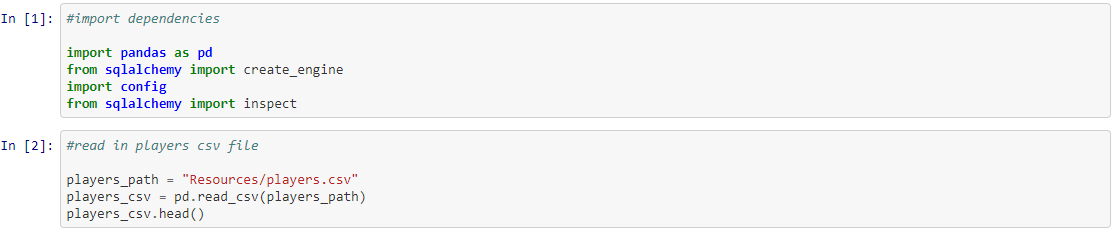
MARY COLASANTO, COLIN O’NEILL, GABRIEL BANG

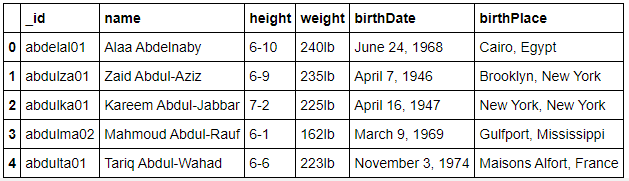
**EXTRACTION:**

We used data.world as a source to extract two raw NBA CSV datasets: [players.csv](https://data.world/datadavis/nba-salaries/workspace/file?filename=players.csv) and [salaries\_1985to2018.csv](https://data.world/datadavis/nba-salaries/workspace/file?filename=salaries_1985to2018.csv).

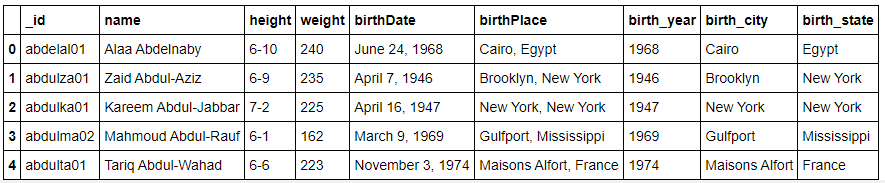
**TRANSFORMATION:**

In Jupyter Notebook, we imported our **pandas** and **sqlalchemy** dependencies. We also created a path to our players.csv file:

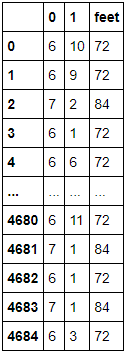
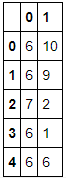
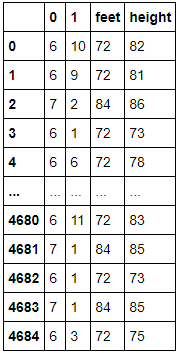
We pulled in the columns from our "demographics" table and consolidated to the columns we needed:



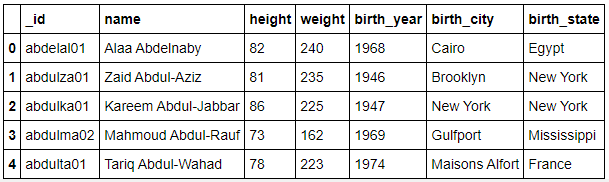
Next, we use the str.split function to remove the ‘lb’ from the weight column, split ‘birthyear’ from date, split birth city from birth state/country, and added them back added both back into the database:



Then, we split height into feet and inches so we could get the total height in inches, converted the height, and multiplied feet by 12 to convert to inches using the str.split function. We then added the inches column to the new ‘feet’ column and saved the total in a new column called height.

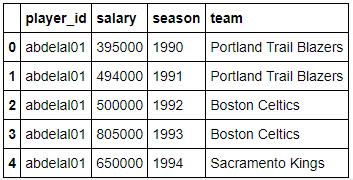
 

With the new height column, we added this back into the database and removed unwanted columns:



We read in our salary table data and created a new dataframe using columns from our salary table. We then renamed the the “season\_start” column to “season” to match our table column name. This process was repeated for the position and draft tables.

Salary:



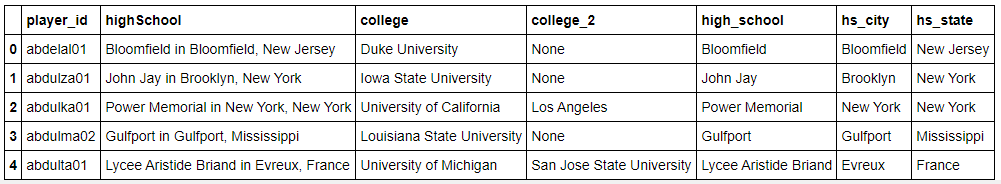
Position:



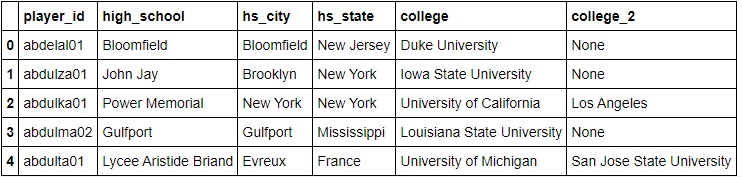
Draft:



For the Schools table, we repeated the process of creating the dataframe using columns from the respective table and renaming the “id” column to “player\_id”. We then split colleges into separate columns for play who enrolled in two schools. We rename each of the columns ‘0’ and ‘1’ respectively. We then split the data in the high school column to extract the high school name to then add it back to the dataframe. We then split the high school city and state/country separately and add it back into the main dataframe.

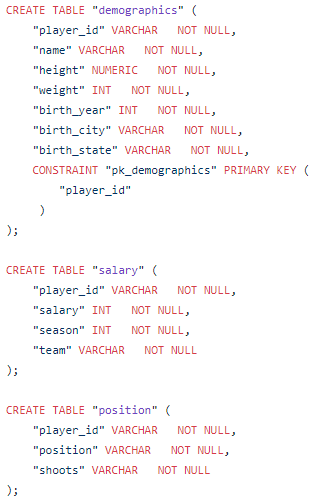
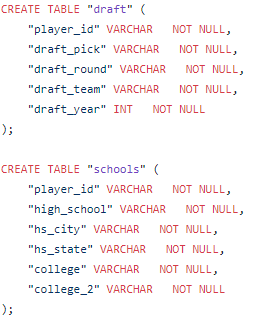


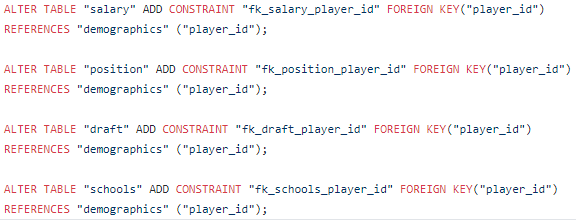
We remake the dataframe without unwanted columns:



**LOAD:**

Using Postgres, we imported our data and created "demographics", "salary", "position", "draft", and "schools" tables to represent our data. Using MySQL we named the PRIMARY KEY CONSTRAINT for each table.



To load our data, we created a local connect to Postgres and verified that the columns are present. The primary key for the demographics is ‘player\_id’ and for the rest of the databases the foreign key is ‘id’. We push these dataframes to a relational database.

