

ETL Project

Members: Tyler Adams, Edgardo Bungay, Marcus DeMarici, Vaibhav Vijaykar

Data Sources:

- Kaggle Data Set
 - NBA Twitter Data
 - NBA Salary
 - NBA 2017 Player Stats
 - NBA Real-Plus-Minus
 - NBA Player Impact Estimate
 - NBA Player Attributes

Extract:

- In this portion of the project we loaded all the CSVs that we gathered from Kaggle
- We analyzed the data to identify how to transform the data based on player name and key metrics

Transform

- Renamed all columns for importing into postgresql tables.
- Dropped all NA values
- Reformatted data frame series using split() and replace()
- Created unique values for 'player_name' to use as a primary key
- Exported to CSV

Load:

- Tried using SQL Alchemy to load, ran into issues due to data set discrepancies.
- Used an ERD to map out primary key connections between databases and to generate SQL code to create and alter Tables.

Conclusion:

- The data warehouse we created was through PgAdmin, merging on the player name to create a master list of all the preferred data we garnered.
- Using names as a primary key is