1. Expanded Sample Datasets

A. Player Performance Data (StatsBomb style)

player_id	player_name	team	matches	goals	assists	minutes_played	passes_completed	shots	season
101	Lionel Messi	PSG	28	25	18	2450	1200	150	2021/22
102	Kylian Mbappé	PSG	30	32	10	2600	950	170	2021/22
103	Neymar Jr.	PSG	25	15	12	2200	1100	140	2021/22
104	Kevin De Bruyne	Man City	31	12	20	2700	1600	100	2021/22
105	Erling Haaland	Man City	29	35	5	2500	700	180	2021/22
106	Vinícius Júnior	Real Madrid	33	20	15	2800	1300	160	2021/22
107	Luka Modrić	Real Madrid	32	6	14	2900	1700	70	2021/22
108	Mohamed Salah	Liverpool	34	27	12	3000	1400	190	2021/22
109	Robert Lewandowski	Barcelona	30	33	7	2750	900	200	2021/22
110	Jude Bellingham	Dortmund	28	10	9	2600	1500	90	2021/22

B. Market Value Data (Transfermarkt style)

player_id	player_name	current_value (€M)	highest_value (€M)	age	contract_until	position
101	Lionel Messi	35.0	180.0	35	2023	Forward
102	Kylian Mbappé	160.0	200.0	24	2024	Forward
103	Neymar Jr.	75.0	150.0	30	2025	Forward
104	Kevin De Bruyne	85.0	100.0	31	2025	Midfielder
105	Erling Haaland	170.0	180.0	22	2027	Forward
106	Vinícius Júnior	120.0	120.0	21	2026	Forward
107	Luka Modrić	10.0	80.0	36	2023	Midfielder
108	Mohamed Salah	90.0	120.0	30	2025	Forward
109	Robert Lewandowski	50.0	100.0	34	2024	Forward
110	Jude Bellingham	110.0	110.0	19	2026	Midfielder

C. Social Media Sentiment Data (Twitter + VADER/TextBlob)

player_id	player_name	tweet_id	tweet_text	sentiment_score	sentiment_label
101	Lionel Messi	t001	"Messi is still magical!"	0.91	Positive
102	Kylian Mbappé	t002	"Mbappé is a future Ballon d’Or"	0.88	Positive
103	Neymar Jr.	t003	"Neymar injured again, unlucky"	-0.40	Negative

player_id	player_name	tweet_id	tweet_text	sentiment_score	sentiment_label
104	Kevin De Bruyne	t004	"De Bruyne controls the midfield"	0.75	Positive
105	Erling Haaland	t005	"Haaland is a goal machine!"	0.92	Positive
106	Vinícius Júnior	t006	"Vinícius has improved a lot"	0.60	Positive
107	Luka Modrić	t007	"Modrić still class at 36"	0.70	Positive
108	Mohamed Salah	t008	"Salah missing easy chances lately"	-0.30	Negative
109	Robert Lewandowski	t009	"Lewandowski is underrated"	0.65	Positive
110	Jude Bellingham	t010	"Bellingham is the next superstar"	0.80	Positive

D. Injury History Data

player_id	player_name	injury_type	start_date	end_date	days_out
103	Neymar Jr.	Ankle Sprain	2021-11-01	2021-12-01	30
101	Lionel Messi	Hamstring Injury	2021-10-10	2021-10-25	15
108	Mohamed Salah	Groin Strain	2022-01-15	2022-02-05	21
109	Robert Lewandowski	Knee Injury	2021-12-05	2022-01-10	36
107	Luka Modrić	Muscle Fatigue	2022-02-20	2022-03-01	9

2. Exploration Coding

```
# Extended Exploration with 10 Players

# Quick overview
print("Performance Data:\n", performance_data.head())
print("\nMarket Value Data:\n", market_value_data.head())
print("\nSentiment Data:\n", sentiment_data.head())
print("\nInjury Data:\n", injury_data.head())

# Summary statistics
print("\nSummary Statistics - Performance Data:\n",
performance_data.describe())
print("\nSummary Statistics - Market Values:\n",
market_value_data.describe())

# Missing data check
print("\nMissing Values:")
for name, df in {
    "Performance": performance_data,
    "Market": market_value_data,
    "Sentiment": sentiment_data,
    "Injury": injury_data
}.items():
    print(f"{name} Data: \n{df.isnull().sum()}\n")

# Correlation between goals & market value
merged = pd.merge(performance_data, market_value_data, on=["player_id",
"player_name"])
sns.scatterplot(data=merged, x="goals", y="current_value (€M)",
hue="player_name")
plt.title("Goals vs Market Value")
plt.xticks(rotation=45)
plt.show()

# Distribution of sentiment
sns.histplot(sentiment_data["sentiment_score"], bins=10, kde=True)
plt.title("Distribution of Sentiment Scores")
plt.show()

# Top 5 most valuable players
top_players = market_value_data.sort_values("current_value (€M)",
ascending=False).head(5)
print("\nTop 5 Most Valuable Players:\n", top_players[["player_name",
"current_value (€M)"]])
```