**ETL Project**

**Topic:** 2017 NBA Season Players Statistics

**Date:** 12/12/2021

**Author:** Darrenmond Chao

**Objective**

The objective of this report is to show a detailed analysis of the extract, transform, and load (ETL) process using the 2017 NBA Season player data.

**ETL Process**

**I. Extraction**

The dataset was provided and downloaded from [this](https://github.com/noahgift/socialpowernba/tree/master/data) Github repository. The file used for extraction was originally formatted as .csv.

* [nba\_2017\_nba\_players\_with\_salary.csv](https://github.com/noahgift/socialpowernba/blob/master/data/nba_2017_nba_players_with_salary.csv)

The .csv was then read into a Pandas data frame in a Jupyter notebook using the AWS SageMaker service.



**II. Transformation**

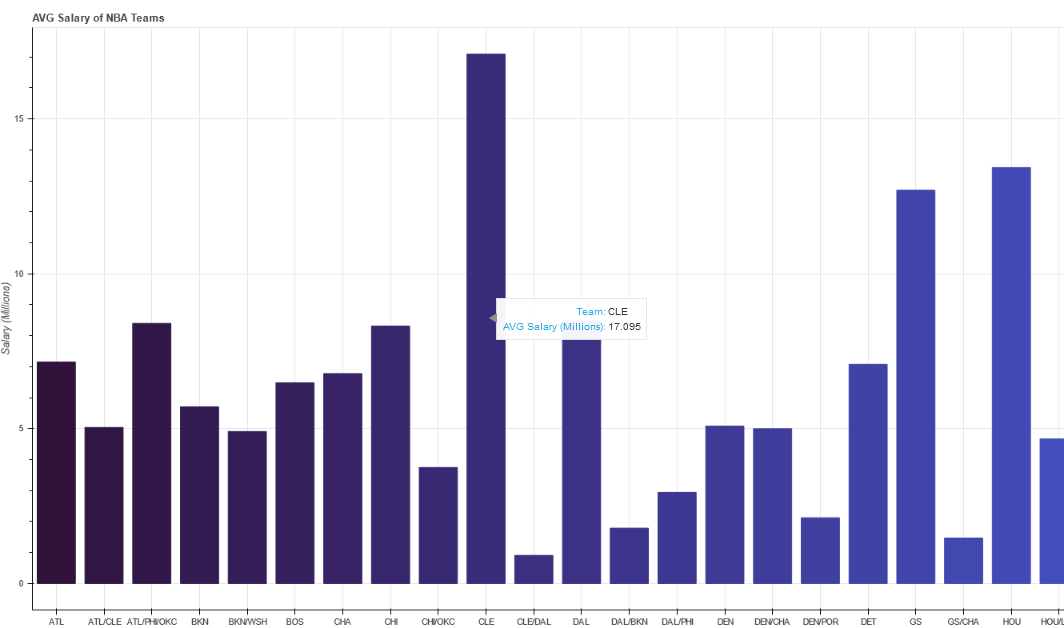
A total of five transformations were performed on the dataset using the Pandas and Bokeh Python libraries. A quick overview of the transformations done is listed below.

* Show the 3-Point average to wins correlation
* NBA player age to shooting percentage correlation
* NBA rank to salary correlation
* NBA players total wins for the 2017 season
* Show a teams average salary

After the transformations, the data is now visualized using HTML. The files can be found in the project Github repository. Samples of the visualizations will be shown below.

* <https://github.com/tsundarren/ETL-Project>

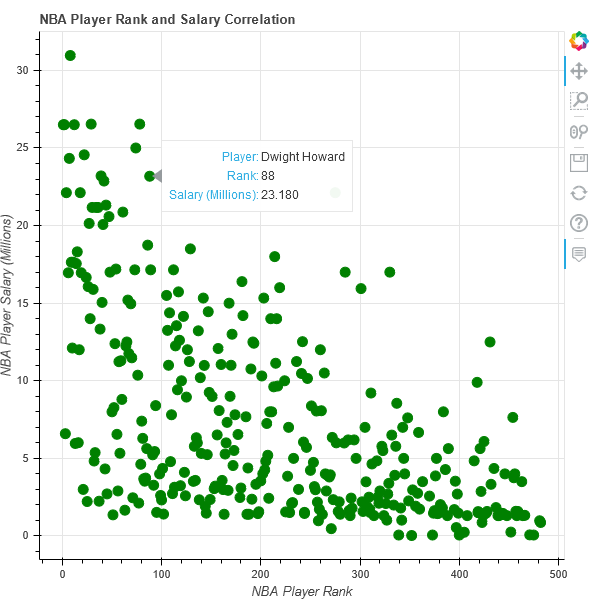
**The average salary of NBA teams histogram:**



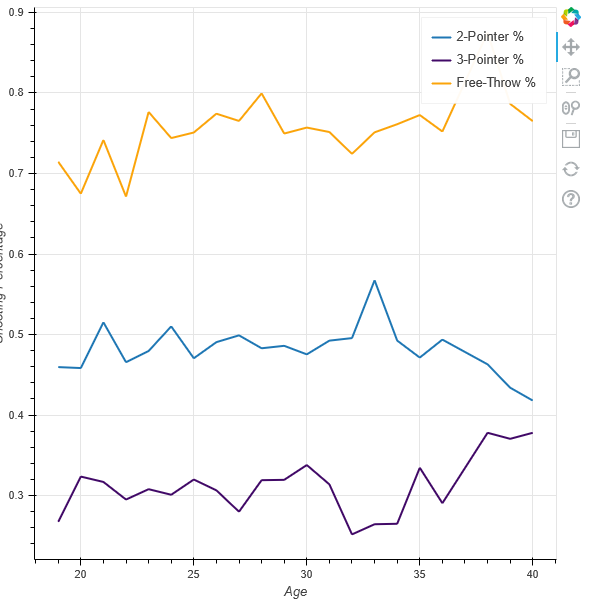
**NBA Players total wins bar chart:**



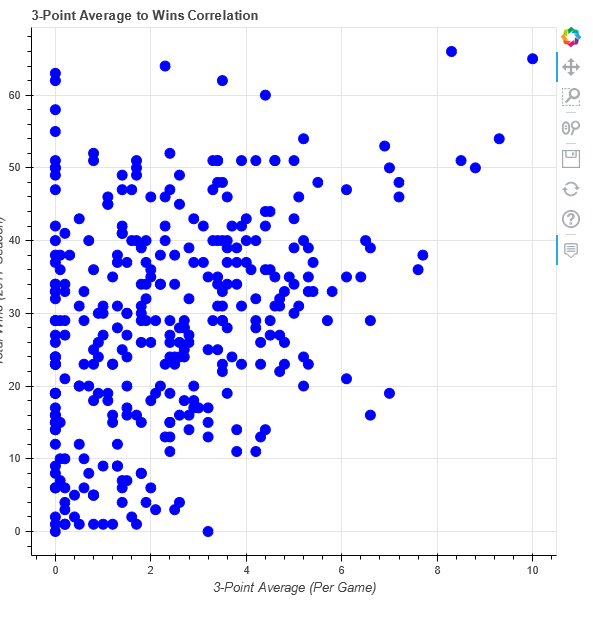
**NBA player rank and salary scatterplot:**



**NBA player age to shooting % line chart:**



**3-Point average to wins scatterplot:**



III. Load

Our transformed data is delivered via this report extract.