Delroy Myrie

Brad Mendolla

Kendall Hodges

ETL Project Report

August 20, 2019

**Extract**

We decided to make a predictive analysis of NBA players’ salaries from their respective game stats. To do this, we found two datasets from Kaggle that were CSVs already luckily, and from there we just needed to clean both datasets then ultimately merge them into one final database.

* We chose the 2016-2017 season to focus of efforts in making a final relational database which from there anyone could be able to use to make a final analysis and come up with predictive strategies using the database created.

**Transform**

* The cleaning process was reasonably straightforward. In Jupyter Notebook, we took off unnecessary information that would add no benefit to a final predictive analysis of the NBA database. With the 2016-2017 salary data, we used the player name, salary, and team name columns. The NBA Stats CSV was much bigger, so it was necessary to cut the extraneous information, duplicative information, and ultimately, any stat that would hinder the overall analysis between NBA stats and player salaries.
* Additionally, we also needed to make our tables within Postgres and use an ERD to visualize the table and possible dependencies to then import into the final tables in Postgres.

**Load**

In loading the final database, we decided to use an ERD visual tool website to import the file into Postgres then (SQL) to fill the tables created. We choose Postgres as the final destination because of its capabilities in querying the tables as need be, and it is also quite straightforward to navigate efficiently.