# **Applied Statistics with R Project Proposal**

By Heber Ng, <insert your name here>

## **Section 1: Abstract / Introduction**

With the rise of E-Sports and the revenue that comes with being a part of the ecosystem, more and more E-Sports teams are looking to leverage every advantage to gain a competitive edge over the opponents in hopes to become major players in the arena, hence gaining profits through various sponsorships, endorsements and merchandises. These attempts in includes specialized training and competition equipment, sports management theories, player drafting etc. In recent years, there is also a rise in attention paid towards the use of data analytics to analyze things such as player performance and competitors’ match strategies.

Particularly, modern games are showing a trend of being willing to allow access to the data it collects during play sessions. Major games, such as Player Unknown Battlegrounds (PUBG), Guild Wars 2, CS:GO and Dota 2 are providing official tools and APIs to retrieve player and match statistics collected by their game servers. This offers the public an opportunity to look at the games they play in a manner that was previously impossible. Already, a number of sites has had created riding on this trend, offering insights such as each games’ respective meta-game and stats based on the data available.

These two trends combine to provide a unique opportunity for data analyst to step in: provided statistical analysis on gameplay data for E-Sports teams based on the publicly available. The level of analysis possible includes descriptive, inferential and predictive, as the nature of E-Sports lends itself readily towards the application of data analysis techniques; the digital competition environment allows all sorts of metrics to be collected, stored and retrieved conveniently, providing a massive source of data to work with.

## **Section 2: Overall Concept (Purpose and Key Idea, Target Audience, Impact etc)**

The purpose of this project is to create a prototype data product that can provide statistical analysis on available Dota 2 gameplay metrics. The target audience of the product includes professional E-Sports teams and players/teams trying to make the transition from amateur to pro level.

Dota 2 is selected as the game to base the product on because it is currently the biggest E-Sports game available, both in terms of monetary and data. It has the biggest prize pool for 5 years in running (since 2014), attracting many teams from many countries to compete in its preliminaries, thereby providing a ready audience for the product. Additionally, its popularity ensures that a large number of games were played, creating a vast trove of data to base any analysis on. Lastly, the game provides an official API to pull data off the game servers, giving any analyst the ease of access to the data.

While a number of similar products has already being developed such as [OpenDota](https://www.opendota.com/) and [DotaBuff](https://www.dotabuff.com/), a deeper look on both sites will show that they are mainly focused on descriptive aspect of the data, such as player performance metrics (number of kills/deaths/assist), match metrics (team gold/experience advantage, event logs) and overall meta-game metrics (hero pick/ban percentiles, win/loss ratios etc.). Very little is shown to be done on the inferential and predictive aspects of data analysis.

The aim of the data product will be to apply inferential statistics techniques on the existing descriptive base to provide comprehensive insights on any match. Particularly, the product aims to provide useful information for esports teams, focusing on aspects such as:

* The Dota 2 version of “Moneyball”-quise sabermetrics of professional players and matches
* Team composition, hero picks and bans predictions
* Lane efficiency, team advantage analysis throughout the course of the match
* Player movement and ward placement patterns

We believe that these statistics will be useful to E-Sports teams in both drafting its players, doing post-game reviews and pre-game strategy planning.

**Section 3: Data Sources (Detailed description of data sources, fields and transformed variables)**

The data source used will be from OpenDota API, the same API that OpenDota is using for its own site. The product will be build using R Shiny and the R package [ROpenDota](https://cran.r-project.org/web/packages/ROpenDota/index.html) to access the OpenDota API. The documentation on the data and variables available through the API can be found [here](https://docs.opendota.com/).

Based on the details on the data variables provided by the OpenDota API documentation, many fields can be readily used for the features planned, such as:

**Professional Match Metrics:** The response json of GET/matches/{match\_id} readily provides a number of variables such as participating team and players, duration, each team’s gold/experience advantage etc.

**Professional Player Metrics:** Likewise, the response json of GET/matches/{match\_id} contains a {players} object list, which contains the data of player’s performance for that match. Summarizing and aggregating this data will allow the product to deliver various player performance metrics.

**Team Composition and Picks/Bans:** The {draft\_timings} variable in GET/matches/{match\_id} stores the hero drafting decision process for the match, which, when aggregated, provides an overview on the hero composition preferences of each team.

The data provided by the API is expansive, as such, throughout the course of development, iterative reviews will be done on the data available and the features adjusted accordingly.

## **Section 4: Specific Methodology (Statistical techniques and analytics employed)**

## **Section 5: Conclusion / Summary**

It should be noted that the scope of the data product is huge, even for a prototype. Given the constrain of time and manpower, it is likely that the prototype will not contain all planned features. As such, the project will be delivered in milestones as listed below:

**Milestone 1:**

* Establish data retrieval protocols with the OpenDota API.
* Create the basic UI layout to display the various statistical analysis features

**Milestone 2:**

* Complete all Descriptive Statistics features on the product

**Milestone 3:**

* Complete all Inferential Statistics features on the product

**Milestone 4 (Target):**

* UI polish

**Milestone 5:**

* Research into possible Predictive Statistical features
* Develop Predictive Statistical features on the product

The team will strive to hit as many milestones as possible in the timeframe given for the project. As much as possible, we will try to achieve milestone 4 for this project.