ETL Project Ian Arch Daniel Buescher Richard Moseley Mohammed Alshamsi

For this project we wanted to look at PPG statistics from the 2019-2020 NBA season, compared to the current salaries for the top players in the NBA. We extracted the data from 2 web sources and then using pandas transformed the data and finally with SQLAlchemy we loaded the data into a SQLite Database.

**Project Write-Up**

**E**

The two sources of data that our group extracted from were :

<https://www.basketball-reference.com/leagues/NBA_2020_per_game.html#per_game_stats::pts_per_g>

<https://www.spotrac.com/nba/contracts/sort-value/drafted/limit-400/>

Our group also created an excel document for the Naming Conventions of each team for easier linking within our tables.

For our salary site for example, our group took the web address and used pandas to read the tables inside HTML, and then chose our table through the index.

We were able to download and export the data from basketball reference to a CSV file.

**T**

For the transformation part of our project, We are first scrubbing and normalizing our data using Pandas so that our naming conventions are uniform across both datasets.

We used pandas to convert excel workbooks into CSV files and then taking those CSV files into data frames.

We used to clean up and reorganize the structure of our data.

**L**

The database we chose to load our data into is Sqlite. We used SqlAlchemy to reference our data frames and compile all the data into one place.

After the database was established, we compiled all our data into one executable file (Main.py)

