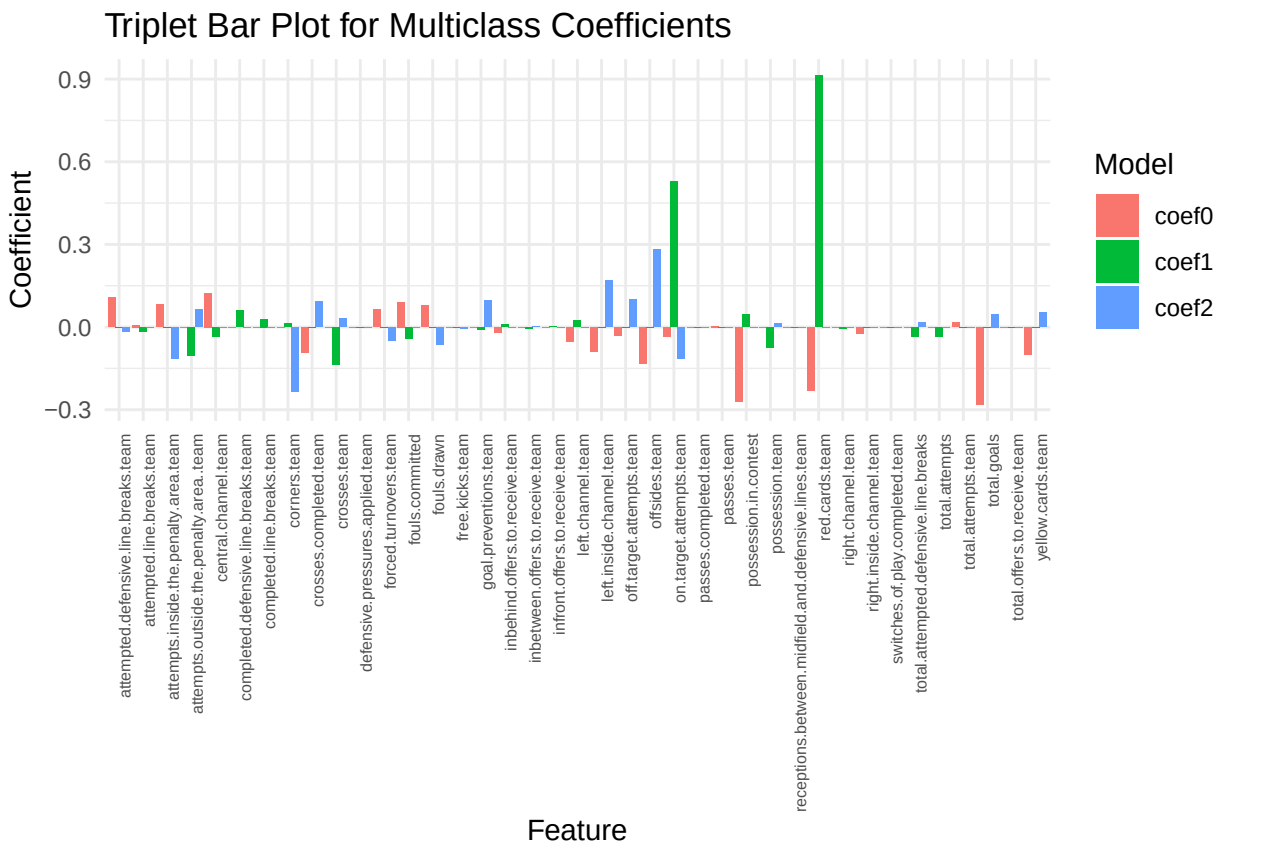
imp

2024-12-07

Which variables are most important for these predictions?

```
library(ggplot2)
long_df <- read.csv("coef_df.csv")


#Plot the triplet bar graph
#Compare coefficient values across all three classes (tie, loss, win)
ggplot(long_df, aes(x = Variable, y = Value, fill = Model)) +
  geom_bar(stat = "identity", position = "dodge") +  # position dodge places bars next to each other
  theme_minimal() +
  labs(title = "Triplet Bar Plot for Multiclass Coefficients",
       x = "Feature",
       y = "Coefficient") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size=6))
```




```
ggsave("en_coef_plot.png")
```

```
## Saving 6.5 x 4.5 in image
```

It looks like red cards and on target attempts were most important for determining the outcome of the games, based on the plot. Surprisingly, the log odds of winning versus tieing increase significantly for each red card a team receives! Although I do not have data to support this, I would suspect that when a winning team received a red card, it often happened near the end of the game. This way, the team would not have to play a man down for a long time. Another surprising result is that possession was not particularly influential in determining match outcome.