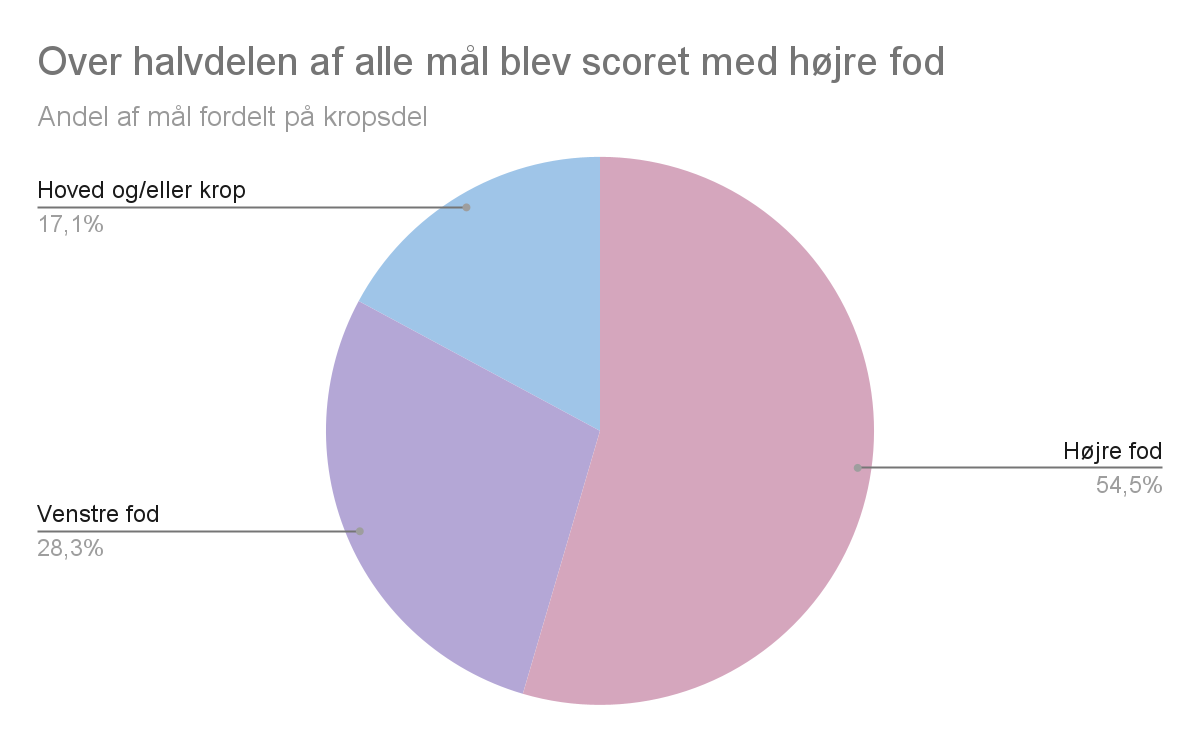
**Expected Goals – 3 ligaer**

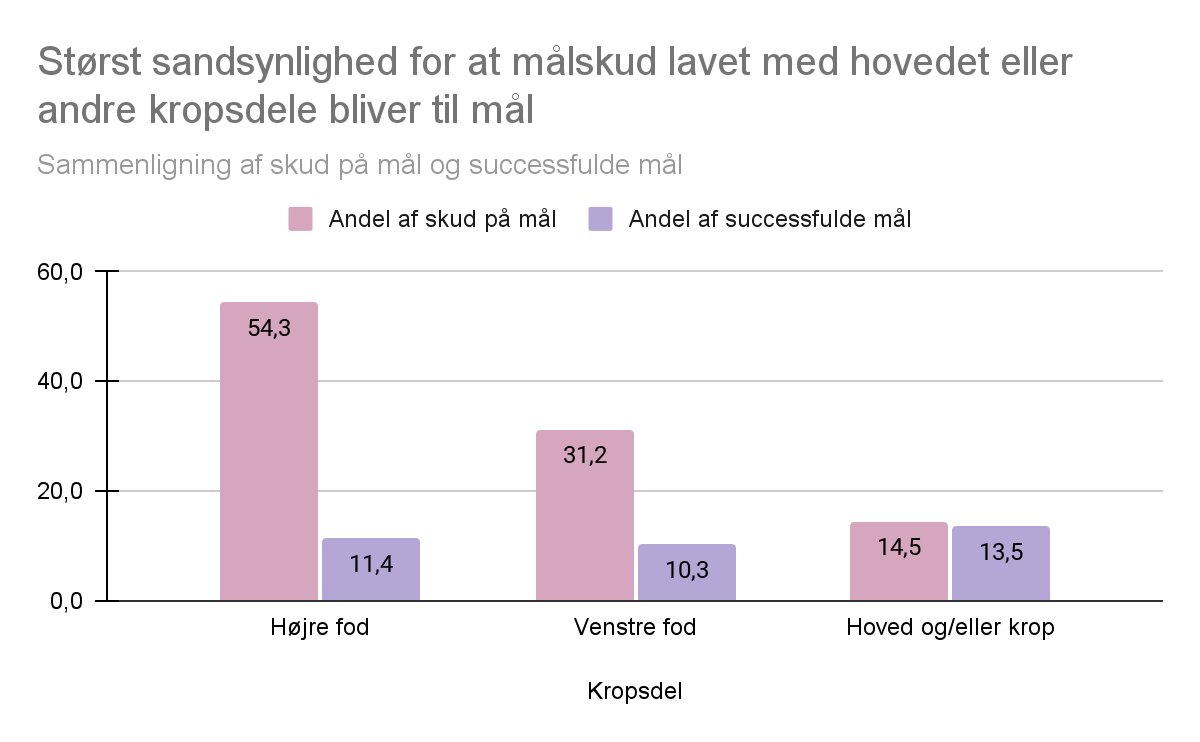
Vi har først samlet alt data for hver liga i én dataframe per liga. Dette giver os 3 datasæt med hver 1000 events (skud) af 47 variabler. Vi har valgt at lægge de tre ligaer sammen, for at undgå over- og underestimering af antal mål, da vi antager at målfrekvensen er forskellig i de forskellige ligaer.

**Feature Engineering**

**Forklarende variabler:**

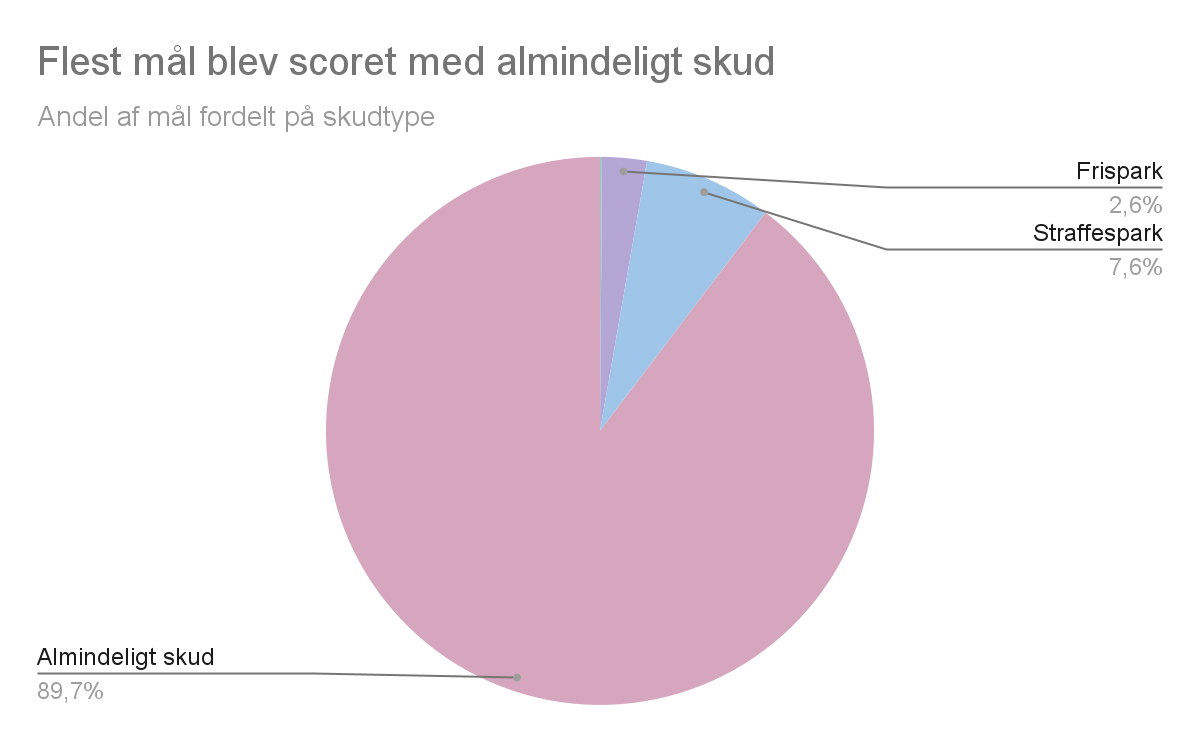
**Kropsdel:** Denne variabel beskriver hvilken kropsdel der blev brugt til at lave skuddet. Der er følgende tre kropsdele: højre fod, venstre fod og hovedet og/eller kroppen. Sidste udfald anvendes hvis det blev brugt mere end én kropsdel.





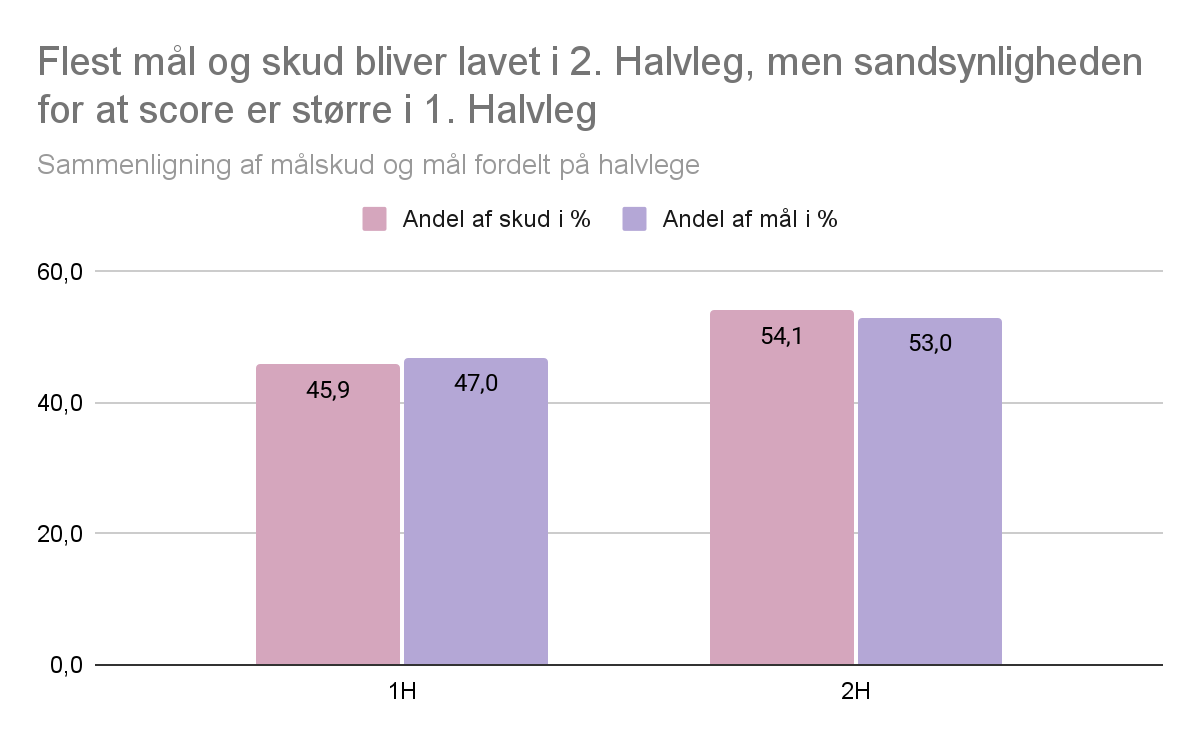
**Skudtype:**

Denne variabel beskriver hvilken type skud på mål der blev lavet: almindeligt skud, hjørnespark, frispark eller straffespark.



**Halvleg:**

Denne variabel beskriver om skuddet blev lavet i første eller anden halvleg.



**Lokation X & Lokation Y:**

Disse variabler beskriver spillerens position på banen, angivet i procent fra 0 til 100, hvor 0 er baglinjen på holdets egen banehalvdel og 100 modstanderens baglinje.

**Mål:**

Gennemsnitlige X position for mål: 84

Gennemsnitlige Y position for mål: 49

(Begge koordinater er afrundet til nærmeste halve tal.)

Diagram

Description automatically generated

**Skud:**

Gennemsnitlige X position for skud: 84,5

Gennemsnitlige Y position for skud: 49,5

(Begge koordinater er afrundet til nærmeste halve tal.)

Der er altså kun 0,5% forskel på den gennemsnitlige lokation for mål og mislykkedes skud på mål. Det er dog interessant at mislykkede skud på mål er 0,5% tættere på målet end succesfulde skud.

En fodboldbane til internationale kampe er 64-75 meter bred og 100-110 meter lang[[1]](#footnote-1). Den gennemsnitlige professionelle bane er 68 \* 105 meter [[2]](#footnote-2).

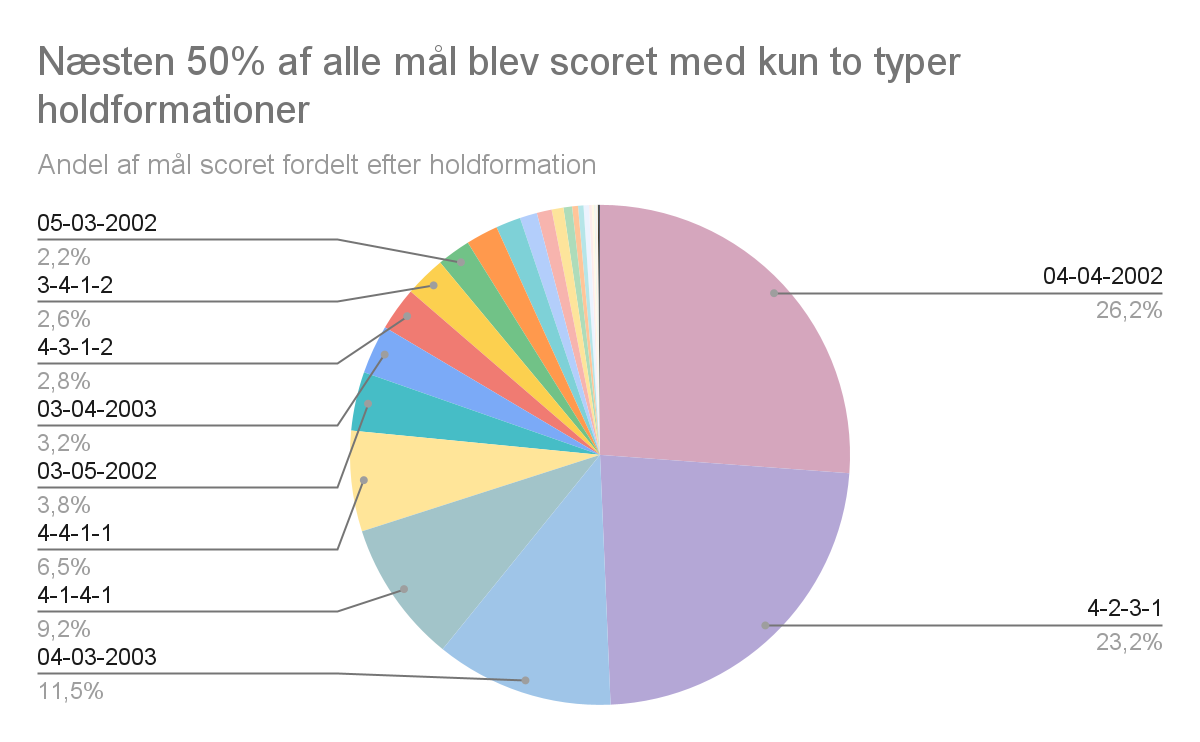
Tager vi udgangspunkt i disse dimensioner, er det gennemsnitlige succesfulde målskud 0,525 m længere væk fra målet end de mislykkede skud. De mislykkede målskud er samtidig 0,34 meter tættere på midten end succesfulde målskud.

**Team formation:**

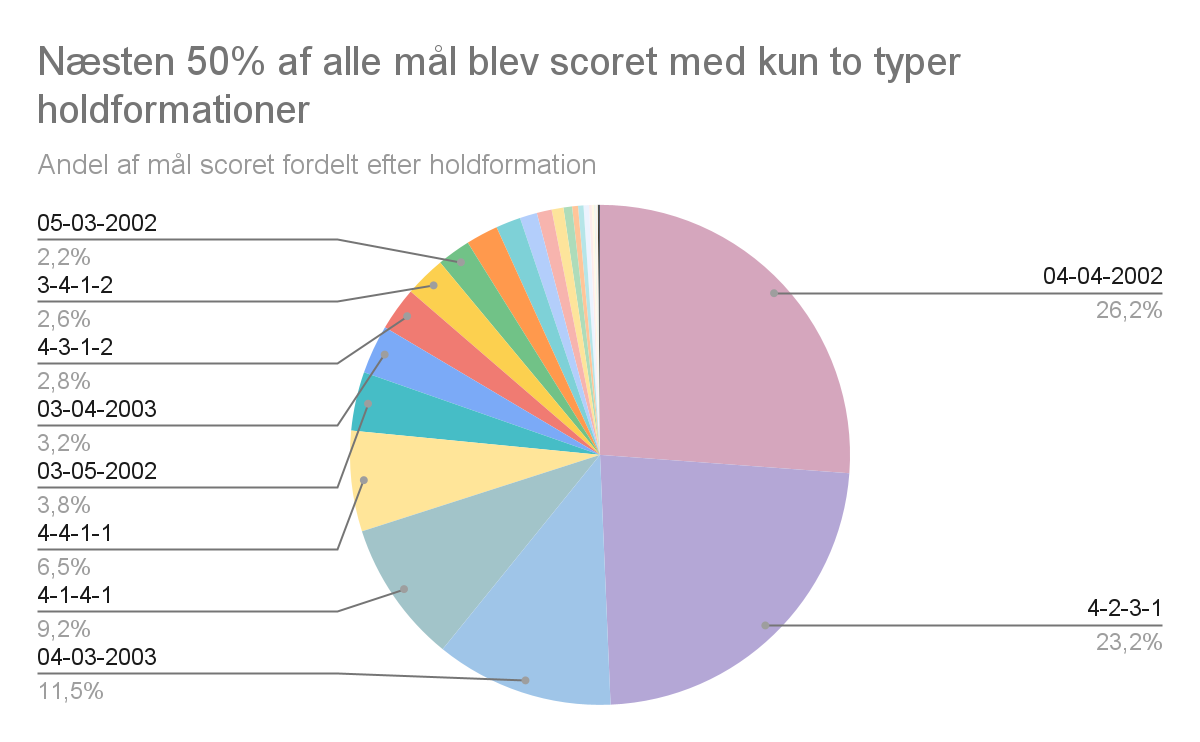
Denne variabel beskriver holdformationen ved mål for det angribende hold.

4-4-2 er standard formationen. Dataen fra Wyscout er formateret således at 4-4-2 hedder “04-04-2002”. Vi tænker at dette er en fejl der er sket ved indtastningen af dataen fra Wyscouts side, da det ses flere steder under “team formations” i rådataen (csv filerne).

**Mål:**

****

**Mislykkede målskud:**

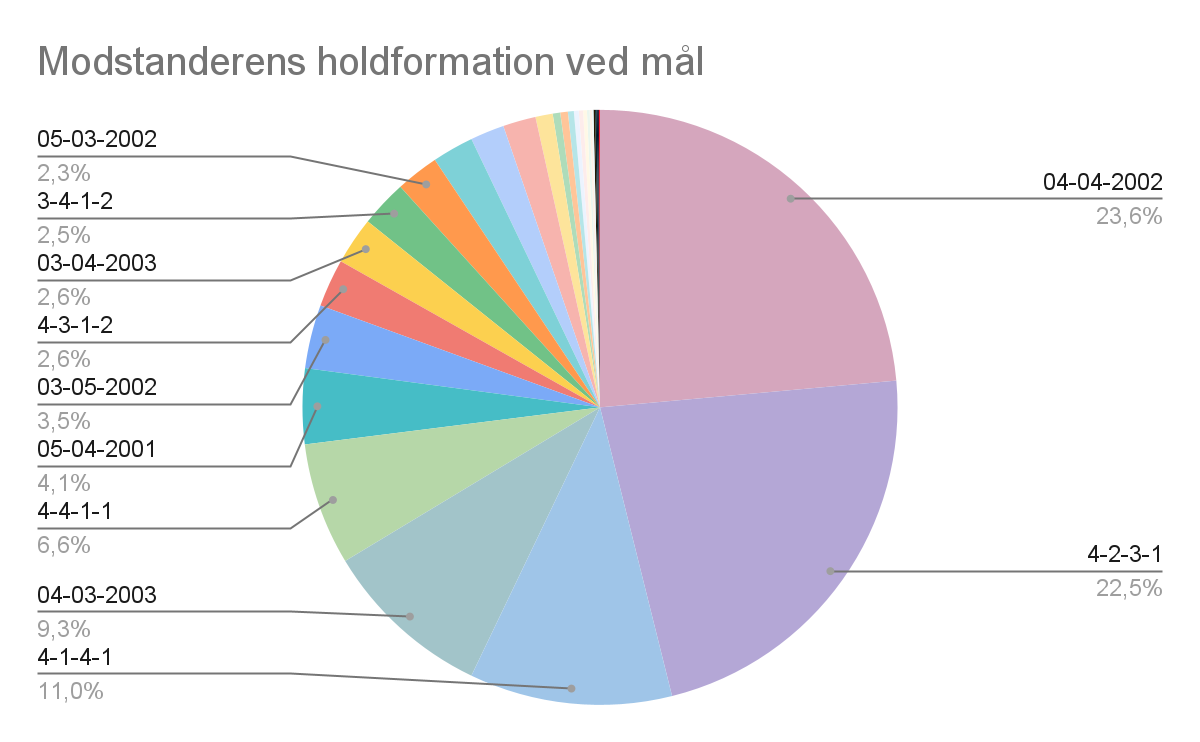
****

Vi kan ud fra ovenstående figurer se at holdformationen umiddelbart ikke har nogen indflydelse på om et målskud er succesfuldt eller ikke.

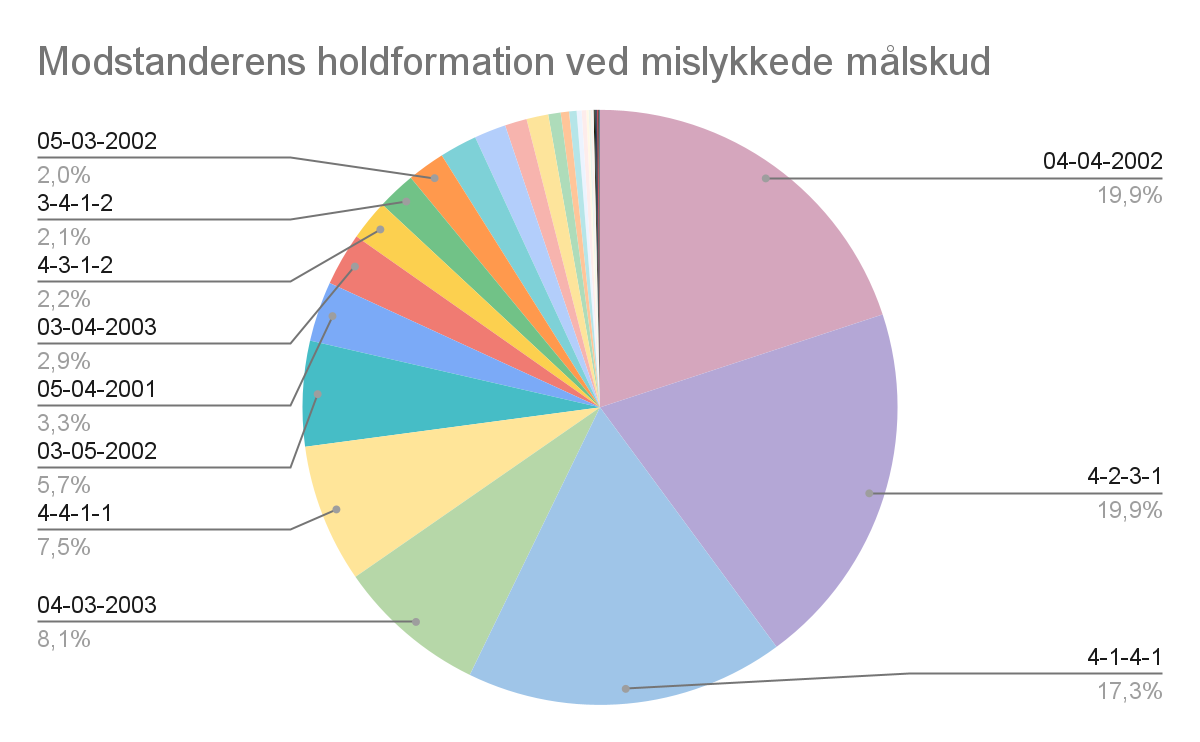
**Opponent team formation:**

Denne variabel beskriver modstanderens holdformation.

**Mål:**

****

**Mislykkede målskud:**



Denne variabel kan bruges til at undersøge hvilken holdformation der er mest defensiv.

**Målzone:**

Denne variabel beskriver hvor på målet skuddet blev lavet. Målet er opdelt således:

Calendar

Description automatically generated

**Mål:**

**Calendar

Description automatically generated**

**Mislykkede målskud:**

27,2% af alle mislykkede målskud blev blokeret. Resten fordeles således:

Calendar

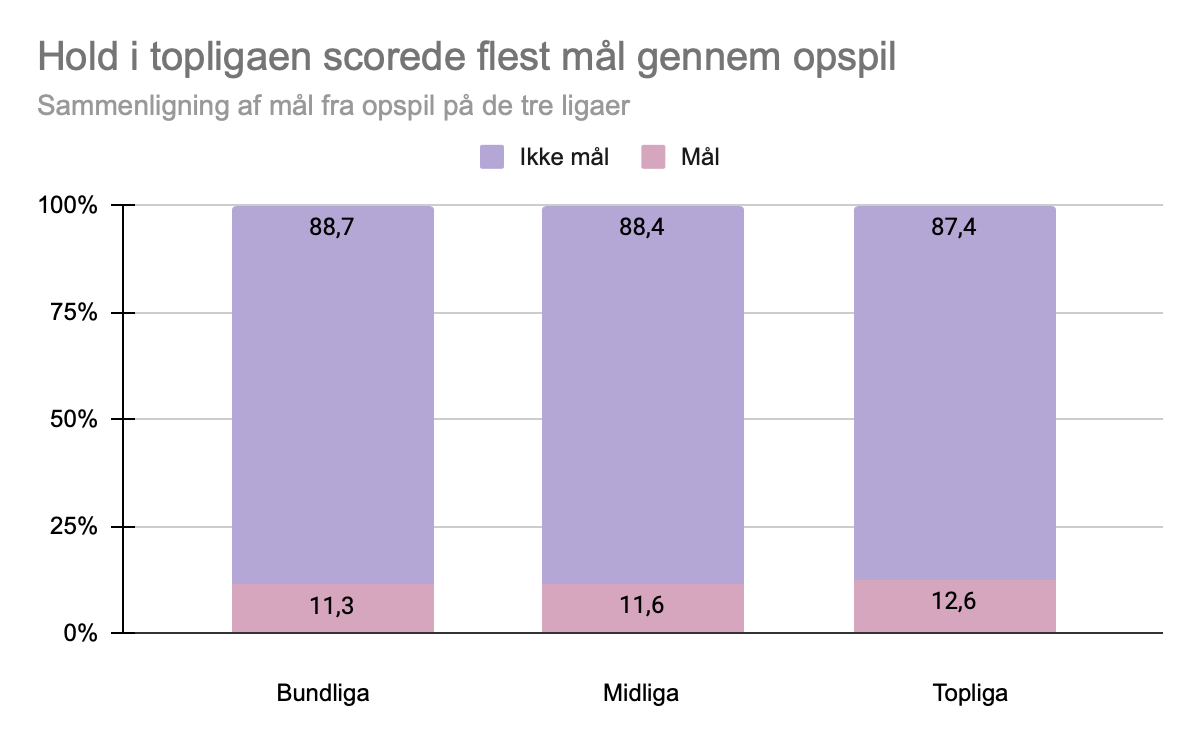
Description automatically generated

De mørkeblå tal er off-target. De lyseblå tal ramte stolpen.

## **Skud og opspil**

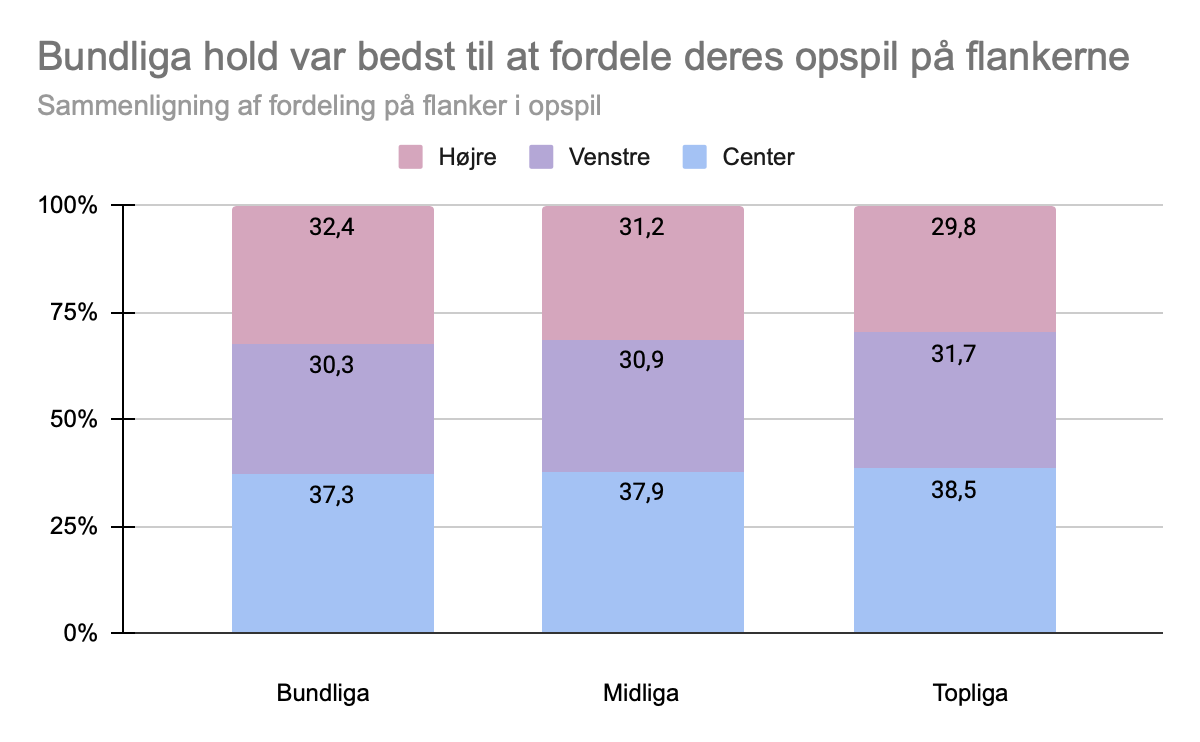
Opspil er en blanding af defensivt spil (lav risiko, spiller dybere på banen, inviterer mere pres fra modspillerne) og short plays (penetrate space, cycling possession, creating space by drawing players out of position). Derudover indeholder opspil ofte lange afleveringer fra defensive spillere, og færre driblinger[[3]](#footnote-3).

**Mål fra opspil:**



**Opspil fordelt på flanke:**

(flanke: side af en fodboldbane hvor et angreb føres frem)[[4]](#footnote-4)



**Længde på opspil:**

Gennemsnitlige længder.

|  |  |  |  |
| --- | --- | --- | --- |
| Liga | Start X | Slut X | Længde |
| Bundliga | 55,4 | 82,7 | 27,2 |
| Midliga | 57,3 | 82,8 | 25,4 |
| Topliga | 57,9 | 82,2 | 24,3 |
|  |  |  |  |
|  | Start Y | Slut Y | Længde |
| Bundliga | 51,3 | 49,6 | 1,6 |
| Midliga | 50,6 | 49,5 | 1,0 |
| Topliga | 49,9 | 49,7 | 0,2 |

Tabellen viser os at bundligahold spiller dybere og længere i opspil.

Holdene i bundligaen er altså bedre til at lave opspil, men det lykkedes dem ikke at score lige så mange mål som mid- og topliga holdene.

**Expected Goals**

For at forsøge at forudsige om et givent skud bliver til et mål for alle tre ligaer, har vi valgt at benytte en random forest model. Fordelen ved at benytte træ-baserede modeller frem for lineære modeller til klassificering er at de ofte er nemmere at forstå, da de nærmere afspejler hvordan mennesker tager beslutninger[[5]](#footnote-5).

Vi har valgt at benytte random forest frem for et classification tree, da det er bedre til større datasæt. Random forest er ofte mere nøjagtigt end et classification tree, dog tager det længere at træne modellen.

Her kan vi se hvordan vores random forest model har vurderet vores forskellige variablers vigtighed. Jo højere en MeanDecreaseGini variablen har, jo vigtigere er den i forudsigelsen af et mål.   
Modellen er lavet på træningsdata af 75% af datasættet.

Table

Description automatically generated

## **Validering af xG model**

For at teste vores xG model har vi lavet en forudsigelse på om skud blev til mål på de resterende 25% af datasættet.   
Vi har derefter lavet en confusion matrix for at sammenligne vores forudsigelser med de egentlige mål/ikke mål fra datasættet.

|  |  |  |
| --- | --- | --- |
|  | Forudsigelse: Ikke mål | Forudsigelse: Mål |
| Egentlige: Ikke mål | 6556 | 854 |
| Egentlige: Mål | 21 | 60 |

|  |  |
| --- | --- |
| Sensitivitet | 0,7407 |
| Præcision | 0,0656 |

Vores xG model er dårlig til at forudsige at skud bliver til mål, men god til at forudsige at skud ikke bliver til mål.

1. <https://downloads.theifab.com/downloads/laws-of-the-game-2021-22?l=en> s. 33 [↑](#footnote-ref-1)
2. https://authoritysoccer.com/official-fifa-soccer-field-dimensions/ [↑](#footnote-ref-2)
3. <https://www.guidetofm.com/tactics/possession/#suggested-formations> [↑](#footnote-ref-3)
4. <https://ordnet.dk/ddo/ordbog?query=flanke> [↑](#footnote-ref-4)
5. “An Introduction to Statistical Learning” s. 339 [↑](#footnote-ref-5)