National Basketball Association Team and Player Analytics

Group 7

Dana Goodenough, Dieneba Diaby, Ken Tseng, Jaeik Park

* Data sources:
  + NBA player salary data (2017 through 2018): <https://www.kaggle.com/koki25ando/salary>
  + NBA players stats (2017 through 2018):<https://www.kaggle.com/mcamli/nba17-18?select=nba_extra.csv>
  + NBA season records from every year: https://www.kaggle.com/boonpalipatana/nba-season-records-from-every-year
* ETL
  + Extraction
    - Use pandas to extract data from CSV files and create data frames
  + Transformation
    - Drop null values to avoid errors
    - Split text column (Player name)
    - Filter team record data for 2017-2018 season
    - Copy only necessary columns
    - Remove duplicate data by “groupby” function
    - Rename the column headers
  + Loading
    - Used pandas and sqlalchemy to upload cleaned data
    - Created a database named Project2 in Pgadmin4 and created the schema for the tables of our data: nba\_stats,nba\_salary and team\_stats. The nba\_stats and nba\_salary tables are connected by the player\_name.
    - Created engine to connect to the database by setting up a connection to the local postgres using pandas
    - Connected our final data to sql by uploading the DataFrames to sql through the engine
    - Finalized everything by closing the engine