

# Analiza najdroższych transferów piłkarskich

**Paweł Gorgolewski**

*AGH, Wydział Informatyki Elektroniki i Telekomunikacji  
Rachunek prawdopodobieństwa i statystyka 2021/2022*

Kraków, 26 stycznia 2022

*Ja, niżej podpisany własnoręcznym podpisem deklaruję, że przygotowałem(tam) przedstawiony do oceny projekt samodzielnie i żadna jego część nie jest kopią pracy innej osoby.*

.....

## 1 Streszczenie raportu

Raport powstał w oparciu o bazę danych: <https://www.kaggle.com/berkayalan/the-most-expensive-football-transfers> i ma na celu przeanalizowanie 54 najdroższych transferów w piłce nożnej

## 2 Opis danych

Dane zawierają 54 najdroższych piłkarzy w historii. Każdy rekord ma następujące cechy:

- **Rank** - numer w rankingu
- **Origin** - kraj pochodzenia piłkarza
- **Player** - imię i nazwisko piłkarza
- **From.Country.** - kraj z której ligi odchodził piłkarz
- **From.Club.** - klub z której ligi odchodził piłkarz
- **To.Country.** - kraj do której ligi przeniósł się piłkarz
- **To.Club.** - klub do którego przeniósł się piłkarz
- **Position** - pozycja na której gra piłkarz
- **Fee...mln.** - cena transferu w Euro

- **Fee.L.mln.** - cena transferu w Funtach
- **Year** - rok transferu
- **Born** - rok urodzenia piłkarza

Część z tych danych nie podobała mi się (nazwy także) więc zdecydowałem się zdefiniować swoje własne dane na podstawie wyżej wymienionych. Każdy rekord z zdefiniowanych danych to:

- **Player** - imię i nazwisko piłkarza
- **Origin** - kraj pochodzenia piłkarza
- **From\_Country** - kraj z której ligi odchodził piłkarz
- **From\_Club** - klub z której ligi odchodził piłkarz
- **To\_Country** - kraj do której ligi przeniósł się piłkarz
- **To\_Club** - klub do którego przeniósł się piłkarz
- **Position** - pozycja na której gra piłkarz
- **Fee\_Euro** - cena transferu w milionach Euro
- **Age** - wiek piłkarza w momencie transferu (czyli 'Year'-'Born')

## 3 Analiza danych

### 3.1 Wstępne rozeznanie się w danych

Na sam początek, warto sprawdzić jak wyglądają podstawowe statystyki związane z naszymi danymi numerycznymi, czyli *Age* oraz *Fee\_Euro*

Poniżej statystyki dla *Age*

	Name	Min	First_Quantile	Median	Third_Quantile	Max	Mean	Variance	Std	Skewness	Kurtosis
1	All	20	23.00	24.0	27.00	33	24.74074	8.233403	2.869391	0.3356312	2.861807
2	Forward	20	21.00	24.0	26.50	33	24.07143	14.640659	3.852358	0.8233145	2.981419
3	Striker	22	25.25	26.5	27.75	29	26.30000	4.900000	2.213594	-0.5286686	2.508507
4	Defender	20	23.00	23.5	25.25	27	23.87500	4.696429	2.167124	-0.2799437	2.550695
5	Goalkeeper	24	24.50	25.0	25.50	26	25.00000	2.000000	1.414214	0.0000000	1.000000
6	Midfielder	21	23.00	24.0	26.25	30	24.75000	6.723684	2.593007	0.3937355	2.166831

Widząc statystyki dla cechy *Age* można zauważyć:

1. Najbardziej rozproszoną wiekowo grupą wokół średniej są zawodnicy ofensywni
2. Najmniej rozporozoną wiekowo grupą wokół średniej są bramkarze (tak na marginesie jest ich tylko 2)

3. Rozkład (ze skośności):

- jest symetryczny dla bramkarzy
- ma długi ogon prawostronny dla napastników pomocników i zestawieniu wszystkich zawodników
- ma długi ogon lewostronny dla obrońców i napastników

4. Rozkład (z kurtozy): dla każdego przypadków w danych istnieje więcej dodatnich wartości odstających niż w przypadku rozkładu normalnego

Teraz spojrzmy na statystyki czegoś bardziej interesującego - pieniędzy, czyli *Fee\_Euro*:

	Name	Min	First_Quantile	Median	Third_Quantile	Max	Mean	Variance	Std	Skewness	Kurtosis
1	All	60.0	64.625	75.00	86.500	222	83.15926	910.4859	30.17426	2.6634505	11.249814
2	Forward	62.5	72.500	100.00	116.250	222	106.52143	2066.5141	45.45893	1.4072592	4.243074
3	Striker	60.0	63.900	67.50	81.725	90	72.05000	124.9672	11.17887	0.3961447	1.574216
4	Defender	60.0	65.150	71.50	81.125	87	73.08750	100.0041	10.00021	0.1715803	1.548953
5	Goalkeeper	62.5	66.875	71.25	75.625	80	71.25000	153.1250	12.37437	0.0000000	1.000000
6	Midfielder	60.0	63.500	73.50	76.250	145	77.58000	464.2227	21.54583	1.9397742	6.200669

Rozważając te same statystyki dla *Fee\_Euro*:

1. Najbardziej rozproszoną grupą pod względem kosztu transferu wokół średniej są zawodnicy ofensywni
2. Najmniej rozporozoną grupą pod względem kosztu transferu wokół średniej są obrońcy
3. Rozkład (ze skośności):
  - jest symetryczny dla bramkarzy
  - ma długi ogon prawostronny dla reszty grup (zauważmy, że obrońcy mają prawie symetryczny rozkład - skośność bliska 0)
4. Rozkład (z kurtozy): dla każdego przypadków w danych istnieje więcej dodatnich wartości odstających niż w przypadku rozkładu normalnego

Łatwo zauważyć, że maksymalna kwota transferu to aż 222 miliony Euro! Patrząc na takie sumy aż chce się zobaczyć TOP 10 najdroższych transferów w piłce nożnej. Oto one:

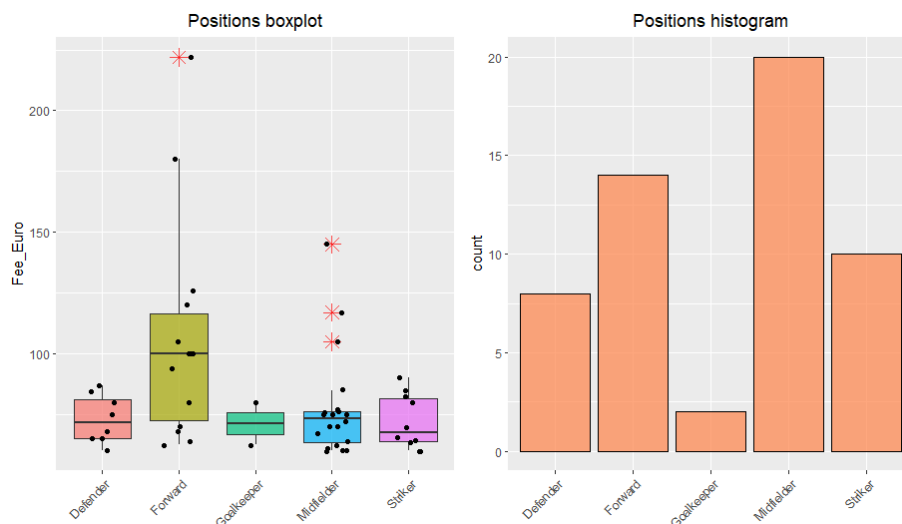
	Player	origin	From_Country	From_Club	To_Country	To_Club	Position	Fee_Euro	Age
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<int>
1	Neymar	Brazil	Spain	Barcelona	France	Paris Saint-Germain	Forward	222	25
2	Kylian Mbappé	France	France	Monaco	France	Paris Saint-Germain	Forward	180	20
3	Philippe Coutinho	Brazil	England	Liverpool	Spain	Barcelona	Midfielder	145	26
4	Joao Félix	Portugal	Portugal	Benfica	Spain	Atlético Madrid	Forward	126	20
5	Antoine Griezmann	France	Spain	Atlético Madrid	Spain	Barcelona	Forward	120	28
6	Jack Grealish	England	England	Aston villa	England	Manchester City	Midfielder	117	26
7	Paul Pogba	France	Italy	Juventus	England	Manchester United	Midfielder	105	23
8	Gusmane Dembélé	France	Germany	Borussia Dortmund	Spain	Barcelona	Forward	105	20
9	Gareth Bale	Wales	England	Tottenham Hotspur	Spain	Real Madrid	Forward	100	24
10	Cristiano Ronaldo	Portugal	Spain	Real Madrid	Italy	Juventus	Forward	100	33

Porównując teraz wcześniejsze statystyki i TOP 10 transferów dzieje się coś zadziwiającego. Jak to się stało, że drugi najdroższy piłkarz świata to najmłodszy w zestawieniu 20-latek!? Tak samo może ktoś spytać, dlaczego na 10 miejscu znajduje się najstarszy piłkarz (33 lat) i to za aż 100 milionów Euro! Dodatkowo, jeśli kogoś nie interesuje tylko wiek piłkarza i kwota transferu - prawie wszyscy zawodnicy w TOP 10 są piłkarzami ofensywnymi.

W tym tkwi sedno problemu, w najdroższych transferach często nie chodzi tylko o umiejętności piłkarza i wiek, ale także jego rozpoznawalność, klub do jakiego oraz z jakiego przychodzi a także pozycja na której występuje. To czego nie ma w bazie, ale też ma znaczenie to długość obecnego kontraktu (im krótszy kontrakt ma piłkarz tym mniej się za niego płaci - po skończeniu kontraktu można go wziąć za darmo. Przykład Lewandowski z Borussi do Bayernu)

### 3.2 Pozycja zawodnika a cena

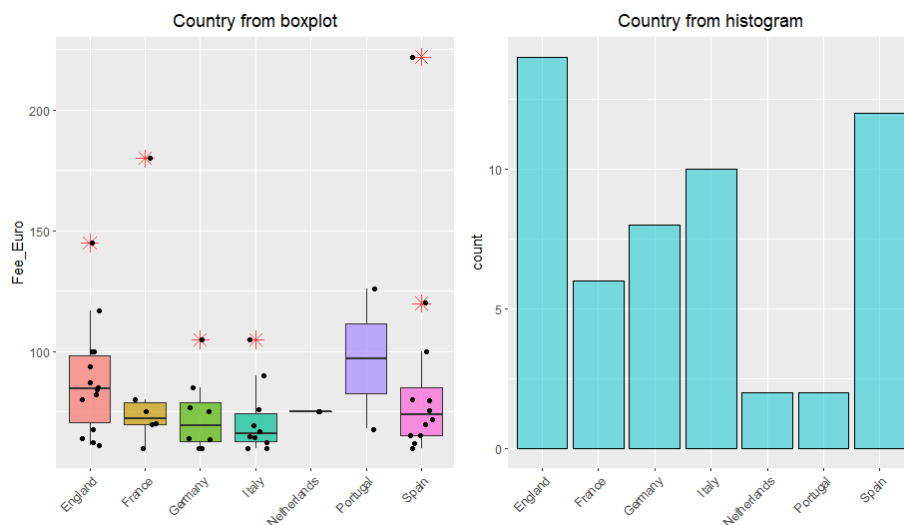
Wiemy, że Neymar i Mbappe to najdrożsi piłkarze na świecie oraz oboje są zawodnikami ofensywnymi. Z poprzednich statystyk wiemy już nieco o podziale na pozycję, lecz mimo wszystko zilustrujmy to sobie:



Tak jak wcześniej można było wywnioskować ze statystyk - zdecydowanie najdrożsi są zawodnicy ofensywni. Reszta pozycji nie ma większej różnicy między sobą - można wyróżnić tylko bramkarzy, ponieważ istnieją tylko 2 transfery z tej pozycji co jest lekko niemiernorodne. Ku mojemu zdziwieniu, kupiono więcej pomocników niż atakujących! Oczywiście tak jak widać istnieją wyjątki od reguły i widać to na boxplotach - są to ceny zawodników które nie pasują do pozostałych w swojej kategorii. Ceny trzech pomocników i jednego z zawodników ofensywnych zdefiniowano jako wartości odstające (czerwone gwiazdki)

### 3.3 Kraj z jakiego przybywa zawodnik a cena

Nasze wiadomości rozszerzyły się o różnice między pozycjami. Jak wygląda sytuacja, gdy w podobny sposób rozważymy kraj z którego przybywa zawodnik?

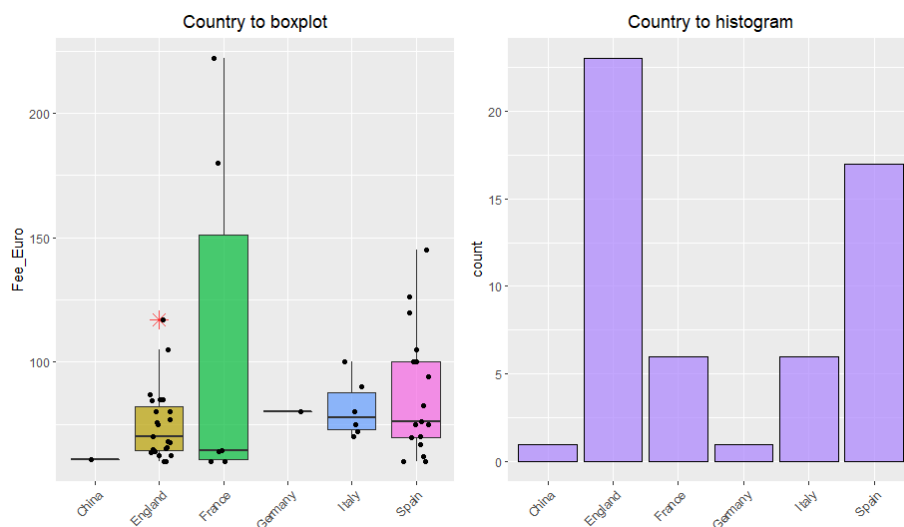


Tak jak widac na histogramie, niemalże wszystkie transfery są z TOP 5 lig europejskich: Anglia, Hiszpania, Niemcy, Włochy, Francja. Są też po dwa transfery z Portugalii oraz Holandii ale można z tego wysunąć prosty wniosek - tych 4 piłkarzy musiało się wybitnie wyróżniać na tle drużyny. A teraz dlaczego głównie z najmocniejszych lig? Odpowiedź jest prosta - piłkarz wyróżniający się w mocnej lidze, najprawdopodobniej bez problemu poradzi sobie w innej (bądź tej samej) mocnej lidze i będzie odgrywał w niej pierwszoplanową rolę.

Mimo jednego 'outliera' z Francji to Anglia i Hiszpania ma najbardziej cenionych piłkarzy na świecie, z przewagą angielską. Hiszpanią rządzi fakt posiadania dwóch klubów: Barcelony oraz Realu Madryt. Anglia ma 6 topowych i bogatych drużyn, zaś reszta ligi to również mocne zespoły. Nie bez powodu kursy u bukmacherów są wysokie na mecze tej ligi - tam każdy może wygrać z każdym (trochę jak nasza Ekstraklasa ale o wiele lepsze mecze)

### 3.4 Kraj do którego przybywa zawodnik a cena

Wiemy, że to piłkarze z Anglii i Hiszpanii są najbardziej cenieni. A gdzie najwięcej piłkarzy odchodzi ze swoich klubów?

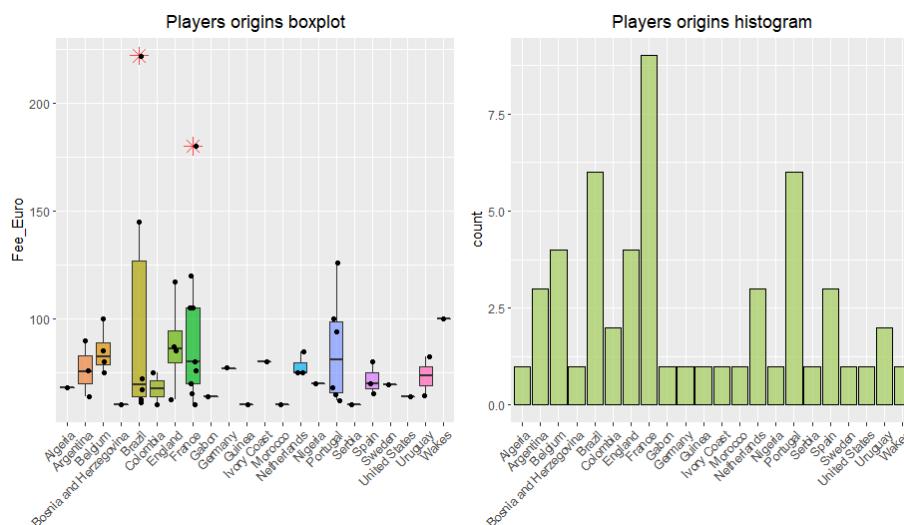


No i tym razem bez zaskoczenia - do Anglii. Pod względem liczebności transferów, tylko Hiszpania jako tako jej dorównuje. Jeżeli jednak popatrzymy na ceny transferów, to Hiszpania i Włochy są na najwyższym poziomie (mimo wszystko faworyzowałbym Hiszpanię przez o wiele większą liczbę transferów)

Patrząc na wykresy, można dostać lekkiego szoku. Dlaczego Neymar i Mbappe, czyli dwóch najdroższych zawodników na świecie, udali się do Francji? Oboje są świetnymi piłkarzami, więc dlaczego nie chcieli iść do Anglii czy Hiszpanii? W praktyce chodziło tu o kwestie osobiste (Neymar chciał być numerem 1, Mbappe chciał zostać we Francji) oraz standardowo, o pieniądze - tylko klub PSG był w stanie zapłacić tak kosmiczne pieniądze za piłkarzy i dać im odpowiednie pensje.

### 3.5 Kraje z których pochodzą zawodnicy a cena

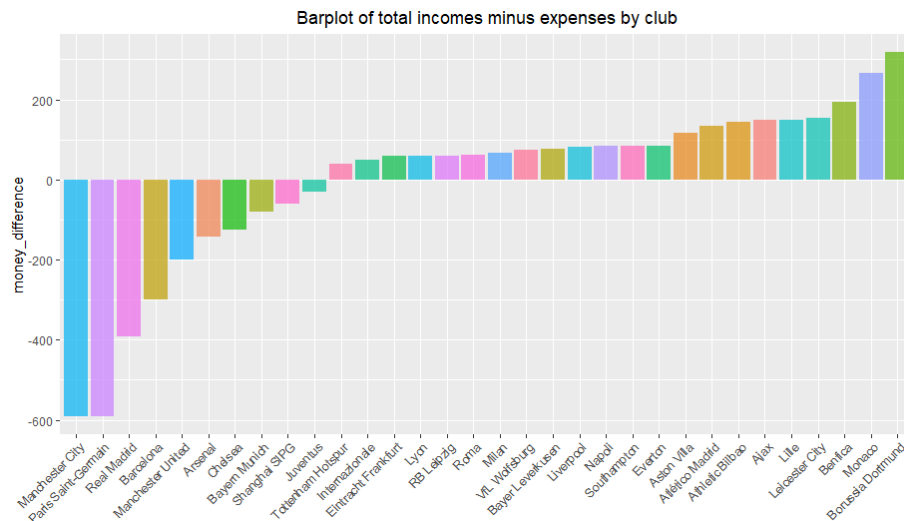
Neymar pochodzi z Brazylii, Mbappe z Francji. Wiemy już że ich ceny są mało porównywalne do reszty, lecz sprawdzić zawsze warto - jakiego pochodzenia piłkarzy kupuje się najdrożej.



Wyróżniającym się krajem jest Francja - ma najwięcej transferów. Brazylia i Portugalia dopełniają podium. Oprócz tego, żaden z krajów nie wyróżnia się niczym szczególnym (poza Brazylią i Francją które mają swoje rekordowe transfery). Można zatem stwierdzić, że kraj pochodzenia nie ma większego znaczenia jeżeli chodzi o cenę transferu (równie dobrze zamiast Neymara za 222 mln mógłby się tam znaleźć super wybitny gracz z Polski czy San Marino). Jedyne co tutaj gra rolę to prawdopodobieństwo wystąpienia tak drogiego (czyli uzdolnionego) zawodnika dla danej narodowości - tak jak każdy się spodziewa dla San Marino na pewno będzie jednym z najmniejszych.

### 3.6 Kluby - przychód i wydatki z wielkich transferów

A co gdyby ktoś się zastanawiał - jak wypadł każdy z klubów pod względem finansowym w tym zestawieniu? Wiedząc, że PSG kupiło Mbappe i Neymara, zakładam, że ten klub będzie miał większy wydatki niż przychody. Zobaczmy jak to się ma dla każdego z klubów:

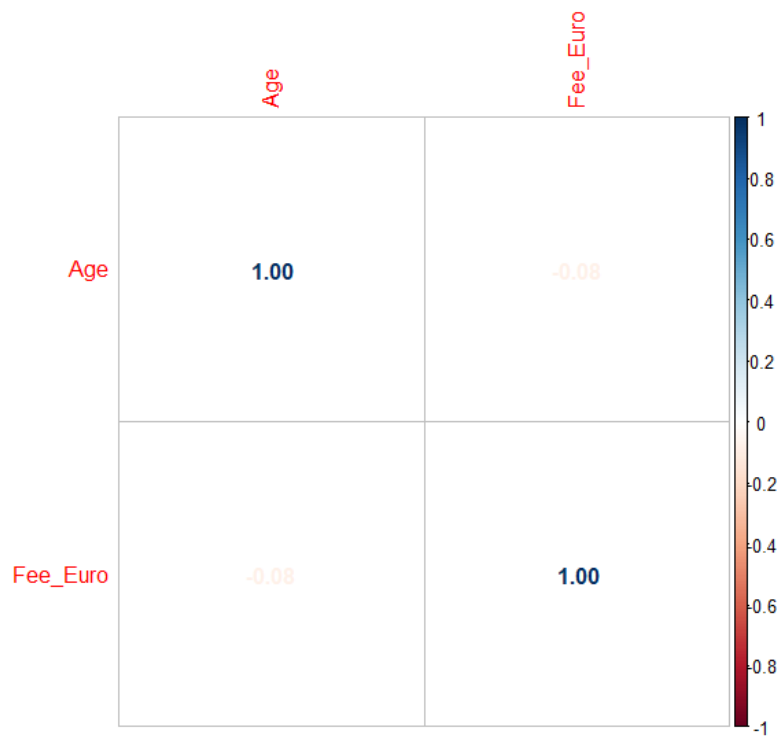


Tak jak to było wcześniej powiedziane, drogie transfery wykonywane są tylko do mocnych lig i/lub bogatych zespołów. Nikogo tu pewnie nie dziwi, że 'na minusie' są takie kluby jak: Manchester City, PSG, Chelsea, Real Madryt, Barcelona. Z drugiej strony na plusie są zespoły które są mocne, grają w dobrych/najlepszych ligach ale nie są w stanie walczyć o ligę mistrzów, zapłacić wymaganej przez piłkarza pensji lub tak jak Ajax są nastawione na szkolenie wybitnych młodych piłkarzy i później ich sprzedawać (np. Ibrahimović).

### 3.7 Korelacja zmiennych

Będąc przy transferach, nie można zapomnieć o rozważeniu czy wiek i cena idą ze sobą w parze. Poniżej przedstawiam macierze korelacji dla wieku i kwoty transferu:

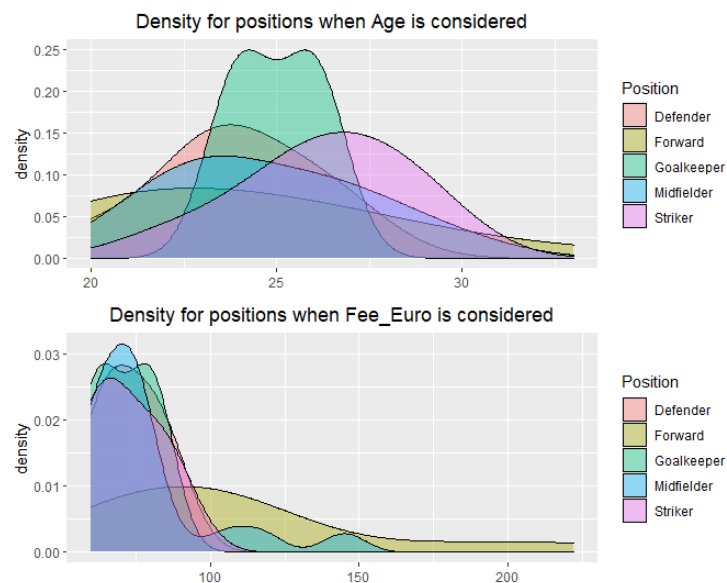




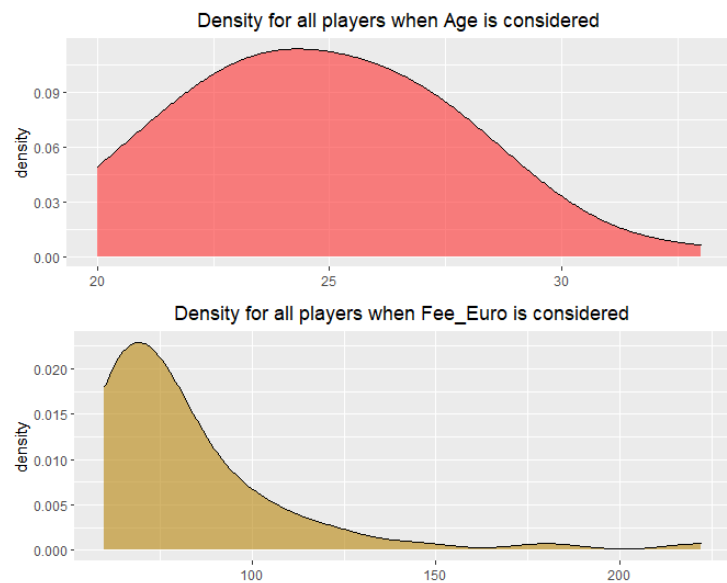
No i tak jak można się było spodziewać - może i jest jakieś powiązanie między tymi zmiennymi, ale na pewno nie jest ono liniowe.

### 3.8 Gęstości i testy przynależności do rozkładu

Rozważmy teraz gęstości naszych rozkładów z podziałem na pozycje:



Nie można ukryć, wyglądają ciekawie. Tak samo interesujące jest, czy po połączeniu (czyli wyliczeniu gęstości dla wszystkich zawodników w bazie) dostaniemy rozkłady podobne do normalnego. Spójrzmy na kolejne dwa wykresy:



Zacznijmy od rozkładu przy rozważanym wieku piłkarza. Aby sprawdzić, czy rozkład jest normalny użyję testu *Shapiro-Wilk'a*. Jeżeli wartość  $p$  będzie większa niż 0.05 (wartość tą zmienia się dla wartości  $W > 0.98$ ) to rozkład

można uznawać za normalny.

```
> shapiro.test(players$Age)

      shapiro-wilk normality test

data:  players$Age
W = 0.96777, p-value = 0.1536
```

Wyniku testu są pozytywne - rozkład wieku piłkarzy można traktować jako rozkład normalny.

Teraz czas na sprawdzenie rozkładu ceny piłkarzy. Wszystko zadziała tutaj w analogiczny sposób, lecz spodziewamy się wartości  $p < 0.05$  - wykres gęstości o wiele mniej przypomina rozkład normalny. Spójrzmy na wyniki:

```
> shapiro.test(players$Fee_Euro)

      shapiro-wilk normality test

data:  players$Fee_Euro
W = 0.69962, p-value = 3.252e-09
```

Wyniki testów potwierdzają przypuszczenia.  $P$  jest zdecydowanie mniejsze od 0.05 więc rozkład ten nie może być rozkładem normalnym.

## 4 Wnioski

Wnioski płynące z przeprowadzonej analizy, są następujące:

- Najbardziej pożądani są gracze ofensywni
- Francja ma najliczniejszą reprezentację w zestawieniu najdroższych zawodników
- Bogate i czołowe kluby wydają najwięcej na rekordowych transferach
- Kluby z dobrą szkółką młodzieżową (np. Ajax) oraz kluby o dużym potencjale (np. Borussia) mają największy porfit z rekordowych transferów
- Tak drodzy piłkarze zazwyczaj nie skończyli 30 lat
- Zależność między wiekiem a ceną na pewno nie jest liniowa
- Rozkład wieku piłkarzy można traktować jako rozkład normalny