DESAUTELS Faculty of Management
Faculté de gestion

McGill

McGill University

Desautels Faculty of Management

Community Project - McGill Football

Written by

Shuxi (261208189)

Hamza Javed (260768221)

JP (260954757)

July 19th, 20

Executive Summary

The Problem:
The McGill Football coaching staff, led by Maxime and Chris, faced two core challenges that were limiting the team's ability to fully leverage data. Maxime's challenge was operational: he spent hours each week manually creating the same game reports in Excel using DV Sport data — a process prone to errors and inefficiencies. Chris's challenge was infrastructural: he aimed to enrich the team's analytics with external data sources like CFL, Genius Sports, and Twitter, but faced access and integration barriers.

Our Solution:
For Maxime, we delivered a suite of 10 automated reports — 8 built directly into the DV Sport system, and 2 (Reports 9 and 10) created in Excel to allow deeper multi-dimensional analysis. These reports capture play tendencies, formation effectiveness, field zone targeting, and yardage attribution between offensive lines and running backs. Our approach emphasized ease of use, visual clarity, and zero ongoing maintenance.

For Chris, we built a modular, Python-based external data pipeline. This included:

- A working Twitter data collection system using `snscrape`, with sentiment-ready CSV outputs.
- A fully built (but not yet deployed) pipeline using the Twitter API v2.
- A future-facing integration framework for CFL and Genius Sports data, pending access credentials.

The Impact & Value Proposition:
Maxime now has a fast, repeatable reporting system embedded directly in his weekly workflow, freeing up hours for game planning and player development. The Excel templates for advanced reports require only a data paste to refresh. For Chris, we laid the groundwork for scalable, external data integration — a future-proof foundation for modeling opponent tendencies, player fatigue, and fan engagement.

Both solutions were designed with long-term sustainability in mind: zero coding knowledge required, no cost, and full alignment with existing tools already used by the coaching staff.

Key Recommendations:

- **For Maxime:** Continue using the DV Sport "defaults" setup for weekly in-platform reports, and paste new game data into the Excel templates for advanced insights. Consider involving DVSport support to fully automate Reports 9 and 10 internally.
- **For Chris:** Secure credentials for Genius and CFL APIs to unlock full potential of the external data framework. Assign a point person to maintain and gradually expand the system's capabilities.

I. Project Overview & Scope

Introduction to the Community Partner/stakeholders

The McGill Redbirds is more than just a varsity program — it's a cornerstone of school spirit and tradition on campus. As one of the oldest football teams in the country, it represents a long-standing legacy of athletic excellence, pride, and community at McGill. The team brings people together — students, alumni, faculty, and fans — creating a shared sense of energy and purpose both on and off the field. Beyond the game itself, it plays a big role in shaping leadership, work ethic, and resilience among student-athletes. As part of this project, we've had the chance to work closely with two of the team's coaches, Maxime and Chris, who have both provided valuable insight and a clear passion for using data to support player development and smarter decision-making. Their involvement has made this initiative feel like a true collaboration between athletics and analytics.

As part of this project, **Coach Maxime** outlined ten specific reports he wanted us to create, each designed to help the coaching staff make more informed, data-driven decisions. His goal was to move beyond just watching film and start uncovering deeper patterns in how the team performs — from player positioning and play types to directional tendencies and field coverage. Each report ties directly into a practical coaching question, like understanding where the ball is typically thrown, how different formations affect outcomes, or how blocking and running responsibilities are split. Maxime's requests were clear, detailed, and grounded in the realities of game preparation, and they gave us a solid roadmap to build out analyses that could actually support coaching strategy. **Chris** also contributed a set of analytical priorities. His request focuses on integrating external data sources, specifically connecting to the CFL and Genius API using existing credentials and joining these databases to existing internal data. He also emphasized the importance of linking CFL and Genius data to enhance player and team level insights. Beyond that, Chris also demonstrates the opportunities for improving the existing database, including building dashboards to support simple query execution and refining existing predictive models. These requests helped us think more holistically about infrastructure, interoperability, and long term scalability of our analytics work.

The Core Challenge - Operational Inefficiency

The project began with two primary stakeholder contacts: Maxime Chaput Dupuy and Christopher Clement. The initial objectives outlined in our SOW included integrating external data APIs (CFL/Genius, PFF), optimizing database performance, refining predictive models, and developing an interactive dashboard. However, a deviation from the original SOW occurred within Chris's project stream. While the SOW outlined a broad set of technical tasks like database optimization and CFL API integration, these were superseded by a direct and more urgent request from the client to perform a social media analysis using the Twitter API. This pivot was a conscious, collaborative decision made to ensure our work remained aligned with the client's most immediate priorities.

This led to the project evolving into two distinct streams:

- **Maxime's Stream**: One of Max's core responsibilities as a coach is to review game film and identify patterns that can inform both game planning and player development. While DV Sport

provides a powerful platform for organizing and tagging plays, the process of pulling meaningful insights from that data has been largely manual. Each week, Max has had to sift through hundreds of clips and datasets, exporting information into Excel, reorganizing it by hand, and building the same types of reports from scratch. This repetitive process has become increasingly time-consuming, often taking hours to complete — especially when trying to cross-reference multiple variables like formations, play direction, player alignment, and field position. The manual nature of the work also leaves room for human error and makes it harder to maintain consistency from week to week. As a result, valuable coaching time is spent wrangling spreadsheets instead of focusing on film analysis or strategic planning.

- **Chris's Stream**: Chris initially focused on integrating external data sources — including Genius, CFL, and PFF — into the existing database to improve model accuracy and expand player and team-level analysis. Midway through the project, however, he requested a shift toward analyzing public sentiment via Twitter. While this added a valuable new perspective, it introduced technical hurdles: Twitter's data access is now highly restricted, with historical data requiring a costly Enterprise API. Although we explored tools like snscrape, they provided limited coverage, making large-scale analysis difficult without significant investment.

Project Goals & Objectives

Given the evolution of the project, our primary goal was refined: to deliver two distinct, client-centric analytics solutions that addressed the most urgent needs of each stakeholder.

- **Goal for Maxime's Stream**: Max's main goal in working with us was to make this entire workflow more efficient. What he needed was a practical and easy-to-use analytics solution that could automate the generation of his weekly reports. Specifically, he wanted to eliminate the time spent copying and filtering data manually and instead have a system that updates instantly when new game data is uploaded. He emphasized that the solution had to be intuitive — something he could easily interact with, even without a technical background. By automating the reporting process, Max hoped to free up time for more high-value tasks like player coaching, film breakdown, and opponent scouting. Ultimately, he wanted a tool that helps him make smarter, faster decisions without getting bogged down in spreadsheets.
- **Goal for Chris's Stream**: Chris's main objective in working with us was to expand the breadth and richness of available data by integrating sources such as CFL's internal platform, Genius Sports, and Twitter. His focus was less on immediate reporting and more on building a strong data foundation that could unlock a wide range of downstream applications. By accessing and consolidating these diverse data streams, Chris aimed to enable more advanced analytics, such as enhancing existing models with external features, uncovering deeper player insights, and even developing predictive models to anticipate game outcomes, player performance, or fan engagement trends. In short, he saw data as the fuel for smarter decisions — both on and off the field — and our goal was to help him access and structure that data in a scalable, actionable way.

II. The Dual-Stream Projects

Part I: The DV Sport Workflow Automation Project (for Maxime)

Project Overview & Purpose

Problem Statement
Offensive Coordinator Maxime Chaput Dupuy and his team faced a significant operational bottleneck in their strategic analysis workflow. To generate insights, they were forced into a redundant and time-consuming process of manually creating and filtering numerous single-use datasets within their DVSport system. This was necessary to accommodate the unique row-level logic each desired report demanded. The existing workflow was not only inefficient and prone to error, but also unsustainable, limiting the coaches' ability to perform timely, in-depth analysis as their needs grew throughout the season.

Expected Impact & Value Proposition
The primary goal of this project stream was to create a suite of 10 standardized, strategically-focused reports and automate their weekly generation within the client's existing DVSport analytics platform. We delivered a fully automated, zero-cost, internal reporting template that could be integrated directly into their workflow, empowered with readable visuals for clear communication among the coaches and players.
We aim to eliminate the coaching staff's manual, multi-hour workflow, empowering the team with faster, more consistent insight. By automating this process, our solution enables coaches to redirect hundreds of labor-hours each season away from redundant data tasks and into high-value activities like deeper game-planning, strategic review, and athlete mentoring.

Methodology & Data-Driven Strategic Plan

To ensure each report directly answers the coaching staff's strategic questions and fits seamlessly into their existing DVSport workflow, we adopted a structured five-phase methodology. We began by working with the client's "Single Competition" dataset to ground our approach in real game context: we explored its schema, verified key field relationships, and built initial Python prototypes. This hands-on foundation then informed our subsequent phases: domain mapping, iterative client validation, hybrid platform implementation, and automated deployment, guaranteeing both accuracy and ease of use.

Phases	Description
1. Domain Research & Data Dictionary	• Reviewed football strategy materials to understand each abbreviated field name (e.g., g_FPOS, O_BOX). • Compiled these mappings into a simple Data Dictionary
2. Logic Deconstruction & Python Prototypes	• Broke down the calculations required for each of the 10 reports, wrote grouped-aggregation scripts, generated

	pivot-table mock-ups, and created simple visualizations.
3. Client Feedback & Alignment	• Presented each prototype and confirmed details such as how percentages should be calculated or be viewed. • Discovered report 9 & 10 required multidimensional grouping beyond the DVSport module's capabilities, we agreed to deliver them via Excel instead.
4. Technical Implementation (In-Platform Reports + Excel Template)	• Implemented 8 reports directly in DVSport, each has 2 versions of views - "Global" vs. "Subgroup" views. • Organized rows and columns to follow a coach's natural workflow (e.g., Personnel → Play Type → Direction). Labels and formatting were chosen for quick interpretation under time constraints. • Built Reports 9 and 10 in a pre-formatted Excel workbook with pivot tables and calculated fields to handle advanced grouping and yardage splits.
5. Maintenance & Future Data Streams	• **In-Platform Reports**: All 8 DVSport scripts are saved ("Defualts") in the system, so coaches can run them weekly without rewriting code. • **Excel template**: Designed a single input tab where coaches paste raw CSV data. All pivot tables and charts update automatically, requiring no additional setup.

Solution

The final solution was shaped by the inherent constraints of the DVSport reporting environment, which does not support custom formulas or nested subtotals and computes all percentage metrics against the total play count. First, a set of **Global reports** was generated to satisfy baseline requirements by displaying aggregate play distributions. Recognizing the coaching staff's need for context-specific insight, we then engineered **Subgroup reports**: by adjusting field configurations and repurposing native percentage metrics (e.g., % RUN), we simulated subtotal calculations within defined segments.

Each core report is available in both Global and Subgroup format, enabling coaches to transition seamlessly between high-level overviews and detailed situational analyses, thereby supporting more nuanced, data-driven strategic planning.

1. **The "Global View"**

   This report is structured like a classic **crosstab** or **matrix**.

- **Metrics:** You select a generic metric pair like (# PLAYS, % PLAYS).
- **Structure:** One variable is placed on the rows (g R/P), and another is placed on the columns (OFF PERS).
- **Percentage Logic (Global):** The system's rule for this generic % PLAYS is that the denominator is **always the grand total of all plays in the entire report**.
    - *Example:* In the cell for OFF PERS = 11 and g R/P = P, the value is 33 plays and 28.2%. This means 33 / (Grand Total of 117 Plays) = 28.2%.

2. **The "Subgroup View"**

   This report is structured as a **breakdown** or **tendency report**.

- **Structure:** The primary variable of interest is placed on the rows (OFF PERS). The columns are not another variable, but rather a set of *specific, pre-defined metric sets* like (# RUN, % RUN), (# PASS, % PASS), etc.
- **Percentage Logic (Within-Group):** The system's rule for a specific metric like % RUN is that the denominator is **the total number of plays for that specific row only**.
    - *Example:* In the row for OFF PERS = 11, the total plays are 52. The % RUN is 26.9%. This means (14 Runs) / (52 Total Plays for 11 Personnel) = 26.9%.

The question "When I do X, what do I do next?" is the most important one for almost all coaching purposes, so the **"Subgroup View" is significantly more valuable and intuitive.**

Feature	"Global View" (Crosstab)	"Subgroup View" (Breakdown)
Structure	Variable vs. Variable Matrix	Variable Breakdown by Metric
Percentage Denominator	Grand Total of All Plays	Subtotal of Plays in that Row
Answer	"What is the frequency/contribution to the whole?"	"What is the tendency/distribution within this group?"

Report 1

Have the # of plays and % for R/P/O/A by PSN

- Group by OFF PERS and g R/P. Count the occurrences for each unique combination.
- Calculate the percentage of each play type (R/P/O/A) within each OFF PERS group.

Report 2 | Report 8 - STR | Report 8 - WK | Report 1 (Global) | Report 1 (subgroups) *

#↓	10		11		12		12T		20		21	
g R/P	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS
TOTALS:	26	22.2%	52	44.4%	4	3.4%	1	0.9%	18	15.4%	16	13.7%
P	20	17.1%	33	28.2%	1	0.9%			14	12.0%	7	6.0%
R	6	5.1%	14	12.0%	3	2.6%	1	0.9%	3	2.6%	9	7.7%
O			5	4.3%								
D									1	0.9%		

Report 2 | Report 8 - STR | Report 8 - WK | Report 1 (Global) | Report 1 (subgroups) *

OFF PERS #↓	# RUN	% RUN	# PASS	% PASS	# OPTION	% OPTION	# PLAYS	% PLAYS
TOTALS:	36	30.8%	75	64.1%	5	4.3%	117	100.0%
11	14	26.9%	33	63.5%	5	9.6%	52	44.4%
10	6	23.1%	20	76.9%			26	22.2%
20	3	16.7%	14	77.8%			18	15.4%
21	9	56.3%	7	43.8%			16	13.7%
12	3	75.0%	1	25.0%			4	3.4%
12T	1	100.0%					1	0.9%

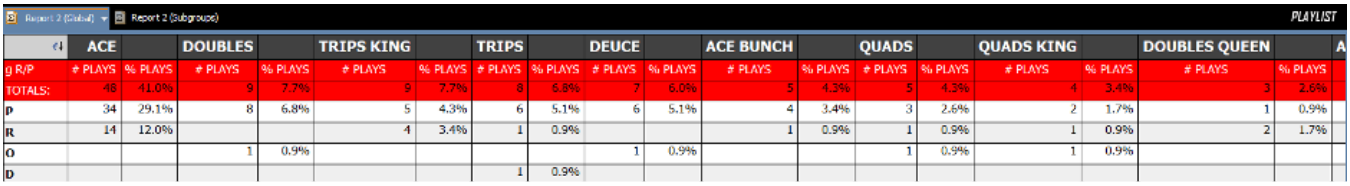
Q:
When I substitute a specific personnel package onto the field (like our '11' or '12' package), what is our primary tendency? Are we telegraphing 'run' or 'pass' just by the players we have out there?

Ans: It reveals our offensive identity and balance for each personnel group. It shows whether we are a run-heavy, pass-heavy, or balanced team out of a specific look, which helps us identify and break our own predictable tendencies.	
Global View	This report answers, "What percentage of our entire game plan consists of passes from 11 personnel?" It shows the overall frequency of each personnel/play-type combination relative to the 117 total plays. This is useful for seeing the "big picture" of your offensive strategy.
Subgroup View	This report answers, "When our 11 personnel is on the field, what is our tendency to run versus pass?" It measures the internal breakdown for each personnel group. This view is crucial for self-scouting to see if a specific personnel package is telegraphing your intentions to the defense.

Report 2

Have the # of plays and % for R/P/O/A by FORM

- Group by FORM and g R/P. Count the occurrences for each unique combination.
- Calculate the percentage of each play type (R/P/O/A) within each FORM group.



Report 2 (Global) | Report 2 (Subgroups) | PLAYLIST

	ACE		DOUBLES		TRIPS KING		TRIPS		DEUCE		ACE BUNCH		QUADS		QUADS KING		DOUBLES QUEEN		A
g R/P	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	
TOTALS:	48	41.0%	9	7.7%	9	7.7%	8	6.8%	7	6.0%	5	4.3%	5	4.3%	4	3.4%	3	2.6%	
P	34	29.1%	8	6.8%	5	4.3%	6	5.1%	6	5.1%	4	3.4%	3	2.6%	2	1.7%	1	0.9%	
R	14	12.0%			4	3.4%	1	0.9%			1	0.9%	1	0.9%	1	0.9%	2	1.7%	
O			1	0.9%					1	0.9%			1	0.9%	1	0.9%			
D							1	0.9%											

Report 2 (Global) | Report 2 (Subgroups)

FORM #↓	# PASS	% PASS	# RUN	% RUN	# OPTION	% OPTION	# PLAYS	% PLAYS
TOTALS:	75	64.1%	36	30.8%	5	4.3%	117	100.0%
ACE	34	70.8%	14	29.2%			48	41.0%
DOUBLES	8	88.9%			1	11.1%	9	7.7%
TRIPS KING	5	55.6%	4	44.4%			9	7.7%
TRIPS	6	75.0%	1	12.5%			8	6.8%
DEUCE	6	85.7%			1	14.3%	7	6.0%
ACE BUNCH	4	80.0%	1	20.0%			5	4.3%
QUADS	3	60.0%	1	20.0%	1	20.0%	5	4.3%
QUADS KING	2	50.0%	1	25.0%	1	25.0%	4	3.4%
ACE QUEEN	1	33.3%	1	33.3%	1	33.3%	3	2.6%
DEUCE KING	1	33.3%	2	66.7%			3	2.6%
DOUBLES QUEEN	1	33.3%	2	66.7%			3	2.6%
ACE KING	1	50.0%	1	50.0%			2	1.7%
ACE SPLIT	2	100.0%					2	1.7%
DOUBLES KING	1	50.0%	1	50.0%			2	1.7%
TRIO			2	100.0%			2	1.7%
TRIPS NUB			2	100.0%			2	1.7%
ACE BUNCH WIDE			1	100.0%			1	0.9%
ACE THICK			1	100.0%			1	0.9%
SNEAK			1	100.0%			1	0.9%

Q:
From a specific offensive formation, what is our most likely play type? Are there formations we only run from, or only pass from?

Ans:
It reveals the strategic purpose of each formation in our playbook. It helps us understand which formations are our 'run' formations, which are our 'pass' formations, and which are truly balanced, allowing us to game plan more effectively.

Global View	This report answers, "**What percentage of our entire game plan is a pass from the ACE formation?**" It shows the overall frequency of each formation/play-type combination relative to the 117 total plays. This is useful for seeing which formations are the true cornerstones of your offense.
Subgroup View	This report answers, "**When we line up in the ACE formation, what is our tendency to run versus pass?**" It measures the internal breakdown for each formation. This view is crucial for self-scouting to see if certain alignments are telegraphing your play call to the defense.

Report 3

Have the # of plays and % for R/P/O/A by D and Dist and Field zones not including the 1st plays of series

1) Filter only for the first play of each series (Filtered by "g SER# 1")
2) Create categorical columns for Field Zone and Down & Distance.
3) Group by FZONE, D and D, and g R/P. Count the occurrences for each unique combination.
4) Calculate the percentage of each play type (R/P/O/A) within its specific situation

Report 3

FZONE	D&D	# RUN	% RUN	# PASS	% PASS	# OPTION	% OPTION	# PLAYS	% PLAYS
TOTALS:		19	23.2%	58	70.7%	2	2.4%	82	100.0%
Middle (-30 to +26)		10	19.6%	37	72.5%	2	3.9%	51	62.2%
	1 & 10	5	26.3%	13	68.4%	1	5.3%	19	37.3%
	2 & 4-6	1	9.1%	9	81.8%	1	9.1%	11	21.6%
	2 & 7-10			9	90.0%			10	19.6%
	2 & 11+	1	20.0%	3	60.0%			5	9.8%
	2 & 1-3	2	50.0%	2	50.0%			4	7.8%
	3 & 2+	1	50.0%	1	50.0%			2	3.9%
Green Zone (+25 to +16)		5	38.5%	7	53.8%			13	15.9%
	1 & 10	4	57.1%	2	28.6%			7	53.8%
	2 & 4-6			3	100.0%			3	23.1%
	2 & 7-10			2	100.0%			2	15.4%
	2 & 1-3	1	100.0%					1	7.7%
Red Zone (+15 to +6)		2	22.2%	7	77.8%			9	11.0%
	2 & 4-6	1	33.3%	2	66.7%			3	33.3%
	1 & 10			2	100.0%			2	22.2%
	2 & 7-10			2	100.0%			2	22.2%
	2 & 1-3	1	100.0%					1	11.1%
	3 & 2+			1	100.0%			1	11.1%
Coming Out (-11 to -29)		2	33.3%	4	66.7%			6	7.3%
	2 & 4-6	1	50.0%	1	50.0%			2	33.3%
	2 & 11+			2	100.0%			2	33.3%
	2 & 1-3	1	100.0%					1	16.7%
	2 & 7-10			1	100.0%			1	16.7%
Backed Up (-1 to -10)				2	100.0%			2	2.4%
	2 & 4-6			1	100.0%			1	50.0%
	2 & 11+			1	100.0%			1	50.0%
Goal Line (+5 to +1)				1	100.0%			1	1.2%
	2 & 1-3			1	100.0%			1	100.0%

Q:

When we are in a high-pressure situation like 3rd & 8, or a favorable one like 2nd & 2, how do our play calls change depending on where we are on the field? What are our tendencies on the 'money downs'?

Ans:
It reveals our play-calling tendencies under pressure or in advantageous situations. It shows what we do on the critical 'money downs' (like 3rd down) and helps identify if we are too predictable when trying to extend a drive.

Report 4

Have the # of plays and % for R/P/O/A by D and Dist and Field zones only for the first plays of series

1) Filter only for the first play of each series. (Filtered by g SER# = 1)
2) Create categorical columns for Field Zone and Down & Distance.
3) Group by FZONE, D and D, and g R/P. Count the occurrences for each unique combination.

4) Calculate the percentage of each play type (R/P/O/A) within its specific situation.

Report 4

FZONE	D&D	# RUN	% RUN	# PASS	% PASS	# OPTION	% OPTION	# PLAYS	% PLAYS
TOTALS:		13	48.1%	12	44.4%	1	3.7%	27	100.0%
Middle (-30 to +26)		4	30.8%	8	61.5%	1	7.7%	13	48.1%
	P+10	4	30.8%	8	61.5%	1	7.7%	13	100.0%
Coming Out (-11 to -29)		6	66.7%	2	22.2%			9	33.3%
	P+10	6	66.7%	2	22.2%			9	100.0%
Green Zone (+25 to +16)		2	66.7%	1	33.3%			3	11.1%
	P+10	2	66.7%	1	33.3%			3	100.0%
Backed Up (-1 to -10)		1	50.0%	1	50.0%			2	7.4%
	P+10	1	100.0%					1	50.0%
	2 & 11+			1	100.0%			1	50.0%

Q:
On the first play of a drive, how does our game plan change based on where we are on the field? Are we conservative when backed up and aggressive at midfield?

Ans:
It reveals our 'drive starter' philosophy. It shows how we try to 'get on schedule' at the beginning of a possession and whether our opening play calls are predictable based on field position

Report 5

Have the # of plays and % for O box Run plays by Psn qualified for Direction (L/R)

1) Filter only RUN plays (Filtered by g R/P = “R”)
2) Group by OFF PERS, O BOX, L/R. Count the occurrences for each unique combination.

3) Calculate percentage within each OFF PERS group.

Report 5 (Global) | Report 5 (Subgroups)

		10		11		12		12T		20		21	
O BOX	L / R	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS
TOTALS:		6	16.7%	14	38.9%	3	8.3%	1	2.8%	3	8.3%	9	25.0%
DUO				4	11.1%	3	8.3%					4	11.1%
	L			1	9.1%	1	9.1%					1	9.1%
	R			3	27.3%	2	18.2%					3	27.3%
GF CT		1	2.8%	3	8.3%					3	8.3%	2	5.6%
	L			3	33.3%					1	11.1%		
	R	1	11.1%							2	22.2%	2	22.2%
IZ READ		3	8.3%										
	L	3	100.0%										
TRIP O												3	8.3%
	L											1	33.3%
	R											2	66.7%
DART		2	5.6%										
	R	2	100.0%										
ISO				2	5.6%								
	L			1	50.0%								
	R			1	50.0%								
IZ BRICK				2	5.6%								
	R			2	100.0%								
FLY SWP				1	2.8%								
	R			1	100.0%								
IZ CUT				1	2.8%								
	R			1	100.0%								
OZ				1	2.8%								
	L			1	100.0%								
SNEAK								1	2.8%				
	R							1	100.0%				

Report 5 (Global) | Report 5 (Subgroups)

		L		R	
OFF PERS	O BOX	# RUN	% RUN	# RUN	% RUN
TOTALS:		13	36.1%	23	63.9%
10		3	50.0%	3	50.0%
	DART			2	100.0%
	GF CT			1	100.0%
	IZ READ	3	100.0%		
11		6	42.9%	8	57.1%
	DUO	1	25.0%	3	75.0%
	FLY SWP			1	100.0%
	GF CT	3	100.0%		
	ISO	1	50.0%	1	50.0%
	IZ BRICK			2	100.0%
	IZ CUT			1	100.0%
	OZ	1	100.0%		
12		1	33.3%	2	66.7%
	DUO	1	33.3%	2	66.7%
12T				1	100.0%
	SNEAK			1	100.0%
20		1	33.3%	2	66.7%
	GF CT	1	33.3%	2	66.7%
21		2	22.2%	7	77.8%
	DUO	1	25.0%	3	75.0%
	GF CT			2	100.0%
	TRIP O	1	33.3%	2	66.7%

Q:
When I have a specific personnel package on the field and the defense presents a certain number of players in the box, what is the directional tendency (Left vs. Right) of my favorite run plays?

Ans:
It reveals HOW we physically run a specific play. It shows the physical direction (Left/Right) of our run concepts, which is crucial for analyzing blocking schemes and identifying if we have a directional 'tell' that a smart defense could exploit.

Global View	This report answers, "What percentage of our entire run game is a DUO play from 11 personnel?" It shows the overall frequency of each specific run play against the grand total of all runs. This is useful for seeing which plays make up the bulk of your running attack.
Subgroup View	This report answers, "When we call the DUO play with our 11 personnel, what is our tendency to run Left versus Right?" It measures the directional breakdown for each specific play concept. This view is crucial for identifying predictable directional 'tells' that a defense can use to align their players correctly before the snap.

Report 6

Have the # of plays and % for O box Run plays Psn qualified for Direction (Fld/Bnd)

1) Filter only RUN plays (Filtered by g R/P = “R”)
2) Group by OFF PERS, O BOX, FLD / BND. Count the occurrences for each unique combination.
3) Calculate percentage within each OFF PERS group

Report 6 (Global) | Report 6 (Subgroups)

		10		11		12		12T		20		21	
O BOX	FLD / BND	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS
TOTALS:		6	16.7%	14	38.9%	3	8.3%	1	2.8%	3	8.3%	9	25.0%
DART		2	5.6%										
	FLD	2	100.0%										
DUO				4	11.1%	3	8.3%					4	11.1%
	BND			2	18.2%	2	18.2%					2	18.2%
	FLD			2	18.2%	1	9.1%					2	18.2%
FLY SWP				1	2.8%								
	FLD			1	100.0%								
GF CT		1	2.8%	3	8.3%					3	8.3%	2	5.6%
	BND	1	11.1%	2	22.2%					2	22.2%		
	FLD			1	11.1%					1	11.1%	2	22.2%
ISO				2	5.6%								
	BND			2	100.0%								
IZ BRICK				2	5.6%								
	BND			2	100.0%								
IZ CUT				1	2.8%								
	FLD			1	100.0%								
IZ READ		3	8.3%										
	BND	3	100.0%										
OZ				1	2.8%								
	BND			1	100.0%								
SNEAK								1	2.8%				
	BND							1	100.0%				
TRIP O												3	8.3%
	BND											2	66.7%
	FLD											1	33.3%

Report 6 (Global) | Report 6 (Subgroups)

		BND		FLD	
OFF PERS	O BOX	# RUN	% RUN	# RUN	% RUN
TOTALS:		22	61.1%	14	38.9%
10		4	66.7%	2	33.3%
	DART			2	100.0%
	GF CT	1	100.0%		
	IZ READ	3	100.0%		
11		9	64.3%	5	35.7%
	DUO	2	50.0%	2	50.0%
	FLY SWP			1	100.0%
	GF CT	2	66.7%	1	33.3%
	ISO	2	100.0%		
	IZ BRICK	2	100.0%		
	IZ CUT			1	100.0%
	OZ	1	100.0%		
12		2	66.7%	1	33.3%
	DUO	2	66.7%	1	33.3%
12T		1	100.0%		
	SNEAK	1	100.0%		
20		2	66.7%	1	33.3%
	GF CT	2	66.7%	1	33.3%
21		4	44.4%	5	55.6%
	DUO	2	50.0%	2	50.0%
	GF CT			2	100.0%
	TRIP O	2	66.7%	1	33.3%

Q:
Are we strategically attacking the wide side of the field (Field) or the short, hash-mark side (Boundary)? Does our run game strategy change based on field geography?

Ans:
It reveals WHERE we strategically attack on the field with our run game. It shows if we prefer to run to the wide 'Field' side or the short 'Boundary' side, helping us understand how we are trying to leverage space and angles against the defense

Global View	This report answers, "What percentage of our total run game uses each O Box concept toward the Boundary (BND) or Field (FLD)?" It shows how often each run concept is used across all offensive personnel, broken down by run direction relative to field alignment. This view is helpful to understand whether certain plays tend to favor BND vs. FLD across your entire offensive package.
Subgroup View	This report answers, "When we run a specific concept like DUO from 11 personnel, what is our directional tendency—Boundary or Field?" It gives a directional breakdown (BND vs. FLD) for each run concept within a personnel group. This helps assess whether any play–personnel combo may be telegraphing directionality and thus could be exploited by defenses pre-snap

Report 7

Have the # of plays and % for O box pass plays Psn and Direction (Fld/Bnd)

1) Filter only PASS plays (Filtered by g R/P = “P”)
2) Group by OFF PERS, O BOX, FLD / BND. Count the occurrences for each unique combination.

3) Calculate percentage within each OFF PERS group

Report 7 (Global) | Report 7 (Subgroups)

		10		11		12		20		21	
O BOX	FLD / BND	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS
TOTALS:		14	24.1%	24	41.4%	1	1.7%	13	22.4%	6	10.3%
51				5	8.6%						
	BND			3	60.0%						
	FLD			2	40.0%						
61		3	5.2%	2	3.4%			1	1.7%		
	BND	1	16.7%	1	16.7%						
	FLD	2	33.3%	1	16.7%			1	16.7%		
62		2	3.4%	3	5.2%			2	3.4%		
	BND			3	42.9%			1	14.3%		
	FLD	2	28.6%					1	14.3%		
64								1	1.7%		
	FLD							1	100.0%		
66		7	12.1%					1	1.7%		
	BND							1	12.5%		
	FLD	7	87.5%								
67		2	3.4%					3	5.2%	2	3.4%
	BND							2	28.6%	2	28.6%
	FLD	2	28.6%					1	14.3%		
71				1	1.7%						
	FLD			1	100.0%						
72				5	8.6%			1	1.7%		
	BND			1	16.7%						
	FLD			4	66.7%			1	16.7%		
75								2	3.4%		
	FLD							2	100.0%		
76						1	1.7%				
	FLD					1	100.0%				
77				1	1.7%					1	1.7%
	BND									1	50.0%
	FLD			1	50.0%						
88				1	1.7%					1	1.7%
	FLD			1	50.0%					1	50.0%
89										1	1.7%
	FLD									1	100.0%

Report 7 (Global) | Report 7 (Subgroups)

		BND		FLD	
OFF PERS	O BOX	# PASS	% PASS	# PASS	% PASS
TOTALS:		21	36.2%	37	63.8%
11		13	54.2%	11	45.8%
	51	3	60.0%	2	40.0%
	61	1	50.0%	1	50.0%
	62	3	100.0%		
	71			1	100.0%
	72	1	20.0%	4	80.0%
	77			1	100.0%
	88			1	100.0%
	DART	1	100.0%		
	GF CT	2	66.7%	1	33.3%
	IZ BRICK	1	100.0%		
	IZ CUT	1	100.0%		
10		1	7.1%	13	92.9%
	61	1	33.3%	2	66.7%
	62			2	100.0%
	66			7	100.0%
	67			2	100.0%
20		4	30.8%	9	69.2%
	61			1	100.0%
	62	1	50.0%	1	50.0%
	64			1	100.0%
	66	1	100.0%		
	67	2	66.7%	1	33.3%
	72			1	100.0%
	75			2	100.0%
	DUO			1	100.0%
	OZ CUT			1	100.0%
21		3	50.0%	3	50.0%
	67	2	100.0%		
	77	1	100.0%		
	88			1	100.0%
	89			1	100.0%
	REVERSE			1	100.0%

Q:
When we decide to pass with a certain personnel group against a specific defensive look, are we favoring the wide side of the field or the short side? Where are our passing concepts designed to go?

Ans:
It reveals the design and strategy of our **pass** game. It shows whether our pass concepts are designed to stretch the defense horizontally by attacking the wide side of the field or to use quick, high-percentage throws to the boundary.

Global View	This report answers, “What percentage of our entire pass game comes from each specific offensive formation (O Box) by personnel group?” It provides a top-down view of how often different pass sets are used, organized by personnel group.
Subgroup View	This report answers, “When we pass from a given formation (O Box), what is our directional tendency: Field vs. Boundary?” It breaks down the directional split (FLD/BND) for each O Box within every personnel group.

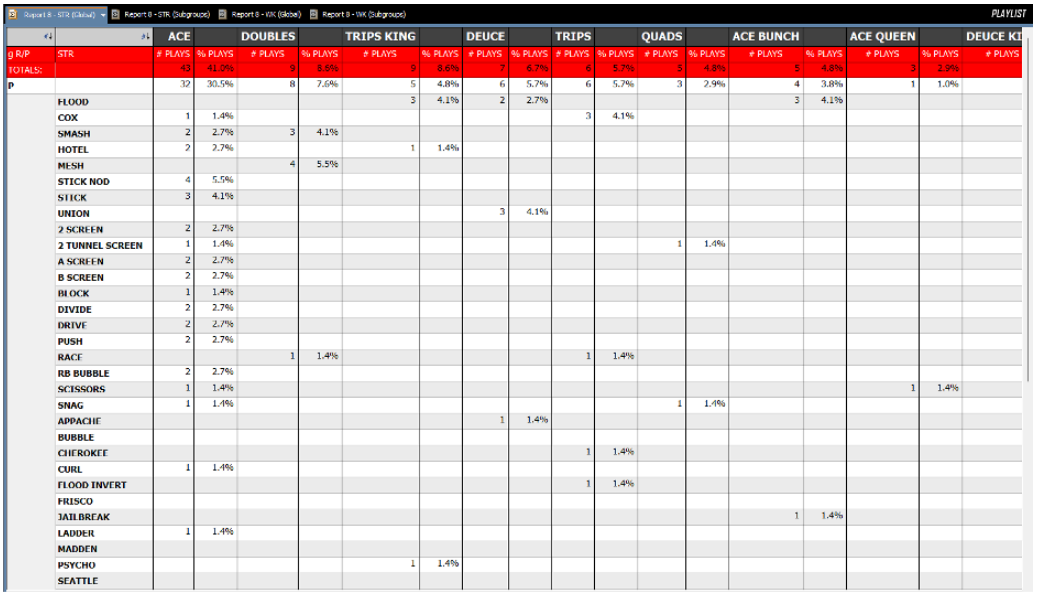
Report 8

Have the # of plays and % for Strong and Weak Concepts grouped by formations and P/O/A

1) Group by FORM and G R/P. Within each group, get the count of plays per STR and per WK
2) Calculate percentage for each FORM × G R/P combination
    a) % of each STR = (count of this STR) / (total count)

    b) % of each WK concept = (count of this WK) / (total count)

STR (Strong Side)

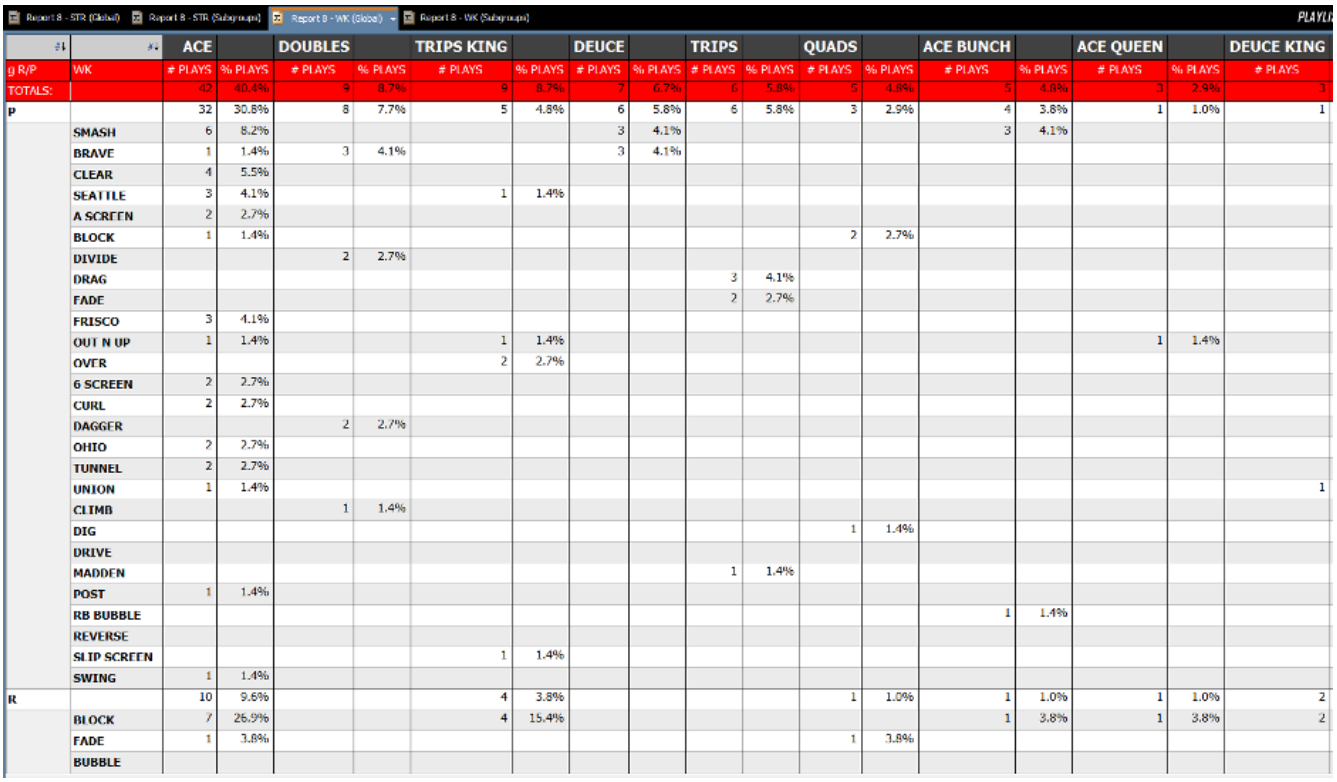


g R/P	STR	ACE		DOUBLES		TRIPS KING		DEUCE		TRIPS		QUADS		ACE BUNCH		ACE QUEEN		DEUCE KI
		# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS	% PLAYS	# PLAYS
TOTALS:		43	41.0%	9	8.6%	9	8.6%	7	6.7%	6	5.7%	5	4.8%	5	4.8%	3	2.9%	
P		32	30.5%	8	7.6%	5	4.8%	6	5.7%	6	5.7%	3	2.9%	4	3.8%	1	1.0%	
	FLOOD					3	4.1%	2	2.7%					3	4.1%			
	COX	1	1.4%							3	4.1%							
	SMASH	2	2.7%	3	4.1%													
	HOTEL	2	2.7%			1	1.4%											
	MESH			4	5.5%													
	STICK NOD	4	5.5%															
	STICK	3	4.1%															
	UNION							3	4.1%									
	2 SCREEN	2	2.7%															
	2 TUNNEL SCREEN	1	1.4%									1	1.4%					
	A SCREEN	2	2.7%															
	B SCREEN	2	2.7%															
	BLOCK	1	1.4%															
	DIVIDE	2	2.7%															
	DRIVE	2	2.7%															
	PUSH	2	2.7%															
	RACE			1	1.4%					1	1.4%							
	RB BUBBLE	2	2.7%															
	SCISSORS	1	1.4%													1	1.4%	
	SNAG	1	1.4%									1	1.4%					
	APPACHE							1	1.4%									
	BUBBLE																	
	CHEROKEE									1	1.4%							
	CURL	1	1.4%															
	FLOOD INVERT									1	1.4%							
	FRISCO																	
	JAILBREAK													1	1.4%			
	LADDER	1	1.4%															
	MADDEN																	
	PSYCHO					1	1.4%											
	SEATTLE																	

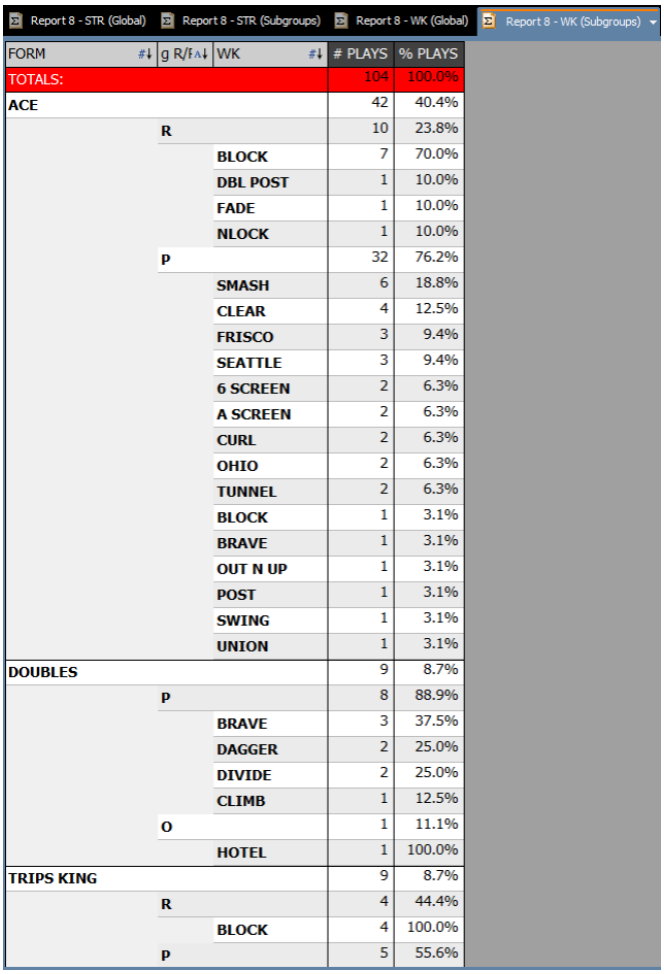
Report 8 - STR (Global) | Report 8 - STR (Subgroups) | Report 8 - WK (Global) | Report 8 - WK (Subgroups)

FORM	g R/P	STR	# PLAYS	% PLAYS
TOTALS:			105	100.0%
ACE			43	41.0%
	R		11	25.6%
		BLOCK	6	54.5%
		BRAVE	1	9.1%
		FLAT	1	9.1%
		FLOOD	1	9.1%
		HOTEL	1	9.1%
		SEATTLE	1	9.1%
	P		32	74.4%
		STICK NOD	4	12.5%
		STICK	3	9.4%
		2 SCREEN	2	6.3%
		A SCREEN	2	6.3%
		B SCREEN	2	6.3%
		DIVIDE	2	6.3%
		DRIVE	2	6.3%
		HOTEL	2	6.3%
		PUSH	2	6.3%
		RB BUBBLE	2	6.3%
		SMASH	2	6.3%
		2 TUNNEL SCREEN	1	3.1%
		BLOCK	1	3.1%
		COX	1	3.1%
		CURL	1	3.1%
		LADDER	1	3.1%
		SCISSORS	1	3.1%
		SNAG	1	3.1%
DOUBLES			9	8.6%
	P		8	88.9%
		MESH	4	50.0%
		SMASH	3	37.5%
		RACE	1	12.5%
	O		1	11.1%
		BLOCK	1	100.0%

WK (Weak Side)



g R/P	WK	ACE # PLAYS	ACE % PLAYS	DOUBLES # PLAYS	DOUBLES % PLAYS	TRIPS KING # PLAYS	TRIPS KING % PLAYS	DEUCE # PLAYS	DEUCE % PLAYS	TRIPS # PLAYS	TRIPS % PLAYS	QUADS # PLAYS	QUADS % PLAYS	ACE BUNCH # PLAYS	ACE BUNCH % PLAYS	ACE QUEEN # PLAYS	ACE QUEEN % PLAYS	DEUCE KING # PLAYS
TOTALS:		42	40.4%	9	8.7%	9	8.7%	7	6.7%	6	5.8%	5	4.8%	5	4.8%	3	2.9%	3
P		32	30.8%	8	7.7%	5	4.8%	6	5.8%	6	5.8%	3	2.9%	4	3.8%	1	1.0%	1
	SMASH	6	8.2%					3	4.1%					3	4.1%			
	BRAVE	1	1.4%	3	4.1%			3	4.1%									
	CLEAR	4	5.5%															
	SEATTLE	3	4.1%			1	1.4%											
	A SCREEN	2	2.7%															
	BLOCK	1	1.4%									2	2.7%					
	DIVIDE			2	2.7%													
	DRAG									3	4.1%							
	FADE									2	2.7%							
	FRISCO	3	4.1%															
	OUT N UP	1	1.4%			1	1.4%									1	1.4%	
	OVER					2	2.7%											
	6 SCREEN	2	2.7%															
	CURL	2	2.7%															
	DAGGER			2	2.7%													
	OHIO	2	2.7%															
	TUNNEL	2	2.7%															
	UNION	1	1.4%															1
	CLIMB			1	1.4%													
	DIG											1	1.4%					
	DRIVE																	
	MADDEN									1	1.4%							
	POST	1	1.4%															
	RB BUBBLE													1	1.4%			
	REVERSE																	
	SLIP SCREEN					1	1.4%											
	SWING	1	1.4%															
R		10	9.6%			4	3.8%					1	1.0%	1	1.0%	1	1.0%	2
	BLOCK	7	26.9%			4	15.4%							1	3.8%	1	3.8%	2
	FADE	1	3.8%									1	3.8%					
	BUBBLE																	



FORM	g R/P	WK	# PLAYS	% PLAYS
TOTALS:			104	100.0%
ACE			42	40.4%
	R		10	23.8%
		BLOCK	7	70.0%
		DBL POST	1	10.0%
		FADE	1	10.0%
		NLOCK	1	10.0%
	P		32	76.2%
		SMASH	6	18.8%
		CLEAR	4	12.5%
		FRISCO	3	9.4%
		SEATTLE	3	9.4%
		6 SCREEN	2	6.3%
		A SCREEN	2	6.3%
		CURL	2	6.3%
		OHIO	2	6.3%
		TUNNEL	2	6.3%
		BLOCK	1	3.1%
		BRAVE	1	3.1%
		OUT N UP	1	3.1%
		POST	1	3.1%
		SWING	1	3.1%
		UNION	1	3.1%
DOUBLES			9	8.7%
	P		8	88.9%
		BRAVE	3	37.5%
		DAGGER	2	25.0%
		DIVIDE	2	25.0%
		CLIMB	1	12.5%
	O		1	11.1%
		HOTEL	1	100.0%
TRIPS KING			9	8.7%
	R		4	44.4%
		BLOCK	4	100.0%
	P		5	55.6%

Q: What is the true identity of our playbook? From our core formations, what are the handful of tactical concepts that we rely on the most for our run, pass, and play-action game?	
Ans: It reveals the core concepts and building blocks of our entire offense. By unifying the Strong and Weak side concepts, it gives a true frequency analysis of what plays are actually in our game plan, versus what's just on paper.	
Global View	This report answers, “What percentage of our overall pass and run game is made up of each route concept across all formations?” It shows the frequency of each concept used within each formation, broken down by R (Run) and P (Pass). This is useful for identifying your most common plays and how your strategy varies by formation groupings.
Subgroup View	This report answers, “When we are in a specific formation, how are our plays distributed between the Run and Pass game, and what are the specific route concepts we favor in each?” It breaks down each formation by play type and then by concept usage, giving insight into whether your tendencies shift within a formation. This view is key for spotting imbalances or tendencies in concept selection by formation that may create calls for a defense.

Report 9 (Excel)

Filtered the dataset to display the rows that correspond to “Complete”, “Incomplete” and “Interception” in the RESULT column.

				Values	
QB	REC	ZONE THROWN	O COVERAGE	Count of RESULT	Sum of Is_Complete
12	1	BND DEEP	PURPLE	1	1
		BND INT	ROVER	1	1
		FLD SHT	DODGE	1	1
			FORD	1	0
		INT SHT	DODGE	1	0
	2	FLD DEEP	SILVER	1	0
		FLD INT	SWIPE HONDA	1	1
		FLD SHT	HONDA	1	1
			IRONMAN	1	1
		MID INT	FORD	1	0
			IRONMAN	1	1
	3	FLD SHT	IRONMAN	1	1
	4	BND SHT	DODGE	2	2
			HONDA	1	1
		FLD SHT	ROLL	1	1
		MID SHT	SWIPE HONDA	1	1
	15	FLD SHT	HONDA	1	1
			READ	1	0
			SWIPE HONDA	1	1
		MID SHT	IRONMAN	1	1
	80	FLD INT	FORD	1	1
	84	BND DEEP	HONDA	1	0
		FLD INT	HONDA	1	0
		MID INT	FORD	1	0
	88	BND DEEP	HONDA	1	0
		BND SHT	FORD	1	1
			IRONMAN	1	1
		FLD SHT	HONDA	1	1
		MID INT	SWIPE HONDA	1	1
		MID SHT	HONDA	1	0
BLOCK	12	(blank)	PIRATE	1	1
FISH	12	1	PIRATE	1	1
			WAVE	1	1
RACE	12	3	(blank)	1	1
		4	TUFF	1	0
SNAG	POST	12	MUG	1	1
UNION	12	80	SHARK	1	1
Grand Total				38	27

This pivot table presents each **pass attempt** made by the quarterback, broken down by:

- **Receiver (REC)** – the player who was targeted
- **Zone Thrown (ZONE THROWN)** – the area of the field the ball was aimed at (e.g., BND DEEP, FLD SHT, MID INT)
- **Defensive Coverage (O COVERAGE)** – the coverage schemes the defense was using (e.g., HONDA, IRONMAN, DODGE)

The table contains two main values:

- **Count of RESULT** – how many passes were attempted for each combination of receiver, zone, and coverage.
- **Sum of Is_Complete** – how many of those passes were actually completed (1 = completed, 0 = incomplete).

How to Interpret This Table

To interpret this table:

1. **Each row** represents a unique combination of QB, receiver, zone, and coverage.
2. Read across each row to see:
    a. How many times that specific combination was used (Count of RESULT)
    b. How many of those throws were successful (Sum of Is_Complete)
3. A combination with **Count of RESULT = 2** and **Sum of Is_Complete = 2** indicates a **100% completion rate**.
4. A combination with **Count = 1** and **Sum = 0** shows a **failed pass attempt**.
5. Patterns across rows can help identify **which zones and matchups were most efficient**.

This allows you to evaluate the effectiveness of specific passing decisions under different defensive conditions — such as which receivers excel against certain coverages, or which zones are high-risk.

The first row of the table has the following interpretation:

Quarterback #12 threw **1 pass** to **Receiver #1**. The throw was targeted to the **BND** zone. The defensive coverage on this play was **PURPLE**. The **pass attempt count was 1**, and it was **completed**.

Insights from the Table

Most successful combinations:

- **Receiver 4** in the **BND SHT** zone against **DODGE** had **2 attempts and 2 completions**, a 100% rate.
- **Receiver 1** had consistent completions in **FLD SHT** and **BND INT** zones against **DODGE**, **ROVER**, and **IRONMAN** coverages.

Least successful throws:

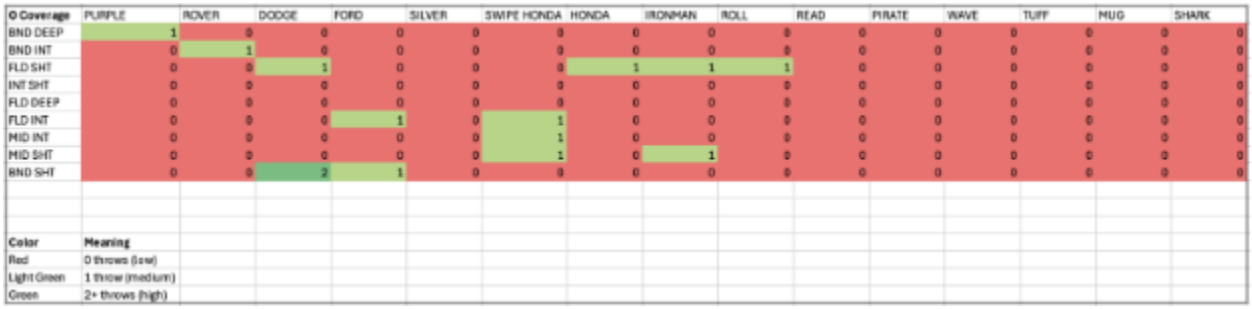
- Several incomplete passes were observed in **INT SHT**, **MID INT**, and **BND DEEP** zones, particularly when facing **FORD** and **HONDA**.

Coverage impact:

- **IRONMAN** and **HONDA** were faced frequently, but completion rates varied. For example, **IRONMAN** saw multiple completions from short throws, while **HONDA** had both successes and failed attempts — suggesting variability based on zone and receiver.

Zone usage:

- Most passes were concentrated in short and intermediate zones like **FLD SHT**, **MID SHT**, and **FLD INT**, which also had the highest success rates.
- Deep zones like **BND DEEP** and **INT SHT** were used less and showed lower completion efficiency.



Coverage	PURPLE	ROVER	DODGE	FORD	SILVER	SWIPE HONDA	HONDA	IRONMAN	ROLL	READ	PIRATE	WAVE	TUFF	MUG	SHARK
BND DEEP	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BND INT	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
FLD SHT	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0
INT SHT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FLD DEEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FLD INT	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
MID INT	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
MID SHT	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
BND SHT	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0

Color	Meaning
Red	0 throws (low)
Light Green	1 throw (medium)
Green	2+ throws (high)

The heatmap created for Report 9 provides a visual representation of the quarterback's passing tendencies based on two key variables: the defensive coverage faced (listed along the rows) and the intended receiver (listed across the columns). Each cell indicates how many times a pass was attempted to a specific receiver under a specific coverage, with a color scale to reflect volume. Red signifies zero throws (low usage), light green indicates one throw (moderate usage), and dark green highlights two or more throws (high usage).

From the visualization, it's clear that the quarterback targeted only a handful of receiver–coverage combinations during the game. Most cells remain red, signaling either conservative play-calling or limited variation in coverage during the sample. The most notable pairing is DODGE against BND SHT coverage, which stands out in green, showing it was targeted twice—more than any other combination on the grid. This suggests a preferred or successful matchup. Additionally, there are several instances of one-off throws (light green), such as PURPLE vs. BND DEEP, ROVER vs. BND INT, and FORD vs. FLD SHT, among others. These may reflect situational or test plays based on coverage looks or offensive alignment.

The large number of red zones implies missed opportunities or unexplored matchups, particularly against coverages like PIRATE, WAVE, TUFF, MUG, and SHARK, which saw no targets to any receiver. This insight can help coaches and analysts identify tendencies, gaps in play selection, and potential areas to diversify the passing strategy in future games. The heatmap ultimately serves as a

quick and intuitive way to assess how coverage schemes influenced throwing decisions and receiver usage.

Report 10 (Excel)

For Report 10, the first step involved preparing and refining the raw dataset to enable more meaningful analysis of rushing plays. We began by filtering the dataset to include only R plays, ensuring consistency in how play types were labeled (e.g., all rushes were marked as “R” under “g R/P”). We then introduced a new variable called “Gain Bucket,” which grouped each rushing attempt into one of four yardage categories based on the value in the “g GAIN” column: 0 yards or less, 1–4 yards, 5–10 yards, and 11+ yards. This allowed us to better segment and analyze the efficiency and productivity of different rushing plays.

Gain Bucket	Yardage Range
0	0 or less
1	1 – 4
2	11+
3	5 – 10

Then, to better understand the contribution of the offensive line versus the running back, we engineered two new metrics: **OL Yards** and **RB Yards**.

The **OL Yards** (offensive line yards) were calculated using the formula:

```
=IF(OR(N2="Penalty", N2="TD", N2=""), "", IF(M2 <= 0, M2, MIN(5, M2)))
```

This logic assumes that for any rush play that results in a gain of 5 yards or fewer, those yards are attributed to the offensive line. If the total gain exceeds 5 yards, only the first 5 are credited to the offensive line. Negative plays (where `g GAIN` ≤ 0) are attributed entirely to the offensive line as well. Plays marked as "Penalty", "TD", or blank in the `RESULT` column are excluded from the calculation.

The **RB Yards** (running back yards) were calculated using:

```
=IF(P2="", "", N2 - P2)
```

This subtracts the OL Yards from the total yards gained (column g GAIN) to isolate the yardage credited to the running back. If the OL Yards cell is blank, the formula also returns a blank, ensuring only valid plays are considered.

Row Labels	Num_Runs	Total_Yards	Sum of OL Yards	Sum of RB Yards
0	3	-12	-14.4	2.4
1	6	20	20	0
2	3	47	14.1	32.9
3	5	31	15.5	15.5
Grand Total	**17**	**86**	**35.2**	**50.8**

This pivot table summarizes rushing performance by **Gain Bucket**, offering a breakdown of how the offensive line and running backs contribute across different yardage outcomes. The rows represent each **Gain Bucket**, where:

- **0** = 0 or fewer yards (negative plays)
- **1** = 1–4 yards
- **2** = 11+ yards
- **3** = 5–10 yards

The table includes the following metrics:

- **Num_Runs**: Total number of rush attempts in each gain bucket.
- **Total_Yards**: Combined rushing yards for all plays in that bucket.
- **Sum of OL Yards**: Total yards credited to the offensive line.
- **Sum of RB Yards**: Total yards credited to the running back.

Interpretation:

- **Gain Bucket 0**: These are unsuccessful plays where the offense lost a total of 12 yards across 3 attempts. Notably, OL Yards totaled -14.4, indicating that the offensive line was consistently overpowered. The running backs managed to salvage a modest 2.4 yards in total.
- **Gain Bucket 1 (1–4 yards)**: These were short-yardage gains across 6 runs, totaling 20 yards—all of which are credited to the offensive line. This suggests these plays were likely designed to gain minimal yardage (e.g., short conversions), with no contribution beyond the line's push.
- **Gain Bucket 2 (11+ yards)**: This small set of explosive plays (3 runs) yielded a remarkable 47 yards. Only 14.1 of those were from the line, while the running backs contributed an impressive 32.9 yards—indicating strong individual performance and field vision beyond the line of scrimmage.

- **Gain Bucket 3 (5–10 yards)**: This group delivered 31 total yards over 5 plays, split almost evenly between OL (15.5) and RB (15.5) yards. These medium-gain plays suggest strong blocking and effective follow-through by the ball carriers.

Findings & Insights

The Pivot from an External to an Internal Solution

Our initial project plan, based on the flexibility of tools like Python and Excel, was to deliver a master Excel template. The proposed workflow involved the client exporting raw DVSport data weekly and pasting it into our file to automatically generate the 10 reports. However, a crucial insight gained during our early client feedback sessions was that the manual export process itself was a significant pain point. Maxime expressed a strong preference for a solution that lived entirely within the DVSport platform to eliminate this recurring, time-consuming task. This finding fundamentally reshaped our project's direction, shifting our focus from building an external tool to optimizing their internal environment.

Deconstructing Report Logic for Strategic Value

We found that the structure of these reports was as important as the data within them. A report's utility depended entirely on whether the row and column hierarchies mirrored a coach's natural thought process. Our initial Python prototypes were essential for this. For example, we tested different layouts for Report 5, discovering that a structure with Personnel as the primary group and O BOX as the secondary group was far more intuitive for a coach than the reverse. This iterative process of standing in their shoes and asking, "Is this the easiest way for me to read this?" allowed us to refine the report structures. Once we figured out the optimal, coach-centric layout for each report, we had a blueprint that could be implemented in DVSport and reused over and over again for all future game analysis.

Uncovering Actionable Tendencies

A core finding of our analysis was the discovery of the DVSport system's dual-percentage capability: the "Global View," for understanding overall frequency, and the "Subgroup View," for analyzing situational tendencies. We quickly realized that the "Subgroup View" was significantly more valuable for coaching decisions.

Our deep dive into the report logic allowed us to frame the final deliverables around the specific strategic questions they answer. This transformed them from simple data tables into true analytical tools. The following examples highlight this:

Report 1 (R/P/O/A by Personnel):

- Coach's Question: "When I substitute my '11' personnel onto the field, what tendency am I showing the defense?"
- Report's Answer (Subgroup View): "When our 11 personnel are on the field, we have a 63.5% tendency to pass. This view is crucial for self-scouting predictable 'tells'."

Implications for the Community

The development of these clear, automated insights provides direct and tangible benefits to the McGill Football community. The reports are no longer just data tables; they are actionable strategic tools. For example, by using the Subgroup view of Report 1 to easily identify a 75% run tendency from 21 personnel, coaches can immediately adjust their play-calling in practices and future games to be less predictable. This data-driven self-awareness directly contributes to improving the team's on-field performance, which in turn enhances the experience for the student-athletes and the broader McGill community that supports them.

Sustainability Analysis

Technical Sustainability & Client Independence

Our final solution was designed for 100% client independence. After our initial client consultation, the project pivoted from an external Excel-based tool to an in-platform DVSport solution. During implementation, we discovered the system's limitation in performing nested subtotal calculations. After proactively consulting with DVSport technical support, we confirmed that the optimal and most sustainable solution was to build the best possible reports and integrate them into the system's "defaults list."

This "defaults" functionality is the key to the project's long-term technical sustainability. It ensures that whenever the coaching staff creates a new game or project in their system, all 10 reports we designed are automatically generated, populated with the new data, requiring zero manual setup or code rewriting from the coaches.

While the course rubric suggests a "single file where clients can easily open and use the solution" like an Excel template, our final hybrid solution achieves this goal even more effectively. For the eight core reports, the infrastructure is built directly into their primary software. For the two more complex reports (9 and 10), which required calculations beyond the system's capabilities, we delivered a pre-formatted Excel workbook. This workbook requires only a simple data paste into a single input tab to automatically update all pivot tables and charts.

Cost, Management, & Future Directions

- **Cost & Management**: There is zero ongoing financial or time cost to maintain this solution. The "defaults list" feature has fully automated the management of the eight in-platform reports. The Excel template for Reports 9 and 10 requires only a simple weekly data paste, a process that is significantly faster than their previous workflow.
- **Future Directions**: The foundational work of defining and building these 10 reports opens the door for more advanced analytics. The most logical future direction is to pursue API integration with the DVSport system. This would allow for the automated export of this clean, structured report data into other platforms (like R, Python, or a BI tool) for predictive modeling or the creation of more advanced, custom data visualizations.

Part II: The Twitter API Social Media Analysis Project (for Chris)

1. Project Overview & Purpose

The goal of Chris's stream was to enable the football program to leverage external data sources—including CFL's internal data portal, Genius Sports, and Twitter—to drive smarter decision-making. Rather than focusing on immediate reporting, this stream was about building a strategic data foundation that could fuel a wide variety of long-term analytics, from performance modeling to opponent scouting and public sentiment tracking.

2. Methodology & Data-Driven Strategic Plan

Chris's stream followed a data-first, modular methodology focused on expanding the scope and utility of external football-related data. The strategy centered around three pillars: identifying richer datasets, designing systems to structure them for future use, and enabling long-term scalability for advanced analytics. Although full access to external sources such as the CFL's internal platform and Genius Sports APIs was not available due to credential limitations, we developed a prototype integration framework outlining how these sources could be incorporated once access is granted. For Twitter, in the absence of official API access, we implemented a functional solution using snscrape, an open-source scraping tool that extracts publicly available tweet data without requiring authentication. The system was designed to be modular and adaptable — all output data was cleaned, labeled, and exported as flat CSV files, making it readily compatible with Excel, Tableau, or future modeling pipelines. All scripts were written for clarity and reusability, allowing staff with minimal technical experience to operate or expand the framework once full data access becomes available.

3. Solution

This report documents the development of two independent pipelines for collecting Twitter data related to Ivy League sports and athletics. The primary goal is to extract tweets mentioning specific keywords or published by verified institutional accounts such as `@IvyLeague`, `@YaleAthletics`, and `@HarvardFootball`, within a specified date range. Two distinct tools are used: `snscrape` and the official Twitter API v2.

1. Data Sources

- **Keywords:** `"Ivy League"`, `"Yale football"`, `"Harvard coach"`
- **Accounts:** `@IvyLeague`, `@YaleAthletics`, `@HarvardFootball`
- **Time Window:** July 1, 2023 – July 1, 2024
- **Language Constraint:** English (`lang:en`)

2. Methodology

`snscrape`-Based Scraper

`snscrape` is a robust, open-source scraper for social networking data. It bypasses the official API by leveraging Twitter's front-end JSON endpoints. This method enables unrestricted historical access but is potentially subject to HTML structure changes.

3. Key Features Implemented:

- Nested scraping over `(keywords × accounts)` matrix

- Use of `since:` and `until:` for date restriction
- Language filtering via `lang:en`
- Metadata extraction:
    - Tweet content, timestamp, username
    - Public metrics: likes, retweets, replies
    - User-level data: followers, display name
- Functional pipeline:
    - `build_query()` → `scrape_tweets()` → `batch_scrape()`
- Output to structured `DataFrame` and CSV

4. Twitter API v2 Scraper

The Twitter API v2 offers precise and officially supported access to tweet data. The `full-archive search` endpoint (Academic Research Track required) provides historical access beyond 7 days.

5. Key Functionalities:

- Secure access via Bearer Token (OAuth2.0)
- Controlled pagination using `next_token`
- Query builder with `from:`, `lang:`, and keyword logic
- Expansion of tweet metadata and user fields
- Output formatting with `tweet.fields`, `user.fields`, and `expansions`
- Rate-limiting and error handling (status code capture)

6. Technical Architecture

Component	snscrape Implementation	Twitter API v2 Implementation

Access Method	Unofficial scraper	Official REST API
Rate Limits	None (unofficial)	1 request/sec (Academic Tier)
Time Range	Any (no restriction)	Full archive with academic access
Language Filter	lang:en in query	lang:en in API parameter
Pagination	Manual loop with break	Controlled via next_token
Output Format	CSV and Pandas DataFrame	CSV and Pandas DataFrame
Robustness	HTML-dependent, may break	Stable and officially supported

7. Implementation Limitation: Twitter API v2

While the Twitter API v2 pipeline was fully developed and tested in a controlled environment, actual data collection could not be completed due to access limitations imposed by Twitter's pricing structure.

As of 2023, Twitter requires a paid subscription (e.g., Academic Research, Pro, or Enterprise tiers) to access the full-archive search endpoint or even to retrieve a significant volume of historical tweets. The pricing for such access is prohibitive for independent or academic use (often exceeding $5,000/month for full access).

As a result, while the codebase is functional and well-documented, the API pipeline remains unexecuted in production due to these financial constraints.

4. Sustainability Analysis

The solution developed for Chris's stream was intentionally designed to ensure long-term technical sustainability and client independence. Although full integration with CFL and Genius APIs was not achieved during the project due to credential restrictions, the final deliverable remains fully usable and adaptable. It consists of a self-contained, clearly organized folder that includes modular Python scripts, preformatted CSV outputs, and a documented example configuration. This setup enables coaching or analytics staff to operate the system independently, without needing support from the student team. The scripts are designed for ease of modification, so future users can easily update data sources, keywords, or configurations with minimal technical expertise. In keeping with the rubric's expectations, this solution functions as a "single file where clients can easily open and use the solution," providing a direct and accessible interface to external data workflows.

There are no ongoing costs associated with the use of this solution. All components rely entirely on open-source software, and no paid APIs or proprietary tools are required. Week-to-week management is simple: once CFL or Genius data access becomes available, staff can paste exported data directly into the workflow or configure the system to automate ingestion with minimal effort. Twitter scraping can continue via the snscrape tool, which operates without authentication or usage fees. Looking ahead, there are several clear paths for future development. Once API credentials are secured, the system can be extended to perform full automated ingestion from CFL and Genius Sports platforms. The existing

template can also be expanded to include additional metrics, broader team comparisons, or custom coaching questions. Over time, as historical data accumulates, the system can support more advanced modeling work such as predicting opponent tendencies, assessing player fatigue risk, or simulating in-game decision scenarios. Overall, the solution is not only sustainable in its current form but also positioned to evolve with the program's growing data capabilities.

III. Challenges & Limitations

Client Requirements (Max's Stream)

Although we initially proposed an external Excel-based master template, the client preferred an in-platform solution. The key challenge became designing and implementing an automated reporting suite directly inside DVSport, respecting its proprietary data schema and scripting limitations.

Technical Constraints (Max's Stream)

The primary challenge was that its reporting module lacks the flexibility of Excel; it has no self-defined formulas for calculating percentages at different row levels and no "subtotal" function. Any metric labeled % PLAYS is automatically calculated on a global basis against the grand total of all plays. To create column-like structures, the system relies on a single "grouping" variable, which further constrains complex, multi-level analysis.

Challenges & Data Dictionary Development (Max's Stream)

A critical early hurdle was the absence of documentation: dozens of abbreviated, domain-specific column names (e.g., g FPOS, O BOX) appeared in the raw feed. With no prior football or DVSport expertise, our team conducted extensive research, reviewing game film, coaching materials, and strategic analyses, to assemble a comprehensive Data Dictionary. This reference underpinned all subsequent automation logic.

Technical Constraints (Chris's Stream)

The main technical limitation was the lack of API access to both the CFL platform and Genius Sports, which prevented full data integration. Without credentials, we were unable to test live endpoints or automate data ingestion. Twitter's API was also inaccessible due to its paid model, so we used snscrape as a workaround to gather public tweet data. However, this method has limitations in depth, speed, and filtering.

Challenges & Early Design Decisions (Chris's Stream)

Limited communication from the stakeholder made it difficult to define concrete priorities or validate assumptions. As a result, the work focused on building a flexible, forward-compatible framework that could easily be adapted once access and clearer direction are provided. Designing without confirmed data schemas also posed a challenge, which we addressed by using modular code and generalized data structures.

IV. Recommendations

Actionable Steps: (Maxime)

- **Immediate Adoption of the Current Solution**: We recommend that Maxime and his staff immediately embed the 10 automated reports (8 in-platform, 2 in Excel) into their weekly workflow. Since the "defaults list" makes the in-platform reports easily repeatable, this will provide immediate time savings and analytical value.
- **Delegate and Empower**: We recommend creating a step-by-step handoff guide based on our documentation for a student assistant or junior analyst. This will delegate the initial weekly review and build the team's internal analytical capacity.
- **Engage DVSport Support for Advanced Calculations**: Our final conversation with Dylan, the DVSport Support Manager, revealed a significant opportunity. He confirmed that the system is SQL-based and that it is possible to add custom formulas and calculations, which could potentially solve the limitations we encountered with Reports 9 and 10. Therefore, our strongest recommendation is for Maxime to schedule a follow-up call with Dylan's team. The goal of this call would be to have DVSport support assist in implementing the custom yardage attribution formulas for Report 10 directly into the platform, potentially eliminating the need for the external Excel file entirely.

Actionable Steps: (Chris)

To fully realize the potential of the external data integration framework, we recommend that Chris and the team take several key steps. First, securing API credentials from the CFL and Genius Sports is essential to activate the prototype scripts and enable automated data ingestion. Next, the team should establish a simple weekly workflow—either by pasting exports into the provided template or automating data pulls—to ensure up-to-date insights are consistently available. We also suggest formalizing a list of recurring coaching questions to guide future metric development. Finally, appointing a point person to manage the system will ensure smooth, ongoing use. These actions will help transition the current solution into a reliable and scalable part of the football program’s analytics toolkit.

Stakeholder Involvement: (Maxime)

- **Involve Positional Coaches for Feedback**: We continue to recommend looping in other key staff members to gather feedback on the existing reports to ensure they remain relevant.
- **Formalize Future Collaborations**: The creation of our Data Dictionary provides a perfect onboarding tool for future analytics-minded student groups. This will allow the team to continuously evolve their analytical capabilities, whether by refining existing reports or pursuing more advanced projects. The immediate next project for such a group could be to work alongside DVSport support to implement the advanced formulas.

Stakeholder Involvement: (Chris)

- Stakeholder engagement in Chris’s stream was limited, which made it challenging to finalize the technical direction of the project. Despite early indications of interest in integrating data from CFL, Genius Sports, and Twitter, follow-up communication was minimal, and API credentials

were not provided. As a result, we focused on building a flexible, forward-compatible framework that could easily be activated and customized once access and clearer direction become available.

V. Conclusion

This project was ultimately about turning hours of manual effort into minutes of insight. Working closely with Coaches Maxime and Chris, we focused on building something practical — a system that could be used week after week without added friction. While the original scope of work imagined things like API integrations and advanced model tuning, the real value came from automating what already mattered most to the coaching staff: the weekly reports they relied on to make strategic decisions.

For strategic data analysis, we streamlined the DV Sport workflow and developed a complete suite of 10 reports — 8 within the platform and 2 in Excel — we've helped create a process that's not only faster but far more consistent and repeatable. From play-type tendencies and formation breakdowns to completion zones and rushing efficiency, each report is now tailored to answer a real coaching question, grounded in data, and delivered in a format that makes sense within their existing routine.

In parallel, we worked on expanding the program's long-term analytics capacity by laying the foundation for external data integration. Chris's stream focused on creating a flexible and scalable framework to bring in data from sources such as CFL's internal system, Genius Sports, and Twitter. While credential limitations prevented full implementation, we built prototype scripts and a modular architecture that can be activated once access is granted. This stream positions the team to unlock new insights around opponent tendencies, player performance, and even fan sentiment — opening the door for predictive modeling and more strategic scouting in future seasons.

The result is a set of tools that support smarter game planning, better self-scouting, and clearer communication between coaches and players. Most importantly, the solution is sustainable: no coding knowledge required, no added cost, and no need to rebuild it week after week. Just actionable insight, ready when they need it.