

Predicting European Football Results

By Ben Royles

Purpose

Betting

- The sports betting market currently has an annual value of around \$85 billion USD.
- Predicted value of \$150 billion by 2030
- Football is 70% of the market

Fantasy football

- Fantasy sports had a value of \$24 Billion USD
- Expected to reach \$75 billion by 2030

Fans

- Fans wishing to understand the game better

End Result

- model that can predict match outcomes with 55% accuracy or more.

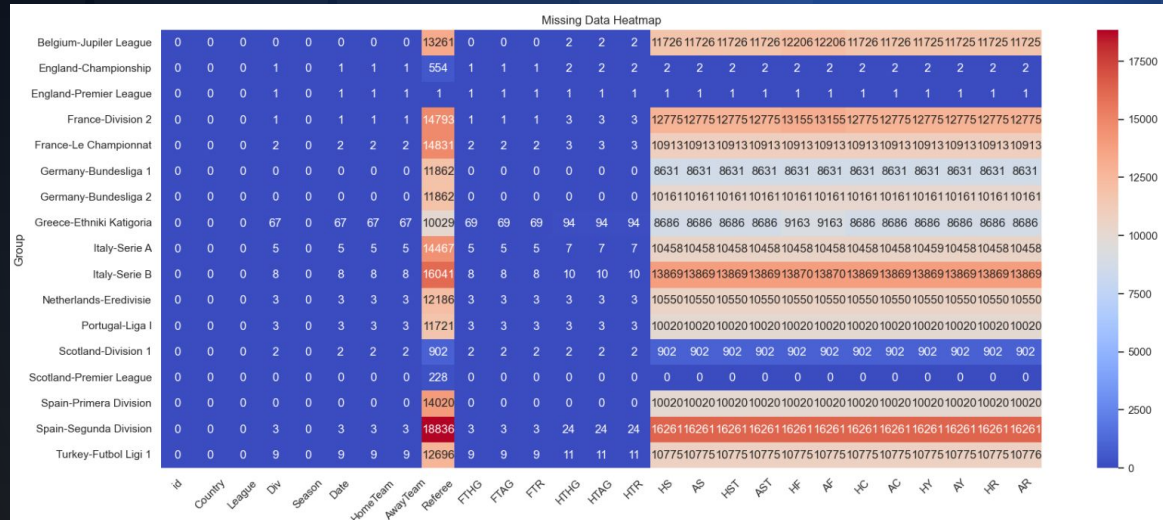
The Data

id	Country	League	Div	Season	Date	HomeTeam	AwayTeam	Referee	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR	AR
dkkXCXT5QHM	England	Premier League	E0	2012-2013	2018-08-12	Arsenal	Sunderland	C Foy	0	0	D	0	0	D	14	3	4	2	12	8	7	0	0	0	0	0
dAmonz9YM7c	England	Premier League	E0	2012-2013	2018-08-12	Fulham	Norwich	M Oliver	5	0	H	2	0	H	11	4	9	2	12	11	6	3	0	0	0	0
YwUhmYU4nas	England	Premier League	E0	2012-2013	2018-08-12	Newcastle	Tottenham	M Atkinson	2	1	H	0	0	D	6	12	4	6	12	8	3	5	2	2	0	0
iEk9YuADjHNVf	England	Premier League	E0	2012-2013	2018-08-12	QPR	Swansea	L Probert	0	5	A	0	1	A	20	12	11	8	11	14	5	3	2	2	0	0
bLQW9sPVeIPZ	England	Premier League	E0	2012-2013	2018-08-12	Reading	Stoke	K Friend	1	1	D	0	1	A	9	6	3	3	9	14	4	3	2	4	0	1
iDes8BMAyKAn	England	Premier League	E0	2012-2013	2018-08-12	West Brom	Liverpool	P Dowd	3	0	H	1	0	H	15	14	10	7	10	11	7	3	1	4	0	1

- Provided by a colleague, initially scrapped from footystats.com
- European Football data from 2012-2023
- 11 countries and 17 leagues
- Data needs transforming from descriptive to predictive
- Scraped from <https://football-data.co.uk>

Basic EDA

- Only 4 columns are complete
- Deleting missing rows could impact season running statistics
- Missing data in every league



Column	Null Values
id	0
Country	0
League	0
Div	105
Season	0
Date	105
HomeTeam	105
AwayTeam	105
Referee	178290
FTHG	107
FTAG	107
FTR	107
HTHG	165
HTAG	165
HTR	165
HS	145750
AS	145750
HST	145750
AST	145750
HF	147088
AF	147088
HC	145750
AC	145750
HY	145750
AY	145749
HR	145749
AR	145750

		id	Country	League	Div	Season	Date	HomeTeam	AwayTeam	Referee	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR	AR
141310	3HauqXQJoU2kaQoAmFpppT		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
92930	E5YhVp4dTVrJeThYtZX5		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
181156	DwLHeFPVJSeoUkqAktWACb		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
89360	CwJxv64nw5gd2o6vPyVcKf		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
185225	67zwDyg7sTr9n9QCSRjCzw		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
189304	WnJrPw9HmZnFWgMRo63KF		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
85792	dU7n5H4EWeeSdVuYktLdS		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
193509	fXSkIzPJSwp3KUaSGUN6nr		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
177087	2mtJRW4KY9mxsDBmRCCDNC		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
82224	no8NaUFGX9e7wx36hJWew		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0
78769	ajf8CkDsPZ7YWGcW2vduUw		England	Championship	E1	2022-2023	02/01/2023	Wigan	Hull	G Eltringham	1.0	4.0	A	0.0	1.0	A	6.0	9.0	1.0	4.0	10.0	8.0	7.0	5.0	1.0	2.0	0.0	0.0

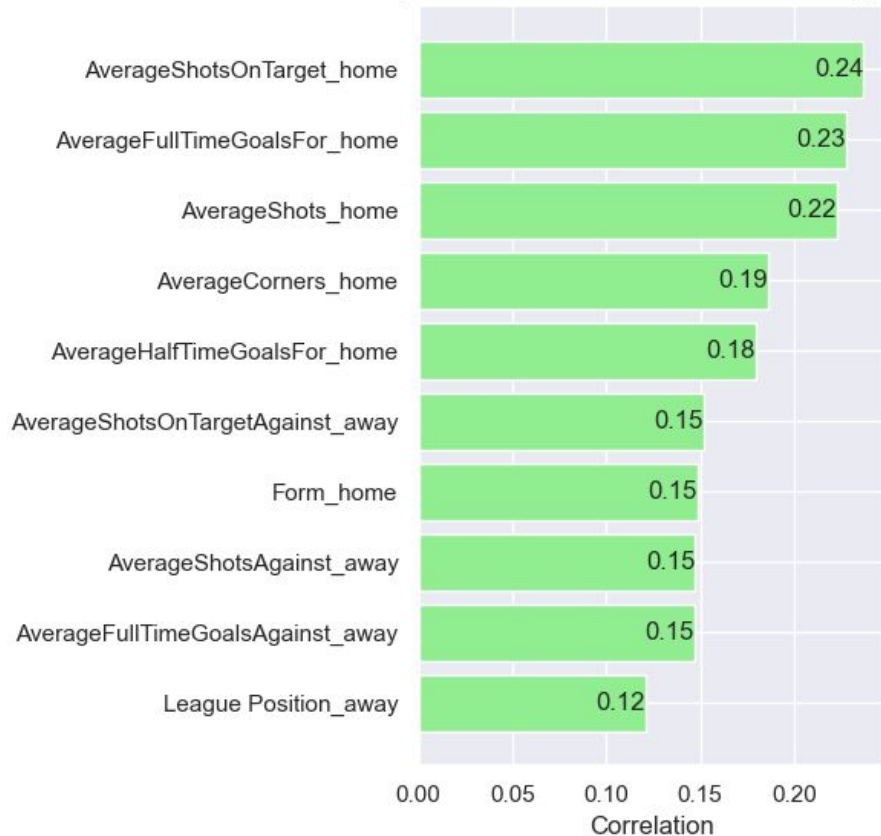
[illegible]

Cumulative statistics

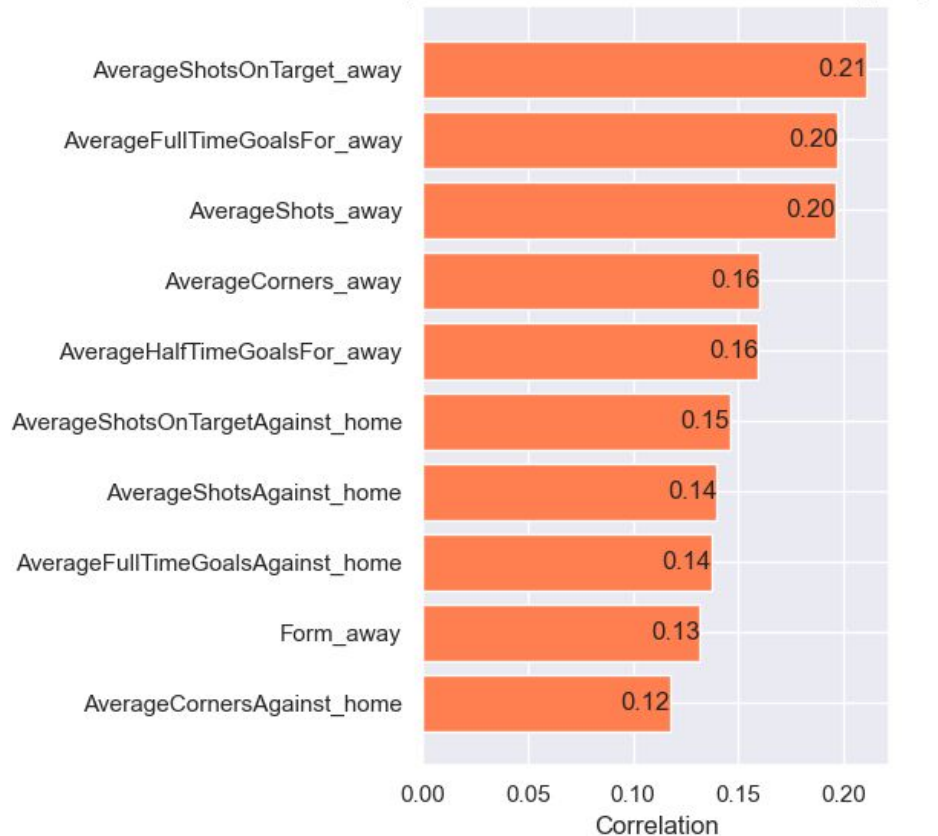
Points	Points before match	Match Week	League Position	Form	AverageFullTimeGoalsFor	AverageFullTimeGoalsAgainst	AverageHalfTimeGoalsFor	AverageHalfTimeGoalsAgainst	AverageShots	AverageShotsAgainst	AverageShotsOnTarget
3	12	8	4	0.533333	1.000000	0.571429	0.714286	0.000000	10.142857	9.000000	3.428571
0	51	30	4	0.266667	1.793103	1.068966	0.724138	0.551724	16.862069	8.758621	5.689655
1	24	23	12	0.533333	1.045455	1.409091	0.409091	0.636364	8.772727	8.272727	4.500000
1	0	4	10	0.000000	0.666667	2.333333	0.000000	1.333333	11.000000	12.333333	3.666667
AverageShotsOnTargetAgainst				AverageFoulsCommitted	AverageFoulsAgainst	AverageCorners	AverageCornersAgainst	AverageYellowCards	AverageYellowCardsAgainst	AverageRedCards	AverageRedCardsAgainst
2.857143				13.142857	13.285714	4.000000	5.142857	1.857143	2.571429	0.142857	0.000000
3.413793				10.586207	10.862069	7.103448	3.931034	1.655172	1.620690	0.000000	0.034483
4.318182				13.545455	14.136364	4.863636	4.227273	1.909091	2.500000	0.045455	0.181818
6.000000				16.000000	10.666667	9.333333	4.666667	1.666667	2.000000	0.000000	0.000000

Key Relationships

Top 10 Correlations with FullTimeGoalsFor_home



Top 10 Correlations with FullTimeGoalsFor_away



Putting the data frame back together

id	League	Season	Date	Team	Against	HomeStadium	HomeCapacity	HomeCity	Match Week	FullTimeResult	DistanceTraveled in Km	Weekday	Month
bULedur5aGZ43fxoN3WfAm	Le Championnat	2013-2014	2013-08-24	Montpellier	Sochaux	Stade de la Mosson	32939.0	Montpellier	3	H	491.6	5	8

Home	Away
Points before match_home	Points before match_away
League Position_home	League Position_away
Form_home	Form_away
AverageFullTimeGoalsFor_home	AverageFullTimeGoalsFor_away
AverageFullTimeGoalsAgainst_home	AverageFullTimeGoalsAgainst_away
AverageHalfTimeGoalsFor_home	AverageHalfTimeGoalsFor_away
AverageHalfTimeGoalsAgainst_home	AverageHalfTimeGoalsAgainst_away
AverageShots_home	AverageShots_away
AverageShotsAgainst_home	AverageShotsAgainst_away
AverageShotsOnTarget_home	AverageShotsOnTarget_away
AverageShotsOnTargetAgainst_home	AverageShotsOnTargetAgainst_away
AverageFoulsCommitted_home	AverageFoulsCommitted_away
AverageFoulsAgainst_home	AverageFoulsAgainst_away
AverageCorners_home	AverageCorners_away
AverageCornersAgainst_home	AverageCornersAgainst_away
AverageYellowCards_home	AverageYellowCards_away
AverageYellowCardsAgainst_home	AverageYellowCardsAgainst_away
AverageRedCards_home	AverageRedCards_away
AverageRedCardsAgainst_home	AverageRedCardsAgainst_away

Country	Div
Country_Belgium	Div_B1
Country_England	Div_D1
Country_France	Div_D2
Country_Germany	Div_E0
Country_Greece	Div_E1
Country_Italy	Div_F1
Country_Netherlands	Div_F2
Country_Portugal	Div_G1
Country_Scotland	Div_I1
Country_Spain	Div_I2
Country_Turkey	Div_N1
	Div_P1
	Div_SC0
	Div_SC1
	Div_SP1
	Div_SP2
	Div_T1

Each observation is in a similar format to the original.

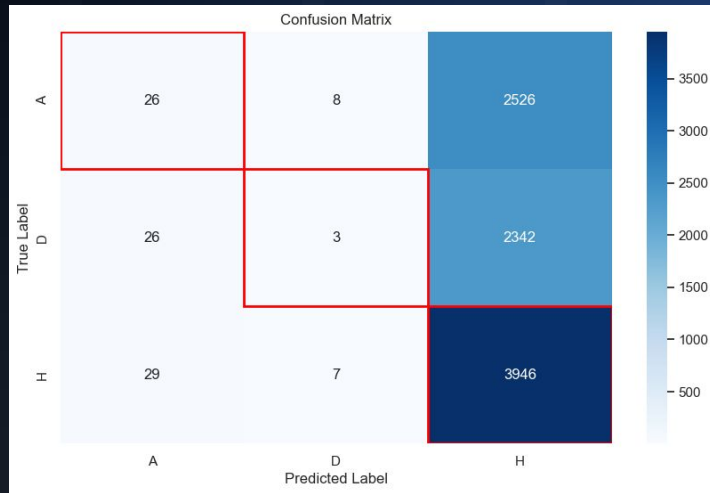
League position and Form run for the entire season.

All of the Average columns work out the Average for that season per match

Initial Modelling

Logistic Regression

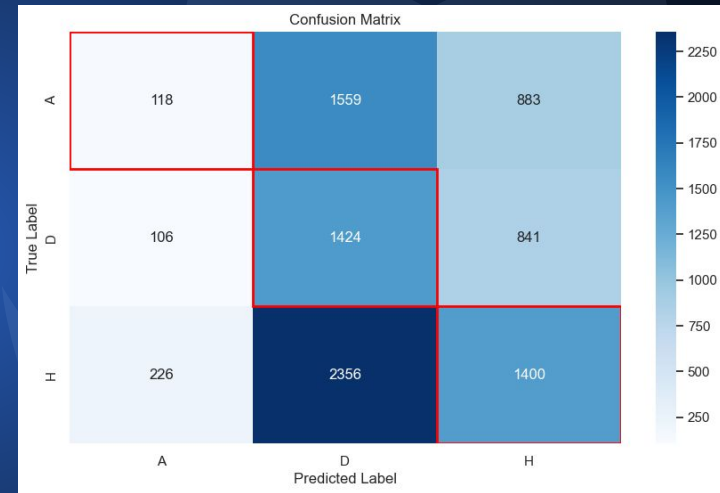
- Training score = 44.9%
- Test Score = 44.6%



	precision	recall	f1-score	support
A	0.32	0.01	0.02	2560
D	0.17	0.00	0.00	2371
H	0.45	0.99	0.62	3982
accuracy			0.45	8913
macro avg	0.31	0.33	0.21	8913
weighted avg	0.34	0.45	0.28	8913

Decision Tree

- Training score = 46.2%
- Test Score = 44.3%
- Depth = 9

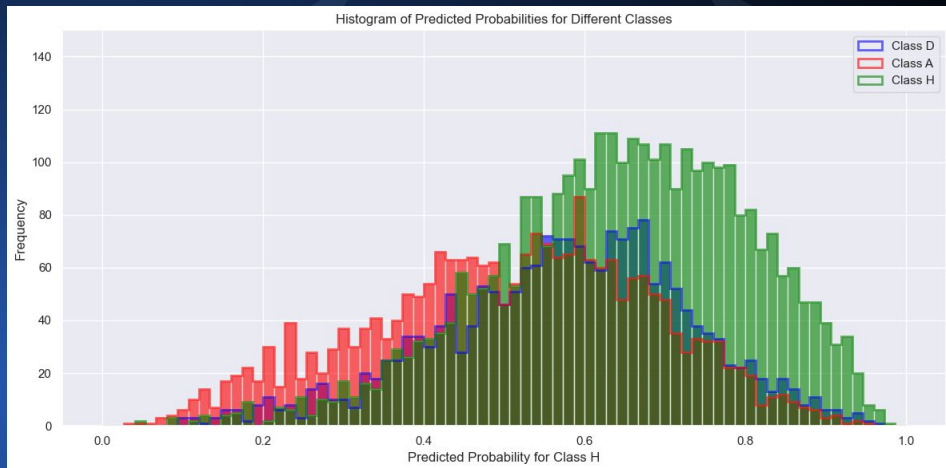


	precision	recall	f1-score	support
A	0.26	0.05	0.08	2560
D	0.27	0.60	0.37	2371
H	0.45	0.35	0.39	3982
accuracy			0.33	8913
macro avg	0.33	0.33	0.28	8913
weighted avg	0.35	0.33	0.30	8913

Advanced Models

Model	Accuracy
Logistic Regression	50.3%
Binary Logistic Regression	50.4%
Binary Logistic Regression & decision boundary	49.6%
Decision Tree	48.9%
Random Forest	50.2%
XGBoost	50.5%

- Draws are the hardest to predict
- Most models over predict class H
- Stratified sampling lead to lower accuracy – class H is the most frequent



Development

I may look to add more features to the dataset.

- Formations used
- Weather
- Time since last match

Advanced feature creation – conversion rate, goal differentials, league differential

Tailoring the model to specific use cases – betting strategy, understanding the game better etc



Thank you for listening

Any questions?

Glossary

Div = League Division

Date = Match Date (dd/mm/yy)

Time = Time of match kick off

HomeTeam = Home Team

AwayTeam = Away Team

FTHG and HG = Full Time Home Team Goals

FTAG and AG = Full Time Away Team Goals

FTR and Res = Full Time Result (H=Home Win, D=Draw, A=Away Win)

HTHG = Half Time Home Team Goals

HTAG = Half Time Away Team Goals

HTR = Half Time Result (H=Home Win, D=Draw, A=Away Win)

Referee = Match Referee

HS = Home Team Shots

AS = Away Team Shots

HST = Home Team Shots on Target

AST = Away Team Shots on Target

HC = Home Team Corners

AC = Away Team Corners

HF = Home Team Fouls Committed

AF = Away Team Fouls Committed

HY = Home Team Yellow Cards

AY = Away Team Yellow Cards

HR = Home Team Red Cards

AR = Away Team Red Cards