Canadian Highschool Football Offensive & Defensive Tendancies

Liam Cook

2025-03-10

Questions to be investigated

1. How do Offensive and Defensive play calling tendencies shift based on field position and game situations. I.E. How often did X team call X Formation/Coverage/Run or Pass on the X Yard Line or X Quarter?
2. Which defensive coverages (e.g., Cover 2, Cover 3, Cover 4) yield the highest success rates in certain situations. I.E. How effective was Cover X in X Down/Yard Line/Quarter?
3. Which offensive formations were primarly used in run situations vs passing situations. I.E. How often did X team run in X formation?

All of the above questions are those of which we study and analyze on a week to week basis during the season for each opponent.

Data to be used

I will use a dataset containing high school football defensive statistics collected from our team's 2023-2025 seasons. This dataset has been exported from QwikCut and includes:

- **Game Information:** Date, opponent.
- **Play Characteristics:** Run vs. pass, play direction (left, right, middle), hash location, pass zone (area of the field), Formations, Coverage, Blitz, etc.
- **Situational Data:** Down and distance, field position, game quarter.

Dataset Details

- **Source:** Our team's game logs and statistical tracking via QwikCut Film review. All data is taken from sideline stat collectors and film review after games of ourselves and future opponents.
- **Size:** Estimated 5000+ plays from multiple games across 2-3 seasons.
- **Notes:** The data still needs a significant amount of time to clean and be expanded upon, so for now this data is from a couple of games with a good amount of data.
- **Types of Data:**
    - Categorical: Defensive coverage type, offensive formation, play direction, game quarter, opponent name.
    - Quantitative: Number of plays per coverage, number of plays and play types per offensive formation, field position, pass zone distribution. Player stats can also be added, however, they are not linked play by play.

```
library(tidyverse)
library(readxl)
library(skimr)
```

```
## Warning: package 'skimr' was built under R version 4.4.3
```

```
loc <- "C:/Users/lm820/OneDrive/Desktop/School/Stats/Project/GameData.xlsx" #This path may ne
ed to be updated depending on where you saved the folder

# Read the Excel file and explicitly set column types to avoid incorrect assumptions
game_data <- read_excel(loc, sheet = 1, col_types = c(rep("text", 67)))

# Convert numeric columns back after loading
numeric_cols <- c("Down", "Distance", "Start Yardline", "End Yardline", "GN/LS", "Pass Zone")

# Convert only if the column exists in the dataset (avoids errors if a column is missing)
numeric_cols <- intersect(numeric_cols, colnames(game_data))
game_data[numeric_cols] <- lapply(game_data[numeric_cols], as.numeric)

# **Automatically detect and exclude empty columns**
empty_cols <- colnames(game_data)[colSums(!is.na(game_data)) == 0]  # Columns where all value
s are NA
game_data_filtered <- game_data %>% select(-all_of(empty_cols))  # Remove them dynamically


glimpse(game_data_filtered)
```

```
## Rows: 909
## Columns: 28
## $ `Clip #`          <chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "…
## $ `Clip Number`     <chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "…
## $ Quarter           <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ ODK               <chr> "K", "O", "O", "O", "O", "O", "O", "D", "D", "D", "K…
## $ Possession        <chr> NA, "Bayview High School", "Bayview High School", "B…
## $ Hash              <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ Down              <dbl> NA, 1, NA, NA, NA, NA, NA, 1, NA, NA, NA, NA, 1, 2, …
## $ Distance          <dbl> NA, 10, NA, NA, NA, NA, NA, 10, NA, NA, NA, NA, 10, …
## $ `Start Yardline`  <dbl> -40, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ `Motion Dir`      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ `Play Dir`        <chr> NA, "Right", "Left", "Right", "Left", "Left", "Left"…
## $ `Play Type`       <chr> "Kickoff", "Pass", "Pass", "Pass", "Pass", "Run", "P…
## $ `GN/LS`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ `Off. Formation`  <chr> NA, "ACE", "ACE", "JOKER", "QUADS", "ACE", "ACE", NA…
## $ Backfield         <chr> NA, NA, NA, NA, NA, NA, NA, "PISTOL", "PISTOL", "PIS…
## $ `Off. Play`       <chr> NA, "REVERSE PASS", "DOMI BRADY", "ZAP WK", "BRADY A…
## $ `Def. Front`      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ Coverage          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ Blitz             <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ Motion            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ `Pass Thrown`     <chr> NA, "Yes", "Yes", "Yes", "Yes", NA, "Yes", NA, NA, "…
## $ `Pass Completion` <chr> NA, "Yes", "Yes", "Yes", "Yes", NA, "Yes", NA, NA, "…
## $ Sack              <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ Turnover          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ Touchdown         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ `FG Made`         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ `Onside Kick`     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ `Kick Blocked`    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
```

```
skim(game_data_filtered)
```

Data summary

Name	game_data_filtered
Number of rows	909
Number of columns	28
_____	
Column type frequency:	
character	24
numeric	4
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Clip #	0	1.00	1	3	0	156	0
Clip Number	0	1.00	1	3	0	156	0
Quarter	471	0.48	1	1	0	4	0
ODK	85	0.91	1	1	0	4	0
Possession	256	0.72	3	36	0	10	0
Hash	688	0.24	1	1	0	3	0
Motion Dir	896	0.01	4	5	0	2	0
Play Dir	493	0.46	4	5	0	2	0
Play Type	381	0.58	2	9	0	8	0
Off. Formation	539	0.41	3	16	0	25	0
Backfield	576	0.37	6	11	0	7	0
Off. Play	644	0.29	4	22	0	143	0
Def. Front	877	0.04	2	11	0	5	0
Coverage	782	0.14	7	18	0	9	0
Blitz	896	0.01	6	10	0	5	0
Motion	817	0.10	6	24	0	40	0
Pass Thrown	780	0.14	3	3	0	1	0
Pass Completion	789	0.13	2	3	0	2	0
Sack	904	0.01	3	3	0	1	0
Turnover	896	0.01	6	12	0	2	0
Touchdown	895	0.02	3	3	0	1	0
FG Made	893	0.02	3	3	0	1	0
Onside Kick	908	0.00	3	3	0	1	0
Kick Blocked	908	0.00	3	3	0	1	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Down	494	0.46	1.53	0.71	1	1	1	2	4	
Distance	521	0.43	9.38	3.12	1	10	10	10	28	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Start Yardline	608	0.33	6.03	36.03	-49	-35	10	40	50	
GN/LS	765	0.16	6.17	11.46	-63	2	5	12	40	

Ideas for visualization

1. **Heatmap of Defensive Success Rates by Field Position**
    - X-axis: Field Position
    - Y-axis: Defensive Call
    - Color: Success rate in preventing 1st downs and/or yards gained
2. **Bar Chart of Play Calls by Situation**
    - X-axis: Down and Distance
    - Y-axis: Frequency of Defensive or Offensive Calls
    - Color: Defensive Coverage Type or Offensive formation
3. **Directional Flow Chart of Plays**
    - Visualizing how often plays go left, right, or middle and defensive response rates
4. **K-means Clustering of Offensive Tendencies**
    - Cluster analysis grouping offensive tendencies based on field position, play direction, and offensive formation

Team planning

- **Liam:** Data cleaning and preparation, exploratory data analysis. I have been working with this data for years.
- **Liam:** Creating and refining visualizations.
- **Liam:** Writing and structuring the final report.
- **Liam:** Preparing presentation. Theres a good chance this gets presented to the team!

Challenges anticipated

- Ensuring data quality and consistency across different sources (games).
- Likely that some games are much more thoroughly analyzied and documented than others.
- Selecting the most informative visualizations for complex relationships.
- Properly interpreting clustering and results for defensive/offensive tendencies.

References

- Dataset source: Exported from QwikCut (2023-2025 team and opponent statistics)
- Football and tendancy knowledge will come from me as a previous player and the defensive cordinator for this team when analyzing data.