



# PROJECT PROPOSAL

GAINING INSIGHTS INTO EVENT DATA - 2022 NRL SEASON

By **SPORTALYTICS** @ UNSW (Data3001):

- Dylan Upton (z5308844)
- Nikhil Khatri (z5311164)
- Kyle Brown (z5308312)
- Diya Patel (z5311113)
- Saumil Talati (z5309456)

## Mission Statement

SPORTALYTICS mission is to provide insights into sporting event data by using leading computational and statistical data science techniques. We strive to create value for our clients by presenting these new insights in a way so they can make informed decisions about future improvements to their sports.

---

## BACKGROUND

The flow of an NRL game is critical for its entertainment level for the spectators. Great Rugby League games are those with the most amount of flow which is defined as the least number of stoppages. Whereby the ruck area is the most liable to decrease the flow of the game which is what happens around the area surrounding a tackled player. This is then used to recover from the tackle and restructure the team's formation. For this reason, ruck area rules have been changed multiple times in recent years such as any ruck infringement or offside conceded by the defending team within the attacking team's 0-40m zone would be penalised instead of a set restart. This is done to deter ruck infringements by the defending team as a penalty would result in a larger loss of field position or an increase in scoring opportunities. This loss in field position may disrupt the flow of the game and dissatisfy consumers creating a negative impact on the NRL.

Thus, it is important that the NRL are informed of the repercussions of these rule changes, and what this means Rugby League. For this reason, we at SPORTALYTICS have established and created this project and the ensuing report that aims to explore the 2022 rule change by analysing and creating conclusions to the provided data to understand how the rule changes to the ruck area in the past couple of years has made an impact on the game. Another aim we at SPORTALYTICS look to examine is determining how the value of field position over possession has changed in previous years, and what this means for the outcome of games.

## Previous work

Due to the rule change being for the current 2022 season, SPORTALYTICS understand that no other work has been done on this problem specifically. We do however acknowledge other NRL analysis that may assist us:

- Factors affecting performance in professional Rugby League - UTS Thesis by Thomas Kempton. Bachelor of Human Movement (<https://opus.lib.uts.edu.au/bitstream/10453/43430/7/02whole.pdf>)
- Conceptualizing Rugby League Performance – Sports Medicine Article by Tannath J. Scott. (<https://sportsmedicine-open.springeropen.com/articles/10.1186/s40798-021-00375-x#citeas>)
- The Rule change affecting time in play that's rarely talked about by League Eye Test (<https://www.rugbyleagueeyetest.com/2020/08/18/nrl-round-14-notes-and-trends/>)

The NRL are a private company, so their data is not public. As a result, very little analysis has been conducted outside of the NRL. We have however considered public news articles to gain a better understanding of Rugby League and the ruck area rules to accommodate the lack of professional work done on this subject matter. These are as follows:

- What is the Ruck in Rugby League? - News article by Alexander Wolf (<https://fluentrugby.com/what-is-the-ruck-in-rugby-league/>)
- Experts view: Who'll benefit from six-again rule – News article by NRL.com (<https://www.nrl.com/news/2020/05/28/experts-view-wholl-benefit-from-six-again-rule/>)
- The top five rule changes that changed the NRL – News article by ZeroTackle (<https://www.zerotackle.com/the-top-five-rule-changes-that-changed-the-nrl-122366/#:~:text=So%2C%20instead%20of%20a%20penalty,to%20do%20with%20the%20ruck>)

---

## OBJECTIVES

Our objectives are:

- Outline and understand detailed patterns, trends, and the effect of ruck infringement rule changes through visualisation and detailed extraction of data findings. Then examining how this affects the game such as distribution with respect to field position and the importance of field position versus possession
- Produce a model, predicting the expected win percentage of a team at any point in the game to highlight the changes in play style due to the rule changes over time. We can then further outline trends in the data such as the number of blow out wins and correlate this to the rule changes.
- The mathematical model alongside the analysed data will serve as a platform to determine the overall rankings of players and teams in the sport as well as identifying the key factors affecting win percentage and scoring.
- Outline a recommended course of action to the NRL.

## SCOPE

The main deliverables of this project are a report and associated presentation that will outline the team's analysis of the proposed NRL request. It will be our goal to present our findings concisely so they can be understood by both technical and non-technical people. The report and presentation will be developed based on the following data provided by the NRL:

- CSV Event Data Files for 2019, 2020, 2021 and 2022 NRL Seasons
  - Each NRL game event is captured by a third party and described in a single row in CSV format.
- Event Data Readme File
  - Documentation on variables in the captured CSV files in addition to an explanation on the field position calculation process and zones. Examples are also provided.
- Match CSV Data
  - Data Captured specifically on matches

From the NRL provided data, our analysis will focus on the ruck area events which detail the type of event, the position of these on the field and the subsequent events that can help us draw conclusions from the data. These events will include:

- EventCodes, specifically
  - A ruck infringement occurring (Code = RINS/RIND)
  - Penalty (Code = PABD)
  - A try being scored (Code = TRY)
  - The end of a team's set (Code = EC1S, EC2S, EC3S, EC4S, EC5S)
  - The end of a team's possession (Code = ENPO)

Qualifier/ QualifierName

- Each of the eight qualifiers give a linear description of the event. Most importantly qualifier1 outlines "Ruck Infringement", the area we are focusing on.
- Xm\YmPlayer, TeamAScores/ TeamBScores, SeqNumber, SeasonId, Set/Tackle and MatchId.

**Note:** As we work through the exploratory analysis, these key features are subject to change.

As part of our analysis, we will need to make certain assumptions. As seen below:

- How we choose to analyse the data files, where we will draw on our knowledge of the domain and then escalate any uncertainties to our NRL contact. Since we are comparing the rule changes and their effect over the years, we will assume that certain statistics and trends changed over the seasons are only related to ruck rule changes.
- COVID19 and its effect on the data. Post pandemic, NRL games were only affected off-field with players testing positive and unable to play games, but when teams took to the field, games were no different to previous years.

The success of SPORTALYTICS project will be assessed by a series of factors:

- Have we found sufficient evidence that addresses the elected problem? If so, are the findings significant? This allows us to measure highly correlated trends in the data.
- Can we create a visualisation, effectively presenting our results?
- Model can be tested using a multitude of ways, some including AUC (Area Under Curve), RMSE (Root Mean Squared Error), ROC (Receiver Operating Characteristic) and F1 Score.
- The strength of the recommendation provided which can be assessed with compelling evidence.

We expect each of the model's performance tests to return high results, splitting the data into a training and testing set to properly explore the success of our model.

Given the scope and importance of the project at hand, we at SPORTALYTICS must consider issues and establish solutions that could interfere or hinder the progress of satisfying the elected aims.

Issues that may affect the project are:

- Unfinished project due to a large scope managed by use of a Gantt chart to plan and follow strict deadlines.
- Uneven and unfair workload within the team causing the loss of expertise in certain areas managed by use of a Gantt chart with assigned roles and deliverables which are required to progress to later stages of the project.
- The lack of understanding of rugby specific knowledge leading to errors in data analysis and modelling countered by individuals with NRL knowledge to do tasks which require domain knowledge.
- Difficulty of communication due to Covid-19 restrictions whereby we have planned frequent meetings to allow for proper team communication and easier means of teamwork.

# PROJECT PLAN AND TIMETABLE

## Approach

Our approach is structured with each member completing assigned tasks in a given timeframe. All team members will come together weekly to make changes and progress the report. Teamwork will be critical to support each member with specific tasks, allocated on an individual's experience and skills. Our approach will be broken down into the following 5 phases:

- **Research** (Due 5<sup>th</sup> of October)

Collectively the team will decide on the problem. This will allow us to identify if we will be predicting through statistical modelling or to only extrapolate findings from the data. Research will then be required into the domain (NRL Rugby League), specifically ruck infringement rule changes. **Dylan** and **Nikhil** will lead the work on this section. Dylan is an avid NRL fan with extensive domain knowledge. Nikhil is a long-term Maroons fan who has experienced years of NRL culture.

- **Data Cleaning and Exploratory Analysis** (Due 1<sup>st</sup> of November)

This stage involves structuring the data in an accessible manner allowing for visualisation and modelling of the project. Using exploratory data analysis, we will create histograms or similar techniques to determine the relevancy of certain variables for future use in modelling and our visualisation process focused on different objectives. We plan to export event data from 2019 to the present to our local systems. By choosing Python, we are enabling the use of SQL and any relevant libraries to help to filter out irrelevant data. Early plans are that we will consider using the likes of GitHub or alternatives to ensure version control and collaboration. **Dylan** and **Kyle** were specifically chosen for this task through their recent expertise and proficiency in Python and associated manipulation of databases having completed a recent subject successfully. Furthermore, they both have external working experience in this area where they can draw inspiration from.

- **Modelling** (Due 1<sup>st</sup> of November)

Being one of our primary objectives, we plan to create a mathematical model using the determined relevant variables from the data cleaning stage, predicting win percentages based on differences in field position and possession. From the planning stage, we see that this may involve a single model that is applicable to every year and game regardless of the rule changes to ruck infringements to create a wider assessment of the importance of these changes. Using our exploratory data analysis, we will select a model to represent our data which will then be trained and built up through further analysis. This will be tested at the half-way point of the existing games to determine correctness with the actual result of these historic games. **Diya**, having studied a mathematical major in her time at UNSW is highly skilled in mathematical modelling. Similarly, **Saumil** has professional experience in modelling which will assist in further building up the model. Finally, **Nikhil** can inspire a more complex model through his personal experience of machine learning.

- **Visualisation** (Due 15<sup>th</sup> of November)

Visualization of the data to determine the impact of ruck infringement changes over the period of 2019-2022 will be critical. This will be achieved by using tools such as Tableau and the Python library MatLib to create many different models including: heatmaps and breakdowns of ruck infringements with respect to field position. Furthermore, using the algorithm created in the modelling stage, we will look to create graphs and tables to determine the rankings of individual players and teams. This will also be completed by the duo of **Diya** and **Saumil** who have both worked on the creation of dashboards and visual data representations. Furthermore, **Kyle** and **Nikhil** will provide inspiration in their use of Tableau from past online courses.

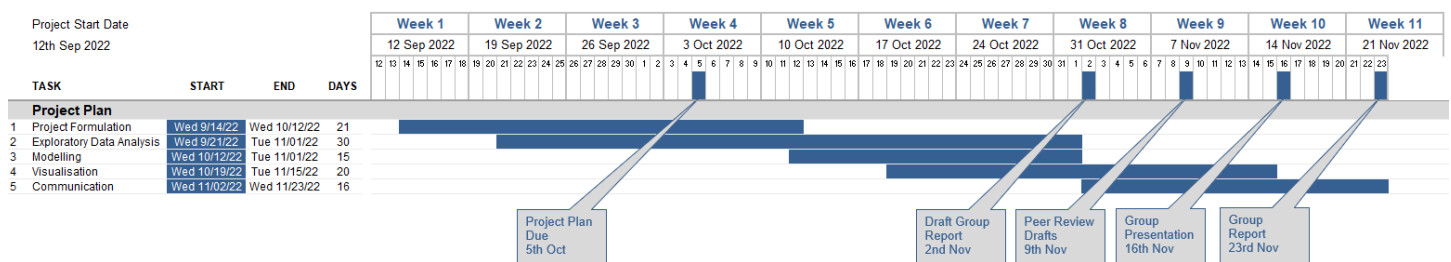
- **Communication** (Due 23<sup>rd</sup> of November)

Led by **Dylan**, all group members will contribute to the creation of the report by adding their insights from the understanding each has gained from their respective roles in the project. The proposal will be created using Microsoft Word with our presentation to be created using either Microsoft PowerPoint or Overleaf.

## Project Schedule

By: Sportalytics

Project Start Date  
12th Sep 2022



## REFERENCES

- Kempton, T. (2016) *Factors affecting performance in professional Rugby League*. UTS. Available at: <https://opus.lib.uts.edu.au/bitstream/10453/43430/7/02whole.pdf>
- Scott, T.J. et al. (2021) *Conceptualising rugby league performance within an Ecological Dynamics Framework: Providing direction for player preparation and development - sports medicine - open*, SpringerOpen. Springer International Publishing. Available at: <https://sportsmedicine-open.springeropen.com/articles/10.1186/s40798-021-00375-x#citeas>
- Test, L.E. (2021) *The rule change affecting time in play that's rarely talked about - NRL round 14 2020 stats and Trends*, *The Rugby League Eye Test*. Available at: <https://www.rugbyleagueeyetest.com/2020/08/18/nrl-round-14-notes-and-trends/>
- NRL.com. (2020) *Experts view: Who'll benefit from six-again rule*, *National Rugby League*. NRL. Available at: <https://www.nrl.com/news/2020/05/28/experts-view-wholl-benefit-from-six-again-rule/>
- Zero Tackle (2022) *The top five rule changes that changed the NRL*, *Zero Tackle*. Available at: <https://www.zerotackle.com/the-top-five-rule-changes-that-changed-the-nrl-122366/#:~:text=So%2C%20instead%20of%20a%20penalty,to%20do%20with%20the%20ruck>
- Wolf, A. (2021) *What is the ruck in rugby league?* *FluentRugby*. FluentRugby. Available at: <https://fluentrugby.com/what-is-the-ruck-in-rugby-league/>