## **Project Overview – NBA ETL Pipeline**

**Summary**The NBA ETL Pipeline project builds a foundational data engineering workflow for extracting, transforming, and storing basketball data using the nba\_api Python package. The goal is to establish a modular and reusable system for collecting player statistics, team performance, game logs, and league standings for downstream analysis and modeling.

**Purpose**This project demonstrates practical ETL (Extract, Transform, Load) skills applied in a sports analytics context. The pipeline pulls live data from the NBA Stats API, processes it into structured formats, and stores it in a clean and query-ready form (CSV). This pipeline supports scenario-based analytics, such as analyzing playoff performance, filtering by team or player, or creating season-based dashboards.

**Tools Used**

* **Python** (Jupyter/Colab-ready)
* **nba\_api** for data extraction
* **pandas** for data transformation and manipulation
* **Google Drive** or local directory for storage

**Folder Structure**

* Code & Notebooks/ – Includes Jupyter and .py files for each data source.
* Data Files/ – Contains saved CSVs (e.g., team stats, player stats, game logs).
* Visuals & Reports/ – Reserved for dashboards, charts, or summaries.
* README – User guidance and usage examples.
* Project Overview – This document.

**Dataset Outputs**

* Team per-game stats for the 2023–24 season
* Career stats for selected players (e.g., Jayson Tatum, Jaylen Brown)
* Game logs for the Boston Celtics across multiple seasons
* Eastern Conference standings
* Full team roster metadata

**Future Additions**

* SQL or SQLite loading layer for querying
* Scheduled automated updates
* Player-level per-game stats or shot chart integration

**Author**Maekala Turner  
GitHub: [github.com/maekala](https://github.com/maekala)