NBA ETL Report

1. Sources of data

* We extracted the CSV files from Kaggle here: <https://www.kaggle.com/drgilermo/nba-players-stats>

2. Extract:

* We used the pd.read\_csv fuction in Pandas to load the CSV Files

3: Transformation:

* We removed “NaN” rows on both dataframes
* Deleted certain columns on “Season\_stats” file because they were deemed unnecessary, such as obscure NBA statistics.
* Renamed percentage columns, such as “3Pt%” to “3Pt\_percent” so the file could correctly be loaded on to the SQL table
* Created College, Team and Position dataframes with unique values for SQL table

4. Load:

* Creation of Database schema based on preview of column names from each of the two CSV Files
* Pandas.to\_Sql feature
* Sqlalchemy to connect to Postgres database
* Relational database was chosen to maintain data integrity