To create our imaginary analysis problem we visited Kaggle and data.world for inspiration. We ultimately chose to compare professional player injury and found an API on <https://www.mysportsfeeds.com/>. We found that they had a vast variety of data spread across four professional sports: NBA, NFL, NHL, and MLB.

Once we were able to obtain a valid API key, we found that we were only able to associate it to one sport. At this point we picked NHL and instead of comparing injuries across sports, we decided to compare injuries internally within the sport.

We were able to download a CSV file that contains player injury data from the 2016 NHL playoffs. We also used an API call to bring in JSON formatted data for all NHL players during the 2016 playoffs.

We created a MySQL database with two tables (**injury** and **player\_stats**) and were able to load the data via Jupyter. With the two files loaded we would be able to perform analysis.

Some of our analysis ideas were:

* Comparing total % of players with injuries vs the player pool.
* Comparing teams and look for injury patterns.
* Comparing player age and injuries (and how does that compare to the pool?)

We added our SQL file (hockey\_stats.sql) that builds the schema, tables, and adds a sample join. The sample join looks at injured players and provides some additional information like date of birth and country of origin.