









Marktwert und Gesamtleistung von Spielern in der UCL 2021/22

Can Bakirci, Ben Gebhardt, Tolga Ipek,
Robin Klinkhammer, Lennard Michel,
Mogens Rosiny

DIS08 Datenmodellierung/Data Modelling
[WS 2024/2025]



Inhaltsverzeichnis

1. Einleitung 
2. Datenquellen 
3. Untersuchungsfragen und Hypothesen 
4. Vorgehensweise 
5. Analyse und Visualisierung der Ergebnisse 
6. Fazit: Planung, Herausforderungen und Reflexion der Gruppenarbeit
(Projektmanagement)   

Einleitung

Motivation

- Der Marktwert eines Spielers beeinflusst Transfers und Strategien
- Ist der Marktwert ein verlässlicher Indikator für die Leistung?

Ziel

- Analyse von Marktwert und Leistung (offensiv/defensiv) in der UCL 2021/22

Fragestellung

- Besteht ein Zusammenhang zwischen Marktwert und Spielerleistung?

Datenquellen

Transfermarkt  – Marktwerte der Spieler für die Saison 2021/22 der Teams Liverpool, Manchester City, Real Madrid, Villarreal

KADER REAL MADRID - SAISON 21/22

Diese Seite enthält die detaillierte Kader-Übersicht eines Vereins. Alle Spieler der jeweiligen Mannschaften werden mit ihrem Alter, der Natio...

Saison auswählen





































21/22

Anzeigen

Kompakt

Erweitert

Galerie

# ↑	Spieler ↓	Alter ↑	Nat.	Aktueller Verein	Marktwert ↓
20	<div><div></div><div>Vinicius Junior Linksaußen</div></div>	21	<div><div></div></div>	<div><div></div></div>	100,00 Mio. €
15	<div><div></div><div>Federico Valverde Zentrales Mittelfeld</div></div>	23	<div><div></div></div>	<div><div></div></div>	70,00 Mio. €
1	<div><div></div><div>Thibaut Courtois Torwart</div></div>	30	<div><div></div></div>	<div><div></div></div>	60,00 Mio. €
3	<div><div></div><div>Éder Militão Innenverteidiger</div></div>	24	<div><div></div></div>	<div><div></div></div>	60,00 Mio. €
21	<div><div></div><div>Rodrygo Rechtsaußen</div></div>	21	<div><div></div></div>	<div><div></div></div>	60,00 Mio. €
4	<div><div></div><div>David Alaba Innenverteidiger</div></div>	30	<div><div></div></div>	<div><div></div></div>	55,00 Mio. €
25	<div><div></div><div>Eduardo Camavinga Zentrales Mittelfeld</div></div>	19	<div><div></div></div>	<div><div></div></div>	55,00 Mio. €
14	<div><div></div><div>Casemiro Defensives Mittelfeld</div></div>	30	<div><div></div></div>	<div><div></div></div>	40,00 Mio. €
23	<div><div></div><div>Ferland Mendy Linker Verteidiger</div></div>	27	<div><div></div></div>	<div><div></div></div>	40,00 Mio. €
11	<div><div></div><div>Marco Asensio Rechtsaußen</div></div>	26	<div><div></div></div>	<div><div></div></div>	40,00 Mio. €
9	<div><div></div><div>Karim Benzema Mittelstürmer</div></div>	34	<div><div></div></div>	<div><div></div></div>	30,00 Mio. €
8	<div><div></div><div>Toni Kroos Zentrales Mittelfeld</div></div>	32	<div><div></div></div>	<div><div></div></div>	20,00 Mio. €

Kaggle  – Spielerstatistiken der UEFA-Champions League 2021/22 der Teams Liverpool, Manchester City, Real Madrid, Villarreal

key_stats.csv (32.57 kB)					
Detail Compact Column					
△ player_name ▾	△ club ▾	△ position ▾	# minutes_pl... ▾	# match_pla... ▾	# goals ▾
Courtois	Real Madrid	Goalkeeper	1230	13	0
Vinicius Júnior	Real Madrid	Forward	1199	13	4
Benzema	Real Madrid	Forward	1106	12	15
Modrić	Real Madrid	Midfielder	1077	13	0
Éder Militão	Real Madrid	Defender	1076	12	0
Alaba	Real Madrid	Defender	1040	12	1
Carvajal	Real Madrid	Defender	959	11	0
Casemiro	Real Madrid	Midfielder	914	11	0
Kroos	Real Madrid	Midfielder	902	12	2
Mendy	Real Madrid	Defender	867	10	0
Valverde	Real Madrid	Midfielder	804	11	0
Nacho	Real Madrid	Defender	574	9	0
Rodrygo	Real Madrid	Forward	505	11	5

Untersuchungsfragen und Hypothesen ?

1. Hauptfrage: *Gibt es eine Korrelation zwischen Marktwert und Spielerleistung in der UCL 2021/22?*

2. Hypothesen:

- **Hypothese 1:** Spieler mit einem höheren Marktwert erzielen mehr Tore und Assists und weisen eine größere Anzahl an gespielten Minuten auf.
- **Hypothese 2:** Spieler mit höherem Marktwert sind stärker in defensiven Metriken wie Tackles und Ballgewinnen
- **Hypothese 3:** Spieler mit höherem Marktwert zeigen insgesamt eine bessere Passgenauigkeit

Vorgehensweise

1. Kaggle-Daten laden und Markwerte scrapen
2. Daten bereinigen
3. Daten verknüpfen
4. Daten analysieren und visualisieren

1. Kaggle-Daten laden und Marktwerte scrapen

- Herunterladen des Kaggle-Datensatzes (UCL | Matches & Players Data)
- Scraping der Marktwerte und Spielerdaten von der Transfermarkt-Website
 - Erstellen einer Funktion zum Scraping
 - Verwendung einer for-Schleife, um die Daten der 4 Teams zu scrapen
 - Speicherung der gescrapten Daten in einer CSV-Datei
 - Verwendete Python Bibliotheken: os, zipfile, requests, BeautifulSoup, pandas

Herausforderungen:

- Wo im HTML-Code befinden sich die relevanten Daten?
- Wie lässt sich die Funktion für die 4 verschiedenen Teams (Liverpool, Manchester City, Real Madrid, Villareal) ausführen?

2. Daten bereinigen

- Entfernen unvollständiger Daten (z. B. Spieler ohne Marktwert)
- Konvertieren der Marktwerte in ein einheitliches Format (Tausend/Millionen)
- Normalisieren von Spaltennamen und Texten
- Extrahieren und Hinzufügen zusätzlicher Spalten (z. B. Vornamen der Spieler)
- Überprüfen der Kaggle-Daten auf grobe Fehler
- Verwendete Python Bibliotheken: pandas, os

Herausforderungen:

- Im Kaggle-Datensatz waren die Angaben zu roten und gelben Karten vertauscht
- Die Spieler wurden in den Kaggle- und Transfermarkt-Datensätzen unterschiedlich benannt, z. B. wurden teilweise nur Vornamen verwendet oder der Vorname wurde nur abgekürzt
- Spieler mit identischen Nachnamen, wie Alberto Moreno und Gerard Moreno, konnten im Transfermarkt-Datensatz nicht eindeutig unterschieden werden

3. Daten verknüpfen

- Kombination der gescrapten Daten mit zusätzlichen Datensätzen (z.B. Key Stats, Defending)
- Verwendete Python Bibliotheken: pandas, re
- Abgleich von Spielernamen und Teams zwischen verschiedenen CSV-Dateien
 - Verwendung textbasierter Algorithmen für die Zuordnung

Herausforderungen:

- Inkonsistente Schreibweise der Spielernamen im Kaggle-Datensatz, was die Entwicklung komplexer Funktionen zur Datenverknüpfung erforderlich machte

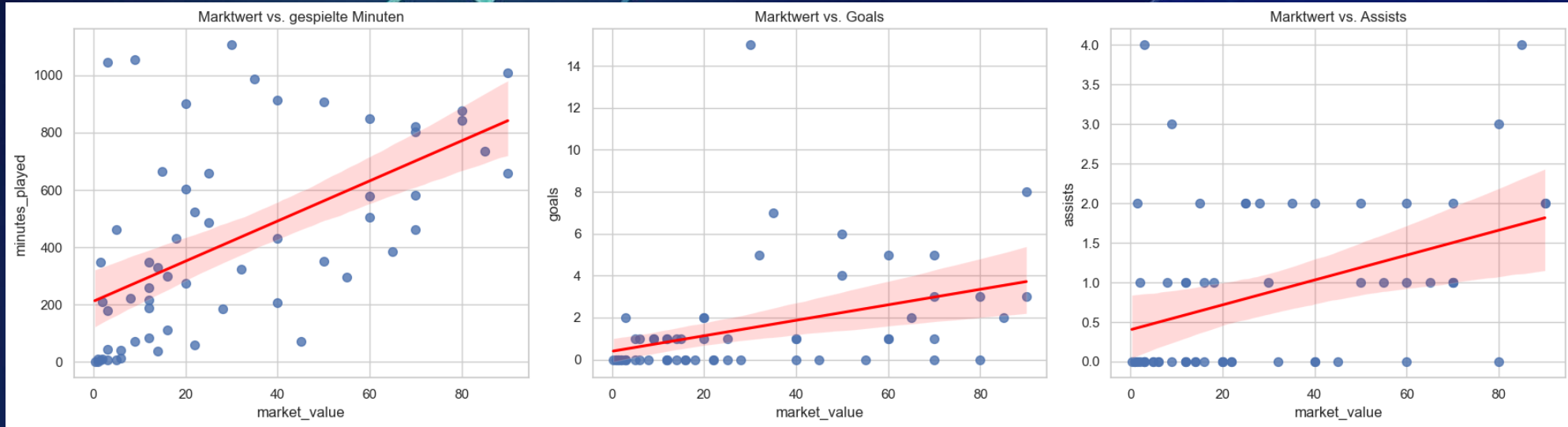
4. Daten analysieren und visualisieren

- **Verschiedene Analysen werden durchgeführt:**
 - Marktwert Zusammenhang mit gespielten Minuten
 - Marktwert Zusammenhang mit Toren und Assists (Korrelation und Regression)
 - Marktwert Zusammenhang mit Defensivleistungen der Verteidiger (Balleroberungen, Tacklings, gewonnene Zweikämpfe)
 - Passgenauigkeit in Relation zum Marktwert
- Verwendete Python Bibliotheken: pandas, matplotlib, seaborn, scipy, statsmodels
- **Visualisierung:**
 - Ergebnisse werden mit Scatterplots und Regressionslinien dargestellt
 - Lineare Regressionsmodelle werden erstellt und interpretiert

Herausforderungen:

- Auswahl der passenden Analysemethode
- Interpretation der Ergebnisse und Werte

Marktwert vs. gespielte Minuten/Tore/Assists



Marktwert vs. gespielte Minuten:

- Positiver Trend: Spieler mit höherem Marktwert tendieren dazu, mehr Minuten zu spielen.
- Breite Streuung bei niedrigen Marktwerten zeigt starke Unterschiede in Einsatzzeit.

Marktwert vs. Goals:

- Leichter Anstieg: Spieler mit höherem Marktwert erzielen im Durchschnitt mehr Tore.
- Viele Spieler mit niedrigem Marktwert erzielen keine oder wenige Tore (Cluster bei 0).

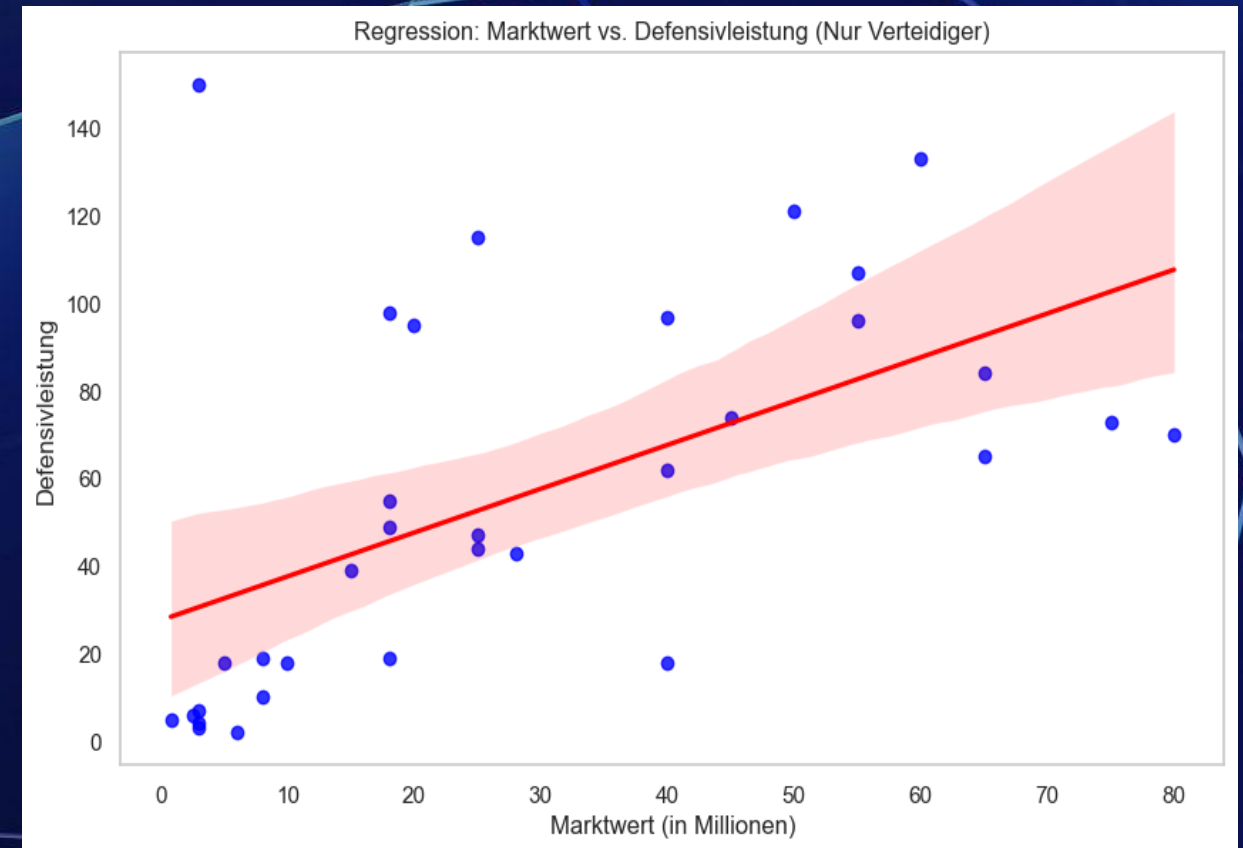
Marktwert vs. Assists:

- Geringfügig positiver Zusammenhang: Spieler mit höherem Marktwert neigen dazu, mehr Assists zu leisten.
- Hohe Streuung deutet auf die Einflussnahme anderer Faktoren hin.

- Fazit: Insgesamt zeigen alle Graphen positive Trends, die jedoch keine starken Zusammenhänge darstellen. Spieler mit höherem Marktwert sind tendenziell stärker an Spielzeit, Toren und Assists beteiligt, aber es gibt zahlreiche Ausnahmen und eine deutliche Streuung.

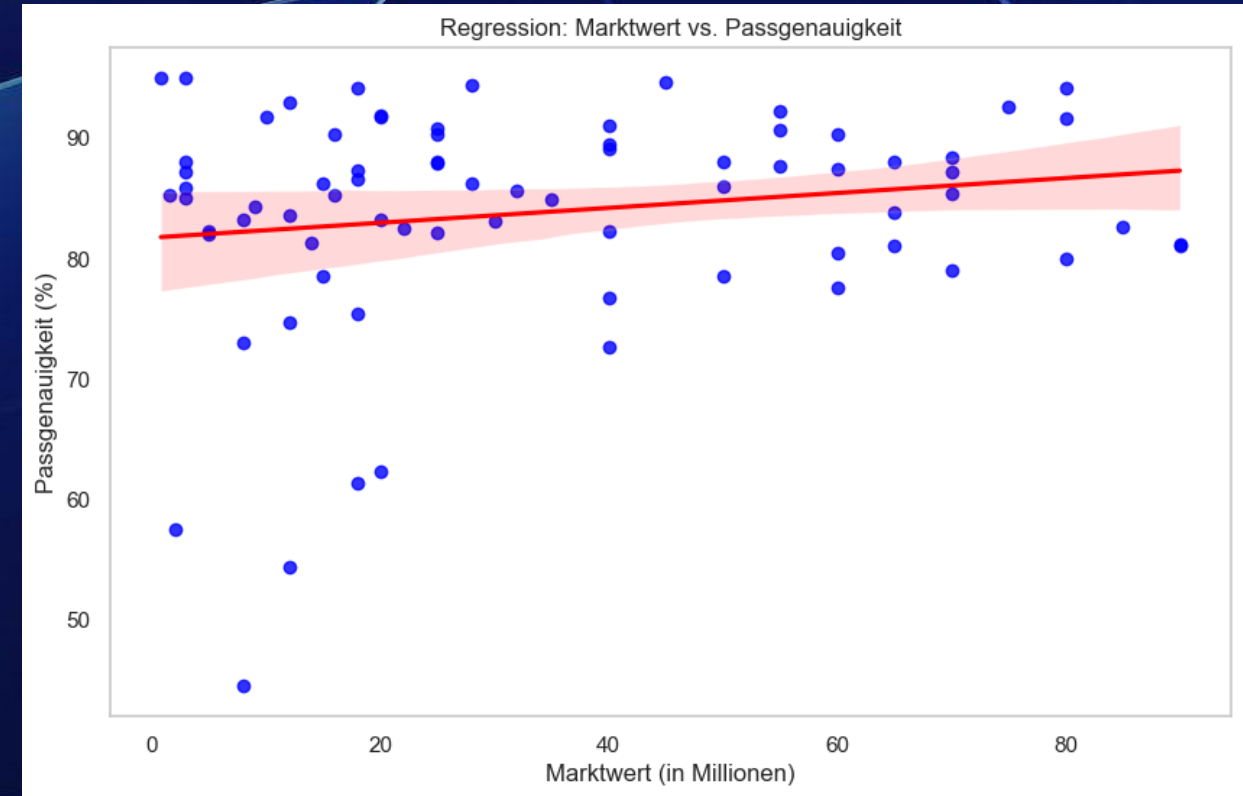
Marktwert vs. Defensivleistung

- **Streuung:** Datenpunkte zeigen eine breite Verteilung, insbesondere bei niedrigen Marktwerten, was auf starke individuelle Unterschiede hinweist.
 - **Regressionslinie:** Positiver Zusammenhang – höhere Marktwerte gehen tendenziell mit besserer Defensivleistung einher.
 - **Konfidenzintervall:** Die rote Schattierung zeigt Unsicherheiten der Schätzung, die bei höheren Marktwerten breiter wird.
 - **Ausreißer:** Einige Verteidiger mit sehr hoher Defensivleistung, aber relativ niedrigem Marktwert, beeinflussen die Streuung.
- ➔ **Fazit:** Verteidiger mit höherem Marktwert erzielen tendenziell bessere Defensivleistungen, jedoch sind Ausreißer und Unsicherheiten erkennbar.



Marktwert vs. Passgenauigkeit

- **Streuung:** Breite Verteilung der Punkte zeigt, dass Marktwert kaum die Passgenauigkeit erklärt.
 - **Regressionslinie:** Minimaler positiver Trend, jedoch zu schwach, um relevant zu sein.
 - **Konfidenzintervall:** Breite Schattierung verdeutlicht Unsicherheit, insbesondere bei hohen Marktwerten.
 - **Ausreißer:** Spieler mit hoher Passgenauigkeit und niedrigem Marktwert (und umgekehrt) betonen die geringe Aussagekraft.
- ➔ **Fazit:** Kein signifikanter Zusammenhang – Marktwert ist kein verlässlicher Prädiktor für Passgenauigkeit.



Planung der Gruppenarbeit

1. Zielsetzung:

- Analyse der Marktwerte und Gesamtleistungen von Spielern in der UCL 2021/22

2. Rollenverteilung:

- Datenbeschaffung und Scraping: Can, Tolga, Ben
- Datenbereinigung und Analyse: Ben, Mogens, Lennard, Robin
- Visualisierung und Präsentation: Lennard, Robin, Mogens, Tolga, Can

3. Zeitplan:

- Woche 1: Datensammlung
- Woche 2: Datenbereinigung und Analyse
- Woche 3: Erstellung der Visualisierungen
- Woche 4: Vorbereitung der Präsentation

Herausforderungen und Lösungen ✂

1. Herausforderungen:

- Uneinheitliche Datenquellen
- Technische Schwierigkeiten beim Scraping
- Zeitmanagement innerhalb der Gruppe

2. Lösungen:

- Regelmäßige Abstimmungen und klare Kommunikationswege
- Einsatz von Standardtools (Python, Pandas, PowerPoint, Github)
- Unterstützung innerhalb der Gruppe bei technischen Problemen

Reflexion der Gruppenarbeit 🧠

1. Was hat gut funktioniert?

- Klare Rollenverteilung und Zielsetzung
- Effektive Kommunikation und Zusammenarbeit
- Pünktliche Fertigstellung aller Meilensteine

2. Was hätte besser laufen können?

- Mehr Zeit für die Erstellung der Präsentation einplanen
- Früheres Testen der Visualisierungen

3. Fazit:

- Insgesamt erfolgreiche Zusammenarbeit mit wertvollen Erkenntnissen



Vielen Dank für Ihre Aufmerksamkeit!
Gibt es noch Fragen?