**# Data-Manager**

A python program that uses openpyxl, pandas and numpy to manipulate data

**DESCRIPTION**

Data Manager is a program that uses NBA stats and data from basketball reference website in form of a CSV file This is a python program that mainly utilizes the openpyxl library to manipulate CSV data and output data in HDF5 files. I used openpyxl because it's the best library to perform data analysis on .xlsx files. I decided to use python because it had numpy, pandas, and openpyxl libraries. It is also generally better for data analysis due to its great scalability and automation. While writing this program, I faced some challenges. One of them is less familiarity with python because I mainly use C++ and Java. Other challenges were handling missing data, making the program efficient in terms of memory usage and runtime speed. Onwards, I look to incorporate this code into a website once I learn more about web development and APIs.

**HOW TO USE THIS PROGRAM**

Currently, the program can be used via IDEs(plans to implement an actual API for the program). The user inputs whatever statistic category they are interested in, the specific player, teams, statistical positions and specific season(can currently only accommodate the 2019, 2020, and 2021 NBA seasons. The program will output the requested start either as an HDF5 file or console output depending on what the user prefers.

**LICENSE**

Apache license 2.0