




Sports Tech World Series

Melbourne Hackathon

Date: 14 Nov 2020

2020 Hackathon Information Sheet

1.1 Introduction

Welcome to the 2020 Sports Tech World Series hackathon. Champion Data is pleased to be on board as a data provider and excited to see what some great minds with fresh ideas can discover and explain.

The hackathon is highly focused on the sport of Australian football, with all data coming from the past three seasons (2018-2020) of the elite men's competition, the AFL. There is a level of assumed knowledge that participants or teams will have a general understanding of the sport, its rules, and its terminology. Where there are any questions, there are staff from Champion Data and STWS on hand to answer throughout the weekend.

1.2 Data Overview

The data provided for the hackathon is a summarised version of our data set. The metrics provided are the same as those used by coaches, analysts, and commentators to tell the story of the game every week during the season. In the provided materials there are multiple files containing the source data for the hackathon, and some explanations to help you get started.

Included Files

- Info Sheet.PDF — This document.
- Metric Definitions — Overview of included metrics.
- Players 2018-2020.csv — Player Details for all AFL listed players in the 2018-2020 seasons.
- Fixtures 2018-2020.csv — Match Details for all matches in the 2018-2020 AFL seasons.
- Team Summary XXXX.csv — Summary Data at team level for the XXXX season.
- Player Summary XXXX.csv — Summary Data at player level for the XXXX season.

Included Metrics

The 'Metric Definitions' file contains a list of all metrics provided. This includes:

- 12 columns for Dimensions. Metrics are grouped by Match, Player/Squad, Zone and Quarter
- 36 columns for metrics that only appear in Player Summary files.
- 111 columns for metrics that only appear in the Squad Summary files.
- 192 columns for metrics that appear in both the Player and Squad Summary files.

Synthetic Metrics

The metrics in the summary files exclude those that we refer to as Synthetic metrics – where the definition is a function of other metrics. Some common examples of Synthetic metrics are:

- Disposal Efficiency = EFFECTIVE_DISPOSAL / DISPOSAL
- Player Tackle Efficiency = TACKLE / PLY_ATTEMPTED_TACKLE

- Squad Tackle Efficiency = TACKLE / SQUAD_ATTEMPTED_TACKLE
- Time on Ground Percentage = TOG_PLAYED_SECS / TOG_PERIOD_SECS
- Hitout Win Percentage = HITOUT / RUCK_CONTEST
- Hitout to Advantage Rate = HIT_OUT_TO_ADVANTAGE / HITOUT
- Mark Play On Percentage = MARK_PLAY_ON / MARK
- Pressure Factor = SQUAD_PRESSURE_POINTS / SQUAD_PRESSURE_CHANCE

This is not an exhaustive list, and exploration of multiple combinations of other metrics is encouraged.

Fixture

This file contains information about every match played from the 2018-2020 seasons. Most columns should be self-explanatory, but those that may require some explanation are:

- MATCH_TIME Local start time for the match, in 24-hour format.
- MATCH_TIME_MELB Start time for the match in the Melbourne time zone.
- *_SQUAD_TRAVEL L = Local (game played within team's home state). I = Interstate/Overseas
- ATTENDANCE Crowd size. Null where unknown.

Player Details

This file contains information about every listed player for the 2018-2020 seasons.

- POSITION_NAME A player's primary position across the entire season
- CAREER_MATCHES_PRE_2018 Included to provide the ability to track experience in games played.

Known Issues

Unfortunately, there are some known issues in the data set that were unable to be resolved before the hackathon.

- Stoppage Exit (STOP_EXIT_*) metrics currently do not include centre bounces.
- RUCK_CONTEST counts in the midfielder (DM and AM) are all assigned to the DM zone.
- Ruck contests for stoppages around the ground (BALL_UP_VERSUS and THROW_IN_VERSUS) are not stored by zone and appear in the X row.

Suggested Topics

This Hackathon is different to most. There is no fixed question, and no formula for assessing responses. A large proportion of the task of performance analysis is working with a blank state, asking questions of yourself and the data to gain a better understanding of how the game works and what sets teams and players apart.

Below are some suggested categories or questions that you may consider tackling, but these are a guide only. The winning response does not need to be revolutionary, conclusive, or narrow in focus. Telling the story of the game is the main priority.

Season

How was 2020 different to the previous two seasons? Noting that games were 16-minute quarters instead of the usual 20 minutes, most counting metrics should be 80% of the value from previous years.

- Were some features of play more/less common than in previous years?
- How did hubs or differences in travel schedules affect match outcomes and/or home-ground advantage?
- How did shortened time between games change the way the game was played?

Team

- Focusing on a single team, what did they do differently to other teams?
- What were their strengths?
- Weaknesses?

Player

- Focusing on a single player, what did they do differently to players in similar roles?
- What were their strengths?
- Weaknesses?
- Did anything change before/after a point in time during the year? Or compared to earlier years?

Match

- Focusing on a single match from the 2020 season, how did the game unfold?
- What did one team control or do to win the game?
- What made this game special?

Other

- Be creative, curious, and ambitious!