**NFL Combine Data Pipeline & Dashboard**

**Project Overview**

This project extracts, stores, and visualizes NFL Combine data from Pro-Football-Reference. It automates data collection, loads it into an SQLite database, and provides interactive insights using Streamlit and Power BI.

**Features**

* **Automated Data Scraping**: Extracts NFL Combine data from Pro-Football-Reference.
* **Database Storage**: Saves structured data in SQLite for efficient querying.
* **Streamlit Dashboard**: Provides an interactive web app to analyze player performance.
* **Power BI Visualization**: Offers advanced analytics and trends visualization.
* **Automated Data Pipeline**: Regularly updates the database and dashboards with new data.

**Technologies Used**

* **Python** (pandas, sqlite3, requests, Streamlit)
* **SQLite** (DB Browser for SQLite)
* **Power BI** (for advanced visualizations)
* **pandas.read\_html()** (for web scraping data)

**Installation & Setup**

**Prerequisites**

Ensure you have the following installed:

* Python 3.x
* SQLite (or DB Browser for SQLite)
* Power BI Desktop
* Required Python libraries:
* pip install pandas sqlite3 streamlit certifi

**Running the Project**

1. **Run Data Extraction & Storage**:
2. python scrape\_combine.py
   * This will scrape the latest NFL Combine data and store it in nfl\_combine.db.
3. **Launch the Streamlit Dashboard**:
4. streamlit run combine\_dashboard.py
5. **Load Data into Power BI**:
   * Open Power BI Desktop.
   * Connect to nfl\_combine.db as a data source.
   * Create visualizations and insights.

**Database Schema**

**combine\_results Table**

| **Column** | **Type** | **Description** |
| --- | --- | --- |
| player | TEXT | Name of the athlete |
| pos | TEXT | Player position |
| college | TEXT | College attended |
| height | TEXT | Height measurement |
| weight | INTEGER | Weight in pounds |
| forty\_yd | FLOAT | 40-yard dash time (seconds) |
| bench | INTEGER | Bench press reps |
| vertical | FLOAT | Vertical jump (inches) |
| broad\_jump | FLOAT | Broad jump distance (inches) |
| three\_cone | FLOAT | 3-cone drill time (seconds) |
| shuttle | FLOAT | Shuttle run time (seconds) |

**Querying the Data**

Example SQL queries to analyze data:

* **Top 10 Fastest Players**:
* SELECT player, pos, forty\_yd FROM combine\_results ORDER BY forty\_yd ASC LIMIT 10;
* **Strongest Players (Bench Press)**:
* SELECT player, pos, bench FROM combine\_results ORDER BY bench DESC LIMIT 10;
* **Average Performance by Position**:
* SELECT pos, AVG(forty\_yd) AS avg\_forty, AVG(bench) AS avg\_bench FROM combine\_results GROUP BY pos;

**Future Improvements**

* Automate daily data updates using a scheduler.
* Enhance Power BI dashboards with more metrics.
* Deploy the Streamlit app for public access.

**Author**

**Rodney Littlejohn**

* GitHub: github.com/rlitt579