**README – NBA ETL Pipeline**

**Description**This project is a modular Python pipeline designed to extract, clean, and store NBA data from the league’s public API. Using the nba\_api package, the pipeline enables season-by-season and player-specific data collection to power analytics projects, dashboards, and simulations.

This project was created as part of Maekala Turner's personal data portfolio and is structured for reuse and extension in future analytics and engineering workflows.

**Setup Instructions**

1. Clone or download the repository or open the .ipynb notebook in Jupyter or Google Colab.
2. Install nba\_api if needed:

| !pip install nba\_api |
| --- |

1. Mount your Google Drive or set a local path for storing output CSVs.

**Key Modules in This Pipeline**

* Team-level stats (per season)
* Full career player stats (per season)
* Game logs by team and season
* League standings
* Roster metadata
* Season phase tagging (preseason, regular season, playoffs)

**How to Use**

*Example 1: Extract Donovan Mitchell’s playoff performances from his seasons with the Cleveland Cavaliers*



| from nba\_api.stats.static import players from nba\_api.stats.endpoints import playergamelog  # Find player ID mitchell\_id = players.find\_players\_by\_full\_name("Donovan Mitchell")[0]['id']  # Pull playoff game logs gamelog = playergamelog.PlayerGameLog(player\_id=mitchell\_id, season\_type\_all\_star='Playoffs') gamelog\_df = gamelog.get\_data\_frames()[0]  # Filter for Cleveland Cavaliers mitchell\_cavs\_playoffs = gamelog\_df[gamelog\_df['TEAM\_ABBREVIATION'] == 'CLE'] |
| --- |

*Example 2: Get play-in game logs from the 2022 season for all teams*

| from nba\_api.stats.endpoints import leaguegamefinder  # Get all games and filter for 2022 Play-In (prefix '3' = Play-In, year 2022 = '32022') gamefinder = leaguegamefinder.LeagueGameFinder() games\_df = gamefinder.get\_data\_frames()[0]  # Clean column names games\_df.columns = games\_df.columns.str.lower().str.replace(' ', '\_')  # Filter play-in games from 2022 playin\_2022 = games\_df[games\_df['season\_id'] == '32022'] |
| --- |

*Example 3: Retrieve all regular season games where the Celtics scored more than 120 points during the 2023–24 season*

| from nba\_api.stats.endpoints import leaguegamefinder  celtics\_id = 1610612738 gamefinder = leaguegamefinder.LeagueGameFinder(team\_id\_nullable=celtics\_id) games\_df = gamefinder.get\_data\_frames()[0] games\_df.columns = games\_df.columns.str.lower().str.replace(' ', '\_')  # Filter for 2023-24 regular season (season\_id '22024') and PTS > 120 high\_scoring\_wins = games\_df[(games\_df['season\_id'] == '22024') & (games\_df['pts'] > 120)] |
| --- |