STM5IPL Initial Report

Interpreting key player and team performances into AFL using data visualizations

Michael Le

August 2024

Abstract

AFL, commonly known as Australian Rule Football, where the sport was drafted in 1859 by "Hammersley, Smith, Thompson and Wills", who constructed the "rules" which has been the sport has been famous for in over "140+ years or more". Where the first clubs were formed in "1860" and "1866" in South Australia and Queensland respectively. Later, continued to expand in Victoria in "1877", forming "eight-team leagues" alongside with South Australia, forming VFA/VFL in "1897" but was renamed to AFL in 1990 expanded more teams up to AFL 2011-2012 Season where two new teams GWS Giants and Gold Coast Suns were added into the AFL official roaster. To understand the "draft system" has been a main goal building relationships between teams and players. The "selection process" requires intensive "scouting reports" and "player assessments" to inform decision-making. When drafting players, we must consider under two conditions fairness and talent amongst across multiple football clubs. One method using recent and up to date data from "previous drafts" to understand patterns of player and team performance. Hence, will be valuable for the player and team success and satisfaction of "drafting outcomes".

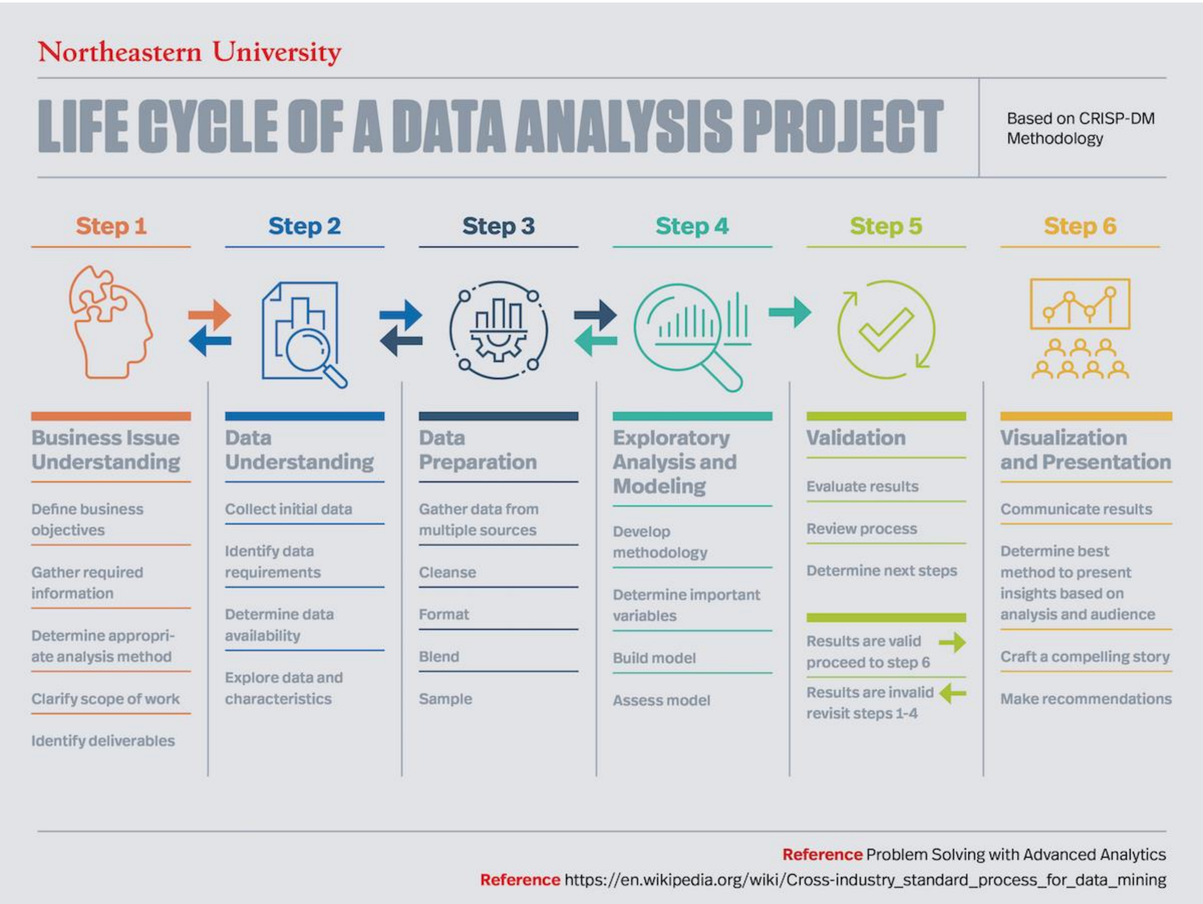

Figure 1. Life Cycle of a Data Analysis Project from Northeastern University

Introduction

The purpose of finding key performances in Australian Rules Football is to determine best talent among player candidates based on their “physical skills”, “past performance in matches”, “athletic abilities”, and “psychological assessments”. Most of the factors may include such as “speed”, “growth”, “endurance” and “strength” that is “quantifiable” to enable passionate and determined players to officially be drafted into the AFL Season. To enable decision making when it comes to their selection of players from the Sandringham Football Club at Foundry Athletic. The data was collected through email to discuss open-minded questions regarding key aspects to finding the best talent to provide an appropriate data visualization tool to demonstrate meaningful insights. Interpreting the “Business issue Understanding” and “Data Understanding” phases to determine on how well the team is performing. Includes, how well players are performing to easily process key talent and team performance during data analysis project. By understanding the data analytics lifecycle consists of six phases “based on the CRISP-DM methodology”. Include, the “business issue”, “understanding the data set, preparing the data, exploratory analysis, validation, visualization and presentation”. To execute a well-informed data analysis project report.

Methodology

During the “Data Preparation” phase **(seen Figure 1),** in performing an explanatory analysis of the data collected on Australian Football games, understanding two key concepts talent and recruitment when drafting only 12 out of the 170 players from the Sandringham Football Club will be drafted into AFL. This can be achieved trends through data preparation, is the process of transforming and cleaning raw data by “standardizing data formats, enriching source data, or removing outliers.” The data collection contains folders containing **(seen Figure 2 below)**. After further discussion with the data team and supervisor during the meeting. We decided to divide the number of folders to focus only the folders from **2km Results, Combine Testing Data, GPS Data, Gym Strength Testing Data, Height & Weight Data, Max Speed Data** and **Player Football Match Statistics**, without losing valuable information containing Excel Spreadsheets.

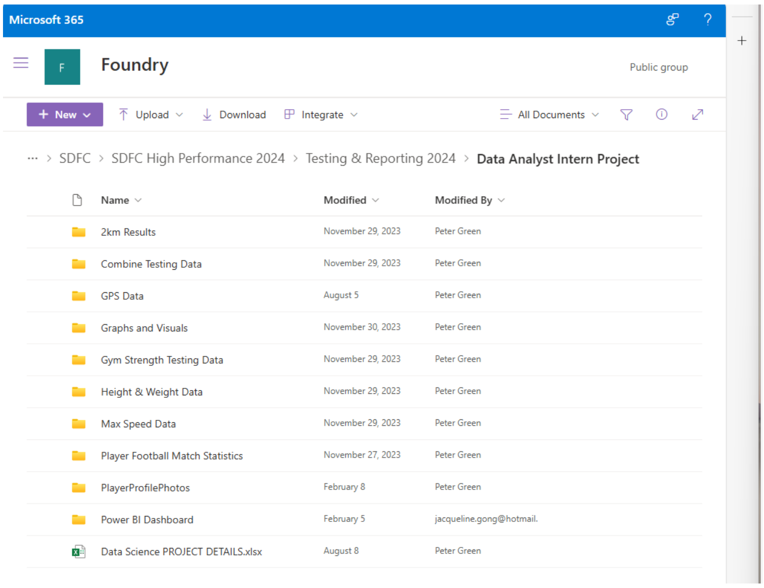

Figure 2. Supervisor's Folders containing all Player and Team Performance data from the Sandringham Football Club

Recommended technologies when necessarily used to perform these tasks using Microsoft Excel, combining multiple tables for each of the following folders without losing valuable information using Python programming language to merge. During the data collection process, up to the end of Week 3 of the placement, the data team have completed transformed and cleaned all raw data for **Max Speed Data, Player Football Match Statistics (in Figures 4,5 and 6 respectively), and Combine Testing Data confirmed via chat while the GPS Data, while folders 2km Results and Heights and Weights** are still in progress and lastly **Gym Strength Testing Data (in Figure 3)** is almost completed, where the cells are filled in red are the remaining entries needed to clean.

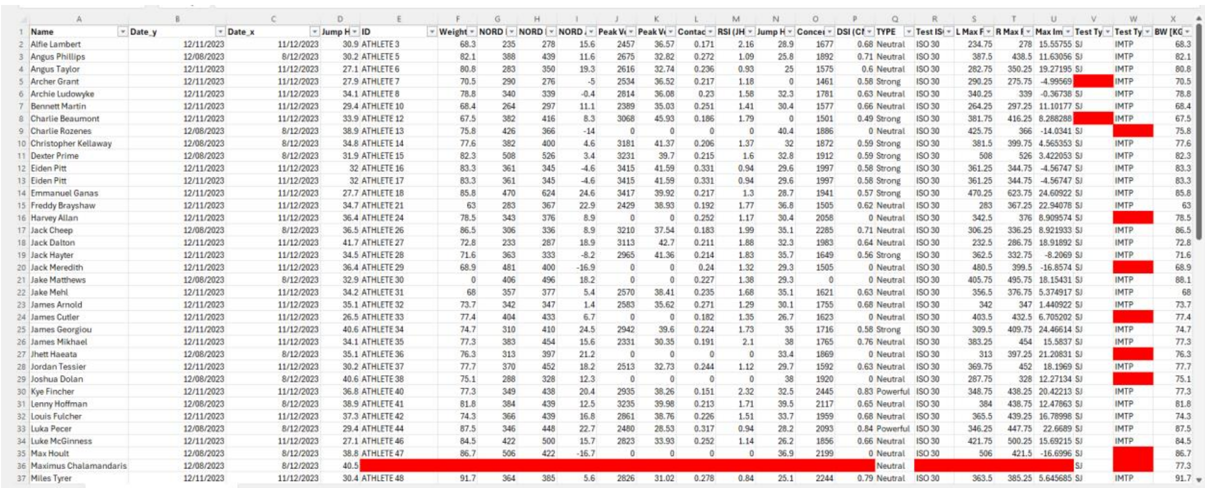

Figure 3. Transformed Gym Strength Testing Data (shown only partially of it)

	A	B	C	D	E	F	G	H	I
1	Player Na	Vmax	D(vmax)	T(vmax)	90% Vmax	D(90%)	T(90%)	Speed (km/hr)	
2	Billy McGe	8.77	38.69	5.4	7.89	17.99	3	31.57	
3	Brodie Fin	8.52	49.48	6.7	7.67	21.1	3.6	30.67	
4	Charlie Ed	8.86	39.6	5.3	7.97	18.77	2.9	31.9	
5	Charlie Ee	8.64	46.61	6.5	7.78	19.3	3.3	31.1	
6	Charlton H	9.12	58.25	7.5	8.21	22.33	3.2	32.83	
7	Cooper Lo	8.83	35.43	4.9	7.95	21.48	3.3	31.79	
8	Ethan Will	8.63	46.96	6.3	7.77	15.83	2.6	31.07	
9	Harrison (	8.6	31.22	4.6	7.74	22.66	3.6	30.96	
10	Harry Sulli	9.17	44.35	5.8	8.25	19.26	3	33.01	
11	Harvey Joh	8.49	38.69	5.5	7.64	18.66	3.1	30.56	
12	Jack Daltc	8.54	24.82	3.7	7.69	16.29	2.7	30.74	
13	Jeremy a'E	8.61	29.08	4.3	7.75	18.77	3.1	31	
14	Joshua Dc	9.19	37.94	5.1	8.27	23.57	3.5	33.08	
15	Joshua Dc	8.97	42.8	5.8	8.07	17.35	2.9	32.29	
16	Justin Kro	8.63	27.83	4.1	7.77	16.8	2.8	31.07	
17	Lachlan V	8.59	41.12	5.8	7.73	21.92	3.5	30.92	
18	Levi Ashcr	9.16	47.32	6.2	8.24	19.4	3.1	32.98	
19	Levi Young	8.16	37.65	5.6	7.34	17.74	3.1	29.38	
20	Luke Kenn	9.03	37.1	5	8.13	20.45	3.2	32.51	
21	Mason Szo	9.28	44.42	5.9	8.35	24.32	3.7	33.41	
22	Mitchell K	8.68	43.31	5.9	7.81	19.14	3.1	31.25	
23	Nathan Su	8.43	34.3	4.9	7.81	20.71	3.8	30.35	
24	Nathaniel	8.9	41.36	5.6	8.01	19.51	3.1	32.04	
25	Ned Magir	8.92	44.97	6.2	8.03	21.37	3.5	32.11	
26	Ollie Murp	8.1	42.06	5.9	7.29	17.23	2.8	29.16	
27	Ollie Wart	8.74	39.49	5.4	7.87	18.99	3	31.46	
28	Riely Colli	9.07	44.56	5.9	8.16	20.55	3.2	32.65	
29	Ryley Sanc	9.2	43.35	5.8	8.28	18.89	3.1	33.12	
30	Samuel Li	8.39	42.98	6.1	7.55	18.34	3.1	30.2	
31	Taj Hottor	8.54	46.29	6.3	7.69	16.92	2.8	30.74	
32	Timothy P.	8.75	47.27	6.4	7.88	20.59	3.3	31.5	
33	Vigo Viser	8.37	40.53	5.6	7.53	16.56	2.7	30.13	
34	William Br	8.57	37.96	5.4	7.71	17.66	3	30.85	
35	William N	9.07	51.16	6.8	8.16	20.58	3.4	32.65	
36									
37	Average	8.75	41.15	5.65	7.88	19.44	3.15	31.5	

Transformed_SDFC_Speed_Data

Figure 4. Transformed Sandringham Dragons Football Club Max Speed Data

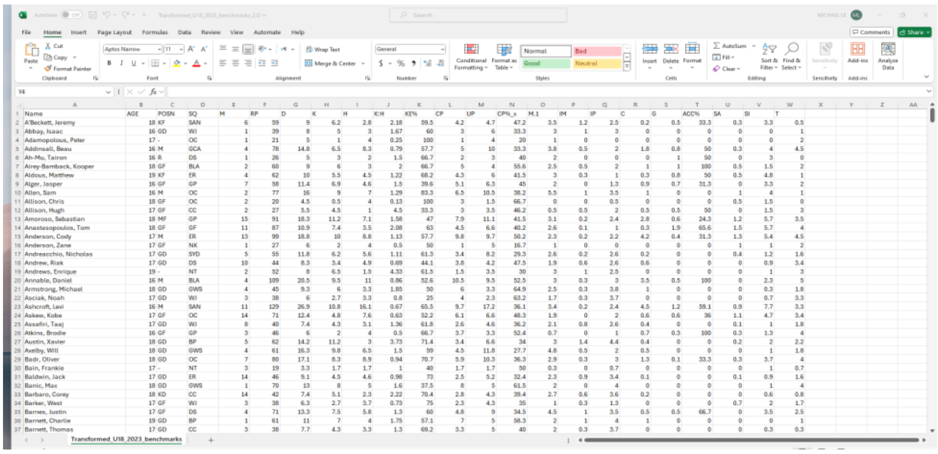

Figure 5. Transformed Player Football Match Statistics: Under 18-year boys 2023 BenchMarks

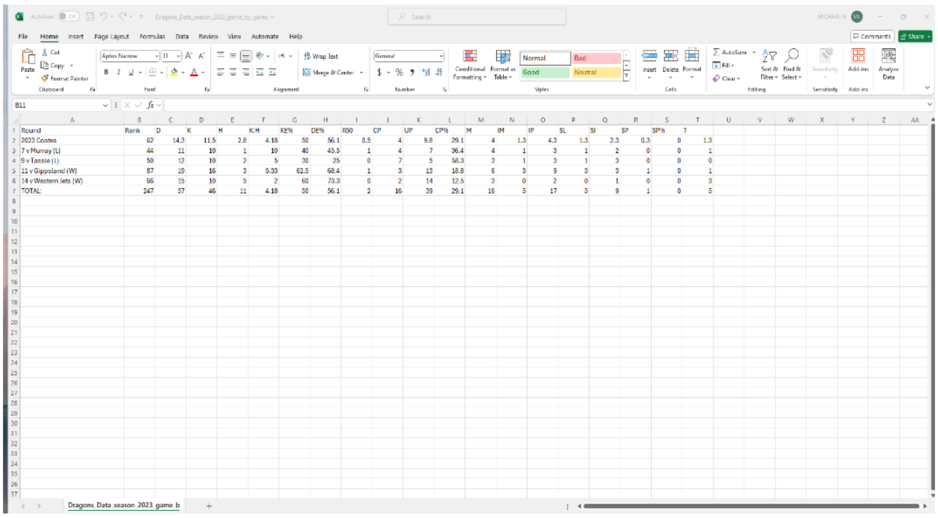

Figure 6. Transformed Player Football Match Statistics: Dragons Data season 2023 game by game

After enhancing and improving the data, we process only the "Exploratory Data Analysis", since we would not be modelling during the data analysis project if necessary, depending on the supervisor's proposal. We wanted to recognise and understand patterns of data using appropriate libraries to perform data manipulation, aggregation methods for statistical analysis such as computing their averages, standard deviations and z-scores to compute for the final dashboard. The supervisor prefers key statistics for computing average values from the GPS Data, 2km Results, and Combine Testing Data. **(Seen in Figure 7 and 8)** below, taking the maximum value as the supervisor proposed for each of the columns to solve for following z-scores, which will be later be used to form into a Rader Chart **(See Figure 11 below)** into the final dashboard.

1. Z-Score Formula

The **Z-Score** formula is used to measure how many standard deviations a specific data point is from the mean of the distribution. It's used to standardize scores on a common scale:

$$z = \frac{x - \mu}{\sigma}$$

Where:

- x is a single data point.
- μ is the mean of the population.
- σ is the standard deviation of the population.

Figure 7. Z-Score Formula

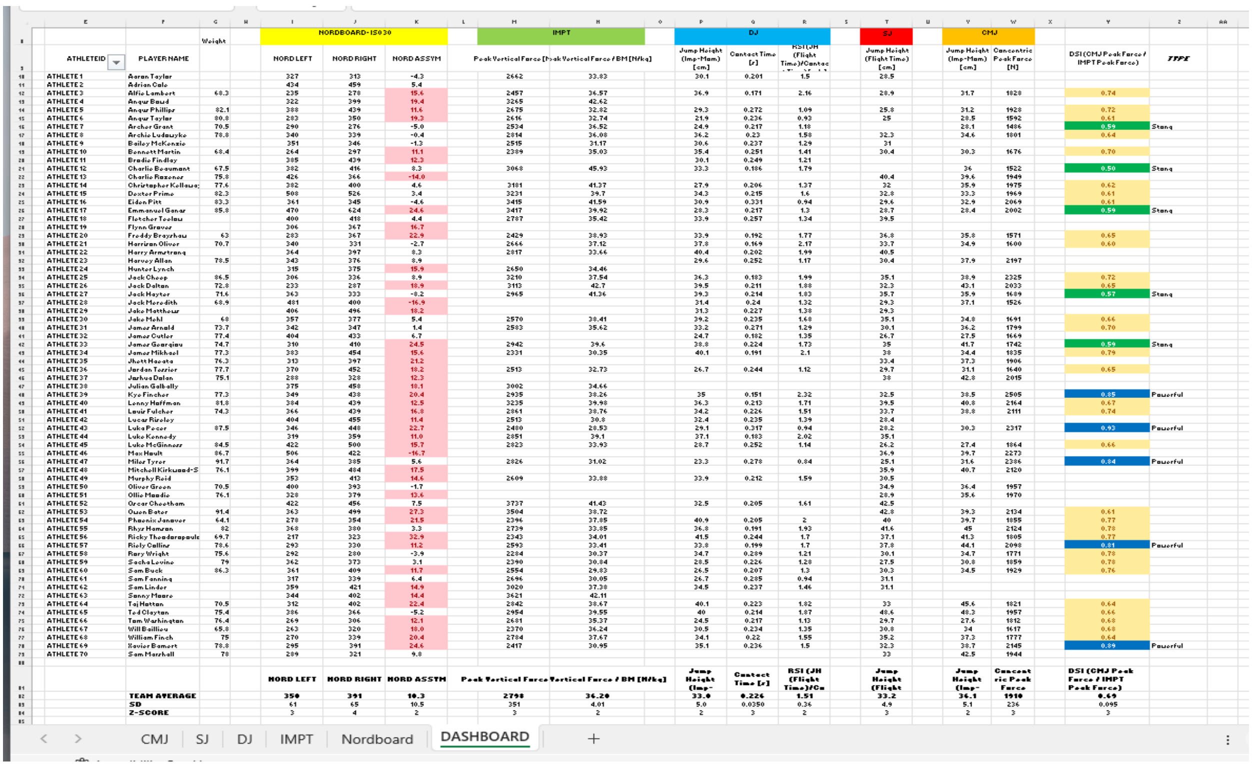

Figure 8. Gym Strength Testing Data: Dashboard Spreadsheet

After careful selection, intensive preparation and exploratory analysis from the First Four Phases during the data analysis project. The data team wanted to make sure all "the correct information" is obtained during the deliverables, ensure we do not have to repeat through an iterative process. Maintain all quality and security issues resolved to guarantee better outcomes when it comes to data visualizations into a dashboard using one of the two data visualization tools from either Tableau or Power-BI. After preparation and exploratory analysis stages is completed, the supervisor initialized what our final dashboards will be, shown in the figures below. During the placement, proposals may change periodically over time, since were still waiting for all data collection is completely finalised, likely due to the expansion for new recruits towards both boys and girls under 16 years of age. This will likely improve our results in terms of GPS averages, and Team GPS averages and comparisons **(shown in Figure 10.)** to "becoming more accurate" avoiding bias.

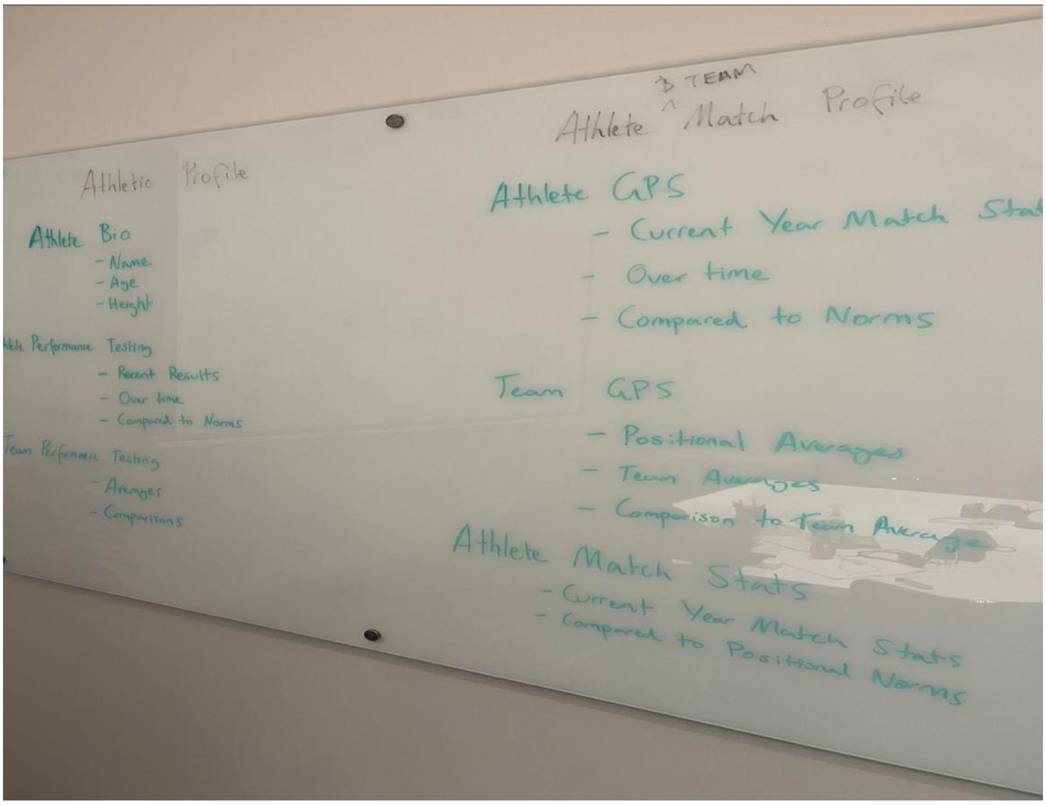

Figure 10. Supervisor's initial dashboard 1. Took place during the first day Week 1 during the placement

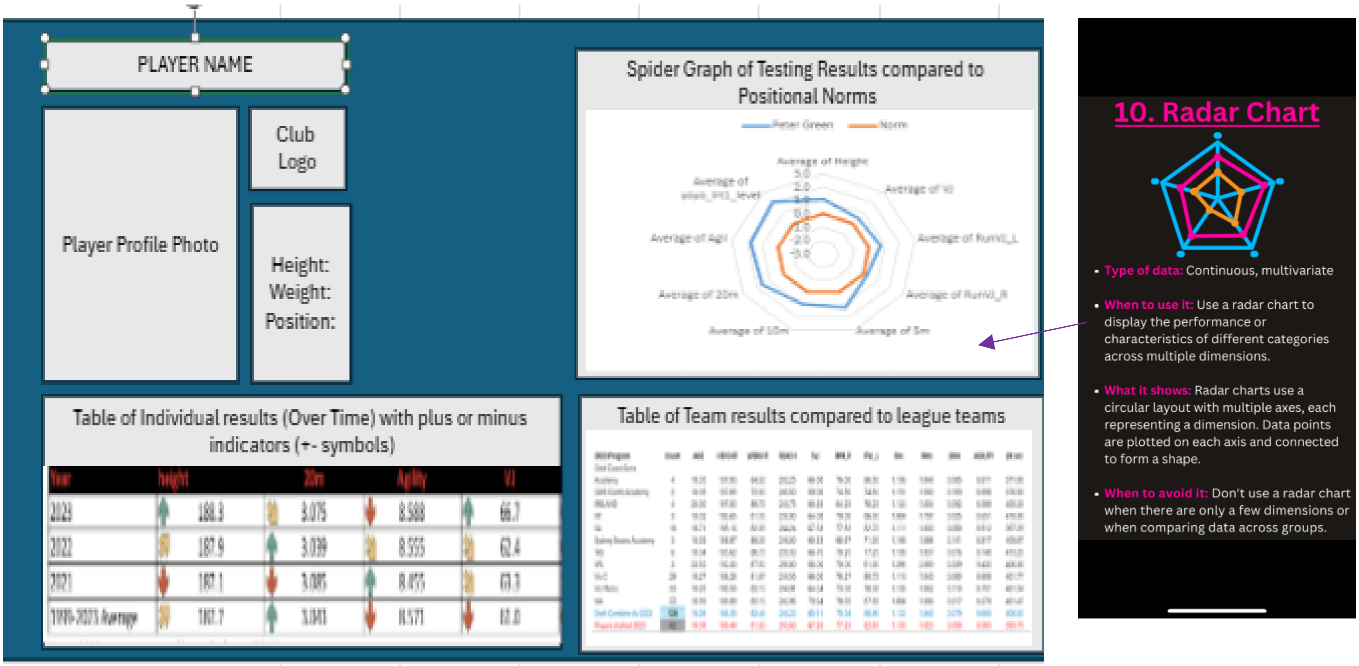

Figure 11. Supervisor's initial dashboard 2. Shown during via live chat during Week 3 of the placement.

Figure 12. Bar chart displaying key metrics for player performance from GPS data using PowerBI. Shown during via live chat during Week 3

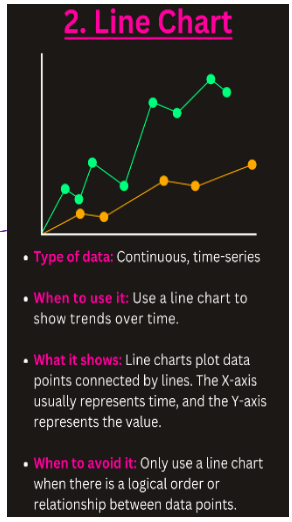

Figure 13. Line chart displaying key metrics for player performance growth from July 2023 to January 2024 using PowerBI. Shown during via live chat during Week 3

Displaying in onto the dashboards **(from Figures 10-13 shown above)** to tell relevant “highlights” of that story for the organization to secure trust and integrity that is required to accomplish their goals. To clarify better interpretations in terms of using the appropriate data to enable decision-making to drive key performances more effectively without being misinterpreted in a kindly manner. Without reinforcing bias and always consider fairness for the clients and what is required to the overall business to succeed for players to be eligible to be drafted into AFL. The organization at Foundry Athletic proposed one of two powerful tools in their domain to enable stakeholders and organizations to interact data. Deciding from Tableau or Microsoft Power BI which data and business analysts frequently use on the daily basis **(seen in Figure 14).**

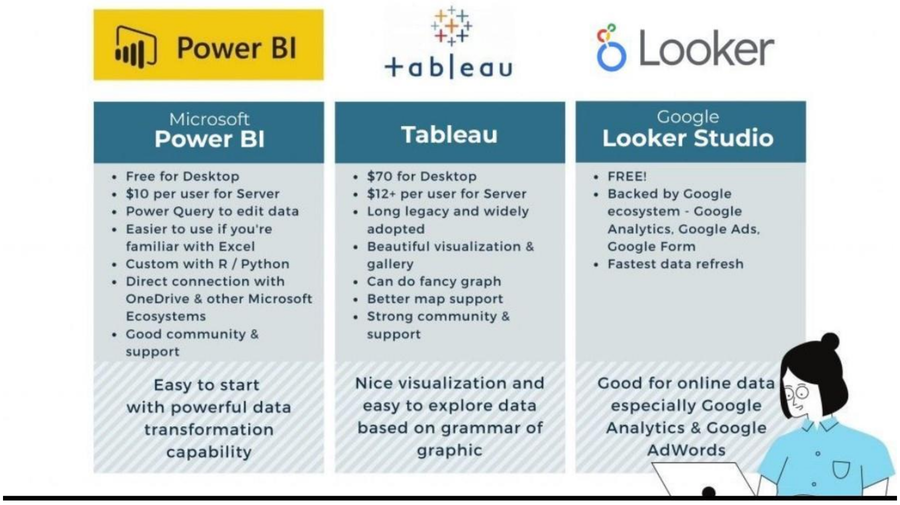

Figure 14. Comparison of data visualization tools

During the placement, when after the first phases of the data analysis project is completed, the data team proposed to use Microsoft Power-BI since it is free and convenient to use, making it suitable. Despite costs for each server are in USD not AUD, the data team will eventually have to pay these servers upfront to get familiarity and adapt with the new data visualization tool. Assuming the validation of the data is updated leaving no flaws or bias of our data during the 5th phase of the project. It has "the ability to analyze and share large datasets" to businesses. Enhances "sharing and collaboration aspects of business intelligence", like Tableau, Power-BI can integrate Microsoft Excel spreadsheets which it can be shared with colleagues and stakeholders effortlessly when creating dashboards and reports. Adjustments can be made with "the data being updated in real-time; businesses can identify and address issues promptly".

Overall, it enables coaches and stakeholders to interact with the visualizations through "charts, graphs and maps that brings data to life". To be satisfied with high expectations with their results in terms of finding key talent and recruitment from player and team performances. The reason the data team prefer Microsoft Power BI over Tableau because we want to perform more reports, since we wanted 2 pages at least to showcase the player key statistics and team performance statistics. Available "in Power BI Desktop" since we can build and view reports. Able to "filter data" to showcase data manipulation and aggregations more effectively and efficiently. Without using programming languages such as Python or R which can be integrated into Power-BI if necessarily to update more data visualizations. Our goal for the data analysis project was to allow more people to access more freedom of how they understand and handle the data to a huger audience and businesses overall. To conclude how the data team can best stories to clients about our data findings of how we can implement further strategies using player and team performances. To allow recruitment and finding key talent more effectively when drafting players into the Australian Football League for the Sandringham Football Club at Foundry Athletic.

References

- **AFL. (n.d.). *History of football - AFL.com.au*. afl.com.au. https://www.afl.com.au/about-afl/history**
- **Collins, B. (2019, November 24). *How do I get drafted by an AFL club? Am I eligible?* afl.com.au. https://www.afl.com.au/news/149356/how-do-i-get-drafted-by-an-afl-club-am-i-eligible**
- **maggiesMSFT. (2023, October 11). *Intro to dashboards for power BI designers - Power BI*. Microsoft Learn: Build skills that open doors in your career. https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboards**
- **McLean, J. (n.d.). *The Science of Success: How AFL Teams Scout and Develop Emerging Talent*. https://preparelikeapro.com/the-science-of-success-how-afl-teams-scout-and-develop-emerging-talent**
- **Staff, G. P. (2023, April 26). *Understanding the lifecycle of a data analysis project*. Graduate Blog. https://graduate.northeastern.edu/resources/data-analysis-project-lifecycle/**

- ***What is data preparation? Processes and example*. Talend - A Leader in Data Integration & Data Integrity. https://www.talend.com/resources/what-is-data-preparation**
- **Charlesworth Author Services. (2022, May 26). *The Importance of Large Sample Sizes in Research | CW Authors*. Www.cwauthors.com; Charlesworth Author Services. https://www.cwauthors.com/article/importance-of-having-large-sample-sizes-for-research**
- ***12 Top Benefits of Power BI - Why Use It In BI? | Multishoring*. (2024, April 11). Multishoring.com. https://multishoring.com/blog/benefits-of-power-bi/**

END OF STM5IPL INITIAL REPORT