# Project Report (Kaiser Haag, Rohan Puthanangady, Jack Salva)

Extract

* DS 1) Madden 20 NFL Ultimate Team
  + Description:
    - All player data from the video game Madden NFL 20, in the Ultimate Team game mode.
  + Format:
    - CSV
  + Source:
    - Kaggle
* DS 2) NFL Football Players Stats
  + Description:
    - Real player data scraped from the website profootballreference.com.
  + Format:
    - CSV
  + Source:
    - Kaggle

Transform

* Join
  + Joined DS1)/DS2) on the column ‘ID’
* Clean
  + DS 2) Split ‘Player ID’ on “/”
  + DS1)/DS2) Renamed column headers
  + DS2) String.split command to split dataframe to create ID field
  + DS1)/DS2) Renamed column names
* Filter
  + Filtered DS1) on a certain number of fields to limit the data that was read in the jupyter notebook file

Load (Postgres)

* Final Databases
  + Madden
* Tables
  + pro\_football\_reference
  + player\_info
* Why we chose Postgres
  + Both our data sources were excel files which were already structured like columns in a table. The datasets were both easily capable of being joined as well.