Cameron Kochi
Danny Pham

Project Report

Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

- Our Original data was acquired through kaggle at the following links..
- We chose our data based off the original data set with NBA Draft players. This data set had both NBA teams and college basketball teams. This prompted us too look for data sets that had statistics for college and NBA teams.
- All these data sets were in a CSV format.

Transform: what data cleaning or transformation was required.

- Pandas
  - In the original datasets there were extra columns such as… to cut down the extra columns that had to be made when importing data we dropped those columns using Pandas
  - To remove a lot of the extra colleges that players did not attend, we used inner merges to only keep colleges that both tables have in common
- mySQL
  - To create the Player_Table I copied the Player, College, and NBA columns into a new table
  - To create the Player_Info table I copied the Player, draft_year, pro_height_inches, pro_weight_lb, and date_of_birth into a new table
  - To create the NBA_Teams table…
    - We used SELECT DISTINCT to create a column of each team
    - We used SELECT COUNT(__) and GROUP BY to create the Player Count and Win_Count columns
  - To create the College_Teams table…
    - We used SELECT COUNT(__) and GROUP BY to create the Player Count and Win_Count columns

Load: the final database, tables/collections, and why this was chosen.

- The final database was created and uploaded to a mySQL database using pgAdmin
- This type of database was chosen because of the rigid nature of mySQL and the use of Primary and functional keys
- On top of that, the queries in mySQL allowed us to try different table configurations before committing to a table format

College_Teams

college	player_count	wins
Nevada	8	7
Clemson	3	16
West Virginia	4	15
South Carolina	2	26
Arkansas	6	16

NBA_Teams

team	player_count	win_count
DEN	30	149
IND	31	173
HOU	26	217
CHI	37	160
GSW	30	265

Player_Info

player	draft_year	pro_height_inches	pro_weight_lb	date_of_birth
lebron-james	2003	81	250	1984-12-30
darko-milicic	2003	84	250	1985-06-20
carmelo-anthony	2003	79	238	1984-05-29
chris-bosh	2003	83	235	1984-03-24
dwyane-wade	2003	76	220	1982-01-17

Player_Table

team	player	college
CLE	lebron-james	
DET	darko-milicic	
DEN	carmelo-anthony	Syracuse
TOR	chris-bosh	Georgia Tech
MIA	dwyane-wade	Marquette