**Statsbomb Test Case**

**Google Sheets for Data Preparing**

Firstly, I load the CSVs to Google Sheets as it’s an easy tool to play around with the data, prepare it for processing, and have an understanding of the provided variables. Also it would be easier to share later. then I sort both sheets, and try Identify potential variables that would determine the probability of scoring a goal by highlighting them.

In shots sheet, I created new field shot\_distance by using both variables location.x and shot.end\_location.x.

For shot\_freeze field, I used lookup to get location of the ball, then calculated distance of player from ball **distance.from.ball.abs**, I’ll then use this newly created variable along with teammate, to decide if an opponent player is near the shot by 1 m. Post that, in shot field, I create “opponents within 1 m” field that counts number of opponents within 1 m range.

Now that I decided on fields that would serve my analysis and created new ones, it’s time to load the newly created dataset into R for further processing and analysis.