



ANALISIS DEL VALOR xG y xGA PARA LAS PRINCIPALES LIGAS EUROPEAS

1. INTRODUCCION

La analítica ha estado presente en el mundo del fútbol desde hace años, y es muy común ver en todos los partidos de fútbol una parametrización de cómo ha ido el mismo mediante el empleo de medidas tales como: disparos a puerta, goles, disparos totales, paradas, porcentaje de posesión, etc.

En la actualidad todas estas métricas nos sirven para analizar de forma descriptiva el rendimiento de un equipo y hasta de un jugador durante, y al final, el partido. Estas métricas, aunque dan información y son muy válidas no acaban de explicar de forma razonada el porqué de los resultados de algunos partidos, por ejemplo, la causa de que un equipo con 9 disparos a puerta y otro con sólo 2 haya perdido, por ejemplo. Para arrojar algo de luz sobre estas incertidumbres analíticas y poder tener una visión más completa del juego del fútbol, desde hace unos años y como consecuencia del boom de datos disponibles (Big data) en cualquier partido se empezaron a definir y utilizar nuevas métricas que nos permiten no sólo hacer un análisis descriptivo de los datos, sino un análisis predictivo de los partidos, incluso en tiempo real; es decir, estamos consiguiendo pasar de una analítica descriptiva de los partidos o un análisis postpartido a una predictiva.

Uno de los parámetros más empleados en la actualidad es el denominado “expected goals” (xG) o goles esperados, que viene a indicarnos en base a una serie de variables que se definen para su construcción, por ejemplo, posición del atacante en el campo, ángulo del disparo, zona de disparo, etc., la probabilidad de que ese disparo acabe en un gol. Por consiguiente, disponiendo de este valor a lo largo de un partido y al final de él, podemos compararlo con los goles marcados realmente y determinar o analizar mejor como ha sido un partido: ¿Ha



tenido muchas oportunidades de gol reales el equipo? ¿Cómo cuantificamos esas oportunidades?

Lógicamente, este no es el único parámetro que se analiza hoy en día, también tenemos el parámetro “expected goals against” (xGA) equivalente a lo anterior con xG pero que analiza la probabilidad de que marquen un gol en contra.

En el estudio que a continuación se presenta se analizan estos 2 valores para los equipos de las principales ligas europeas: Ligue1, Serie A, Bundesliga, Premier, La Liga y la RFPL (rusa), y se pone foco en la Liga Española.

Posteriores estudios podrían ampliar éste, mediante el uso de más parámetros.

Los datos se han obtenido mediante web scrapping de la web understat.com, y toda la visualización se ha realizado mediante Power BI.

2. ANALISIS LIGAS EUROPEAS PERIODO 2014-2020

En el siguiente gráfico se observa los valores acumulados de xG para todas las temporadas y de cada una de las ligas:

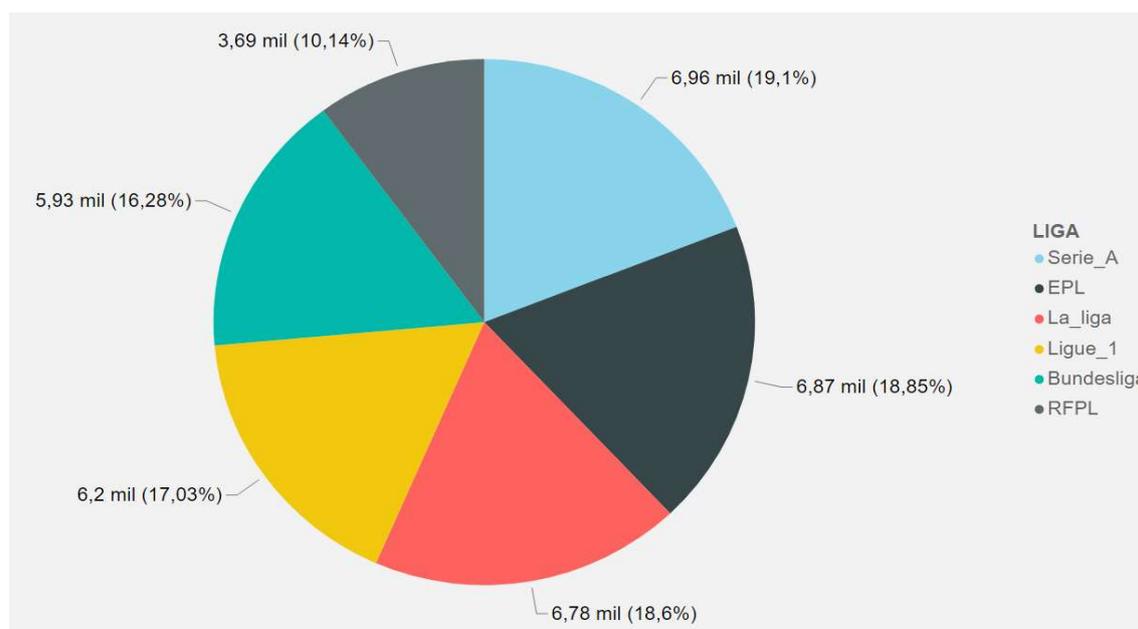


Gráfico 1: Valor acumulado de xG y porcentaje sobre el total, periodo 2014-2020



El menor valor se da para la RFPL, mientras que el resto de las ligas, salvo la Bundesliga presentan valores muy similares, siendo el valor más alto el de la Serie A. Parece contradictoria que una liga aparentemente más defensiva como es la Serie A presente valores más altos de xG que otra más ofensiva como la Bundesliga o la Liga española, veremos algunas causas que lo justifican

Hasta el 2018 era líder la Liga española

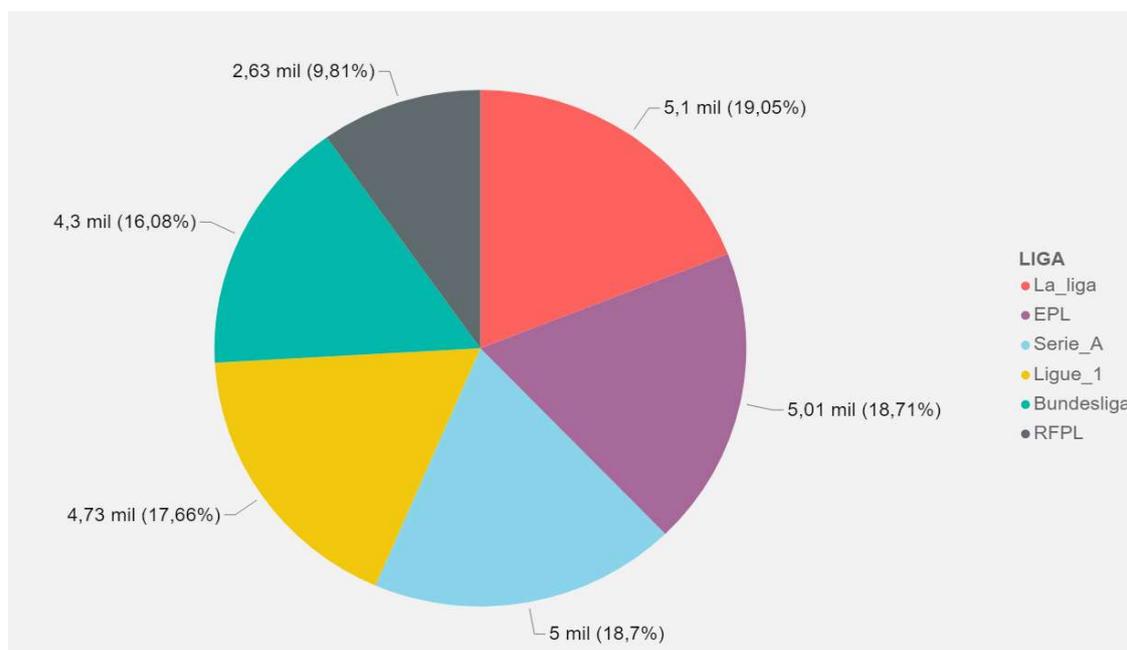


Gráfico 2: Valor acumulado de xG y porcentaje sobre el total, periodo 2014-2018

Pero observamos que a partir de ahí, el líder es la Serie A. Analicemos los factores que han podido influir en este cambio:



Barcelona y Real Madrid tienen una influencia sobre el xG total de La Liga del 50%, ver gráfico número 4, aproximadamente en cada temporada y en la suma agregada, sus valores de xG empezaron a disminuir desde el año 2018, inclusive, sobre todo el Real Madrid, ¿marcha de Cristiano Ronaldo?

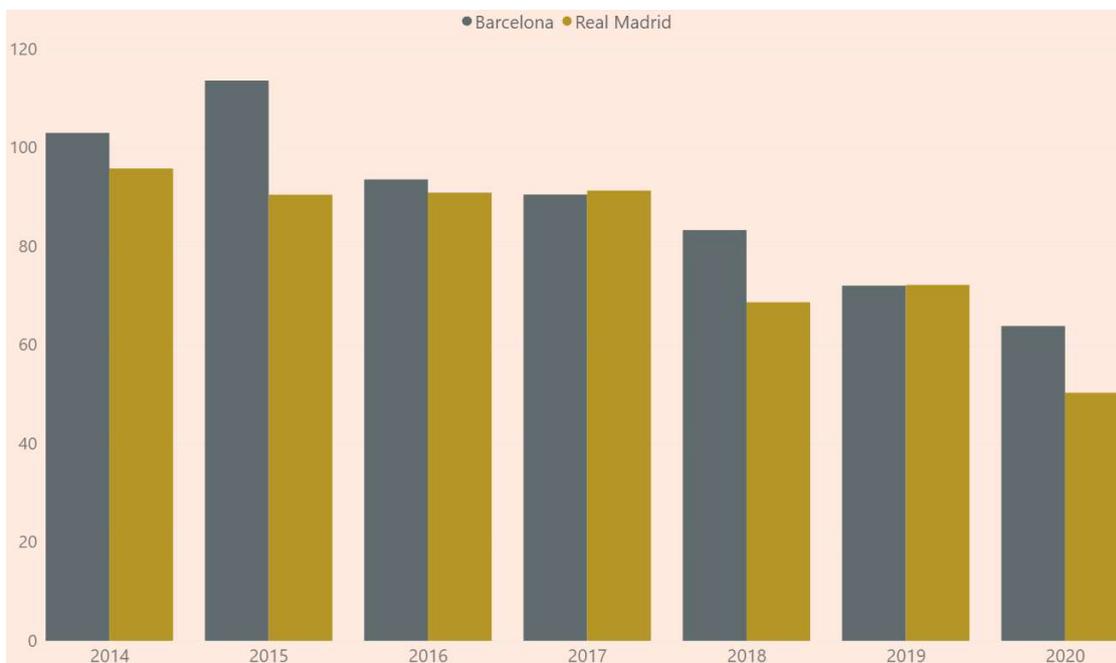


Gráfico 3: Valor de xG temporadas 2014-2020, FC Barcelona y Real Madrid

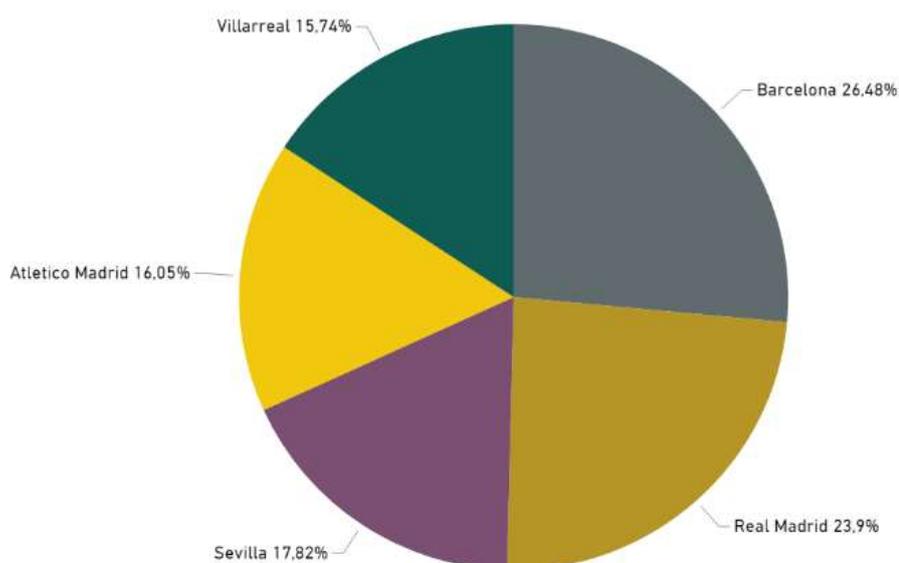


Gráfico 4: Porcentaje sobre el total de xG periodo 2014-2020, 5 equipos TOP.



En cambio, en el caso de la liga Italiana, la entrada en juego desde el año 2018 de otro equipo miuy ofensivo como el Atalanta ha hecho aumentar notablemente el valor de xG acumulado haciendo que esa liga ocupe ahora la primera posición.

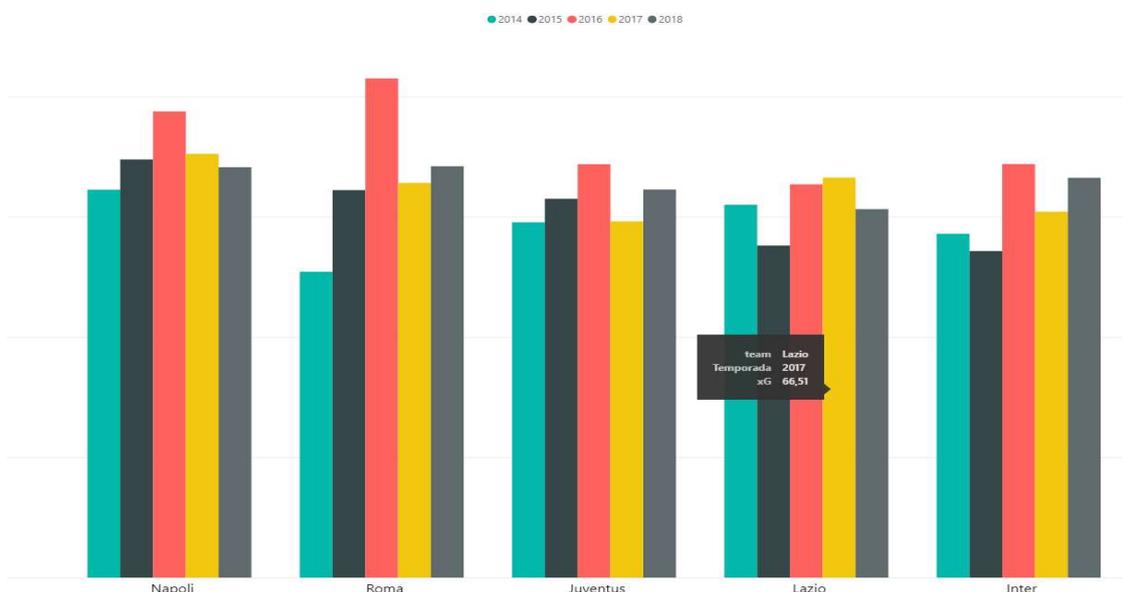


Gráfico 5: Valor de xG temporadas 2014-2018. 5 equipos TOP de la Serie A

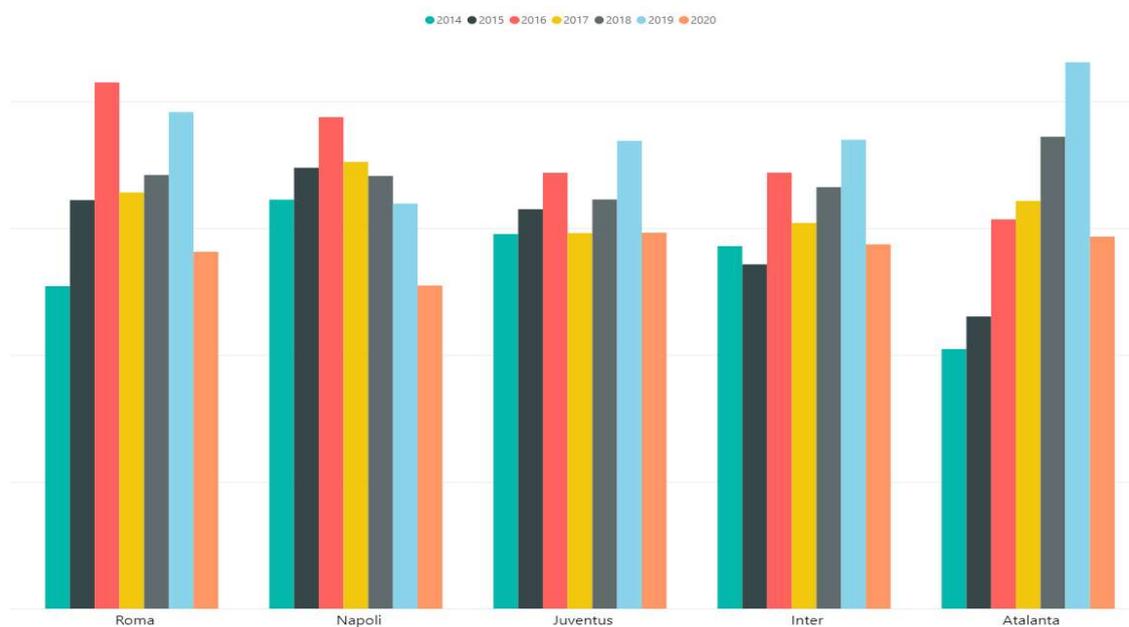


Gráfico 6: Valor de xG temporadas 2014-2020. 5 equipos TOP de la Serie A

La entrada del Atalanta en juego en vez de la Lazio, equipo menos ofensivo, hace aumentar notablemente el valor de xG, no en vano en 2018 y 2019 es líder destacado en esa faceta, y en el 2020 va igualado con la Juventus. También



parece haber una influencia aunque menor, del fichaje de Cristiano Ronaldo por la Juventus.

Si hacemos un análisis de la totalidad de los equipos de las ligas y dibujamos el top ten para el total de temporadas 2014-2020, el líder en xG es el FC Barcelona, pero si hacemos zoom desde el 2018 vemos la entrada con fuerza del Manchester City, líder, y del Atalanta equipos con gran tendencia ofensiva.

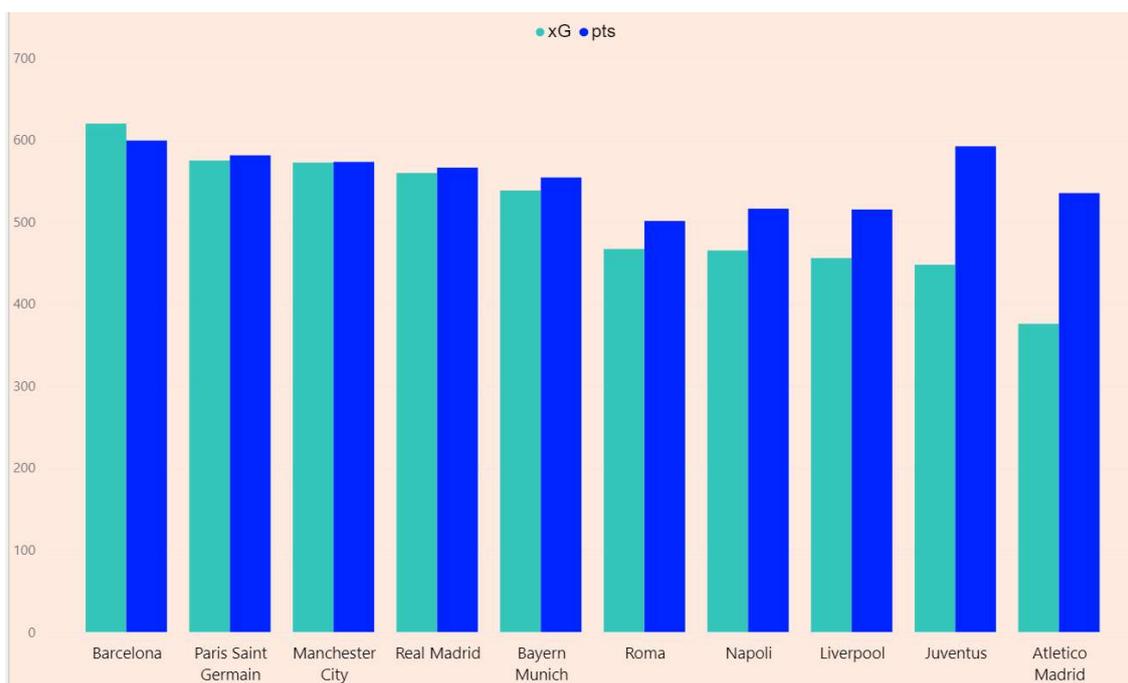


Gráfico 7: Valor de xG y puntos reales, temporadas 2014-2020. TOP 10 equipos



Para el periodo 2018-2020, tendríamos la siguiente gráfica:

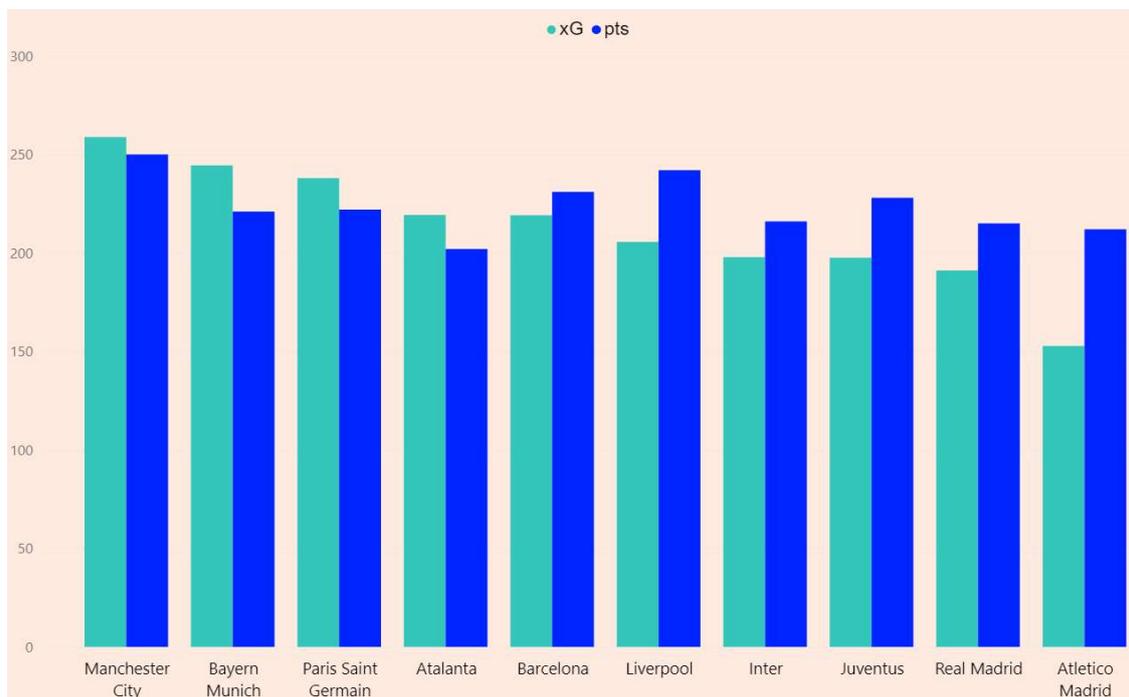


Gráfico 8: Valor de xG y puntos reales, temporadas 2018-2020. TOP 10 equipos

FC Barcelona pasa a la 5ª posición, mientras que el Real Madrid se queda en una novena posición. Está claro que nuestra Liga ha bajado de nivel en cuanto a acciones ofensivas se refiere, mientras que la Premier y la Serie A han subido en estos últimos años debido a equipos como el Manchester City y la Atalanta.



Por último se ha efectuado un análisis de clusterización para los equipos de todas las ligas y periodo 2014-2020, encontrándose la siguiente distribución para los top 50 de promedio xG durante este periodo:



Gráfico 9: Clúster por valores de xG y xGA, temporadas 2014-2020. TOP 50 equipos

Como se observa hay 3 clusters o agrupaciones, siendo la de mayor nivel futbolístico, esto es, alto xG y bajo xGA la de color rojo, formada por los equipos Top: Real Madrid, FC Barcelona, Manchester City, Bayern de Munich y Paris Saint Germain. De todos ellos, el Real Madrid es el que tiene valor más elevado de xGA, es decir le generan muchas oportunidades de gol comparativamente al resto. Destacar también equipos como la Juventus y el Atlético de Madrid que están en un mismo cluster con valores bajos de xGA, pero intermedios de xG, lo que a pesar de todo les ha llevado a ganar varios títulos e incluso ocupar posiciones destacadas en sus ligas, es decir hay diferentes formas de jugar que te pueden llevar a ganar títulos, pero en todas ellas hay dos claros denominadores, crear muchas oportunidades de gol xG, tener muy pocas oportunidades de gol en contra, bajo xGA.

En el siguiente gráfico podemos ver esas diferencias entre los equipos TOP del clúster 1, en donde destaca el alto valor de xGA para el Real Madrid:



También es destacable el Rayo Vallecano, equipo con muy alto valor de xGA, pero que está en el top 50 de todas las ligas en cuanto a general ocasiones de gol. Es conocida la influencia del juego de Paco Jémez, ofensivo pero con problemas atrás, aquí queda reflejado.

La evolución del City es extraordinaria en generar xG, veámoslo:

Gráfico 9: Clúster por valores de xG y xGA, temporadas 2014-2017. TOP 50 equipos





Observamos que el City estaba muy alejado de equipos como el Real Madrid y el FC Barcelona y a partir de 2018, entrenado por Guardiola empieza su evolución positiva y su cambio de juego. Algo similar tenemos con el Bayern de Munich que ha mejorado notablemente sus prestaciones, comprobado con el último título de la liga de campeones. El cambio de entrenador benefició a ambos equipos para bien, la marcha de Guardiola para el Bayern, la llegada de Guardiola para el City.



3. ANÁLISIS DE LA LIGA ESPAÑOLA PERIODO 2014-2020

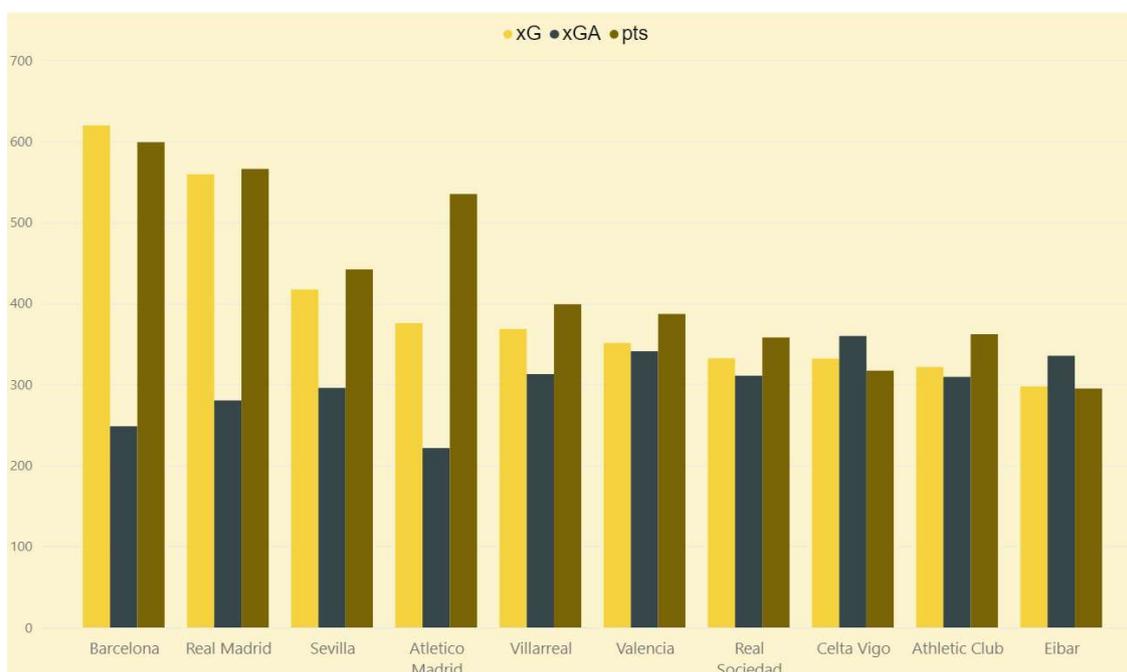


Gráfico 10: Valores de xG, xGA y puntos, temporadas 2014-2020. TOP 10 equipos

- En el análisis de todas las temporadas se observa como el FC Barcelona tiene casi tantos puntos como xG genera, igual ocurre con el Real Madrid. La presencia de jugadores con alta eficacia hace que las oportunidades xG se traduzcan en goles y puntos. La defensa del RM concede también más oportunidades que la del FCB.
- Hay que destacar el caso del Atlético de Madrid con muchos más puntos que xG, es decir aprovecha notablemente sus oportunidades además de tener un gran sistema defensivo como se ve con el valor más bajo de xGA.
- Sevilla FC tercer equipo con mayor xG.
- Hay que destacar el caso del Eibar que ocupa la décima posición, el problema del Eibar es que tiene valores muy altos de xGA, lo que hace que los puntos reales que al final se obtienen sean bajos con respecto a otros equipos.



En el siguiente gráfico vamos a hacer un zoom sobre la temporada 2020-21

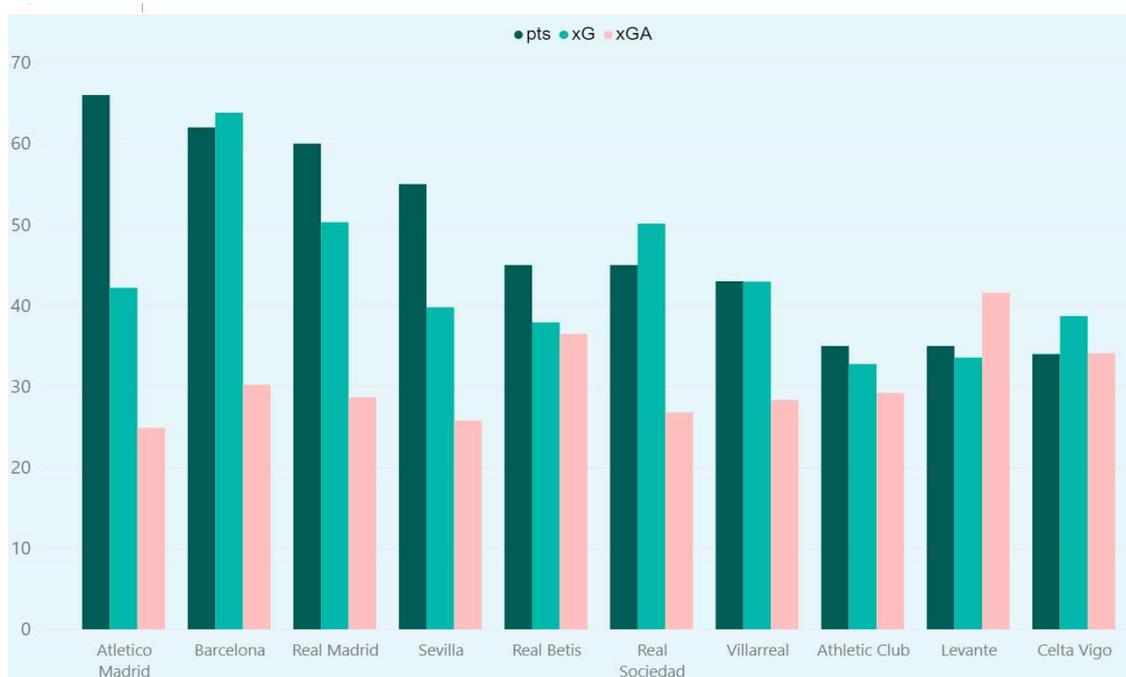


Gráfico 11: Valores de xG, xGA y puntos (ordenado por este valor), temporada 2020-21. TOP 10 equipos

- El Atlético de Madrid ocupa la primera posición principalmente por el bajo valor de xGA, pocas oportunidades de gol en su contra con alta probabilidad de tener éxito. Tiene un valor xG mucho más bajo que el FC Barcelona y el Real Madrid. Aprovecha muy bien sus oportunidades y concede muy pocas.
- FC Barcelona tiene un valor de puntos obtenidos y xG muy similares, incluso menos puntos de los que dice xG. El valor alto de xGA comparativamente con los 4 primeros equipos de la clasificación hace que esté con menos puntos de los que debería por su xG.
- Hay que destacar el Levante con el valor más alto de xGA de los 10 primeros clasificados, a pesar de lo cual tiene muchos puntos por su alto xG.



- El Valencia es un caso negativo pues sale del top ten de equipos en esta temporada cuando en el periodo 2014-2020 ocupa la sexta posición global.