Sydney Lockwood

The Relationship Between Sacks and Turnovers on Win Probability in the NFL

Question Snapshot

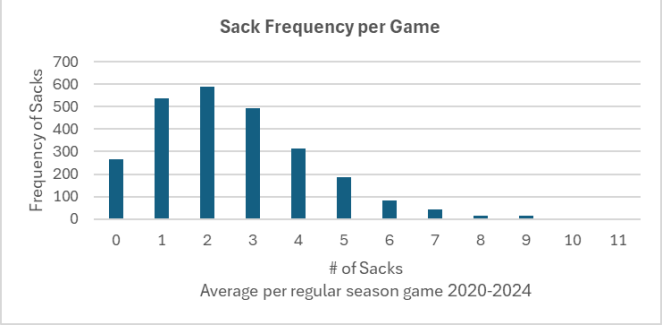
This project examines how sacks and defensive turnovers contribute to a team's win probability in the NFL. By comparing defensive pressure metrics to a team's win success, the analysis can help identify the hidden value of defensive efficiency beyond traditional metrics. This is a useful insight for players, stakeholders, and coaches to leverage in games.

Data Used

In the cleaned file used for data exploration, one row represents a team in a single game, season, total sacks, total turnovers caused, and a win or loss. The data exploration is all regular-season games from 2020-2024 with no major filters applied.

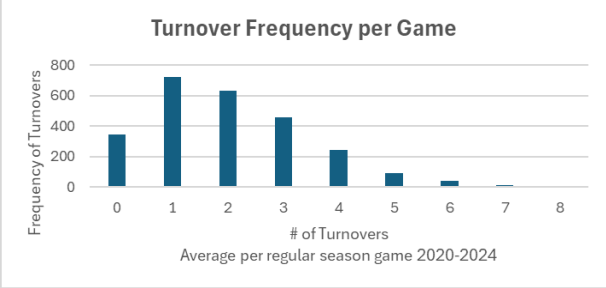
Univariate Distributions

Sacks



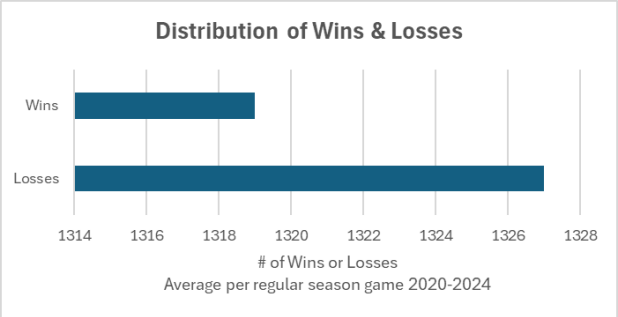
Shape is slightly right-skewed. Most teams record 1-4 sacks, with a long tail for big defensive games. No negative data issues for sacks.

Turnovers



Shape is also right-skewed. Turnovers tend to cluster around 0-2, with occasional high turnover games to skew the data. The graph suggests that turnovers are a relatively rare event, but the data has no obvious issues.

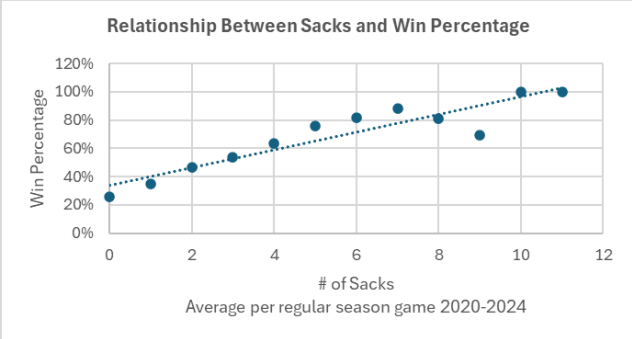
Win/Loss



The distribution of wins and losses is nearly even, there are slightly more losses than wins. The distribution should be exactly equal since the distribution of any game is 50%. With this in mind, some differences could come from missing data rows, ties (rare but neither counts as a loss or a win), or data cleaning differences.

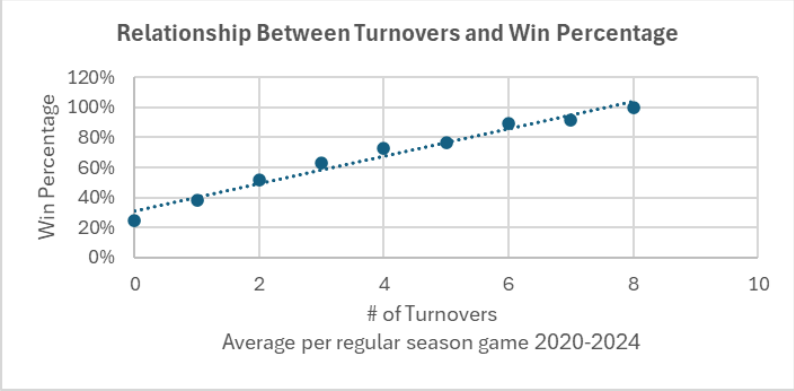
Relationships (Pairs)

Sacks vs. Win Percentage



The scatterplot shows a slight upward trend. Games with higher sack totals correspond more frequently to wins.

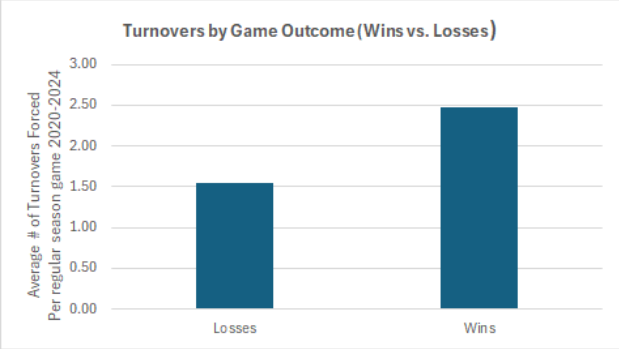
Turnovers vs. Win Percentage



A clear positive relationship, as turnovers gained increase, win frequency rises. Evidence suggests turnovers have a stronger influence on wins than sacks.

Subgroup Comparisons

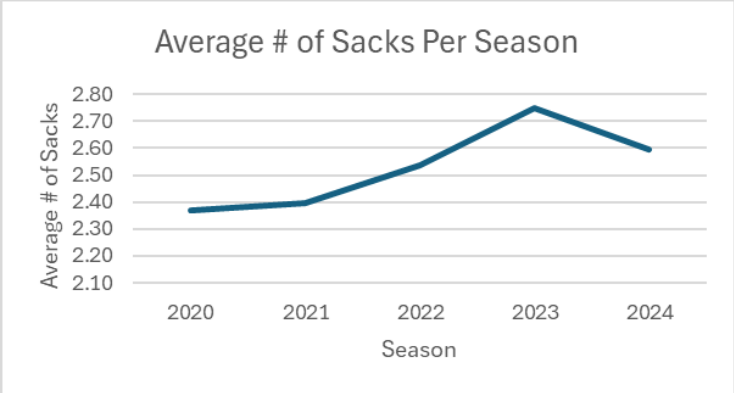
Comparing turnovers by win vs loss



Winning teams show significantly higher takeaways, suggesting turnovers are a key differentiator.

Time or Sequence View

Sacks over time (by season)



Sack totals show small year-to-year variation without a strong upward/downward trend. Some fluctuation may be due to schedule differences.

Ouliters & Anomalies

For sacks, there are some outliers. These are from games with 7-10 sacks; these are legit NFL events, not errors. There are also a few turnover-heavy games with 6-8 turnovers, these were also high-producing events.

Missingness & Coverage

Missing values are minimal and do not cluster by season or team, so they are unlikely to bias results. The missing values are mainly limited to the season entries or sacks/turnovers that were never recorded.

Early Takeaways

- Turnovers show the strongest relationship with winning, even 1 turnover drastically improves the win rate
- Sack totals show a mild positive relationship with win probability, but with greater noise
- Year-to-year patterns appear stable, suggesting these defensive metrics are consistent predictors

Next Steps

- Build a simple logistic regression model predicting wins using sacks and turnovers
- Layering in the win percentage from another source may help the analysis