

# Visualisierung des Fussballtransfermarktes

Data Visualization

vorgelegt am 30. Juni 2025

Fakultät Wirtschaft und Gesundheit

Studiengang Wirtschaftsinformatik – Data Science

Kurs WWI2024F

von

Caleb Boakye-Yiadom

Matrikelnummer: 4351473

DHBW Stuttgart:

Florian Eicher

**Inhaltsverzeichnis**

<b>Vorwort.....</b>	<b>III</b>
<b>1 Einleitung.....</b>	<b>1</b>
<b>2 Formulating your brief.....</b>	<b>2</b>
2.1 Ausgangspunkt und Zielsetzung.....	2
2.2 Relevanz und Motivation.....	2
2.3 Leitfragen und Zielgruppe .....	2
2.4 Interaktivität und Funktionalität .....	2
2.5 Datengrundlage .....	2
2.6 Visualisierungskonzept .....	3
<b>3 Working with Data.....</b>	<b>4</b>
3.1 Datenquellen und -struktur.....	4
3.2 Datenbereinigung.....	4
3.3 Transformation und Aggregation .....	5
<b>4 Establish your Editorial Thinking .....</b>	<b>6</b>
4.1 Perspektivenvielfalt und Erkenntnisinteresse .....	6
4.2 Erkenntnisziel und Interaktivität .....	6
4.3 Struktur- und Designkonzept .....	6
<b>5 Develop the Design Solution .....</b>	<b>8</b>
5.1 Choroplethen-Karte für Median der Marktwerte pro Nation bzw. Transferausgaben pro Land.....	8
5.2 Sankey-Diagramm für Transferströme zwischen den Fußballnationen.....	9
5.3 Streudiagramm für Marktwertentwicklung: Junge vs. ältere Spieler .....	9
5.4 Säulendiagramm für durchschnittliche Ablösesumme je Position .....	10
5.5 Liniendiagramm für Transferausgaben pro Verein und Saison .....	11
<b>Anhang.....</b>	<b>IV</b>
<b>Literaturverzeichnis .....</b>	<b>VIII</b>

## Vorwort

Zum Starten des Dashboards müssen die Datensätze im Ordner „**data**“ belassen werden. Anschließend genügt es, die Datei „**app.py**“ auszuführen. In der Kommandozeile erscheint dann ein Link, über den das Dashboard im Browser geöffnet werden kann.

Wichtig ist außerdem, dass der Ordner „**pages**“ vollständig vorhanden ist, da er die einzelnen Unterseiten des Dashboards enthält.

Sollte das lokale Starten wider Erwarten nicht funktionieren, kann das gesamte Projekt inklusive eines Screen-Recordings unter folgendem Link auf GitHub eingesehen werden:

👉 <https://github.com/calebboa/Football-Dashboard.git>

Das Screen-Recording befindet sich dort unter „**Screen-Rec-Dash**“.

## 1 Einleitung

In dieser Dokumentation wird die Entwicklung eines interaktiven Dashboards zur Visualisierung von Fußball-Transferdaten erläutert. Der Prozess folgt den vier Phasen des Data-Visualization-Workflows: Formulating your brief, Working with Data, Establish your editorial thinking und Develop the design solution. Ziel ist es, fünf aussagekräftige Visualisierungen zu erstellen, die unterschiedliche Aspekte des Transfergeschehens zeigen (z. B. Ablösesummen nach Position, Marktwerte nach Nation, Transferströme, Altersgruppenvergleiche, Vereineinnahmen). Dabei werden die Inhalte und Strukturierung so gewählt, dass sie ein zusammenhängendes Gesamtbild vermitteln und typische Gestaltungsprinzipien wie Farbhierarchie, Interaktivität und Lesbarkeit berücksichtigen.

## **2 Formulating your brief**

### **2.1 Ausgangspunkt und Zielsetzung**

Das vorliegende Visualisierungsprojekt wurde im Rahmen des Moduls Data Visualization als individuelle Prüfungsleistung konzipiert. Ziel ist es, aus einem umfangreichen Fußball-Transferdatensatz praxisrelevante Erkenntnisse zu gewinnen, diese adressatengerecht aufzubereiten und durch geeignete Visualisierungstechniken zu vermitteln.

### **2.2 Relevanz und Motivation**

Durch die Anwendung erlernter Konzepte auf ein realitätsnahes Szenario wird das theoretische Wissen im Bereich der explorativen Datenanalyse und Visualisierung vertieft. Zugleich zeigt das Projekt exemplarisch, wie aus quantitativen Informationen visuelle Entscheidungsgrundlagen und Handlungsperspektiven abgeleitet werden können.

### **2.3 Leitfragen und Zielgruppe**

Zentrale Fragestellungen lauten:

- Wie verteilen sich Transferausgaben und Marktwerte nach verschiedenen Dimensionen?
- Welche Länder und Positionen dominieren im internationalen Transfermarkt?
- Wie haben sich Marktwerte über die Jahre hinweg für verschiedene Altersgruppen entwickelt?

Die Zielgruppe umfasst fußballinteressierte Nutzer und datenaffine Analysten. Das Dashboard soll explorativ nutzbar sein und individuelle Analysepfade ermöglichen.

### **2.4 Interaktivität und Funktionalität**

Das Dashboard ist keine statische Darstellung, sondern eine interaktive Anwendung. Nutzer können eigene Filter setzen (z. B. Saisons, Positionen, Länder, Altersgrenzen), um Hypothesen zu überprüfen. Die Bedienbarkeit und Nutzerfreundlichkeit stehen dabei im Vordergrund.

### **2.5 Datengrundlage**

Verwendet wurden zwei kombinierte Datensätze zu Spielertransfers. Diese enthalten Informationen zu Transfergebühren, Marktwerten, Positionen, Alter und Nationalitäten. Ein wichtiges Ziel bestand darin, diese Quellen zu verknüpfen und Kennzahlen wie Summen oder Durchschnitte zu berechnen.

## **2.6 Visualisierungskonzept**

Das Dashboard enthält fünf (bis 6) verschiedene Visualisierungen, davon mindestens eine aufwändigere. Alle Darstellungen sind interaktiv und bieten Filterfunktionen (z. B. Slider, Dropdowns). Die Präsentation ist kompakt, achtet aber auf eine ausreichende Skalierung der einzelnen Grafiken.

Technologisch basiert das Projekt auf Python unter Verwendung der Bibliotheken Pandas, Plotly und Dash, um maximale Flexibilität und visuell ansprechende Darstellungen zu gewährleisten.

### 3 Working with Data

#### 3.1 Datenquellen und -struktur

Für das Dashboard wurden zwei komplementäre CSV-Datensätze verwendet:

„*merged\_transfers.csv*“: Enthält umfangreiche Transferinformationen wie Saison, Position, Transfergebühren, Vereinswechsel und Altersdaten.

„*cleaned\_transfersGit.csv*“: Enthält bereinigte Marktwerte sowie Nationalitätsinformationen.

Typische Spalten der Datensätze sind:

- `transfer_fee` (Ablösesumme in Euro)
- `market_value` (Marktwert in Euro)
- `position`, `season`, `age`
- `team_herkunft_country`, `team_ziel_country`, `nationality`

Durch das Merging beider Quellen anhand gemeinsamer Spielerinformationen wurde eine erweiterte Analyse ermöglicht, z. B. zum Zusammenhang von Marktwert, Alter und Nationalität.

#### 3.2 Datenbereinigung

Die Bereinigung der Datensätze erfolgte in der Datei `datasets.py`. Dort wurden irrelevante Spalten entfernt, fehlerhafte oder doppelt vorkommende Einträge gelöscht sowie zentrale Felder wie Spielername, Position und Nationalität standardisiert. Spielerpositionen wurden über ein Mapping vereinheitlicht (z. B. „Centre-Forward“ zu „CF“), Nationalitäten mithilfe der `pycountry`-Bibliothek auf gültige Ländernamen normalisiert. Außerdem wurden unrealistische Ablösesummen über 200 Mio. € ausgeschlossen und bei Mehrfachnennungen pro Saison jeweils nur der Datensatz mit dem höchsten Marktwert beibehalten. Schließlich wurden beide Hauptdatensätze (`df1` und `df2`) anhand von Spielername und Saison zusammengeführt. Die bereinigten Versionen wurden als `cleaned_transfersGit.csv`, `cleaned_players_transfers.csv` und `merged_transfers.csv` gespeichert.

Fehlende oder ungültige Werte – etwa bei transfer\_fee – wurden entfernt. Um Ausreißer zu vermeiden, wurden Transfers unter 1 Mio. € aus bestimmten Auswertungen ausgeschlossen. Spieler wurden in Altersgruppen (z. B. „jung“ vs. „alt“) eingeteilt, basierend auf einem flexiblen Schwellenwert. Zudem konzentrierte sich die Analyse internationaler Bewegungen auf Transfers zwischen den Top-5-Fußballnationen.

### 3.3 Transformation und Aggregation

Die Rohdaten wurden zu folgenden Analyseformen weiterverarbeitet:

- **Durchschnittliche Ablösesumme pro Position:** Mittelwert der Transfergebühren, umgerechnet in Mio. €, sortiert nach Höhe
- **Median des Marktwertes pro Nationalität:** Median der Spieler-Marktwerte, gefiltert nach Mindestwert
- **Transferausgaben pro Land:** Aggregierte Summen, beschränkt auf die Top-N-Ausgabennationen
- **Transfers zwischen Ländern:** Zählung von Wechseln zwischen den Top-5-Fußballnationen zur Darstellung im Sankey-Diagramm
- **Marktwertentwicklung nach Altersgruppe und Saison:** Durchschnittlicher Marktwert je Saison für „jung“ und „alt“
- **Transferausgaben pro Verein und Saison:** Summierte Ausgaben, Darstellung der ausgabestärksten Clubs

Die Kombination und Transformation der beiden Datensätze ermöglichte sowohl finanzielle als auch demografische Einblicke, wie die Altersverteilung und Ausgaben, in den internationalen Fußballtransfermarkt.



## **4 Establish your Editorial Thinking**

### **4.1 Perspektivenvielfalt und Erkenntnisinteresse**

Das Dashboard kombiniert verschiedene Betrachtungsebenen des internationalen Transfermarkts. Spielerpositionen (z. B. typische Ablösesummen je Rolle), geografische Verteilungen (z. B. Transferausgaben oder Marktwerte nach Nation), länderübergreifende Transferströme, zeitliche Entwicklungen (z. B. Investitionen pro Saison) und die finanzielle Sicht einzelner Vereine werden dabei miteinbezogen.

Diese Kombination erlaubt eine ganzheitliche Analyse zentraler Marktmechanismen im Profifußball.

Beispielsweise zeigt das Säulendiagramm teure Positionen, die Choroplethenkarte geografische Transferzentren, das Sankey-Diagramm Ströme zwischen Top-Nationen, während Marktwert- und Vereinsanalysen strukturelle und zeitliche Dynamiken sichtbar machen.

### **4.2 Erkenntnisziel und Interaktivität**

Jede Visualisierung verfolgt ein spezifisches Analyseziel und beleuchtet dabei einen gezielten Aspekt des Transfergeschehens.

Mittels interaktiver Elemente, wie Dropdown-Menüs, Schieberegler oder Saisonfiltern, können Nutzende eigene Fragestellungen explorieren, etwa:

- Wie haben sich Ablösesummen für Innenverteidiger über die Jahre verändert?
- Welches Land hatte im Median die höchsten Marktwerte?

Ziel ist eine explorative Nutzererfahrung, bei der individuelle Perspektiven möglich sind, ohne den Blick für übergreifende Muster zu verlieren.

### **4.3 Struktur- und Designkonzept**

Die Anwendung ist als mehrseitiges Dashboard konzipiert. Den Einstieg bildet eine Startseite, die spielerisch durch einen zufälligen Fußballfakt auf das Thema einstimmt. Im Anschluss sind die Visualisierungen in einer thematisch sinnvollen Abfolge angeordnet: Zunächst erfolgt die Weltkarte mit geographischen Verteilungen, gefolgt von dem Sankey-Diagramm zu internationalen Transferströmen, der Marktwertentwicklung nach Altersgruppen, der Analyse nach Spielerpositionen und abschließend der Analyse der Transferausgaben einzelner Vereine. Die Navigation zwischen den Seiten erfolgt intuitiv über ein zentrales Menü.

Alle Seiten folgen einem konsistenten visuellen Konzept: Ein dunkler Hintergrund sorgt in Kombination mit kontrastreichen Farben für eine gute Lesbarkeit, während eine einheitliche Typografie die Wiedererkennbarkeit und Klarheit im Design unterstützt. Dieses Gestaltungskonzept verbessert die Orientierung und Benutzerfreundlichkeit über alle Ansichten hinweg.

## 5 Develop the Design Solution

Für alle Visualisierungen wurde ein konsistentes, modernes Design gewählt. Der schwarze Hintergrund in Kombination mit weißer Schrift erzeugt einen kontrastreichen „Dark Mode“-Look. Akzentfarben aus Plotly's Farbpaletten heben zentrale Datenpunkte hervor, während einheitliche Typografie und Schriftgrößen für Lesbarkeit und Wiedererkennung sorgen.

Das Dashboard setzt bewusst auf interaktive Elemente wie Slider, Dropdown-Menüs und Filter. Diese erlauben es den Nutzenden, Inhalte individuell zu explorieren und unterschiedliche Perspektiven einzunehmen. Alle Diagrammtypen wurden gezielt ausgewählt, um den jeweiligen Datentyp optimal zu visualisieren – stets mit Fokus auf Klarheit und Funktionalität.

Die zugrundeliegenden Datensätze stammen aus einem GitHub-Repository sowie einem Mendeley-Datensatz. Beide Quellen enthalten umfassende Informationen zu Transfers, Marktwerten, Spielern, Vereinen und Saisons. Die verwendeten Daten sind im Code dokumentiert und wurden zur Analyse vorbereitet und zusammengeführt.

### 5.1 Choroplethen-Karte für Median der Marktwerte pro Nation bzw. Transferausgaben pro Land

Eine Weltkarte (Choropleth) eignet sich, um geographische Unterschiede zu verdeutlichen. Je nach Einstellung (Radio-Button) zeigt die Karte entweder den median der Marktwerte je Nationalität oder die gesamten Transferausgaben je Zielland.

Jeder Eintrag (Land) bekommt eine Farbe entsprechend seinem Wert in Mio. €. Die Skala zeigt graduelle Unterschiede. Auf der Karte bleibt der Hintergrund durch `geo.bgcolor` transparent, was bei schwarzem Gesamtdesign einen „Nachtmodus“-Effekt erzeugt. Ein Farbbalken (Colorbar) mit Titel („Marktwert“ oder „Transferausgaben (€M)“) ermöglicht dem Nutzer, Wertebereiche abzuschätzen.

Der Slider „*Top-N Länder*“ limitiert die Anzeige auf die  $N$  Länder mit höchsten Werten, damit der Fokus auf den wichtigsten Akteuren liegt. Ein Eingabefeld („Mindestwert“) erlaubt es, Länder unter einem bestimmten Schwellenwert auszublenden. Diese interaktiven Regler tragen dazu bei, die Karte nicht zu überladen und die wichtigsten Daten hervorzuheben.

Eine Kartenprojektion stellt die Länder annähernd korrekt dar. Die Farbauswahl reflektiert intuitiv die Größenordnung. Beschriftungen direkt auf der Karte wurden vermieden, um Platz zu sparen – stattdessen liefert der Tooltip (Hover) exakte Werte.

## 5.2 Sankey-Diagramm für Transferströme zwischen den Fußballnationen

Um die Flüsse von Spielern zwischen Ländern darzustellen, wurde ein Sankey-Diagramm gewählt. Dieses Diagrammtyp ist aufwändiger, eignet sich aber gut für die Visualisierung von Strömen und Mengen. Hier repräsentieren die Knoten die Top 5 Fußballnationen. Die Verbindungen zeigen, wie viele Spieler von einem Herkunftsland (team\_herkunft\_country) in ein anderes Zielland (team\_ziel\_country) gewechselt sind.

Die Anzahl der Transfers von Land A nach Land B wurde aggregiert (Count). Diese Zahlen bestimmen die Dicke der Links im Sankey-Chart. Die Knoten sind eindeutig beschriftet, und die Farben der Links basieren auf die RGB-Farbe der jeweiligen Flagge, z. B. sind alle Links aus Spanien rot. Dies erleichtert es, den Ursprung einer Spielerbewegung optisch zu erkennen.

Die Knotenfarben sind einheitlich hellgrau mit dunklem Rand, damit sie im Schwarzweiß-Hintergrund gut sichtbar sind. Die Linkfarben sind kräftig (rot, gelb, blau, etc.), um die Nationalität hervorzuheben. Der Titel „Häufigkeit der Transfers zwischen Ländern“ gibt klar an, was gemessen wird und der Ländername steht direkt am Knoten, weshalb auf Farbe verzichtet werden kann.

Zwei Dropdown-Filter („Herkunftsländer“ und „Zielländer“) erlauben es, die angezeigten Flüsse einzugrenzen. Zum Beispiel kann man sich alle Transfers aus Frankreich anzeigen lassen oder nur Transfers, die nach Deutschland gehen. Diese Filter führen zu sofortiger Aktualisierung, wodurch der Nutzer die Sankey-Ströme in verschiedenen Szenarien erkunden kann.

## 5.3 Streudiagramm für Marktwertentwicklung: Junge vs. ältere Spieler

Ein Streudiagramm (Scatter-Plot) visualisiert den durchschnittlichen Marktwert nach Saison, getrennt nach Altersgruppe (jung bzw. alt). Dies verdeutlicht Trends im Zeitverlauf. Jede Altersgruppe wird durch eine eigene Farblinie mit Markierungen repräsentiert (rot für jung, grün für alt).

Die X-Achse ist die Saison, die Y-Achse ist der durchschnittliche Marktwert in Mio. €. Ein einheitliches Symbol (Marker) wird eingesetzt, um einzelne Datenpunkte hervorzuheben. Die Markierungen sind für bessere Sichtbarkeit in einem entfernten Layout etwas groß (Size=10).

Zwei deutliche Farben (Rot/Grün) kontrastieren jung vs. alt. Diese Zuordnung wird im Code festgelegt. Die Legende nennt die Altersgruppen. Durch das Zusammenführen beider Gruppen in einer Grafik kann der Betrachter sofort erkennen, ob „junge“ Spieler im Durchschnitt als teurer oder billiger bewertet wird als „alte“ in einem bestimmten Jahr.

Ein Range-Slider für die Saisons lässt den Nutzer einen bestimmten Zeitraum auswählen (z. B. nur die letzten 5 Jahre). Ein Schieberegler („Altersgrenze“) legt fest, ab welchem Alter ein Spieler in die „alten“ Kategorie fällt. Durch Änderung dieser Grenze kann man interaktiv ausloten, wie sensibel die Ergebnisse auf die Definition von „jung“ und „alt“ reagieren.

Das Diagramm hat einen schwarzen Hintergrund und einen weißen Titel. Der Titel „Ø Marktwert nach Saison (Mio €)“ fasst das Thema zusammen. Der Scatter wurde statt Linien gewählt, um einzelne Saisonpunkte zu betonen und Überlagerungen zu vermeiden.

#### 5.4 Säulendiagramm für durchschnittliche Ablösesumme je Position

Ein vertikales Säulendiagramm eignet sich, um kategoriale Daten, wie Positionen, zu vergleichen. Die Höhe der Säulen zeigt den Durchschnitt der Transfergebühren in Mio. € je Position. Dieses Format erlaubt einen schnellen visuellen Vergleich zwischen dem Zusammenhang zwischen Transferausgabe und Position. Die Säulen sind nach absteigender Höhe sortiert, um den Fokus sofort auf die höchsten Ausgaben zu lenken. Ein Feststext auf jeder Säule (mit zwei Dezimalstellen) zeigt den exakten Wert, was genaue Interpretation unterstützt. Die Y-Achse ist als „Durchschn. Ablösesumme in Mio €“ beschriftet.

Die Balkenfarben stammen aus einer fortlaufenden Palette namens „qualitative Plotly-Farben“, sodass jede Position anders eingefärbt ist. Der Hintergrund ist Schwarz, die Schrift ist weiß, um den Kontrast zu maximieren. Eine leichte Transparenz und ein weicher Rand aus dem CSS-Overlay erzeugen ein modernes Aussehen. Der Nutzer kann mit einem Range-Slider die Saisons eingrenzen und Positionen in einem Dropdown selektieren. Dies aktualisiert dynamisch die Darstellung, um den Zeitraum oder den Vergleich auf bestimmte Positionen zu beschränken.

### 5.5 Liniendiagramm für Transferausgaben pro Verein und Saison

Ein mehrliniiges Diagramm visualisiert die Summe aller Transferausgaben pro Saison für ausgewählte Clubs. Jeder Verein ist eine eigene Linie. Da es sich um zeitliche Verlaufsdaten handelt, eignen sich Linienplots gut, um Trends zu zeigen.

Die Daten wurden nach „*team\_name*“ und „*season*“ gruppiert, summiert und in Millionen Euro umgerechnet. Zur besseren Übersicht werden die zehn Vereine mit den höchsten Gesamtausgaben dargestellt.

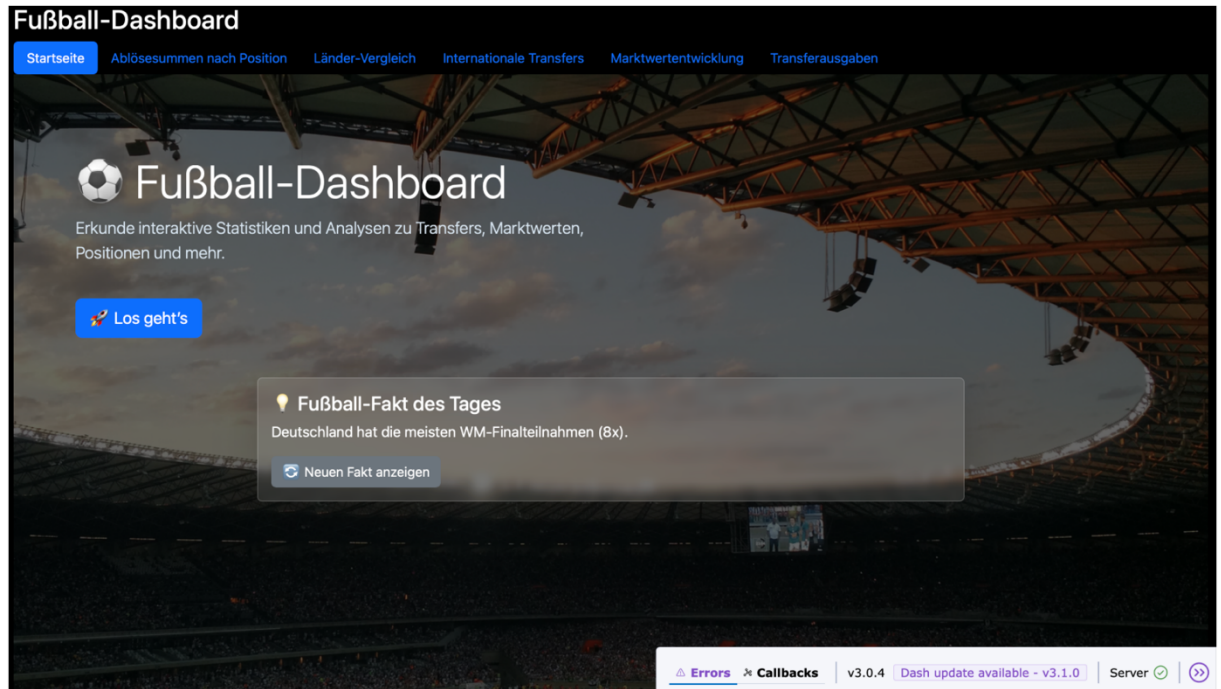
Ein Liniendiagramm bildet die Summen pro Saison ab, wobei jeder Verein durch eine klar unterscheidbare Farbe dargestellt wird. Die automatische Farbvergabe durch Plotly sowie Markierungen und Tooltips, wie „*px.line*“ ermöglichen eine präzise Auswertung.

Drei Filter-Dropdowns erlauben es, Saisons, Ligen oder einzelne Clubs auszuwählen. So kann man z. B. nur Champions-League-Vereine oder nur die letzten 3 Saisons betrachten und das Diagramm aktualisiert sich entsprechend. Linien sind mit Punkten markiert, damit einzelne Jahreswerte deutlich sichtbar sind. Die Hintergrundfarben und Schriftart sind konsistent mit dem Rest des Dashboards (schwarz/weiß) und mit einem Doppelklick auf einem der Vereine in der Legende, lässt sich dieses isoliert anzeigen.

## Anhang

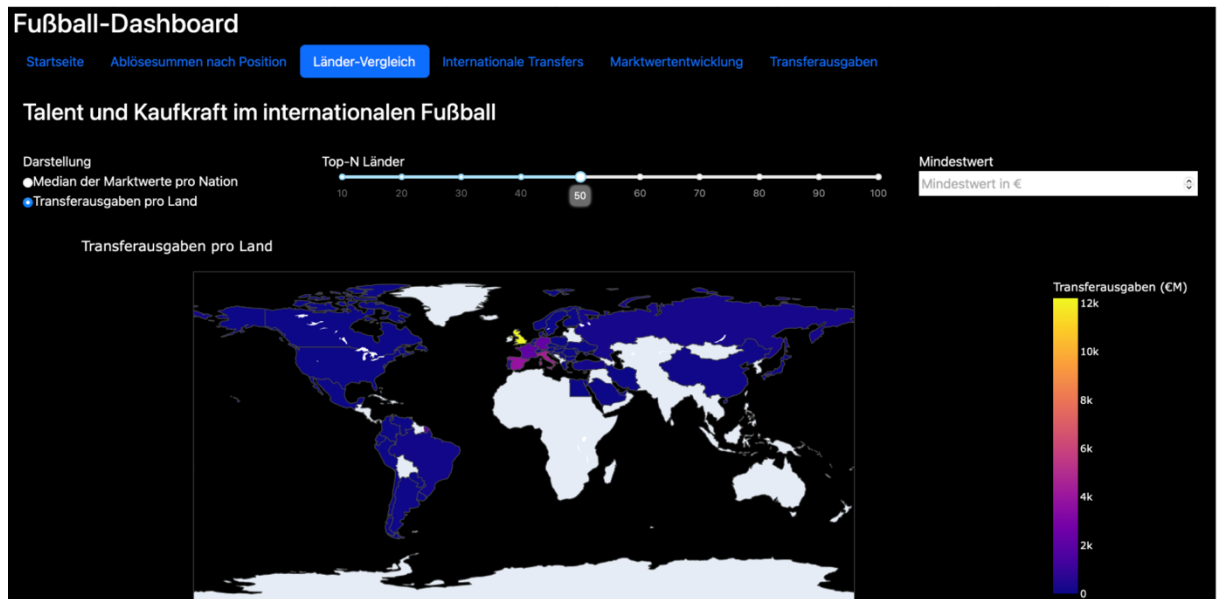
### Screenshots der Visualisierungen

#### 1. Startseite

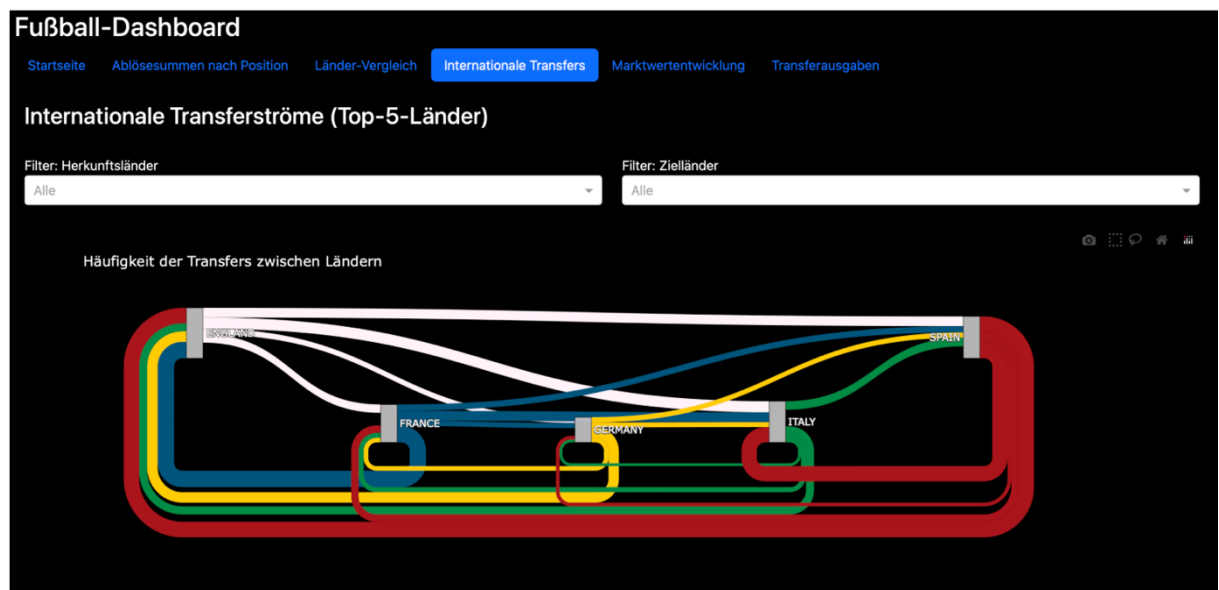


#### 2. Ländervergleich: Choropleth-Karten für Marktwerte und Transferausgaben nach Nationen





### 3. Internationale Transfers: Sankey-Diagramm für Transferströme zwischen Top-Ligen





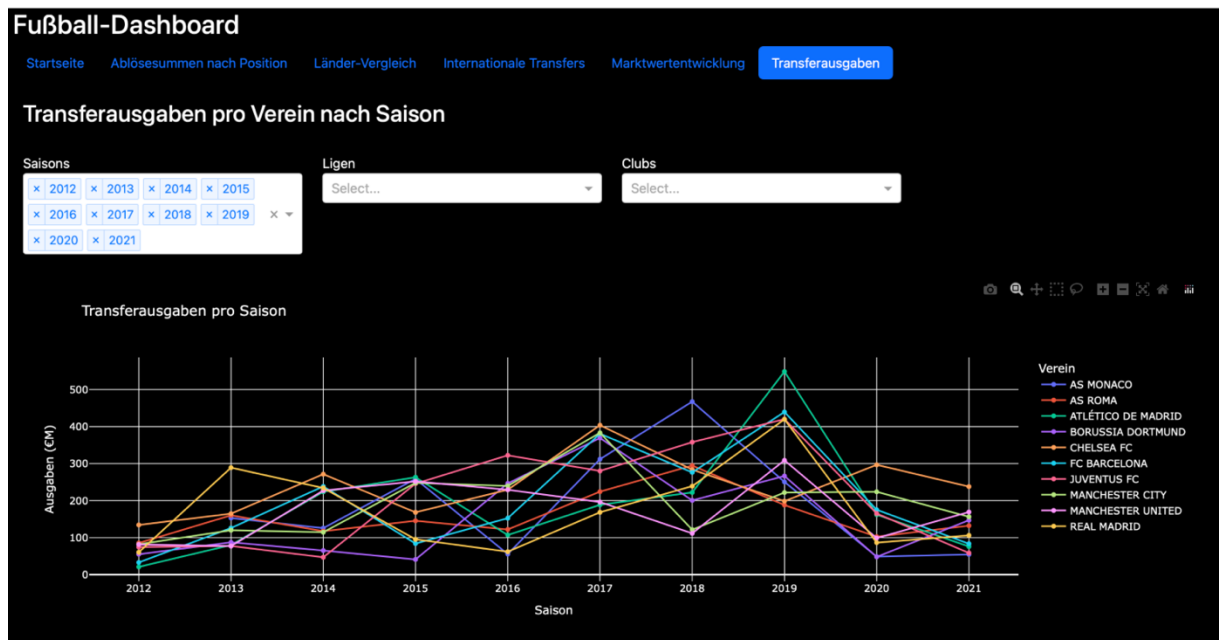
#### 4. Marktwertanalyse: Vergleich der Marktwerte junger vs. älterer Spieler nach Position



#### 5. Positionen: Durchschnittliche Ablösesummen nach Spielerpositionen



## 6. Transferausgaben: Visualisierung der Transferausgaben von Vereinen und Ligen



## Literaturverzeichnis

### Datensätze von Github und Mendeley Data Stand 27.06.2025

Football-transfers-data/dataset/

<https://github.com/d2ski/football-transfers-data/tree/main/dataset>

Network dynamics of the football transfer market in the 10 major European leagues from 1992 to 2022

<https://data.mendeley.com/datasets/rv7hktj86h/1?.com>

### Ehrenwörtliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe und diese Arbeit bei keiner anderen Prüfung mit gleichem oder vergleichbarem Inhalt vorgelegt habe und diese bislang nicht veröffentlicht wurde. Des Weiteren versichere ich, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

---

(Ort, Datum)

---

(Unterschrift)

