Data Scraping

Brian Fernando   Aarush Maddela   Nixon Tan
Sharona Yang

Wednesday 17th December, 2025

```
import os
from nba_scraper.data_scraping import scrape_and_merge_stats, scrape_salaries
```

1 Scrape player gameplay stats

```
# looping through the range of years (2002 to 2025) and scraping/merging stats for each seas
# 2002 was Lebron's rookie season (oldest player in NBA currently), used to calculate Experi
data_directory = "data/nba_merged_stats"
for year in range(2002, 2026):
    scrape_and_merge_stats(year, data_directory)
```

```
Merged stats saved for 2002: nba_merged_2002.csv
Merged stats saved for 2003: nba_merged_2003.csv
Merged stats saved for 2004: nba_merged_2004.csv
Merged stats saved for 2005: nba_merged_2005.csv
Merged stats saved for 2006: nba_merged_2006.csv
Merged stats saved for 2007: nba_merged_2007.csv
Merged stats saved for 2008: nba_merged_2008.csv
Merged stats saved for 2009: nba_merged_2009.csv
Merged stats saved for 2010: nba_merged_2010.csv
Merged stats saved for 2011: nba_merged_2011.csv
Merged stats saved for 2012: nba_merged_2012.csv
Merged stats saved for 2013: nba_merged_2013.csv
Merged stats saved for 2014: nba_merged_2014.csv
Merged stats saved for 2015: nba_merged_2015.csv
Merged stats saved for 2016: nba_merged_2016.csv
Merged stats saved for 2017: nba_merged_2017.csv
Merged stats saved for 2018: nba_merged_2018.csv
Merged stats saved for 2019: nba_merged_2019.csv
Merged stats saved for 2020: nba_merged_2020.csv
Merged stats saved for 2021: nba_merged_2021.csv
```

```
Merged stats saved for 2022: nba_merged_2022.csv
Merged stats saved for 2023: nba_merged_2023.csv
Merged stats saved for 2024: nba_merged_2024.csv
Merged stats saved for 2025: nba_merged_2025.csv
```

2 Scrape player salaries

```
# player salary data URLs for 2023, 2024, 2025 seasons
salaries_2223_url = "https://web.archive.org/web/20230516051257/https://www.basketball -refe
salaries_2324_url = "https://web.archive.org/web/20240605124914/https://www.basketball -refe
salaries_2425_url = "https://web.archive.org/web/20250305232518/https://www.basketball -refe
salaries_2526_url = "https://web.archive.org/web/20250906094013/https://www.basketball -refe

# scraping salary data for the three seasons and saving them as CSVs
data_directory = "data/nba_player_contracts"
os.makedirs(data_directory, exist_ok=True)
salaries_2223_df = scrape_salaries(salaries_2223_url)
salaries_2223_df.to_csv(data_directory + "/nba_contracts_2023.csv", index=False)
salaries_2324_df = scrape_salaries(salaries_2324_url)
salaries_2324_df.to_csv(data_directory + "/nba_contracts_2024.csv", index=False)
salaries_2425_df = scrape_salaries(salaries_2425_url)
salaries_2425_df.to_csv(data_directory + "/nba_contracts_2025.csv", index=False)
```