

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science e Big Data*

Thiago Garcia Molina Pinto

**Título: Modelagem e estruturação dos
processos de análise da liga BFA**

Curitiba
2022

Thiago Garcia Molina Pinto

**Título: Modelagem e estruturação dos processos de
análise da liga BFA**

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Setor de Ciências Exatas, Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Fernando de Pol Mayer.

Curitiba

2022

Título: Modelagem e estruturação dos processos de análise da liga BFA

Aluno: Thiago Garcia Molina Pinto
Orientador: Fernando de Pol
Mayer

Resumo

O trabalho e o desenvolvimento da análise de dados nos esportes revolucionaram a evolução dos jogos ou da estrutura dos negócios ao seu redor. As ligas profissionais pelo mundo mais bem desenvolvidas tem ótimos setores de análise do seu esporte e do seu negócio.

A liga BFA, (liga brasileira de futebol americano) é uma liga recente, que tem times semiprofissionais a times completamente amadores e ainda não tem um setor dedicado a análise. Este estudo teve por finalidade construir processo de coleta de variáveis dos jogos da liga BFA, criar ambientes de armazenamento dos dados de jogos e emitir relatório com análises dos jogos. Desta forma, de imediato, a Liga obtém um sistema de coleta padrão e de relatório dos jogos, além de construir um banco de dados históricos das partidas, ferramentas e processos que a liga ainda não tem, auxiliando no desenvolvimento da liga com o público e seus clubes.

Palavras-chave: Futebol americano, ambiente de armazenamento, sistema de coleta, análise estratégica.

Abstract

The work and development of data analytics in sports revolutionized the evolution of games or the business structure around them. The most well-developed professional leagues around the world have great sectors for analyzing your sport and your business.

The BFA league, (Brazilian American football league) is a recent league, which has semi-professional teams to completely amateur teams and still does not have a dedicated sector for analysis. This work aimed to build a process for collecting variables from the BFA league games, create game data storage environments and issue a report with analysis of the games. In this way, the League immediately obtains a standard collection and game reporting system, in addition to building a historical database of matches, tools and processes that the league does not yet have, helping to develop the league with the public and their clubs.

Keywords: American football, storage environment, collection system, strategic analysis.

I. Introdução

A National Football League (NFL), liga profissional de futebol americano nos EUA, é a liga esportiva mais popular do mundo, com milhões de espectadores a cada jogo e bilhões de dólares gerados a cada temporada. As estatísticas são uma parte importante do modelo operacional de negócios de uma equipe da NFL e contribuem muito para a tomada de decisões.[1]

Segundo a Forbes [2] o Dallas Cowboys, time da NFL, está no topo da lista de equipes esportivas mais valiosas no mundo, com uma avaliação de valor em torno de 5,7 bilhões de dólares, o top 10 ainda tem mais três times da NFL e o top 50 contem 26 equipes de futebol americano.

Uma das tarefas mais importantes para o desempenho aperfeiçoamento de uma equipe esportiva é analisar jogos passados de várias perspectivas. Usando a análise de jogos anteriores, a diferença entre o desempenho atual da equipe e o ideal desempenho pode ser identificado. [3]

Como visto na revisão sistemática recente [4] o desenvolvimento dos setores de análise, tornam as equipes muito mais competitivas dentro e fora do campo, o time passa a ter vantagem no ganho técnico da equipe, passa a preparar seus atletas melhor fisicamente, prepara melhor a comissão técnica para analisar os adversários, o negócio ao redor também passa a ter ganhos maiores com entendimento melhor dos consumidores.

As equipes e ligas profissionais preparam relatórios detalhados dos tipos de jogadas preferidos dos adversários em diversas situações do jogo. Falando de análise de dados no futebol americano profissional, é possível encontrar coleta de dados na década de 1930 [5], se concentrando no uso de probabilidades estatísticas para entendimento melhor do jogo e assim fornecer conselhos para o time tomar as melhores decisões.

Uma análise interessante para o público sobre como as equipes jogam é a análise de correspondência. A análise de correspondência serve para classificar graficamente, por coordenadas, variáveis categóricas, serve para explorar associação entre variáveis qualitativas, agrupando as informações

em um gráfico de duas dimensões [6]. Na prática ela pode responder quais elementos foram mais associados com uma categoria do que com outra. Desta forma pode-se entender, quais variáveis estão mais associadas a situações importantes ou ao resultado do jogo.

No Brasil o interesse pelo futebol americano vem crescendo. Uma pesquisa realizada pelo Ibope Repucom em 2016 apontava que 15,2 milhões de pessoas se declaravam fãs de futebol americano e em 2021 esse número saltou para 33 milhões de torcedores. O número de equipes federadas à Confederação Brasileira de Futebol Americano (CBFA) em 2020 era de 442 equipes e com 18 mil atletas. ***

A Liga Brasileira de futebol americano (BFA), foi criada em 2016 e é para ela que o produto deste estudo se destina. Ao contrário das experiências internacionais descritas sobre uso de análise dos dados estatísticos no esporte a Liga BFA tem dificuldade em construir o mesmo processo.

Nesse estudo a proposta é criar um processo de estrutura para analisar os jogos e demonstrar como as equipes jogam. Armazenar os dados gerados pelos jogos no futebol americano brasileiro em um único local para consulta histórica. E coletar as variáveis de forma sistematizada.

2 Materiais e Métodos

2.1 Caracterização do jogo

É preciso fazer uma breve descrição do futebol americano e como ele funciona. Existem algumas regras e versões diferentes do mesmo esporte, a NFL, o college futebol americano ou a liga profissional do Canada jogam o mesmo esporte, porém com algumas regras, dimensões do campo e desempenhos diferentes. Uma relação parecida como o vôlei de quadra e o vôlei de areia. Para uma descrição mais concisa do esporte e para falar do que é enfrentado nas regras e modo de jogo da IFAF (International Federation of American Football), que são as regras usadas na liga BFA.

O objetivo do jogo é ter a posse da bola na zona do campo chamada de *endzone*, alcançar esse objetivo vale 6 pontos, gerando uma tentativa adicional de mais 1 ponto ao tentar um chute pelo *fieldgoal* (traves em Y onde a bola chutada deve passar entre elas) ou 2 pontos adicionais tentando entrar novamente na *endzone*. As maneiras de alcançar esse objetivo tem diversas variações, as mais comuns são correndo com a bola ou passando a bola para frente.

Cada time posiciona 11 jogadores de cada lado, um para defender seu campo, e o outro para atacar, o time com a posse da bola tem 4 tentativas para avançar com a bola 10 jardas, não alcançando as 10 jardas o time então perde a posse da bola onde a bola parou em seu último avanço, na 4ª e última tentativa, é normal o time devolver a bola para o outro time o mais longe da sua *endzone* com um chute, ou tentar marcar 3 pontos chutando um *fieldgoal*. [7]

O campo contém 360 pés por 160 pés, como visto na Figura 1

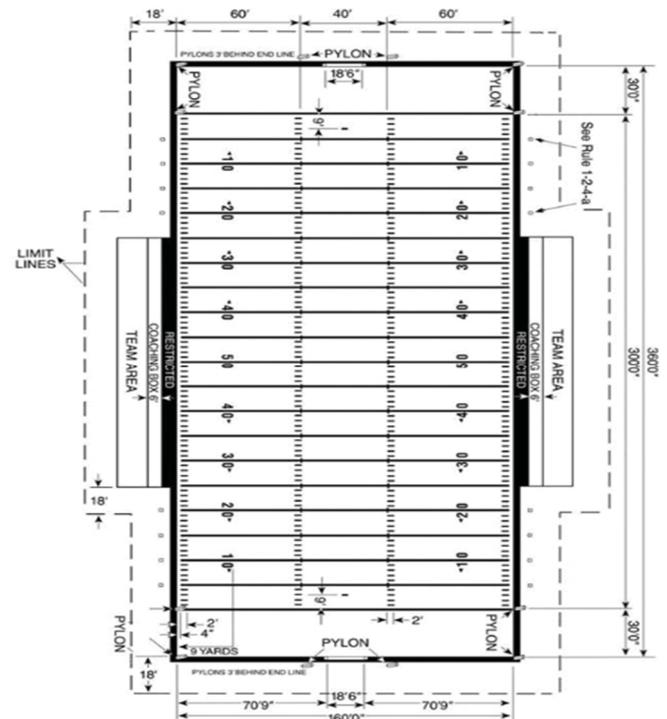


Figura 1: dimensões do campo obtida no manual de pintura de campo liga BFA

O presente estudo trata da realidade brasileira no principal campeonato organizado pela liga BFA deste esporte.

2.2 Padronização da coleta dos dados

Na realidade brasileira do esporte, não existia um padrão de coleta de variáveis. Então, uma planilha foi criada.

Na ocorrência dos jogos, eles são transmitidos no *Youtube* e ficam gravados. Estes vídeos serviram como fonte de coleta

A escolha das variáveis seguiu a dinâmica do jogo. Um jogo é dividido em jogadas, essas jogadas podem ser divididas em 3 etapas. A primeira etapa pré snap (comando que coloca a bola em jogo), a segunda etapa em que a jogada se desenrola e a última etapa e o que acontece depois que a jogada foi feita. Então pode-se fazer um apanhado de informações antes do início da jogada como:

-Primeira etapa pré snap: Posição do campo que a jogada ocorre, formação que os times estão dispostos em campo, qual o momento em que o jogo se encontra.

-Segunda etapa da jogada: Qual jogada foi executada, quem iniciou ou finalizou a jogada.

- Terceira e última etapa pós a jogada: Qual o ganho ou perda de jardas ou pontos, penalidades.

Então existe um ciclo contínuo para coleta, um processo de coleta pode ser criado para agilizar essa tarefa e preencher as informações do total do jogo.

É preciso que o entendimento das mais diversas interpretações de alguns pontos da coleta sejam o mesmo. É crítico ao processo que a coleta seja feita de forma padronizada, portanto idealizou-se um manual de coleta, para que o entendimento de como alimentar a planilha seja o mesmo por todos os coletores.

Todas as informações de todos os jogos foram unificadas em um único local, onde possa ser consultado sem a necessidade de novas coletas, por isto um banco de dados histórico passa a ser criado a partir das variáveis coletados nas planilhas padrão. A automação deste processo é importante. E um código foi escrito que coleta todas essas planilhas padronizadas de um local e crie esse banco de dados único com todos os jogos coletados.

Essa unificação foi elaborada no ambiente "R" [10].

2.3 Relatório dos jogos

Ao final do jogo e da coleta de dados unificada, a visualização dos resultados foi padronizada em um relatório único dinâmico.

O relatório do jogo tem por objetivo entregar um

apanhado de informações que apresente com dados como o jogo aconteceu. Então, além da informação do resultado, é interessante apresentar outras informações, como mandante, visitante, vencedor, produção de jardas avançadas pelo ataque, número de jogadas, número de vezes que um time perdeu a posse da bola, tipo de jogada executada (passe ou corrida).

2.3.1. Análises exploratórias

Dentro do relatório, foram inseridas análises exploratórias, que descrevem como foi a partida, essas análises precisam mostrar o time que mais pontuou para indicar o vencedor da partida. Apresentar quando os pontos ocorreram, em qual quarto cada time pontuou para mostrar qual foi a dinâmica das pontuações no jogo. Quantas jardas cada ataque dos times produziu, é interessante separar essas jardas por passe e corrida e mostrar em quais quartos dos jogos essas jardas foram conquistadas. Número de fist Down de cada equipe, indicando quantas oportunidades cada equipe teve para iniciar uma série de ataque. Número de turnover para dar um panorama de quantas vezes cada time perdeu a bola para o adversário. Além das informações de quem era o mandante da partida e quando jogo aconteceu.

Essas análises dão um panorama geral de como o jogo ocorreu e quais foram as formas que as equipes jogaram.

2.3.2. Análises múltiplas

Além das análises exploratórias, foram realizadas análises que apresentem como as equipes vem se comportando durante a temporada. As perguntas respondidas por tais análises foram:

1° O que mais contribui para a pontuação do time?

2° O que mais contribui para a decisão do time correr ou passar a bola?

3° O que mais contribui para o time acertar seus passes?

Para responder essas perguntas, foi optado por um modelo de análise de correspondência, desta forma consegue-se associar quais variáveis correspondem mais a variáveis importantes ao jogo.

Esse modelo apresenta resultados interessantes para os espectadores da liga e outras equipes entenderem como as equipes vem jogando. Na prática ela pode responder quais elementos foram mais associados com uma categoria do que com outra. Para exemplificar, pode-se entender com a análise quando uma equipe prefere correr ou passar. E desta forma um espectador novo ao esporte, passa a entender quais são as variáveis importantes para o desenvolvimento do jogo e a maneira dos times jogarem, aprofundando o conhecimento dele ao esporte.

Para a pergunta 1 foram utilizadas as seguintes variáveis: Down (número da tentativa de jogada, categorizada em 1,2,3,4), Distancia (distancia para a linha de execução da jogada, categorizada em 1 a 4 jardas de distância, 5 a 9 jardas e mais de 10 jardas), Formação do ataque (disposição dos jogadores de ataque no campo, categorizada em 4 formações distintas), Frente da defesa (como os jogadores da linha e atrás da linha de defesa estão dispostos, categorizada em 2 formações), cobertura da defesa (como os jogadores do fundo do campo de defesa estão dispostos, categorizados em 3 formações) e Pontos (categorização binária de pontos ou não pontos). Com a análise de correspondência, verificou-se a associação entre essas variáveis, observando qual delas está mais associada a variável ponto.

Para as perguntas 2 e 3 foram utilizadas as mesmas variáveis, trocando a variável ponto pelo resultado da jogada (categorizada em passe completo ou não completo) para a pergunta 2. Na 3, a troca foi pelo tipo da jogada (categorizada em passe ou corrida). Com a análise de correspondência, verificou-se na pergunta 2 a associação entre essas variáveis, observando qual delas está mais associada a variável passe completo ou aos não completos e na 3 qual delas está mais associada a variável passe ou aos de corrida.

No relatório as análises de correspondência vão levar em conta os dados dos jogos daquela temporada, ao contrário das outras análises que levam em conta as variáveis do jogo em si, a análise de correspondência é feita com todos os jogos que

o time jogou naquela temporada.

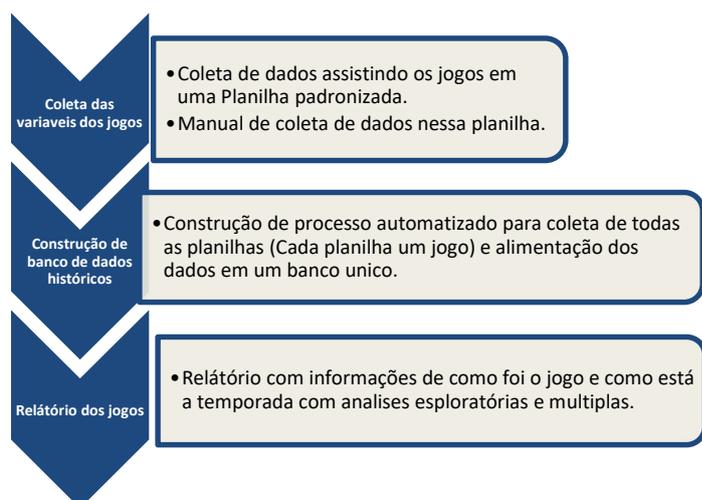
Essas análises foram elaboradas no ambiente R utilizando funções básicas e os pacotes “ca” [11] “FactoMineR” [12] “factoextra” [13]

As análises do relatório são apresentadas em um padrão. Deste modo, da coleta, armazenamento dos dados até o relatório, a automatização do processo facilita para que mesmo na ausência de um programador experiente, o processo possa ser executado. O R markdown [14] foi utilizado para construir um padrão de entrega tanto das análises como do visual da apresentação.

Os processos até aqui podem ser sintetizados desta forma na Figura 2.

Os resultados apresentados utilizam dados simulados e reais de jogos da liga BFA. Isso acontece porque embora a Liga tenha organizado alguns anos de campeonato, nenhuma dessas variáveis foram coletadas em nenhum momento. Por isso foi simulada a coleta na planilha padrão com o auxílio do manual criado neste estudo. O processo de unificação elaborado para criar o banco de dados únicos foi validado usando 4 jogos simulados. As análises exploratórias do relatório contêm um jogo com variáveis reais e simuladas e a análise múltipla utilizou os dados do jogo do relatório da exploratória mais os dados dos outros jogos simulados.

Figura 2. Síntese dos processos criados para coleta, armazenamento e análise de jogos de futebol americano.



3 Resultados e discussão

3.1 Padronização da coleta dos dados

3.1.1. Planilha de coleta de variáveis

Levando-se em consideração o processo das jogadas durante o jogo, a fim de carregar o máximo de informações, foram selecionadas as variáveis a segundo a ordem que a jogada vai ocorrendo. Primeiro, as variáveis de como os times se encontram em campo, depois como a jogada ocorreu e por fim os resultados que a jogada entregou. Na tabela 1 se encontram as variáveis na sequência criada para a planilha padrão de coleta.

Tabela 1: Dicionário do conjunto de variáveis.

Varável	Descrição
RODADA	Número da rodada
MANDANTE	Nome do time mandante
VISITANTE	Nome do time visitante
DATA	Data do jogo
PLAY #	jogada
OFENSE	Nome do time que está com ataque em campo
DEFFENSE	Nome do time que está com defesa em campo
SPECIAL_RETURN	Nome do time que está com time de especialista de retorno em campo
SPECIAL_KICK	Nome do time que está com time de especialista de chute ou punt em campo
KICK	Nome do time que está com time de chute de fielgoal em campo
QTR	Tempo de jogo Qtr
DN	Down 1,2,3,4
DIST	Distância para linha de goal
HASH	lado esquerdo L, Lado Direito D meio M
QB jersey	Número do qb
Qbname	Nome do qb
YARD LN	posição no campo, positivos no campo de ataque negativos no campo de defesa
PLAY TYPE	Corrida Run, Passe pass,Ponto extra Extra point,Punt punt, Field goal FG
RESULT	Pass, completo ou incompleto, Rush positivo ou negativo, penalty
GN/LS	Jardas ganhas + ou perdidas -
OFF FORM	Formação do ataque
OFF PLAY	Movimentação pré snap que modifica a formação
OFF STR	Lado forte TE ou RB
PLAY DIR	Direção onde a jogada foi, esquerdo L ou direito D
GAP	Gap de entrada da corrida A,B, C
PASS ZONE	Zona do campo onde a jogada ocorreu 1 a 9
PLAYER_Jersey	Número do jogador que correu ou passou
PLAYER_Name	nome do jogador que recebeu ou passou
DEF FRONT	Formação do "front" DL e LB
COVERAGE	Cobertura da secundaria CB E S's
BLITZ	Quantos jogadores foram para blitz
INTERCEPTED BY_Jersey	Número do jogador que interceptou
INTERCEPTED BY_Name	nome do jogador q interceptou
TACKLER1_Jersey	Número do jogador que fez o TKL
TACKLER1_Name	Nome do jogador que fez o tkl
sack1_Jersey	Número do primeiro jogador a sakear
sack2_Jersey	Número do segundo se houver jogador a sakear
sack1_Name	Nome do primeiro jogador a sakear
sack2_Name	Nome do segundo se houver jogador a sakear

3.1.2. Manual de coleta de variáveis

A necessidade de um manual vem da complexidade do jogo e das diversas jogadas e formações que podem ter interpretações diferentes. Para que a coleta na planilha padrão não se limite a um indivíduo, um manual foi criado contendo instruções sobre a maneira da coleta de dados dos vídeos dos jogos, assim qualquer um pessoa com entendimento do jogo pode usando o manual pegar a planilha padrão e assistindo o jogo coletar as variáveis na planilha . O manual está apresentado na íntegra no Anexo 1.

O manual visou aplicar instruções passo a passo de como alimentar a planilha. Na Figura 3, foi selecionado um exemplo que mostra como identificar uma jogada de Passe. Na Figura 4 outro exemplo, que mostra como identificar em qual *Hash* a posição da bola se encontra.

Passo 13:

Qual foi o tipo de jogada, se foi corrida Rum, ou passe Pass, aqui é preciso dar play e aguardar o desenrolar da jogada, aqui passe Pass.



	M	N	O	P	Q	R
1	DIST	HASH	Qb jersey	Qb name	YARD LN	PLAY TYPE
2						
3	10	M	7	Bassani	-32	Pass

Figura 3. Manual de coleta de Variáveis Passo 13.

Como visto na figura 3, o jogador com a bola está fazendo o movimento de jogar a bola para frente, portanto a identificação da jogada na coluna *Play type* da tabela 1 e da planilha padronizada, é a de *Pass*.

Passo 10.:

Hash é a posição que a bola se encontra, pode estar a direita R, a esquerda L ou no meio M, na jogada esta no meio M.



J	K	L	M	N
KICK	QTR	DN	DIST	HASH
	1			
	1	1	10	M

Figura 4. Manual de coleta de Variáveis Passo 10

Como observado na Figura 4, as duas linhas pontilhadas no meio do campo, a posição da bola esta no meio entre elas, portanto o preenchimento da coluna *Hash* descrita na tabela 1 ou na planilha

3.1.3. Banco de dados histórico

A criação de um banco de dados históricos único foi necessária para operacionalizar o armazenamento das variáveis dos jogos. Assim, todos os jogos e toda a série temporal ficaram no mesmo local, evitando um processo manual de copiar e colocar todas as informações de uma planilha para outra. Para isso, optou-se por elaborar um código na linguagem R, que copia as variáveis de todas as planilhas em um local e cria um arquivo com todos os dados conjuntos.

Como finalização, um arquivo texto padrão está sendo utilizado para armazenamento.

3.2. Relatório dos jogos

O jogo precisa de um relatório para divulgação do resultado e dos aspectos do jogo, pensando nisso resolveu elaborar um relatório com as seguintes análises e informações da partida na tabela 2:

Tabela 2: Inputs e outputs do relatório.

Input (Variáveis coletadas dos jogos)	Output (análises presentes no relatório)
---------------------------------------	--

Rodada, Id, Poit_team, Points, Qtr	Soma dos pontos por time e quarto
Rodada, Id, gn_ls, offense, special_kick	Jardas totais por time
Rodada, Id, gn_ls, Play_type, offense, special_kick	Jardas totais por corrida e time
Rodada, Id, gn_ls, Play_type, offense, special_kick	Jardas totais por passe e time
Rodada, Id, dn, Ofense, kick, special_kick	NÚMERO DE FIRST DOWN POR EQUIPE
Rodada, Id, Ofense, kick, special_kick, special_return, result	NÚMERO DE TURNOVERS POR EQUIPE
DN, DIST, OFF FORM, DEF FRONT, COVERAGE, Point	O que mais contribui para a pontuação? por equipe
DN, DIST, OFF FORM, DEF FRONT, COVERAGE, RESULT	O que mais contribui para o passe ser completo ou incompleto? Por time
DN, DIST, OFF FORM, DEF FRONT, COVERAGE, Play type	O que mais contribui para a jogada do ataque ser corrida ou passe? Por time

Os resultados são apresentados em um relatório padrão em R markdown onde com um clique, um relatório completo em .pdf é gerado.

3.2.1. Análises exploratórias

As análises exploratórias dão aos interessados na liga, um panorama geral de como o jogo ocorreu. O relatório está na íntegra no Anexo 2 com os resultados obtidos com os dados simulados já descritos.

Na figura 5, estão apresentadas as informações técnicas de quando o jogo ocorreu e quais times jogaram como mandante e como visitante. Na figura 6 uma tabela com o resultado do jogo via pontuação por quarto. A figura 7 trata do número de jardas produzidos por time, com uma tabela e uma representação gráfica. Na figura 8 é a tabela no relatório que divide a produção de jardas por corrida e por passes dos times. Outra análise representada no relatório representada graficamente é a figura 9 que mostra os ganhos de jardas divididos por quarto. A figura 10 traz as últimas análises exploratórias do relatório, com a informação de números de first down por equipe em uma tabela, em um gráfico o número de first down por quarto em um gráfico e em uma tabela os turnovers das equipes

RELATÓRIO DE JOGO

RELATOR: Thiago Garcia Molina Pinto
2022-06-16

JOGO

Data:	20/05/2022
MANDANTE:	CROCO
VISITANTE:	BS

Figura 5. Cabeçalho e início do relatório.

PONTOS POR QUARTO

TEAM	1	2	3	4	FINAL
BS	3	6	3	0	9
CROCO	7	0	0	13	13

Figura 6. Apresenta pontuação dos times por quarto do jogo e o resultado final

NÚMERO DE JARDAS TOTAIS DO ATAQUE

TEAM	TOTAL OFFENSE
BS	17
CROCO	133



Figura 7. Número de jardas totais do ataque, tabela e gráfico.

NÚMERO DE JARDAS DA FORMAÇÃO DE ATAQUE: CORRIDA

TEAM	YARDS
BS	-6
CROCO	47

NÚMERO DE JARDAS DA FORMAÇÃO DE ATAQUE: PASSE

TEAM	YARDS
BS	23
CROCO	91

Figura 8. Conquistas de jardas por equipe, divididas pelos ganhos de passe e corrida.

NÚMERO DE JARDAS POR QUARTO POR EQUIPE

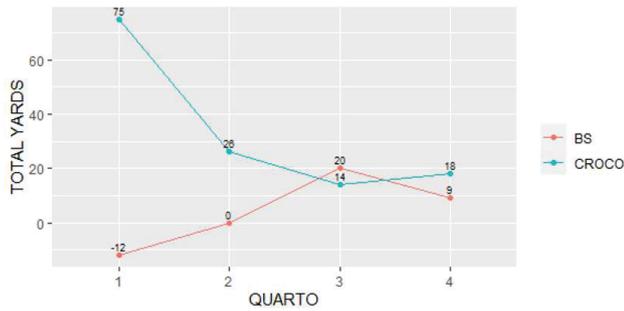


Figura 9. Ganhos de jardas representada graficamente por equipe e quarto.

NÚMERO DE FIRST DOWN POR EQUIPE

TEAM	TOTAL FIRST DOWN
BS	19
CROCO	21

NÚMERO DE TURNOVERS POR EQUIPE

TURNOVER	TEAM	TOTAL
Fumble	BS	2
Interception	BS	1
Fumble	CROCO	0
Interception	CROCO	1

NÚMERO DE FIRST DOWN POR QUARTO POR EQUIPE

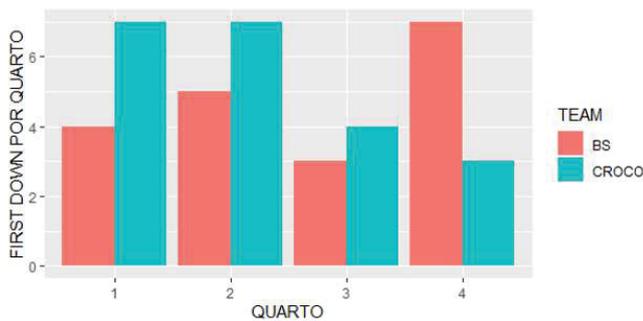


Figura 10. Turnovers por equipe e gráfico de *first down* por equipe e quarto.

3.2.2. Análises múltiplas

As análises múltiplas serviram para responder perguntas de como a temporada vem se desenrolando e como as equipes vem jogando. Os consumidores do material gerado pelo relatório podem interpretar o gráfico da análise observando os vetores que se aproximam entre si e entender que essas variáveis estão associadas. Nas figuras 11,12 e 13 estão os exemplos construídos com os dados de testes e a interpretação destes resultados se dão da seguinte forma.

A figura 11 responde o que mais contribuiu para as pontuações do time. Pode-se observar que os pontos se relacionam com a variável da formação do front da defesa do adversário com 4 jogadores na linha e 3 Linebackers (4_3) destacadas com tracejado azul e a letra A.

A figura 12 responde à pergunta de quais variáveis contribuem para escolha do time em passar ou correr. Uma das relações é a marcada como A e tracejado azul que aponta que quando o jogo está empatado e o time está com a sua formação "I" de ataque em campo, a jogada mais frequente é corrida.

Na Figura 13, foram analisadas as variáveis e como se relacionam com passes serem completos ou incompletos. Pôde-se observarno destaque em tracejado azul e letra A que quando a formação da defesa adversária estava com 4 jogadores na linha e 3 linebackers (4_3), cobertura com 2 jogadores no fundo do campo C2 e a formação do ataque era F4, essas condições resultam mais frequentemente em passes completos.

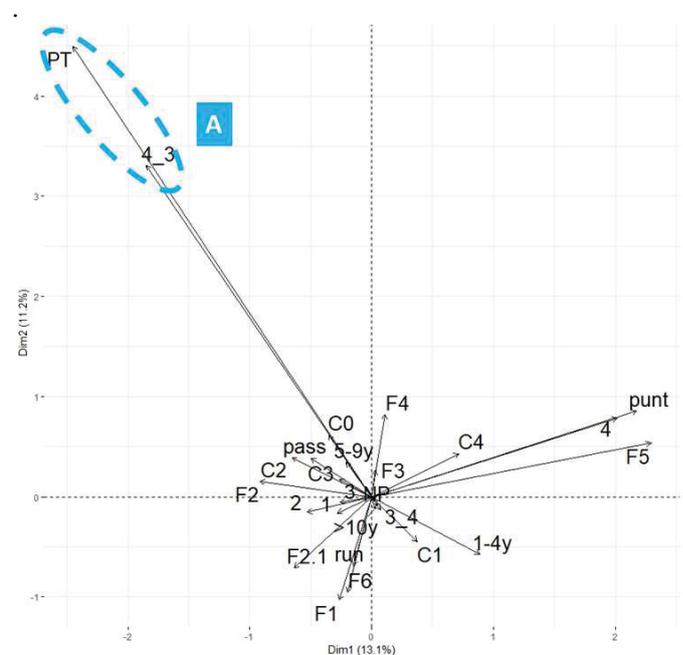


Figura 11. Gráfico da análise de correspondência múltipla. O tracejado azul e a letra A destacam quais variáveis mais contribuíram para a equipe pontuar.

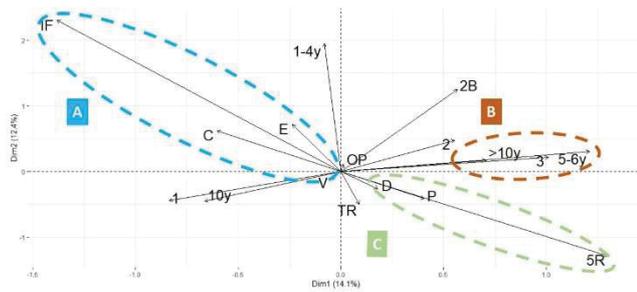


Figura 12. Gráfico da análise de correspondência múltipla. Marcando as associações das variáveis com o tracejado azul, marrom e verde com as letras respectivamente A, B e C mostram quais variáveis mais contribuíram para a equipe decidir correr ou passar.

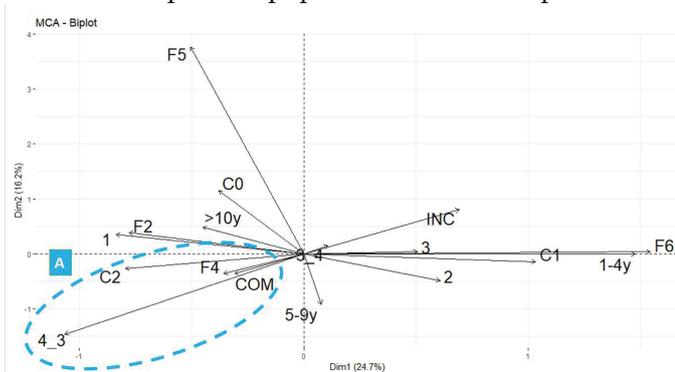


Figura 13. Gráfico da análise de correspondência múltipla. Marcando a associação das variáveis com o tracejado azul e a letra A mostram quais variáveis mais contribuíram para os passes completados de um time.

4. Conclusões

Foi elaborado neste estudo um processo estruturado de análise que vai da coleta estruturada e padronizada, ao armazenamento dos dados históricos de forma automatizada e relatório dos jogos.

Este estudo servira de base para a construção do banco de dados da liga BFA e geração de relatórios automáticos e padronizados.

Desta forma a liga BFA passa ter um processo estruturado, elaborado do zero, desta forma a liga que apenas armazenava os resultados dos jogos, passa a poder produzir relatório dos jogos, da temporada ou do total dos dados armazenados. Além disso com o modelo de análise de correspondência que passa dar panoramas jogo a jogo de como as equipes se comportam durante o campeonato.

Agradecimentos

Gostaria de agradecer meu orientador Fernando de Pol Mayer pela orientação e ajuda na construção deste trabalho de conclusão, pela clareza e objetividade na orientação. Agradecer aos professores da especialização em data Science e big data que aceitaram o desafio de continuar perante as adversidades que a pandemia de COVID19 trouxe. E para concluir, deixar meu agradecimento e admiração a minha namorada e companheira de vida, Camila Marinelli Martins, seu apoio nos momentos difíceis que foram esses últimos anos.

Referências

- [1] Porter, L.. Decision Making within an NFL Context Using Multiple Objective Decision Analysis. *Industrial Engineering Undergraduate Honors Theses* Retrieved from Disponível em: <https://scholarworks.uark.edu/ineguht/77> acesso em 13/07/2022
- [2] Ozanian, Mike As equipes esportivas mais valiosas do mundo 2021 Disponível em: <https://forbes.com.br/listas/2021/05/as-equipes-esportivas-mais-valiosas-do-mundo-2021/#foto1> acesso em 13/07/2022
- [3] TANI, Toshihiro; HUANG, H.; KAWAGOE, Kyoji. Sports play visualization system for American football. In: **Proceedings of the international multiconference of engineers and computer scientists**. 2015
- [4] HERBERGER, Tim A.; LITKE, Christoph. The Impact of Big Data and Sports Analytics on Professional Football: A Systematic Literature Review. **Digitalization, Digital Transformation and Sustainability in the Global Economy**, p. 147-171, 2021.
- [5] CARROLL, Bob Newhardt; PALMER, Pete; THORN, John. **The hidden game of football**. Warner books, 1988 *data science: import, tidy, transform, visualize, and model data*. O'Reilly Media, Inc., 2016
- [6] SUJATHA, K.; GODHAVARI, T.; BHAVANI, Nallamilli PG. Football match statistics prediction using artificial neural networks. **International Journal of Mathematical**
- [7] Lacalle, Danilo O crescimento do Futebol Americano no Brasil depende, também, do fã da NFL no país Disponível em: www.lance.com.br/futebol-

- [americano/crescimento-brasil-depnde-tambem-nfl-pais.html](#) acesso em 13/07/20
- [8] FLORENSA, Dídac et al. The use of multiple correspondence analysis to explore associations between categories of qualitative variables and cancer incidence. **IEEE Journal of Biomedical and Health Informatics**, v. 25, n. 9, p. 3659-3667, 2021.
- [9] STERN, H. American football. **Statistics in sport**, p. 3-23, 1998.
- [10] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [11] Wickham H, Hester J, Bryan J (2022). `_readr: Read Rectangular Text Data_`. R package version 2.1.2, <https://CRAN.R-project.org/package=readxl>
- [12] Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>.
- [13] Kassambara A, Mundt F (2020). `factoextra: Extract and Visualize the Results of Multivariate Data Analyses`. R package version 1.0.7, <<https://CRAN.R-project.org/package=factoextra>>.
- [14] Allaire J, Horner J, Xie Y, Marti V, Porte N (2019). `markdown: Render Markdown with the C Library 'Sundown'`. R package version 1.1, <<https://CRAN.R-project.org/package=markdown>>.

Anexo I

MANUAL DE COLETA DE VARIÁVEIS LIGA BFA



PLANILHA PADRÃO DE COLETA DE DADOS

Com o intuito de aprimorar os dados coletados e transformados em informação, foi estipulado uma planilha padrão de coleta de dados, desta forma otimizando a análise e o sistema de armazenamento desses dados. As variáveis são as descritas abaixo e neste documento iremos fazer um passo a passo de como fazer a coleta.

Varável	Descrição
RODADA	Número da rodada
MANDANTE	Nome do time mandante
VISITANTE	Nome do time visitante
DATA	Data do jogo
PLAY #	jogada
OFENSE	Nome do time que está com ataque em campo
DEFENSE	Nome do time que está com defesa em campo
SPECIAL_RETURN	Nome do time que está com time de especialista de retorno em campo
SPECIAL_KICK	Nome do time que está com time de especialista de chute ou punt em campo
KICK	Nome do time que está com time de chute de fielgoal em campo
QTR	Tempo de jogo Qtr
DN	Down 1,2,3,4
DIST	Distancia para linha de goal
HASH	lado esquerdo L, Lado Direito D meio M
QB jersey	Numero do qb
Qbname	Nome do qb
YARD LN	posição no campo, positivos no campo de ataque negativos no campo de defesa
PLAY TYPE	Corrida Run, Passe pass, Ponto extra Extra point, Punt punt, Field goal FG Pass, completo ou incompleto+ TD, Rush positivo ou negativo +TD, Extra point good , penaltys
RESULT	penaltys
GN/LS	Jardas ganhas + ou perdidas -
OFF FORM	Formação do ataque
OFF PLAY	Movimentação pré snap que modifica a formação
OFF STR	Lado forte TE ou RB
PLAY DIR	Direção onde a jogada foi, esquerdo L ou direito D
GAP	Gap de entrada da corrida A,B, C
PASS ZONE	Zona do campo onde a jogada ocorreu 1 a 9
PLAYER_Jersey	Numero do jogador que correu ou passou
PLAYER_Name	nome do jogador que recebeu ou passou
DEF FRONT	Formação do "front" DL e LB
COVERAGE	Cobertura da secundaria CB E S's
BLITZ	Quantos jogadores foram para blitz
INTERCEPTED	
BY_Jersey	Numero do jogador que interceptou
INTERCEPTED	
BY_Name	nome do jogador q interceptou
TACKLER1_Jersey	Numero do jogador que fez o TKL
TACKLER1_Name	Nome do jogador que fez o tkl
sack1_Jersey	Numero do primeiro jogador a sackear
sack2_Jersey	Numero do segundo se houver jogador a sackear
sack1_Name	Nome do primeiro jogador a sackear
sack2_Name	Nome do segundo se houver jogador a sackear

PASSO A PASSO DE COLETA

Passo 1:

Identificação da rodada, no caso desse jogo do manual se trata da **FINAL**.

	A
1	RODADA
2	FINAL

Passo 2:

Identificação do mandante, aqui denominado **REX**.

	A	B
1	RODADA	MANDANTE
2	FINAL	REX

Passo 3:

Identificação do visitante, aqui denominado **ESPECTROS**.

	A	B	C
1	RODADA	MANDANTE	VISITANTE
2	FINAL	REX	ESPECTROS

Passo 4:

Data do jogo, aqui 15/12/2019.

	A	B	C	D
1	RODADA	MANDANTE	VISITANTE	DATA
2	FINAL	REX	ESPECTROS	15/12/2019

Passo 5:

Número da jogada, aqui primeira jogada.

	A	B	C	D	E
1	RODADA	MANDANTE	VISITANTE	DATA	PLAY # <input type="text"/>
2	FINAL	REX	ESPECTROS	15/12/2019	1

Passo 6:

Hora de identificar em qual posição os times estão em campo.



É preciso preencher na planilha qual time esta em qual posição, como podemos observar o time de branco (Espectros) esta com o time de SPECIAL_KIK e o time de vermelho com o time de Special_return

	B	C	D	E	F	G	H	I	J
1	MANDANTE	VISITANTE	DATA	PLAY #	OFENSE	DEFFENSE	SPECIAL_RETURN	SPECIAL_KICK	KICK
2	REX	ESPECTROS	15/12/2019	1			REX	ESPECTROS	

Quando um time estiver no ataque preenche seu nome na coluna F, defesa na G ou com time de chute de field goal no J

Passo 7:

Qual o quarto do jogo está, aqui no início do jogo no 1º quarto.

	A	B	C	D	E	F	G	H	I	J	K
1	RODADA	MANDANTE	VISITANTE	DATA	PLAY #	OFENSE	DEFFENSE	SPECIAL_RETURN	SPECIAL_KICK	KICK	QTR
2	FINAL	REX	ESPECTROS	15/12/2019	1			REX	ESPECTROS		1

A partir de agora, a menos que não tenha uma pontuação, não iremos preencher mais nenhuma informação, continuaremos daqui para frente na jogada 2

Passo 1, 2, 3, 4, 5, 6 e 7:



Rex no ataque, espectros na defesa no primeiro quarto.

	A	B	C	D	E	F	G	H	I	J	K
1	RODADA	MANDANTE	VISITANTE	DATA	PLAY #	OFENSE	DEFFENSE	SPECIAL_RETURN	SPECIAL_KICK	KICK	QTR
2	FINAL	REX	ESPECTROS	15/12/2019	1			REX	ESPECTROS		1
3	FINAL	REX	ESPECTROS	15/12/2019	2	REX	ESPECTROS				1

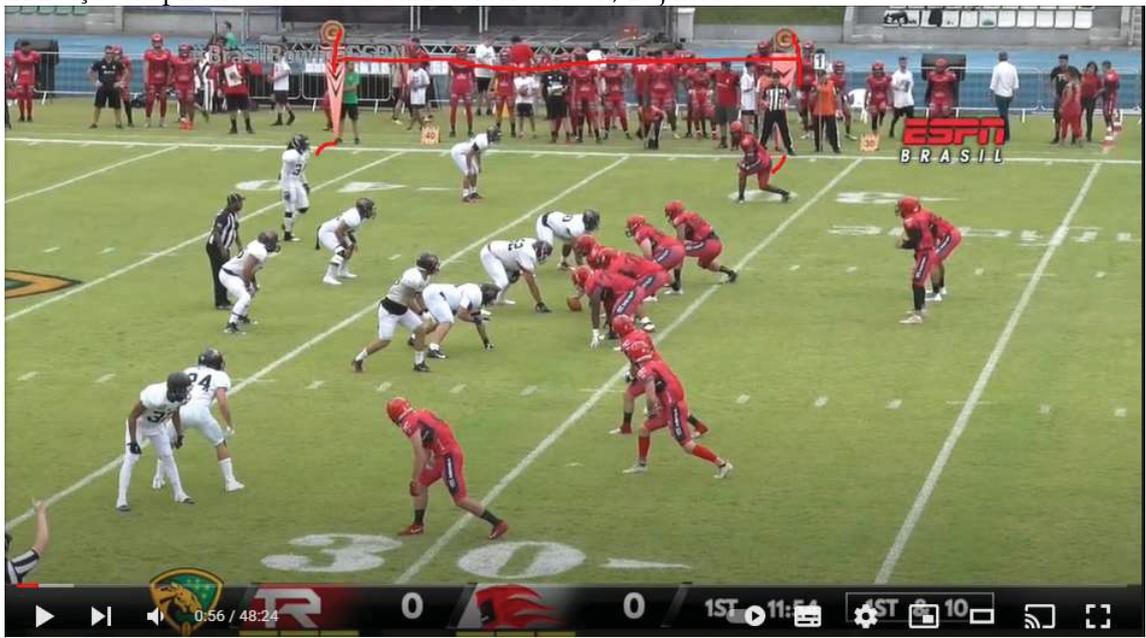
Passo 8.:

Qual down está a jogada, aqui 1º down.

	A	B	C	D	E	F	G	H	I	J	K	L
1	RODADA	MANDANTE	VISITANTE	DATA	PLAY #	OFENSE	DEFFENSE	SPECIAL_RETURN	SPECIAL_KICK	KICK	QTR	DN
2	FINAL	REX	ESPECTROS	15/12/2019	1			REX	ESPECTROS		1	
3	FINAL	REX	ESPECTROS	15/12/2019	2	REX	ESPECTROS				1	1

Passo 9.:

Qual a distância para o first down, aqui podemos contar na imagem quantas jardas da linha da bola até o a indicação do pirulito onde está a linha de fist down, 10 jardas.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	RODADA	MANDANTE	VISITANTE	DATA	PLAY #	OFENSE	DEFFENSE	SPECIAL_RETURN	SPECIAL_KICK	KICK	QTR	DN	DIST
2	FINAL	REX	ESPECTROS	15/12/2019	1			REX	ESPECTROS		1		
3	FINAL	REX	ESPECTROS	15/12/2019	2	REX	ESPECTROS				1	1	10

Passo 10.:

Hash é a posição que a bola se encontra, pode estar a direita R, a esquerda L ou no meio M, na jogada esta no meio M.



J	K	L	M	N
KICK	QTR	DN	DIST	HASH
	1			
	1	1	10 M	

Passo 11.:

Qual o número do QB e qual seu nome, no caso aqui, o número 7 e nome de Bassani.

K	L	M	N	O	P
QTR	DN	DIST	HASH	Qb jersey	Qb name
1					
1	1	10 M		7 Bassani	

Passo 12:

Qual a posição do campo a jogada está, para facilitar e utilizar o campo todo, quando um time está no seu campo de defesa, usamos a métrica no negativo, e no ataque sua posição no positivo, no caso aqui a jogada está na linha de 32 jardas na defesa, portanto -32 jardas.



	M	N	O	P	Q
1	DIST ▾	HASH ▾	Qb jersey ▾	Qb name ▾	YARD LN ▾
2					
3	10 M		7 Bassani		-32

Passo 13:

Qual foi o tipo de jogada, se foi corrida Rum, ou passe Pass, aqui é preciso dar play e aguardar o desenrolar da jogada, aqui passe Pass.



	M	N	O	P	Q	R
1	DIST ▾	HASH ▾	Qb jersey ▾	Qb name ▾	YARD LN ▾	PLAY TYPE ▾
2						
3	10 M		7 Bassani		-32	Pass

Passo 14:

Agora o resultado da jogada, é preciso aguardar o fim da jogada para preencher o resultado de passe completo, incompleto incompleto ou corrida como Rush, interceptação ou Funble. Aqui o resultado foi passe

incompleto.



N	O	P	Q	R	S
HASH	Qb jersey	Qb name	YARD LN	PLAY TYPE	RESULT
M	7	Bassani	-32	Pass	incomplete

Passo 15:

Ganho ou perda de jardas na jogada, como foi passe incompleto o ganho foi de 0 jardas

	M	N	O	P	Q	R	S	T
1	DIST	HASH	Qb jersey	Qb name	YARD LN	PLAY TYPE	RESULT	GN/LS
2								
3	10 M		7 Bassani		-32	Pass	incomplete	0

Passe 16:

Se houve pontuação na jogada e quem fez esse ponto preenchendo com o nome do time, nessa jogada não houve pontuação.

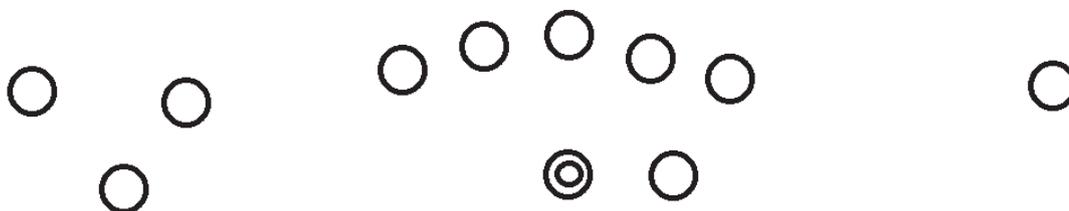
	S	T	U	V
1	RESULT	GN/LS	POINTS	POINT_TEAM
2			0	
3	incomplete	0	0	

Passo 17:

Qual formação o ataque levou a campo e afim de não gerar má interpretação, a ideia é preencher com um nome genérico como Form 1 e depois apresentar qual jogada vc se referiu como Form 1, e no campo offence play, se houve alguma movimentação pré snap que nessa jogada não houve



E a jogada Form1 é :



S	T	U	V	W	X
RESULT	GN/LS	POINTS	POINT_TEAM	OFF FORM	OFF PLAY
		0			
incomplete	0	0		Form 1	

Passo 18:

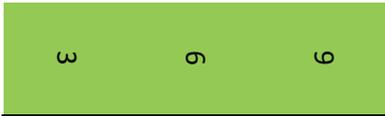
Aqui preenchemos quando for corrida, qual lado da corrida e seu gap, como foi passe fica sem preenchimento, caso tivesse seria a indicação de qual lado a corrida foi, esquerda L ou direita R e qual Gap, A, B, C OU D.

W	X	Y	Z	AA
OFF FORM	OFF PLAY	OFF STR	PLAY DIR	GAP
Form 1				

Passo 19:

Qual a zona do campo o passe foi feito, de 1 a 9 como a tabela abaixo

7	10 ao final
4	8
5 a 10	5
1	2
1 a 4	
jardas	



No caso desta jogada o passe foi incompleto mas o passe foi na zona 7



	W	X	Y	Z	AA	AB
1	OFF FORM ▾	OFF PLAY ▾	OFF STR ▾	PLAY DIR ▾	GAP ▾	PASS_ZONE ▾
2						
3	Form 1					7

Passo 20:

Aqui é onde preenchemos quem recebeu o passe ou correu com a bola, nesta jogada ninguém recebeu, mas o passe era pro jogador de numero 4 com o nome de Milrer.

	W	X	Y	Z	AA	AB	AC	AD
1	OFF FORM ▾	OFF PLAY ▾	OFF STR ▾	PLAY DIR ▾	GAP ▾	PASS_ZONE ▾	PLAYER_Jersey ▾	PLAYER_Name ▾
2								
3	Form 1					7		4 Milrer

Passo 21:

Nesse passo preenchemos qual o fornt da defesa, cobertura e se houve blitz, o fornt limitando aos jogadores de box, no ocaso dessa jogada 3-4, com ninguém em blitz, e um jogador no fundo do campo



AE	AF	AG
DEF FRONT ▾	COVERAGE ▾	BLITZ ▾
3-4	COVER 1	0

Passo 22:

Aqui coloca o nome e número do jogador de defesa que fez o tack ou interceptou o passe, nessa jogada não houve nenhuma dessas opções então não preenche.

AH	AI	AJ	AK
INTERCEPTED BY_Jersey ▾	INTERCEPTED BY_Name ▾	TACKLER1_Jersey ▾	TACKLER1_Name ▾

APÓS PREENCHIMENTO

Após encerrar a coleta dos dados na planilha é preciso enviar a planilha para a Bfa com o seguinte nome padronizado:

MANDANTE VS VISITANTE RODADA XX.xls

Depois disto é só aguardar o relatório da rodada.

Anexo II

RELATÓRIO DE JOGO

RELATOR: Thiago Garcia Molina Pinto

2022-07-26

JOGO

Data:	20/05/2022
MANDANTE:	CROCO
VISITANTE:	BS

PONTOS POR QUARTO

TEAM	1	2	3	4	FINAL
BS	3	6	3	0	9
CROCO	7	0	0	13	13

NÚMERO DE JARDAS TOTAIS DO ATAQUE

TEAM	TOTAL OFFENSE
BS	17
CROCO	133



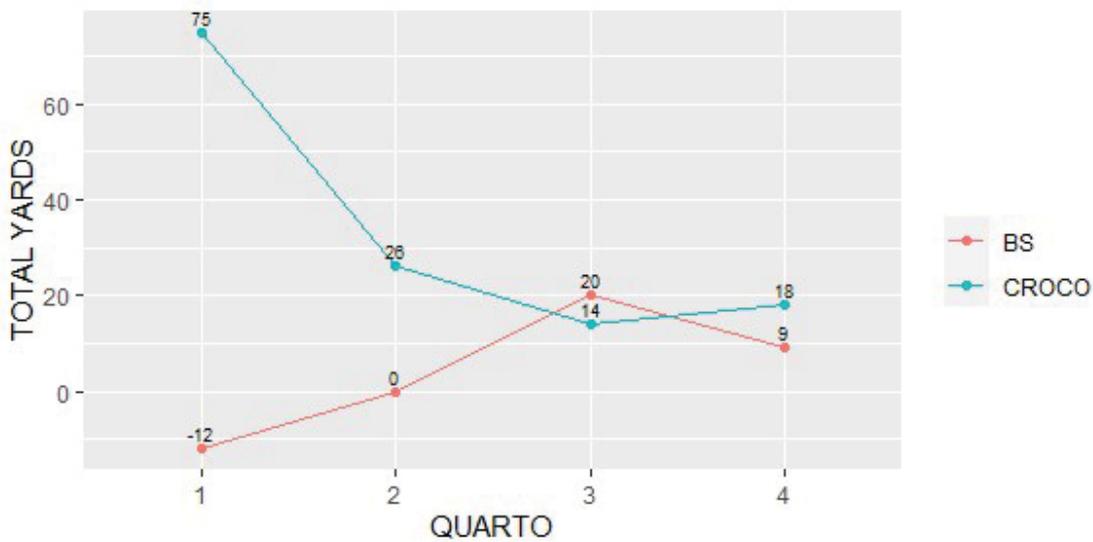
NÚMERO DE JARDAS DA FORMAÇÃO DE ATAQUE: CORRIDA

TEAM	YARDS
BS	-6
CROCO	47

NÚMERO DE JARDAS DA FORMAÇÃO DE ATAQUE: PASSE

TEAM	YARDS
BS	23
CROCO	91

NÚMERO DE JARDAS POR QUARTO POR EQUIPE



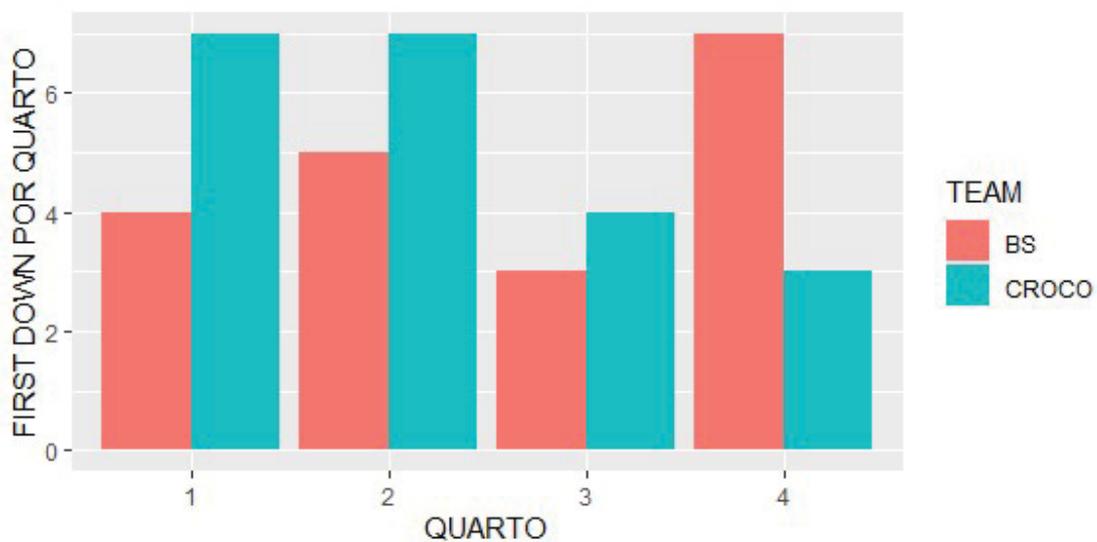
NÚMERO DE FIRST DOWN POR EQUIPE

TEAM	TOTAL FIRST DOWN
BS	19
CROCO	21

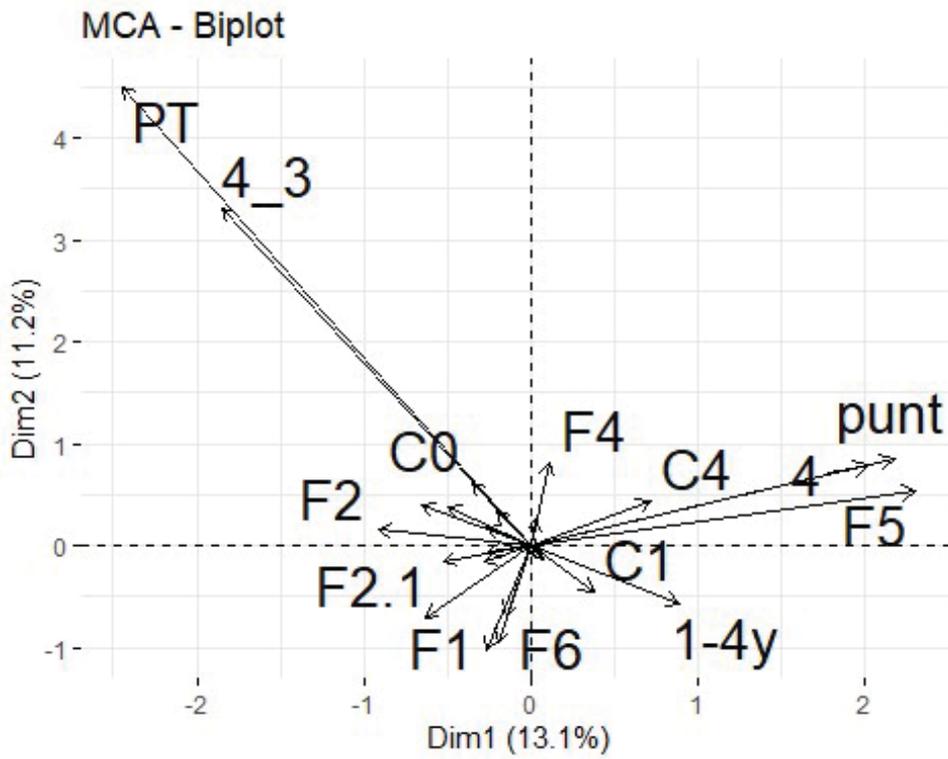
NÚMERO DE TURNOVERS POR EQUIPE

TURNOVER	TEAM	TOTAL
Fumble	BS	2
Interception	BS	1
Fumble	CROCO	0
Interception	CROCO	1

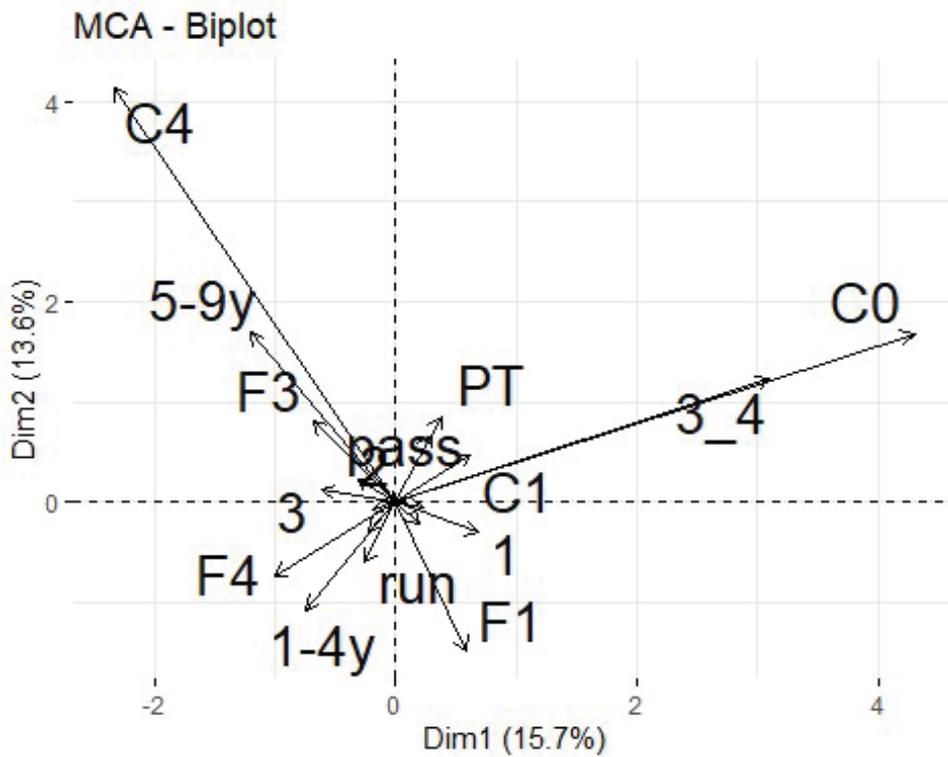
NÚMERO DE FIRST DOWN POR QUARTO POR EQUIPE



ANÁLISES MÚLTIPLAS: O QUE MAIS CONTRIBUIU PARA A PONTUAÇÃO DO ATAQUE DO TIME VENCEDOR?



ANÁLISES MÚLTIPLAS: O QUE MAIS CONTRIBUIU PARA A PONTUAÇÃO DA DEFESA DO TIME VENCEDOR?



##O QUE MAIS CONTRIBUIU PARA PASSES COMPLETOS DO TIME VENCEDOR?

