3	33
	PF	3	8	24	32	18	85
	C	0	0	1	2	26	29
							276

JUGADORES CLASIFICADOS CORRECTAMENTE	137	0,49638
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA	244	0,88406

2.2.2. 2010

2.2.2.1 Pruebas con misma temporada

Tabla 61: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 37 26 0 1 0
PF 5 13 0 6 1
PG 0 0 42 0 2
SF 3 11 0 31 9
SG 0 0 15 11 39

Overall Statistics

Accuracy : 0.6429
95% CI : (0.5803, 0.702)
No Information Rate : 0.2262
    
```


Gráfico 63: Gráfico del árbol de decisión C5.0 creado a partir del modelo generado para una muestra del 2% de jugadores de la temporada 2009/2010 para las variables de estadística avanzada. Generado con R Studio.

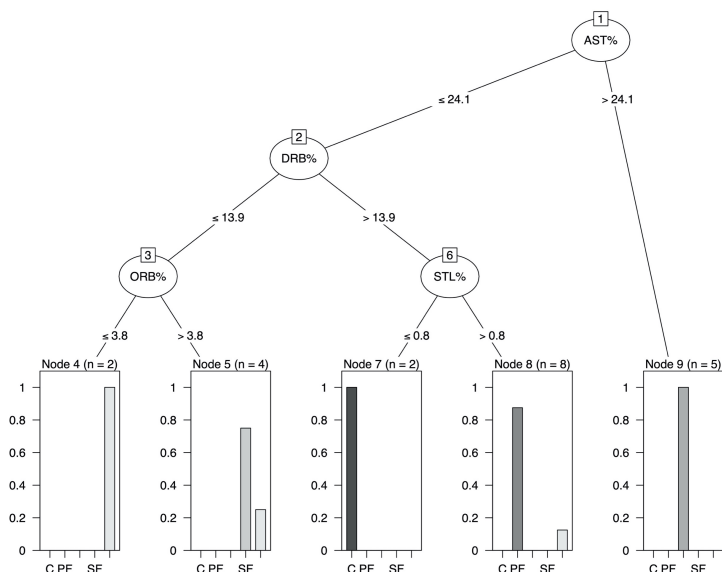
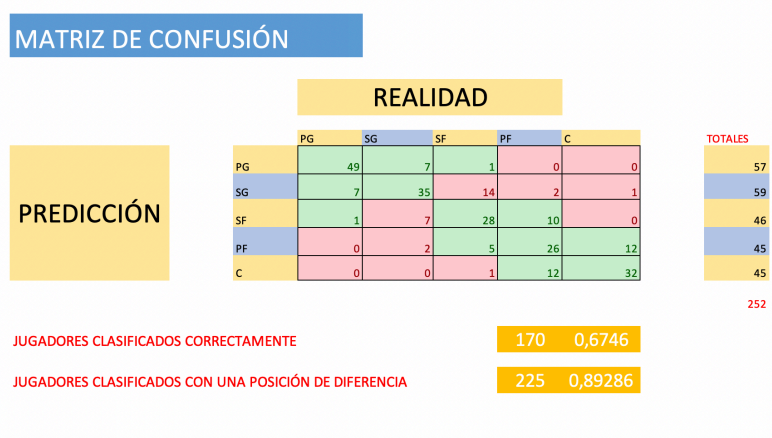


Tabla 63: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2009/2010. Generada con R Studio.

		Reference				
		C	PF	PG	SF	SG
Prediction	C	32	12	0	1	0
	PF	12	26	0	5	2
	PG	0	0	49	1	7
	SF	0	10	1	28	7
	SG	1	2	7	14	35
	SG	1	2	7	14	35

Overall Statistics	
Accuracy	: 0.6746
95% CI	: (0.613, 0.7321)
No Information Rate	: 0.2262
P-Value [Acc > NIR]	: < 2.2e-16

Tabla 64: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2009/2010. Generada con Excel.



2.2.2.2 Pruebas con distintas temporadas

Tabla 65: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		27	28	0	2	1
PF		10	11	0	11	0
PG		0	0	37	1	3
SF		4	9	0	15	5
SG		0	1	11	15	36

Overall Statistics

Accuracy : 0.5551
 95% CI : (0.4879, 0.6208)
 No Information Rate : 0.2159
 P-Value [Acc > NIR] : < 2.2e-16

Tabla 66: Matriz de confusión entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *R Studio*.

		Reference				
Prediction		C	PF	PG	SF	SG
C		24	14	0	3	0
PF		15	26	0	6	0
PG		0	0	42	1	8
SF		2	8	0	20	11
SG		0	1	6	14	26

Overall Statistics

Accuracy : 0.6079
 95% CI : (0.5411, 0.6719)
 No Information Rate : 0.2159
 P-Value [Acc > NIR] : < 2.2e-16

Tabla 67: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *Excel*.



Tabla 68: Matriz de confusión ajustada 2 entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	42	8	1	0	0	51
	SG	6	26	14	1	0	47
	SF	0	11	20	8	2	41
	PF	0	0	6	26	15	47
	C	0	0	3	14	24	41
							227
JUGADORES CLASIFICADOS CORRECTAMENTE		138		0,60793			
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA		220		0,96916			

Tabla 69: Matriz de confusión entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 22 0 0 0 0
PF 23 30 3 10 3
PG 0 1 27 1 6
SF 3 18 5 25 13
SG 0 2 23 11 50

Overall Statistics

Accuracy : 0.558
95% CI : (0.4972, 0.6175)
No Information Rate : 0.2609
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 70: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 31 4 0 0 0
PF 14 24 3 8 5
PG 0 2 43 2 16
SF 3 15 2 18 13
SG 0 6 10 19 38

Overall Statistics

Accuracy : 0.558
95% CI : (0.4972, 0.6175)
No Information Rate : 0.2609
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 71: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	27	6	1	1	0	35
	SG	23	50	11	2	0	86
	SF	5	13	25	18	3	64
	PF	3	3	10	30	23	69
	C	0	0	0	0	22	22
						276	

JUGADORES CLASIFICADOS CORRECTAMENTE	154	0,558
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA	240	0,8696

Tabla 72: Matriz de confusión ajustada 2 entre la posición real de los datos de la temporada 2018/2019 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2009/2010. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	27	6	1	1	0	35
	SG	23	50	11	2	0	86
	SF	5	13	25	18	3	64
	PF	3	3	10	30	23	69
	C	0	0	0	0	22	22
						276	

JUGADORES CLASIFICADOS CORRECTAMENTE	154	0,55797
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA	258	0,93478

2.2.3. 2019

2.2.3.1 Pruebas con misma temporada

Tabla 73: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 32 2 2 0 0
PF 14 32 1 10 1
PG 0 1 43 4 10
SF 2 16 3 28 36
SG 0 0 9 5 25

Overall Statistics

Accuracy : 0.5797
95% CI : (0.5191, 0.6386)
No Information Rate : 0.2609

```


Tabla 74: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *Excel*.

		REALIDAD					TOTALES
		PG	SG	SF	PF	C	
PREDICCIÓN	PG	43	10	4	1	0	58
	SG	9	25	5	0	0	39
	SF	3	36	28	16	2	85
	PF	1	1	10	32	14	58
	C	2	0	0	2	32	36
						276	

JUGADORES CLASIFICADOS CORRECTAMENTE	160	0,57971
JUGADORES CLASIFICADOS CON UNA POSICIÓN DE DIFERENCIA	228	0,82609

Gráfico 64: Gráfico del *árbol de decisión* creado a partir del modelo generado para la temporada 2018/2019 para las variables de estadística avanzada. Generado con *R Studio*.

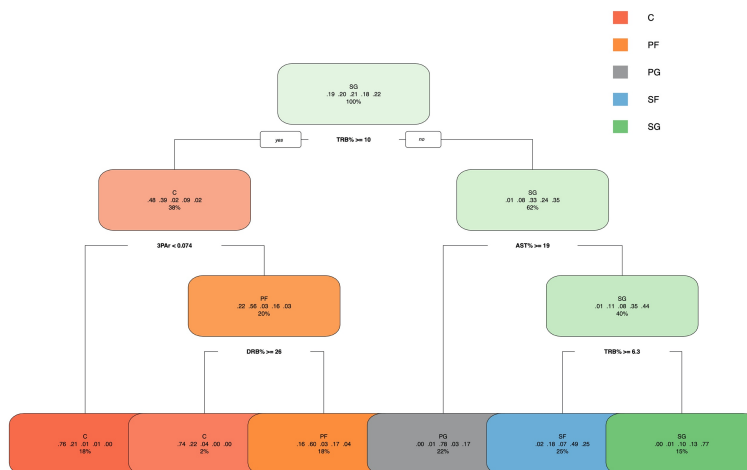


Gráfico 65: Gráfico del árbol de decisión C5.0 creado a partir del modelo generado para una muestra del 2% de jugadores de la temporada 2018/2019 para las variables de estadística avanzada. Generado con R Studio.

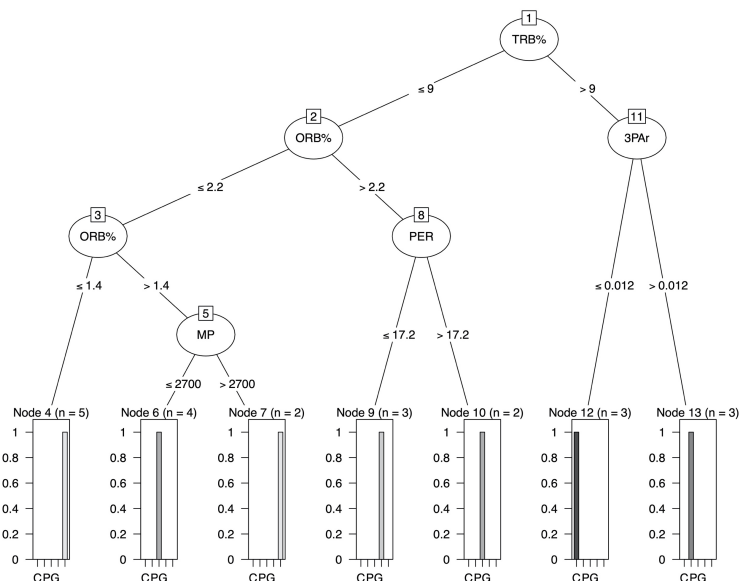


Tabla 75: Matriz de confusión entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con R Studio.

		Reference				
Prediction		C	PF	PG	SF	SG
C		41	7	0	1	0
PF		6	29	0	10	4
PG		0	1	47	2	10
SF		1	10	0	21	20
SG		0	4	11	13	38

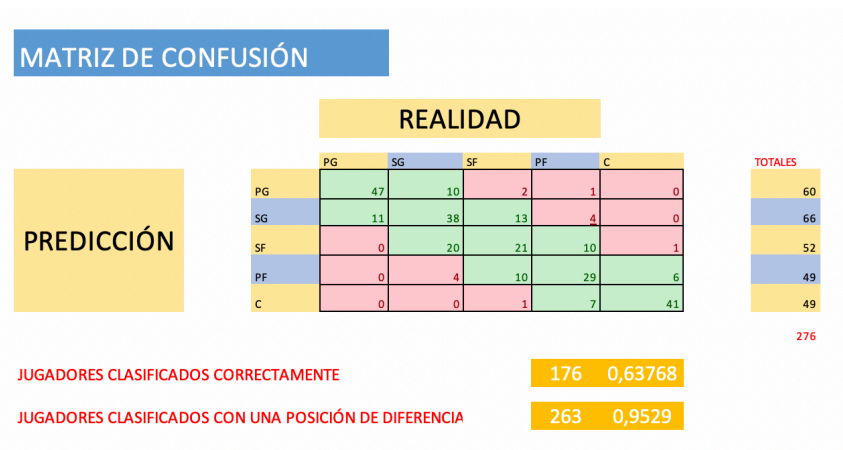
Overall Statistics

Accuracy : 0.6377
 95% CI : (0.5779, 0.6944)
 No Information Rate : 0.2609

Tabla 76: Matriz de confusión ajustada 1 entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con Excel.



Tabla 77: Matriz de confusión ajustada 2 entre la posición real y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *Excel*.



2.2.3.2 Pruebas con distintas temporadas

Tabla 78: Matriz de confusión entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 38 37 0 5 1
PF 0 7 0 9 0
PG 0 0 43 4 7
SF 3 5 0 19 21
SG 0 0 5 7 16

Overall Statistics

Accuracy : 0.5419
95% CI : (0.4746, 0.6079)
No Information Rate : 0.2159
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 79: Matriz de confusión entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 30 24 0 5 0
PF 10 18 0 10 3
PG 0 1 45 3 7
SF 1 4 0 18 8
SG 0 2 3 8 27

Overall Statistics

Accuracy : 0.6079
95% CI : (0.5411, 0.6719)
No Information Rate : 0.2159
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 80: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *Excel*.

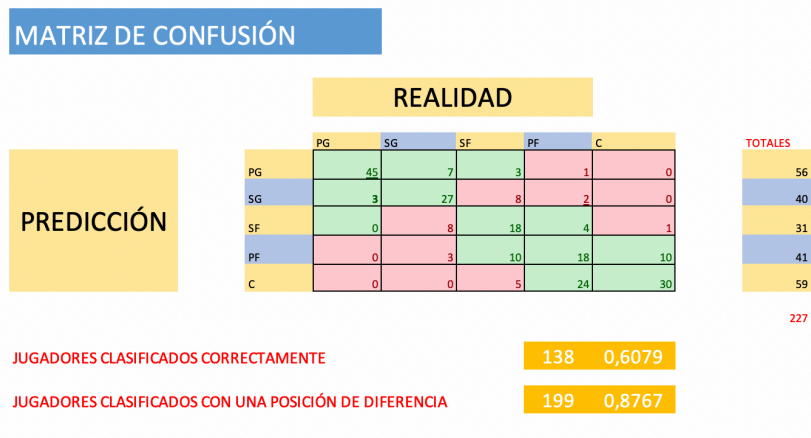


Tabla 81: Matriz de confusión ajustada 2 entre la posición real de los datos de la temporada 1994/1995 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *Excel*.



Tabla 82: Matriz de confusión entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en el *árbol de decisión* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 38 34 0 1 0
PF 5 9 0 9 1
PG 0 0 50 2 11
SF 2 7 0 35 22
SG 0 0 7 2 17

Overall Statistics

Accuracy : 0.5913
95% CI : (0.5278, 0.6526)
No Information Rate : 0.2262
    
```

Tabla 83: Matriz de confusión entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *R Studio*.

```

Reference
Prediction C PF PG SF SG
C 33 26 0 2 0
PF 11 20 0 11 0
PG 0 0 46 0 7
SF 0 4 0 30 14
SG 1 0 11 6 30

Overall Statistics

Accuracy : 0.631
95% CI : (0.5681, 0.6907)
No Information Rate : 0.2262
P-Value [Acc > NIR] : < 2.2e-16
    
```

Tabla 84: Matriz de confusión ajustada 1 entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *Excel*.

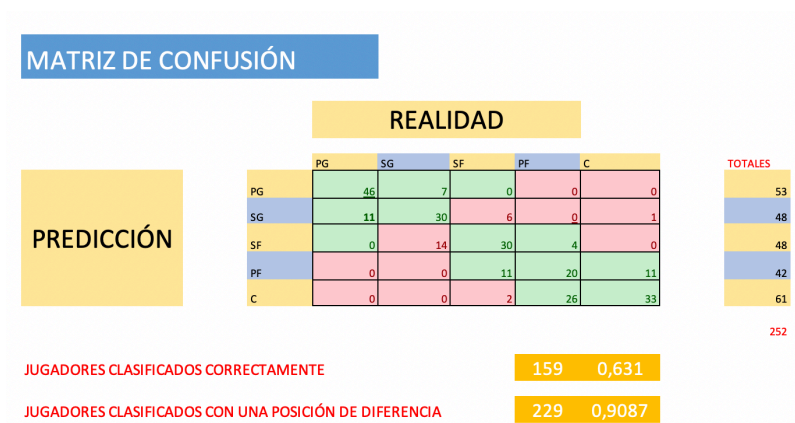
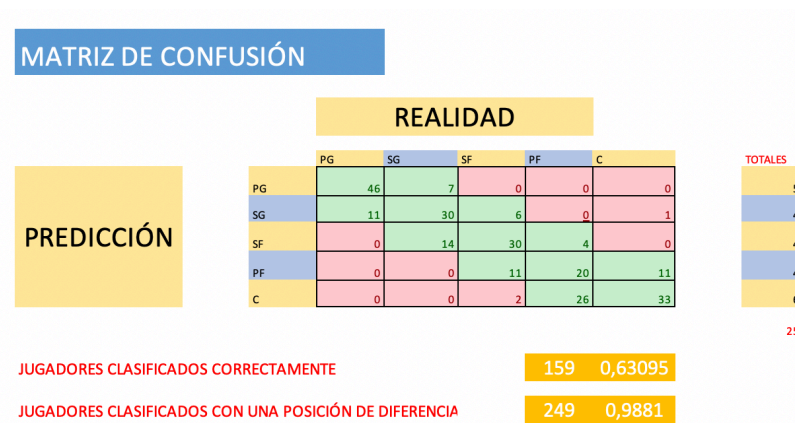


Tabla 85: Matriz de confusión ajustada 2 entre la posición real de los datos de la temporada 2009/2010 y la predicción creada por el modelo basado en los *k-vecinos más cercanos* para las variables de estadística avanzada en la temporada 2018/2019. Generada con *Excel*.



3. REGRESIÓN

3.1. Todos los tiros

3.1.1. Variables iniciales

Tabla 86: Modelo de regresión logística múltiple con las variables iniciales para todo los. datos. Generada con *R Studio*.

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.855104	0.129361	-14.34	< 0.0000000000000002 ***
period	-0.016590	0.011919	-1.39	0.16394
minutes_remaining	-0.001210	0.003986	-0.30	0.76150
seconds_remaining	0.000806	0.000784	1.03	0.30366
action_typeAlley Oop Layup shot	1.303730	0.154750	8.42	< 0.0000000000000002 ***
action_typeCutting Dunk Shot	-1.647151	0.278962	-5.90	0.000000035356131 ***
action_typeCutting Finger Roll Layup Shot	-0.307852	0.287289	-1.07	0.28391
action_typeCutting Layup Shot	0.610726	0.133440	4.58	0.0000047220700374 ***
action_typeDriving Bank Hook Shot	1.052859	0.373586	2.82	0.00483 **
action_typeDriving Bank shot	1.217152	0.153545	7.93	0.0000000000000022 ***
action_typeDriving Dunk Shot	-0.087785	0.168211	-0.52	0.60176
action_typeDriving Finger Roll Layup Shot	0.894460	0.137744	6.49	0.000000000837815 ***
action_typeDriving Floating Bank Jump Shot	1.227933	0.187227	6.56	0.000000000543495 ***
action_typeDriving Floating Jump Shot	1.682822	0.141252	11.91	< 0.0000000000000002 ***
action_typeDriving Hook Shot	1.610080	0.151196	10.65	< 0.0000000000000002 ***
action_typeDriving Jump shot	1.794671	0.826718	2.17	0.02994 *
action_typeDriving Layup Shot	1.484962	0.121667	12.21	< 0.0000000000000002 ***
action_typeDriving Reverse Dunk Shot	1.440519	0.639569	2.25	0.02430 *
action_typeDriving Reverse Layup Shot	1.485222	0.141575	10.49	< 0.0000000000000002 ***
action_typeDunk Shot	0.223312	0.136110	1.64	0.10086
action_typeFadeaway Bank shot	0.993235	0.278596	3.57	0.00036 ***
action_typeFadeaway Jump Shot	1.731046	0.146071	11.85	< 0.0000000000000002 ***
action_typeFinger Roll Layup Shot	1.255455	0.166695	7.53	0.0000000000000502 ***
action_typeFloating Jump shot	1.829484	0.130742	13.99	< 0.0000000000000002 ***
action_typeHook Bank Shot	1.116250	0.251836	4.43	0.0000093172771081 ***
action_typeHook Shot	1.994416	0.127732	15.61	< 0.0000000000000002 ***
action_typeJump Bank Shot	1.502789	0.142515	10.54	< 0.0000000000000002 ***
action_typeJump Shot	2.251984	0.157081	14.34	< 0.0000000000000002 ***
shot_type3PT Field Goal	-0.246375	0.117928	-2.09	0.03669 *
shot_distance	0.029136	0.006108	4.77	0.0000018440107954 ***
x	-0.000200	0.000236	-0.85	0.39571
y	-0.000575	0.000486	-1.18	0.23663
dribbles	-0.029258	0.009797	-2.99	0.00282 **
touch_time	0.051783	0.011498	4.50	0.0000066822019318 ***
shot_clock	-0.016791	0.002370	-7.08	0.0000000000013958 ***

Tabla 87: Anova para el modelo de regresión logística múltiple con las variables iniciales para todo los. datos. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			25930	34453	
period	1	1	25929	34452	0.317
minutes_remaining	1	0	25928	34451	0.504
seconds_remaining	1	0	25927	34451	0.789
action_type	24	3012	25903	31439	< 0.0000000000000002 ***
shot_type	1	0	25902	31439	0.573
shot_distance	1	30	25901	31409	0.0000000446613 ***
x	1	1	25900	31408	0.396
y	1	2	25899	31407	0.211
dribbles	1	9	25898	31398	0.003 **
touch_time	1	28	25897	31370	0.0000001375278 ***
shot_clock	1	50	25896	31320	0.0000000000014 ***

Gráfico 66: Representación gráfica de la importancia de las variables iniciales para todos los datos. Generado con *R Studio*.

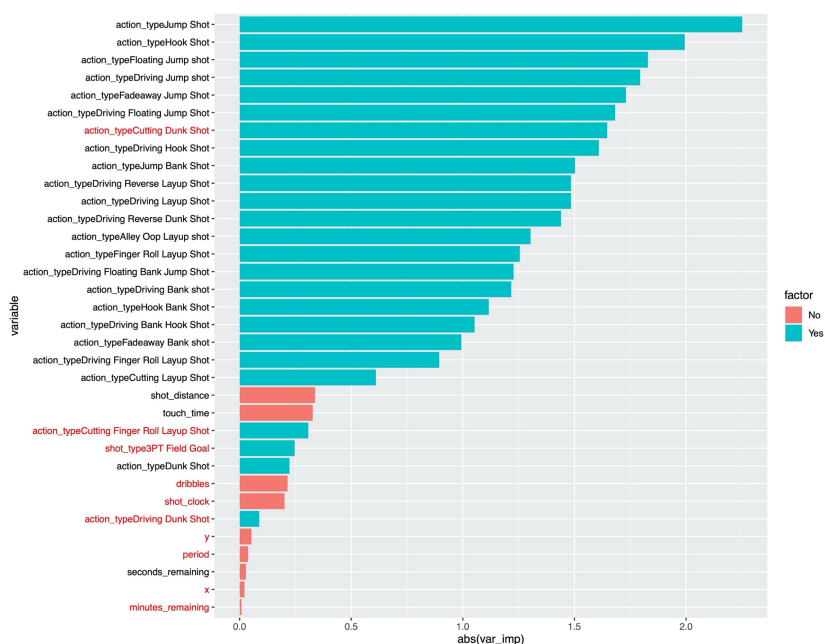


Tabla 88: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables iniciales y la realidad del tiro para todos los datos con un *threshold* de 0.4. Generada con *R Studio*.

Confusion Matrix and Statistics

		Reference	
Prediction	made	missed	
made	9787	3362	
missed	6277	6505	

Accuracy : 0.628
 95% CI : (0.622, 0.634)
 No Information Rate : 0.619
 P-Value [Acc > NIR] : 0.00179

Kappa : 0.254

Tabla 89: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables iniciales y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

Confusion Matrix and Statistics

		Reference	
Prediction	made	missed	
made	13091	6133	
missed	2973	3734	

Accuracy : 0.649
 95% CI : (0.643, 0.655)
 No Information Rate : 0.619
 P-Value [Acc > NIR] : <0.0000000000000002

Kappa : 0.206

Tabla 90: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables iniciales y la realidad del tiro para todos los datos con un *threshold* de 0.6. Generada con *R Studio*.

Confusion Matrix and Statistics

		Reference	
Prediction		made	missed
made	15530	8932	
missed	534	935	

Accuracy : 0.635
 95% CI : (0.629, 0.641)
 No Information Rate : 0.619
 P-Value [Acc > NIR] : 0.00000014

Kappa : 0.074

3.1.2. No lineales añadidas

Tabla 91: Modelo de regresión logística múltiple añadiendo las variables no lineales para todo los. datos. Generada con *R Studio*.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.85276473	0.12937268	-14.32	< 0.0000000000000002 ***
period	-0.01614534	0.01192420	-1.35	0.17574
minutes_remaining	-0.00161922	0.00398992	-0.41	0.68487
seconds_remaining	0.00073380	0.00078467	0.94	0.34970
action_typeAlley Oop Layup shot	1.27460438	0.15495606	8.23	< 0.0000000000000002 ***
action_typeCutting Dunk Shot	-1.64866141	0.27896806	-5.91	0.00000000342404 ***
action_typeCutting Finger Roll Layup Shot	-0.33689811	0.28743468	-1.17	0.24116
action_typeCutting Layup Shot	0.57653668	0.13373729	4.31	0.00001625441207 ***
action_typeDriving Bank Hook Shot	0.96757015	0.37414087	2.59	0.00971 **
action_typeDriving Bank shot	1.10907247	0.15613337	7.10	0.00000000000122 ***
action_typeDriving Dunk Shot	-0.08708873	0.16821647	-0.52	0.60465
action_typeDriving Finger Roll Layup Shot	0.86501727	0.13799435	6.27	0.00000000036455 ***
action_typeDriving Floating Bank Jump Shot	1.11907369	0.18948173	5.91	0.00000000350574 ***
action_typeDriving Floating Jump Shot	1.57722425	0.14453662	10.91	< 0.0000000000000002 ***
action_typeDriving Hook Shot	1.52339975	0.15336491	9.93	< 0.0000000000000002 ***
action_typeDriving Jump shot	1.70587977	0.82733459	2.06	0.03922 *
action_typeDriving Layup Shot	1.45019425	0.12199404	11.89	< 0.0000000000000002 ***
action_typeDriving Reverse Dunk Shot	1.44511208	0.63953388	2.26	0.02384 *
action_typeDriving Reverse Layup Shot	1.45265436	0.14183468	10.24	< 0.0000000000000002 ***
action_typeDunk Shot	0.22437683	0.13611648	1.65	0.09927 .
action_typeFadeaway Bank shot	0.90195742	0.27984558	3.22	0.00127 **
action_typeFadeaway Jump Shot	1.63698192	0.14931375	10.96	< 0.0000000000000002 ***
action_typeFinger Roll Layup Shot	1.22613009	0.16693816	7.34	0.000000000000021 ***
action_typeFloating Jump shot	1.71915814	0.13465031	12.77	< 0.0000000000000002 ***
action_typeHook Bank Shot	1.00031181	0.25363054	3.94	0.00008014291774 ***
action_typeHook Shot	1.89781449	0.13045846	14.55	< 0.0000000000000002 ***
action_typeJump Bank Shot	1.39897113	0.14541761	9.62	< 0.0000000000000002 ***
action_typeJump Shot	2.16570451	0.15998831	13.54	< 0.0000000000000002 ***
shot_type3PT Field Goal	-0.00305598	0.13303730	-0.02	0.98167
shot_distance	0.07312263	0.01282075	5.70	0.00000001174004 ***
x	0.00004239	0.00031198	0.14	0.89191
y	-0.00275701	0.00098653	-2.79	0.00520 **
dribbles	-0.02927702	0.00980007	-2.99	0.00281 **
touch_time	0.05126535	0.01150232	4.46	0.00000831308336 ***
shot_clock	-0.01682313	0.00237116	-7.09	0.00000000000129 ***
`x^2`	-0.00002223	0.00000583	-3.81	0.00014 ***
`y^2`	-0.00000655	0.00000267	-2.46	0.01406 *
`x*y`	-0.00000438	0.00000338	-1.30	0.19504

Tabla 92: Anova para el modelo de regresión logística múltiple añadiendo las variables no lineales para todo los. datos. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			25930	34453	
period	1	1	25929	34452	0.3167
minutes_remaining	1	0	25928	34451	0.5039
seconds_remaining	1	0	25927	34451	0.7889
action_type	24	3012	25903	31439	< 0.0000000000000002 ***
shot_type	1	0	25902	31439	0.5735
shot_distance	1	30	25901	31409	0.0000000446613 ***
x	1	1	25900	31408	0.3958
y	1	2	25899	31407	0.2109
dribbles	1	9	25898	31398	0.0030 **
touch_time	1	28	25897	31370	0.0000001375278 ***
shot_clock	1	50	25896	31320	0.0000000000014 ***
`x^2`	1	10	25895	31310	0.0019 **
`y^2`	1	7	25894	31304	0.0104 **
`x*y`	1	2	25893	31302	0.1939

Gráfico 67: Representación gráfica de la importancia de las variables añadiendo las no lineales para todos los datos. Generado con *R Studio*.

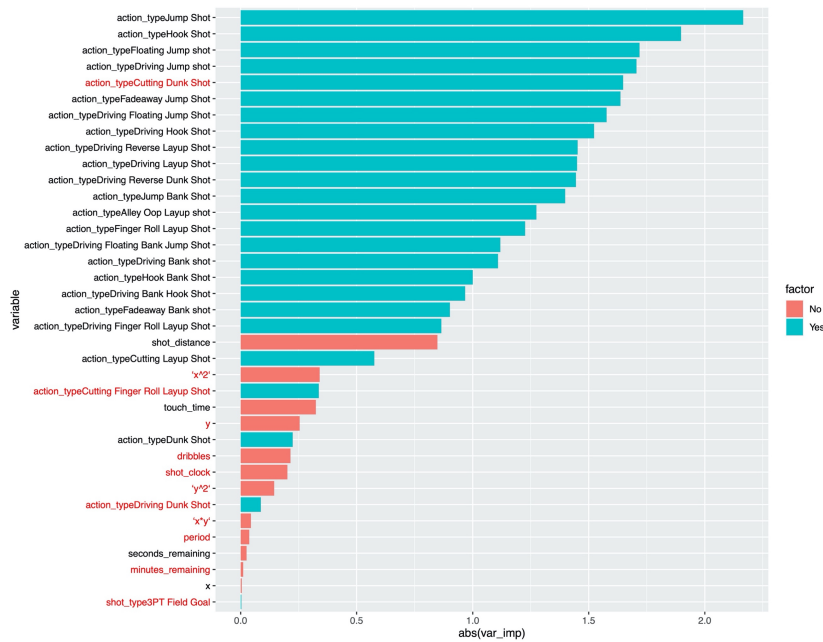


Tabla 93: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables no lineales añadidas y la realidad del tiro para todos los datos con un *threshold* de 0.4. Generada con *R Studio*.

Confusion Matrix and Statistics

		Reference	
		made	missed
Prediction made	made	9748	3310
	missed	6316	6557

Accuracy : 0.629
 95% CI : (0.623, 0.635)
 No Information Rate : 0.619
 P-Value [Acc > NIR] : 0.00103

Kappa : 0.256

Tabla 94: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables no lineales añadidas y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

Confusion Matrix and Statistics

		Reference	
		made	missed
Prediction made	made	13082	6127
	missed	2982	3740

Accuracy : 0.649
 95% CI : (0.643, 0.655)
 No Information Rate : 0.619
 P-Value [Acc > NIR] : <0.0000000000000002

Kappa : 0.206

Tabla 95: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables no lineales añadidas y la realidad del tiro para todos los datos con un *threshold* de 0.6. Generada con *R Studio*.

Confusion Matrix and Statistics

		Reference	
		made	missed
Prediction	made	15521	8914
	missed	543	953

Accuracy : 0.635
 95% CI : (0.629, 0.641)
 No Information Rate : 0.619
 P-Value [Acc > NIR] : 0.0000000753

Kappa : 0.075

3.1.3. Alturas y pesos añadidas

Tabla 96: Modelo de regresión logística múltiple añadiendo las variables de alturas y pesos de los jugadores para todo los datos. Generada con *R Studio*.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.02061823	0.40341230	-5.01	0.0000005476573523 ***
period	-0.01428509	0.01199341	-1.19	0.23362
minutes_remaining	-0.00276883	0.00400830	-0.69	0.48971
seconds_remaining	0.00064537	0.00078783	0.82	0.41269
action_typeAlley Oop Layup shot	1.22855370	0.15562780	7.89	0.0000000000000029 ***
action_typeCutting Dunk Shot	-1.67841245	0.27931497	-6.01	0.0000000018663481 ***
action_typeCutting Finger Roll Layup Shot	-0.45108190	0.28844129	-1.56	0.11785
action_typeCutting Layup Shot	0.43448040	0.13506196	3.22	0.00130 **
action_typeDriving Bank Hook Shot	0.73741097	0.37658423	1.96	0.05021 .
action_typeDriving Bank shot	0.91476787	0.15797658	5.79	0.0000000070165400 ***
action_typeDriving Dunk Shot	-0.10968187	0.16911790	-0.65	0.51663
action_typeDriving Finger Roll Layup Shot	0.70317173	0.14012960	5.02	0.0000005220944410 ***
action_typeDriving Floating Bank Jump Shot	0.92383883	0.19147454	4.82	0.0000014009811194 ***
action_typeDriving Floating Jump Shot	1.40613358	0.14648909	9.60	< 0.0000000000000002 ***
action_typeDriving Hook Shot	1.30414137	0.15452248	8.44	< 0.0000000000000002 ***
action_typeDriving Jump shot	1.51443663	0.82934528	1.83	0.06784 .
action_typeDriving Layup Shot	1.25513473	0.12414521	10.11	< 0.0000000000000002 ***
action_typeDriving Reverse Dunk Shot	1.33206280	0.64428512	2.07	0.03869 *
action_typeDriving Reverse Layup Shot	1.27144503	0.14378969	8.84	< 0.0000000000000002 ***
action_typeDunk Shot	0.24704104	0.13665028	1.81	0.07063 .
action_typeFadeaway Bank shot	0.75187334	0.28136509	2.67	0.00753 **
action_typeFadeaway Jump Shot	1.53410093	0.15036204	10.20	< 0.0000000000000002 ***
action_typeFinger Roll Layup Shot	1.10005111	0.16810211	6.54	0.00000000059165 ***
action_typeFloating Jump shot	1.57279046	0.13649269	11.52	< 0.0000000000000002 ***
action_typeHook Bank Shot	0.76140795	0.25452588	2.99	0.00278 **
action_typeHook Shot	1.69700325	0.13154784	12.90	< 0.0000000000000002 ***
action_typeJump Bank Shot	1.24273833	0.14663286	8.48	< 0.0000000000000002 ***
action_typeJump Shot	2.15920489	0.16121785	13.39	< 0.0000000000000002 ***
shot_type3PT Field Goal	0.03705085	0.13429817	0.28	0.78264
shot_distance	0.08195576	0.01291606	6.35	0.0000000002220478 ***
x	0.00001428	0.00031375	0.05	0.96369
y	-0.00256084	0.00099233	-2.58	0.00986 **
dribbles	-0.03171922	0.01010080	-3.14	0.00169 **
touch_time	0.04765385	0.01163328	4.10	0.0000419736713365 ***
shot_clock	-0.01480599	0.00239629	-6.18	0.000000006463097 ***
`x^2`	-0.00002149	0.00000588	-3.66	0.00026 ***
`y^2`	-0.00000684	0.00000271	-2.53	0.01143 *
`x*y`	-0.00000492	0.00000342	-1.44	0.14989
shooter_height	-0.00832111	0.05250546	-0.16	0.87408
shooter_weight	-0.00224175	0.00067152	-3.34	0.00084 ***
defender_distance	-0.11218479	0.00903326	-12.42	< 0.0000000000000002 ***
defender_height	0.05348588	0.04437089	1.21	0.22804
defender_weight	0.00350973	0.00062039	5.66	0.0000000153799105 ***

Tabla 97: Anova para el modelo de regresión logística múltiple añadiendo las variables de alturas y pesos de los jugadores para todo los. datos. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			25930	34453	
period	1	1	25929	34452	0.3167
minutes_remaining	1	0	25928	34451	0.5039
seconds_remaining	1	0	25927	34451	0.7889
action_type	24	3012	25903	31439	< 0.000000000000002 ***
shot_type	1	0	25902	31439	0.5735
shot_distance	1	30	25901	31409	0.000000446613 ***
x	1	1	25900	31408	0.3958
y	1	2	25899	31407	0.2109
dribbles	1	9	25898	31398	0.0030 **
touch_time	1	28	25897	31370	0.000001375278 ***
shot_clock	1	50	25896	31320	0.000000000014 ***
`x^2`	1	10	25895	31310	0.0019 **
`y^2`	1	7	25894	31304	0.0104 *
`x*y`	1	2	25893	31302	0.1939
shooter_height	1	0	25892	31302	0.6228
shooter_weight	1	2	25891	31300	0.1230
defender_distance	1	155	25890	31145	< 0.000000000000002 ***
defender_height	1	17	25889	31127	0.0000350638475 ***
defender_weight	1	32	25888	31095	0.000000149453 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Gráfico 68: Representación gráfica de la importancia de las variables añadiendo las de las alturas y pesos de los jugadores para todos los datos. Generado con *R Studio*.

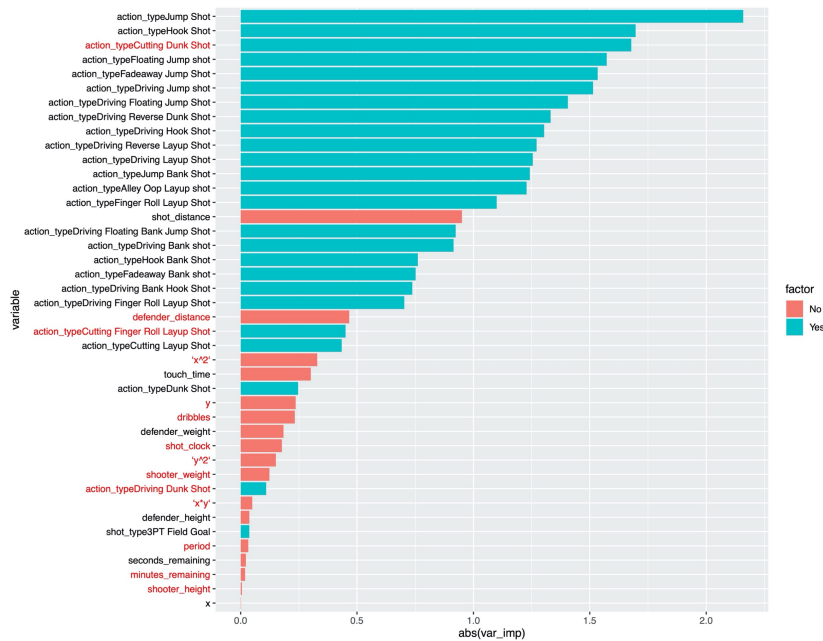


Tabla 98: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de pesos y alturas de los jugadores añadidas y la realidad del tiro para todos los datos con un *threshold* de 0.4. Generada con *R Studio*.

Confusion Matrix and Statistics

	Reference made	Reference missed
Prediction made	9617	3119
missed	6447	6748

Accuracy : 0.631
95% CI : (0.625, 0.637)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.0000589

Kappa : 0.265

Tabla 99: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de pesos y alturas de los jugadores y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      13168   6122
missed    2896   3745

Accuracy : 0.652
95% CI : (0.646, 0.658)
No Information Rate : 0.619
P-Value [Acc > NIR] : <0.0000000000000002

Kappa : 0.213

```

Tabla 100: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de pesos y alturas de los jugadores y la realidad del tiro para todos los datos con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      15420   8719
missed     644   1148

Accuracy : 0.639
95% CI : (0.633, 0.645)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.0000000000518

Kappa : 0.091

```

3.1.4. Tiros añadidos/todas las variables

Tabla 101: Anova para el modelo de regresión logística múltiple añadiendo las variables de el volumen de tiros y porcentaje de acierto por distancia para todo los. datos. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			25930	34453	
period	1	1	25929	34452	0.31668
minutes_remaining	1	0	25928	34451	0.50385
seconds_remaining	1	0	25927	34451	0.78894
action_type	24	3012	25903	31439	< 0.000000000000002 ***
shot_type	1	0	25902	31439	0.57347
shot_distance	1	30	25901	31409	0.000000446613 ***
x	1	1	25900	31408	0.39585
y	1	2	25899	31407	0.21094
`x^2`	1	10	25898	31396	0.00121 **
`y^2`	1	7	25897	31390	0.00886 **
`x*y`	1	2	25896	31388	0.20807
dribbles	1	8	25895	31380	0.00424 **
touch_time	1	27	25894	31353	0.0000001761999 ***
shot_clock	1	50	25893	31302	0.0000000000013 ***
shooter_height	1	0	25892	31302	0.62276
shooter_weight	1	2	25891	31300	0.12304
defender_distance	1	155	25890	31145	< 0.000000000000002 ***
defender_height	1	17	25889	31127	0.0000350638475 ***
defender_weight	1	32	25888	31095	0.0000000149453 ***
`LESS THAN 5FT FGM`	1	31	25887	31064	0.0000000236707 ***
`LESS THAN 5FT FGA`	1	33	25886	31031	0.0000000080721 ***
`LESS THAN 5FT FG%`	1	20	25885	31011	0.0000066939053 ***
`5-9 FT FGM`	1	40	25884	30970	0.0000000002081 ***
`5-9 FT FGA`	1	1	25883	30969	0.27481
`5-9 FT FG%`	1	14	25882	30955	0.00023 ***
`10-14 FT FGM`	1	0	25881	30955	0.53486
`10-14 FT FGA`	1	2	25880	30953	0.17079
`10-14 FT FG%`	1	0	25879	30953	0.61961
`15-19 FT FGM`	1	5	25878	30948	0.02768 *
`15-19 FT FGA`	1	3	25877	30945	0.08116 .
`15-19 FT FG%`	1	2	25876	30943	0.15917
`20-24 FT FGM`	1	6	25875	30937	0.01185 *
`20-24 FT FGA`	1	2	25874	30935	0.17769
`20-24 FT FG%`	1	0	25873	30935	0.50625
`25-29 FT FGM`	1	29	25872	30905	0.0000000619691 ***
`25-29 FT FGA`	1	4	25871	30902	0.05496 .
`25-29 FT FG%`	1	3	25870	30899	0.10819

Gráfico 69: Representación gráfica de la importancia de las variables añadiendo las de el volumen de tiros y porcentaje de acierto por distancia para todos los datos. Generado con *R Studio*.

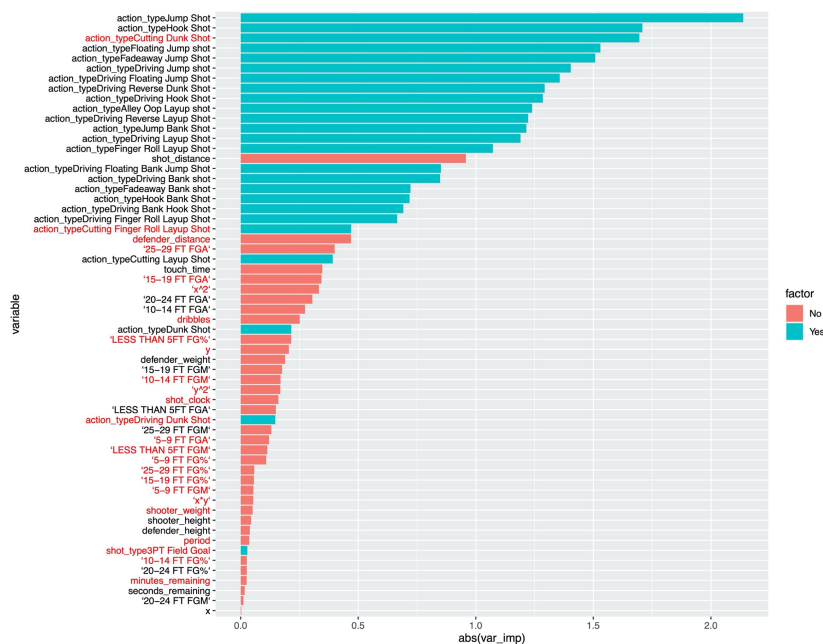


Tabla 102: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de volumen de tiro y porcentaje de acierto por zonas y la realidad del tiro para todos los datos con un *threshold* de 0.4. Generada con *R Studio*.

```
Confusion Matrix and Statistics

              Reference
Prediction made missed
made      9821  3109
missed   6243  6758

Accuracy : 0.639
95% CI : (0.633, 0.645)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.00000000000201

Kappa : 0.279
```

Tabla 103: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de volumen de tiro y porcentaje de acierto por zonas y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

```
Confusion Matrix and Statistics

              Reference
Prediction made missed
made      13184  5958
missed   2880  3909

Accuracy : 0.659
95% CI : (0.653, 0.665)
No Information Rate : 0.619
P-Value [Acc > NIR] : <0.000000000000002

Kappa : 0.231
```

Tabla 104: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de volumen de tiro y porcentaje de acierto por zonas y la realidad del tiro para todos los datos con un *threshold* de 0.6. Generada con *R Studio*.

```
Confusion Matrix and Statistics

              Reference
Prediction made missed
made      15246  8447
missed     818  1420

Accuracy : 0.643
95% CI : (0.637, 0.649)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.000000000000559

Kappa : 0.109
```


3.1.5. *Mínimo de variables importantes*

Tabla 105: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable del tipo de tiro y la realidad del tiro para todos los datos con un *threshold* de 0.4. Generada con *R Studio*.

```
Confusion Matrix and Statistics

      Reference
Prediction made missed
made      15654  9096
missed    410    771

Accuracy : 0.633
95% CI : (0.628, 0.639)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.0000019

Kappa : 0.063
```

Tabla 106: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable del tipo de tiro y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

```
Confusion Matrix and Statistics

      Reference
Prediction made missed
made      11210  4439
missed   4854   5428

Accuracy : 0.642
95% CI : (0.636, 0.647)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.000000000000894

Kappa : 0.246
```

Tabla 107: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable del tipo de tiro y la realidad del tiro para todos los datos con un *threshold* de 0.6. Generada con *R Studio*.

```
Confusion Matrix and Statistics

      Reference
Prediction made missed
made      12501  5536
missed   3563   4331

Accuracy : 0.649
95% CI : (0.643, 0.655)
No Information Rate : 0.619
P-Value [Acc > NIR] : <0.000000000000002

Kappa : 0.226
```

Tabla 108: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable de la distancia de tiro y la realidad del tiro para todos los datos con un *threshold* de 0.4. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      12409   5740
missed    3655   4127

      Accuracy : 0.638
      95% CI   : (0.632, 0.644)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.000000000726

      Kappa : 0.199

```

Tabla 109: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable de la distancia de tiro y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      14559   8013
missed    1505   1854

      Accuracy : 0.633
      95% CI   : (0.627, 0.639)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.00000396

      Kappa : 0.108

```

Tabla 110: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable de la distancia de tiro y la realidad del tiro para todos los datos con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      15361   8880
missed     703   987

      Accuracy : 0.63
      95% CI   : (0.625, 0.636)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.00014

      Kappa : 0.067

```

Tabla 111: Anova para el modelo de regresión logística múltiple con el menor número de variables posibles para obtener los resultados apropiados para todos los datos. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			25930	34453	
defender_distance	1	120	25929	34332	<0.0000000000000002 ***
`x^2`	1	648	25928	33684	<0.0000000000000002 ***
`y^2`	1	335	25927	33349	<0.0000000000000002 ***
dribbles	1	120	25926	33229	<0.0000000000000002 ***
defender_weight	1	78	25925	33150	<0.0000000000000002 ***
touch_time	1	115	25924	33035	<0.0000000000000002 ***
`LESS THAN 5FT FG`	1	163	25923	32872	<0.0000000000000002 ***

Gráfico 70: Representación gráfica de la importancia del menor número de variables posibles para obtener los resultados apropiados para todos los datos. Generado con *R Studio*.

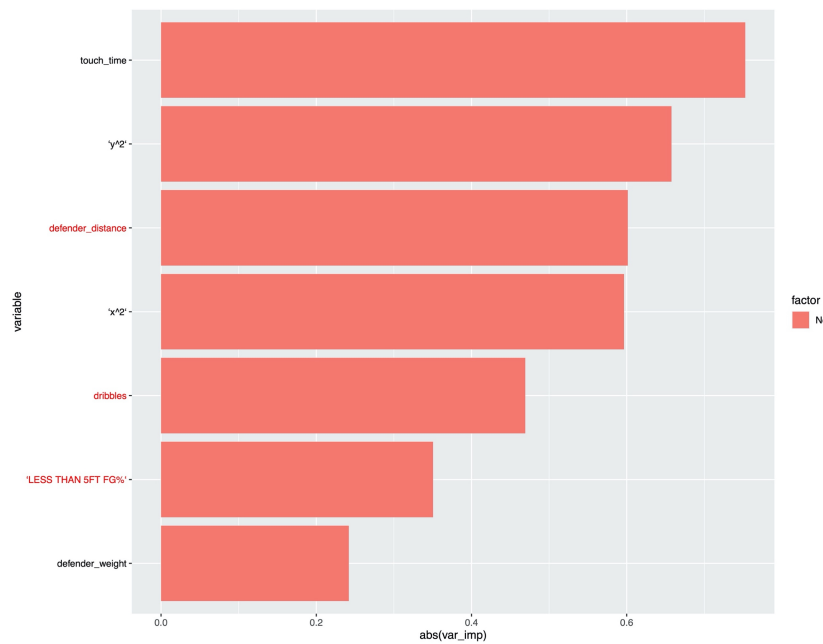


Tabla 112: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con el mínimo número de variables con el que obtener resultados oportunos y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

Confusion Matrix and Statistics

```

Reference
Prediction made missed
made 14581 8095
missed 1483 1772

Accuracy : 0.631
95% CI : (0.625, 0.637)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.000109

Kappa : 0.1

McNemar's Test P-Value : < 0.0000000000000002

Sensitivity : 0.908
Specificity : 0.180
Pos Pred Value : 0.643
Neg Pred Value : 0.544
Prevalence : 0.619
Detection Rate : 0.562
Detection Prevalence : 0.874
Balanced Accuracy : 0.544

'Positive' Class : made
    
```

Tabla 113: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con el mínimo número de variables con el que obtener resultados oportunos y la realidad del tiro para todos los datos con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      15548   9130
missed     516    737

Accuracy : 0.628
 95% CI : (0.622, 0.634)
No Information Rate : 0.619
P-Value [Acc > NIR] : 0.00237

Kappa : 0.051

McNemar's Test P-Value : < 0.0000000000000002

Sensitivity : 0.9679
Specificity : 0.0747
Pos Pred Value : 0.6300
Neg Pred Value : 0.5882
Prevalence : 0.6195
Detection Rate : 0.5996
Detection Prevalence : 0.9517
Balanced Accuracy : 0.5213

'Positive' Class : made

```

3.2. Tiros a más de 5 pies

3.2.1. Variables iniciales

Tabla 114: Modelo de regresión logística con las variables iniciales para los datos de tiros a más de 5 pies. Generada con *R Studio*.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -13.397909  266.871716  -0.05  0.960
period       0.014526   0.020259   0.72  0.473
minutes_remaining
seconds_remaining
action_typeCutting Finger Roll Layup Shot -0.155213  598.235425  0.00  1.000
action_typeCutting Layup Shot             13.835179  266.872793  0.05  0.959
action_typeDriving Bank Hook Shot         13.472413  266.872236  0.05  0.960
action_typeDriving Bank shot              12.977487  266.871730  0.05  0.961
action_typeDriving Dunk Shot              -0.163098  376.119364  0.00  1.000
action_typeDriving Finger Roll Layup Shot 13.037215  266.872349  0.05  0.961
action_typeDriving Floating Bank Jump Shot 12.864047  266.871758  0.05  0.962
action_typeDriving Floating Jump Shot     13.392750  266.871711  0.05  0.960
action_typeDriving Hook Shot              13.215958  266.871735  0.05  0.961
action_typeDriving Jump shot              14.156274  266.874515  0.05  0.958
action_typeDriving Layup Shot             13.535588  266.871978  0.05  0.960
action_typeDriving Reverse Layup Shot     -0.165236  598.235429  0.00  1.000
action_typeDunk Shot                      0.215928  598.235426  0.00  1.000
action_typeFadeaway Bank shot            12.798793  266.871854  0.05  0.962
action_typeFadeaway Jump Shot            13.566854  266.871709  0.05  0.959
action_typeFinger Roll Layup Shot        14.198308  266.873109  0.05  0.958
action_typeFloating Jump shot            13.539029  266.871703  0.05  0.960
action_typeHook Bank Shot                13.143993  266.871878  0.05  0.961
action_typeHook Shot                     13.788031  266.871705  0.05  0.959
action_typeJump Bank Shot                 13.329581  266.871711  0.05  0.960
action_typeJump Shot                     14.059094  266.871717  0.05  0.958
shot_type3PT Field Goal                  0.065229   0.127277   0.51  0.608
shot_distance                             0.002652   0.008073   0.33  0.743
x                                          -0.000148   0.000245  -0.60  0.546
y                                          -0.000255   0.000507  -0.50  0.616
dribbles                                  -0.014568   0.016056  -0.91  0.364
touch_time                                0.035606   0.018770   1.90  0.058
shot_clock                               -0.016257   0.004105  -3.96 0.000075 ***

```

Tabla 115: Anova para el modelo de regresión logística con las variables iniciales para los datos de tiros a más de 5 pies. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			7781	10759	
period	1	1.0	7780	10758	0.313
minutes_remaining	1	2.7	7779	10756	0.102
seconds_remaining	1	0.0	7778	10756	0.965
action_type	21	163.6	7757	10592	< 0.0000000000000002 ***
shot_type	1	0.3	7756	10592	0.559
shot_distance	1	0.1	7755	10592	0.727
x	1	0.4	7754	10591	0.543
y	1	0.2	7753	10591	0.626
dribbles	1	4.0	7752	10587	0.045 *
touch_time	1	5.2	7751	10582	0.023 *
shot_clock	1	15.7	7750	10566	0.000073 ***

Gráfico 71: Representación gráfica de la importancia de las variables iniciales para los datos de tiros a más de 5 pies. Generado con *R Studio*.

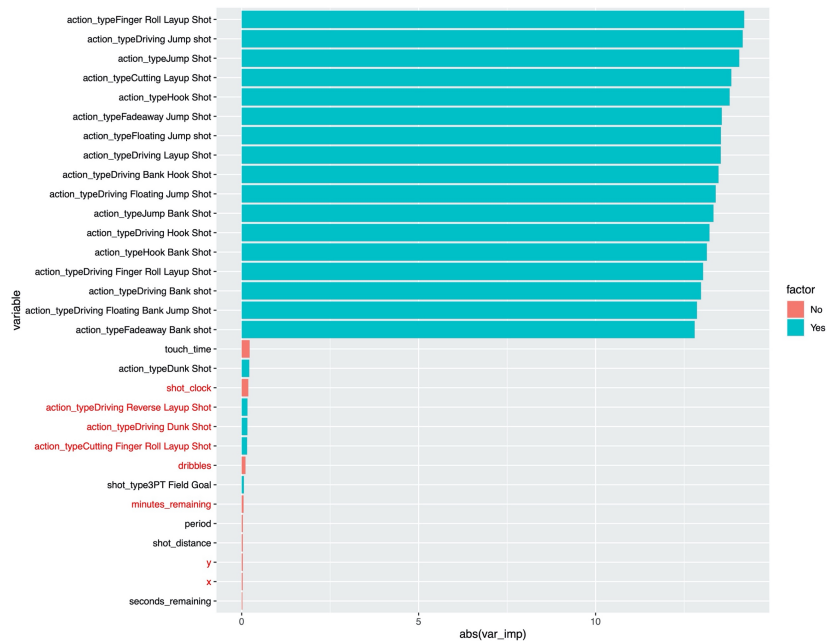


Tabla 116: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables iniciales y la realidad del tiro para todos los datos con un *threshold* de 0.4. Generada con *R Studio*.

Confusion Matrix and Statistics

		Reference	
Prediction made		made	missed
made	265	143	
missed	3390		3984

Accuracy : 0.546
 95% CI : (0.535, 0.557)
 No Information Rate : 0.53
 P-Value [Acc > NIR] : 0.00288

Kappa : 0.04

Tabla 117: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables iniciales y la realidad del tiro para todos los datos con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      1391  1161
missed    2264  2966

Accuracy : 0.56
95% CI : (0.549, 0.571)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.0000000893

Kappa : 0.101
    
```

Tabla 118: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables iniciales y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      3145  3204
missed    510   923

Accuracy : 0.523
95% CI : (0.512, 0.534)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.912

Kappa : 0.081
    
```

3.2.2. No lineales añadidas

Tabla 119: Modelo de regresión logística añadiendo las variables no lineales para los datos de tiros a más de 5 pies. Generada con *R Studio*.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -13.29820938 266.67153690 -0.05 0.960
period       0.01430194  0.02026773  0.71 0.480
minutes_remaining -0.00757880  0.00679050 -1.12 0.264
seconds_remaining  0.00071890  0.00131901  0.55 0.586
action_typeCutting Finger Roll Layup Shot -0.12974885 598.14614870  0.00 1.000
action_typeCutting Layup Shot 13.84267538 266.67260609  0.05 0.959
action_typeDriving Bank Hook Shot 13.48504300 266.67204810  0.05 0.960
action_typeDriving Bank shot 12.99830867 266.67154260  0.05 0.961
action_typeDriving Dunk Shot -0.12108173 375.32002085  0.00 1.000
action_typeDriving Finger Roll Layup Shot 13.06504171 266.67216173  0.05 0.961
action_typeDriving Floating Bank Jump Shot 12.88913159 266.67157060  0.05 0.961
action_typeDriving Floating Jump Shot 13.42765551 266.67152348  0.05 0.960
action_typeDriving Hook Shot 13.23556208 266.67154722  0.05 0.960
action_typeDriving Jump shot 14.17536086 266.67432874  0.05 0.958
action_typeDriving Layup Shot 13.54914707 266.67179005  0.05 0.959
action_typeDriving Reverse Layup Shot -0.18897054 598.14615549  0.00 1.000
action_typeDunk Shot  0.24190965 598.14614984  0.00 1.000
action_typeFadeaway Bank shot 12.84942317 266.67166698  0.05 0.962
action_typeFadeaway Jump Shot 13.62353109 266.67152258  0.05 0.959
action_typeFinger Roll Layup Shot 14.19874321 266.67292231  0.05 0.958
action_typeFloating Jump shot 13.57857757 266.67151512  0.05 0.959
action_typeHook Bank Shot 13.15488029 266.67169037  0.05 0.961
action_typeHook Shot 13.80826338 266.67151676  0.05 0.959
action_typeJump Bank Shot 13.37012436 266.67152378  0.05 0.960
action_typeJump Shot 14.12301591 266.67153095  0.05 0.958
shot_type3PT Field Goal  0.00845456  0.14083483  0.06 0.952
shot_distance -0.01892383  0.02112790 -0.90 0.370
x              0.00024101  0.00033519  0.72 0.472
y             -0.00059290  0.00117952 -0.50 0.615
dribbles     -0.01430724  0.01605457 -0.89 0.373
touch_time   0.03558326  0.01876785  1.90 0.058
shot_clock  -0.01609573  0.00410812 -3.92 0.000089 ***
`x^2`        0.00000697  0.00000791  0.88 0.379
`y^2`        0.00000769  0.00000446  1.72 0.085
`x*y`       -0.00000608  0.00000358 -1.70 0.090
    
```

Tabla 120: Anova para el modelo de regresión logística añadiendo las variables no lineales para los datos de tiros a más de 5 pies. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			7781	10759	
period	1	1.0	7780	10758	0.313
minutes_remaining	1	2.7	7779	10756	0.102
seconds_remaining	1	0.0	7778	10756	0.965
action_type	21	163.6	7757	10592	< 0.000000000000002 ***
shot_type	1	0.3	7756	10592	0.559
shot_distance	1	0.1	7755	10592	0.727
x	1	0.4	7754	10591	0.543
y	1	0.2	7753	10591	0.626
dribbles	1	4.0	7752	10587	0.045 *
touch_time	1	5.2	7751	10582	0.023 *
shot_clock	1	15.7	7750	10566	0.000073 ***
`x^2`	1	0.1	7749	10566	0.780
`y^2`	1	3.5	7748	10563	0.062 .
`x*y`	1	2.9	7747	10560	0.089 .

Gráfico 72: Representación gráfica de la importancia de las variables añadiendo las no lineales para los datos de tiros a más de 5 pies. Generado con *R Studio*.

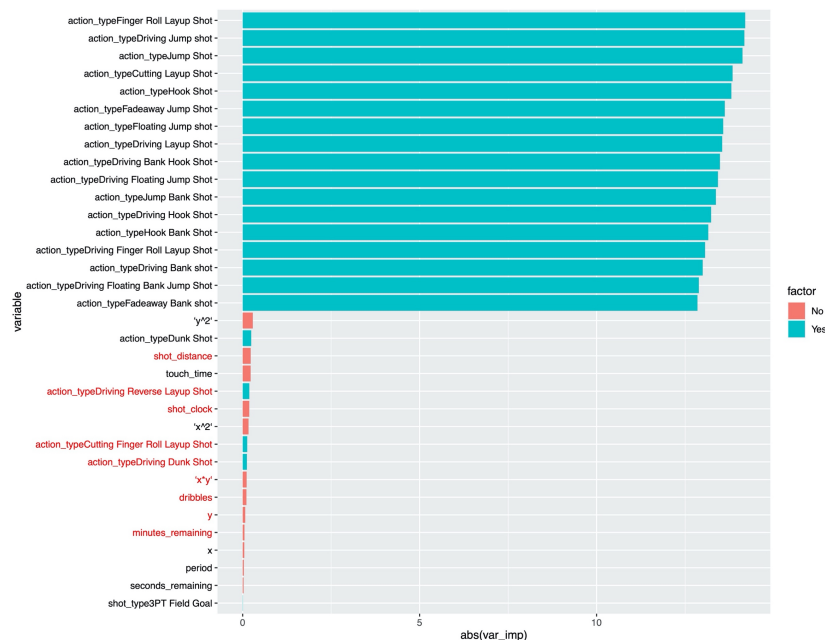


Tabla 121: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables no lineales añadidas y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.4. Generada con *R Studio*.

Confusion Matrix and Statistics

	Reference	
Prediction made	missed	made
made	267	140
missed	3388	3987

Accuracy : 0.547
 95% CI : (0.536, 0.558)
 No Information Rate : 0.53
 P-Value [Acc > NIR] : 0.00202
 Kappa : 0.041

Tabla 122: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables no lineales añadidas y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      1396   1182
missed    2259   2945

Accuracy : 0.558
95% CI : (0.547, 0.569)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.0000006

Kappa : 0.097
    
```

Tabla 123: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables no lineales añadidas y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      3151   3224
missed     504    903

Accuracy : 0.521
95% CI : (0.51, 0.532)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.952

Kappa : 0.078
    
```

3.2.3. Alturas y pesos añadidas

Tabla 124: Modelo de regresión logística añadiendo las variables de alturas y pesos de los jugadores para los datos de tiros a más de 5 pies. Generada con *R Studio*.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.35717138 266.07748470 -0.05 0.96296
period       0.01548811  0.02032248  0.76 0.44599
minutes_remaining
seconds_remaining
action_typeCutting Finger Roll Layup Shot -0.29266806 597.88119995  0.00 0.99961
action_typeCutting Layup Shot 13.76659557 266.07778665  0.05 0.95874
action_typeDriving Bank Hook Shot 13.44278930 266.07722517  0.05 0.95971
action_typeDriving Bank shot 12.94630574 266.07671675  0.05 0.96119
action_typeDriving Dunk Shot -0.02446471 373.41448374  0.00 0.99995
action_typeDriving Finger Roll Layup Shot 12.98506923 266.07733608  0.05 0.96108
action_typeDriving Floating Bank Jump Shot 12.83350797 266.07674546  0.05 0.96153
action_typeDriving Floating Jump Shot 13.39356448 266.07669790  0.05 0.95985
action_typeDriving Hook Shot 13.17048626 266.07672089  0.05 0.96052
action_typeDriving Jump shot 14.13303978 266.07951150  0.05 0.95764
action_typeDriving Layup Shot 13.44576319 266.07696713  0.05 0.95970
action_typeDriving Reverse Layup Shot -0.30059606 597.88120171  0.00 0.99960
action_typeDunk Shot 0.05649340 597.88119868  0.00 0.99992
action_typeFadeaway Bank shot 12.83524977 266.07684065  0.05 0.96153
action_typeFadeaway Jump Shot 13.62286081 266.07669556  0.05 0.95917
action_typeFinger Roll Layup Shot 14.26133831 266.07810635  0.05 0.95726
action_typeFloating Jump shot 13.55219656 266.07668928  0.05 0.95938
action_typeHook Bank Shot 13.08424977 266.07686384  0.05 0.96078
action_typeHook Shot 13.76498346 266.07669049  0.05 0.95874
action_typeJump Bank Shot 13.34960927 266.07669694  0.05 0.95999
action_typeJump Shot 14.19098844 266.07670417  0.05 0.95747
shot_type3PT Field Goal 0.02940477 0.14058391  0.21 0.83432
shot_distance -0.01169676 0.02098066 -0.56 0.57718
x 0.00022588 0.00033561  0.67 0.50092
y -0.00049584 0.00118245 -0.42 0.67497
dribbles -0.01246633 0.01651159 -0.76 0.45025
touch_time 0.02959988 0.01892113  1.56 0.11773
shot_clock -0.01540860 0.00413154 -3.73 0.00019 ***
`x^2` 0.00000634 0.00000784  0.81 0.41911
`y^2` 0.00000675 0.00000426  1.59 0.11289
`x*y` -0.00000638 0.00000359 -1.78 0.07533 .
shooter_height -0.18798853 0.08813256 -2.13 0.03292 *
shooter_weight 0.00012835 0.00118017  0.11 0.91340
defender_distance -0.06487485 0.01422252 -4.56 0.0000051 ***
defender_height -0.01199578 0.08188810 -0.15 0.88353
defender_weight 0.00192369 0.00109345  1.76 0.07853 .
    
```

Tabla 125: Anova para el modelo de regresión logística añadiendo las variables de alturas y pesos de los jugadores para los datos de tiros a más de 5 pies. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			7781	10759	
period	1	1.0	7780	10758	0.313
minutes_remaining	1	2.7	7779	10756	0.102
seconds_remaining	1	0.0	7778	10756	0.965
action_type	21	163.6	7757	10592	< 0.000000000000002 ***
shot_type	1	0.3	7756	10592	0.559
shot_distance	1	0.1	7755	10592	0.727
x	1	0.4	7754	10591	0.543
y	1	0.2	7753	10591	0.626
dribbles	1	4.0	7752	10587	0.045 *
touch_time	1	5.2	7751	10582	0.023 *
shot_clock	1	15.7	7750	10566	0.000073 ***
`x^2`	1	0.1	7749	10566	0.780
`y^2`	1	3.5	7748	10563	0.062 .
`x*y`	1	2.9	7747	10560	0.089 .
shooter_height	1	4.5	7746	10555	0.033 *
shooter_weight	1	0.3	7745	10555	0.611
defender_distance	1	18.7	7744	10536	0.000016 ***
defender_height	1	0.5	7743	10536	0.487
defender_weight	1	3.1	7742	10533	0.078 .

Gráfico 73: Representación gráfica de la importancia de las variables añadiendo las de las alturas y pesos de los jugadores para los datos de tiros a más de 5 pies. Generado con *R Studio*.

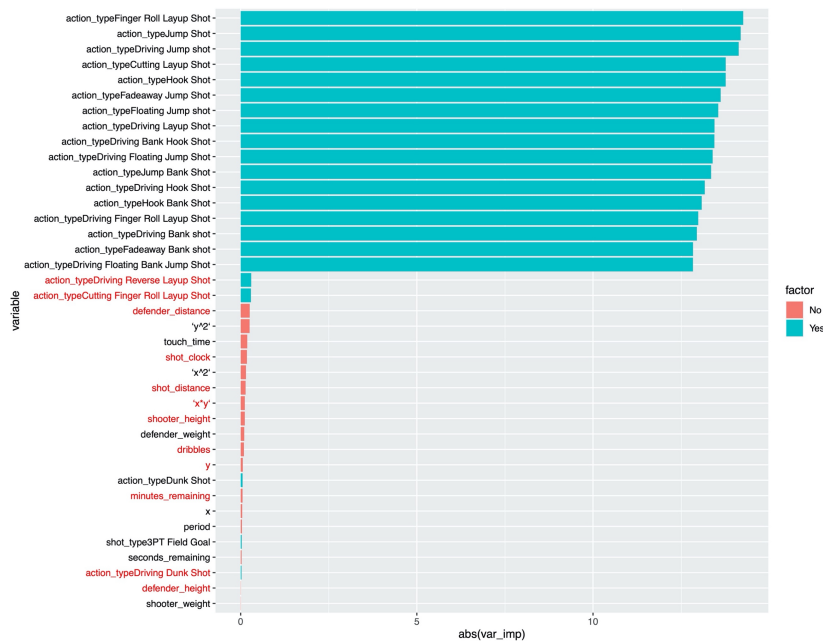


Tabla 126: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de pesos y alturas de los jugadores y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.4. Generada con *R Studio*.

Confusion Matrix and Statistics

	Reference	
Prediction made	missed	made
made	295	162
missed	3360	3965

Accuracy : 0.547
 95% CI : (0.536, 0.559)
 No Information Rate : 0.53
 P-Value [Acc > NIR] : 0.0013

Kappa : 0.044

Tabla 127: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de pesos y alturas de los jugadores y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      1433    1185
missed    2222    2942

      Accuracy : 0.562
      95% CI   : (0.551, 0.573)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.00000000898

      Kappa : 0.107
    
```

Tabla 128: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de pesos y alturas de los jugadores y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      3142    3156
missed     513     971

      Accuracy : 0.529
      95% CI   : (0.517, 0.54)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.629

      Kappa : 0.091
    
```

3.2.4. Tiros añadidos/todas las variables

Tabla 129: Modelo de regresión logística añadiendo las variables de volumen de tiro y porcentaje de acierto por zonas para los datos de tiros a más de 5 pies. Generada con *R Studio*.

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-12.56238820	265.33587476	-0.05	0.9622
period	0.01808218	0.02050305	0.88	0.3778
minutes_remaining	-0.00881487	0.00685160	-1.29	0.1983
seconds_remaining	0.00036797	0.00133062	0.28	0.7821
action_typeCutting Finger Roll Layup Shot	-0.40947510	597.55125702	0.00	0.9995
action_typeCutting Layup Shot	13.51248688	265.33557666	0.05	0.9594
action_typeDriving Bank Hook Shot	13.34368698	265.33499644	0.05	0.9599
action_typeDriving Bank shot	12.84586115	265.33448312	0.05	0.9614
action_typeDriving Dunk Shot	-0.30789814	369.07199861	0.00	0.9993
action_typeDriving Finger Roll Layup Shot	12.91138932	265.33510988	0.05	0.9612
action_typeDriving Floating Bank Jump Shot	12.72301944	265.33451262	0.05	0.9618
action_typeDriving Floating Jump Shot	13.31029519	265.33446425	0.05	0.9600
action_typeDriving Hook Shot	13.13324881	265.33448572	0.05	0.9605
action_typeDriving Jump shot	14.07913486	265.33729479	0.05	0.9577
action_typeDriving Layup Shot	13.29571122	265.33474202	0.05	0.9600
action_typeDriving Reverse Layup Shot	-0.66868274	597.55125977	0.00	0.9991
action_typeDunk Shot	-0.12695842	597.55125872	0.00	0.9998
action_typeFadeaway Bank shot	12.73833672	265.33460629	0.05	0.9617
action_typeFadeaway Jump Shot	13.56242628	265.33446052	0.05	0.9592
action_typeFinger Roll Layup Shot	14.07834857	265.33588946	0.05	0.9577
action_typeFloating Jump shot	13.48417271	265.33445480	0.05	0.9595
action_typeHook Bank Shot	13.00003393	265.33463342	0.05	0.9609
action_typeHook Shot	13.76342508	265.33445624	0.05	0.9586
action_typeJump Bank Shot	13.26007710	265.33446119	0.05	0.9601
action_typeJump Shot	14.12096221	265.33446702	0.05	0.9576
shot_type3PT Field Goal	-0.00582776	0.14289691	-0.04	0.9675
shot_distance	-0.00991900	0.02116208	-0.47	0.6393
x	0.00022166	0.00033899	0.65	0.5132
y	-0.00040599	0.00119369	-0.34	0.7338
`x^2`	0.00000569	0.00000791	0.72	0.4721
`y^2`	0.00000632	0.00000429	1.47	0.1409
`x*y`	-0.00000666	0.00000362	-1.84	0.0658
dribbles	-0.01085129	0.01667883	-0.65	0.5153
touch_time	0.03043480	0.01919414	1.59	0.1128
shot_clock	-0.01360194	0.00417713	-3.26	0.0011 ***
shooter_height	-0.02131398	0.09685645	-0.22	0.8258
shooter_weight	-0.00064226	0.00132349	-0.49	0.6275
defender_distance	-0.06447628	0.01438719	-4.48	0.000074 ***
defender_height	-0.00320343	0.08249084	-0.04	0.9690
defender_weight	0.00183908	0.00110367	1.67	0.0956
`LESS THAN 5FT FGM`	0.02339703	0.18494078	0.13	0.8993
`LESS THAN 5FT FGA`	0.00355987	0.10857498	0.03	0.9738
`LESS THAN 5FT FG%`	0.00039454	0.00679408	0.06	0.9537
`5-9 FT FGM`	0.05735706	0.39041534	0.15	0.8832
`5-9 FT FGA`	-0.07150372	0.18692201	-0.38	0.7021
`5-9 FT FG%`	-0.01786092	0.00390969	-4.57	0.000049 ***
`10-14 FT FGM`	-0.80206852	0.42085233	-1.91	0.0567
`10-14 FT FGA`	0.47817513	0.19737248	2.42	0.0154 *
`10-14 FT FG%`	-0.00216422	0.00336888	-0.64	0.5206
`15-19 FT FGM`	-0.31832254	0.26588932	-1.20	0.2312
`15-19 FT FGA`	0.04562368	0.11605113	0.39	0.6942
`15-19 FT FG%`	0.00019403	0.00281120	0.07	0.9450
`20-24 FT FGM`	-0.21283868	0.27552614	-0.77	0.4398
`20-24 FT FGA`	0.16359294	0.10702095	1.53	0.1264
`20-24 FT FG%`	0.00179420	0.00216277	0.83	0.4068
`25-29 FT FGM`	0.11949388	0.30242615	0.40	0.6928
`25-29 FT FGA`	-0.12271357	0.12209586	-1.01	0.3149
`25-29 FT FG%`	-0.00066845	0.00175998	-0.38	0.7041

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tabla 130: Anova para el modelo de regresión logística añadiendo las variables de volumen de tiro y porcentaje de acierto por zonas para los datos de tiros a más de 5 pies. Generada con *R Studio*

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			7781	10759	
period	1	1.0	7780	10758	0.31265
minutes_remaining	1	2.7	7779	10756	0.10201
seconds_remaining	1	0.0	7778	10756	0.96507
action_type	21	163.6	7757	10592	< 0.0000000000000002 ***
shot_type	1	0.3	7756	10592	0.55943
shot_distance	1	0.1	7755	10592	0.72707
x	1	0.4	7754	10591	0.54287
y	1	0.2	7753	10591	0.62603
`x^2`	1	0.1	7752	10591	0.74943
`y^2`	1	3.8	7751	10587	0.05076 .
`x*y`	1	2.7	7750	10585	0.09864 .
dribbles	1	4.2	7749	10580	0.04115 *
touch_time	1	5.1	7748	10575	0.02347 *
shot_clock	1	15.4	7747	10560	0.00008777 ***
shooter_height	1	4.5	7746	10555	0.03340 *
shooter_weight	1	0.3	7745	10555	0.61144
defender_distance	1	18.7	7744	10536	0.00001569 ***
defender_height	1	0.5	7743	10536	0.48733
defender_weight	1	3.1	7742	10533	0.07844 .
`LESS THAN 5FT FGM`	1	0.0	7741	10533	0.98489
`LESS THAN 5FT FGA`	1	0.3	7740	10532	0.56112
`LESS THAN 5FT FG%`	1	1.0	7739	10531	0.31772
`5-9 FT FGM`	1	13.9	7738	10518	0.00019 ***
`5-9 FT FGA`	1	11.7	7737	10506	0.00062 ***
`5-9 FT FG%`	1	26.4	7736	10479	0.0000028 ***
`10-14 FT FGM`	1	0.5	7735	10479	0.46650
`10-14 FT FGA`	1	9.9	7734	10469	0.00168 **
`10-14 FT FG%`	1	0.3	7733	10469	0.58276
`15-19 FT FGM`	1	7.2	7732	10462	0.00745 **
`15-19 FT FGA`	1	0.1	7731	10462	0.72321
`15-19 FT FG%`	1	0.0	7730	10461	0.90971
`20-24 FT FGM`	1	1.0	7729	10460	0.30754
`20-24 FT FGA`	1	1.8	7728	10459	0.18362
`20-24 FT FG%`	1	0.7	7727	10458	0.38762
`25-29 FT FGM`	1	4.6	7726	10453	0.03215 *
`25-29 FT FGA`	1	0.9	7725	10452	0.34675
`25-29 FT FG%`	1	0.1	7724	10452	0.70416

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Gráfico 131: Representación gráfica de la importancia de las variables añadiendo las de volumen de tiro y porcentaje de acierto por zonas para los datos de tiros a más de 5 pies. Generado con *R Studio*.

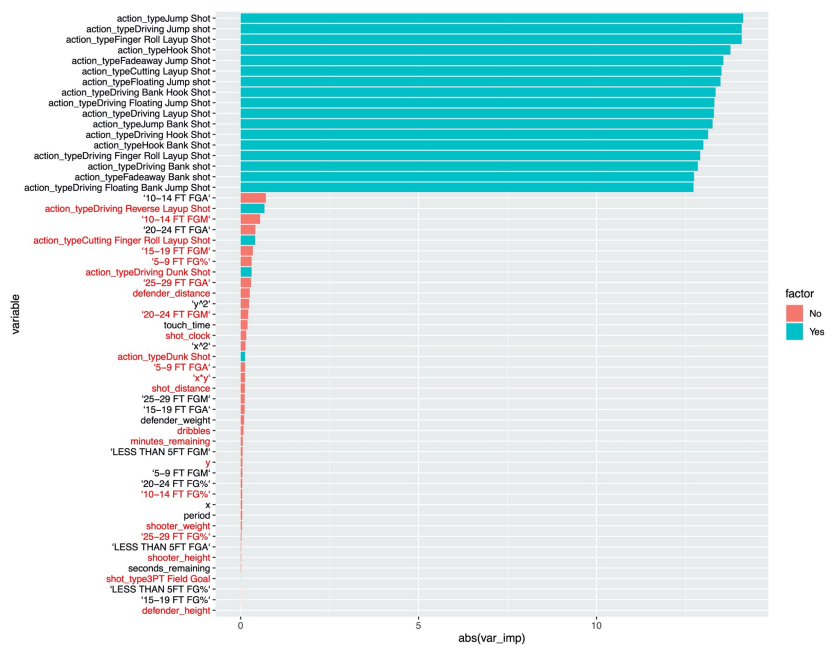


Tabla 132: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de volumen de tiro y porcentaje de acierto por zonas y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.4. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      429    233
missed 3226   3894

Accuracy : 0.556
95% CI : (0.544, 0.567)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.00000438

Kappa : 0.064
    
```

Tabla 133: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de volumen de tiro y porcentaje de acierto por zonas y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      1628   1268
missed 2027   2859

Accuracy : 0.577
95% CI : (0.566, 0.588)
No Information Rate : 0.53
P-Value [Acc > NIR] : <0.0000000000000002

Kappa : 0.14
    
```

Tabla 134: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de volumen de tiro y porcentaje de acierto por zonas y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      3053   2955
missed   602   1172

Accuracy : 0.543
95% CI : (0.532, 0.554)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.0134

Kappa : 0.115
    
```

3.2.5. *Mínimo de variables importantes*

Tabla 135: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable del tipo de tiro y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.4. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      289    166
missed 3366   3961

Accuracy : 0.546
95% CI : (0.535, 0.557)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.00268

Kappa : 0.041
    
```

Tabla 136: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable del tipo de tiro y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      1135    918
missed    2520   3209

Accuracy : 0.558
95% CI : (0.547, 0.569)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.000000424

Kappa : 0.09
    
```

Tabla 137: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable del tipo de tiro y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      3259   3398
missed     396    729

Accuracy : 0.512
95% CI : (0.501, 0.524)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.999

Kappa : 0.065
    
```

Tabla 138: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable de la distancia de tiro y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.4. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made         0         0
missed    3655    4127

Accuracy : 0.5303
95% CI : (0.5192, 0.5415)
No Information Rate : 0.5303
P-Value [Acc > NIR] : 0.5046

Kappa : 0
    
```

Tabla 139: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable de la distancia de tiro y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.5. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      3541   3917
missed     114    210

Accuracy : 0.482
95% CI : (0.471, 0.493)
No Information Rate : 0.53
P-Value [Acc > NIR] : 1

Kappa : 0.019
    
```

Tabla 140: Matriz de confusión entre la predicción creada por el modelo de regresión logística simple con la variable de la distancia de tiro y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```
Confusion Matrix and Statistics

      Reference
Prediction made missed
made      3655  4127
missed     0     0

Accuracy : 0.47
95% CI : (0.459, 0.481)
No Information Rate : 0.53
P-Value [Acc > NIR] : 1

Kappa : 0
```

Tabla 141: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de tiros intentados y acertados por partido para zonas a 10-14 pies de distancia a la canasta y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.4. Generada con *R Studio*.

```
Confusion Matrix and Statistics

      Reference
Prediction made missed
made         0     0
missed    3655  4127

Accuracy : 0.5303
95% CI : (0.5192, 0.5415)
No Information Rate : 0.5303
P-Value [Acc > NIR] : 0.5046

Kappa : 0

McNemar's Test P-Value : <2e-16

Sensitivity : 0.0000
Specificity : 1.0000
Pos Pred Value : NaN
Neg Pred Value : 0.5303
Prevalence : 0.4697
Detection Rate : 0.0000
Detection Prevalence : 0.0000
Balanced Accuracy : 0.5000

'Positive' Class : made
```

Tabla 142: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de tiros intentados y acertados por partido para zonas a 10-14 pies de distancia a la canasta y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.5. Generada con *R Studio*.

```
Confusion Matrix and Statistics

      Reference
Prediction made missed
made        617  667
missed    3038  3460

Accuracy : 0.5239
95% CI : (0.5127, 0.5351)
No Information Rate : 0.5303
P-Value [Acc > NIR] : 0.8743

Kappa : 0.0075

McNemar's Test P-Value : <2e-16

Sensitivity : 0.16881
Specificity : 0.83838
Pos Pred Value : 0.48053
Neg Pred Value : 0.53247
Prevalence : 0.46967
Detection Rate : 0.07929
Detection Prevalence : 0.16500
Balanced Accuracy : 0.50360

'Positive' Class : made
```


Tabla 143: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con las variables de tiros intentados y acertados por partido para zonas a 10-14 pies de distancia a la canasta y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made      3541  3917
missed    114   210

Accuracy : 0.482
95% CI   : (0.4709, 0.4932)
No Information Rate : 0.5303
P-Value [Acc > NIR] : 1

Kappa : 0.0186

McNemar's Test P-Value : <2e-16

Sensitivity : 0.96881
Specificity : 0.05088
Pos Pred Value : 0.47479
Neg Pred Value : 0.64815
Prevalence : 0.46967
Detection Rate : 0.45502
Detection Prevalence : 0.95837
Balanced Accuracy : 0.50985

'Positive' Class : made
    
```

Tabla 144: Anova para el modelo de regresión logística múltiple con el menor número de variables posibles para obtener los resultados apropiados para los datos de tiros a más de 5 pies. Generada con *R Studio*.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			7781	10759	
shot_distance	1	33.4	7780	10726	0.000000076 ***
`10-14 FT FGM`	1	7.2	7779	10719	0.00719 **
`10-14 FT FGA`	1	11.9	7778	10707	0.00056 ***
`5-9 FT FG%`	1	36.6	7777	10670	0.000000015 ***
shot_type	1	3.5	7776	10667	0.06315 .
`25-29 FT FGA`	1	8.0	7775	10659	0.00465 **
defender_distance	1	13.4	7774	10646	0.00025 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Gráfico 75: Representación gráfica de la importancia del menor número de variables posibles para obtener los resultados apropiados para los datos de tiros a más de 5 pies. Generado con *R Studio*.

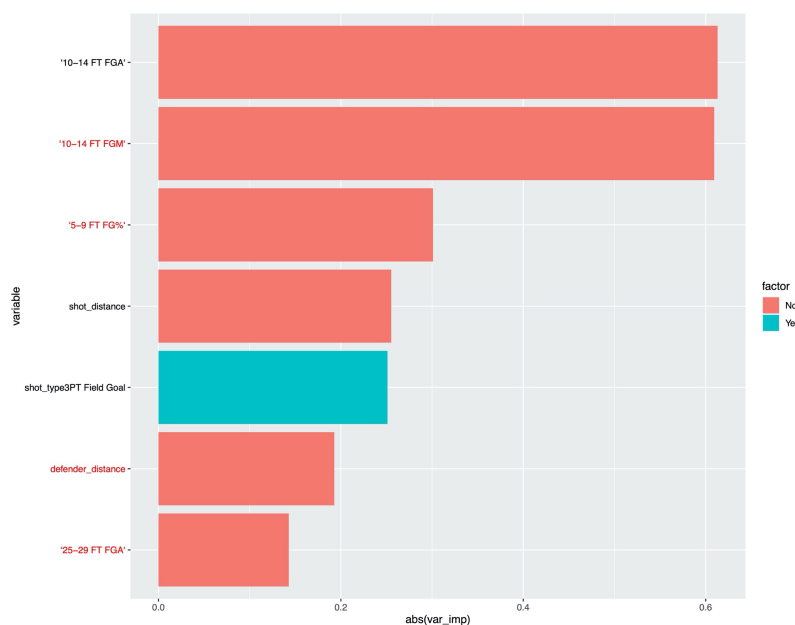


Tabla 145: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con el mínimo número de variables con el que obtener resultados oportunos y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.4. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made          54      49
missed       3601   4078

Accuracy : 0.531
95% CI : (0.52, 0.542)
No Information Rate : 0.53
P-Value [Acc > NIR] : 0.459

Kappa : 0.003

Mcnemar's Test P-Value : <0.000000000000002

Sensitivity : 0.01477
Specificity : 0.98813
Pos Pred Value : 0.52427
Neg Pred Value : 0.53106
Prevalence : 0.46967
Detection Rate : 0.00694
Detection Prevalence : 0.01324
Balanced Accuracy : 0.50145

'Positive' Class : made

```

Tabla 146: Matriz de confusión entre la predicción creada por el modelo de regresión logística múltiple con el mínimo número de variables con el que obtener resultados oportunos y la realidad del tiro para los datos de tiros a más de 5 pies con un *threshold* de 0.6. Generada con *R Studio*.

```

Confusion Matrix and Statistics

              Reference
Prediction made missed
made          3338   3560
missed         317    567

Accuracy : 0.502
95% CI : (0.491, 0.513)
No Information Rate : 0.53
P-Value [Acc > NIR] : 1

Kappa : 0.048

Mcnemar's Test P-Value : <0.000000000000002

Sensitivity : 0.913
Specificity : 0.137
Pos Pred Value : 0.484
Neg Pred Value : 0.641
Prevalence : 0.470
Detection Rate : 0.429
Detection Prevalence : 0.886
Balanced Accuracy : 0.525

'Positive' Class : made

```

4. CÓDIGO EN R

4.1. Clustering

```
library(readxl)
library(tidyverse)
library(ggpubr)
library(stats)
library(cluster)
library(mclust)
library(factoextra)
library(dendextend)
library(igraph)
library(tidygraph)
library(ggraph)
library(writexl)

# Lectura de las bases de datos #

X1980 <- read_excel("~/Desktop/TFG/CLUSTERING/1980.xlsx")
X1995 <- read_excel("~/Desktop/TFG/CLUSTERING/1995.xlsx")
X2010 <- read_excel("~/Desktop/TFG/CLUSTERING/2010.xlsx")
X2019 <- read_excel("~/Desktop/TFG/CLUSTERING/2019.xlsx")

datos <- X1980 # Introducir aquí el set de datos con el que quieres trabajar #

# Definir los niveles de las variables categóricas y ajustar los datos #

datos$Pos = factor(datos$Pos)
levels(datos$Pos)= c("C","PF","PG","SF","SG")

datos$`EXT/INT` = factor(datos$`EXT/INT`)
levels(datos$`EXT/INT`)= c("EXT","INT")

datos <- as.data.frame(datos)

rownames(datos) <- datos[,1]

# Preparación de los subsets de datos #

misvars1 <- c("Player","Pos","3P","2P","FT", "3PA", "2PA","FTA")
misvars2 <- c("Player","Pos","3P/MIN","2P/MIN", "3PA/MIN", "2PA/MIN","EXT/INT")
misvars3 <- c("Player","Pos","3PA", "2PA","FTA","EXT/INT")
misvars4 <- c("Player","Pos","3PA","3PA/MIN")

datosTODOS = datos[,misvars1]
datosMINUTO = datos[,misvars2]
datosINTENTADOS = datos[,misvars3]
datosTRIPLES = datos[,misvars4]
```

```

datos <- datosTODOS # Introducir aquí el subset de datos con el que quieres trabajar #
datos$Pos <- NULL
datos$Player <- NULL

# Comprobar número óptimo de clusters #

fviz_nbclust(x = datos, FUNcluster = kmeans, method = "wss", k.max = 15,
             diss = get_dist(datos, method = "euclidean"), nstart = 50)

#####CLUSTERS POR K MEANS #####

# Generar los clusters #

set.seed(123)
km_clusters <- kmeans(x = datos, centers = 5, nstart = 50)

# Representar los clusters gráficamente #

fviz_cluster(object = km_clusters, data = datos, show.clust.cent = TRUE,
              ellipse.type = "euclid", star.plot = TRUE, repel = TRUE, labelsize=5) +
  labs(title = "Resultados clustering K-means") +
  theme_bw() +
  theme(legend.position = "none")

# Comprobar la matriz de confusión #

table(km_clusters$cluster, datosTRIPLES["Pos"], #Especificar el subset con el que estás trabajando#
      dnn = list("cluster", "grupo real"))

# Generar una lista que añada la posición de la predicción a los datos originales #

LISTA2 <- cbind(datosTRIPLES[,c("Player", "Pos", "3PA", "3PA")], km_clusters$cluster)
write_xlsx(LISTA2, "~/Desktop/\\LISTA.xlsx")

##### PAM #####

# Generar los clusters #

pam_clusters <- pam(x = datos, k = 5, metric = "manhattan")

# Representar los clusters gráficamente #

fviz_cluster(object = pam_clusters, data = datos, ellipse.type = "t",
              repel = TRUE) +
  theme_bw() +
  labs(title = "Resultados clustering PAM") +
  theme(legend.position = "none")

# Comprobar los jugadores que hacen de medoids #

```

```

pam_clusters$medoids

# Comprobar la matriz de confusión #

table(pam_clusters$cluster, datos[, "Pos"],
      dnn = list("cluster", "grupo real"))

# Generar una lista que añada la posición de la predicción a los datos originales #

LISTA3 <- cbind(datos[, c("Player", "Pos", "3PA", "3PA")], pam_clusters$cluster)
write_xlsx(LISTA3, "~/Desktop/\\LISTA2.xlsx")

##### Hierarchical clustering #####

# Matriz de distancias euclídeas #

mat_dist <- dist(x = datos, method = "euclidean")

# Dendrogramas con linkage complete y average #

hc_euclidea_complete <- hclust(d = mat_dist, method = "complete")
hc_euclidea_average <- hclust(d = mat_dist, method = "average")
cor(x = mat_dist, cophenetic(hc_euclidea_complete))

hc_euclidea_completo <- hclust(d = dist(x = datos, method = "euclidean"),
                             method = "complete")

# Representación del dendograma gráficamente #

fviz_dend(x = hc_euclidea_completo, k = 5, cex = 0.3) +
  geom_hline(yintercept = 0.25, linetype = "dashed") + #Ajustar este valor para el corte visual del
dendograma#
  geom_hline(yintercept = 0.30, linetype = "dashed") + #Ajustar este valor para el corte visual del
dendograma#
  labs(title = "Herarchical clustering",
       subtitle = "Distancia euclídea, Lincage complete, K=5")

# Usar el dendograma para crear 5 clusters (Posiciones) #

clusters5 <- cutree(tree = hc_euclidea_completo, k = 5)

# Comprobar la matriz de confusión #

table(clusters5, datosMINUTO[, "Pos"],
      dnn = list("cluster", "grupo real"))

# Usar el dendograma para crear 2 clusters (exterior/interior) #

```

```
clusters2 <- cutree(tree = hc_euclidea_complete, k = 2)
```

```
# Comprobar la matriz de confusión #
```

```
table(clusters2, datosMINUTO["EXT/INT"], #Especificar el subset con el que estás trabajando#  
      dnn = list("cluster", "grupo real"))
```

4.2. Clasificación

```
library(readxl)  
library(tidyverse)  
library(ggpubr)  
library(stats)  
library(cluster)  
library(mclust)  
library(factoextra)  
library(dendextend)  
library(igraph)  
library(tidygraph)  
library(ggraph)  
library(rpart)  
library(rpart.plot)  
library(caret)  
library(C50)  
library(writexl)
```

```
# Lectura de las bases de datos que van a ser usadas para crear el modelo#
```

```
X1994 <- read_excel("~/Desktop/TFG/CLASIFICACIÓN/90-94 TIRO.xlsx")
```

```
X2009 <- read_excel("~/Desktop/TFG/CLASIFICACIÓN/05-09 TIRO.xlsx")
```

```
X2018 <- read_excel("~/Desktop/TFG/CLASIFICACIÓN/14-18 TIRO.xlsx")
```

```
# Selección de datos que se van a clasificar #
```

```
Clasif <- read_excel("~/Desktop/TFG/CLASIFICACIÓN/2019 CLASIF TIRO.xlsx")
```

```
Clasif2 <- read_excel("~/Desktop/TFG/CLASIFICACIÓN/2019 CLASIF TIRO.xlsx")
```

```
# Selección y tratamiento de datos con los que generar el modelo #
```

```
datos <- X2018
```

```
Clasif$Pos = factor(Clasif$Pos)  
levels(Clasif$Pos)= c("C","PF","PG","SF","SG")
```

```
datos$Pos = factor(datos$Pos)  
levels(datos$Pos)= c("C","PF","PG","SF","SG")
```

```
datos <- as.data.frame(datos)  
Clasif <- as.data.frame(Clasif)  
rownames(Clasif) <- Clasif$Player
```

```

datos <- datos[,2:11]
Clasif <- Clasif[,2:11]

set.seed(1234)

# Diseñamos el modelo y analizamos su varianza con un anova #
modelo = glm(Pos~.,data = datos, family="binomial")
summary(modelo)
anova(modelo,test = "Chisq")

##### ÁRBOL DE CLASIFICACIÓN #####

# Generar el modelo #
arbol_1 <- rpart(formula = Pos ~ ., data = datos)

# Gráfica del modelo #
rpart.plot(arbol_1)

# Predicción del modelo #
prediccion_1 <- predict(arbol_1, newdata = Clasif, type = "class")

# Comprobar la Matriz de Confusión #
confusionMatrix(prediccion_1, Clasif[["Pos"]])

# Generar una lista que añada la posición de la predicción a los datos originales #
LISTA2 <- cbind(Clasif2[,c("Player","Pos")],prediccion_1)
write_xlsx(LISTA2,"~/Desktop/\\LISTA.xlsx")

##### ÁRBOL C5.0 #####

# Elección de una muestra de los datos #
set.seed(12)
muestra <- sample_frac(datos,.02)

# Generar el modelo #
modelo <- C5.0(Pos ~ .,data = datos)

# Gráfica del modelo #

```

```

plot(modelo)

# Predicción del modelo #

prediccion <- predict(modelo,newdata=Clasif)

# Comprobar la Matriz de Confusión #

confusionMatrix(prediccion, Clasif[["Pos"]])

##### K-VECINOS MÁS CERCANOS #####

# División de datos en aprendizaje y los que se van a usar de test #

testing <- Clasif
aprendizaje <- datos

# Generar el modelo #

suppressWarnings(suppressMessages(library(kknn)))
modelo <- train.kknn(Pos ~ ., data = aprendizaje, kmax = 9)
modelo

# Generar las predicciones #

pred <- predict(modelo, testing)
pred

# Comprobar la matriz de confusión #

confusionMatrix(pred, testing[, 1])

```

4.3. Regresión

```

library(readxl)
library(caret)
library(stats)
library(ggplot2)
library(dplyr)
library(tibble)
library(tidyverse)

# Lectura de la base de datos #

X2016 <- read_xlsx("~/Desktop/TFG/REGRESION/2016.xlsx")

# todas las variables posibles y las que se elijan para cada modelo #

todas_vars <- c("period","minutes_remaining","seconds_remaining","shot_result","action_type",

```



```

"shot_type","shot_distance","x","y","dribbles","touch_time","shot_clock","x^2","y^2","x*y",
"shooter_height","shooter_weight","defender_distance","defender_height","defender_weight",
"LESS THAN 5FT FGM","LESS THAN 5FT FGA","LESS THAN 5FT FG%","5-9 FT
FGM","5-9 FT FGA","5-9 FT FG%",
"10-14 FT FGM","10-14 FT FGA","10-14 FT FG%","15-19 FT FGM","15-19 FT FGA","15-
19 FT FG%",
"20-24 FT FGM","20-24 FT FGA","20-24 FT FG%", "25-29 FT FGM","25-29 FT FGA","25-
29 FT FG%")

```

```

vars_usar <- c("period","minutes_remaining","seconds_remaining","shot_result","action_type",
"shot_type","shot_distance","x","y","dribbles","touch_time","shot_clock","x^2","y^2","x*y",
"shooter_height","shooter_weight","defender_distance","defender_height","defender_weight")

```

```

vars_usar_largos <- c("shot_result","shot_distance","10-14 FT FGM","10-14 FT FGA","5-9 FT
FG%","shot_type",
"25-29 FT FGA","defender_distance")

```

```

# Selección de variables y tratamiento de los datos #

```

```

datos <- X2016[,vars_usar]

```

```

datos <- filter(datos, shot_distance >= 5.01)

```

```

datos$shot_distance <- NULL

```

```

datos$shot_result = factor(datos$shot_result)
levels(datos$shot_result)=c("made","missed")

```

```

datos$action_type = factor(datos$action_type)

```

```

datos$shot_type = factor(datos$shot_type)

```

```

# Generar el modelo de regresión logística #

```

```

modelo <- glm(shot_result~.,data = datos, family = "binomial")

```

```

summary(modelo)

```

```

anova(modelo,test = "Chisq")

```

```

# Comprobar la importancia de las variables definiendo las funciones necesarias #

```

```

varimp.logistic <- function(modelFit) {
  UseMethod("varimp.logistic", modelFit)
}

```

```

#' @export

```

```

#' @describeIn varimp.logistic Method for glm()

```

```

varimp.logistic.glm <- function(modelFit) {

```

```

  if(modelFit$method != "glm.fit") {

```

```

    stop("This function only works when you call glm() or caret train() with method set to glm or
glmnet.")

```

```

  }

```

```

result <- cbind( coef(summary(modelFit)), confint.default(modelFit) )
colnames(result)[4] <- c("p_value")
tmp_coefs <- coef(modelFit)
tmp_newvars <- names(tmp_coefs)[-1]
used.dataframe <- eval(modelFit$call$data)
if(is.null(used.dataframe)){
  stop("This function only works when you explicitly pass in data as a data.frame when you call glm()
or caret train() with method set to glm or glmnet.")
}

is.01 <- function(x){ su <- sort(unique(x)); length(su)==2 && all(su==c(0,1)) }
allvars_factor <- c(names(used.dataframe[sapply(used.dataframe, is.factor)]),
  names(used.dataframe[sapply(used.dataframe, is.character)]),
  names(used.dataframe[sapply(used.dataframe, is.01)]))
allvars_factor2 <- paste(allvars_factor, collapse="|" )

factor_merge <- enframe(tmp_newvars[grepl(allvars_factor2,tmp_newvars)], name = "fac", value =
"variable") %>%
  mutate(fac=1)

logit_temp <- glm(modelFit$formula, data=used.dataframe, family=binomial, x=TRUE)
sd_merge <- as_tibble(logit_temp$x) %>%
  summarize_all(sd) %>%
  gather(key = "variable", value = "sd")

final_result <- as_tibble(result, rownames="variable") %>%
  dplyr::slice(-1) %>%
  left_join(factor_merge,by = "variable") %>%
  mutate(p_value=round(p_value,3),
  fac = replace_na(fac, 0),
  factor = ifelse(fac==1, "Yes", "No"),
  sign=sign(Estimate)) %>%
  left_join(sd_merge, by = "variable") %>%
  mutate(var_imp = ifelse(fac==1, Estimate, Estimate*2*sd),
  var_imp_lower = ifelse(fac==1, `2.5 %`, `2.5 %`*2*sd),
  var_imp_upper = ifelse(fac==1, `97.5 %`, `97.5 %`*2*sd)) %>%
  arrange(-abs(var_imp)) %>%
  select(variable, factor, var_imp, var_imp_lower, var_imp_upper, p_value)

options(scipen=999, digits =3)
return(final_result)
}

#' @export
#' @describeIn varimp.logistic Method for either glm or glmnet in caret. This works regardless of
whether the data is pre-processed with "scale"
varimp.logistic.train <- function(modelFit) {
  if( !(modelFit$method %in% c("glm","glmnet"))) {
    stop("This function only works when you call glm() or caret train() with method set to glm or
glmnet.")
  }

  if(modelFit$method=="glm"){
    glmFlag <- TRUE
    result <- cbind( coef(summary(modelFit)), confint.default(modelFit$finalModel) )
    colnames(result)[4] <- c("p_value")
    tmp_coefs <- coef(modelFit$finalModel)

```

```

tmp_newvars <- names(tmp_coefs)[-1]
used.dataframe <- eval(modelFit$call$data)
}

if(modelFit$method=="glmnet"){
  glmFlag <- FALSE
  tmp.fm <- getform.glmnet(modelFit)
  tmp.df <- modelFit$trainingData
  colnames(tmp.df)[1] <- as.character(tmp.fm)[2]
  tmp.lr <- glm(tmp.fm, data=tmp.df, family=binomial)
  tmp.coef <- coef(modelFit$finalModel,modelFit$finalModel$lambdaOpt)[names(coef(tmp.lr)),1]
  tmp.lr$coefficients <- tmp.coef
  oldmodelFit <- modelFit
  modelFit <- tmp.lr

  result <- coef(summary(modelFit))
  colnames(result)[4] <- c("p_value")
  result[,4] <- NA
  tmp_coefs <- coef(modelFit)
  tmp_newvars <- names(tmp_coefs)[-1]
  used.dataframe <- eval(modelFit$call$data)
}

if(is.null(used.dataframe)){
  stop("This function only works when you explicitly pass in data as a data.frame when you call glm()
or caret train() with method set to glm or glmnet.")
}

is.01 <- function(x){ su <- sort(unique(x)); length(su)==2 && all(su==c(0,1)) }
allvars_factor <- c(names(used.dataframe[sapply(used.dataframe, is.factor)]),
  names(used.dataframe[sapply(used.dataframe, is.character)]),
  names(used.dataframe[sapply(used.dataframe, is.01)]))
allvars_factor2 <- paste(allvars_factor, collapse="|")

factor_merge <- enframe(tmp_newvars[grepl(allvars_factor2,tmp_newvars)], name = "fac", value =
"variable") %>%
  mutate(fac=1)

logit_temp <- glm(modelFit$terms, data=used.dataframe, family=binomial, x=TRUE)
sd_merge <- as_tibble(logit_temp$x) %>%
  summarize_all(sd) %>%
  gather(key = "variable", value = "sd")

scaleFlag <- length(modelFit$preProcess$method$scale)>0

final_result <- as_tibble(result, rownames="variable") %>%
  dplyr::slice(-1) %>%
  left_join(factor_merge,by = "variable") %>%
  mutate(fac = replace_na(fac, 0),
  factor=ifelse(fac==1, "Yes", "No"),
  sign=sign(Estimate)) %>%
  left_join(sd_merge, by="variable") %>%
  mutate(var_imp = case_when(
  scaleFlag == TRUE ~ ifelse(fac==1, Estimate/sd, Estimate*2),
  scaleFlag == FALSE ~ ifelse(fac==1, Estimate, Estimate*2*sd) )
) %>%
  arrange(-abs(var_imp))

```

```

if(glmFlag){
  final_result <- final_result %>%
    mutate(p_value=round(p_value,3),
           var_imp_lower = case_when(
             scaleFlag == TRUE ~ ifelse(fac==1, `2.5 %`/sd, `2.5 %`*2),
             scaleFlag == FALSE ~ ifelse(fac==1, `2.5 %`, `2.5 %`*2*sd) ),
           var_imp_upper = case_when(
             scaleFlag == TRUE ~ ifelse(fac==1, `97.5 %`/sd, `97.5 %`*2),
             scaleFlag == FALSE ~ ifelse(fac==1, `97.5 %`, `97.5 %`*2*sd) )) %>%
    select(variable, factor, var_imp, var_imp_lower, var_imp_upper, p_value)
}else{
  final_result <- final_result %>% select(variable, factor, var_imp)
}

options(scipen=999, digits =3)
return(final_result)
}

# A function to create a new formula after glmnet in caret
#
# This function takes non-zero variables after glmnet and returns a
# new formula. If any dummy created from a factor is non-zero, it will
# include the factor variable
# @param modelFit required: trained caret glmnet model
# @param lambda optional: a lambda penalty (default is modelFit$bestTune$lambda)
# @keywords glmnet LASSO
# @export
# @return A formula
# @examples
# getform.glmnet(glmnetFit)
# @export

getform.glmnet <- function(modelFit, lambda = modelFit$bestTune$lambda) {
  if(modelFit$method != "glmnet") {
    stop("This function only works when you call caret train() with method set to glmnet.")
  }
  tmp_coeffs <- coef(modelFit$finalModel, lambda)
  tmp_newvars <- names(tmp_coeffs[, 1])[which(tmp_coeffs[, 1] != 0)][-1]
  used.dataframe <- eval(modelFit$call$data)
  if(is.null(used.dataframe)){
    stop("This function only works when you explicitly pass in data as a data.frame when you call caret
train() with method set to glmnet.")
  }

  allvars_factor <- enframe(attr(modelFit$terms, "dataClass"), name = "variable", value = "type") %>%
filter(type != "numeric") %>% select(variable) %>% unlist %>% unname()

  allvars_not_factor <-
  enframe(attr(modelFit$terms, "dataClass"), name = "variable", value = "type") %>% filter(type
=="numeric") %>% select(variable) %>% unlist %>% unname()

  new_factor <- NULL
  for (i in allvars_factor) {
    temp <- unique(str_extract(tmp_newvars, i))
    temp2 <- temp[!is.na(temp)]

```

```

  new_factor <- c(new_factor, temp2)
}

new_not_factor <- NULL
for (i in allvars_not_factor) {
  temp <- unique(str_extract(tmp_newvars, i))
  temp2 <- temp[!is.na(temp)]
  new_not_factor <- c(new_not_factor, temp2)
}

indvars <- c(new_not_factor, new_factor)
allvars <- attr(modelFit$terms, "term.labels")
removedvars <- setdiff(allvars, indvars)
removedvars_disp <- paste(removedvars, collapse=" ")
message(paste("glmnet removed variables:", removedvars_disp, sep=" "))
tmp_newvars3 <- paste(indvars, collapse = "+")
tmp_depvar <- paste0(all.vars(eval(modelFit$call$form)[[2]]), "~")
glmnet.fm <- as.formula(paste0(tmp_depvar, tmp_newvars3))
return(glmnet.fm)
}

# A function to create a plot after varimp.logistic()
#
# This function takes the dataframe created by varimp.logistic() and returns a
# plot. It will also pass on the dataset, so it can be used in the middle or
# the end of a pipe.
# @param .data a dataframe created by varimp.logistic() (can be piped)
# @keywords plot variable importance
# @export
# @return a plot and the unmodified dataframe created by varimp.logistic()
# @examples
# varimp.logistic(modelFit) %>% filter(var_imp > 1) %>% plotimp.logistic()
# @export

plotimp.logistic <- function(.data) {
  plotdata <- .data %>%
    mutate(variable=fct_reorder(variable, abs(var_imp)))

  temp <- plotdata %>% arrange(variable) %>% select(var_imp)
  a <- ifelse(temp >= 0, "black", "red")

  print(ggplot(data=plotdata) +
    geom_col(aes(x=variable, y=abs(var_imp), fill=factor)) +
    theme(axis.text.y = element_text(colour = a)) +
    coord_flip())
  return(.data)
}

importancia <- varimp.logistic.glm(modelo)

plotimp.logistic(importancia)

# Generar las predicciones en base al modelo #

pq = predict.glm(modelo,newdata = datos, type = "response" )

```

```
# Ajustar las predicciones a valores 0 o 1 por los thresholds y darle los niveles #

pred=as.numeric(pq>.5)
pred2=as.numeric(pq>.4)
pred3=as.numeric(pq>.6)

pred=factor(pred)
levels(pred)= c("made","missed")

pred2=factor(pred2)
levels(pred2)= c("made","missed")

pred3=factor(pred3)
levels(pred3)= c("made","missed")

# Comprobar las matrices de confusión para cada threshold #

confusionMatrix(pred, datos$shot_result)

confusionMatrix(pred2, datos$shot_result)

confusionMatrix(pred3, datos$shot_result)
```