

ANALYSE DES WELTWEITEN FUSSBALMARKTES

Justin Mester

Matrikelnummer : 30016330 |

Inhaltsverzeichnis

Einleitung	2
Analyse	3
<i>Describe</i>	<i>3</i>
<i>Boxplots</i>	<i>4</i>
<i>Bar Graph</i>	<i>6</i>
<i>Korrelation</i>	<i>8</i>
<i>Marktwertkarte</i>	<i>8</i>
<i>KMeans</i>	<i>9</i>
<i>Pie Chart</i>	<i>10</i>
<i>Relativ/Absolut Scatter dot Chart</i>	<i>11</i>

Einleitung

In der folgenden Analyse werden Statistische Verfahren auf die Daten des Fußballtransfermarktes angewandt. Hieraus sollte Rückschlüsse gezogen werden können, welchen Einfluss verschiedene Spielerfakten auf den Marktwert eines Spielers haben. Hierzu wurden die Daten der Website Transfermarkt.de mithilfe eines Crawlrs gespeichert. Abgerufen wurden die Daten am 20.06.2020 weshalb die Aktualität abweichen kann.

Die abgerufenen Daten wurden mithilfe des Jupyter Notebooks 2_clean_transfermarkt aufgearbeitet, sodass eine reibungslose Analyse möglich ist. Die Analyse selber erfolgt anschließend in dem Jupyter Notebook 3_analyse_transfermarkt.

Der Marktwert eines Spielers ist zu vergleichen mit dem Marktwert eines „Markenunternehmens“. Bestimmt wird der Marktwert aus verschiedenen Analysen und Studien die Experten durchführen. Diese nehmen Einfluss auf das Talent, welches ein Spieler hat, wie alt dieser ist, seine Leistungsdaten, die er vorweisen kann, in welchem Land der Spieler spielt und gespielt hat sowie sich der Name des Spielers vermarkten lässt. Das heißt, wie viele Trikots oder Ähnliches mithilfe dieses Spielers verkauft werden können. Nach dem Marktwert und dem dann noch bestehenden Vertrag den ein Spieler besitzt wird bei einem „Kauf“ eines Spielers so auch die Ablösesumme bestimmt die fällig wird, wenn ein Spieler aus seinem laufenden Vertrag austreten will um dann für einen anderen Verein spielen zu können..

Analyse

Describe

Zunächst wurde auf die bereinigten Daten die Describe Funktion angewandt. Diese brachte folgende Ergebnis:

	Values	Size	Age	Lon	Lat
count	20914.000000	17816.000000	20914.000000	20914.000000	20914.000000
mean	3.064455	182.398181	25.978866	12.200817	41.717974
std	9.359818	6.677803	4.721866	28.632888	18.934170
min	0.025000	158.000000	16.000000	-123.113953	-41.500083
25%	0.150000	178.000000	22.000000	4.269680	40.792949
50%	0.325000	183.000000	26.000000	14.248783	46.314395
75%	1.200000	187.000000	29.000000	23.728305	51.507322
max	180.000000	204.000000	45.000000	176.167505	70.049628

Abbildung 1 / df.describe von analyse_transfermarkt.ipynb

Anhand dieser Tabelle kann man bereits erste Statistische Werte ablesen. So zeigte diese zum einen mit dem Wert 20914 bei **count** An wie viele Spielerdaten mithilfe des Crawlers erfasst wurden. Die Differenz der **count** Kennzahl zwischen **Values** und **Size** kommt dadurch zustande, dass nicht alle Größen der Spieler bekannt sind und somit eine Differenz von 3098 entsteht. Die **Lon** und **Lat** brauchen hierbei nicht beachtet werden, da diese ausschließlich zur Standort Bestimmung eines jeden Spielers dient. Der Mittelwert der Daten lässt sich in der **mean** Spalte ablesen. So liegt die Durchschnittsgröße der Spieler bei ca. 182,40, das Durchschnittsalter bei 25,98 Jahren und die Durchschnittliche Marktwert bei 3,064 Mio €. Zusätzlich kann man bereits die Obere und Untere Grenzen der Daten ablesen. So ist der Kleinste Spieler 158,00 cm Groß und der größte 204,00 cm. Bei den Marktwerten liegt die untere Grenze bei 0,025 Mio € und der höchste Marktwert liegt bei 180 Mio. €. Um die Maximal und Minimalwerte sowie die Durchschnittswerte nochmals grafisch darstellen zu können wurden Box Plots sowie ein Punktediagramm geplottet.

Boxplots

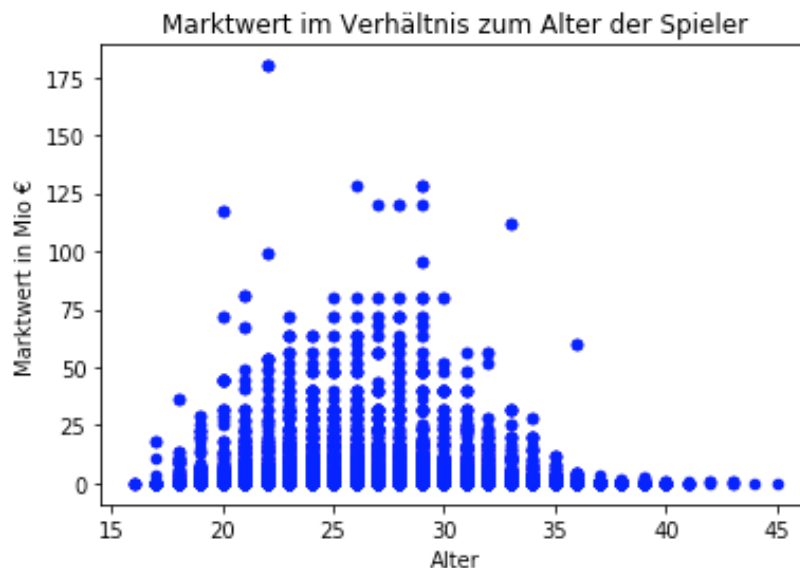


Abbildung 2 Scatter Plot Marktwerte und Alter *analyse_transfermarkt.ipynb*

Das Punktediagramm zeigt nochmals die Verteilung der Jahre in der Fußballbranche sowie die Marktwert Höhen. Man kann hier auch bereits erste Erkenntnisse daraus ziehen, dass die höchsten Marktwerte der Spieler im Alter von 22 bis 30 erzielt werden. Des weiteren kann man sehen, dass der Spieler mit 175 Mio. € Marktwert einen großen Abstand zu den darunter liegenden hat.

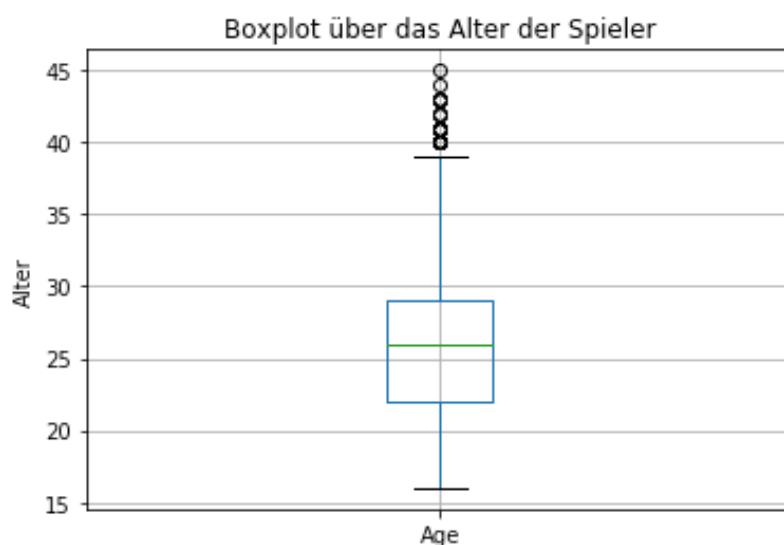


Abbildung 3 Boxplot über das Alter der Spieler von *analyse_transfermarkt.ipynb*

Die Abbildung 3 visualisiert nochmals das Alter der Spieler mithilfe eines Box Plots. Dieses bildet das untere und obere Quartil ab. Die Quartile weisen dabei folgende Eigenschaften aus:

- Unteres Quartil: Es sind maximal ein Viertel aller Werte unterhalb dieses Quartils und maximal drei Viertel über diesem Wert. Im Schaubild (Abbildung 3) ist dies die untere Waagerechte Linie des Rechtecks.
- Oberes Quartil: Hier sind maximal drei Viertel aller Werte unterhalb dieses Wertes sowie maximal ein Viertel über ihn. Im Schaubild (Abbildung 3) ist dies die obere Waagerechte Linie des Rechtecks.

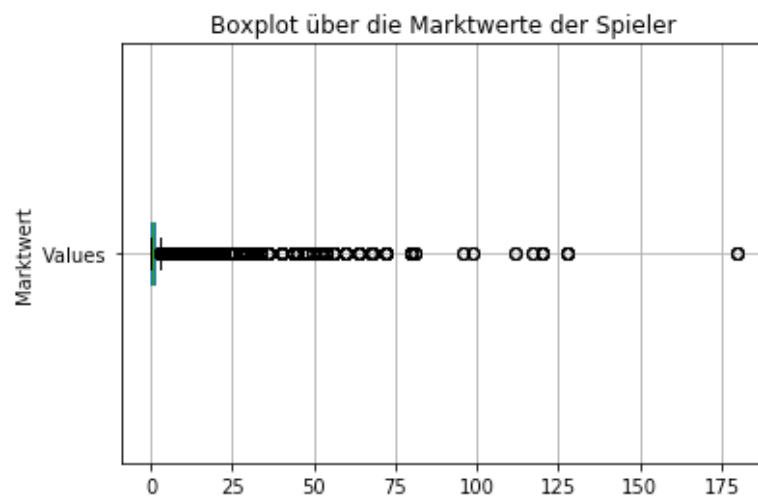


Abbildung 4 Boxplot über den Marktwert der Spieler von `analyse_transfermarkt.ipynb`

Bar Graph

Die folgenden Balkendiagramme (Bar Graph) sollen die Top 10 an Clubs sowie Spieler von „analyse_transfermarkt.ipynb“ visualisieren. Das Balkendiagramm eignet sich gut, um die hier bestehenden Differenzen der einzelnen Elemente anzeigen zu lassen.

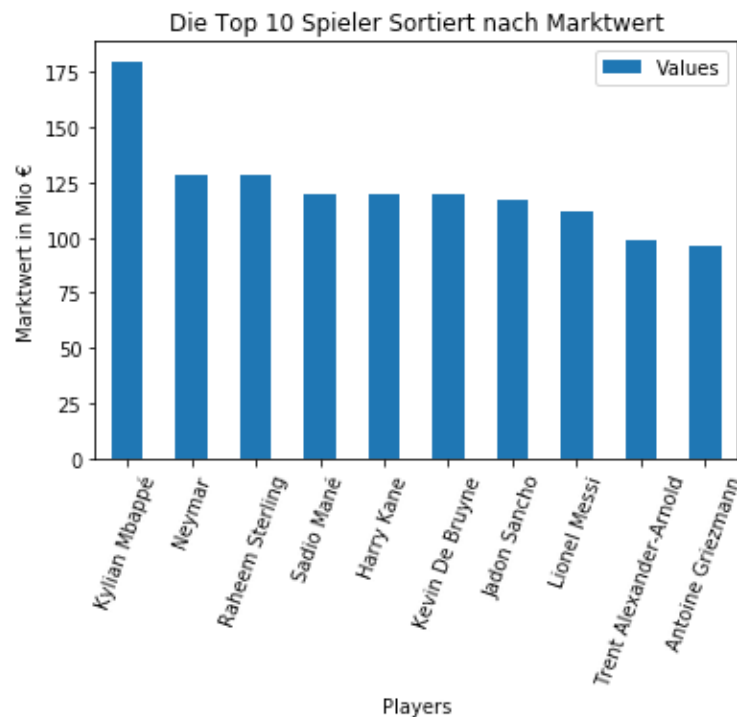


Abbildung 5 Bar Graph über die höchsten Marktwerte der Spieler von analyse_transfermarkt.ipynb

Man sieht hier bereits den großen Abstand welcher Platz 1 und 2 trennt. Hier ist eine Differenz von knapp 50 Mio. € zuerkennen. Im Gegensatz dazu ist die Differenz zwischen den nachfolgenden Plätzen relativ gering. Was die Besonderheit des Spielers „Kylian Mbappe“ andeutet.

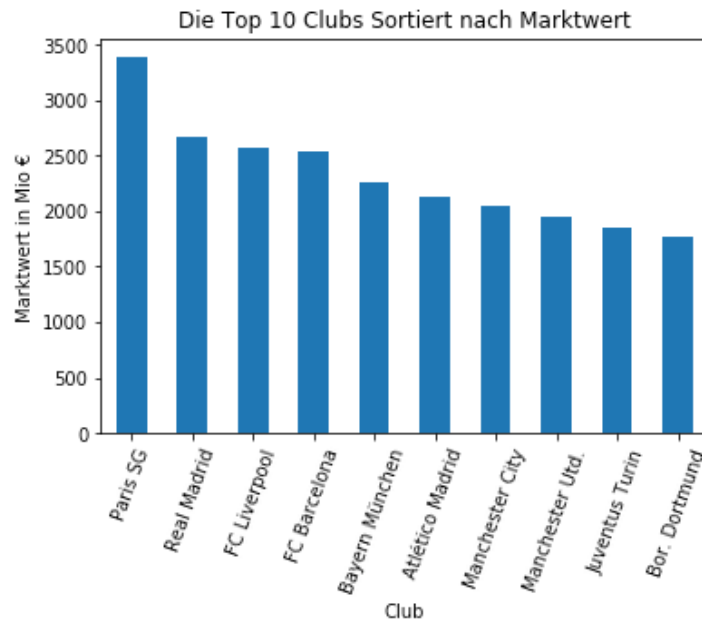


Abbildung 6 Bar Graph über die Summe der Marktwerte je Club von analyse_transfermarkt.ipynb

Auch die Sortierung nach Clubs gibt es ein Team was hervorsteht dies ist hier Paris SG. Eine weitere Auffälligkeit fällt auf, wenn man sich das Herkunftsland der Teams anguckt.

Frankreich	Spanien	England	Deutschland	Italien
Paris SG	Real Madrid, FC Barcelona, Atletico Madrid	FC Liverpool, Manchester City, Manchester United	Bayern München, Bor. Dortmund	Juventus Turin

Paris ist auf Platz 1. Der Marktwerte jedoch auch der einzige Vertreter der Französischen Liga (Ligue 1). Die Spanische, Englische und auch die Deutsche Liga haben mindestens 2 Vertreter in den Top 10. Zeigen tut dies, das in diesen Ligen mehrere Top Clubs vorhanden sind und somit ein größerer Konkurrenz Kampf im Bezug zur Meisterschaft besteht.

Korrelation

Eine Korrelation hilft dabei Beziehungen zwischen gewissen Merkmalen zu analysieren. Dabei wird ein Korrelationskoeffizient berechnet welcher einen Wert zwischen -1 und +1 annehmen kann, und dabei eine Aussage über die Beziehung der Merkmale gibt. Die Folgende Matrix visualisiert das Ergebnis der Korrelation nochmals mit Farbe. Hierdurch lässt sich die Ausprägung der Beziehungen Übersichtlich darstellen. Je heller ein Quadrat in der Matrix ist desto stärker ist die Beziehung zwischen den Merkmalen.

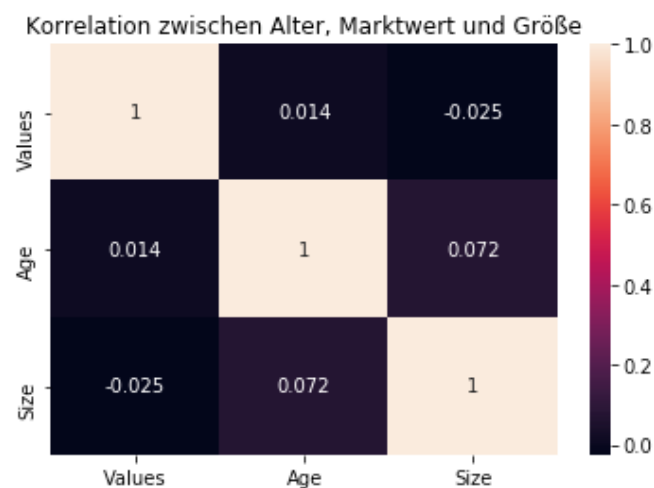


Abbildung 7 Korrelation zwischen Alter, Marktwert und Größe von `analyse_transfermarkt.ipynb`

Anhand der obigen Grafik lässt sich erkennen, dass die Beziehungen zwischen den Merkmalen nicht sehr stark ist und somit kein direkter Zusammenhang zwischen den Marktwert und dem Alter oder der Größe eines Fußballspielers.

Marktwertkarte

Mithilfe der Marktwert Karte soll eine Übersicht darüber gegeben werden, in welchem Verhältnis der Marktwert mit dem Herkunftsland der Spieler steht. So kann man anhand der Färbung des jeweiligen Landes erkennen wie hoch die Summe der Marktwerte ist, aus die ein Spieler stammt. Die Karte deckt dabei alle 195 Staaten der Erde ab und ist dabei interaktiv steuerbar. So kann man sich innerhalb der Karte bewegen.

KMeans

Durch den KMean Algorithmus lassen sich Gruppierungen aus Datensätze bilden. Dabei ist der KMean Algorithmus dafür gemacht großer Datenmenge sehr effizient zu bearbeiten und somit Perfekt für die hier angewandten Daten geeignet.

Ziel einer Clusteranalyse ist es Gruppen aus den Datensätzen zu bilden und so ähnliche Werte zusammenzufassen. Die Reihenfolge wie der KMean Algorithmus erfolgt ist dabei immer gleich und erfolgt in 4 Schritten:

1. Zunächst werden die sogenannten K-Punkte gewählt diese dienen als Anfangszentren der Berechnung
2. Im 2. Schritt werden die Datenpunkte dann den verschiedenen Clustern zugeordnet. Dies erfolgt dabei auf den Abstand zu den verschiedenen Zentren.
3. Nun folgt eine Neu Berechnung der Clusterzentren, um diese noch genauer zuzuordnen.
4. Nun werden die Schritte ab Schritt 2 solange wiederholt, bis keine Änderung der Zentren mehr erfolgt.

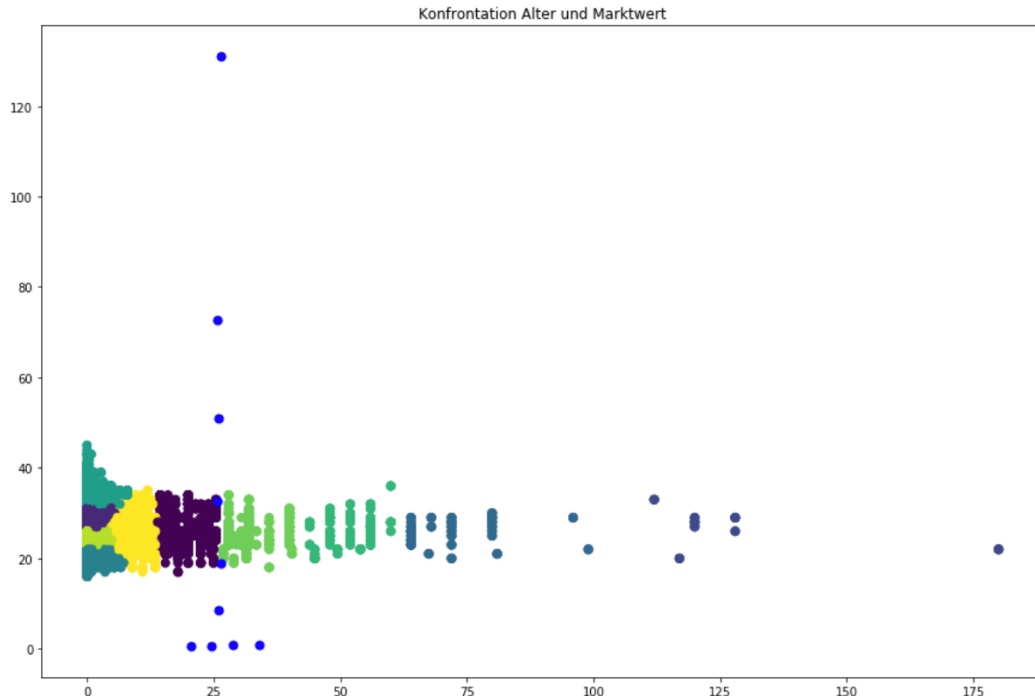


Abbildung 8 KMean angewandt auf Transfermarkt Daten von `analyse_transfermarkt.ipynb`

In der obigen Abbildung wurde der KMean auf den Datensatz angewandt und in einem Diagramm dargestellt. Auf der Y-Achse befindet sich hierbei das Alter und auf der X-Achse der Marktwert der Spieler.

Pie Chart

Da im Fußball die Position der Spieler sehr unterschiedlich sind und oftmals auch unterschiedlich bewertet werden, wurde hier zunächst ein Kreisdiagramm erstellt um alle Positionen, welche im Mittelfeld existieren, anzuzeigen und diese gleichzeitig mit der dazugehörigen Häufigkeit visualisierbar zu machen.

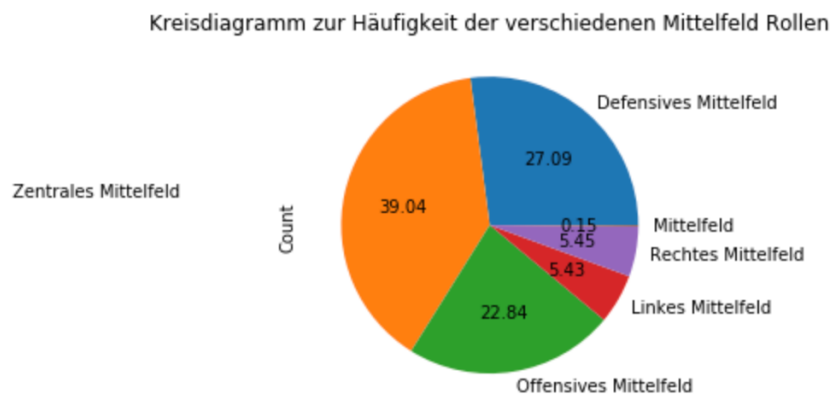


Abbildung 9 Pie Chart Häufigkeit der verschiedenen Mittelfeld Rollen von `analyse_transfermarkt.ipynb`

	Position	Sum_Value	Count	Mean
7	Defensives Mittelfeld	4480.790	1586	2.825214
8	Zentrales Mittelfeld	9702.550	2286	4.244335
9	Offensives Mittelfeld	4388.175	1337	3.282105
37	Linkes Mittelfeld	721.025	318	2.267374
365	Rechtes Mittelfeld	629.550	319	1.973511
3924	Mittelfeld	1.050	9	0.116667

Abbildung 10 Tabelle zur Häufigkeit der verschiedenen Mittelfeld Rollen von `analyse_transfermarkt.ipynb`

So kann man erkennen, dass aus allen Spielern, welche im Mittelfeld spielen, das Zentrale Mittelfeld mit 39,04 % am meisten gespielt wird. Zieht man nun die Tabelle aus Abbildung 10 dazu erkennt man das auf dieser Position auch der höchste Durchschnittliche Marktwert pro Spieler vorliegt. Ähnlich sieht das auch bei den zweit und dritt meistgespielten Mittelfeld Positionen aus. Hier sind nämlich auch die durchschnittlichen Marktwerte auf Platz 2 und 3.

Relativ/Absolut Scatter dot Chart

In der Fußball Welt sind die Rückennummer auf dem Trikot des Spielers nicht nur eine Nummer, sondern sind eine Art Statussymbol, welche Rolle man in einem Verein spielt.

Dabei stehen die Nummern für die Stammposition, die ein Spieler spielt. So steht die Rückennummern 1 z. B. für den Stammtorhüter einer Mannschaft oder die Nummer 9 für den Stamm Stürmer einer Mannschaft.

Mithilfe denn folgenden Grafiken soll demnach geprüft werden, welche Marktwertsummen die Spieler mit der jeweiligen Rückennummer haben und wie oft die Rückennummer

vergeben ist.

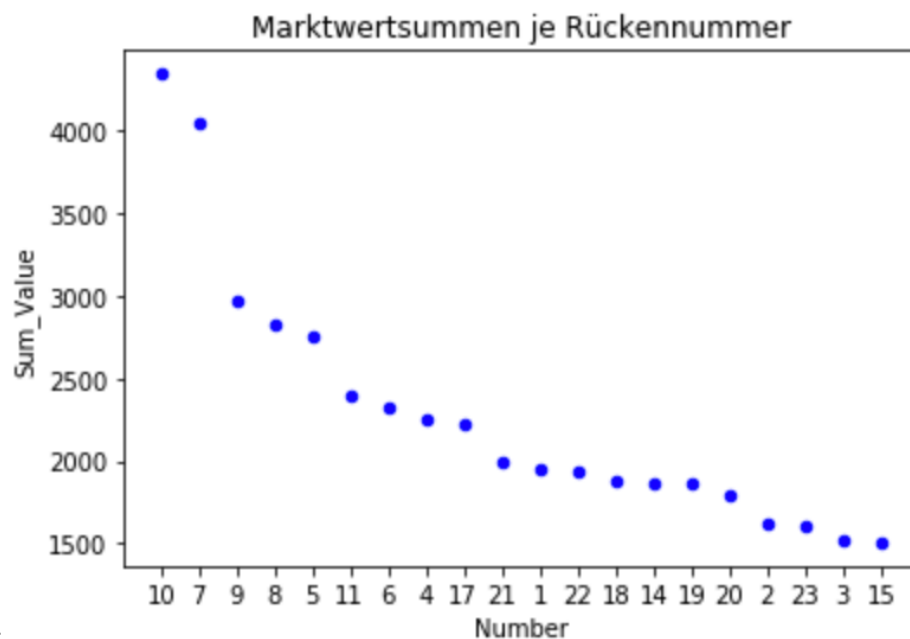


Abbildung 11 Marktwertsummen je Rückennummer von analyse_transfermarkt.ipynb

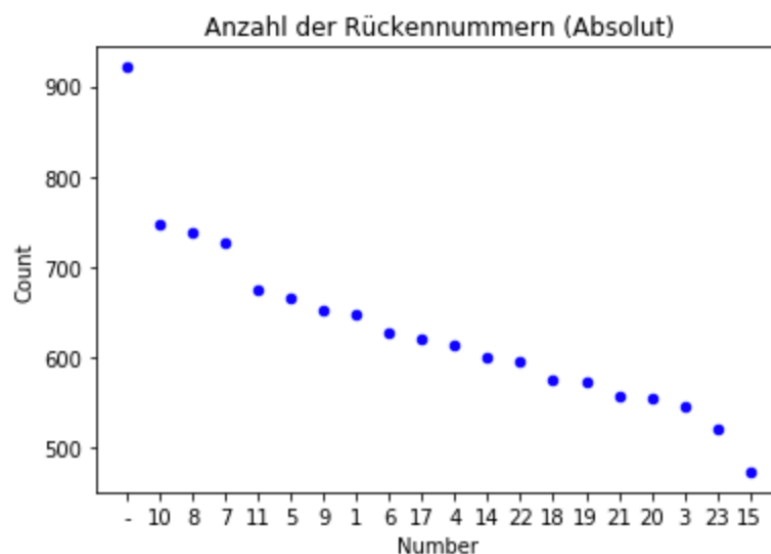


Abbildung 12 Anzahl der Rückennummern (Absolut) von analyse_transfermarkt.ipynb

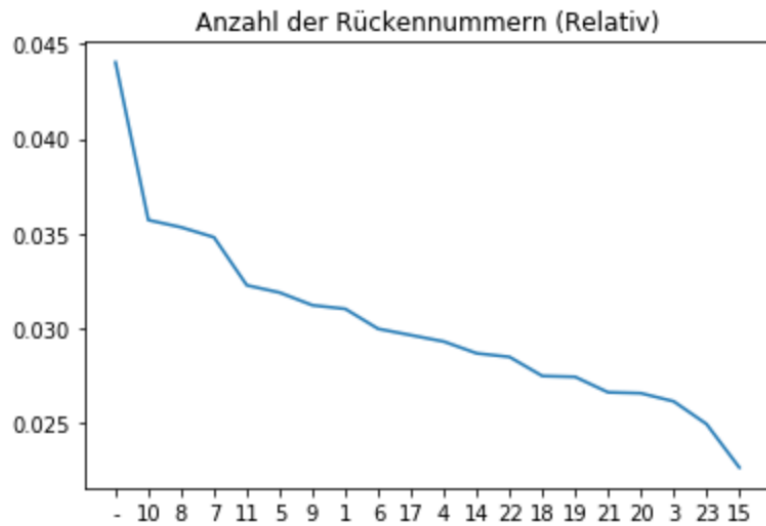


Abbildung 13 Anzahl der Rückennummern (Relativ) von `analyse_transfermarkt.ipynb`