We retrieved two CSV files related to soccer. One of our CSV files was from Kaggle and it looked at player performances at the international level prior to the 2018 World Cup. The other CSV retrieved from Footystats detailed out player performances in the English Premier League for the 2018-19 season. Our goal was to see how much players performances on the international stage correlates with their clubs.

For the Data Cleaning stage, we identified an index for the tables. In order to fully complete the transformation, we had to identify and eliminated several columns in both the tables that we identified irrelevant. We also renamed a few columns so the final table would be more concise after merging.

The final database connects Python and Jupyter to SQL. We used SQL as we were using two CSV files which categorize as relational database. Using Pandas we had to confirm that two databases were pre-existing and we inserted data from Pandas to our SQL files. The final Table looks at International performances of players side-by-side their performance in the English Premier League. It helps to pinpoint player performance with club against country and shows how the numbers differ on either side.